Appendix A
Test Items for Chapter 3 and Solutions

Ray School, Chicago, 1986

Name ____ Date ______________

Science Exam

Please answer the following statements with a 'T' if it is true or an 'F' if it is false

1. ______ All living things are made of protoplasm.
2. ______ Cells make more of themselves through cell division.
3. ______ There are 40 pairs of chromosomes in the common cell.
4. ______ Plants and plant cells are not alive.
5. ______ Groups of the same types of cells form together to make tissue.

---

6. List four characteristics that living things have that make them living.

   1. ______________________________
   2. ______________________________
   3. ______________________________
   4. ______________________________

Please circle the best answer to each of the following questions.

7. Chromosomes are found

   A. In the cytoplasm

   B. In the nucleus

   C. In the plasma membrane.

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8

8. After cell division, each of the cells is

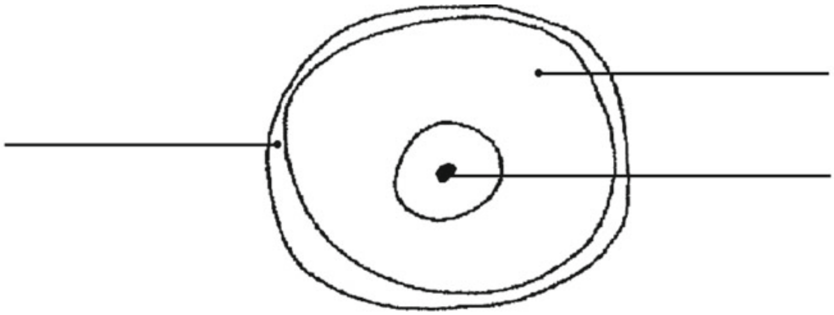
   A. 1/4 the size of the original cell.

   B. Twice the size of the original cell.

   C. 1/2 the size of the original cell.

---

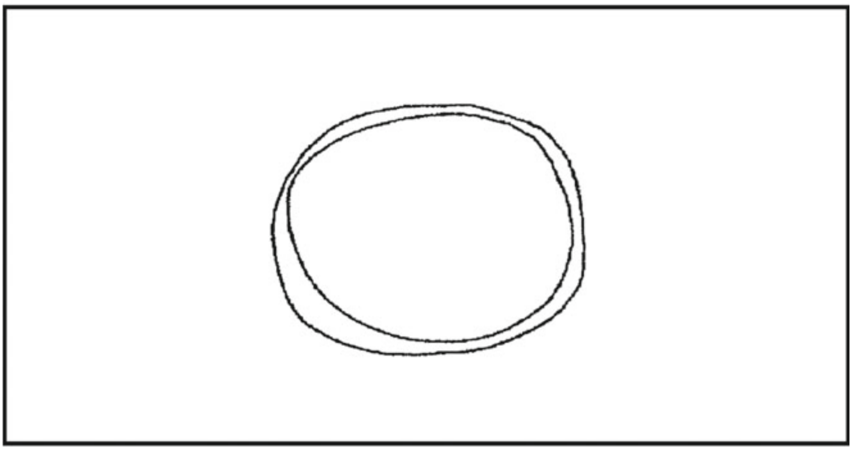
9. Please select and label the three major parts of a cell from the words listed



   Nucleus cytoplasm plasma membrane protoplasm

10. Extra credit

   Draw and label cell division taking place. Identify: chromosomes, asters, spindle, nuclei.



Name ____ Date _____________

Science Exam Solutions

Please answer the following statements with a 'T' if it is true or an

'F' if it is false

1. ___**T**___ All living things are made of protoplasm.

2. ___**T**___ Cells make more of themselves through cell division.

3. ___**F**___ There are 40 pairs of chromosomes in the common cell.

4. ___**F**___ Plants and plant cells are not alive.

5. ___**T**___ Groups of the same types of cells form together to make tissue.

- -

6. List four characteristics that living things have that make them living.

   1. **They breathe**
   2. **Respond to stimuli**
   3. **Eat**
   4. **Give off waste**

- -

Please circle the best answer to each of the following questions.

7. Chromosomes are found

   A. In the cytoplasm
   B. In the nucleus
   C. In the plasma membrane.

8. After cell division, each of the cells is

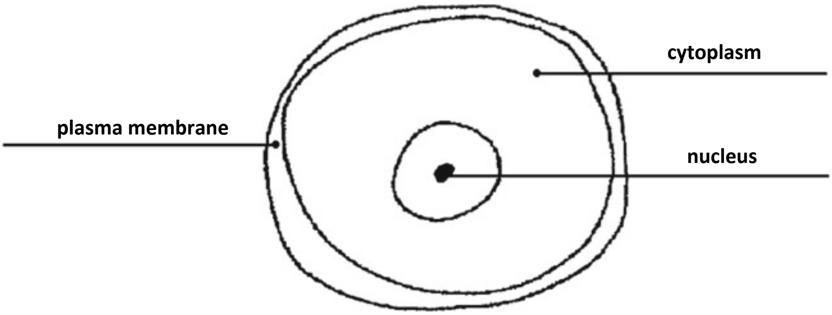
   A. 1/4 the size of the original cell.

   B. Twice the size of the original cell.

   C. 1/2 the size of the original cell.

---

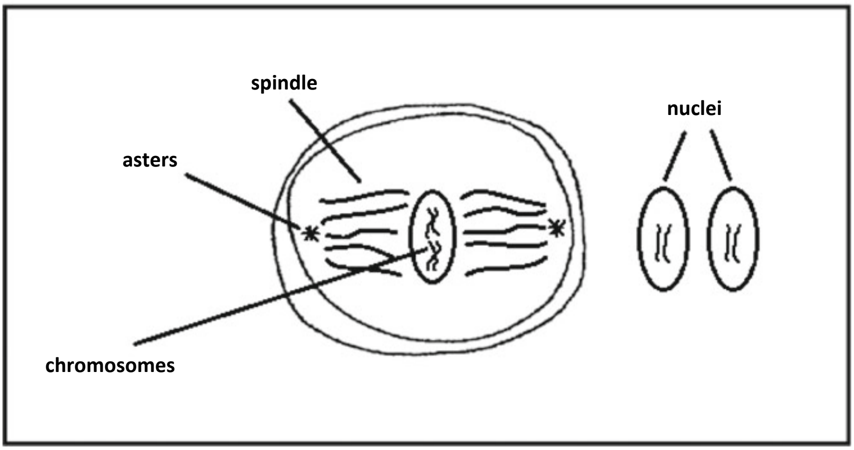
9. Please select and label the three major parts of a cell from the words listed



Nucleus cytoplasm plasma membrane protoplasm

10. Extra credit

    Draw and label cell division taking place. Identify: chromosomes, asters, spindle, nuclei.



Appendix B
Chapter Exercises Solutions

B.1 Chapter Exercises Solutions

Chapter 1

a. Ratio,
b. Nominal,
c. Ordinal,
d. Ordinal,
e. Interval.

Chapter 3

1. Item 7: facility = 38, discrimination = 0.72
   Item 8: facility = 26, discrimination = 0.35
   Item 8 is more difficult and Item 7 discriminates more,
2. 11.24,
3. Mean = 6.68, SD = 4.06, SE = 1.82,
4. 8.26–14.23,
5. 13.18.

Chapter 4

1.

s_1^2	s_2^2	s_3^2	s_4^2	s_5^2	s_6^2	s_7^2	s_8^2	s_x^2
0.11	0.26	0.19	0.24	0.24	0.58	0.61	3.84	16.48

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8

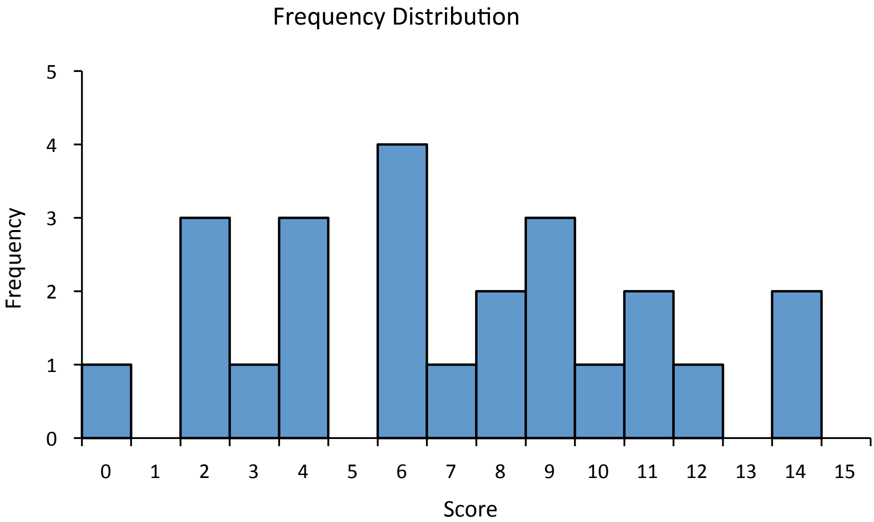
2. 0.72.
3. Highly acceptable for a teacher-made test. If a student's score is interpreted with other information about the student, then it might be very useful.
4. For example, a Year 12 external examination in mathematics used for university entry in Australia. This examination is based on a very tightly controlled syllabus understood by teachers, and therefore would show excellent content validity. It should have reasonable predictive validity because students would need to have the knowledge of the content to proceed to studies in the same area at university level. If item level data are used to calculate coefficient α, the expected reliability would be at least 0.8.

Chapter 5

1.

	Items	1	3	4	5	2	6	7	8	R	F	CF	
Person	Maximum score	1	1	1	1	1	2	2	6				
7		0	0	0	0	0	0	0	0	0	2	2	Lower group 9
18		0	0	0	0	0	0	0	0	0			
15		1	0	0	0	0	1	0	0	2	3	5	
19		1	0	0	1	0	0	0	0	2			
23		1	1	0	0	0	0	0	0	2			
8		1	1	0	0	1	0	0	0	3	1	6	
2		1	0	0	0	1	0	0	2	4	3	9	
14		1	1	1	1	0	0	0	0	4			
20		1	1	0	1	0	1	0	0	4			
11		1	1	1	1	0	1	1	0	6	4	13	Middle group 7
12		0	1	0	0	1	1	1	2	6			
24		1	1	1	1	0	1	0	1	6			
25		1	1	1	1	0	1	0	1	6			
5		1	1	1	1	0	1	1	1	7	1	14	
1		1	1	1	1	1	1	1	1	8	2	16	
17		1	1	1	0	1	2	2	0	8			
9		1	1	1	1	1	2	2	0	9	3	19	Upper group 9
10		1	1	1	1	1	1	2	1	9			
21		1	1	1	1	1	1	1	2	9			
6		1	0	1	1	0	1	1	5	10	1	20	
4		1	1	1	1	0	2	1	4	11	2	22	
22		1	1	1	1	1	2	1	3	11			
13		1	1	1	0	1	2	1	5	12	1	23	
3		1	1	1	1	1	2	2	5	14	2	25	
16		1	1	1	1	0	2	2	6	14			
Total Score		22	19	16	16	11	25	19	39				
Items		1	3	4	5	2	6	7	8				
% maximum		88	76	64	64	44	50	38	26				

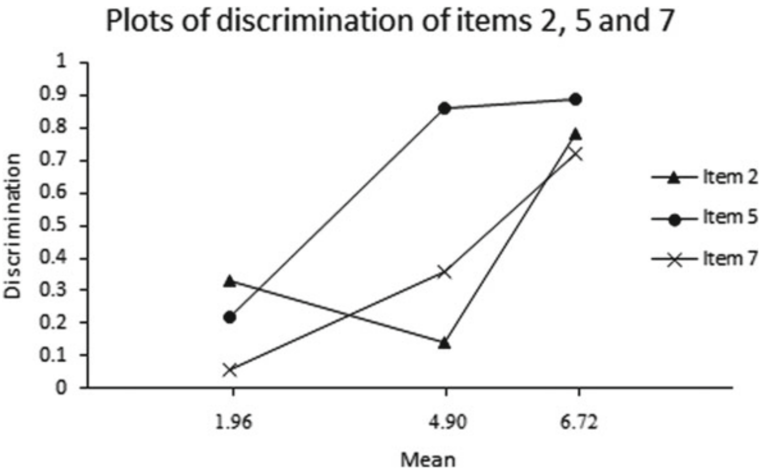
2.



3.

	Items	1	3	4	5	6	2	7	8	R	R*	
Person	Maximum score	1	1	1	1	2	1	2	6			
7		0	0	0	0	0.00	0	0.00	0.00	0.00	0.00	Lower group 9
18		0	0	0	0	0.00	0	0.00	0.00	0.00	0.00	
15		1	0	0	0	0.50	0	0.00	0.00	2.00	1.50	
19		1	0	0	1	0.00	0	0.00	0.00	2.00	2.00	
23		1	1	0	0	0.00	0	0.00	0.00	2.00	2.00	
2		1	0	0	0	0.00	1	0.00	0.33	4.00	2.33	
8		1	1	0	0	0.00	1	0.00	0.00	3.00	3.00	
12		0	1	0	0	0.50	1	0.50	0.33	6.00	3.33	
20		1	1	0	1	0.50	0	0.00	0.00	4.00	3.50	
14		1	1	1	1	0.00	0	0.00	0.00	4.00	4.00	Middle group 7
24		1	1	1	1	0.50	0	0.00	0.17	6.00	4.67	
25		1	1	1	1	0.50	0	0.00	0.17	6.00	4.67	
6		1	0	1	1	0.50	0	0.50	0.83	10.00	4.83	
11		1	1	1	1	0.50	0	0.50	0.00	6.00	5.00	
5		1	1	1	1	0.50	0	0.50	0.17	7.00	5.17	
17		1	1	1	0	1.00	1	1.00	0.00	8.00	6.00	
1		1	1	1	1	0.50	1	0.50	0.17	8.00	6.17	Upper group 9
4		1	1	1	1	1.00	0	0.50	0.67	11.00	6.17	
13		1	1	1	0	1.00	1	0.50	0.83	12.00	6.33	
21		1	1	1	1	0.50	1	0.50	0.33	9.00	6.33	
10		1	1	1	1	0.50	1	1.00	0.17	9.00	6.67	
9		1	1	1	1	1.00	1	1.00	0.00	9.00	7.00	
16		1	1	1	1	1.00	0	1.00	1.00	14.00	7.00	
22		1	1	1	1	1.00	1	0.50	0.50	11.00	7.00	
3		1	1	1	1	1.00	1	1.00	0.83	14.00	7.83	
Total Score		22	19	16	16	12.50	11	9.50	6.50			
Items		1	3	4	5	6	2	7	8			
% maximum		88	76	64	64	50	44	38	26			

4. DI (Item 2) = 0.45, DI (Item 5) = 0.67, DI (Item 7) = 0.66.
5.



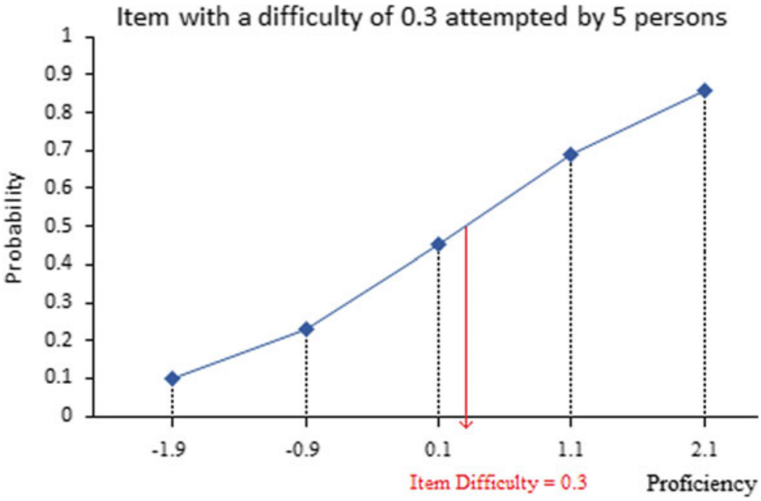
6. Item 7 is the best discriminating item.
   Item 5 exhibits a ceiling effect, which means the item does not discriminate well between the middle and upper groups.
   Item 2 is the worst discriminating item. The middle group has the lowest proportion correct score out of the three proficiency groups, a pattern which is very inconsistent with expectations.

Chapter 6

1. (a) True score (T_v)
   (b) The person parameter is the value on a continuum that governs whether or not the person responds positively or negatively to an item. Therefore, it explains the performance on the item which assesses a particular variable. The greater the proficiency relative to the item's difficulty, the greater the probability of success.
   (c) (i) Only the proficiency of persons, and no other person characteristics, affects their position on the latent trait.
   (ii) Only the difficulty of the items, and no other item characteristics, affects their position on the latent trait.
   (iii) Persons are ordered according to proficiency and items are ordered according to difficulty on the same, single continuum.

2. (a) $\Pr\{x_{n1} = 1\} = 0.90; \Pr\{x_{n2} = 1\} = 0.50; \Pr\{x_{n3} = 1\} = 0.31$
   (b) This 'probability' is a theoretical proportion of the number of times that a person with a fixed proficiency would answer correctly many items of exactly the same difficulty.

3. (a)
   $\Pr\{x_{1i} = 1\} = 0.10; \Pr\{x_{2i} = 1\} = 0.23; \Pr\{x_{3i} = 1\} = 0.45; \Pr\{x_{4i} = 1\} = 0.69;$
   $\Pr\{x_{5i} = 1\} = 0.86.$
   (b)



Chapter 7

1. Item 2,
2. $\Pr\{x_{n1} = 1\} = 0.62; \Pr\{x_{n2} = 1\} = 0.38,$
3. 0.73,
4. 0.73,
5. The probability is the same. It is independent of the person's proficiency and depends only on the relative difficulties of the items.

Chapter 8

The Rasch model formalizes explicitly that the total score is a sufficient statistic for the person parameter, but only if the responses fit the model. If the responses deviate from the pattern required by the model beyond the error implied by the model, then that is evidence that the response pattern cannot really be summarized by the total score and that the pattern of responses should possibly be examined for some systematic, that is, non-random effects.

Chapter 9

1. Item 1 = −1.24, Item 2 = 1.24,
2. Item 1 = −2.48, Item 2 = 2.48,
3. Item 1 = 7.52, Item 2 = 12.48.

Chapter 10

1. (a) 5.92, greater than 0.45,
   (b) 6.08, less than 0.55,
   (c) 6.00, 0.50 is the best estimate.

2. 0.79.

Chapter 19

1. A paradigm involves the taken-for-granted assumptions in carrying out research or applications of some area of the field. These assumptions are generally not made explicit unless they are challenged in some way.
2. The usually unstated assumption in the IRT paradigm is that the model chosen should fit the data. Therefore, if a simple model does not fit the data, a more complex model with more parameters is tried.
   The assumption in the RMT paradigm is that the Rasch model provides a criterion for invariance and therefore measurement. In this paradigm, the data should not only be valid on all other relevant characteristics, but in addition, it should fit the Rasch model. If the data do not fit the relevant model, then the task is not to find a model with more parameters that fit the data, but to investigate the data and improve the instrument.

Chapter 27

(a) −0.50, −0.10, 0.60,
(b)

y_{niP}	y_{niC}	y_{niD}	x		
0.378	**0.475**	**0.646**	**0**		
0.622	**0.475**	**0.646**	**1**		
0.622	**0.525**	**0.646**	**2**	Ω^G	
0.622	**0.525**	**0.354**	**3**		
0.378	0.525	0.646	1		Ω
0.378	0.475	0.354	1		
0.622	0.475	0.354	2		
0.378	0.525	0.354	2		

(c)

$x = 0$	Pr{(0, 0, 0)}	0.116
$x = 1$	Pr{(1, 0, 0)}	0.191
$x = 2$	Pr{(1, 1, 0)}	0.211
$x = 3$	Pr{(1, 1, 1)}	0.116

(d)

$x = 0$	0.183
$x = 1$	0.301
$x = 2$	0.333
$x = 3$	0.183

(e)

$x = 1$	0.622
$x = 2$	0.525
$x = 3$	0.354

(f) $y_{niP} = 1, y_{niC} = 1, y_{niD} = 1.$

Chapter 28

Sufficient statistic for item and person parameters—PRM has sufficient statistics, non-Rasch models do not.

Partitioning in forming categories—GRM partitions the frequency distribution of responses to generate adjacent categories, PRM partitions the latent continuum to create adjacent categories.

Appendix C
RUMM2030 Exercises and Solutions

Data sets A, B, C, D, E, F and G can be located at the book's webpage [https://doi.org/10.1007/978-981-13-7496-8]

Exercise 1: Interpretation of RUMM2030 Printout

This exercise essentially involves being able to interpret the results from RUMM2030. The example analysed according to the program is the one that you analysed by hand in earlier exercises (starting in Chap. 3), so the example is familiar to you. The test items are in Appendix A.

The exercise involves three distinct tasks given below:

Part A: identifying the relevant statistical information displayed in the computer printout and reporting this information. This is essentially a descriptive task.
Part B: carrying out some calculations based on the information available from the computer printout. The task should consolidate some of the key aspects of Rasch measurement theory.
Part C: interpreting the statistics that are presented in the analysis.

RUMM2030 analysis is a very general program and it has features that have not been included in the output here. This output is edited to include only the material which has been covered. Nevertheless, you are advised to look through the whole output which follows.

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8

Part A: Identifying the relevant statistical information displayed in the computer printout.

Summary of the computer output is as follows:

1. Items

   1.1 Number of items ________,
   1.2 Number of categories per item ______,

2. Sample persons

   2.1 Number of persons entered in the data analysis __________,
   2.2 Number of persons eliminated for various reasons __________,
   2.3 Number of persons retained for analysis __________,

3. Item and proficiency parameter estimates.

   Complete the following table for items and proficiencies:

Items			Person Scores		
Item Number	Difficulty $\hat{\delta}_i$	Standard Error $\hat{\sigma}_i$	Raw Score	Proficiency Estimate	Estimated Standard Error
1			2		
4			7		
8			13		

4. Person parameters estimates including fit statistics

   Complete the following table for the persons indicated below.
   To report the observed response pattern, reorder the items by their difficulty.
   Note, we are referring to the ID, the original number allocated to the person, and be careful to note this. When persons have been eliminated, the original ID number can be different from the PERSON NUMBER.

Person ID	Total score	Proficiency estimate	Fit statistic	*Observed response pattern (order items by difficulty)
3				, , , , , , ,
12				, , , , , , ,
19				, , , , , , ,

*To complete this, you will need to look at the 'Guttman Pattern' section in the computer output.

5. Item parameter estimates including fit statistics.

Complete the following table for selected items:

Item number	Location estimate $\hat{\delta}_i$	Fit chi-square	Probability
1			
4			
7			

6. Item characteristic curves

6.1 Complete the following table for item 1:

Item 1 Location $\hat{\delta}_i$ =_______	*Class interval* 1	 2	 3
Mean proficiency			
Observed mean (OM)			
Estimated mean or expected value (EV)			

6.2 Complete the following table for item 4:

Item 4 Location $\hat{\delta}_i$ =_______	*Class interval* 1	 2	 3
Mean proficiency			
Observed mean (OM)			
Estimated mean or expected value (EV)			

6.3 Complete the following table for item 7:

Item 7 Location $\hat{\delta}_i$ =_______	*Class interval* 1	 2	 3
Mean proficiency			
Observed mean (OM)			
Estimated mean or expected value (EV)			

7. Traditional reliability and separation indices

   7.1 Report the traditional coefficient α reliability together with the relevant components for calculating this reliability.
   Coefficient α _________
   Variance (total score) _________
   Variance (items) _________

   7.2 Report the index of person separation r_β together with the relevant components for calculating this index.
   Index r_β _________
   Variance of $\hat{\beta}$ _________
   Average error variance _________

Part B: Calculations based on the information available from the computer printout.

8. Item characteristic curves

   8.1 Form axes for plotting item characteristic curves for dichotomous items. The horizontal axis refers to proficiency, the vertical to the probability that any item will be answered correctly.

   8.2 On these axes, plot the item characteristic curve for items 1 and 4.
   Use information in questions 6.1 and 6.2 to plot the expected values (probabilities) as a function of the proficiency for each of the three class intervals, and then join these points by a smooth curve.
   Use different symbols for items 1 and 4, say a small circle for item 1 and a small square for item 4.

   8.3 On the same graph, plot the observed means in the class intervals for each item. Again use different symbols. Say a '+' sign for item 1 and a small filled circle for item 4.

   8.4 Form axes for plotting item characteristic curves for an item with three categories, that is, possible scores of 0, 1 and 2. The horizontal axis refers to proficiency and the vertical to the expected value according to the model.

   8.5 On these axes, plot the item characteristic curve for item 7.
   Use information in question 6.3 to plot the expected values as a function of the proficiency for each of the three class intervals, and then join these points by a smooth curve.

   8.6 On the same graph, plot the observed means in the class intervals for item 7 using a '+' sign.

9. Display of person proficiencies

   9.1 Form axes for plotting the relationship between a total score and the proficiency estimate. The horizontal axis should be the proficiency axis and the vertical axis the raw score.
   9.2 On these axes, plot the actual proficiency estimates against the raw scores.

10. Calculation of reliability indices
    In question 7 above you reported the values of coefficient α and of the separation index r_β.

    10.1 Verify the value of coefficient α by showing how it is calculated from the relevant information in question 7.1.
    10.2 Verify the value of the separation index r_β by showing how it is calculated from the relevant information in question 7.2.

Part C: Interpreting the analysis

11. Two persons are eliminated from the analysis. Why were these persons eliminated? (Look at the 'Individual Person fit' section and note the two persons marked 'Extm').
12. One of the curves in your graph of question 8.2 is to the right of the other,

    12.1 What relative property of two items puts one to the right of the other?
    12.2 What information in the raw data reflects this relative property for these two items?

13. Your graph of question 8.2 shows observed as well as theoretical values,

    13.1 Which of the two items in your graph of question 8.2 shows the worse actual discrimination?
    13.2 Why do you come to this conclusion?

14. In question 5 you also reported the statistical index of fit for item 7, and in question 8.5 you plotted the expected and observed values in three class intervals for this item.

    14.1 From the graphical display, does the item discriminate more than expected, worse than expected, or about as expected? Explain your answer.
    14.2 Is the difference between the theoretical discrimination and the actual discrimination statistically significant at the 5% level? Explain your answer.

15. In your graph of question 9, you had proficiency estimates for scores ranging from 1 to 14, even though the maximum possible score was 15, and the minimum was 0.

    15.1 What is the theoretical estimate of proficiency for someone with a score of 15?

    ____

    15.2 What is the theoretical estimate of proficiency for someone with a score of 0?

    ____

16. In question 4 you listed information about persons with ID number 3, 12 and 19.

    16.1 According to the fit index, which of these people has a response pattern that is closest to that expected from the model and the furthest from that expected from the model?
    Closest:_________ Furthest ___________

    16.2 What aspect of the response pattern that you recorded in question 4 is consistent with your conclusion from the statistical information in 16.1?

17. Overall summary statements

    17.1 You reported the traditional reliability indices for this test. Comment as to whether these values are reasonably good, bad or moderate considering the length of the test.

    17.2 On the computer printout (Individual Person-fit), the fit-residual for every person is shown. This statistic is theoretically a standard normal deviate, that is, with a mean of 0.0 and a standard deviation of 1.0. Explain why the values obtained suggest either that on the whole the responses of the persons are consistent with the model, or why they are not?

    17.3 On the computer printout (Summary Test-of-fit Statistics), there is an overall test of fit statistic for the items (Total Item Chi-Square). Explain why the values of this statistic suggest that on the whole the responses across items are consistent with the model, or why they are not?

```
ANALYSIS  TITLE     RUN1

** SPECIAL  COMMENTS
          1. Derived from the Default Project Settings

** TEST  STRUCTURE
Analysis Type Polytomous/Extended Response Category test format
No. of Items  8
No. of Categories   Different across Items
Score Range [All Items]    14
Some Items Anchored No
Subtests created    No

** CALIBRATING  SAMPLE
No. of Persons:
  * entered Project 25
  * invalid records  0
  * extreme scores   2
  * valid scores    23 [available for analysis]
Missing data detected      None

SUMMARY TEST-OF-FIT STATISTICS

                        ITEM-PERSON INTERACTION
====================================================================
                          ITEMS                       PERSONS
                 Location  Fit Residual      Location  Fit Residual
--------------------------------------------------------------------
Mean               0.000      0.057            0.353     -0.191
SD                 1.372      0.792            1.730      0.786
Skewness                      0.269                       0.836
Kurtosis                     -1.091                      -0.101
Correlation                  -0.021                      -0.017

Complete data DF =            0.810
--------------------------------------------------------------------

====================================================================
        ITEM-TRAIT INTERACTION           RELIABILITY INDICES
--------------------------------------------------------------------
Total Item Chi Squ           20.118     Separation Index  0.865
Total Deg of Freedom         16.000     Cronbach Alpha    0.722
Total Chi Squ Prob         0.214958
--------------------------------------------------------------------

====================================================================
        LIKELIHOOD-RATIO TEST            POWER OF TEST-OF-FIT
--------------------------------------------------------------------
Chi Squ                                 Power is EXCELLENT
Degrees of Freedom                      [Based on SepIndex of 0.865]
Probability
--------------------------------------------------------------------

Cronbachs alpha = 0.722  (Variance[X] = 16.477, Variance Items = 6.067)
Separation index = 0.865 (Variance Beta = 2.99, Av Error Variance = 0.405)
```

GUTTMAN DISTRIBUTION

Serial	PerLocn	1	3	4	5	6	2	7	8	ID
1	1.116	1	1	1	1	1	1	1	1	1
2	-0.335	1	0	0	0	0	1	0	2	2
3	3.122	1	1	1	1	2	1	2	5	3
4	1.675	1	1	1	1	2	0	1	4	4
5	0.914	1	1	1	1	1	0	1	1	5
6	1.471	1	0	1	1	1	0	1	5	6
7	-3.551	0	0	0	0	0	0	0	0	7
8	-0.917	1	1	0	0	0	1	0	0	8
9	1.292	1	1	1	1	2	1	2	0	9
10	1.292	1	1	1	1	1	1	2	1	10
11	0.640	1	1	1	1	1	0	1	0	11
12	0.640	0	1	0	0	1	1	1	2	12
13	1.956	1	1	1	0	2	1	1	5	13
14	-0.335	1	1	1	1	0	0	0	0	14
15	-1.582	1	0	0	0	1	0	0	0	15
16	3.122	1	1	1	1	2	0	2	6	16
17	1.116	1	1	1	0	2	1	2	0	17
18	-3.551	0	0	0	0	0	0	0	0	18
19	-1.582	1	0	0	1	0	0	0	0	19
20	-0.335	1	1	0	1	1	0	0	0	20
21	1.292	1	1	1	1	1	1	1	2	21
22	1.675	1	1	1	1	2	1	1	3	22
23	-1.582	1	1	0	0	0	0	0	0	23
24	0.640	1	1	1	1	1	0	0	1	24
25	0.640	1	1	1	1	1	0	0	1	25

COMPLETE DATA ESTIMATES

TotSc	Frequency	CumFreq	CumPerCent	Estimate	StdErr
0	2	2	8.0	-3.551	1.434
1	0	2	8.0	-2.442	1.064
2	3	5	20.0	-1.583	0.882
3	1	6	24.0	-0.917	0.792
4	3	9	36.0	-0.336	0.731
5	0	9	36.0	0.207	0.659
6	4	13	52.0	0.639	0.568
7	1	14	56.0	0.914	0.508
8	2	16	64.0	1.116	0.480
9	3	19	76.0	1.296	0.472
10	1	20	80.0	1.470	0.484
11	2	22	88.0	1.677	0.516
12	1	23	92.0	1.956	0.588
13	0	23	92.0	2.415	0.723
14	2	25	100.0	3.122	0.941
15	0	25	100.0	4.213	1.294

Separation Index = 0.865 Mean = 0.693
Cronbach Alpha = 0.722 Std Dev = 1.327

INDIVIDUAL PERSON-FIT - Serial Order

ID	Total	Max	Miss	Extrm	Locn	SE	Residual	DegFree	Data Pts	personid
1	8	15	8		1.116	0.48	-1.119	6.50	8	1
2	4	15	8		-0.335	0.73	1.511	6.50	8	2
3	14	15	8		3.122	0.94	-0.499	6.50	8	3
4	11	15	8		1.675	0.52	-0.285	6.50	8	4
5	7	15	8		0.914	0.51	-1.053	6.50	8	5
6	10	15	8		1.471	0.48	0.828	6.50	8	6
7	0	15	8	extm	-3.551	1.43				7
8	3	15	8		-0.917	0.79	0.339	6.50	8	8
9	9	15	8		1.292	0.47	0.133	6.50	8	9
10	9	15	8		1.292	0.47	-0.256	6.50	8	10
11	6	15	8		0.640	0.57	-0.939	6.50	8	11
12	6	15	8		0.640	0.57	1.686	6.50	8	12
13	12	15	8		1.956	0.59	0.231	6.50	8	13
14	4	15	8		-0.335	0.73	-0.235	6.50	8	14
15	2	15	8		-1.582	0.88	-0.560	6.50	8	15
16	14	15	8		3.122	0.94	0.204	6.50	8	16
17	8	15	8		1.116	0.48	0.467	6.50	8	17
18	0	15	8	extm	-3.551	1.43				18
19	2	15	8		-1.582	0.88	0.021	6.50	8	19
20	4	15	8		-0.335	0.73	-0.637	6.50	8	20
21	9	15	8		1.292	0.47	-1.237	6.50	8	21
22	11	15	8		1.675	0.52	-0.706	6.50	8	22
23	2	15	8		-1.582	0.88	-0.713	6.50	8	23
24	6	15	8		0.640	0.57	-0.782	6.50	8	24
25	6	15	8		0.640	0.57	-0.782	6.50	8	25
				Mean:	0.353		-0.191			
				SD :	1.730		0.786			

Key: extm : location value is an extrapolation based on actual estimates
\# : fit residual value exceeds limit set for test-of-fit

Cronbach Alpha = 0.722 Mean Err Var = 0.405
Separation Index = 0.865 Est True Var = 2.589

INDIVIDUAL ITEM-FIT - Serial Order

Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob
1	I0001	Poly	-2.484	0.840	0.160	18.63	4.172	2	0.124179
2	I0002	Poly	0.744	0.481	1.466	18.63	1.148	2	0.563217
3	I0003	Poly	-1.289	0.612	0.263	18.63	1.673	2	0.433129
4	I0004	Poly	-0.337	0.523	-1.080	18.63	3.123	2	0.209872
5	I0005	Poly	-0.108	0.509	0.622	18.63	0.514	2	0.773259
6	I0006	Poly	0.513	0.382	-0.478	18.63	3.570	2	0.167772
7	I0007	Poly	1.392	0.364	-0.611	18.63	3.965	2	0.137714
8	I0008	Poly	1.569	0.167	0.114	18.63	1.952	2	0.376823

INDIVIDUAL ITEM-FIT - Location Order

Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob
1	I0001	Poly	-2.484	0.840	0.160	18.63	4.172	2	0.124179
3	I0003	Poly	-1.289	0.612	0.263	18.63	1.673	2	0.433129
4	I0004	Poly	-0.337	0.523	-1.080	18.63	3.123	2	0.209872
5	I0005	Poly	-0.108	0.509	0.622	18.63	0.514	2	0.773259
6	I0006	Poly	0.513	0.382	-0.478	18.63	3.570	2	0.167772
2	I0002	Poly	0.744	0.481	1.466	18.63	1.148	2	0.563217
7	I0007	Poly	1.392	0.364	-0.611	18.63	3.965	2	0.137714
8	I0008	Poly	1.569	0.167	0.114	18.63	1.952	2	0.376823

INDIVIDUAL ITEM-FIT - Item-Person Fit Residual Order

Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob
4	I0004	Poly	-0.337	0.523	-1.080	18.63	3.123	2	0.209872
7	I0007	Poly	1.392	0.364	-0.611	18.63	3.965	2	0.137714
6	I0006	Poly	0.513	0.382	-0.478	18.63	3.570	2	0.167772
8	I0008	Poly	1.569	0.167	0.114	18.63	1.952	2	0.376823
1	I0001	Poly	-2.484	0.840	0.160	18.63	4.172	2	0.124179
3	I0003	Poly	-1.289	0.612	0.263	18.63	1.673	2	0.433129
5	I0005	Poly	-0.108	0.509	0.622	18.63	0.514	2	0.773259
2	I0002	Poly	0.744	0.481	1.466	18.63	1.148	2	0.563217

INDIVIDUAL CHI SQUARE TEST-of-FIT

Descriptor for Item 1 [I0001]: Locn = -2.484

```
   GROUP         LOCATION            COMPONENT                    Category
Responses
 No.  Size     Max     Mean       Residual        ChiSqu            0       1

 1       7     -.335    -.953        1.326         1.759  OBS.P    .00    1.00
                                                          EST.P    .18     .82
 [OM = 1.00  EV =  .81  OM-EV =  .19  ES =   .50]         OBS.T           1.00

 2       7     1.116     .815       -1.508         2.275  OBS.P    .14     .86
                                                          EST.P    .04     .96
 [OM =  .86  EV =  .96  OM-EV = -.11  ES =  -.57]         OBS.T            .86

 3       9     3.122    1.877         .372          .138  OBS.P    .00    1.00
                                                          EST.P    .01     .99
 [OM = 1.00  EV =  .98  OM-EV =  .02  ES =   .12]         OBS.T           1.00

 Whole sample expected value =    .96

                          ITEM: ChiSqu =   4.172 on  2.00 df [prob =.124179]
```

Descriptor for Item 2 [I0002]: Locn = 0.744

```
   GROUP         LOCATION            COMPONENT                    Category
Responses
 No.  Size     Max     Mean       Residual        ChiSqu            0       1

 1       7     -.335    -.953         .836          .698  OBS.P    .71     .29
                                                          EST.P    .85     .15
 [OM =  .29  EV =  .17  OM-EV =  .12  ES =   .32]         OBS.T            .29

 2       7     1.116     .815        -.474          .224  OBS.P    .57     .43
                                                          EST.P    .48     .52
 [OM =  .43  EV =  .52  OM-EV = -.09  ES =  -.18]         OBS.T            .43

 3       9     3.122    1.877        -.475          .226  OBS.P    .33     .67
                                                          EST.P    .24     .76
 [OM =  .67  EV =  .73  OM-EV = -.07  ES =  -.16]         OBS.T            .67

 Whole sample expected value =    .48

                          ITEM: ChiSqu =   1.148 on  2.00 df [prob =.563217]
```

Descriptor for Item 3 [I0003]: Locn = -1.289

```
________________________________________________________________________________

   GROUP          LOCATION             COMPONENT                   Category
Responses
 No.  Size      Max      Mean       Residual        ChiSqu           0      1
________________________________________________________________________________
 1       7     -.335    -.953        -.031           .001  OBS.P    .43    .57
                                                           EST.P    .42    .58
 [OM =   .57  EV =   .58  OM-EV = -.01  ES =  -.01]        OBS.T           .57

 2       7     1.116     .815         .934           .872  OBS.P    .00   1.00
                                                           EST.P    .11    .89
 [OM = 1.00  EV =   .89  OM-EV =  .11  ES =   .35]         OBS.T          1.00

 3       9     3.122    1.877        -.895           .801  OBS.P    .11    .89
                                                           EST.P    .04    .96
 [OM =   .89  EV =   .95  OM-EV = -.06  ES =  -.30]        OBS.T           .89

 Whole sample expected value =    .83
________________________________________________________________________________
                          ITEM: ChiSqu =   1.673 on  2.00 df [prob =.433129]
________________________________________________________________________________
```

Descriptor for Item 4 [I0004]: Locn = -0.337

```
________________________________________________________________________________
   GROUP          LOCATION             COMPONENT                   Category
Responses
 No.  Size      Max      Mean       Residual        ChiSqu           0      1
________________________________________________________________________________

 1       7     -.335    -.953       -1.250          1.561  OBS.P    .86    .14
                                                           EST.P    .65    .35
 [OM =   .14  EV =   .36  OM-EV = -.22  ES =  -.47]        OBS.T           .14

 2       7     1.116     .815         .616           .380  OBS.P    .14    .86
                                                           EST.P    .24    .76
 [OM =   .86  EV =   .76  OM-EV =  .10  ES =   .23]        OBS.T           .86

 3       9     3.122    1.877        1.087          1.182  OBS.P    .00   1.00
                                                           EST.P    .10    .90
 [OM = 1.00  EV =   .89  OM-EV =  .11  ES =   .36]         OBS.T          1.00

 Whole sample expected value =    .70
________________________________________________________________________________
                          ITEM: ChiSqu =   3.123 on  2.00 df [prob =.209872]
________________________________________________________________________________
```

Descriptor for Item 5 [I0005]: Locn = -0.108

```
______________________________________________________________________________
   GROUP          LOCATION              COMPONENT                   Category
Responses
 No.  Size      Max      Mean       Residual        ChiSqu            0       1
______________________________________________________________________________

 1       7      -.335    -.953          .677            .458  OBS.P    .57     .43
                                                              EST.P    .70     .30
 [OM =  .43  EV =  .31  OM-EV = .11  ES =   .26]             OBS.T            .43

 2       7      1.116     .815          .004            .000  OBS.P    .29     .71
                                                              EST.P    .28     .72
 [OM =  .71  EV =  .71  OM-EV = .00  ES =   .00]             OBS.T            .71

 3       9      3.122    1.877          .237            .056  OBS.P    .11     .89
                                                              EST.P    .12     .88
 [OM =  .89  EV =  .86  OM-EV = .03  ES =   .08]             OBS.T            .89

 Whole sample expected value =    .70
______________________________________________________________________________
                          ITEM: ChiSqu =   0.514 on  2.00 df [prob =.773259]
______________________________________________________________________________
```

Descriptor for Item 6 [I0006]: Locn = 0.513 Sprd = 1.444

```
______________________________________________________________________________

   GROUP          LOCATION              COMPONENT                   Category
Responses
 No.  Size      Max      Mean       Residual        ChiSqu         0      1    2
______________________________________________________________________________
 1       7      -.335    -.953        -1.272        1.617  OBS.P    .71    .29 .00
                                                           EST.P    .49    .48 .03
 [OM =  .29  EV =  .54  OM-EV = -.26  ES =  -.48]          OBS.T           .29 .00

 2       7      1.116     .815          .214         .046  OBS.P    .00    .86 .14
                                                            EST.P   .12    .67 .21
 [OM = 1.14  EV = 1.10  OM-EV =  .05  ES =   .08]          OBS.T         1.00  .14

 3       9      3.122    1.877         1.381        1.907  OBS.P    .00    .33 .67
                                                           EST.P    .03    .50 .47
 [OM = 1.67  EV = 1.42  OM-EV =  .25  ES =   .46]          OBS.T          1.00 .67

 Whole sample expected value =   1.09
______________________________________________________________________________
                          ITEM: ChiSqu =   3.570 on  2.00 df [prob =.167772]
______________________________________________________________________________
```

Descriptor for Item 7 [I0007]: Locn = 1.392 Sprd = 1.038

```
______________________________________________________________________________
   GROUP         LOCATION           COMPONENT                     Category
Responses
 No.  Size     Max     Mean       Residual        ChiSqu          0    1    2
______________________________________________________________________________

 1      7     -.335    -.953       -1.478       2.183  OBS.P   1.00   .00 .00
                                                       EST.P    .78   .21 .01
 [OM =  .00  EV =  .25  OM-EV = -.25  ES =  -.56]      OBS.T          .00 ***

 2      7     1.116     .815         .387        .150  OBS.P    .29   .57 .14
                                                       EST.P    .34   .55 .11
 [OM =  .86  EV =  .77  OM-EV =  .09  ES =   .15]      OBS.T          .67 .20

 3      9     3.122    1.877        1.278       1.632  OBS.P    .00   .56 .44
                                                       EST.P    .12   .56 .32
 [OM = 1.44  EV = 1.18  OM-EV =  .26  ES =   .43]      OBS.T         1.00 .44

 Whole sample expected value =    .83
______________________________________________________________________________
                         ITEM: ChiSqu =   3.965 on  2.00 df [prob = .137714]
______________________________________________________________________________
```

Descriptor for Item 8 [I0008]: Locn = 1.569 Sprd = 0.113 Skew = 0.024 Kurt = 0.003

```
______________________________________________________________________________

   GROUP         LOCATION           COMPONENT                     Category
Responses
No. Size  Max  Mean   Residual  ChiSqu      0    1    2    3    4    5    6
______________________________________________________________________________

 1   7  -.335  -.953   .914     .835  OBS.P .86  .00  .14  .00  .00  .00 .00
                                      EST.P .90  .09  .01  .00  .00  .0  .00
 [OM =  .29  EV =  .14  OM-EV =  .14  ES =   .35]
                                       OBS.T .00  1.00 .00    ****    **** ****

2   7  1.116  .815   -.905    .819  OBS.P .29  .57  .14  .00  .00  .00 .00
                                      EST.P .41  .25  .14  .10  .07  .03 .00
 [OM =  .86  EV = 1.35  OM-EV = -.49  ES =  -.34]
                                       OBS.T .67  .20  .00    ****    **** ****

3   9   3.122  1.877   -.546    .298  OBS.P .11  .11  .11  .11  .11  .3  .11
                                      EST.P .02  .04  .07  .13  .26  .34 .14
 [OM = 3.44  EV = 3.72  OM-EV = -.27  ES =  -.18]
                                       OBS.T .50  .50  .50  .50  .75  .25

 Whole sample expected value =   1.70
______________________________________________________________________________

                          ITEM: ChiSqu =   1.952 on  2.00 df [prob =.376823]
______________________________________________________________________________
```

CATEGORY RESPONSE PROPORTIONS

Seq	Item Label	0	1	2	3	4	5	6
1	I0001 Descriptor for Item 1	.04	.96					
2	I0002 Descriptor for Item 2	.52	.48					
3	I0003 Descriptor for Item 3	.17	.83					
4	I0004 Descriptor for Item 4	.30	.70					
5	I0005 Descriptor for Item 5	.30	.70					
6	I0006 Descriptor for Item 6	.22	.48	.30				
7	I0007 Descriptor for Item 7	.39	.39	.22				
8	I0008 Descriptor for Item 8	.39	.22	.13	.04	.04	.13	.04

ITEM SPECIFICATIONS

Code	Seq No.	Test	Key	Categ	Thresh	Param
I0001	1	Poly		2	1	1
I0002	2	Poly		2	1	1
I0003	3	Poly		2	1	1
I0004	4	Poly		2	1	1
I0005	5	Poly		2	1	1
I0006	6	Poly		3	2	2
I0007	7	Poly		3	2	2
I0008	8	Poly		7	6	4

Exercise 2: Basic Analysis of Dichotomous and Polytomous Responses

Use simulated Data Set A which contains the responses of 400 persons to 15 items. Ten items are multiple-choice items and are scored as dichotomous items (0, 1) and five are polytomous with three ordered categories (0, 1, 2). You will be required to run the data using RUMM2030 and write a report on your analysis completing the tasks below. Two groups, boys and girls, need to be identified.

1. Identify the three best fitting dichotomous items statistically and confirm this fit graphically.
2. Identify the three most misfitting dichotomous items statistically and explain in which way they misfit.
3. Identify the three best fitting polytomous items statistically and confirm this fit graphically.
4. Identify the three most misfitting polytomous items statistically and explain in which way they misfit.
5. Identify the three best fitting persons statistically and explain why their patterns fit so well to the Rasch model.
6. Identify the three most misfitting persons and explain in which way they misfit.
7. Report on the distribution of the persons in relation to the targeting of the items and comment on the traditional reliability of the responses.
8. Report on the operation of the response categories for items and explain which categories in which items are not working as intended.

9. Suppose that you were considering making two test forms, one containing only the multiple-choice items (dichotomous) and one containing only the polytomous items. Report on the raw score equivalence between the two forms—that is, what are equivalent total scores on the polytomous items for each of the scores 1 to 9 on the 10 multiple-choice items.
10. Investigate whether any items show DIF for gender.
11. Compare the overall performance of the two groups: boys and girls.

Exercise 3: Advanced Analysis of Dichotomous Responses

Description of Data Set B—dichotomous responses

The data set contains responses of 1000 year 6 and 1000 year 7 students to a multiple-choice test. The year 6 students' test consisted of items 1–36, and the year 7 students' test consisted of items 31–66. Items 31–36 were link items and both year groups responded to these items. These data have been simulated to fit the Rasch model, but with some deviation. Apart from the student ID, a student's year group is also indicated. You are to run these data using RUMM2030, and you may use the template files provided when creating the project. After considering the graphical and statistical evidence together, answer the following questions.

Part A

First analyse *only the year 6 data*. Delete all the year 7 data by choosing the *Delete sample—person factor* option.

1. Summary test of fit statistics—provide the summary table.
   (a) How do the values for the item fit-residual mean and SD deviate from the expected? What does that mean?
   (b) How does the value for the item–trait interaction chi-square statistic deviate from the expected? What does that mean?

2. Item tests of fit—year 6
   (a) Which items misfit according to the fit-residual or chi-square fit statistics or the ICC or a combination of these criteria? Provide both graphical and statistical evidence. State which items over-discriminate and which items under-discriminate.
   (b) Which items misfit when the Bonferroni correction is used? (1 mark) Use the probability of 0.05.
   (c) Which items misfit according to the ANOVA fit statistic (with and without the Bonferroni correction)? Justify your answer with the relevant statistics.

3. Power of tests of fit—sample size, etc.

   (a) Use the *Delete sample—random select* option in RUMM2030 to reduce the sample size to 400. How does that change the pattern of misfit?
   (b) Is the power to detect misfit greater for item 35 (close to person locations) or item 20 (further from person locations)? Is the error of measurement greater for item 35 (close to person locations) or item 20 (further from person locations)?
   (c) Return to using the original sample size of 1000. What is the number of persons in class interval 10 of item 20? Change the number of class intervals to 5—does that change the fit/misfit? If so, how? (e.g. is the value of the fit-residual or chi-square smaller or larger than before the change? Has it changed the significance of the chi-square test?)

   Now analyse *only the year 7 data*. Delete all the year 6 data by choosing the *Delete sample—person factor* option as before.
4. Item tests of fit—year 7

   (a) Which items misfit according to the fit-residual, in conjunction with the chi-square fit statistics and the ICC? Provide both graphical and statistical evidence. State which items over-discriminate and which items under-discriminate. Which items misfit when the Bonferroni correction is used?
   (b) Which items misfit according to the ANOVA fit statistic (with and without the Bonferroni correction)?

Part B

Analysis of guessing in Data Set B

Analyse the year 6 and 7 data together.

1. Which items under-discriminate? Provide fit-residual, chi-square and graphical evidence. Use the Bonferroni correction.
2. Do the ICCs suggest guessing played a role in these items? Why or why not?
3. Analyse the data with a tailored analysis choosing a cut-off of 0.25. In order to compare the original with the tailored item estimates, anchor the original data set on the mean of the 10 easiest item estimates from the tailored analysis [for instructions on how to run an Anchor analysis, see Chap. 2 of the 'Extending the RUMM2030 Analysis' manual (maRmExtend.pdf) which was automatically installed into the RUMM directory when you installed RUMM]. Draw a graph plotting the item estimates from the tailored analysis against the item estimates from the anchored analysis. According to the graph did any items change location? Were they easier or more difficult in the tailored analysis?

Exercise 4: Advanced Analysis of Polytomous Responses

Description of Data Set C—Polytomous responses

The data set contains responses of 250 males and 250 females to a 10 item questionnaire. All items have 5 response categories. These data have been simulated to fit the Rasch model, but with some deviation. Apart from an ID for each person, an M or F indicates whether the person is male or female.

Part A

You are to run these data on RUMM2030, with the option of using the template files provided when creating the project. After considering the graphical and statistical evidence together, answer the following questions:

1. Report and interpret the summary statistics, including the item and person fit-residual means and SDs, the overall item–trait chi-square and the person separation index.
2. Are any items misfitting according to the chi-square and fit-residual fit statistics (provide the values of the statistics and use the Bonferroni correction)?
3. Are there any items with disordered thresholds? Provide either the threshold values or the threshold map as evidence.
4. Show the category probability curves *including observed categories* for any item(s) identified in question 3. Describe how these graphs show that the response categories are not working as intended.
5. Show the threshold probability curves *including observed thresholds* for any item(s) identified in question 3. Describe how these graphs show that the response categories are not working as intended.
6. Is there sufficient evidence to combine response categories for any of the item(s) identified in question 3? If so, rescore the item(s) and compare the individual item fit before and after rescoring. Is the fit improved?
7. In future administrations of the questionnaire would you change the response format for any of the questions? State the reason(s) why or why not, as well as how you would change it if you responded yes

Part B

The default setting in RUMM2030 is the partial credit parameterization of the model. This was the setting you used to analyse the data in part A of the exercise. The rating scale parameterization can be selected by creating a new analysis, then selecting *Edit analysis specifications* on the Analysis Control form. Then Choose *Rating* instead of *Unrestricted* model structure (Items are not grouped into rating subsets). Also note the other options on the *Edit analysis specifications* screen. *Converge limits* and *Number of loops* can be specified for both item and person estimations. If the estimation does not converge using the default settings you can change these, for example, the converge limits can be made less stringent or the number of loops can be increased.

1. Use the rating scale parameterization of the model to analyse the data again.
   (a) Compare the fit of the data to the two different parameterizations of the model (report the overall item trait chi-square as well as any individual items that misfit according to the chi-square fit statistic) (*Display specifications screen—Summary Statistics and Individual Item fit*).
   (b) Compare the threshold parameters from the two different parameterizations (*Display specifications screen—Item details—Thresholds*).
   (c) Compare the spread component for the items for the two different parameterizations of the model (*Display specifications screen—Item details—Guttmann Principal components*).
2. Consider the partial credit and rating scale parameterizations of the model.
   (a) In your own words describe the difference between the partial credit and rating scale parameterizations of the model.
   (b) Give an example of where each might be an advantage.
3. What is the source of difference between the graded response model and polytomous Rasch model? More specifically, what is partitioned in forming contiguous categories?

Exercise 5: Analysis of Data with Differential Item Functioning

In this exercise you do DIF analysis on Data Set B.

1. What is the difference in the person location means for the two year groups? Is the difference statistically significant?
2. DIF analysis:
   (a) Are any items showing DIF for year group? Provide graphical evidence. Also provide statistical evidence (DIF summary results with Bonferroni correction)—show the relevant table.
   (b) Which item would you split first, if any? Splitting the item would mean that the item is no longer a link item. Proceed with splitting the item and provide statistical evidence (DIF summary results with Bonferroni correction) after splitting. Redo (a) and (b) until no items show statistically significant DIF.
3. What is the difference in the person location means for the two year groups after you have split the items, that is, when no items show statistically significant DIF? Is this difference statistically significant? Compare the means with the means before any items were split (question 1).
4. Explain which item(s) showed real DIF and which item(s) showed artificial DIF?

Exercise 6: Analysis of Data with Dependence

1. Interpreting tests of fit
   (a) If all items fitted the model does that confirm the unidimensionality of the construct measured? Why or why not?
   (b) How do the residual and chi-square item fit statistics relate to the infit and outfit statistics reported by some other Rasch analysis software packages?
   (c) There is an option in RUMM2030 to specify an effective sample size for the calculation of the chi-square fit statistic (*Amend sample size* and *Create adjusted chi-square* on the Display Control form). Use these options in RUMM2030 to specify an effective sample size of 200 for Data Set B. Does that change the results of fit according to the chi-square statistic? If so, how?

Description of Data Set D—Polytomous responses and Multidimensionality

The data set contains responses of 400 persons to a 10 item questionnaire. All items have 5 response categories. These data have been simulated to fit the Rasch model, but with some deviation. Analyse this data set using RUMM2030 and answer questions 2, 3 and 4 below:

2. Analysis of residuals
   (a) Provide the matrix of correlations between item residuals. Highlight correlations above 0.3. Is there any evidence of violations of the assumption of independence?
   (b) Do a principal components analysis (PCA) of the residuals. Provide the PC Loadings with items sorted according to their loadings on PC1. Is there any evidence of multidimensionality?
   (c) Provide the principal component summary. What percentage of the total variance is accounted for by the first component? What conclusion can you reach regarding the dimensionality of this instrument?

3. Using the factor loadings from the PCA divide the items into two subscales. Use the paired t-test option in RUMM2030 to test whether person estimates derived from the two subscales are significantly different from each other.
4. Divide the items into two subscales using evidence from the PCA and t-test. Perform a subtest analysis and report the following results from the summary statistics. What can you conclude regarding the amount of multidimensionality in these data?

RELIABILITY INDICES

	run1	subtest	c*c	c	r	A
Per Sep Idx:						
Coef Alpha :						

5. Using Data Set B (year 6 only) answer the following questions:
   a. Provide the item residual correlation matrix. Is there any evidence of response dependence?
   b. Use the procedure described in the Andrich and Kreiner (2010) paper (*Item split* on *item dependence* option in RUMM2030) to estimate the amount of response dependence.

Exercise 7: Analysis of More than Two Facets and Repeated Measurements

Part A

Description of Data Set E—Responses to the same instrument at two time points

The data set contains responses of persons to a questionnaire at two time points. Analyse this data set using RUMM2030 and answer the questions given below:

1. With the data racked what is the mean item location at time 1 and time 2?
2. With the data stacked what is the mean person location at time 1 and time 2?

Part B

Description of Data Set F—Ratings of raters judging persons

The data set contains ratings of raters grading persons on 6 criteria. Analyse this data set using a three-facet analysis in RUMM2030 and answer the questions given below:

1. Provide the rater locations and fit statistics. Do any raters misfit?
2. Provide the criteria locations and fit statistics. Is there any evidence of misfit?

Exercise 8: Writing Up a Rasch Model Analysis

Read:
Tennant, A. & Conaghan, G. (2007). The Rasch Measurement Model in Rheumatology: What is it? Why use it? When should it be applied and what should one look for in a Rasch paper? *Arthritis & Rheumatism, 57*(8), 1358–1362.

Analyse your own data set. Write up the results in the form of a short paper following the guidelines in the paper above. You should have the following sections: introduction including a brief literature review, a statement of aims and hypotheses, method, results and brief discussion (even if using a simulated data set).

The preferred option for this exercise is for you to use your own data set. If you do not have access to a data set you can use simulated Data Set G. If you are using a simulated data set please state that at the start of your write-up. Create a context of your own for the data set, e.g. educational test with two groups of students (two countries, male/female, private school/public school, etc.). Write the introduction (including a brief literature review) and the discussion with this context in mind.

The world limit is 3000 words (excluding Abstract, References, Tables and Figures). Consistent with preparing a publishable manuscript, please ensure that you follow APA guidelines and minimize the number of tables and figures (e.g. only present one or two example Item characteristic curves, rather than all of them).

Remember to address in the analysis the following aspects, if relevant to the data:

- Description of the instrument and its purpose, and aims of analysis.
- Which parameterization of the Rasch Model used, why?
- Targeting, reliability and sample size.
- Model fit, including the approach to model fit and the application of multiple criteria.
- Working of response categories.
- Differential item functioning (DIF): real and artificial DIF. In real data, the analysis of DIF may follow some item reconstruction or elimination. For the purpose of this exercise, carry out the DIF analysis on the original set of items.
- Violations of the assumption of local independence: multidimensionality and response dependence—detection and estimation of magnitude.
- Guessing, if relevant.
- Conclusions about the operation and appropriateness of the instrument.

RUMM2030 Exercises Solutions

Exercise 1

Part A:

1. 1.1 8
   1.2 Different across items.

2. 2.1 25
   2.2 2
   2.3 23

3.

	Items			Person	Scores
Item number	Difficulty $\hat{\delta}_i$	Standard error $\hat{\sigma}_i$	Raw score	Proficiency estimate	Estimated standard error
1	−2.484	0.840	2	−1.583	0.882
4	−0.337	0.523	7	0.914	0.508
8	1.569	0.167	13	2.415	0.723

4.

Person ID	Total score	Proficiency estimate	Fit statistic	Observed response pattern (order items by difficulty)
3	14	3.122	−0.499	1, 1, 1, 1, 2, 1, 2, 5
12	6	0.640	1.686	0, 1, 0, 0, 1, 1, 1, 2
19	2	−1.582	0.021	1, 0, 0, 1, 0, 0, 0, 0

5.

Item number	Location estimate $\hat{\delta}_i$	Fit chi-square	Probability
1	−2.484	0.160	0.124
4	−0.337	−1.080	0.210
7	1.392	−0.611	0.138

6.

6.1

Item 1	*Class interval*		
Location $\hat{\delta}_i = -2.484$	1	2	3
Mean proficiency	−0.953	0.815	1.877
Observed mean (OM)	1.00	0.86	1.00
Estimated mean or expected value (EV)	0.81	0.96	0.98

6.2

Item 4	*Class interval*		
Location $\hat{\delta}_i = -0.337$	1	2	3
Mean proficiency	−0.953	0.815	1.877
Observed mean (OM)	0.14	0.86	1.00
Estimated mean or expected value (EV)	0.36	0.76	0.89

6.3

Item 7	*Class interval*		
Location $\hat{\delta}_i = 1.392$	1	2	3
Mean proficiency	−0.953	0.815	1.877
Observed mean (OM)	0.00	0.86	1.44
Estimated mean or expected value (EV)	0.25	0.77	1.18

7.

7.1

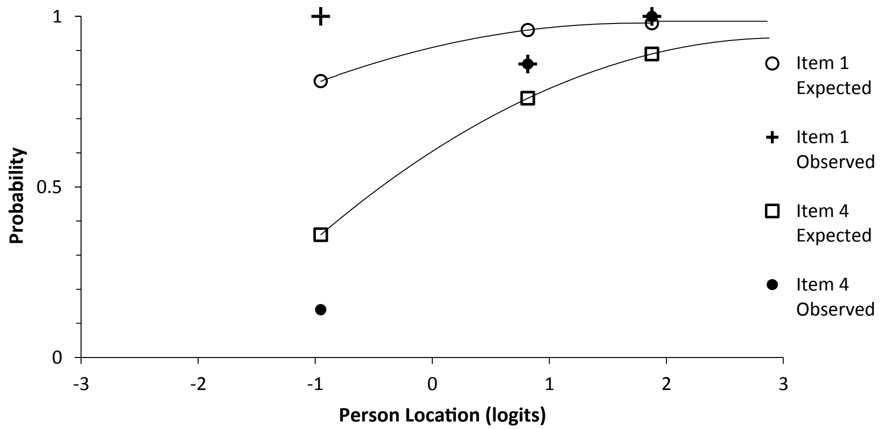
Coefficient α	0.722
Variance (total score)	16.477
Variance (items)	6.067

7.2

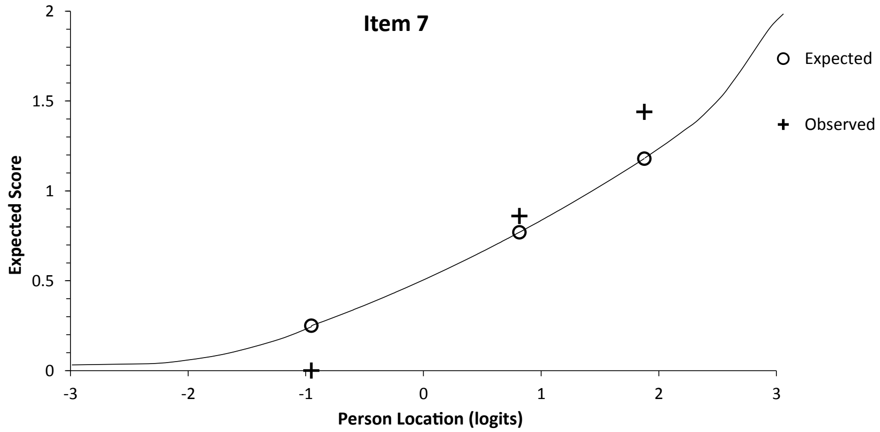
Index r_β	0.865
Variance of $\hat{\beta}$	2.99
Average error variance	0.405

Part B:

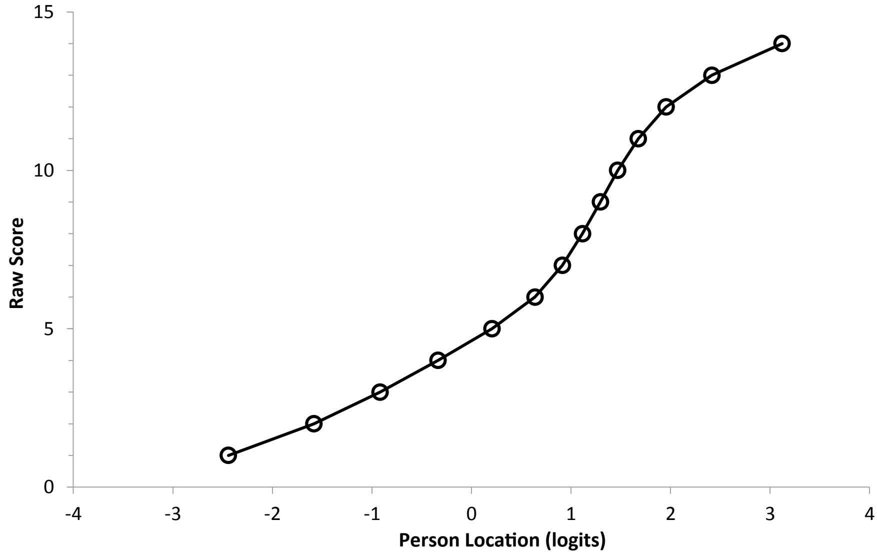
8. 8.1, 8.2, 8.3



8.4 8.5 8.6.



9. 9.1 9.2.



10.

10.1 8/7 * (16.477 − 6.067)/16.477 = 0.722
10.2 (2.99 − 0.405)/2.99 = 0.865.

Part C:

11. They answered all items incorrectly; therefore, no relative item information can be obtained from their responses.
12.
    - 12.1 Difficulty,
    - 12.2 Percentage correct on an item.
13.
    - 13.1 Item 1
    - 13.2 The slope of the observed means in Item 1 is flatter than its expected curve (does not discriminate well), while the slope of Item 4 is steeper than expected (highly discriminative).
14.
    - 14.1 More than expected, the slope of observed means is steeper than its expected curve.
    - 14.2 No, the probability of the chi-square test (0.138) is greater than 0.05.
15.
    - 15.1 4.213
    - 15.2 −3.551
16.
    - 16.1 Closest: 19 Furthest: 12
    - 16.2 Person 19 scored their two points on items of relatively low difficulty. Person 12 scored five of their six points on items of relatively high difficulty, yet answered easier items wrongly (not a Guttman structure).
17.
    - 17.1 Moderately good.
    - 17.2 Consistent because the person fit-residual mean (−0.191) and SD (0.786) are close to the theoretical ones.
    - 17.3 Consistent because the total item chi-square probability (0.215) is greater than 0.05.

Exercise 2

1. Items 8, 9 and 10 have fit-residuals closest to zero, non-significant chi-square statistics and ICCs showing almost perfect fit.
2. Items 4, 5 and 6 have fit-residuals furthest from zero. Item 4 has the worst fit and its ICC indicates under-discrimination. Items 5 and 6 show less extreme misfit with non-significant chi-square statistics and some under-discrimination and over-discrimination, respectively.

3. Items 11, 12 and 15 are sufficiently close to zero that the chi-square test does not indicate misfit. The ICCs show the observed means are very close to the expected values.
4. Items 12, 13 and 14 have fit-residuals furthest from zero. Item 14 has a significant chi-square statistic, it does not fit the Rasch model. Items 12 and 13 show non-significant misfit.
5. Persons 313, 115 and 339 answered the easier items correctly and made mistakes answering the more difficult items (close to Guttman pattern).
6. Person 277 and 331 had high negative fit-residuals and a Guttman pattern that may be too good to be true. Person 394 had a high positive fit-residual and a response pattern which is very different from a Guttman pattern (answered some easier items wrong and some more difficult items correct).
7. Items are relatively well targeted as there is a spread of item difficulties with most items clustered where the majority of persons are located on the scale. The test is of moderate difficulty and captures most proficiency levels. However, some respondents are below and above the range of measurement covered by the items, indicating floor and ceiling effects.
8. Item 12 has disordered thresholds. With increasing proficiency, it is not more likely for persons to score highly. There is also no person proficiency level for which a score of 1 is most likely. All the other graded response items work reasonably well.
9.

Item location	MC test (dichotomous items)	Graded response items
−3.279	1.0	0.0
−2.393	2.0	0.7
−1.577	3.0	1.7
−0.770	4.0	3.4
0.017	5.0	5.1
0.807	6.0	6.6
1.584	7.0	8.1
2.347	8.0	9.2
3.200	9.0	9.9

10. Items 5 and 6 show uniform DIF for gender. Boys have a higher probability of success on item 5 and girls on item 6. The ANOVA found significant uniform DIF on these items and non-uniform DIF on item 6.
11. There are more boys in the low-proficiency group and fewer in the middle to high-proficiency groups. The boys form a flat, almost bimodal distribution of proficiency levels and girls form an almost normal distribution. The ANOVA found a significant difference with girls performing better than boys.

Exercise 3

Part A

1.

(a) The observed mean (−0.399) and SD (1.250) deviate, but not substantially, from the theoretical. The pattern of responses is consistent with the model (very slight over-fit).

(b) The significant item–trait interaction chi-square statistic (435.139, p = 0.000) suggests that overall the responses are not consistent with the model (contradicting part a). The hierarchical ordering of the items varies across the trait. However, it is noted that the chi-square statistic is affected by the sample size and the sample size in each interval.

2.

(a) Items 21, 28 and 29 over-discriminate. They have large negative fit-residuals, significant chi-square fit statistics and their observed means are steeper than the theoretical curve.
Item 15 under-discriminates. It has a positive fit-residual (quite large but within the acceptable range), significant chi-square fit statistic and the observed proportions are flatter than the theoretical curve.

(b) Item 21.

(c) Items 16, 21, 28 and 29 show misfit without the Bonferroni correction (significant F statistic at the 0.01 level). Only items 29 and 21 show misfit with the Bonferroni correction.

3.

(a) Item 29 shows misfit according to the fit-residual, chi-square, ANOVA fit statistics and the ICC. With the Bonferroni correction, only the ANOVA fit statistic is significant. Items 21 and 28 have acceptable fit-residuals, but chi-square and ANOVA fit statistics show poor fit.

(b) Item 20. The greater the distance of a person from an item, the greater the power to detect misfit and the greater the error of measurement.

(c) 180
It does not change considerably, the fit-residual remains the same while the chi-square and ANOVA fit statistics improve slightly since their degrees of freedom are reduced.

4.

(a) Item 63 has a large positive fit-residual, significant chi-square fit statistic with Bonferroni correction and the observed proportions are flatter than the ICC (under-discriminates).

(b) Item 63 misfits according to the ANOVA fit statistic without the Bonferroni correction. With the Bonferroni adjustment, none of the items misfit.

Part B

1. Item 63 under-discriminates, it has a large positive fit-residual, significant chi-square statistic and the observed proportions are flatter than the theoretical curve.
2. Item 63 could be prone to guessing from lower proficiency students. With 5 class intervals, the observed mean in the least proficient group is greater than expected and in the more proficient groups are lower than expected (ICC is located further to the left because item appears easier).
3. The items were slightly harder in the tailored compared to the anchored analysis, but the estimates are very similar (R-squared = 0.998). The deviation from the regression line ($y = 1.011x$) is mainly a function of the relative difficulty of the items (the harder the item, the more likely that guessing could have played a role).

Exercise 4

1. The observed item fit-residual mean (−0.052) and SD (1.261) are very close to the theoretical values. The pattern of responses is consistent with the model.
   The observed person fit-residual mean (-0.198) and SD (0.810) deviate, but not substantially, from the expected values. The response profiles across persons are as expected (slight over-fit).
   The item–trait interaction chi-square statistic (100.169, p = 0.217) could have occurred by chance. The responses fit the model and the hierarchical ordering of items is consistent across all levels of the underlying trait.
   The person separation index (0.93) is very high (over 0.7 is acceptable), indicating good internal consistency and that persons can be reliably differentiated.
2. No items misfit according to both fit statistics. Item 5 has a large positive fit-residual (2.812) but its chi-square statistic is not significant with the Bonferroni correction.
3. Item 5 shows disordered thresholds. The thresholds for items 1, 6, 7, 8 and 10 are not equally spaced and some are located very close to each other (poor discrimination between categories).
4. The thresholds for item 5 are not correctly ordered, the third threshold is lower than the second threshold and there is no region where a score of 2 is most likely. The thresholds for items 1, 6, 7, 8 and 10 are correctly ordered, but a score of 3 (for items 1 and 6), 1 (item 7) and 2 (items 8 and 10) is almost never most likely.
5. The second and third thresholds of item 5 are reversed and the observed proportions deviate substantially from the theoretical at the third threshold.
   The other items have thresholds which lie very close to each other showing very little discrimination between pairs.
6. Item 5 misfits according to its fit-residual and reversed thresholds, and therefore it is reasonable to collapse categories 2 and 3 for an exploratory analysis. This resulted in properly ordered categories and considerably improved fit.
   Items 1, 6, 7, 8 and 10 have thresholds that are almost overlapping, and therefore it would be worth collapsing categories. Items 7 and 8 improved in fit.

For the remaining items, the trend towards over-fit evident in the initial analysis becomes even more apparent.

7. Ambiguous response category labels and/or too many response options would make it difficult to distinguish between options, resulting in disordered thresholds and a lack of discrimination. Successive categories might not necessarily imply successively more of a property and reducing the number of categories in a revised instrument might solve the problem.

Part B

1.

(a) The significant item–trait chi-square suggests that the responses are not consistent with the rating scale model (in contrast to the partial credit model) and the hierarchical ordering of items varies across the trait. Items 3, 6, 7 and 8 show misfit according to their chi-square fit statistics with the Bonferroni correction. None of the items misfit using the partial credit model. The fit-residual statistics are all within the acceptable range.

(b) None of the items had disordered thresholds when the rating scale model was employed, but item 5 had reversed thresholds with the partial credit model. The threshold locations differ between the rating scale (same for all items) and the partial credit (different across items) parameterisations.

(c) The spread parameter differs for items and on average is large with the partial credit model, indicating that the majority of responses appear in the middle categories (narrower spread). The spread parameter is the same for all items and smaller with the rating scale model, indicating that the majority of responses are in the extreme categories (wider spread).

2.

(a) The rating scale model assumes that the distance between thresholds across all items are the same because they share a common response format (number of categories and descriptors). Therefore, both the centralized thresholds and the spread are the same for all items.

In the partial credit model, the numbers of categories differs across items because each has its own response structure. The thresholds differ across items, so additional parameters have to be estimated, resulting in a more complex model.

(b) Partial credit is useful in educational testing situations where some responses are given partial credit (scoring rubric) and each item has its own response format. The rating scale model is useful in attitudinal surveys where all items share the same number of categories with identical descriptors (Likert-type items).

3. The graded response model partitions the frequency distribution of the responses while the polytomous Rasch model partitions the underlying continuum.

Exercise 5

1. Year 6 = 0.40, Year 7 = 0.38. The difference of 0.02 is not statistically significant (p = 0.70).
2.
   (a) Items 33, 34 and 35 have statistically significant F statistics (p = 0.000) and their ICCs show uniform DIF for year group.
   (b) Item 35 is split first because it has the highest Mean Square. Item 34 is split next because it continues to show DIF. Then, no DIF remains.
3. Year 6 = 0.21, Year 7 = 0.53. The difference of 0.32 is statistically significant (p = 0.00). The mean of the year 6s decreased and the year 7s increased after splitting two items.
4. Items 34 and 35 show real DIF and Item 33 shows artificial DIF.

Exercise 6

1.
   (a) If all items fitted the Rasch model, there would be strong evidence for unidimensionality of the construct measured, but not confirmation.
   (b) Outfit = fit-residual, Infit = weighted.
   (c) The chi-square statistic is no longer statistically significant.

2.
   (a)

Item	I0001	I0002	I0003	I0004	I0005	I0006	I0007	I0008	I0009	I0010
I0001	1									
I0002	−0.314	1								
I0003	0.2	−0.344	1							
I0004	−0.285	**0.314**	−0.417	1						
I0005	0.131	−0.399	0.152	−0.411	1					
I0006	−0.326	0.195	−0.315	0.176	−0.426	1				
I0007	0.201	−0.293	0.151	−0.356	0.23	−0.442	1			
I0008	−0.281	0.203	−0.306	0.195	−0.381	**0.305**	−0.413	1		
I0009	0.126	−0.372	0.259	−0.364	0.266	−0.357	**0.329**	−0.39	1	
I0010	−0.279	0.157	−0.327	0.19	−0.286	0.175	−0.249	0.075	−0.26	1

There are a number of relatively high correlations (pairs of items have something in common in addition to what all the items have in common).

   (b)

Item	PC1
I0009	0.639
I0005	0.631
I0007	0.626

(continued)

(continued)

Item	PC1
I0003	0.577
I0001	0.504
I0010	−0.473
I0008	−0.605
I0002	−0.608
I0004	−0.632
I0006	−0.635

Half the items load positively and half load negatively on the first principle component.

(c)

PC	Eigen	Percent (%)	CPercent (%)	StdErr
PC001	3.547	35.47	35.47	0.491
PC002	1.068	10.68	35.47	0.144
PC003	0.941	9.41	55.56	0.125
PC004	0.849	8.49	64.05	0.108
PC005	0.826	8.26	72.31	0.106
PC006	0.744	7.44	79.75	0.103
PC007	0.705	7.05	86.80	0.096
PC008	0.665	6.65	93.44	0.089
PC009	0.585	5.85	99.29	0.088
PC0010	0.071	0.71	100.00	0.06

35.47% suggesting the instrument is multidimensional.

3. The two subscales are significantly different. Independent t-tests for each person showed that 26% of the values exceed the 5% level and 10% exceed the 1% level.

4.

```
-----------------------------------------------------------------------------
              run1   subtest  c*c    c      r      A
Per Sep Idx:  0.872  0.554    1.291  1.136  0.437  0.635
Coef Alpha:   0.871  0.607    0.980  0.990  0.505  0.697
-----------------------------------------------------------------------------
```

Reliability estimates (PSI and Coefficient Alpha) decrease, c is large and the correlation is small, which all suggest multidimensionality.

5.

(a) Item 28 and item 29 show response dependence (r = 0.876).
(b) d = 3.047
Students that scored 0 on item 28 found item 29 more difficult by 3 logits than it would otherwise be, and those who scored 1 found it easier by 3 logits.

Exercise 7

Part A

1. Time 1 = 0.097, Time 2 = −0.097.
2. Time 1 = −0.744, Time 2 = −0.559.

Part B

1.

Rater	Location	Fit residual
1	0.228	1.008
2	0.025	0.834
3	−0.253	−1.318

None of the raters misfit.

2.

Criteria	Location	Fit residual
1	−0.53	0.252
2	−0.28	−0.022
3	−0.095	0.654
4	0.158	0.838
5	0.261	−0.145
6	0.486	−0.253

No evidence of misfit.

Appendix D
Statistics Reviews, Exercises and Solutions

Statistics Review 1: Sigma Notation, Mean and Variance

Key words: Symbols X_i, N, $\overline{X}$, s^2, s, Greek letter sigma $\sum$, Mean, Median, Mode, Range, Interquartile range, Variance, Standard deviation (SD)

Key points: The distribution of a set of N scores can be characterized by measures of central tendency (e.g. mean) and variability [e.g. variance. standard deviation (SD)]. The symbol X_i is often used to refer to individual scores and the Greek letter sigma $\sum$as the sum of the scores. The mean is then indicated by $\overline{X}$, the variance by $s^2$and the SD by s.

If a set of scores has been obtained by giving a test to a group of students, it is often helpful to be able to speak about the pattern or distribution of all the scores rather than to have to give the whole set. The two characteristics of this distribution in which we are usually most interested are the position of its 'middle' and the amount of 'spread' in the scores within it.

To illustrate this, we can take the scores obtained by six students (that's a good student–staff ratio) on a test of ten items scored simply correct (1) and incorrect (0). Their scores were 6, 4, 7, 10, 7 and 2.

If we want to refer to this set of numbers without actually writing down any one of them, we can use the symbol X instead of the number, with the subscript i indicating it is the i th number, X_i. The first number can be referred to as X_1, the second as X_2, etc.

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8

1. **Measures of central tendency**

A number of measures have been developed to show where the 'middle' of a distribution is. These are often referred to as 'measures of central tendency'. They are

Median The score above which half the group scored and below which the other group scores.

Mode The score obtained by the largest number of people. (This is useful only where a lot of people are involved and where large numbers of people obtain the available scores around the middle.)

Mean The arithmetic average, that is, the sum of the scores divided by the number of scores. This can be represented as

$$\text{Mean} = \frac{\Sigma X}{N},$$

where the greek letter sigma (Σ) is used to indicate 'sum of'. Actually, to be precise, we could say X_i is the score for student i (in our class of 6) and represent the sum of the six scores by

$$\sum_{\mathrm{i}=1}^{6} \mathrm{X_i} = X_1 + X_2 + X_3 + X_4 + X_5 + X_6.$$

Later in this Statistics Review we set out some rules for calculations with the sigma notation.

A notation frequently used to indicate a mean is a bar over the symbol for the score. The mean score in our case could be written as $\bar{X}_.$, with the dot replacing the i because the mean represents an average taken over all N (in our case, 6) people. We can best write all this as

$$\begin{aligned}\bar{X}_. &= \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{X_i}}{\mathrm{N}} = \frac{6+4+7+10+7+2}{6}\\ &= \frac{36}{6}\\ &= 6.0.\end{aligned}$$

Of these measures, the median is appropriate in cases where the scale of measurement used is only ordinal (see Chap. 1) or where, even if the scale is interval (or ratio), there are a few very extreme cases which will distort the mean. To see this last point, check that, for this set of scores (2, 5, 7, 8, 10, 11, 40) the median is 8 but the mean is 11.9 (i.e. higher than every score but the top one). The salaries earned by Australians provide a good example of a distribution in which the mean is pulled up above the median by a relatively small number of very high salaries.

For all of the statistics we will be using in this book, the mean is the measure of central tendency we will need.

2. Measures of variability

A number of measures of variability have also been developed. These include the following:

Range — The difference between the highest and the lowest scores. In our original case, this would be 10 − 2 = 8.

Inter quartile Range — The difference between the score which was exceeded by 75% of the students and the score which was exceeded by 25% of them (notice how this pinpoints either side of the median and indicates the degree of variability in terms of the distance between them).

Average Deviation — The average deviation of scores from the mean. If we calculate the deviation of student i as $(X_i - \bar{X}_.)$ it will be positive if the student scores above the mean and negative if he scores below the mean. If we then average all the deviations calculated like this the average will be zero (of course!). Take an example and convince yourself of that. Instead of taking this value for the deviation we take its absolute value. That is, we consider all the deviations to be positive since we are interested in their average size. To indicate that we want the absolute value we write $|X_i - \bar{X}_.|$. The sum of these for all the N students (6 in our example), we would show as

$$\sum_{i=1}^{N} |X_i - \bar{X}_.|.$$

See that you can get this to be 12 for our example.
This then gives us an average deviation of 2.0.

Variance — The average of the square of the deviations of the scores from the mean. For student i, the square deviation will be $(X_i - \bar{X}_.)^2$. For all N students, the sum of squares (SS) will be represented by

$$\sum_{i=1}^{N} (X_i - \bar{X}_.)^2.$$

To take our earlier set of data again and to make this explicit, we could set out the information as

X_i	$(X_i - \bar{X}_.)$	$(X_i - \bar{X}_.)^2$
6	0	0
4	−2	4
7	1	1
10	4	16
7	1	1
2	−4	16
Sum = 36	0	38

Since the sum of squares is 38, the average square deviation is (38/6) = 6.33.
Actually, for small samples we have to divide, not by N (6 in our case), but by N − 1 (i.e. 5), so the variance is 38/5 = 7.6. The reason we divide by (N − 1) is that we are usually not so much trying to work out what the variance is in a particular sample of students, but to estimate what it really is for the population of students which they represent. Dividing by N − 1 gives a better estimate than dividing by N (check that out in one of the books if you want to follow it up). In the meantime, note that the formula for calculating the variance of a set of N scores is

$$\text{Variance} = s^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X}_.)^2}{N - 1},$$

where the symbol s^2 stands for variance.
The calculation can be performed in a slightly simpler way. The sum of squared deviations can be calculated differently because the following relationship holds:

$$\sum_{i=1}^{N} (x_i - \bar{X})^2 = \sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}.$$

Check this out using our example. Square each of the scores and add the squares. That will give you the following:

$$\sum_{i=1}^{N} X_i^2 = 254.$$

You already know that the sum of the numbers is 36, so the formula becomes

$$\sum_{i=1}^{N}(X_i-\bar{X})^2 = 254-\frac{(36)^2}{6} = 254-216 = 38,$$

which is what we got by actually calculating the deviations, squaring them, and adding the squares. In our case, it doesn't matter much which way you do it. If the mean were not a whole number, however, it becomes a bit messy to compute deviations and square them. In such cases the alternate formula just given is easier to use. Try out some examples.

Standard deviation is the square root of the variance. So, where we used the symbol s^2 for variance, we use the symbol s for standard deviation. To return again to our case, where the variance was 7.6, the standard deviation we can see to be the square root of this, which is 2.76.

Of these measures of variability, the one on which we will most depend in this book on measurement is variance. Since the standard deviation is simply the square root of variance we will also be able to use it.

To increase your 'feel' for the sort of calculations we have talked about so far, try adding a constant (say 5) to every student's score and then seeing what the mean and variance will now be compared with the 6.0 and 7.6 we originally obtained (Answer: mean = $\bar{X}$ = 11.0; variance = s^2 = 7.6).

Try again, this time doubling each score and showing that the mean is now twice as high, **viz** 12.0, but that the variance is four times as large, **viz** 30.4.

3. Calculation with the sigma notation

We saw earlier that the greek letter sigma (Σ) is used to indicate 'sum of'. Following are a set of rules for calculations with the sigma notation. These are necessary, for example, when you calculate indices of reliability, fit statistics, etc.

Consider a set of six scores as before:

$$\sum_{\mathrm{i}=1}^{6}\mathrm{X_i} = X_1+X_2+X_3+X_4+X_5+X_6.$$

Rule 1: If c is any constant number then

$$cX_1 + cX_2 + \cdots + cX_n = \sum_{i=1}^{n} cX_i = c\sum_{i=1}^{n} X_i.$$

Rule 2: Also,

$$\begin{aligned}(X_1 + c) + (X_2 + c) + \cdots + (X_n + c) &= \sum_{i=1}^{n} (X_i + c) \\ &= X_1 + X_2 + \cdots X_n) + (c + c + \cdots + c) \\ &= \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} c = \sum_{i=1}^{n} X_i + nc.\end{aligned}$$

Rule 3: Similarly,

$$(X_1 \times X_1) + (X_2 \times X_2) + \cdots + (X_n \times X_n) = X_1^2 + X_2^2 + \cdots X_n^2 = \sum_{i=1}^{n} X_i^2.$$

Rule 4: And

$$\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} X_i\right) = \left(\sum_{i=1}^{n} X_i\right)^2.$$

Rule 5: Also,

$$\sum_{i=1}^{n} (X_i + c)^2 = (X_1 + c)^2 + (X_2 + c)^2 + \ldots (X_n + c)^2.$$

But $(X_i + c)^2$ can be written as $X_i^2 + 2cX_i + c^2$, so

$$\sum_{i=1}^{n} (X_i + c)^2 = \sum_{i=1}^{n} (X_i^2 + 2cX_i + c^2) = \sum_{i=1}^{n} X_i^2 + 2c\sum_{i=1}^{n} X_i + nc^2).$$

Exercises

1. Let $X_1 = 2$, $X_2 = 2$, $X_3 = 3$, $X_4 = 3$, $X_5 = 4$. Calculate each of the following:

   a. $\sum_{i=1}^{5} X_i$,
   b. $\sum_{i=2}^{4} X_i$,
   c. $\sum_{i=1}^{5} X_i - \sum_{i=3}^{5} X_i$.

2. Write out the following expressions:

   a. $\sum_{i=2}^{3} X_i$,
   b. $(\sum_{i=1}^{4} X_i)^2$,
   c. $\sum_{i=1}^{5} X_i^2$.

3. Change the following expressions into Σ notation:

   a. $5X_1 + 5X_2 + 5X_3 + 5X_4$,
   b. $(X_1 + X_2 + X_3 + X_4)^2$.

Statistics Review 2: The Normal Distribution

Key words: Normal curve, Standard normal curve, Skewness, Kurtosis, Standard score *z*

Key points: The normal curve is bell shaped and symmetrical. The standard normal curve has a mean of 0 and SD of 1. Original scores, *x*, in a normal distribution can be converted to standard *z*-scores. The area under the normal curve has been calculated and is both meaningful and useful. In comparing a distribution to the symmetrical normal distribution two terms are sometimes used. Skewness refers to the degree to which the data is distributed either to the left or right of the mean. Kurtosis refers to the 'peakedness' of the data.

The normal curve arises theoretically as the shape when many components, each with an error, are added together. There are many chapters written in statistics texts on the normal distribution, and it is the most important distribution in statistics. The chapter by Roscoe listed under **Further Reading** gives a sound summary of its use. The frequency distributions of many events found in nature closely approximate the normal distribution. Psychological test scores often, but not always, approximate the normal distribution.

Roscoe summarizes the properties of the normal curve as follows:

- A normal curve is a graph of a particular mathematical function—a model for events whose outcome is left to chance.
- There is an infinite number of normal curves, each determined by the values of the mean and SD. The *standard normal curve* has a mean of 0 and an SD of 1.
- A normal curve is symmetrical and bell shaped with its maximum height at the mean.
- A normal curve is continuous.
- The value of Y is positive for all values of X.
- The curve approaches but never touches the X-axis.
- The inflection points of the curve occur 1 SD on either side of the mean.
- The total area between the curve and score scale is 1 and areas under portions of the curve may be treated as relative frequencies.

Converting an original score x_i to a standard score z_i of the standard normal distribution can be done by

$$z_i = \frac{x_i - \bar{x}}{s_x},$$

where $\bar{x}$ and s_x are the mean and SD of the original distribution.
The area under the normal curve has been calculated and tabled. Table 1 at the end of this review provides the proportion of the total area under a normal curve between the mean and any score expressed as a standard score z_i. So given a standard score z_i the percentage of scores between this score and the mean can be determined, or the percentage of scores to the right or left of this score or the percentile rank of the score. The reverse can also be done: given the percentage of scores or a percentile a standard score can be determined.

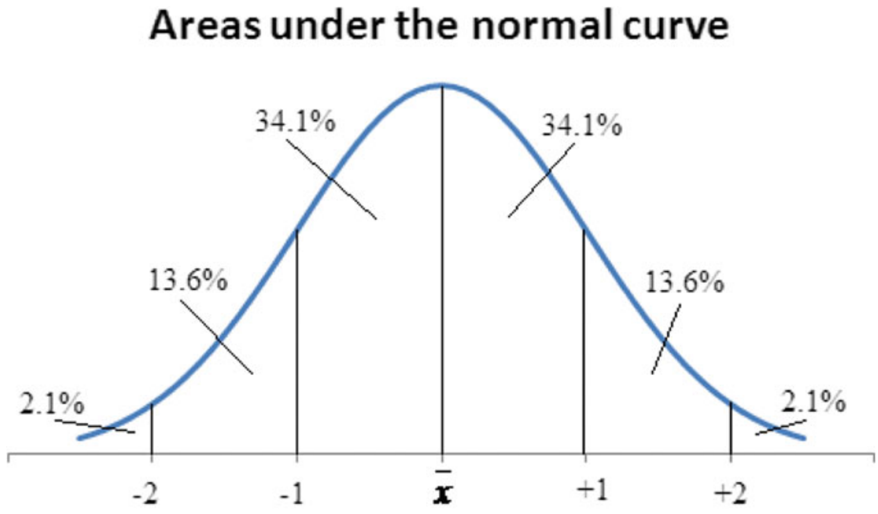
Figure 1 shows two normal curves, differing in their SDs, with the area under them.

The areas under the normal curve can also be found using the Microsoft Excel function normdist(). The function returns the probability of an x-score for a normal distribution. For example, what percentage of scores fall below 65 when $\bar{x} = 50$ *and* $s_x = 10$? *= normdist(65, 50, 10, true) will return the value 0.933. So 93.3% of scores fall below a score of 65. The z-score doesn't have to be calculated when using this function—the z-score is (65 − 50)/10=1.5. So 93.3% of scores fall below a z-score of 1.5. 6.7% of scores fals above a z-score of 1.5. Because the total area under the curve is 1, we need to subtract the area to the left from 1 to get 1 − 0.933 = 0.067. Figure 2 shows the area under the curve graphically.*

It is also very important to be able to use the normal curve to establish *confidence intervals* for the range of observed scores. For example, suppose we know that scores have a mean of 50 and a standard deviation of 10, then we can determine the percentage of scores within any interval, and the interval within which we can confidently say any percentage of scores will fall.

Table 1 Areas under the standard normal distribution between the mean and the z-score

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.004	0.008	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.195	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.219	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.258	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.291	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.377	0.379	0.381	0.383
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.437	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.485	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.489
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.494	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.496	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.497	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.498	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.499	0.499



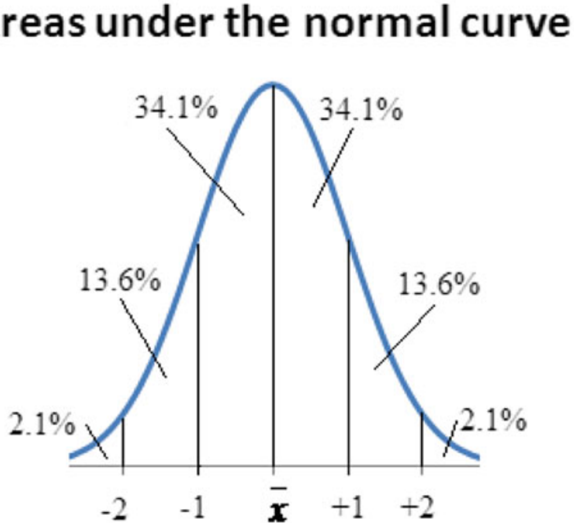


Fig. 1 Two normal curves with the area under them, the curve at the top having a bigger SD

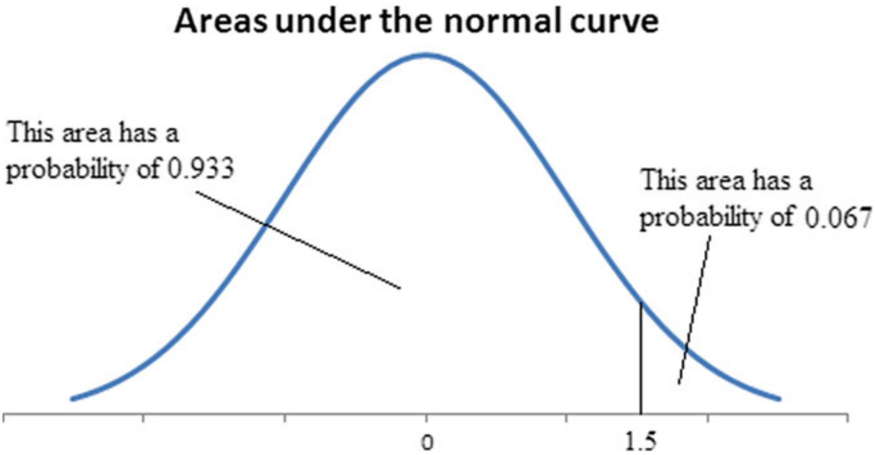


Fig. 2 Areas under the standard normal curve

Example 1. For the case above of the variable where $\bar{x} = 50, s_x = 10$, what percentage of scores falls within the range of 35–65? To be able to look this percentage up from the table, we convert $x_i = 35$ and $x_i = 65$ to standard scores according to

$$\begin{aligned} z_i &= \tfrac{x_i - \bar{x}}{s_x} & z_i &= \tfrac{x_i - \bar{x}}{s_x} \\ &= \tfrac{35-50}{10} & &= \tfrac{65-50}{10} \\ &= \tfrac{-15}{10} & &= \tfrac{15}{10} \\ &= -1.5 & &= 1.5. \end{aligned}$$

From the table for $z_i = 1.5$, the area is 0.4332. Therefore, 43.32% of cases are between the mean and $x_i = 65$, and likewise between the mean and $x_i = 35$. Therefore, the percentage of scores is 43.32 + 43.32 = 86.64.

Example 2. For the case above of the variable where $\bar{x} = 50, s_x = 10$, find the interval within which 90% of the scores fall if the distribution is normal.

We first compute a standardized z-score so that we can refer to the table.

To form 90%, 45% = 0.45 cases are on either side of the mean. For an area of 0.4495 $z_i = 1.64$, for an area of 0.4505 $z_i = 1.65$.

Therefore, we can approximate 1.645 for an area of 0.45. A value of $z_i = -1.645$ would take in the other 0.45 (45%) cases.

We now convert z_i to x_i from

$$
\begin{aligned}
z_i &= \frac{x_i - \bar{x}}{s_x} \\
z_i &= \frac{x_i - 50}{10} \\
x_i - 50 &= 10z_i \\
x_i &= 50 + 10z_i \\
\text{For } z_i = 1.645 \quad x_i &= 50 + 10(1.645) \\
&= 50 + 16.45 \\
&= 66.45 \\
\text{For } z_i = -1.645 \quad x_i &= 50 - 10(1.645) \\
&= 50 - 16.45 \\
&= 34.55.
\end{aligned}
$$

Therefore, the 90% confidence interval is 34.55–66.45.

In comparing a distribution to the symmetrical normal distribution two terms are sometimes used: *skewness* and *kurtosis*. Skewness refers to the degree to which the data is distributed either to the left or right of the mean. Kurtosis refers to the 'peakedness' of the data.

Further Reading

Roscoe, J. T. (1975) *Fundamental Research Statistics for the Behavioural Sciences,* (2nd ed.), New York: Holt, Reinhart and Winston.

Exercises

1. Given a score point located 1.54 SD above the mean in a normal distribution, determine the following:

   a. The area between the mean and the score.
   b. The total area to the left of the score.
   c. The total area to the right of the score.
   d. The percentile rank of the score.

2. Given a percentile rank of 20 in a normal distribution, determine the corresponding z-score.
3. Given a percentile rank of 95 in a normal distribution, determine the corresponding z-score.

Statistics Review 3: Covariance and the Variance of a Sum of Two Variables

Key words: Covariance, Deviation from the mean, Correlated variables, Uncorrelated variables

Key points: Covariance gives an indication of how the variations in a set of scores relate to variations in another set of scores. When variables are correlated, the variance of the sum of two variables is the sum of the variances of the variables and twice the covariance. When variables are not correlated, the variance of the sum of two variables is only the sum of the variances of the variables.

In Statistics Review 1 you learnt about measures of variability, for example, variance. In this Statistics Review, we take the concept of variance further.

1. **Measure of covariability**

In many instances, we are not interested as much in how scores on one test vary as we are in how the variations in scores on that test relate to variations in scores on another test. For example, do the students who get a high score on one test get a high score on the other test too? That is, how do scores on the two tests covary?

When we had scores on one test, we used the square of the deviation from the mean as the basis for our calculation of variance. If we now consider a case where we have two test results, for example, a small 10-item test on arithmetic (that's the one we've been using all along) and a small 5-item test on algebra. Suppose the scores obtained by the six students in the class were

Arithmetic test	Algebra test
X_i	Y_i
6	2
4	0
7	2
10	4
7	3
2	1

To calculate the variance on the arithmetic test we need, for person i, the square deviation $(X_i - \bar{X}_.)^2$. Similarly, for the algebra test, we need $(Y_i - \bar{Y}_.)^2$. Notice that the mean on the algebra test $(\bar{Y}_.)$ is equal to 2.0. To calculate covariance we need, not the square of either one, but their product, i.e. $(X_i - \bar{X}_.)\ (Y_i - \bar{Y}_.)$.

We can set out the necessary calculations using the following table:

	X_i	$(X_i - \bar{X}_.)$	$(X_i - \bar{X}_.)^2$	Y_i	$(Y_i - \bar{Y}_.)$	$(Y_i - \bar{Y}_.)^2$	$(X_i - X_.)(Y_i - \bar{Y}_.)$
	6	0	0	2	0	0	0
	4	−2	4	0	−2	4	4
	7	1	1	2	0	0	0
	10	4	16	4	2	4	8
	7	1	1	3	1	1	1
	2	−4	16	1	−1	1	4
Sum	36	0	38	12	0	10	17

From these you will remember we calculated the variance of X as follows:

$$\begin{aligned} s_x^2 &= \frac{\sum_{i=1}^{N} (X_i - \bar{X}.)^2}{N-1} \\ &= \frac{38}{5} \\ &= 7.6. \end{aligned}$$

We can also calculate the variance for the algebra test (Y) as follows:

$$\begin{aligned} s_y^2 &= \frac{\sum_{i=1}^{N} (Y_i - \bar{Y}.)^2}{N-1} \\ &= \frac{10}{5} \\ &= 2.0. \end{aligned}$$

The covariance between the two tests, X and Y, will be

$$\begin{aligned} C_{xy} &= \frac{\sum_{i=1}^{N} (X_i - \bar{X}.)(Y_i - \bar{Y}.)}{N-1} \\ &= \frac{17}{5} \\ &= 3.4. \end{aligned}$$

Remember the easier calculation procedure for variance? Well, there is a corresponding one for covariance. Instead of actually calculating the deviations from the means and then computing the sum of products of the deviations, you can use the following relationship:

$$\begin{aligned} \sum_{i=1}^{N} (X_i - \bar{X}.)(Y_i - \bar{Y}.) &= \sum_{i=1}^{N} X_i Y_i - \frac{\left(\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{X_i}\right)\left(\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{Y_i}\right)}{\mathrm{N}} \\ &= 89 - \frac{(36)(12)}{6} \\ &= 17. \end{aligned}$$

Check to see that you can get this last result yourself, that is, to make sure that you know what each of the symbols means.

The results which we now have can be summarized in a table form called a variance–covariance matrix. It would be set out as

$$
\begin{array}{lll}
 & & \begin{array}{cc}(\mathrm{X}) & (\mathrm{Y})\end{array} \\
\text{Arithmetic} & (\mathrm{X}) & \multirow{2}{*}{$\begin{bmatrix} 7.6 & 3.4 \\ 3.4 & 2.0 \end{bmatrix}$} \\
\text{Algebra} & (\mathrm{Y}) &
\end{array}
$$

You can read from this matrix that the variance of X (or, if you like, the covariance of X with X) is 7.6, the variance of Y is 2.0, the covariance of X with Y (top right-hand corner) is 3.4 and the covariance of Y with X (bottom left-hand corner) is 3.4 (of course).

2. **Variance of a sum of two variables**

We are sometimes interested in adding together scores obtained on two different tests. It is important to see what happens to the variance when this is done. We can take our two tests, X and Y, to see this [note that the mean of the new variable, (X + Y), is 8.0].

	X_i	Y_i	$(X_i + Y_i)$	$\{(X_i + Y_i) - (\bar{X}_. + \bar{Y}_.)\}$	$\{(X_i + Y_i) - (\bar{X}_. + \bar{Y}_.)\}^2$
	6	2	8	0	0
	4	0	4	−4	16
	7	2	9	1	1
	10	4	14	6	36
	7	3	10	2	4
	2	1	3	−5	25
Sum	36	12	48	0	82

The variance of this variable (X + Y) will therefore be

$$
\begin{aligned}
S^2_{x+y} &= \frac{\sum_{i=1}^{N} \{(X_i + Y_i) - (\bar{X}_. + \bar{Y}_.)\}^2}{N-1} \\
&= \frac{82}{5} \\
&= 16.4.
\end{aligned}
$$

The important question is, how is this related to the variance of X and the variance of Y. We will tell you first and then, for those who are interested in following it, we will prove it. The relationship is

$$
\begin{aligned}
s^2_{x+y} &= s^2_x + s^2_y + 2c_{xy} \\
&= 7.6 + 2.0 + (2 \times 3.4) \\
&= 7.6 + 2.0 + 6.8 \\
&= 16.4.
\end{aligned}
$$

The proof that this relationship always holds is simple enough. If we leave out the (N − 1), by which we divide to convert the sum of squares to the variance, we can develop the algebra for the sum of squares (SS) as follows:

$$\begin{aligned}
SS_{x+y} &= \sum_{i=1}^{N} [(X_i + Y_i) - (\bar{X}_. + \bar{Y}_.)]^2 \\
&= \sum_{i=1}^{N} [(X_i - \bar{X}_.) + (Y_i - \bar{Y}_.)]^2 \\
&= \sum_{i=1}^{N} \left[(x_i - \bar{X}_.)^2 + (Y_i - \bar{Y}_.)^2 + 2(X_i - \bar{X}_.)(Y_i - \bar{Y}_.)\right] \\
&= \sum_{i=1}^{N} (X_i - \bar{X}_.)^2 + \sum_{i=1}^{N} (Y_i - \bar{Y}_.)^2 + 2\sum_{i=1}^{N} (X_i - \bar{X}_.)(Y_i - \bar{Y}_.) \\
&= SS_x + SS_y + 2SP_{xy}.
\end{aligned}$$

In this last expression, SP_{xy} represents the sum of products of deviations for X- and Y-scores about their respective means. If we now divide each item by N − 1, we get the variances from the sums of squares and the covariance from the sum of products and so we have

$$s^2_{x+y} = s^2_x + s^2_y + 2c_{xy}.$$

From this it should be clear that the variance of a set of scores obtained by summing scores is **not** equal to the sum of the variances of the original sets of scores unless, of course, their covariance is zero (More of that, in fact much more of that, in the next section!).

3. Variance of the sum of uncorrelated variables

To understand this section you need to understand the concept of correlation, explained in *Statistics Review 4 Regression and Correlation*. Therefore, the concepts in Statistics Reviews 3 and 4 should be reviewed simultaneously. We begin this section by making tangible the variance of the sum of two uncorrelated variables. It is essential to understand this effect because in TTT we imagine that an observed score is the sum of two variables, the true score and the error, where these two variables are not correlated.

At the end of the previous section, the following result is derived:

$$s^2_{x+y} = s^2_x + s^2_y + 2c_{xy},$$

where c_{xy} is the covariance between two variables.

The covariance is large when the scores on the X variable are related to scores on the Y variable. In particular, if when a score on the X variable is above the mean, the corresponding score on the Y variable is likely to be above the mean, and when a score on the X variable is below the mean, the corresponding score on the Y variable is also

Table 2 Example of uncorrelated variables

X_i	Y_i	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$
10	6	5	1	5
8	4	3	−1	−3
3	7	−2	2	−4
4	4	−1	−1	1
1	5	−4	0	0
4	4	−1	−1	1
$\bar{X} = 5$	$\bar{Y} = 5$	$\sum_{i=1}^{6} (X_i - \bar{X}) = 0$	$\sum_{i=1}^{6} (Y_i - \bar{Y}) = 0$	$\sum_{i=1}^{6} (X_i - \bar{X})(Y_i - \bar{Y}) = 0$

likely to be below the mean, then the covariance will be large and positive. If the scores on the two variables go in the opposite directions, that is, when a score on the X variable is above the mean, the corresponding score on the Y variable is likely to be below the mean, and when a score on the X variable is below the mean, the corresponding score on the Y variable is also likely to be above the mean, then the covariance will be large and negative.

When one cannot tell which direction the value of the second variable will take from the value of the first variable, then the covariance will be close to 0: $c_{xy} = 0$. To consolidate this point, Table 2 shows an example of two variables that are uncorrelated.

From section 1 where you learned the definition of covariance, you can see from the above table that

$$c_{xy} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{N - 1} = 0. \qquad r_{xy} = \frac{c_{xy}}{s_x s_y} = \frac{0}{s_x s_y} = 0.$$

Of course, the correlation r_{xy} is also 0 as shown.

In this case, from the equation the variance for the sum of two variables, which you learned in the previous section to be given by

$$\begin{aligned} s_{x+y}^2 &= s_x^2 + s_y^2 + 2c_{xy} \quad \text{reduces to} \\ s_{x+y}^2 &= s_x^2 + s_y^2 \qquad\qquad (C_{xy} = 0). \end{aligned}$$

This is a very important result: *when variables are not correlated, the variance of the sum of two variables is the sum of the variances of the variables.*

This can be checked readily in the above example.

Let Z be a new variable which is the sum of the original variables X and Y:Z = X + Y. Table 3 shows the variable Z as well as the squares about the mean for each of the variables from which the variances are calculated.

Table 3 Sum of two uncorrelated variables

$Z_i = X_i + Y_i$	$Z_i - \bar{Z}$	$(Z_i - \bar{Z})^2$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
16	6	36	25	1
12	2	4	9	1
10	0	0	4	4
8	−2	4	1	1
6	−4	16	16	0
8	−2	4	1	1
$\bar{Z} = 10$	$\sum_{i=1}^{6} (Z_i - \bar{Z}) = 0$	$\sum_{i=1}^{6} (Z_i - \bar{Z})^2 = 64$	$\sum_{i=1}^{6} (X_i - \bar{X})^2 = 56$	$\sum_{i=1}^{6} (Y_i - \bar{Y})^2 = 8$

It is evident that from Z = X + Y, when X and Y are not correlated that 64 = 56 + 8, that is,

$$\sum_{i=1}^{6} (Z_i - \bar{Z})^2 = \sum_{i=1}^{6} (X_i - \bar{X})^2 + \sum_{i=1}^{6} (Y_i - \bar{Y})^2,$$

$$\text{i.e. } \mathrm{SS}_{\mathrm{z=x+y}} = \mathrm{SS}_{\mathrm{x}} + \mathrm{SS}_{\mathrm{y}},$$

$$\text{i.e. } \frac{SS_{z=x+y}}{5} = \frac{SS_x}{5} + \frac{SS_y}{5},$$

$$\text{i.e. } s^2_{z=x+y} = s^2_x + s^2_y.$$

Exercises

1. Suppose the variables of height and ability on a non-verbal test of intelligence are uncorrelated in an *adult* population.
   (a) If the standard deviation of height is 10 cm and the standard deviation on the non-verbal test of intelligence is 15 units, what would be the variance of the sum of height and the variable on non-verbal intelligence in this population?
   (b) Would adding these two variables to form the sum make any sense? Explain your answer in one sentence.

2. Suppose the variables of height and ability on the same non-verbal test of intelligence is correlated in a school-age population (during compulsory years of schooling age approximately 5–15 years).
   (a) If the correlation is 0.5, the standard deviation of heights is 12 and the standard deviation on the non-verbal test of intelligence is 18 units, what would be the variance of the sum of height and the variable on non-verbal intelligence in this population?
   (b) Would adding these two variables to form the sum make any sense? Explain your answer in no more than one sentence.
   (c) Why do you think that there is a correlation between height and non-verbal intelligence in the school-age population and not in the adult population?

Statistics Review 4: Regression and Correlation

Key words: Relationship between variables, Linear relationship, Regression, regression coefficients, Prediction, Error of prediction, Correlation, Proportion of variance accounted for, Strength of relationship

Key points: Regression and correlation give an indication of the relationship between variables. If there is a linear relationship between two variables a score on one variable can be predicted based on a score on the other variable through regression. Correlation gives an index of the strength between two variables. Correlation also gives an indication of the proportion variance in one variable that can be accounted for by the other variable. Correlation is the standardized covariance between two variables.

Relationship between two variables: In educational (or any social or behavioural) research, one of the things we often want to do is to make predictions. For example, if we want to allocate places in a tertiary institution among a large number of applicants, we will usually want to allocate places to those students for whom we predict the greatest likelihood of success in tertiary study. We can check out the accuracy of our predictions by getting a measure of tertiary success and seeing how well it does actually relate to our basis of prediction.

1. **A simple linear model**

To illustrate this, let's stick with our simple data for our six students on the arithmetic and algebra test (used in Statistics Review 1) and let's imagine that we are interested in seeing what predictions of algebra performance we can make from arithmetic performance.

At first sight, it might seem worthwhile aiming for a deterministic model such as those developed in the physical sciences. Newton's law, for example, declares that a force, F, where exerted on a body of mass, M, moves it with acceleration, a, where the relationship between these is F = Ma. This is a statement of a precise relationship.

We could try an equation of this form for our prediction, say

$$Y = \beta_0 + \beta_1 X \tag{1}$$

but the problem with such a model is that it suggests that, for a particular value of X, we can predict precisely the value of Y. If we graph our data from the arithmetic and algebra test we can see the problem (Fig. 3).

If Eq. (1) were to apply, it would be possible to draw a line on which all of the points would lie. The best we can do is to draw a line which fits the points as well as possible, and then to use this line for making our predictions. That is, for student i, with an arithmetic score of X_i (a score of 10 for student 4) we could predict an algebra score $\hat{Y}_i$ (with the ^ indicating that we are speaking of a prediction, or an estimate) using the following equation:

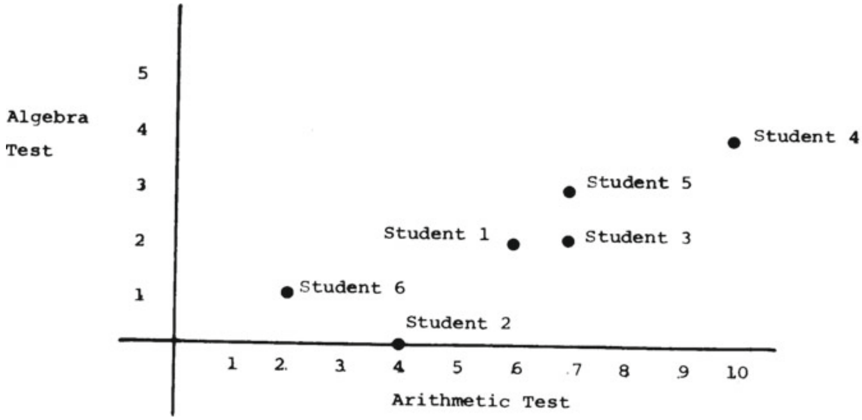


Fig. 3 Distribution of sample scores on arithmetic and algebra tests

$$\hat{Y}_i = \beta_0 + \beta_1 X_i. \tag{2}$$

Jumping ahead a bit, let us tell you that the prediction we will get for student 4, based on his score of 10 on the arithmetic test, will be 3.79 whereas that student's algebra score was actually 4. So our prediction would be out a bit, by an error of prediction of 0.21. We can say this more formally as

$$Y_i - \hat{Y}_i = e_i. \tag{3}$$

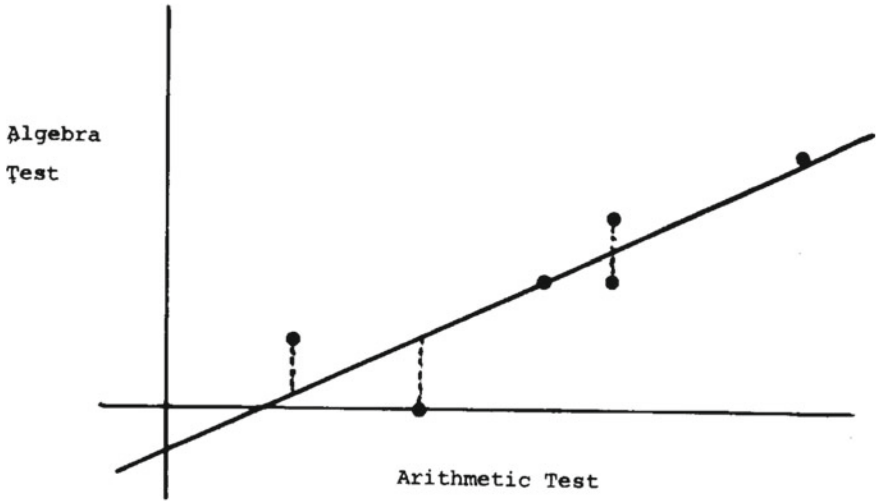


Fig. 4 Regression line for predicting algebra tests scores from arithmetic tests scores

For the case of student 4, this Eq. (3) would be

$$\begin{aligned} Y_4 - \hat{Y}_4 &= 4.00 - 3.79 \\ &= 0.21 \\ &= e_4. \end{aligned}$$

The sort of prediction we are actually making is one in which we freely admit that there will be errors of prediction. We can make that clear by writing Eq. (1), not as a deterministic model, but as a probabilistic model, with the error of prediction included

$$Y = \beta_0 + \beta_1 X + e. \tag{4}$$

The question we haven't considered yet is how the values of β_0 and β_1 are determined. The full details of that are given in the statistics texts, to which you should refer if you want to follow the derivation. Let us just say here that the values we use are the **best** ones—best in the sense that when we use them we keep our errors of prediction as small as possible. Actually, what we keep as small as possible is the sum of squares of our errors, i.e. $\sum e_i^2$ or $\Sigma\left(\mathrm{Y_i} - \hat{\mathrm{Y}}_\mathrm{i}\right)^2$.

All you really need to know is how to calculate these **best** values of β_0 and β_1. They will only be estimates of the values for the whole population of students, because we derive them from the sample of data available (and our sample is small enough, containing only six students). We could write $\hat{\beta}_0$ and $\hat{\beta}_1$ to make it clear we are working with estimates of the population values but a simpler convention is to use the Roman letter b which corresponds to the Greek letter β.

The estimates, then, are

$$b_1 = \frac{c_{xy}}{s^2 x} \tag{5}$$

and

$$b_0 = \bar{Y}_. - b_1 \bar{X}_.. \tag{6}$$

For our sample, then,

$$b_1 = \frac{3.4}{7.6} = 0.447$$

and

$$b_0 = 2.0 - (0.447 \times 6.0) = -0.68.$$

If we draw the line $\hat{Y} = -0.68 + 0.447X$ on the graph, the graph is Fig. 4.

We could read the errors of prediction from the graph, if we wanted to do it carefully, or we can calculate them directly. They are given in the following table, in which $\hat{Y}_i$ has been calculated from

$$\hat{Y}_i = -0.68 + 0.447\mathrm{X_i}. \tag{7}$$

X_i	$\hat{Y}_i$	Y_i	$(Y_i - \hat{Y}_i)$	$(Y_i - \hat{Y}_i)^2$
6	2.00	2	0.00	0.00
4	1.10	0	−1.10	1.22
7	2.45	2	−0.45	0.20
10	3.79	4	0.21	0.04
7	2.45	3	0.55	0.31
2	0.21	1	0.79	0.62

The errors of prediction $(Y_i - \hat{Y}_i)$ shown in this table correspond with the dotted lines shown in the graph to indicate the distance a particular point is from the prediction line. For example, for a score of 7 on the arithmetic test, we would predict a score of 2.45 on the algebra test. One student with 7 on the arithmetic test actually got 3 on the algebra test so we would have under-estimated his score (error of prediction 0.55). The other student with 7 on the arithmetic test got 2 on the algebra test, so we would have over-estimated his score (error of prediction −0.45).

1.1. **Error variance in prediction**

If you look back to the last table, you will see that the right-hand column gives the square of the deviation of each score from the regression line, i.e. the square of the error of prediction. The sum of these squares is

$$\sum_{i=1}^{6} (Y_i - \hat{Y}_i)^2 = 2.39.$$

To save the separate calculation of a predicted score, and error of prediction, for each student, the following formula can be used:

$$\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 = SS_y - \frac{SP_{xy}^2}{SS_x}. \tag{8}$$

You can find the proof of this in a statistics text if you would like to see it.

Referring back to Statistics Review 1, you will find that $SS_y = 10$, $SS_x = 38$, and $SP_{xy} = 17$. Substituting these in Eq. (8) we get

$$\begin{aligned} SS_y - \frac{\left(SP_{xy}\right)^2}{SS_x} &= 10 - \frac{(17)^2}{38} \\ &= 10 - 7.61 \\ &= 2.39 \end{aligned},$$

which is exactly what we got from the full calculation of the sum of squares.

A useful notation for this sum of squared errors of prediction is $SS_{y.x}$. Whereas SS_y indicated the sum of squared deviations of individual Y_i scores around $\bar{Y}_.$, the symbol $SS_{y.x}$ indicates the sum of squared deviations of individual Y_i scores that could have been expected on the basis of the individual's X_i score.

If we now talk in terms of variance and covariance rather than sums of squares and products, we can express a similar relationship to (8) as

$$s_{y.x}^2 = s_y^2 - \frac{c_{xy}^2}{s_x^2} \tag{9}$$

in which $s_{y.x}^2$ is the variance error of prediction of Y from X. From the sum of squares $\left(SS_y = 2.39\right)$ we can, by dividing by N − 1 (i.e. 5), obtain

$$s_{y.x}^2 = 0.48.$$

We could also get this value by substituting directly into Eq. (9).

1.2. A curvilinear relationship

If the regression weight β_1 is zero, it means that, using our simple linear model, there can be no useful prediction of Y-scores from the X-scores. It may mean that there is, in fact, no relationship between the two variables at all. On the other hand, it is always possible that there is a curvilinear relationship of a type which shows no linear relationship. In Fig. 5 such a relationship is shown. The linear regression line is horizontal $(\beta_1 = 0)$ and suggests no relationship on which to base predictions. A more complex model would be needed to deal with this case.

This type of example illustrates the value of inspecting data carefully and not just pumping it through statistical analyses.

2. Correlation

2.1. Proportion of variance accounted for

When we are studying the relationship between two variables, it is often helpful to have an index of the strength of the relationship between the two variables. Obviously, $s_{y.x}^2$ gives us a very good indication of how strong the relationship is because it is a direct measure of variance of those components of each person score on one variable (Y) which cannot be predicted from (i.e. explained by) the other variable (X).

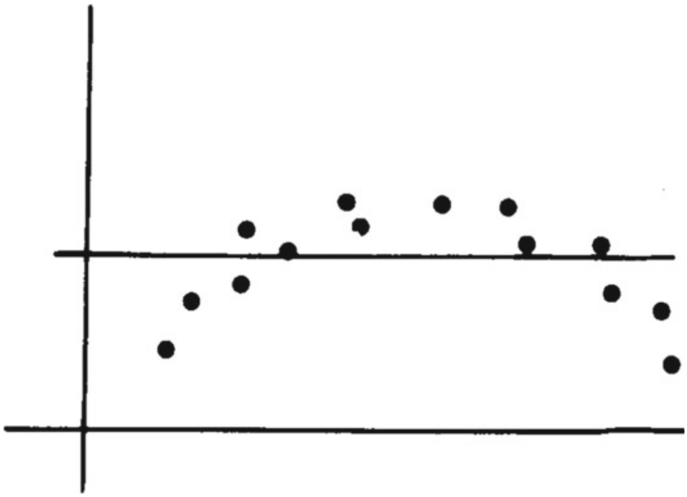
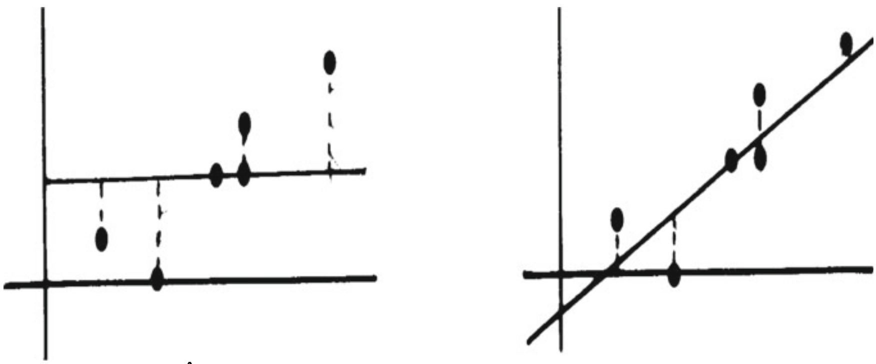


Fig. 5 Data revealing a curvilinear relationship

If the relationship between the two variables were perfect, every person's point on a graph such as the one in Fig. 4 would lie on the line. A person's Y-score could be predicted precisely from his X-score in such a case so there would be no error of prediction. In such a case, $s^2_{y.x} = 0$.

We are very unlikely to find $s^2_{y.x} = 0$, so we really need to have some idea how to judge the values we actually obtain. The important thing is to compare $s^2_{y.x}$ with s^2_y. This sort of comparison is provided, for our continuing example of arithmetic and algebra scores, in Fig. 6.



(a) $\hat{Y} = \overline{Y}_{\bullet}$

$= 2.0$

(b) $\hat{Y} = b_0 + b_1 X$

$= -0.68 + 0.447X$

Fig. 6 Variance accounted for by correlation

In graph (a) in Fig. 6, the variability of each student's Y-score from the mean Y-score $(\bar{Y} = 2.0)$ is shown. We know, from our calculation in Statistics Review 1, that the variance of these scores is

$$s_y^2 = 2.0.$$

In graph (b) in Fig. 6, the variability of each student's Y-score from the score which could have been predicted for him (i.e. $Y_i - \hat{Y}_i$) is shown. We know from our earlier calculation in this Statistics Review that the variance of these deviations from the regression line is

$$s_{y.x}^2 = 0.48.$$

The proportion of error variance to total variance is

$$\frac{s_{y.x}^2}{s_y^2} = \frac{0.48}{2.00} = 0.24.$$

This is unexplained variance. The proportion of explained variance is then given by

$$\mathrm{r}^2 = 1 - \frac{\mathrm{s}_{\mathrm{y.x}}^2}{\mathrm{s}_{\mathrm{y}}^2} \left(\text{you will see why } \mathrm{r}^2 \text{ is used next page}\right). \tag{10}$$

For our example, this is

$$\begin{aligned} \mathrm{r}^2 &= 1 - \frac{0.48}{2.00} \\ &= 0.76. \end{aligned}$$

This tells us that 0.76 of all the variance in the algebra scores (Y) can be explained by the arithmetic scores (X). The proportion of the variance unexplained is

$$\frac{s_{y.x}^2}{s_y^2} = 0.24.$$

If the relationship between the two variables were perfect, $s_{y.x}^2 = 0$ and $r^2 = 1$. In this case, the proportion of the variance in Y accounted for by X would be 1.0.

If there were absolutely no (linear) relationship between the two variables, $s_{y.x}^2 = s_y^2$ and the proportion of variance in Y accounted for by X would be 0.0.

2.2. **Index of strength of relationship**

The coefficient of correlation between two variables has been developed as an index of the strength of the relationship between the variables. The most commonly used coefficient is the Pearson product moment correlation coefficient, for which the formula is

$$r = \frac{c_{xy}}{s_x s_y}. \tag{11}$$

From Statistics Review 3 you will remember that c_{xy} is a measure of the co-variation of two variables. The difficulty in interpreting it is that we have no rules for judging its size. If variable X or Y is measured on a large scale (say one with 100 points) the covariance could be large, but so would the variance be. What the formula (11) does is to standardize the covariance so that the effects of the scale size are removed. You will notice that the covariance is divided by the product of the standard deviations of the two variables.

To return to our example, again, we find the correlation to be

$$\begin{aligned} r &= \frac{3.4}{(2.76)(1.41)} \\ &= 0.87. \end{aligned}$$

If two variables are perfectly related, then $r = 1.0$ (or -1.0 if high scores on one match low scores on the other and vice versa). If there is no relationship between the variables, $c_{xy} = 0$ and so $r = 0$.

Note that r^2 calculated from above is $r^2 = 0.87^2 = 0.7569 \simeq 0.76$. This is identical to r^2 in the previous section.

Exercises

The data in the table below represent the scores of a small group of students on a selection text (X) and a subsequent criterion test (Y).

Student number	Selection test X	Criterion test Y
1	45	69
2	33	55
3	31	46
4	27	35
5	48	58
6	40	60
7	51	72
8	37	53
9	42	64
10	36	48

1. Calculate the covariance and the correlation of the scores on the two tests.
2. Calculate the best estimates for the regression coefficients for the prediction of the criterion test scores from the selection test scores.
3. What would be the variance of the errors of prediction made with the regression equation in the answer to 3?
4. What would be the error of prediction for Student 5?

Statistics Review 5: Probability

Key words: Set, Members, Sample space, Event, Compound events, Probability of compound events, Probability, Probability of a compound event, Odds

Key points: A *sample space* is the set of all possible outcomes. An *event* is a set of outcomes which is a subset of the sample space. A *probability* of an outcome is the *theoretical proportion* of times that we would expect that outcome to occur on a very large number of replications. The probability of an event occurring is the *ratio* of the number of occurrences of the event and the total number of equally likely occurrences. The probabilities of compound events, such as the union of two sets or the intersection of two sets, can be determined. *Odds* and *probabilities* are analogous to *proportions* and *ratios*.

Reviews 5 and 8 deal with probability, a topic which is best handled using set theory so Statistics Review 5 commences with an introduction to the language of set theory.

1. **Set Theory**

1.1. **Definition of Set and Members**

Any well-defined collection (or list) of things can be called a **set**. The ‘things’ in the set are its **elements** or **members**. A set can be defined by listing its elements or stating its properties. So,

$$\mathrm{A} = \{2, 4, 6, 8\}$$

and A = {x: x is a positive even number, x < 10}

both define the same set. We can note that 4 is a member of this set but that 7 is not. These observations can be expressed with the nomenclature

$$4 \in \mathrm{A} \quad \text{and} \quad 7 \notin \mathrm{A}.$$

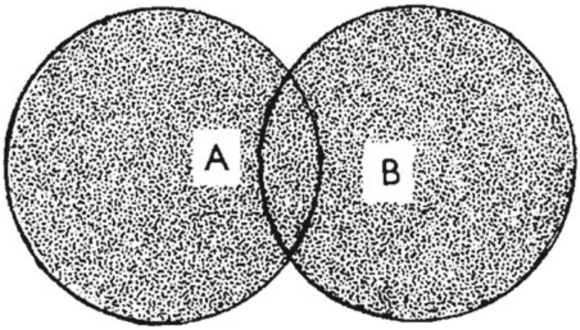
If all the elements of A also belong to some other set B, then A will be a subset of B. This can be expressed as

$$\mathrm{A} \subset \mathrm{B}.$$

Of course A may actually equal B in this case in which it would also be true to say $\mathrm{B} \subset \mathrm{A}$.

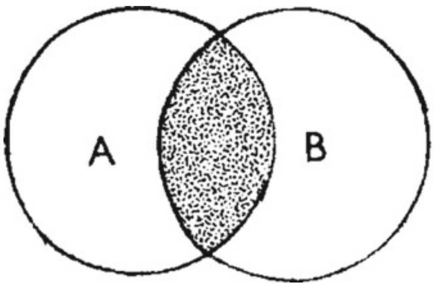
1.2. **Set Operations**

Union: The union of two sets A and B is the set of all elements which belong to A or B or both A and B. It is denoted $A \cup B$ and can be illustrated by the following Venn diagram:



$A \cup B$ is shaded

Intersection: The intersection of two sets A and B is the set of all elements which belong to both A and B. It is denoted $A \cap B$ and can be illustrated by the following Venn diagram:



$A \cap B$ is shaded

Under these operations, various algebraic laws hold. They are
Associative laws:

$$(A \cup B) \cup C = A \cup (B \cup C) \qquad (A \cap B) \cap C = A \cap (B \cap C). \tag{12}$$

Commutative laws:

$$A \cup B = B \cup A \qquad A \cap B = B \cap A. \tag{13}$$

Distributive laws:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C). \quad (14)$$

Draw Venn diagrams to illustrate each of these laws to make sure that you see the point of each one of them.

2. **Probability**

2.1. **Notion of Probability**

Probability involves the study of random events and began as the study of games of chance. A probability of an outcome is the *theoretical proportion* of times that we would expect that outcome to occur on a very large number of replications.

The probability of some event A occurring was defined in terms of the total number of equally likely occurrences (n) and the total number of them (a) which would be considered occurrences of A. The probability of event A then was defined as

$$P\{A\} = \frac{a}{n}.$$

For example, if a six-sided die was tossed, there would be six equally likely occurrences (provided the die was not loaded). Any one of the numbers 1–6 could come up. Thus, n = 6.

If we were interested in the throw of an odd number, such an event (A) could occur as the throw of 1, 3 or 5. Thus, a = 3 and so the probability of throwing an odd number would be

$$P\{A\} = \frac{3}{6} = \frac{1}{2}.$$

If we were interested in the throw of a number greater than 4, such an event (B) could occur as the throw of 5 or 6. Thus, b = 2 so the probability of throwing a number greater than 4 would be

$$P\{B\} = \frac{2}{6} = \frac{1}{3}.$$

[NOTE: There is a circularity in these classical definitions of probability. The probability of throwing an odd number is said to be $\frac{1}{2}$ because it can occur in any one of 3 ways out of 6 'equally likely' ways. But equally likely events are events with equal probability of occurrence, and hence the circularity. In modern probability theory, the development is axiomatic. Probabilities of certain events must satisfy certain axioms. Within this framework, classical probability theory covers just the special cases for which all individual events are equiprobable.]

2.2. **Sample Spaces and Events**

The *sample space* is the set S of all possible outcomes. It is sometimes called an *outcome space*. An *event* A is a set of outcomes which is a subset of the sample space. That is, $A \subset S$.

Two events are *mutually exclusive* if they are disjoint, i.e. if their intersection is the empty set Ø. $A \cap B = \emptyset$ indicates that A and B are mutually exclusive.

If we were interested in the probability of throwing an odd number with a six-sided die, we can define

$$\begin{aligned} S &= \{1, 2, 3, 4, 5, 6\} \\ A &= \{1, 3, 5\}. \end{aligned}$$

Since all occurrences are equiprobable we need only the number of occurrences in A and S to calculate the probability of event A occurring. That is,

$$\begin{aligned} P\{A\} &= \frac{\text{number of elements in A}}{\text{number of elements in S}} \\ &= \frac{n\{A\}}{n\{S\}} \\ &= \frac{3}{6} \\ &= \frac{1}{2}. \end{aligned} \tag{15}$$

2.3. **Probability of Compound Events**

In the discussion of set theory, you encountered two types of compound events, viz. the union of two sets and the intersection of two sets. We now consider the probabilities of such compound events.

Probability of $A \cap B$

If you were selecting cards from a deck of 52 cards, for sample space S would be the whole deck of 52 cards. That is, n{s} = 52. You may be interested in two particular events, viz.

$$A = \{\text{card is a spade}\}$$

for which n{A} = 13. This is because there are 13 spades in the deck, and

$$B = \{\text{card is a face card, i.e. jack, queen or king}\}$$

for which n{B} = 12. This is because there are three face cards in each of the four suits.

The probability of selecting a spade would be

$$P\{A\} = \frac{n\{A\}}{n\{S\}} = \frac{13}{52} = \frac{1}{4}.$$

The probability of selecting a face card would be

$$P\{B\} = \frac{n\{A\}}{n\{S\}} = \frac{12}{52} = \frac{3}{13}.$$

The intersection of these two events, $A \cap B$, is the set containing all elements which belong in both A and B. This is the set containing the jack, queen and king of spades. The fact that there are 3 occurrences in this set can be declared as

$$n\{A \cap B\} = 3.$$

The probability of this compound event is

$$\begin{aligned} P\{A \cap B\} &= \frac{n\{A \cap B\}}{n\{S\}} \\ &= \frac{3}{52}. \end{aligned} \tag{16}$$

Probability of $A \cup B$

The union of the two events A and B is the set containing all 13 spades plus the 9 face cards from the other three suits.

That there are 22 occurrences in this union of the two events could be represented as

$$n\{A \cup B\} = 22.$$

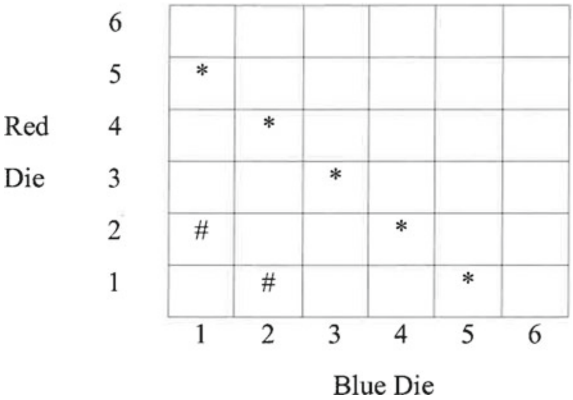
Notice that the number of occurrences is not 13 + 12 = 25 because that would count the three spade face cards twice.

The probability of this compound event is

$$\begin{aligned} P\{A \cup B\} &= \frac{n\{A \cup B\}}{n\{S\}} \\ &= \frac{22}{52} = \frac{11}{26}. \end{aligned} \tag{17}$$

A second example, which can readily be diagrammed, may reinforce the points made in the last example.

If you were throwing two dice, one red and one blue, what would be the probability of the total score on the two dice being 3 or 6? We could call event C a score of 3 and event D a score of 6. The relevant score combinations for both events are marked on the diagram below, an '#' for event C and a '*' for event D.



The diagram makes it clear that the sample space S contains 36 possible combinations of scores on the two dice, i.e. n{S} = 36. The event C is

$$C = \{(1,2),(2,1)\},$$

which is the set containing the two ways in which a score of 3 can be thrown. That is n{C} = 2 so the probability of throwing a score of 3 is

$$P\{C\} = \frac{2}{36}.$$

Event D is the set of outcomes giving a total score of 6. These are shown in the diagram with asterisks (*). Event D thus is the set:

$$D = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}.$$

For this event n{D} = 5, so the probability of throwing a total score of 6 is

$$P\{D\} = \frac{5}{36}.$$

The union of these two events, $C \cup D$, will be the set of all 7 occurrences marked on the diagram. That is, $n\{C \cup D\} = 7$. So the probability of a score of 3 *or* 6 being thrown will be

$$\begin{aligned} P\{C \cup D) &= \frac{7}{36} \\ &= \frac{2}{36} + \frac{5}{36}. \end{aligned}$$

In this case, since no occurrences are common to both events, the simple rule which applies is

$$P\{C \cup D\} = P\{C\} + P\{D\}. \tag{18}$$

Events C and D are *mutually exclusive.* In the earlier case of selecting a spade or a face card, the two events were *not mutually exclusive* since there were occurrences common to both events. Indeed, we have seen that

$$P\{A \cap B) = \frac{3}{52}.$$

In fact, we could have obtained the probability of the union of the two events A and B as

$$\begin{aligned} P\{A \cup B\} &= P\{A\} + P\{B\} - P\{A \cap B) \\ &= \frac{13}{52} + \frac{12}{52} - \frac{3}{52} \\ &= \frac{22}{52}. \end{aligned} \tag{19}$$

In the case with the red and blue dice, throwing 3 or 6 was mutually exclusive events so $P\{C \cap D\} = 0$ which allowed us to assert that $P\{C \cup D\} = P\{C\} + P\{D\}$.

Exercises

1. If A = {1, 2, 3, 4}, B = {2, 4, 5, 8} and C = {3, 4, 5, 6}, what are
   (a) $A \cup B$,
   (b) $B \cap C$,
   (c) $A \cup (B \cap C)$,
   (d) $(A \cup B) \cap C$.
2. If a coin and a die are tossed, the sample space would consist of 12 elements

$$S = (H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6).$$

   (a) How would you represent the following events, including '1' as a prime number so that prime numbers are {1, 2, 3, 5}?

      A = (a prime number appears on the die),
      B = (heads and an odd number appear),
      C = (tails and an even number appear),

   (b) What is the probability that A or B will occur?
   (c) Are any of the events, A, B or C, mutually exclusive?

3. If three coins are tossed
   (a) How many possible outcomes are there, assuming no coin stands on its edge?
   (b) What is the probability of
      (i) three tails,
      (ii) two tails,
      (iii) at least two tails,
   (c) Why is the answer to (b)(iii) the sum of the answers to (b)(i) and (b)(ii)?

Statistics Review 6: Indices

Key words: Index, Raise to the power, Base number, Base number e, Exponential power series

Key points: Numbers can be expressed a powers of a base. An *index* indicates the power to which a number has been raised. One non-integer base of special interest is e. Its value is approximately 2.71828. The exponential power series $y = e^x$ is useful in describing numerous natural processes. If two numbers are expressed as powers of the same base, their *multiplication* can be achieved by *addition of indices*. If two numbers are expressed as powers of the same base, their *division* can be achieved by *subtraction of indices*.

The material in this Statistics Review is straightforward. It should serve simply to remind you of some relationships which are used in logarithms.

1. **Expressing Numbers as Powers of a Base**

The simplest cases of indices are those which are whole numbers (integers). For example, if we square 4 and obtain 16, we can represent this as

$$4^2 = 16$$

where 2 is the index indicating the power to which 4 is raised.
The series of integer indices with 4 as the base is

$$4^1 = 4; \quad 4^2 = 16; \quad 4^3 = 64; \quad 4^4 = 256; \qquad \text{etc.}$$

With 10 as the base, the series is

$$10^1 = 10$$
$$10^2 = 100$$
$$10^3 = 1000$$
$$10^4 = 10{,}000$$
etc

Any base can be raised to powers other than integers, though we don't have easy ways to work out the answers except perhaps by graphing the function. Taking 2 as the base, we could draw a graph using the following series:

$$2^1 = 2; \quad 2^2 = 4; \quad 2^3 = 8; \quad 2^4 = 16; \quad 2^5 = 32; \qquad \text{etc.}$$

Such a graph is shown in Fig. 7.

If we wanted to know the value of $2^{3.4}$, we could use this graph to read it off. Following the line drawn vertically from 3.4 on the x-axis we see that it strikes the curve at about 10.5 or 10.6 on the y-axis. In fact, there are calculators from which this could be obtained exactly as

$$2^{3.4} = 10.556.$$

The point of this illustration is to show that a base can be raised to a power which is not an integer. We have seen it for the base 2, but we could also express other bases to non-integer powers. For example,

$$\begin{aligned}
10^2 &= 100\\
10^{2.1} &= 125.89\\
10^{2.2} &= 158.49\\
10^{2.3} &= 199.53\\
10^{2.4} &= 251.19\\
10^{2.5} &= 316.23\\
10^{2.6} &= 398.11\\
&\text{etc}
\end{aligned}$$

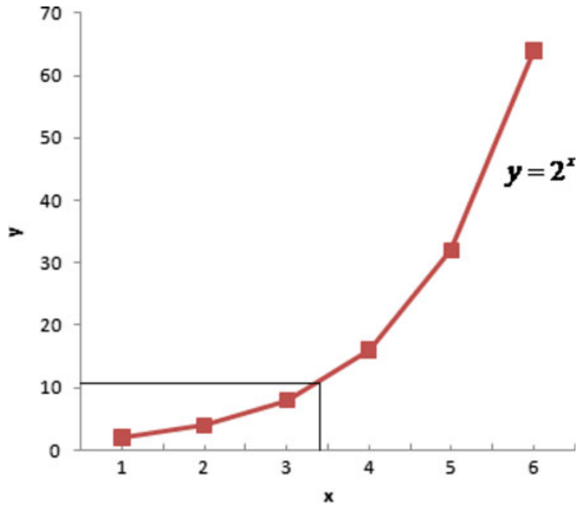


Fig. 7 Graph of powers of two

2. **Multiplying Numbers or Adding Indices**

If two numbers are expressed as powers of the same base, their multiplication can be achieved by addition of indices. For example,

$$4 \times 16 = 64,$$

which could instead be represented as

$$\begin{aligned} 2^2 \times 2^4 &= 2^{(2+4)} \\ &= 2^6 \\ &= 64. \end{aligned}$$

It is not immediately obvious why one might want to add indices rather than to multiply the numbers, but there are advantages in aspects of latent trait test theory.

3. **Dividing Numbers or Subtracting Indices**

For division, we reverse the process of multiplication and, with indices, subtract. Using the example above, we have

$$\frac{64}{16} = 4,$$

which could be represented instead as

$$\begin{aligned} \frac{2^6}{2^4} &= 2^{(6-4)} \\ &= 2^2 \\ &= 4. \end{aligned}$$

By considering examples of division, we can arrive at straight forward definitions of the power of zero and negative powers. We know that, from our rules for division

$$\frac{64}{64} = 1.$$

From our rules of subtraction of indices, we would express this as

$$2^{(6-6)} = 2^0 = 1.$$

This would happen whatever base we used. That is, $10^0 = 1$ and $4^0 = 1$ and so on. More generally, $x^0 = 1$ where x is any number except zero.

From our rules of division, we would also write

$$\frac{16}{64} = \frac{1}{4} = \frac{1}{2^2}.$$

From our rules of subtraction of indices, we would represent this as

$$2^{(4-6)} = 2^{-2}.$$

Since the two are equivalent we can say, by definition, that

$$2^{-2} = \frac{1}{2^2}.$$

With these definitions of negative and zero powers, we could extend the graph in Fig. 7. The curve would extend to the left, flattening out and reaching y = 0 when x = –∞. The value of y will never be negative.

4. **Using the Base e**

In the examples so far we have used indices which are not integers, but the base has always been an integer. It need not be. For example, we could write relationships such as the following:

$2^3 = 8$	$2.1^3 = 9.261$	$2.326^3 = 12.584$
$2^{3.4} = 10.556$	$2.1^{3.4} = 12.461$	$2.326^{3.4} = 17.639$
$2^4 = 16$	$2.1^4 = 19.448$	$2.326^4 = 29.271$

These illustrations were selected arbitrarily to reveal the pattern. They were obtained with a calculator.

There is one non-integer base which is of special interest. It is 2.71828 (approx.) and is represented by e. The exponential power series

$$y = e^x$$

is useful in describing numerous natural processes, including rates of radioactive decay. Taking just some powers, the values are

$$e^0 = 1$$
$$e^1 = 2.718$$
$$e^2 = 7.389$$
$$e^3 = 20.086$$
$$e^4 = 54.598$$

Exercises

Express the following terms in their simplest form:

(a) $3^x \cdot 3^{2x}$,

(b) $\left((a)^4\right)^2$,

(c) $4^a \div 2^a$,

(d) $10^{2n} \cdot 25^n \div 20^n$,

(e) $\frac{2^{(n+3)}.6^{(2-n)}}{15^{(-n-1)}.5^{(n+1)}}$.

Statistics Review 7: Logarithms

Key words: Index, Logarithm, Natural logarithm, Logarithmic scale, Ratio, interval

Key points: A logarithm is an index. Logarithms to base e are referred to as natural logarithms. Instead of $\log_e$, the abbreviation ln is often used. A logarithmic scale expresses equal ratios on the original scale with equal intervals on the logarithmic scale. Indices and logarithms are added when multiplication is required and indices and logarithms are subtracted when division is required.

1. **Definition of Logarithm**

A logarithm is an index. For example, since

$$4^2 = 16,$$

we can say either that the power to which the base 4 must be raised to become 16 is 2 or that the logarithm of 16 to the base 4 is 2. We would write it as

$$\log_4 16 = 2.$$

Similarly, we could write

$\log_4 64 = 3$ because $4^3 = 64$
$\log_4 256 = 4$ because $4^4 = 256$
and $\log_4 1024 = 5$ because $4^5 = 1024$

2. **Logarithms to Different Bases**

Just as we can use different bases raised to various powers, so we can in reverse express numbers in terms of the powers to which various bases must be raised to produce them. For example,

$$100 = 4^{3.322} \text{ so } \log_4 100 = 3.322$$
$$100 = e^{4.605} \text{ so } \log_e 100 = 4.605$$
$$\text{and } 100 = 2^{6.644} \text{ so } \log_2 100 = 6.644$$

(We presume you can see why $\log_2 100 = 2\log_4 100$). Unless you have a fairly sophisticated calculator, and know one extra trick, you couldn't obtain all of these yourself.

3. **Multiplication Using Logarithms**

In Statistics Review 6, we used addition of indices to a common base as an alternative to multiplication of the numbers themselves. Our example there was

$$4 \times 16 = 64.$$

Taking logarithms to base 2, we get

$$\begin{aligned} \log_2 4 + \log_2 16 &= 2 + 4 \\ &= 6 \\ &= \log_2 64. \end{aligned}$$

If we were to take logarithms, say to base 10, we could note that

$$\log_{10} 4 + \log_{10} 16 = \log_{10} 64$$
$$0.602 + 1.204 = 1.806.$$

Having got the answer 1.806 we could take its antilogarithm, i.e. find out what number is equal to 10 raised to the power 1.806. It is, of course, 64. We could get the same answer using logarithms to base e, that is,

$$\log_e 4 + \log_e 16 = \log_e 64$$
$$1.386 + 2.773 = 4.159.$$

The antilogarithm to base e of 4.159 is, of course, 64.

4. **Division Using Logarithms**

Just as we add indices and logarithms when multiplication is required, so we subtract indices and logarithms when division is required. For example,

$$\frac{85}{6} = 14.167$$

could be solved using

$$\log_{10} 85 - \log_{10} 6 = 1.92942 - 0.77815$$
$$= 1.15127.$$

The antilogarithm to base 10 of 1.15127 is 14.167.

The same result would have been obtained using logarithms to the base e, since

$$\log_e 85 - \log_e 6 = 4.44265 - 1.79176$$
$$= 2.65089$$

and the antilogarithm to base e of 2.65089 is 14.167.

Just to make again the connection with indices as presented in Statistics Review 7, these two examples could be written as

$$\frac{85}{6} = 10^{(1.92942-0.77815)}$$
$$= 10^{1.15127}$$
$$= 14.167$$

and

$$\frac{85}{6} = e^{(4.44265-1.79176)}$$
$$= e^{2.65089}$$
$$= 14.167.$$

5. **Natural Logarithms**

Logarithms to base e are referred to as 'natural logarithms'. As a shorthand, instead of $\log_e$, the abbreviation ln is often used. Thus,

$$\ln\ 85 = 4.44265$$

with no need to indicate in the symbol that e is the base.

6. **The Logarithmic Scale**

A special feature of logarithms is that equal differences on a logarithmic scale reflect a constant ratio in the corresponding numbers of which the logarithms are taken. Since logarithms are simply ratios, this property of the logarithmic scale is a property of an index scale. This property can be seen in Table 4.

Consider the base 10 logarithmic transformation first. On the original scale the score 10 is 5 times the score 2 and they are $10 - 2 = 8$ points apart on the scale. The score 100 is 5 times the score 20 and they are $100 - 20 = 80$ points apart on the scale. On the base 10 logarithmic scale, the difference between $\log_{10} 10$ and $\log_{10} 2$ is $(1.000 – 0.301) = 0.699$. Similarly, for other numbers in the same original ratio of 5:1 such as 30 and 6 the difference on the $\log_{10}$ scale is also 0.699 since $\log_{10} 30 = 1.477$ and $\log_{10} 6 = 0.778$. Notice that $\log_{10} 5 = 0.699$.

Table 4 Comparison of original and logarithmic scales

Original scale	$\log_{10}$	$\log_e$
1	0.000	0.000
2	0.301	0.694
5	0.699	1.609
10	1.000	2.303
20	1.301	2.997
50	1.699	3.912
100	2.000	4.606
200	2.301	5.300
500	2.699	6.215
1000	3.000	6.909

The same pattern is evident with the $\log_e$ scale. Pairs of numbers in the ratio 5:1 have $\log_e$ values differing by $\log_e 5 = 1.609$. Pairs in the ratio 10:1 have $\log_e$ values differing by 2.303 and so on.

Because a logarithmic scale expresses with equal intervals, equal ratios on the original scale, it compresses the original scale, which is clear from Table 4.

Exercises

Write the following in the form of a single log expression (i.e. log A):

1. $\log 5 + \log 3$,
2. $\log 18 - \log 9$,
3. $4 \log 2$,
4. $2 + \log_{10} 3$.

Statistics Review 8: Conditional Probability

Key words: Conditional probability

Key points: The conditional probability of an event is the probability of the event under the condition that another event occurs.

1. **Conditional Probability**

In Statistics Review 5, we used a diagram to display the sample space for throws of a red and a blue die and highlighted the occurrences of certain events. Event D = {a total score of six appears} was shown as

$$\mathrm{D} = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}.$$

If we now asked, what is the probability that one of the dice is a 2, *given that their total is 6,* we could see that there are only two occurrences matching this requirement, i.e. {(2, 4), (4, 2)}. The event that *one die is a 2, given that the total on both dice is 6,* is simply the intersection of E = {one die is 2} and D = {total is 6}. That is,

$$\mathrm{E} \cap \mathrm{D} = \{(2,4),(4,2)\}.$$

The probability of this event can be stated as a *conditional probability*. The probability that one die is a 2, given that (under the condition that) both dice total 6 will be

$$\begin{aligned} \mathrm{P}\{\mathrm{E}|\mathrm{D}) &= \frac{\mathrm{n}\{\mathrm{E} \cap \mathrm{D}|}{\mathrm{n}\{\mathrm{D}\}} \\ &= \frac{2}{5}. \end{aligned} \tag{20}$$

The separate probabilities are

$$\begin{aligned} \mathrm{P}\{\mathrm{E} \cap \mathrm{D}\} &= \frac{2}{36} \\ \mathrm{P}\{\mathrm{D}\} &= \frac{5}{36}. \end{aligned}$$

So, the conditional probability could be written as

$$\begin{aligned} \mathrm{P}\{\mathrm{E}|\mathrm{D}) &= \frac{\mathrm{P}\{\mathrm{E} \cap \mathrm{D}\}}{\mathrm{P}\{\mathrm{D}\}} \\ &= \frac{2/36}{5/36} = \frac{2}{5}. \end{aligned} \tag{21}$$

The term P{E | D) is read as the 'probability of E given D' or 'the probability that E occurs given that D occurs'.

A rearrangement of the terms in Eq. (21) produces the multiplication theorem for conditional probability. The equation becomes

$$\mathrm{P}\{\mathrm{E} \cap \mathrm{D}\} = \mathrm{P}\{\mathrm{D}\}\mathrm{P}\{\mathrm{E}|\mathrm{D}). \tag{22}$$

2. **Example: conditional probabilities with two coins**

Before proceeding, let us remind ourselves what we mean by a *probability*: a probability of an outcome is the *theoretical proportion* of times that we would expect that outcome to occur on a very large number of replications.

Suppose that a person tosses two coins and that both coins are somewhat biased, with one being substantially biased[1]: let the probability of Coin 1 coming up Heads

---

[1]No real coin is perfectly unbiased. In developing statistical theory, and in making it concrete with examples of coins (and dice), it is usually assumed that the coins (and dice) are unbiased. These are theoretical coins and dice. Real ones cannot be expected to be perfectly unbiased, though it might be assumed that they are not very biased and that if one coin is biased one way, then another coin might be biased a bit the other way.

Table 5 Probabilities of outcomes of two coins

Coin 1 (Prob)	Coin 2 (Prob)	Joint outcomes (Prob)
H_1 (0.60)	H_2 (0.45)	(0.60)(0.45) = 0.27
T_1 (0.40)	H_2 (0.45)	(0.40)(0.45) = 0.18
H_1 (0.60)	T_2 (0.55)	(0.60)(0.55) = 0.33
T_1 (0.40)	T_2 (0.55)	(0.40)(0.55) = 0.22
		Total = 1.00

(H_1) be 0.60. Then its probability of coming up Tails (T_1) is $1 - 0.60 = 0.40$. Let the probability of Coin 2 coming up Heads (H_2) be 0.45. Then its probability of coming up Tails (T_2) is $1 - 0.45 = 0.55$. The probabilities of each pair of outcomes, on the assumption that the outcome of one toss does not affect the outcome of any other toss of the same coin or of the other coin, can be summarized in Table 5.

The Joint Probabilities of the pair of outcomes are shown in the last column of Table 5, and these probabilities sum to 1.00.

Now a convenient way to compare the relative outcomes of the two coins is to consider only those outcomes in which one coin is a Head and the other is a Tail. If both are heads or both are tails, there is no distinction in outcomes and therefore in information about them. If only one of the coins comes up Head and the other Tail, then we would expect that the one which is biased towards Heads would come up Heads more often. In the theoretical case above, let us calculate what proportion of the time we would expect the first to be Heads and the second to be Tails when only one of them is Heads.

Considering only one of them to be Heads and the other Tails involves a conditional probability: the condition that only one of the Coins is a Head. This means that we do not consider when both are Heads or both are Tails, which in turn means that we consider only the subset of outcomes shown in Table 6.

We can see from Table 6 that the probability of one of the coins being Heads and the other Tails, irrespective of which one is Head or Tails, is 0.51.

Given that only one coin is a Head, what is the probability that it is the second one: that is, *relative* to this probability (of 0.51), we require the probability that the event T_1 and H_2 occurs: since (T_1 and H_2) occurs with probability 0.18 in the whole set, the conditional probability of (T1 and H_2), given (T_1 and H_2) OR (H_1 and T_2), is given by the ratio

$$\mathrm{Prob}\{(\mathrm{T}_1\ \mathrm{and}\ \mathrm{H}_2)|(\mathrm{T}_1\ \mathrm{and}\ \mathrm{H}_2)\mathrm{OR}(\mathrm{H}_1\ \mathrm{and}\ \mathrm{T}_2)\} = \frac{0.18}{0.51} = 0.353.$$

Table 6 Probabilities of outcomes of one head and one tail

Coin 1 (Prob)	Coin 2 (Prob)	Joint outcomes (Prob)
T_1 (0.40)	H_2 (0.45)	(0.40)(0.45) = 0.18
H_1 (0.60)	T_2 (0.55)	(0.60)(0.55) = 0.33
		Total = 0.51

Thus, the probability of the second being a Head, when only one of them is a Head and the other is a Tail, is 0.353—less than 0.50 which would be unbiased. This is as expected since the second coin is biased to being a Tail and the first is biased to being a Head.

The probability of the first being a Head and the second a Tail is the complement of this

$$\text{Prob}\{(\text{H}_1 \text{ and } \text{T}_2)|(\text{T}_1 \text{ and } \text{H}_2)\text{OR}(\text{H}_1 \text{ and } \text{T}_2)\} = 1.00 - 0.353 = 0.647.$$

This probability can also be calculated from first principles as given below:

$$\text{Prob}\{(\text{H}_1 \text{ and } \text{T}_2)|(\text{T}_1 \text{ and } \text{H}_2)\text{OR}(\text{H}_1 \text{ and } \text{T}_2)\} = \frac{0.33}{0.51} = 0.647.$$

Thus, the probability of the first being a Head, when only one is a Head and the other is a Tail, is even greater than 0.60 which is the probability of the first being a Head on its own.

If we did not know the bias of one coin relative to the other, we could toss the coins as a pair 200 times say, and in every case that we have a Head or a Tail, note which one it is. Then suppose one comes up Heads and the other Tails on 110 occasions, we would note which is Heads and which is Tails. If Coin 1 appears 60 times and Coin 2 50 times, then the estimated probability of Coin 1 coming up Heads when the other is Tails in the future would be as follows:

$$\text{Estimated} \quad \text{Prob}\{(\text{H}_1 \text{ and } \text{T}_2)|(\text{T}_1 \text{ and } \text{H}_2)\text{OR}(\text{H}_1 \text{ and } \text{T}_2)\} = \frac{60}{110} = 0.545.$$

Perhaps you could try this argument with 50 tosses of a pair of coins. It does not take long to toss two coins 50 times. Is one coin more likely to come up Heads than Tails, and how likely? Make sure that you mark the coins so that you know which coin is which in the outcomes.

Exercises

1. Suppose one tosses two coins C_1 and C_2, and that the respective probabilities of coming up Heads is

$$\Pr\{C_1 = H\} = 0.55, \quad \Pr\{C_2 = H\} = 0.45.$$

1. What is the probability that C_1 = T and what is the probability that C_2 = T?
2. What is the probability that C_2 = H (and C_1 = T) if only one of the coins comes up H?

Statistics Review 9: Independence

Key words: Independence

Key points: An event A is said to be *independent* of event B if the probability of A occurring is not influenced by whether B occurs or not.

An event A is said to be *independent* of event B if the probability of A occurring is not influenced by whether B occurs or not. That is,

$$P\{A\} = P\{A|B\}. \tag{23}$$

Now, we know from (3) that

$$P\{A \cap B\} = P\{B\}P\{A|B\}$$

so we can substitute P{A} for P{A | B} and to assert as a *formal definition of independence* that

$$\text{Events A and B are independent,} \tag{24}$$

$$\text{if } P\{A \cap B\} = P\{A\}P\{B\}.$$

As an example consider the situation when a coin is tossed three times. The sample space will be

$$S = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}.$$

We can now examine the following events:

$$\begin{aligned} A &= \{\text{first toss is heads}) \\ &= \{HHH,\ HHT,\ HTH,\ HTT], \\ B &= \{\text{second toss is heads}\} \\ &= [HHH,\ HHT,\ THH,\ THT], \\ C &= \{\text{exactly two heads are tossed in a row}\} \\ &= \{HHT,\ THH\}. \end{aligned}$$

The probabilities of these events are

$$P\{A\} = \frac{4}{8} = \frac{1}{2},$$

$$P\{B\} = \frac{4}{8} = \frac{1}{2},$$

$$P\{C\} = \frac{2}{8} = \frac{1}{4}.$$

Independence of A and B?

The intersection of A and B is $A \cap B = \{HHH, HHT\}$ and its probability is

$$\begin{aligned} P\{A \cap B\} &= \frac{2}{8} = \frac{1}{4} \\ &= P\{A\}P\{B\} \end{aligned}$$

so A and B are independent.

Independence of A and C?

The intersection of A and C is $A \cap C = HHT$ and its probability is

$$\begin{aligned} P\{A \cap C\} &= \frac{1}{8} \\ &= P\{A\}P\{C\} \end{aligned}$$

so A and C are independent.

Independence of B and C?

The intersection of B and C is $B \cap C = HHT, THH$ and its probability is

$$\begin{aligned} P\{B \cap C\} &= \frac{2}{8} = \frac{1}{4} \\ &\neq P\{B\}P\{C\} \end{aligned}$$

so B and C are not independent.

For our final comment refer back to Eq. (8) in Statistics Review 3. There we noted that, with two events A and B, the probability of one or the other occurring is

$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}. \tag{25}$$

We noted in that Statistics Review that *if A and B are mutually exclusive*

$$P\{A \cup B\} = P\{A\} + P\{B\}. \tag{26}$$

We can now note that *if A and B are independent*

$$\mathrm{P}\{\mathrm{A} \cup \mathrm{B}\} = \mathrm{P}\{\mathrm{A}\} + \mathrm{P}\{\mathrm{B}\} - \mathrm{P}\{\mathrm{A}\}\mathrm{P}\{\mathrm{B}\}. \tag{27}$$

As an illustration of the use of Eq. (8), we can take a case with two events we know to be independent. Suppose two friends A and B were about to play separate singles tennis matches against another pair they had often played before. On the basis of past performance A could be judged to have a probability of success of 0.5 against her opponent and B a probability of 0.7 against hers. The probability that at least one will win is

$$\begin{aligned}\mathrm{P}\{\mathrm{A} \cup \mathrm{B}\} &= 0.5 + 0.7 - (0.5)(0.7) \\ &= 0.85.\end{aligned}$$

Exercises

Let A be the event that a family has children of both sexes and let B be the event that a family has at most one boy. This is a relatively difficult problem so do not spend too much time on it.

1. Show that A and B are independent events if a family has three children.
2. Show that A and B are dependent events if a family has two children.

Statistics Review 10: Bernoulli and Binomial Random Variables

Key words: Random variable, Bernoulli variable, Binomial experiment, Binomial variable, Average or Observed proportion, Theoretical proportion, Probability

Key points: A variable which takes on different values with certain probabilities in an equation or model, and not with certainty, is called a *random variable*. A random variable which has only two outcomes (0 and 1) is known as a *Bernoulli random variable*. Repeated independent trials of a Bernoulli variable denote a *Binomial experiment*. The *Binomial variable X* can then be defined as the number of 1s in *n* trials. The *observed proportion* of times a value of a Bernoulli random variable occurs in a Binomial experiment, i.e. the *average* is an estimate of the *probability* (*theoretical proportion*) of the value.

A variable which takes on different values with certain probabilities in an equation or model, and not with certainty, is called a *random variable*. It is random because its outcome is not determined even if all of the values in the model are known.

A random variable which can take on only the two values 0 and 1 is known as a *Bernoulli random variable*. It is a variable with only two outcomes, usually defined as ‘success’ and ‘failure’. Two outcome situations are common, for example, Heads/Tails, Male/Female, Win/Loss, Correct/Incorrect, etc.

Repeated independent trials of a Bernoulli variable denote a *Binomial experiment*. The Binomial variable X can then be defined as the number of 'successes', or 1s, in n trials. So the Binomial variable is the sum of the Bernoulli variables. The trials are independent in the sense that the probability of a success or failure remains constant from trial to trial.

The average of a Bernoulli random variable can be found very easily. The Binomial random variable is the replication of these dichotomous values when the parameters are the same. For example, we would have to imagine that the same person is given the same item on repeated occasions and that the person does not remember the previous outcome. This of course is not possible, but we use the idea.

1. **An example where the probability of each Bernoulli response is the same on replications (Binomial experiment)**

A better example to continue the development of a Bernoulli random variable is to imagine tossing the same coin and assigning 1 for a Head and 0 for a Tail. If the same coin is tossed, we can assume that the probability of a Head (and therefore a Tail) remains the same from toss to toss.

Now suppose that the coin is tossed 10 times and that the outcome is 6 Heads (1s) and 4 Tails (0s). The list of numbers in order in which they may appear is shown in Table 7.

What is the average of these scores? This is simply $\bar{X} = \frac{\sum_{i=1}^{10} X_i}{10} = \frac{6}{10} = 0.6$.

However, this average is also the proportion of times that a 1 appears. Thus, the *average* is the *proportion* of times a positive response (scored 1) appears.

Now suppose that we thought of this proportion, not as an observed proportion, but as a theoretical proportion. As a theoretical proportion it becomes a probability. If we denote this probability as P, then in this case $P = \bar{X}$. We use whichever is convenient. It is often convenient to think of probabilities rather than averages, although in this case they are the same.

Thus, in the case of a Bernoulli random variable, where the only two possible values are 0 or 1, we can think of the probability as a theoretical proportion of times that the response would be a 1. If we consider that the responses are to be used to estimate the theoretical proportion (probability) from the observed proportion, the number 0.6 above would be our best estimate of the probability that an outcome would be a '1'.

Although this is our best estimate of the probability, we might have hypothesized in advance of any tosses being made that the coin is not biased and that the probability of a Head is 0.5. Then, we could carry out a statistical test to determine if this number of Heads is significantly different from a probability being 0.5. In the case of 10 tosses, it would not be significant. However, that is a different matter

Table 7 A set of outcomes for 10 tosses of a coin

Random variable X	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	$\sum_{i=1}^{10} X_i$
Value X_i	1	1	1	0	1	1	0	0	1	0	6

Table 8 Estimated probabilities of each toss being a Head

Random variable X	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	$\sum_{i=1}^{10} X_i$
Value X_i	1	1	1	0	1	1	0	0	1	0	6
P_i	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	$\sum_{i=1}^{10} P_i = 6$

which involves having some a priori hypothesis about what the probability might be. On the other hand, if in 100 tosses there were 60 heads, then we would have to conclude that it is very unlikely that the probability of a Head is 0.5. If we had no a priori idea, then 0.6 would be our best estimate.

Using the 'hat' notation for an estimate, we can write $\hat{P} = 0.6$.

With this theoretical average or probability, we can arrange the information in a different way that is important conceptually. Each of the replications of the toss of the coin can be thought of as having a probability of coming up Heads, and this is the average of the times that it would do so. Augmenting the table above with this set of numbers gives Table 8.

Note that now we think of each toss as having an average although in this case it is the same for each toss. We subscript each toss accordingly as P_i. Note further that the probabilities in this case also add up to 6. This is an important idea which will be used in Chap. 10 in the case of responses of persons to items. Table 8 reflects this idea.

2. **An example where the probability of each response is different from one Bernoulli response to the next**

This is the situation when we have the response of one person to, say, 10 items, where each item has a different difficulty. Let the random variable be X_{ni} for each person and each item. If we wanted to be careful, or pedantic, we might represent the specific values in each case as a lower case variable $x_{ni} \in \{0, 1\}$, where the symbol '$\in$' indicates that the response belongs to the set of possible values $\{0, 1\}$.

The table below shows the same responses, but now the symbols have been changed.

Random variables	X_{n1}	X_{n2}	X_{n3}	X_{n4}	X_{n5}	X_{n6}	X_{n7}	X_{n8}	X_{n9}	X_{n10}	Total score r_n
Value x_{ni}	1	1	1	0	1	1	0	0	1	0	6

In this case, although we have the same person, we do not have the same item. Because the items have different difficulties, the probability of success of the person on each item will vary according to the difficulties of the item.

Suppose we knew the person's proficiency and the item's difficulty. We would then know the probability that the person has for getting each item right. We might have something like the table below.

However, we usually do not know both of these. Suppose we know the item difficulties from an analysis of data, and now wish to estimate the person's proficiency.

To estimate the person's proficiency, it is necessary to find the value that could be entered into the Rasch model equation so that the sum of the probabilities is 6. Then, we would end up with a table as below.

Random variable	X_{n1}	X_{n2}	X_{n3}	X_{n4}	X_{n5}	X_{n6}	X_{n7}	X_{n8}	X_{n9}	X_{n10}	Total score r_n
Observed value x_{ni}	1	1	1	0	1	1	0	0	1	0	6
Average $\bar{X}_n$	0.95	0.92	0.88	0.81	0.73	0.50	0.38	0.37	0.27	0.19	6.0
Probability $\hat{P}_{ni}$	0.95	0.92	0.88	0.81	0.73	0.50	0.38	0.37	0.27	0.19	6.0

The average and the probability, which are the same, are now an abstraction—it is a conceptualization of the proportion of success in a series of infinite replications of the person with the same proficiency completing each item. Clearly, if it really were the same person and the same item, the person would give the same response. So we can imagine that it is an infinite number of *people* of exactly the same proficiency, responding to each item.

If we knew the proficiency of the person and the difficulty of each item from some other information, then it is most unlikely that the sum of the probabilities would be 6. However, in estimating the person's proficiency, the constraint is that the sum of the probabilities is 6.

Statistics Review 11: The Chi-Square Distribution and Test

Key words: χ^2 (chi-square) distribution, χ^2 (chi-square) test, Degrees of freedom

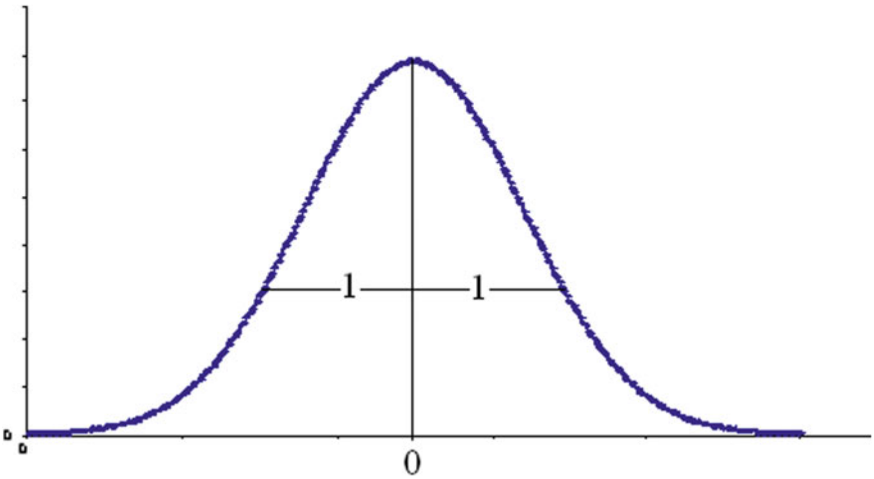
Key points: The distribution of the sum of squares of *n* random normal deviates is given by summing squaring and summing *n* random normal deviates. This is called the χ^2 *distribution* on *n degrees of freedom*. The χ^2 test is used to test how well the frequency distribution of observed data, for which each observation may fall into one of several categories, matches the theoretical χ^2 distribution. The *probability* of obtaining another chi-square value greater than an obtained chi-square value is the area under the χ^2 *distribution* curve to the right of the obtained value.

1. The χ^2 (chi-square) distribution

Suppose that we have a random normal distribution X. It may have come from the sampling distribution of means, or it might be a hypothetical distribution. In general, by similar reasoning to that of the sampling distribution of means, random errors are taken to be normally distributed. Depending on the precision, some will have a greater variance than the others.

The *standard* normal distribution denoted generally Z is obtained by constructing its values as follows:

$z_i = \frac{x_i - E[X]}{\sqrt{V[X]}}$ or in terms of the random variable Z, $Z = \frac{X - E[X]}{\sqrt{V[X]}}$. The curve below is the shape of a normal distribution:



1.1. The expected value and variance of a standard random normal deviate.

You do NOT have to follow the proofs below, which use properties of the operator notation, but you have to know the results.

The variable Z is called a random normal deviate. Then its expected value (theoretical mean) is given by

$$E[Z] = E[\frac{X - E[X]}{\sqrt{V[X]}}] = \frac{E[X - E[X]]}{\sqrt{V[X]}} = \frac{E[X] - E[E[X]]}{\sqrt{V[X]}}$$
$$= \frac{E[X] - E[X]}{\sqrt{V[X]}}$$
$$= 0.$$

Essentially, if you subtract the mean from a set of numbers, then the mean of the new numbers must be zero.

$$E[Z] = 0.$$

The variance is given by

$$V[Z] = V[\frac{X - E[X]}{\sqrt{V[X]}}] = \frac{V[X - E[X]]}{V[X]} = \frac{E[(X - [E[X])^2]}{V[X]}$$
$$= \frac{V[X]}{V[X]}$$
$$= 1.$$

Essentially, if you divide the standard deviation into a set of deviates from the mean of a set of numbers, then the variance of the new numbers must be one.

$$V[Z] = 1.$$

Furthermore, its square root, the standard deviation, will also be 1.

The standard normal deviate is used as a reference point to check if some number is significantly different from the value that might be expected under only random variation. Thus, if one computes a prediction of some number, then in probabilistic models it is not expected that the prediction will be perfect.

Suppose we have an observed value Y_i and we know the theoretical mean μ_Y and variance σ_Y^2 of this value (We come to how we might know these values). Then, we can compute the standard normal deviate

$$Z_i = \frac{Y_i - \mu_Y}{\sigma_Y}.$$

Then to check if the value Y_i is a good prediction of μ_Y, we can compare the value of Z_i with might arise from random normal variation. You have learned that if the value is between -1.96 and 1.96, then that means it is in the range where 95% of cases under random normal variation would fall. In that case, we might consider it a good prediction.

1.2. The χ^2 distribution on 1 degree of freedom

The χ^2 distribution arises from the need to consider more than one prediction simultaneously. We begin by considering just one prediction, and in some sense it is redundant with the random normal deviate. However, it lends itself to generalization when there is more than one simultaneous prediction involved.

When we have one prediction, we consider as a frame of reference one standard normal deviate, and square it: Z_i^2. We now imagine taking many, many such deviates from a standard normal distribution and squaring them.

The distribution of the squares of these random normal deviate is a χ^2 distribution on 1 degree of freedom notated χ_1^2.

Then to check our prediction, we could square the calculated standard normal deviate and check if it falls between the value of 0 and $1.96^2 = 3.8416$.

1.3. The χ^2 distribution on 2 degree of freedom

If we have two simultaneous predictions, we can imagine using as the frame of reference to check its accuracy two simultaneous random normal deviates. However, we should combine these somehow to give a single summary value. This is done by imagining taking many standard random normal deviates, squaring them, and summing them. This is the χ^2 distribution on 2 degrees of freedom:

$$\chi_2^2 = Z_1^2 + Z_2^2.$$

1.4. The χ^2 distribution on n degree of freedom.

The distribution of the sum of squares of n random normal deviates is χ^2 on n degrees of freedom is given by summing squaring and summing n random normal deviates as

$$\chi_n^2 = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 + \cdots + Z_n^2.$$

2. **The χ^2(chi-square) test**

The χ^2(chi-square) test is used to test how well the frequency distribution of observed data, for which each observation may fall into one of several categories, matches the theoretical chi-square probability distribution. The chi-square statistic is calculated for each category first, and then summed across all categories. The chi-square test always tests the **null hypothesis,** which states that there is no significant difference between the expected and observed result.

The formula for calculating the chi-square statistic for a category is

$$\chi^2 = \frac{(o - e)^2}{e},$$

where o is the observed number of observations in that category and e the expected number of observations in that category. That is, chi-square is the squares of the difference between the observed and expected value divided by the expected value.

Then the final chi-square statistic is the sum of the chi-squares of all possible categories k.

$$\chi_k^2 = \sum_{j=1}^{k} \chi_j^2.$$

A chi-square test is really any statistical hypothesis test in which the sampling distribution of the test statistic is a chi-square distribution when the null hypothesis is true, or any in which this is asymptotically true.

Degrees of freedom

Degrees of freedom are used to then determine whether a particular null hypothesis can be rejected based on the number of parameters to be estimated and size of the sample. We start with as many degrees of freedom as there are observations in a data set, n. Then we subtract 1 for each parameter being estimated. So, in general, if two parameters need to be estimated the degrees of freedom are n – 2.

For a chi-square test, we said earlier that each observation may fall into one of several categories. So for this test we start with the number of categories k. One parameter, the total chi-square χ_k^2, is estimated. So the degrees of freedom are $k - 1$.

Probability

The area under the curve for the chi-square distribution is set to one unit so that it is a probability distribution. Then the probability of obtaining a value of chi-square between two values is the area under the curve between the values. The probability of obtaining another chi-square value greater than an obtained chi-square value is the area under the curve to the right of the obtained value.

In hypothesis testing, if the probability of obtaining another chi-square value greater than an obtained chi-square value is less than 0.05 (or sometimes 0.01) then we accept the null hypothesis, that is, there is no significant difference between the expected and observed results and the observed difference is due to chance.

Statistics Review 12: Analysis of Variance (ANOVA)

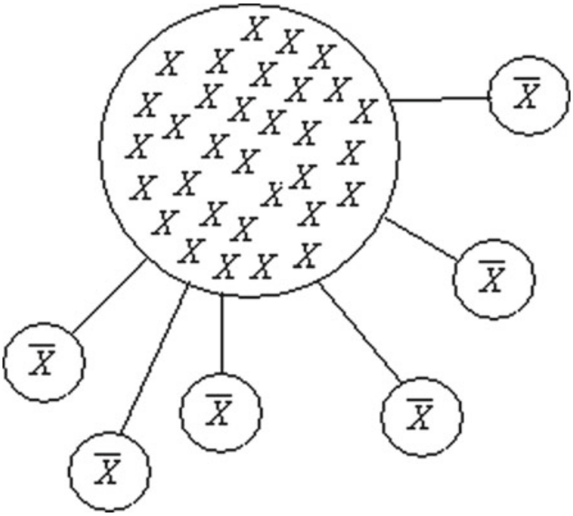
Key words: Sampling distribution of means, Within-groups variance, Between-group variance, F-ratio, Standard error of the mean

Key points: Analysis of Variance (ANOVA) is used to test whether differences among sample group *means* are statistically significant. To do that the *population variance* is analysed, as estimated from the samples. The F-ratio is the ratio of between-group variance and within-groups variance. The larger the F-ratio the greater the likelihood that the differences between the means of two samples are due to something other than chance alone, namely, real effects.

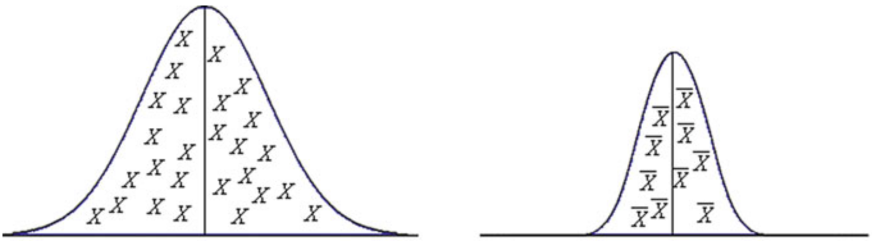
Make sure you are familiar with the sampling distribution of means and the Central Limit Theorem as summarized in section 1. This will help you understand the explanation of analysis of variance in section 2.

1. **The sampling distribution of the means**

Consider the population of scores X below with population mean = μ and standard deviation = σ. We can take many samples of size n and calculate the mean for each one of them as the diagram below illustrates:



We can create a distribution of the sample means. This distribution is called the sampling distribution of means. It has a mean of μ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$. The diagram below shows an example of a distribution of X-scores as well as the sampling distribution of the means.



(a) population distribution $N(\mu, \sigma)$ (b) sampling distribution of means $N(\mu, \frac{\sigma}{\sqrt{n}})$

The Central Limit Theorem states that if a population is normally distributed with a mean of μ and a standard deviation of σ then the distribution of sample means of size n is normally distributed with mean μ and standard deviation of $\frac{\sigma}{\sqrt{n}}$. For large n (20 or so), the sampling distribution of the mean tends to the normal *irrespective of whether or not* the population itself is normally distributed. The standard deviation of the sampling distribution of the sample means is also called the *standard error of the mean*:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}.$$

2. Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is used to test differences among groups, more specifically whether differences among group *means* are statistically significant. To do that *variance* is analysed, as the name analysis of variance suggests. So to decide whether group *means* are different we analyse the *population variance*, as we can best estimate it with the information we get from samples of the population. We make *two estimates of the population variance*. If the groups are from the same population, the two estimates should be close to each other in value. So, if the ANOVA results in two estimates close in value, we conclude that the groups are from the same population and the difference in the means is due to chance or error and hence not significant. If the groups are from different populations, the two estimates will differ significantly. So, if the ANOVA results in two very different estimates, we conclude that the groups are from different populations and hence the means differ significantly. The two estimates of the population variance that are calculated are

(1) The variability of subjects *within* each group and
(2) The variability *between* the different groups.

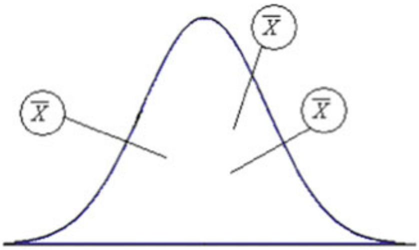
Variance was earlier defined as simply the sum of squared deviations from the mean $\left(\sum_{i=1}^{N}(X_i - \bar{X})^2\right)$ divided by n − 1 (page 10).

$$s^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}.$$

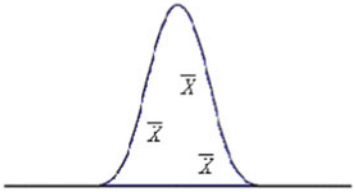
The variability of subjects *within* each group is the variability of scores about their subgroup sample means. This will be reflected by the deviation $X - \overline{X}$ (as in the formula above), where $\overline{X}$ is the mean of the subgroups that contains X. To estimate the population variance, we calculate the average of the group variances.

The variability *between* the different groups is the variability of subgroup sample means about the grand mean of all scores. This will be reflected by the deviation $\overline{X} - \overline{\overline{X}}$. Notice here that we are dealing with the *sampling distribution of the means* and our estimate of population variance is the variance of the sampling distribution of the means multiplied by n (explained in more detail later).

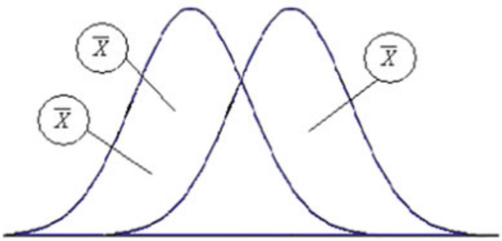
The diagrams below illustrate two cases: (1) where three samples are taken from the same population and (2) where three samples are taken, two from the same population and one from a different population. Diagrams 1b and 2b show the sampling distribution of the means in each case. Note that the sampling distribution 2b has a bigger variance than the sampling distribution 1b, resulting in a bigger *between-groups variance* for the case where the samples are not from the same population.



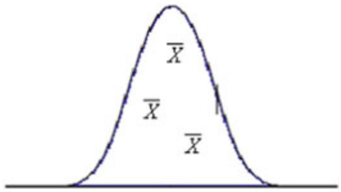
(1a) population distribution



(1b) sampling distribution of means



(2a) population distributions



(2b) sampling distribution of means

Suppose we have three samples of size 3. They are given three different treatments and the results are shown in Table 9 [Note: We would usually have a much larger sample size].

Table 9 X-scores for 3 groups

Group 1	Group 2	Group 3
4	5	6
3	3	4
2	4	5
$\bar{X}_1 = 3$	$\bar{X}_2 = 4$	$\bar{X}_3 = 5$

The means for the three groups are different: 3, 4 and 5, respectively. To find out whether they differ significantly, we follow the following procedure:

Step 1: Calculate the first estimate of population variance: the variability of subjects *within* each group

If each sample is from the same population, then each can have its variance as an estimate of the population variance. We therefore estimate the population variance from each group. Since each of the sample variances may be considered an independent estimate of the parameter σ, finding the mean of the variances provides a method of combining the separate estimates of σ into a single value (Table 10).

(i) From group 1 $s_1^2 = \frac{\sum_{1=1}^{N} (X_i - \bar{X},)^2}{N-1} = \frac{2}{2} = 1.$

(ii) From group 2 $s_2^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N-1} = \frac{2}{2} = 1.$

(iii) From group 3 $s_3^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X}.)^2}{N-1} = \frac{2}{2} = 1.$

Thus, the average estimate is $\sigma^2 = \frac{1+1+1}{3} = 1.$

The number of degrees of freedom here is $n_1 - 1 + n_2 - 1 + n_3 - 1 = 3 - 1 + 3 - 1 + 3 - 1 = 6.$

An assumption here is that the treatment population variances, from which each sample is a random sample, and are homogenous; otherwise, pooling the estimates could not be justified.

Step 2: Calculate the second estimate of population variance: the variability of subjects *between* the different groups

In deriving the second estimate of the population variance the logic is slightly less straightforward and employs both the concept of the sampling distribution and the Central Limit Theorem. The relationship expressed in the Central Limit Theorem may now be used to obtain an estimate of σ^2.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}$$

Table 10 Calculations for the variance of each group

Group 1			Group 2			Group 3		
X	$X - \overline{X_1}$	$(X - \overline{X_1})^2$	X	$X - \overline{X}_2$	$(X - \overline{X}_2)^2$	X	$X - \overline{X_3}$	$(X - \overline{X}_3)^2$
4	1	1	5	1	1	6	1	1
3	0	0	3	−1	1	4	−1	1
2	−1	1	4	0	0	5	0	0
	$\sum(X - \overline{X}_1)^2 = 2$			$\sum(X - \overline{X}_2)^2 = 2$			$\sum(X - \overline{X}_3)^2 = 2$	

Table 11 Calculations for the variance of sample means

	$\overline{X}$	$\overline{X} - \overline{\overline{X}}$	$(\overline{X} - \overline{\overline{X}})^2$
$\overline{X_1}$	3	−1	1
$\overline{X_2}$	4	0	0
$\overline{X_3}$	5	1	1
			$\sum\left(\overline{X} - \overline{\overline{X}}\right)^2 = 2$

$$n * \sigma^2_{\overline{X}} = \sigma^2.$$

Thus, the variance of the population may be found by multiplying the standard error of the mean squared $\left(\sigma^2_{\overline{X}}\right)$ by the size of each sample (n). In our example, $\sigma^2_{\overline{X}}$ is given by calculating the variance of the three sample means. (If they are from the same population we will have an estimate of the population variance.) The grand mean of all scores is $\overline{\overline{X}} = 4$. Let the number of groups be c (here c = 3) (Table 11).

Estimated population variance of means:

$$\sigma^2_{\overline{X}} = \sum \frac{(\overline{X} - \overline{\overline{X}})^2}{c-1} = \frac{2}{2} = 1.$$

$$\begin{aligned} \text{The estimate is } \sigma^2 &= \mathrm{n}\left(\mathrm{est}\,\sigma^2_{\overline{X}}\right) \\ &= 3(1) \\ &= 3. \end{aligned}$$

This formula is satisfactory here only because all sample sizes are the same. Otherwise, the formula needs to be adjusted. The principle, however, is the same. The number of degrees of freedom in this estimate is c – 1 = 2.

Step 3: Form the F-ratio and test the null hypothesis

We need to decide whether at least one mean differs from any other. The null hypothesis is that no mean differs significantly from any other.

H_0 : Between-groups variance – within-groups variance = 0,

H_1 : Between-groups variance > within-groups variance.

The *within-groups* estimate of the population variance is based on the assumption that all treatments have the same effect on the variances, even if the effect on the means is not the same. This is usually zero. Therefore, this variance is thought of as variance one could expect anyway even if the treatments were not applied and is generally known as an estimate of *error variance*. The *between-groups* estimate of the population variance is an appropriate estimate if the treatments had no effect on the means. The differences in the sample means will be due to chance and will therefore be error variance. But if the treatments have an effect, then the between-groups variance will not only contain error variance, but variance due to differences between sample means due to treatment effects.

A new statistic called the F-ratio is computed by dividing the between-groups variance by within-groups variance.

$$F = \frac{\text{Between-groups variance}}{\text{Within-groups variance}}.$$

The F-ratio can be thought of as a measure of how different the means are relative to the variability within each sample. The larger this value, the greater the likelihood that the differences between the means are due to something other than chance alone, namely, real effects.

$$F = \frac{\text{Error Variance} + \text{Treatment Variance}}{\text{Error Variance}}.$$

If $F < 1$, then we consider that treatment variance = 0 and that the error variance estimated from between groups is less than the error variance estimated from within groups simply by chance. Thus, if $F < 1$ we accept H_0. If H_0 is rejected H_1 is accepted, i.e. Error variance + Treatment Variance > Error Variance so treatment variance > 0. If the difference between the means is only due to error variance then the expected value of the F-ratio would be very close to 1 because both the numerator and the denominator of the F-ratio are estimates of the same parameter, σ^2. However, because the numerator and the denominator are estimates rather than exact values the F-ratio will seldom be exactly 1.

In our example,
$$F = \frac{3}{1} = 3(\text{df: } 2, 6).$$

Consult a table of the F-distribution, a theoretical probability distribution characterized by two parameters: degrees of freedom 1(for the numerator in the F-ratio) and degrees of freedom 2 (for the denominator in the F-ratio). For different values of df1 and df2, the tables provides F-ratios, which is also called F(critical) values.

In our example, F(critical) at 5% level for degrees of freedom (2, 6) = 5.1 Since F (observed) < F(critical) we accept H_0. So the difference in the sample means is not significant.

Assumptions for ANOVA

1. All sample groups were originally drawn at random from the same population [After administration of the treatments each group is then regarded as a random sample from a different treatment population.].
2. The variance of measures of the treatment populations is the same, i.e. homogenous.

3. The distribution of measures of the treatment is normal.
4. There is no difference in the means of the measures between the treatment populations. This is the null hypothesis and if assumptions 1, 2 and 3 are satisfied we may accept or reject this condition on the basis of the obtained F value.

Statistics Review 13: Distribution Theory

1. **Theoretical distributions**

Here, we abstract the idea of a theoretical distribution from an observed or empirical distribution of numbers. With this abstraction, we introduce the operator notation used in summarizing properties of a distribution. The idea is not difficult, but it is very important in understanding and carrying out tests of fit.

Consider observations belonging to the set of numbers {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10} These might be possible scores on a 10 item test where each item is scored 0 or 1. The number of possible scores is 11.

Suppose the actual observations in some sample were $X_i, i = 1, 2, \ldots, 10$: {6, 6, 7, 8, 9, 5, 10, 4, 1, 4}. Then the *mean* is given by

$$\bar{X} = (X_1 + X_2 + \cdots X_i \cdots + X_n)/n,$$

which in the particular data we have is given by

$$\bar{X} = \frac{6+6+7+8+9+5+10+4+1+4}{10} = \frac{60}{10} = 6.$$

The responses are shown in a table on the next page.

2. **Formulae for the mean and variance of the observed responses**

Each of the possible responses is represented by the variable Y. The variable can take on the values 0 to 10, and so there are 11 possible values: $Y_i, i = 1, 2, \ldots., 11$. These are shown in Table 12. Please study this table and understand it.

Table 12 Frequency distribution of responses

i	Yi	f_i	f_iY_i	p_i	p_iY_i	$(Y_i - \bar{Y})^2$	$p_i(Y_i - \bar{Y})^2$
1	0	0	0	0.0	0.0	$(-6)^2$	0.0(36) = 0.0
2	1	1	1	0.1	0.1	$(-5)^2$	0.1(25) = 2.5
3	2	0	0	0.0	0.0	$(-4)^2$	0.0(16) = 0.0
4	3	0	0	0.0	0.0	$(-3)^2$	0.0(9) = 0.0
5	4	2	8	0.2	0.8	$(-2)^2$	0.2(4) = 0.8

(continued)

i	Yi	f_i	f_iY_i	p_i	p_iY_i	$(Y_i - \bar{Y})^2$	$p_i(Y_i - \bar{Y})^2$
6	5	1	5	0.1	0.5	$(-1)^2$	0.1(1) = 0.1
7	6	2	12	0.2	1.2	$(0)^2$	0.2(0) = 0.0
8	7					$(1)^2$	
9	8	1	8	0.1	0.8	$(2)^2$	0.1(4) = 0.4
10	9	1	9	0.1	0.9	$(3)^2$	0.1(9) = 0.9
11	10	1	10	0.1	1.0	$(4)^2$	0.1(16) = 1.6
Sum		10	60	1.0	6.0		= 6.4

Then the mean and variances of the values of Y can be calculated as follows:

$$\bar{Y} = \frac{\sum_i f_i Y_i}{\sum_i f_i} = \frac{\sum_i f_i Y_i}{N} = \sum_i \left(\frac{f_i}{N}\right) Y_i = \sum_i p_i Y_i = 6.0$$

$$S_y^2 = \frac{\sum_i f_i (Y_i - \bar{Y})^2}{\Sigma f_i} = \Sigma \left(\frac{f_i}{N}\right) (Y_i - \bar{Y})^2 = \sum p_i (Y_i - \bar{Y})^2 = 6.4.$$

The key feature of this calculation is that we have brought into it the proportion of persons p_i who have obtained each possible score Y_i.

Now suppose that we know the possible scores for some set of responses, but that we have from outside the data, a theoretical basis for knowing or computing the proportion of times each score would occur. This theoretical proportion we call a *probability*.

3. **Frequency distribution of the responses**

Complete the cells that are missing from the set of numbers above and check the 'sum' row.

$$p_i = \frac{f_i}{N} = \frac{f_i}{10}$$

4. **Formulae and operator notation for the mean and variance of a variable**

To calculate the theoretical mean and variance, we simply substitute for the observed proportion of cases we had in the above formula the *theoretical* proportion. To distinguish the observed proportion from a theoretical proportion we use the Greek letter π. The theoretical mean is notated $E[Y]$ and is read as 'Expected Value of Y'. Notice that in this expression, the subscript i which identifies actual observations is not present. This is called the operator notation and variable rules follow from using it.

Likewise the variance is notated $V[Y]$ and read the 'variance of Y'.

$$\text{Theoretical mean: } E[Y] = \sum_i \pi_i Y_i = \mu_Y$$
$$\text{Theoretical variance: } V[Y] = \sum_i \pi_i (Y_i - \mu_Y)^2 = \sigma_Y^2.$$

The term in the middle of the three terms in each equation gives the instruction as to how to calculate the theoretical mean or theoretical variance, while the last part gives the value of the theoretical mean or theoretical variance. These three terms are interchangeable, and each is used when, in the context, it is most convenient. It is important to appreciate that just because we say 'Expected Value' does not mean we expect that value as the most likely to occur or any such thing. In fact, mostly it is not going to appear.

5. **Some Examples with Theoretical Probabilities**

We will now do some exercises where the theoretical proportion, or probability, comes from outside any particular data set.

(i) Consider tossing a coin.

Probability of a Heads is given by π_1, probability of Tails by π_2. We assume the coin is unbiased, and so each has a probability of 0.5 of occurring.

$$\begin{array}{ccc} \text{Eventor Outcome} & \text{Random Variable} & \text{Probabilities} \\ & Y_i & \pi_i \\ \begin{bmatrix} H \\ T \end{bmatrix} & \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \end{array},$$

$$E[Y] = \sum_{i=1}^{2} \pi_i Y_i = \pi_1 Y_1 + \pi_2 Y_2 = (0.5)(1) + (0.5)(0)$$
$$= 0.5\,(= \pi_1 \text{in the dichotomous case}).$$

NB (i) Do not 'expect' to get 0.5.

(ii) Y is the 'Random Variable'. Specific values are Y_i.

(ii) What if the coin were biased?

$$E[Y] = \sum_k \pi_k Y_k = \left(\frac{3}{4}\right)(1) + \left(\frac{1}{4}\right)(0) = \frac{3}{4}(= \pi_1 \text{in the dichotomous case}).$$

Event	Value	π_k
H	1	$\frac{3}{4}$
T	0	$\frac{1}{4}$

(iii) Tossing a die (please complete the table).

Outcomes	Y_i	π_k	$\pi_k Y_k$			$(Y_i - \mu)$	$(Y - \mu)^2$
a $\rightarrow$	1	$^1/_6$	$(^1/_6)(1)$	=	$^1/_6$	-2.5	6.25
b $\rightarrow$	2	$^1/_6$	$(^1/_6)(2)$	=	$^2/_6$	-1.5	2.25
c $\rightarrow$	3	$^1/_6$		=	$^3/_6$	-0.5	0.25
d $\rightarrow$	4	$^1/_6$		=	$^4/_6$	0.5	0.25
e $\rightarrow$	5	$^1/_6$		=	$^5/_6$	1.5	2.25
f $\rightarrow$	6	$^1/_6$	$(^1/_6)(6)$	=	$^6/_6$	2.5	6.25
					$^{21}/_6$		17.50

$$E[Y] = \sum_k \pi_k Y_k = \frac{21}{6} = 3.5$$

$$V[Y] = \left(\frac{1}{6}\right)(6.25) + \frac{1}{6}(2.25) + \frac{1}{6}(.25) + \cdots (6.25)$$
$$= 2.92.$$

(iv) Exercises where different outcomes are assigned the same numerical value.

	Outcome		Y	π_k
Event	a	$\rightarrow$	1	$^1/_6$
Event	b	$\rightarrow$	2	$^1/_6$
.	c	$\rightarrow$	2	$^1/_6$
.	d	$\rightarrow$	3	$^1/_6$
.	e	$\rightarrow$	3	$^1/_6$
Event	f	$\rightarrow$	3	$^1/_6$

Events	X_i	π_X	$(X - \mu)$	$(X - \mu)^2$
a	1	$^1/_6$	$^{-8}/_6$	$^{64}/_{36}$
b, c	2	$^2/_6$	$^{-2}/_6$	$^4/_{36}$
d, e, f	3	$^3/_6$	$^4/_6$	$^{16}/_{36}$

Now, we refer to the variable with possible values as X.

$$\mathrm{E}[X] = \sum_{x=1}^{3} x\pi_x = (1/6)(1) + (2/6)(2) + (3/6)(3)$$
$$= 1/6 + 4/6 + 9/6$$
$$= 14/6 = 7/3,$$

$$V[X] = \sum_{x=1}^{3} (\mathrm{X} - \mu)^2 \pi_x = \left(\frac{1}{6}\right)\left(\frac{64}{36}\right) + \frac{2}{6}\left(\frac{4}{36}\right) + \left(\frac{3}{6}\right)\left(\frac{16}{36}\right)$$
$$= \frac{64 + 8 + 48}{(6)(36)} = \frac{120}{(6)(36)}$$
$$= \frac{5}{9}.$$

(v) Tossing Two Dice—consider all possible combinations (please complete the table).

D2

		1	2	3	4	5	6
	1	(1,1)	(1,2)	(1,3)			(1,6)
	2	(2,1)	(2,2)				
D1	3						
	4						
	5						
	6	(6,1)					(6,6)

Say take $D_1 + D_2$ (please complete the table).

D2

X= D1 + D2		1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6		8
D1	3	4	5	6			9
	4	5	6		8	9	10
	5	6					11
	6	7	8	9	10	11	12

From Table 13 you should be able to write the expected value, E[X], and the variance, V[X], of the sum, $X = D_1 + D_2$, of two dies.

6. Sampling distributions

Statistics involves sampling. If one takes samples of given sizes from a population, and one calculates their mean and variance, what can one expect their distributions to look like? One way to try to begin to answer this question is to actually examine the samples from a small population.

Consider a population of just 4 scores: {2, 2, 3, 5}.

Let $Y_1 = 2$, $Y_2 = 2$, $Y_3 = 3$, $Y_4 = 5$.

Then $\mu = \frac{\sum_{v=1}^{4} Y_v}{4} = \frac{2+2+3+5}{4} = \frac{12}{4} = 3$ and

Table 13 (Please complete the table, including the sums in the last row of the table)

Possible scores x	π_x	$\pi_x x$	$x-\mu$	$(x-\mu)^2$	$\pi_x(x-\mu)^2$
2	1/36	2/36	−5	25	25/36
3	2/36	6/36	−4	16	32/36
4	3/36	12/36	−3	9	27/36
5	4/36	20/36	−2	4	16/36
6	5/36		−1	1	
7	6/36	42/36	0	0	0/36
8	5/36	40/36	1	1	5/36
9	4/36	36/36	2	4	16/36
10	3/36	30/36	3	9	27/36
11	2/36	22/36	4	16	32/36
12	1/36	12/36	5	25	25/36
Sum	1	252/36 = 7			210/36

$$\sigma^2 = \frac{\sum_{v=1}^{4}(Y_v-\mu)^2}{4} = \frac{(2-3)^2+(2-3)^2+(3-3)^2+(5-3)^2}{4} = \frac{6}{4} = 3/2.$$

Note that because we have a population of scores that we divide the sum of squares by n = 4 and not n = 4 − 1 = 3. This part of the lecture explains why we use n − 1, which you would have seen, in samples.

Now suppose that we take random samples of size 2 (with replacement). Table 14 shows all the possible samples of size 2. There are 16 of them. Then the

Table 14 All possible samples of size 2, their means and three different calculations of variance

j	Sample j (n = 2)	π_j	Y_j	$\bar{Y}_j$	S_j^2	s_j^2	$\hat{\sigma}_j^2$
1	(2, 2)	1/16	4	2	0	0	1.0
2	(2, 2)	1/16	4	2	0	0	1.0
3	(2, 3)	1/16	5	2.5	0.25	0.5	0.5
4	(2, 5)	1/16	7	3.5	2.25	4.5	2.5
5	(2, 2)	1/16	4	2	0	0	1.0
6	(2, 2)	1/16	4	2	0	0	1.0
7	(2, 3)	1/16	5	2.5	0.25	0.5	0.5
8	(2, 5)	1/16	7	3.5	2.25	4.5	2.5
9	(3, 2)	1/16	5	2.5	0.25	0.5	0.5
10	(3, 2)	1/16	5	2.5	0.25	0.5	0.5
11	(3, 3)	1/16	6	3	0	0	0
12	(3, 5)	1/16	8	4	1.0	2.0	2.0
13	(5, 2)	1/16	7	3.5	2.25	4.5	2.5
14	(5, 2)	1/16	7	4	2.25	4.5	2.5
15	(5, 3)	1/16	8	4	1.0	2.0	2.0
16	(5, 5)	1/16	10	5	0	0	4.0

probability, that is, the theoretical proportion of each sample j being chosen is equal to $\pi_j = 1/16$. Let $Y_i = \sum_{v=1}^{2} Y_v$ the sum of each sample.

Then $\bar{Y}_i = \left(\sum_{v=1}^{2} Y_v\right)/2$ is the mean of each sample.

We can estimate the variance from each sample in three ways as given below:

(i) $S_j^2 = \frac{\sum_{v=1}^{2}\left(Y_{jv}-\bar{Y}_j\right)^2}{2}$,

(ii) $s_j^2 = \frac{\sum_{v=1}^{2}\left(Y_{jv}-Y_j\right)^2}{2-1}$,

(iii) $\hat{\sigma}_j^2 = \frac{\sum_{v=1}^{2}\left(Y_{jv}-\mu\right)^2}{2}$,

For example, for $j = 1$

$Y_{11} = 2, Y_{12} = 2,$

$$\bar{Y}_j = (2+2)/2 = 2,$$

$$S_j^2 = \left[(2-2)^2 + (2-2)^2\right]/2$$

$$= [0+0]/2 = 0.$$

(In general, instead of the sample size being 2, it would be denoted by n.)

For each of the samples, these values are also shown in Table 14.

It is evident that the same sample mean appears from different samples. The distribution of the means can be summarized as in Table 15. The probability (theoretical proportion) of times each mean appears is also shown in Table 15.

7. **Using the operator notation to get the expected value and variance of sample means of the means**

The expected value of the mean.

The mean of this theoretical distribution of sample means is given by

$$\begin{aligned} E[\bar{X}] &= \sum_{i=1}^{6} \pi_i \bar{X}_i \\ &= \frac{4}{16}(2) + \frac{4}{16}(2.5) + \frac{1}{16}(3) + \frac{4}{16}(3.5) + \frac{2}{16}(4) + \frac{1}{16}(5) \\ &= 3 = \mu. \end{aligned}$$

In general, $E[\bar{X}] = \mu$. Thus, we have one sample mean, $\bar{X} = \hat{\mu}$, which is an unbiased estimate of μ.

That is, the theoretical mean of the sample means is the population mean.

The variance of the means of all the possible samples.

Table 15 Distribution of means

i	$\bar{X}_i$	π_i	
1	2.0	4/16	There are only six different values for the sample means. Those with the same value are combined and probabilities added to give the probability of each value of the mean $\overline{X}_i$.
2	2.5	4/16	
3	3.0	1/16	
4	3.5	4/16	
5	4.0	2/16	
6	5.0	1/16	
Sum = 1			

The variance of the sample means is given by

$$
\begin{aligned}
V[\bar{X}] &= E\left[(\bar{X} - E[\bar{X}])^2\right] \\
&= \sum_{i=1}^{6} \pi_i (\bar{X}_i - \mu)^2 \\
&= \sum_{i=1}^{6} \pi_i (\bar{X}_i - 3)^2 \\
&= \frac{4}{16}(2-3)^2 + \frac{4}{16}(2.5-3)^2 + \frac{1}{16}(3-3)^2 + \frac{4}{16}(3.5-3)^2 + \frac{2}{16}(4-3)^2 + \frac{1}{16}(5-3)^2 \\
&= \frac{3}{4} = \frac{3/2}{2} = \frac{\sigma^2}{2}.
\end{aligned}
$$

In general, $V[\bar{X}] = \frac{\sigma^2}{n}$, where n is the sample size. That is, the theoretical variance of the means is the population variance divided by the sample size.

8. Expected value of the sample variances

We now examine the estimate imagining we just have one sample. The characteristic of the variance of a set of numbers is that it is an index of the dispersion or spread of the numbers. It involves subtracting the mean from each number, and then to avoid summing the resultant numbers whose sum will always be zero, each of these numbers is squared—this makes them all positive. The numbers resulting from subtracting a number from the mean is called a *deviate* or sometimes to stress the point, *mean deviate*. Then so that the index is not greater simply because there are more numbers, the mean or average of the sum of these squares is taken. However, an interesting thing happens if we simply take the mean.

No doubt you have come across the idea of dividing the sum of squares of numbers subtracted from their mean by $n-1$ rather n where n is the sample size. Here, we will review this effect in order to consolidate the idea of (i) the degrees of freedom and (ii) bias in estimates. Already in Table 15, where all possible samples of size 2 were listed, the variance of each sample was calculated in three ways.

Can you tell in advance what the differences are in calculating them?

We will calculate the expected value, that is, the theoretical mean across the samples of the first two estimates of the variance. You will do the same for the last one as an exercise.

To consolidate the point of this exercise, we

- consider each possible sample,
- calculate the mean of each sample,
- calculate the sum of squares of the deviates within each sample,
- divide the number of deviates by either 2 (n) or 1 $(n-1)$, where $(n = 2)$.

The last value is an estimate of the variance of all numbers from the sample. In general, we have only one sample, but by considering this theoretical example we can consider what would happen if we did have all the samples. From this

Table 16 Two estimates of the variance

i	π_i	S_i^2	s_i^2	$\pi_i S_i^2$	$\pi_i s_i^2$
1	6/16	0.0	0.0	0	0
2	4/16	0.25	0.5	1/16	2/16
3	2/16	1.0	2.0	2/16	4/16
4	4/16	2.25	4.5	9/16	18/16
				$\sum_i \pi_i S_i^2 = \frac{3}{4}$	$\sum_i \pi_i s_i^2 = \frac{3}{2}$

consideration, we can decide which is the better of these two ways of calculating the estimate of the variance of the whole set of numbers from just one sample.

The criterion is that if we could take many samples, then the calculation that gives the best value in the long run is the one to be taken. The best value is ideally the one whose expected value is the actual variance. We do not expect that every sample can give exactly the correct variance, but it would be helpful if in the long run, and on the average, we would get that value. For this purpose, we obtain the theoretical mean of the variances of the samples, that is, the *expected value* of the variances.

Table 16 shows the distribution of the first two estimates of the variance. In each case, there are only four different values.

The expected value of the first of these distributions is given by

$$\begin{aligned} E\left[S^2\right] &= \sum_{i=1}^{4} \pi_i S_i^2 \quad \left[S_j^2 \text{ involves the sample mean } \bar{X}_j\right] (\textit{dividing by}\, n) \\ &= \frac{6}{16}(0) + \frac{4}{16}(0.25) + \frac{2}{16}(1.0) + \frac{4}{16}(2.25) \\ &= \frac{3}{4} = \frac{3/2}{2} \neq \sigma^2 = 3/2. \end{aligned}$$

Thus, the theoretical average of these sample variances, ¾, is *not* the variance of the population, which is 3/2. This would suggest that the variance calculated for each sample, from which we want to infer the variance of the original set of numbers, is not a good estimate of the variance.

The expected value of the second distribution is given by

$$\begin{aligned} E\left[s^2\right] &= \sum_{i=1}^{4} \pi_i s_i^2 \quad \left[\text{This also involves the mean } \bar{X}_j\right] (\textit{dividing by}\, n-1) \\ &= \frac{6}{16}(0) + \frac{4}{16}(0.5) + \frac{2}{16}(2.0) + \frac{4}{16}(4.5) = \frac{3}{2} = \sigma^2. \end{aligned}$$

In this case, the theoretical means of the sample variances, 3/2, is the variance of the population of numbers. This would suggest that this is a good estimate of the variance of the population. We will use this example to review why we use $(n-1)$

to divide into the sum of squares to estimate the variance and to review the idea of 'degree of freedom'.

Indeed this is the general case. If the sum of squares is divided by n, the sample size, then this estimate of the variance is said to be *biased*. If it is divided by $(n-1)$, then it is not biased. As the sample size increases, this effect of bias becomes smaller.

The reason for the bias is that each sample has its own mean, and this can *vary* from sample to sample.

You will calculate the theoretical mean of $\hat{\sigma}^2$ as an exercise [$\hat{\sigma}^2$ involves the population mean μ].

9. The sampling distribution of the mean of a set of numbers

Return to the case of the example in Table 15, that is, $V[\bar{X}] = \frac{\sigma^2}{2}$, from which in general, we have that $V[\bar{X}] = \frac{\sigma^2}{n}$, where n is the sample size.

Then the standard deviation of the sampling distribution of the mean is simply given by $\sigma_{\bar{X}} = \sqrt{V[\bar{X}]} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$, where σ is the population standard deviation.

This is also known as the *standard error* of the mean.

The above results are a consequence of the central limit theorem, which in words states that

(a) The theoretical mean of the distribution of all sample means of samples of size n is the population mean μ.

(b) The standard deviation of the sample means of all samples of size n is $\frac{\sigma}{\sqrt{n}}$, where σ is the population standard deviation.

(c) As n gets larger and larger, then, irrespective of the shape of the distribution of the original numbers, the distribution of means becomes more and more normal with mean μ and variance $\frac{\sigma^2}{n}$.

An inspection of the distributions and graphs on the following pages should convince you of the reasonableness of this theorem.

In each case, the mean of the sample means is the same as the population mean.

The distribution of the sample means is positively skewed for small sample sizes, which reflects the shape of the population, but as the sample size increases, the shape becomes more and more normal. Note that for finite samples, the distribution is still discrete.

The approximation to normality can be checked by examining the proportion of means which is within 1 standard error of the mean away from the mean (in both directions). In the normal distribution, this should be approximately 68%. For the case of n = 15 above, approximately 65% of the distribution is within one standard error of the mean.

Sampling Means: Samples of size 3 (Pop: {2,2,3,5})

Sum	Mean	Freq	Prop	Cum.Prop
6.000	2.000	8.000	0.12500	0.12500
7.000	2.333	12.000	0.18750	0.31250
8.000	2.667	6.000	0.09375	0.40625
9.000	3.000	13.000	0.20313	0.60938
10.000	3.333	12.000	0.18750	0.79688
11.000	3.667	3.000	0.04688	0.84375
12.000	4.000	6.000	0.09375	0.93750
13.000	4.333	3.000	0.04688	0.98438
15.000	5.000	1.000	0.01563	1.00000

Number of Patterns=64
Number of Sample Means=9
Mean of Sample Means=3.00000
Variance of Sample Means=0.50000
Population Mean=3.00000
Population variance=1.50000

Sampling Means: Samples of size 5 (Pop: {2,2,3,5})

Sum	Mean	Freq	Prop	Cum.Prop
10.000	2.000	32.000	0.03125	0.03125
11.000	2.200	80.000	0.07813	0.10938
12.000	2.400	80.000	0.07813	0.18750
13.000	2.600	120.000	0.11719	0.30469
14.000	2.800	170.000	0.16602	0.47070
15.000	3.000	121.000	0.11816	0.58887
16.000	3.200	120.000	0.11719	0.70605
17.000	3.400	125.000	0.12207	0.82813
18.000	3.600	60.000	0.05859	0.88672
19.000	3.800	50.000	0.04883	0.93555
20.000	4.000	40.000	0.03906	0.97461
21.000	4.200	10.000	0.00977	0.98438
22.000	4.400	10.000	0.00977	0.99414
23.000	4.600	5.000	0.00488	0.99902
25.000	5.000	1.000	0.00098	1.00000

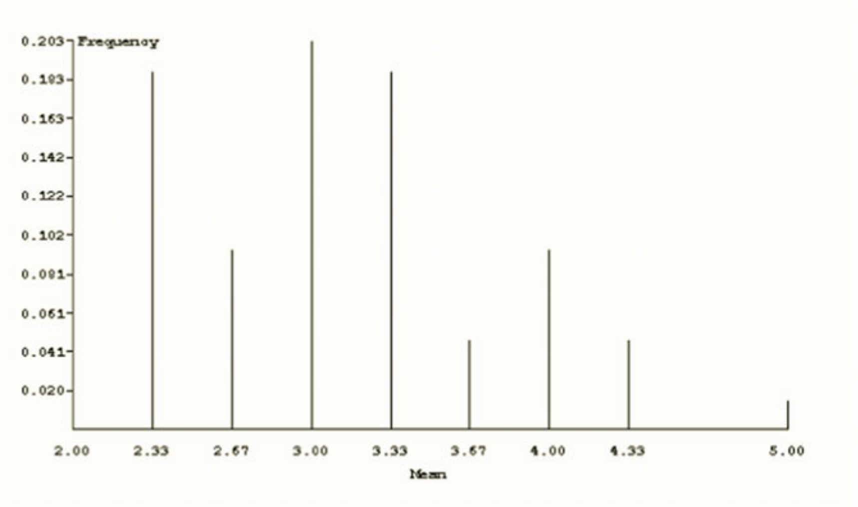
Number of Patterns=1024
Number of Sample Means=15
Mean of Sample Means=3.00000
Variance of Sample Means=0.30000
Population Mean=3.00000
Population variance=1.50000

Sampling Means: Samples of size 10 (Pop: {2,2,3,5})

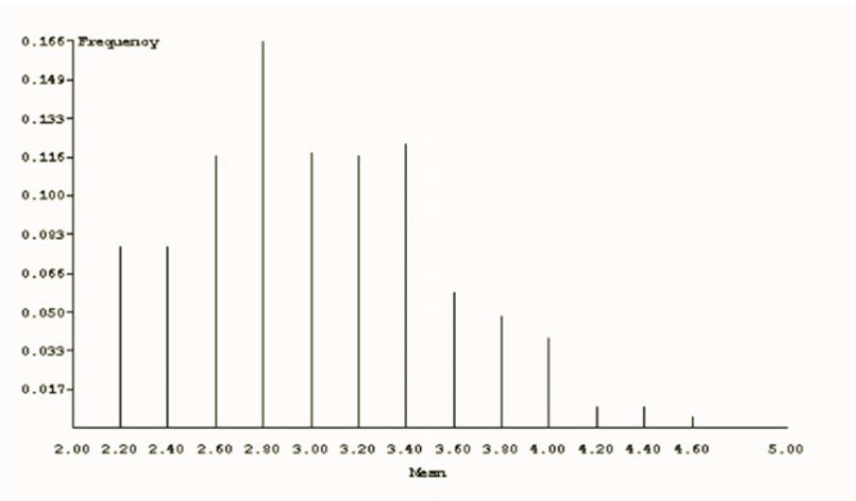
Sum	Mean	Freq	Prop	Cum.Prop
20.000	2.000	1024.000	0.00098	0.00098
21.000	2.100	5120.000	0.00488	0.00586
22.000	2.200	11520.000	0.01099	0.01685
23.000	2.300	20480.000	0.01953	0.03638
24.000	2.400	36480.000	0.03479	0.07117
25.000	2.500	54144.000	0.05164	0.12280
26.000	2.600	68640.000	0.06546	0.18826
27.000	2.700	87360.000	0.08331	0.27158
28.000	2.800	100980.000	0.09630	0.36788
29.000	2.900	102740.000	0.09798	0.46586
30.000	3.000	105601.000	0.10071	0.56657
31.000	3.100	100980.000	0.09630	0.66287
32.000	3.200	85690.000	0.08172	0.74459
33.000	3.300	74640.000	0.07118	0.81577
34.000	3.400	60525.000	0.05772	0.87349
35.000	3.500	43344.000	0.04134	0.91483
36.000	3.600	32880.000	0.03136	0.94619
37.000	3.700	22680.000	0.02163	0.96782
38.000	3.800	13650.000	0.01302	0.98083
39.000	3.900	9240.000	0.00881	0.98965
40.000	4.000	5292.000	0.00505	0.99469
41.000	4.100	2640.000	0.00252	0.99721
42.000	4.200	1650.000	0.00157	0.99878
43.000	4.300	720.000	0.00069	0.99947
44.000	4.400	300.000	0.00029	0.99976
45.000	4.500	180.000	0.00017	0.99993
46.000	4.600	45.000	0.00004	0.99997
47.000	4.700	20.000	0.00002	0.99999
48.000	4.800	10.000	0.00001	1.00000
50.000	5.000	1.000	0.00000	1.00000

```
Number of Patterns=1048576
Number of Sample Means=30
Mean of Sample Means=3.00000
Variance of Sample Means=0.15000
Population Mean=3.00000
Population variance=1.50000
```

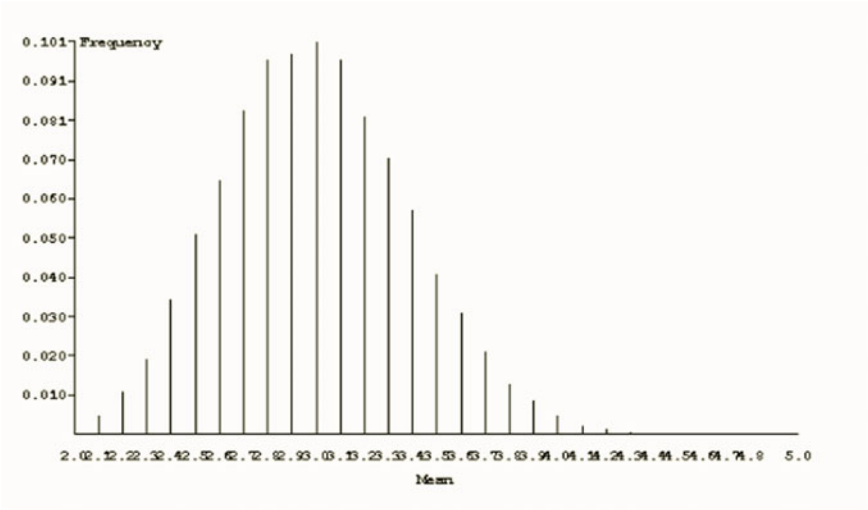
Distribution of Sampling Means: Samples of size 3 (Pop: {2,2,3,5})



Distribution of Sampling Means: Samples of size 5 (Pop: {2,2,3,5})



Distribution of Sampling Means: Samples of size 10 (Pop: {2,2,3,5})



10. **Sampling Distribution of the Difference Between Two Means**

If testing hypotheses about the difference between two means, it is necessary to know how the differences between two means are *distributed*, i.e. we need the sampling distribution of the difference between two means.

Consider the situation where samples of size n = 2 are drawn independently and at random from the two populations, P_1 and $P_{2.}$ P_1 is the same population as we used above.

$$P_1: (2, 2, 3, 5) \text{ and } P_2: (1, 2, 3).$$

The population parameters are easily calculated to be

$$\begin{aligned} \mu_1 &= 3 \qquad \mu_2 = 2 \\ \sigma_1^2 &= 1.5 \qquad \sigma_2^2 = 2/3. \end{aligned}$$

The details of the sampling distribution of means and samples of size 2 for P_1 are shown in Table 15. We will subscript the statistics of population 1 by 1 everywhere (e.g. $\bar{X}_1$) and of population 2 by 2 (e.g. $\bar{X}_2$).

$$E[\bar{X}_1] = \mu_1 = 3,$$

$$V[\bar{X}_1] = \frac{\sigma_1^2}{n} = \frac{3/2}{2} = \frac{1.50}{2} = 0.75.$$

For P_2, listing the various samples gives

i	$\bar{X}_{2i}$	π_i	
1	1	1/9	Using this (collapsed) distribution, you can easily find: $E[\bar{X}_2] = \mu_2 = 2$ $V[\bar{X}_2] = \frac{\sigma_2^2}{n} = 1/3$
2	1.5	2/9	
3	2	3/9	
4	2.5	2/9	
5	3	1/9	

Since there are 16 different samples of size 2 that can be drawn from P_1 and 9 different samples of size 2 that can be drawn from P_2, there are 9 × 16 = 144 different *pairs* of samples that can be drawn from P_1 and P_2.

In examining the distribution of their differences, we shall consider $\bar{x}_1 - \bar{x}_2$ for each possible pair of samples. The *collapsed* distribution of these differences between means can be obtained from Table 17.

This distribution can be collapsed by inspection as in Table 18.

Table 17 ($\overline{X}1 - \overline{X}2$, frequency occurrences (out of 144) of this pair of values $\bar{x}_1, \bar{x}_2$)

$(\bar{X}_1 - \bar{X}_2, f_{12})$	1	1.5	2	2.5	3
2	(1, 4)	(0.5, 8)	(0, 12)	(−0.5, 8)	(−1, 4)
2.5	(1.5, 4)	(1, 8)	(0.5, 12)	(0, 8)	(−0.5, 4)
3	(2, 1)	(1.5, 2)	(1, 3)	(0.5, 2)	(0, 1)
3.5	(2.5, 4)	(2, 8)	(1.5, 12)	(1, 8)	(0.5, 4)
4	(3, 2)	(2.5, 4)	(2, 6)	(1.5, 4)	(1, 2)
5	(4, 1)	(3.5, 2)	(3, 3)	(2.5, 2)	(2, 1)

Table 18 Distribution of the difference between means

i	$\left(\overline{X_1} - \overline{X_2}\right)_i$	π_i
1	4	1/144
2	3.5	2/144
3	3	5/144
4	2.5	10/144
5	2	16/144
6	1.5	22/144
7	1	25/144
8	0.5	26/144
9	0	21/144
10	−0.5	12/144
11	−1	4/144

The mean of this (theoretical) distribution is

$$\begin{aligned} E[\bar{X}_1 - \bar{X}_2] &= \sum_{i=1}^{11} \pi_i (\bar{X}_1 - \bar{X}_2)_i \\ &= \frac{1}{144}(4) + \frac{2}{144}(3.5) + \cdots + \frac{4}{144}(-1) \\ &= 144/144 = 1 = 3 - 2 = \mu_1 - \mu_2. \end{aligned}$$

In general,

$$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2.$$

The variance is

$$\begin{aligned} V[\bar{X}_1 - \bar{X}_2] &= E\big[(\bar{X}_1 - \bar{X}_2) - E[\bar{X}_1 - \bar{X}_2]\big]^2 \\ &= E\big[(\bar{X}_1 - \bar{X}_2) - 1\big]^2 \\ &= \sum_{i=1}^{11} \pi_i \big[(\bar{X}_1 - \bar{X}_2)_i - 1\big]^2 \\ &= \frac{1}{144}(4-1)^2 + \frac{2}{144}(3.5-1)^2 + \cdots + \frac{4}{144}(-1-1)^2 \\ &= 156/144 \\ &= 1\,1/12 = 3/4 + 1/3 = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}. \end{aligned}$$

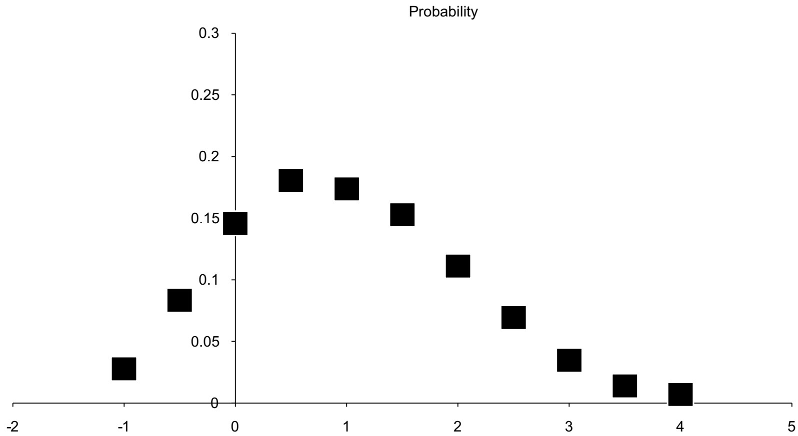
In general,

$$V[\bar{X}_1 - \bar{X}_2] = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}$$

or

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}.$$

Graphically, we have



The above results hold as n becomes larger. In fact, the distribution of the differences between the means is *normal*, for samples of a reasonably large (greater than about 30) size.

i.e. $\bar{x}_1 - \bar{x}_2 \approx N\left(\mu_{\bar{x}_1-\bar{x}_2}, \sigma^2_{\bar{x}_1-\bar{x}_2}\right)$.

If the samples are of equal size (n_1 and n_2) and drawn independently and at random from their respective populations, then

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2,$$

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2},$$

$\sigma_{\bar{x}_1-\bar{x}_2}$ is referred to as the *standard error of the difference* between the two means.

Exercises

Part A

Below is a detailed example involving the mean value operator notation. Suppose a cube (six-sided object like a die) has the letters a, b, c, d, e, f on the respective sides. Suppose the die is tossed in a way that ensures that each face has equal probability of being on top. Assign numbers to the face according to the following procedure:

$$\mathrm{a} \to 0, \mathrm{b} \to 1, \mathrm{c} \to 1, \mathrm{d} \to 2, \quad \mathrm{e} \to 2, \mathrm{f} \to 2.$$

Then the expected value and variance of the appropriate random variable can be calculated as follows. Let X be the random variable representing the numerical outcome. Then for events or outcomes, we have the probabilities shown in Table 19.

Table 19 Outcomes, random variable values and their probabilities

Outcome	X	Probability
a	0	1/6
b	1	1/6
c	1	1/6
d	2	1/6
e	2	1/6
f	2	1/6

Table 20 Values, probabilities, deviations and sum of squares

X_i	π_i	$X_i - \mu_X$	$(X_i - \mu_X)^2$
$X_1 = 0$	$\pi_1 = 1/6$	-4/3	16/9
$X_2 = 1$	$\pi_2 = 2/6$	-1/3	1/9
$X_3 = 2$	$\pi_3 = 3/6$	2/3	4/9
	$\sum_{i=1}^{3} \pi_i = 1$		

(a) Mean Value (Expected Value)

$$E[X] = \mu_X = \sum_{i=1}^{3} \pi_i X_i$$
$$= \frac{1}{6}(0) + \frac{2}{6}(1) + \frac{3}{6}(2)$$
$$= 0 + \frac{2}{6} + \frac{6}{6} = \frac{4}{3} = 1\frac{1}{3}$$

(b) Variance

$$V[X] = \sigma_X^2 = \sum_{x=1}^{3} \pi_i (X_i - \mu_X)^2$$
$$= \frac{1}{6}(\frac{16}{9}) + \frac{2}{6}(\frac{1}{9}) + \frac{3}{6}(\frac{4}{9})$$
$$= \frac{30}{54} = \frac{5}{9}$$

For the values of the random variable X, the probabilities (relative proportions) are shown in Table 20. Also shown in Table 20 are the deviation scores and their squares after the mean μ_X (expected value) is calculated.

Suppose a second die of the same kind is thrown but that the association of numbers with the outcomes is as follows:

$$a = 0, \quad b = 0, \quad c = 0, \quad d = 0, \quad e = 1, \quad f = 2.$$

1. Define a random variable Y to take these numerical values and set these out as in Tables 19 and 20.
2. Calculate the Expected Value E[Y] and the Variance V[Y].

3. Suppose the two random variables, X from the first part and Y above, are independently thrown and that the outcomes are summed to form a new variable. Calculate the Expected Value and the Variance of this new variable. That is, calculate E[W] and V[W].

Hint. Let W = X + Y be the new variable. Since X and Y are independent, every value of the X may be paired and summed with every value of the Y. The probability of a pair of outcomes is then the product of the individual probabilities. By setting up the following table, the values of the sum and their probabilities may be calculated.

Values

		Y 0	Y 1	Y 2
X	0	(0,0)	(0,1)	(0,2)
X	1	(1,0)	(1,1)	(1,2)
X	2	(2,0)	(2,1)	(2,2)

		W = X+Y 0	W = X+Y 1	W = X+Y 2
X	0	0	1	2
X	1	1	2	
X	2			

Probabilities

		Y 0	Y 1	Y 2
X	0	$\left(\frac{1}{6}\right)\left(\frac{4}{6}\right)$	$\left(\frac{1}{6}\right)(\ )$	$\left(\frac{1}{6}\right)(\ )$
X	1	$\left(\frac{2}{6}\right)(\ )$	$\left(\frac{2}{6}\right)(\ )$	
X	2	$\left(\frac{3}{6}\right)(\ )$		$\left(\frac{3}{6}\right)\left(\frac{1}{6}\right)$

		W = X+Y 0	W = X+Y 1	W = X+Y 2
X	0	$\frac{4}{36}$		
X	1			
X	2			$\frac{3}{36}$

4. Is E[X] + E[Y] = E[W]?
5. What do you notice about the relationship between V[W] and V[X] + V[Y]? Can you explain this relationship?

Part B

The estimates of the variance σ_i^2 from each possible sample of size 2 have been shown in Table 14.

Construct the frequency distribution of these samples from first principles, that is, find their expected value $E[\hat{\sigma}^2]$.

What do you notice about this value? Is it the same as σ^2, and what were the sums of squares within each sample divided by, n or n − 1? (5 marks)

Statistics Review 14: Basic Distributions for Tests of It

In this review, we review the Bernoulli and Binomial random variables and the chi-square (χ^2)distribution.

1. **Expectation and variance of a Dichotomous Random Variable**

A dichotomous random variable Xwhich takes on only the values $x = 0,\ x = 1$ is called a Bernoulli variable. The table below shows these values and their general probabilities $\pi_0,\ \pi_1$. Clearly, $\pi_0 + \pi_1 = 1$. It also shows the calculation of the expected value (theoretical mean) and variance.

x_i	π
0	π_0
1	π_1

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{2} \pi_i x_i \\
&= \pi_o + \pi_1(1) \\
&= 0 + \pi_1 \\
&= \pi_1 = \pi \\
V[X] &= \sum_{i=1}^{2} \pi_i (x_i - \pi)^2 \\
&= \pi_o(0-\pi_1)^2 + \pi_1(1-\pi_1)^2 \\
&= \pi_0\pi_1^2 + \pi_1\pi_0^2 \\
&= \pi_0\pi_1[\pi_1 + \pi_0] \\
&= \pi_0\pi_1(1) \\
&= \pi_1\pi_0 \\
&= \pi(1-\pi)
\end{aligned}
$$

Because the two values of the probability are complementary, it is common to simply write $\pi_1 = \pi$, and it is then understood that $\pi_0 = 1 - \pi$.

Suppose there are n dichotomous independent replications to give n Bernoulli variables and these are summed to give a new variable Y= $[X_1 + X_2 + \ldots X_n]$.

then

$$
\begin{aligned}
E[Y] &= E[X_1 + X_2 + \cdots, X_n] = E[X_1] + E[X_2] + \cdots E[X_n] \\
&= E[X] + E[X] + \cdots E[X] \\
&= nE[X] = n\pi.
\end{aligned}
$$

The second step above holds because the variables $X_1, X_2, \ldots, X_n$ have subscripts to distinguish them, but because they are replicates of the same variable, they have the same expected value $E[X]$.

Likewise, the variance of this sum is given by

$$\begin{aligned} V[Y] &= V[X_1 + X_2 + \cdots X_n \\ &= V[X_1] + V[X_2] + \cdots V[X_n] \\ &= nV[X] = n\pi(1 - \pi). \end{aligned}$$

The second step above holds because the variables are independent.

2. **The binomial distribution**

When n Bernoulli variables are summed, the new variable is called a *binomial* distribution. For example, suppose two Bernoulli variables are summed. Then the possible scores are

$$Y_1 = 0, Y_2 = 1, Y_3 = 2.$$

If π is the probability of $X_1 = 1,\ X_2 = 1$, then the table below shows the construction of and probabilities of Y.

X_1 (Prob)	X_2 (Prob)	$Y = X_1 + X_2$ (Prob)
0 (1-π)	0 (1-π)	0 [(1-π)(1-π)]
1 (π)	0 (1-π)	1 [(π)(1-π)]
0 (1-π)	1 (π)	1 [(1-π)(π)]
1 (π)	1 (π)	2 [(π)(π)]

The probabilities may be summarized as below:

Y	$\Pr\{Y\}$	$\Pr\{Y\}$	$\Pr\{Y\}$
0	0 $(1 - \pi)(1 - \pi)$	$(1 - \pi)^2$	$\pi^0(1 - \pi)^2$
1	$(\pi)(1 - \pi) + (1 - \pi)(\pi)$	$2[(\pi)(1 - \pi)]$	$2\left[\pi^1(1 - \pi)^{2-1}\right]$
2	$(\pi)(\pi)$	π^2	$\pi^2(1 - \pi)^0$

The last column is written in symmetric form. It can be readily generalized in the case of n replications to the equation

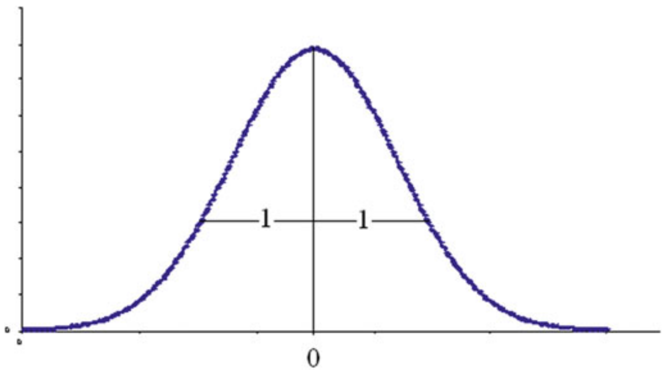
$$\Pr\{Y = y\} = \binom{n}{y}\pi^y(1 - \pi)^{n-y},$$

3. The χ^2 (chi-square distribution)

Suppose that we have a random normal distribution X. It may have come from the sampling distribution of means or it might be a hypothetical distribution. In general, by similar reasoning to that of the sampling distribution of means, random errors are taken to be normally distributed. Depending on the precision, some will have a greater variance than the others.

The *standard* normal distribution denoted generally as Z is obtained by constructing its values as follows:

$z_i = \frac{x_i - E[X]}{\sqrt{V[X]}}$ or in terms of the random variable Z, $Z = \frac{X - E[X]}{\sqrt{V[X]}}$. The curve below is the shape of a normal distribution.



3.1. The expected value and variance of a standard random normal deviate.

You do NOT have to follow the proofs below, which use properties of the operator notation, but you have to know the results.

The variable Z is called a random normal deviate. Then its expected value (theoretical mean) is given by

$$\begin{aligned} E[Z] = E[\frac{X - E[X]}{\sqrt{V[X]}}] = \frac{E[X - E[X]]}{\sqrt{V[X]}} &= \frac{E[X] - E[E[X]]}{\sqrt{V[X]}} \\ &= \frac{E[X] - E[X]}{\sqrt{V[X]}} \\ &= 0. \end{aligned}$$

Essentially, if you subtract the mean from a set of numbers, then the mean of the new numbers must be zero.

$$E[Z] = 0.$$

The variance is given by

$$V[Z] = V\left[\frac{X - E[X]}{\sqrt{V[X]}}\right] = \frac{V[X - E[X]]}{V[X]} = \frac{E[(X - [E[X])^2]}{V[X]}$$
$$= \frac{V[X]}{V[X]}$$
$$= 1.$$

Essentially, if you divide the standard deviation into a set of deviates from the mean of a set of numbers, then the variance of the new numbers must be one.

$$V[Z] = 1.$$

Furthermore, its square root, the standard deviation, will also be 1.

The standard normal deviate is used as a reference point to check if some number is significantly different from the value that might be expected under only random variation. Thus, if one computes a prediction of some number, then in probabilistic models it is not expected that the prediction will be perfect.

Suppose we have an observed value Y_i and we know the theoretical mean μ_Y and variance σ_Y^2 of this value (we come to how we might know these values). Then we can compute the standard normal deviate as follows:

$$Z_i = \frac{Y_i - \mu_Y}{\sigma_Y}.$$

Then to check if the value $Y_i$is a good prediction of μ_Y, we can compare the value of Z_i with might arise from random normal variation. You have learned that if the value is between −1.96 and 1.96, then that means it is in the range where 95% of cases under random normal variation would fall. In that case, we might consider it a good prediction.

3.2. The χ^2 distribution on 1 degree of freedom

The χ^2 distribution arises from the need to consider more than one prediction simultaneously. We begin by considering just one prediction, and in some sense it is redundant with the random normal deviate. However, it lends itself to generalization when there is more than one simultaneous prediction involved.

When we have one prediction, we consider as a frame of reference one standard normal deviate, and square it: Z_i^2. We now imagine taking many such deviates from a standard normal distribution and squaring them.

The distribution of the squares of these random normal deviate is a χ^2 distribution on 1 degree of freedom notated χ_1^2.

Then to check our prediction, we could square the calculated standard normal deviate and check if it falls between the value of 0 and $1.96^2 = 3.8416$.

3.3. The χ^2 distribution on 2 degree of freedom

If we have two simultaneous predictions, we can imagine using as the frame of reference to check its accuracy two simultaneous random normal deviates. However, we should combine these somehow to give a single summary value. This is done by imagining taking many standard random normal deviates, squaring them and summing them. This the χ^2 distribution on two degrees of freedom:

$$\chi_2^2 = Z_1^2 + Z_2^2.$$

3.4. The χ^2 distribution on n degree of freedom

The distribution of the sum of squares of n random normal deviates is χ^2 on n degrees of freedom is given by summing squaring and summing n random normal deviates as

$\chi_n^2 = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 + \cdots + Z_n^2.$

Pages 228–232 are excerpts from Glass, G. V and Stanley, J. C. (1970) Statistical Methods in Education and Psychology. New Jersey: Prentice Hall.

Statistics Review 15: Odds and Ratios

Key words: Odds, Probability, Multiplicative and logarithmic metrics, Multiplicative parameters B and D, Logarithmic parameters β and δ

Key points: Odds and probabilities are analogous to ratios and proportions. The Rasch model for dichotomous responses can be expressed in terms of the odds of a correct response or the probability of a correct response. In the former, it is in terms of multiplicative parameters B and D on the multiplicative metric. In the latter, it is in terms of logarithmic parameters β and δ on the logarithmic metric.

1. **Odds and probabilities and the relationship between the responses and the model**

Odds

'Odds' is a term which is simply a ratio, which is generally used in a context where there is a likely element involved. Thus, it is an abstraction, not something as concrete as six units of some kind of wine blended with four units of another kind of wine. For example, bookmakers offer odds of horses winning races.

$$\text{e.g.}\quad 6{:}4; 7{:}1; 1{:}10; 2{:}5; \quad \text{or} \quad 6/4; 7/1; 1/10; 2/5.$$

In the case of book makers, odds of 6:4 means that if you put \$4 that the horse will win, you will receive \$10, your \$4 plus \$6.

Probabilities (Recall probabilities are theoretical proportions)

A probability of one or the other response is relative to the total number of responses, and thus the probabilities in the above cases of the first response are

$$\text{e.g.}\quad \frac{6}{4+6}; \quad \frac{7}{1+7}; \frac{1}{10+1}; \frac{2}{5+2}.$$

Converting odds to probabilities

If in each case above we divide the numerator and the numerator by the other term of the total, we obtain

$$\frac{6/4}{1+6/4}; \frac{7/1}{1+7/1}; \frac{1/10}{1+1/10}, \frac{2/5}{1+2/5},$$

which are expressions in terms of the odds themselves, e.g. in both the numerator there are the odds of 6/4 in the first example.

In general,

$$\mathrm{P} = \frac{\mathrm{O}}{1+\mathrm{O}} \quad \text{where P is a probability and O is the related odds.}$$

2. **The Rasch model in terms of the odds of the correct response**

$$P_{ni} = \frac{O_{ni}}{1+O_{ni}} \text{ where } O_{ni} = \frac{B_n}{D_i}.$$

Here, we use the Latin letters with B the proficiency of the person and D the difficulty of an item, and we start with the multiplicative metric. We later change to the familiar logarithmic metric.

The above expression, P_{ni}, is the probability of the correct response. Thus, as the proficiency B increases the odds of answering an item correctly *increases*.

As the item difficulty increases, the odds of answering an item correctly *decreases.*

The odds are always some positive number, but can be infinitely large or small.

In terms of these parameters, the Rasch model for dichotomous responses is simply given as

$$P_{ni} = \frac{O_{ni}}{1 + O_{ni}} = \frac{B_n/D_i}{1 + B_n/D_i}.$$

Because there are only two responses, and the sum of the probabilities of the correct and incorrect responses has to be 1, we can write that the probability of an incorrect response is given by $1 - P_{ni}$. Generally, because the incorrect response is complementary to the correct response, in the case of dichotomous responses we only consider the probability of the correct response. However, for completeness we can denote

$$P_{ni1} = correct\ response; P_{ni0} = incorrect\ response.$$

Then

$$P_{ni1} = \frac{O_{ni}}{1 + O_{ni}},$$

$$\begin{aligned} P_{ni0} &= 1 - \frac{O_{ni}}{1 + O_{ni}} \\ &= \frac{1 + O_{ni} - O_{ni}}{1 + O_{ni}} \\ &= \frac{1}{1 + O_{ni}}. \end{aligned}$$

We see that the denominator is the same term in both responses. Often the denominator is presented as a single term, e.g. G_{ni}, where $G_{ni} = 1 + O_{ni}$. Then

$$P_{ni} = P_{ni1} = \frac{O_{ni}}{G_{ni}} = \frac{B_n/D_i}{G_{ni}}.$$

In the metric of the natural logarithms, $B_n = e^{\beta_n};\ D_i = e^{-\delta_i}$ and $\gamma_{ni} = 1 + e^{\beta_n - \delta_i}$ giving

$$P_{ni} = P_{ni1} = \frac{e^{\beta_n - \delta_i}}{\gamma_{ni}}.$$

Thus, written out in full, the estimation equation for the case of a person who has scored 6 on the short test of 10 dichotomous items is

$$\sum_{i=1}^{10} P_{ni} = \sum_{i=1}^{10} \frac{e^{\beta_n - \delta_i}}{\gamma_{ni}} = 6..$$

3. **Multiplicative and logarithmic metrics**

When we make use of the log metric, then

$$\begin{aligned} \beta_n &= \log B_n \quad \text{or} \quad B_n = e^{\beta_n} \\ \delta_i &= \log D_i \quad \text{or} \quad D_i = e^{\delta_i}. \end{aligned}$$

In the log metric, the odds of person υ getting item i right is $e^{\beta_n - \delta_i}$
but $e^{\beta_n - \delta_\iota} = \frac{B_n}{D_i}$
(log metric) (multiplicative metric)
or $\beta_n - \delta_i = \log \frac{(B_n)}{D_i} = \log B_n - \log D_i$.

Therefore, the probability of person n getting item i right in the log metric is given by $\frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_\iota)}}$.

Statistics Reviews Exercises Solutions

Statistics Review 1

1.

(a) 14,
(b) 8,
(c) 4.

2.

(a) $X_2 + X_3$
(b)
$$(X_1 + X_2 + X_3 + X_4)^2$$
$$= X_1^2 + X_2^2 + X_3^2 + X_4^2 + 2X_1X_2 + 2X_1X_3 + 2X_1X_4 + 2X_2X_3 + 2X_2X_4 + 2X_3X_4$$
(c) $X_1^2 + X_2^2 + X_3^2 + X_4^2 + X_5^2$

3.

(a) $5 \sum_{i=1}^{4} X_i$
(b) $\left(\sum_{i=1}^{4} X_i \right)^2$

Statistics Review 2

1.

(a) 43.82%,
(b) 93.82%,
(c) 6.18%,
(d) 94.

2. −0.84.
3. 1.645.

Statistics Review 3

1.

   (a) 325,
   (b) No, the meaning of the variable is not obvious. A high score can be obtained by a tall person with not so high non-verbal intelligence as well as a short person with high non-verbal intelligence.

2.

   (a) 684
   (b) Yes, the added variable may be some sort of 'development'. A higher score tends to be obtained by a tall person with a high non-verbal intelligence level, and a lower score tends to be obtained by a shorter person with a lower non-verbal intelligence level.
   (c) In the school-age population, the physical development and the intellectual development tend to go together in terms of 'growing'. Non-verbal tests of intelligence (encoding and decoding ability) is related to mathematical ability that increase as children grow. In developing countries, a good nutritious environment tends to bring about (i) better physical development and (ii) better intellectual development, and thus these two variables are correlated.

Statistics Review 4

1. $c_{xy} = 76$, $r = 0.89$,
2. $\hat{Y}_i = 5.48 + 1.30X_i$,
3. 26.43,
4. −9.66.

Statistics Review 5

1.

   (a) {1, 2, 3, 4, 5, 8},
   (b) {4, 5},
   (c) {1, 2, 3, 4, 5},
   (d) {3, 4, 5}.

2.

   (a) A = {H1, H2, H3, H5, T1, T2, T3, T5}
   B = {H1, H3, H5}
   C = {T2, T4, T6},
   (b) 0.667,
   (c) B and C because $P\{B\mathrm{U}C\} = P\{B\} + P\{C\}$.

3.

(a) S = {(H, H, H), (H, H, T), (H, T, H), (H, T, T), (T, H, H), (T, H,T), (T, T, H), (T, T, T)},

(b)

(i) 0.125,
(ii) 0.375,
(iii) 0.5.

(c) Mutually exclusive.

Statistics Review 6

1. 3^{3x},
2. $(a)^{8}$,
3. 2^{a},
4. 125^{n},
5. 864.

Statistics Review 7

1. log 15,
2. log 2,
3. log 16,
4. $\log_{10} 300$.

Statistics Review 8

1. $\Pr\{C_1 = T\} = 0.45$, $\Pr\{C_2 = T\} = 0.55$.
2. 0.40.

Statistics Review 9

1. $P\{A \cap B\} = P(A) \text{x} P(B)\, 0.375 = 0.75 \text{ x } 0.5$,
2. $P\{A \cap B\} \neq P(A) \text{x} P(B)\, 0.5 \neq 0.5 \text{ x } 0.75$.

Statistics Review 13

Part A

1.

Outcome	Y	Probability
a	0	1/6
b	0	1/6
c	0	1/6
d	0	1/6
e	1	1/6
f	2	1/6

Y_i	π_i	$Y_i - \mu_Y$	$(Y_i - \mu_Y)^2$
$Y_1 = 0$	$\pi_1 = 4/6$	−1/2	1/4
$Y_2 = 1$	$\pi_2 = 1/6$	1/2	1/4
$Y_3 = 2$	$\pi_3 = 1/6$	3/2	9/4
	$\sum_{i=1}^{3} \pi_i = 1$		

2. E[Y] = 0.5, V[Y] = 0.584,
3. E[W] = 1.833, V[W] = 1.139,
4. Yes, 1.333 + 0.5 = 1.833,
5. Variables X and Y are independent
   V[W] = V[X] + V[Y]
   0.555 + 0.584 = 1.139.

Part B

i	π_i	$\hat{\sigma}_i^2$	$\pi_i \hat{\sigma}_i^2$
1	1/16	0.0	0
2	4/16	0.5	2/16
3	4/16	1.0	4/16
4	2/16	2.0	4/16
5	4/16	2.5	10/16
6	1/16	4.0	4/16
		Sum	3/2

$E[\hat{\sigma}^2] = 3/2 = \sigma^2$ is the population variance. Divided by n-1.

Index

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8