

# Glossary of Terms

**Alpha level** The level of statistical significance selected prior to a test for incorrectly rejecting a true null hypothesis, e.g., .05 alpha level of significance. (See Type I error)

**Alternative hypothesis** A statistical hypothesis that indicates a difference in population parameters. For example, the means of two populations are different, i.e., possible outcomes not stated in the null hypothesis

**Analysis of variance** A technique that tests whether the dependent variable means of three or more mutually exclusive groups are statistically significantly different at a specified level of significance. The F-test is a ratio of MS Between Groups divided by MS Within Groups

**Bell shaped curve** Describes a normal or symmetrical distribution of data in which intervals around the mean are known

**Bimodal** A frequency distribution of data that has two modes, i.e., two scores that occur most frequently in a set of data

**Binomial distribution** A probability distribution generated by taking  $(a + b)$  to the  $n$ th power. Used in a binomial test to determine whether the probability of two outcomes exceed the chance level of occurrence

**Binomial test** A non-parametric test, which doesn't depend on any population data characteristics and measures whether a distribution of scores results in a binomial distribution (each outcome is equally likely). For example, if you tossed an unbiased coin 1,000 times it should land approximately 500 heads and 500 tails

**Bootstrap** An approach that samples with replacement to generate a sampling distribution of a statistic that serves as the population distribution. The mean of the bootstrap sampling distribution or bootstrap estimate is used to determine the amount of bias in the random sample

**Central limit theorem** A theorem that provides a mathematical basis for using the normal distribution, as a sampling distribution of a statistic for a given sample size, to test a statistical hypothesis. For example, the theorem states that a sampling distribution of means for a given sample size is (1) normally distributed, (2) the sampling distribution mean is equal to the population mean, and (3) the sampling distribution variance is equal to the variance divided by the sample size

**Central tendency** A concept that implies most scores fall in the middle of a symmetrical distribution with the scores spreading out evenly toward both tails of the distribution

**Chi-square distribution** A probability distribution or family of curves generated by the difference between observed and expected frequencies. The sampling distribution of chi-square values is used in both parametric and non-parametric tests of significance

**Chi-square statistic** A non-parametric test that measures the difference between the observed frequencies and expected frequencies in two or more groups

**Combinations** The number of ways in which different subsets of events or numbers can be selected

**Conditional probability** The probability of an event B is based upon the occurrence or non-occurrence of event A

**Confidence interval** A high and low value which forms an interval around the sample statistic that should contain the population parameter. The interval will be different depending upon the percentage used, i.e., 68, 95, or 99 %

**Confidence level** A percentage that indicates how certain we are that the interval around the sample statistic contains the population parameter (see alpha level)

**Correlation** A statistic that indicates the strength and direction of association between two sets of scores. The strength is indicated by a correlation value closer to 1.0 and the direction indicated by a  $\pm$  sign. A positive correlation indicates that both variables increase in value across the range of scores while a negative correlation indicates that one set of scores increases as the other set of scores decreases

**Cross-validation** An original sample of data is randomly split into two equal samples, then a sample statistic is computed using one sample of data and applied to the other sample of data

**Cumulative frequency distribution** A frequency distribution of raw scores that indicates successive addition of the number of events, individuals, or objects up to the total number or 100 %

**Degrees of freedom** The number of observations or groups minus the restrictions placed upon them. For example, if four out of five sample means are known, the one remaining unknown sample mean can be determined, hence  $df = 5 - 1 = 4$

**Dependent t-test** A statistical test of whether two sample means from the same subjects or group are significantly different. Also called a paired t-test or correlated t-test

**Dichotomous population** A population of data that can be divided into two mutually exclusive categories

**Directional hypothesis** A hypothesis that states one population parameter is greater than the other. The direction can be stated in a positive or negative direction, e.g., boys' verbal scores will be lower on average than girls' verbal scores. A one-tailed test because the region of rejection is only in one tail of the sampling distribution

**Effect size** Conveys the magnitude of difference in standard units between the mean of the experimental group and the mean of the control group. Used in conjunction

with sample size, alpha level, and direction of the statistical hypothesis to select a value for power

**Equally likely events** Events, individuals, or objects that have the same chance of being selected

**Exponential function** A relationship between two sets of data points that does not have a constant rate of change for a random variable  $X$ , i.e.,  $Y = 2^X$

**F-curve** A positively skewed frequency distribution of  $F$  values for specific degrees of freedom

**F-distribution** A probability distribution or family of curves that require two degrees of freedom. The normal,  $t$ , and chi-square distributions are special cases of the  $F$ -distribution

**F-test** In Analysis of Variance, the test to determine if sample means are different beyond chance expectation. The  $F$ -test is the ratio of  $MS$  Between Groups divided by the  $MS$  Within Groups

**Factoring** A product of sequential numbers that indicate the total number of choices possible (see factorial notation)

**Factorial notation** Indicated as  $n!$ , i.e.,  $3! = 3 \times 2 \times 1 = 6$

**Finite distribution** A population of data where the number of individuals, objects, or events is known, hence exact probabilities of occurrence can be computed

**Frequency distribution** A tabulation of data that indicates the number of times a score or value occurs

**Hartley F-max test** A test of whether three or more sample variances are statistically different. The largest sample variance is divided by the smallest sample variance to form an  $F$ -ratio with degrees of freedom from the two sample sizes

**Heterogeneity** Refers to a grouping of dissimilar individuals, objects, or events

**Histogram** A bar chart that indicates the frequency of numbers on the  $Y$  axis and the mutually exclusive groups or categories on the  $X$  axis. (Also, see Pie Chart)

**Homogeneity** Refers to a grouping of similar individuals, objects, or events

**Independent t-test** A statistical test of whether two independent sample means are significantly different implying that the two population means are different

**Infinite population** A population of data where the number of individuals, objects, or events are too numerous to count, hence exact probabilities of occurrence cannot be computed

**Intercept** The intercept is the point in a linear equation where the line of best fit crosses the  $Y$ -axis. The intercept is the predicted value of  $Y$  when the  $X$  variable equals zero. The value  $a$  in the linear regression equation:  $Y = a + bX + e$ .

**Interquartile range** A score that represents the distance between the first and third quartile. It indicates the range of scores in the middle 50 % of a frequency distribution

**Jackknife** An approach that uses a single sample of data and computes sample statistics based on different  $n - 1$  sample sizes

**Joint probability** The probability of two events occurring that is determined by multiplying the independent probability of each event

- Kurtosis** A measure that indicates the flatness or peakedness of the frequency distribution of scores. Leptokurtic implies a peaked distribution, mesokurtic a bell-shaped normal distribution, and platykurtic a flattened distribution of scores
- Law of complements** Given the probability of event A,  $P(A)$ , the complement is  $1 - P(A)$ , or the remaining probability since  $P(A) + [1 - P(A)] = 1$
- Leaves** The numbers to the right of the vertical line in a stem-and-leaf plot
- Level of significance** The probability of making a Type I error (see alpha level)
- Linear function** An angled straight line of data points that indicate a constant rate of change for a random variable, X, i.e.,  $y = bX$
- Linear regression (equation)** A statistical technique designed to predict values of Y (dependent variable) from one or more X variables (independent predictor variables). The regression equation:  $Y = a + bX + e$ .
- Line of best fit** In linear regression, the line formed by the predicted Y values that pass through the scatterplot of X and Y values. The line indicates the best prediction that minimizes the sum of squared errors of prediction
- Mean** The arithmetic mean computed as the sum of a set of scores divided by the number of scores. Typically referred to as a measure of central tendency
- Mean square** A variance estimate computed by dividing the sum of squares by the degrees of freedom
- Median** The middle score in a distribution of odd-numbered scores or the midpoint in an even-numbered set of scores. Typically referred to as a measure of central tendency
- Meta-analysis** A statistical procedure that averages the effect sizes across several studies to determine the overall significance of a large number of research studies on the same topic
- Mode** The most frequently occurring score in a set of scores. It is possible to have a single modal score (unimodal), two scores that occur the most (bimodal), or even three or more scores that occur the most. Typically referred to as a measure of central tendency
- Monte Carlo** An approach that describes a statistical technique that simulates data and approximates probability density functions of population distributions to study the robustness and properties of statistical tests
- MS between groups** The sum of the squared deviations of group means around the grand mean weighted (multiplied) by the sample size of each group and divided by the number of groups minus one. Indicates whether the group means are similar or different based on how much they vary
- MS within groups** The sum of squared deviations of individual scores around each group mean divided by the number of scores in each group minus the number of groups. Indicates how much the scores vary within each group
- Multiplication law** The independent probabilities of two events can be multiplied to obtain their probability of joint occurrence, i.e.,  $P(J) = P(A) * P(B)$
- Non-directional hypothesis** A hypothesis that states two population parameters are different, rather than one population parameter is greater than the other. A two-tailed test because the region of rejection is in both tails of the sampling distribution

- Normal curve** A symmetric distribution of data based on a mathematical equation formulated by DeMoivre in 1733 and further developed by Carl Fredrick Gauss
- Normal distribution** A frequency distribution of scores that when graphed produces a symmetrical, bell-shaped distribution with skewness and kurtosis of zero. Sometimes referred to as a mesokurtic distribution
- Null hypothesis** A statistical hypothesis that indicates no difference in population parameters. For example, the means of two populations are equal. The null hypothesis is either retained or rejected in favor of an alternative hypothesis
- Ogive** A graph of the cumulative frequency distribution of data that has a characteristic S-shaped curve
- One-sample t-test** A statistical test of whether a sample mean is significantly different from a population mean
- One-way analysis of variance** A statistical test that is an extension of the independent t-test to test whether three or more independent sample means are statistically different implying that the population means are different
- Outlier** An extreme or influential score or data value that affects the sample statistic, e.g., sample mean
- Parameter(s)** Population values or characteristics that are estimated by sample statistics, e.g., population mean or population correlation
- Parametric statistics** Parametric or inferential statistics are based upon being able to randomly draw a sample from a well defined population, estimate the sample statistic, and make an inference about the population parameter. For example, the sample mean is an estimate of the population mean
- Permutations** A technique used to determine the number of different ways individuals, objects, or events can be ordered
- Pie chart** A circle with portions or parts of the circle indicated for each mutually exclusive group or category. (Also, see Histogram)
- Population** A set of individuals or scores that are well defined and share some characteristic in common. Typically, population data is randomly sampled and sample statistics computed to estimate the population values because the population is typically too large to measure all the data. (See Parametric Statistics)
- Power** The probability of rejecting the null hypothesis when it is false. The expression,  $1 - \beta$ , is used to indicate the level of power. Values of .80 are typically selected for power; power is a function of sample size, alpha level, effect size, and directional nature of the statistical hypothesis (one-tailed or two-tailed test)
- Probability** The ratio of the number of favorable outcomes to the total possible number of outcomes
- Properties of estimators** Important characteristics we want sample statistics to possess as estimates of population parameters, i.e., unbiased, consistent, efficient, and sufficient
- Pseudo random numbers** Numerical values typically generated by a random number generator on a computer, but not truly independent or unbiased because they will eventually correlate and repeat

**Quartile** A score that divides a set of data into four equal divisions, i.e., first quartile is a score that separates the bottom 25 % of the data in a frequency distribution from the other data values

**Random assignment** The random process of assigning individuals, objects, or events to a group, i.e., random assignment of individuals to experimental and control groups

**Random numbers** Independent, unbiased numerical values that have an equally likely chance of being selected

**Random number generator** A mathematical algorithm in a software program that is run on a computer to generate pseudo random numbers

**Random sample** A sample of data from a well-defined population where every individual, object, or event has an equally likely chance of being selected

**Random sampling** The process of selecting individuals, objects, or events from a well-defined population in which all members have an equal and independent chance of being selected. Not the same as random assignment

**Range** A score that indicates the distance between the highest and lowest data value in a set of data

**Region of rejection** The area under a sampling distribution where sample statistics fall that is highly improbable if the null hypothesis is true

**Regression weight** In regression analysis, the regression coefficient or slope of the line of best fit that passes through the predicted  $Y$  values. The value  $b$  in the linear regression equation:  $Y = a + bX + e$ . A weight computed by the least squares method of minimizing the sum of squared errors of prediction

**Repeated measures ANOVA** A statistical procedure in which subjects are measured two or more times and the total variation of scores is partitioned into three components: (1) variation among subjects, (2) variation among occasions (time), and (3) residual variation

**Sample** A random selection of individuals, objects, or events from a well-defined population of data

**Sample error** The difference between a sample statistic and the population parameter

**Sampling error** The error in using a sample statistic as an estimate of a population parameter

**Sampling distribution** A probability frequency distribution of a sample statistic formed for all possible random samples of a given sample size. Examples of sampling distributions include: (1) sampling distribution of means, (2) t-distribution, (3) chi-square distribution, and (4) F-distribution

**Sampling with replacement** Each randomly sampled data point is returned to the population before another data point is randomly sampled, therefore it is possible for a data point to be selected more than once

**Sampling without replacement** Each randomly sampled data point is not returned to the population before another data point is randomly sampled, therefore each data point is uniquely drawn and can not be selected again

**Scheffe post-hoc test** A type of post-hoc “t-test” for conducting multiple comparisons of group mean differences after an analysis of variance F-test

- Skewness** A measure of deviation from symmetry in a frequency distribution of scores. Negative skew indicates a distribution with more scores above the mean. Positive skew indicates a distribution with more scores below the mean
- Slope** The amount of change in Y that corresponds to a change of one unit in X. (See Regression Weight)
- Standard deviation** The square root of the average squared deviations of scores around the mean. A measure of how much the individual scores deviate from the mean
- Standard errors of statistic** The standard deviation of the sampling distribution of the statistic that indicates the amount of error in estimating the population parameter
- Standard score** A score computed by taking the deviation of the raw score from the group mean divided by the group standard deviation, i.e., z-score
- Statistic** A sample value that estimates a population parameter
- Stem and leaf** A graphical display that illustrates the shape of a distribution of scores
- Sum of squared deviations** The deviations of each score from the group mean that is squared and then summed for all scores
- Symmetric distribution** A sampling distribution or frequency distribution of scores that is the same on either side of the median value. The normal distribution is an example of a symmetric distribution
- t-distribution** A probability distribution or family of t-curves for different degrees of freedom that is used to determine whether an obtained t value between two sample means is statistically significant at a specified alpha level
- Tchebysheff inequality theorem** A theorem that indicates the percentage of data between intervals around the mean regardless of the shape of the frequency distribution of data
- Type I error** The rejection of the null hypothesis of no difference in population parameters when it is true, i.e., the probability that a null hypothesis would be rejected in favor of an alternative hypothesis. The probability is set by selection of an alpha level (see Alpha level). If the alpha level is set at .05, then 5% of the time a true null hypothesis would be incorrectly rejected in favor of the alternative hypothesis. The symbol,  $\alpha$ , is used to refer to this type of error
- Type II error** The retention of a null hypothesis of no difference in population parameters when it is false, i.e., the probability that we failed to reject the null hypothesis in favor of an alternative hypothesis. The symbol,  $\beta$ , is used to refer to this type of error
- Uniform distribution** A rectangular distribution of scores that are evenly distributed in the range of possible values
- Unimodal distribution** A symmetrical distribution with a single mode
- Variance** A positive value that measures how scores vary around a group mean. If all scores are the same, then the variance is zero. Calculated as the sum of squared deviations around the group mean divided by the number of scores
- z-score** Sometimes called a standard score. A frequency distribution of raw scores that have been standardized to a mean of zero and a standard deviation of one.

A z score indicates the direction and degree to which a score deviates from the mean of a distribution of scores

**z-test** A statistical test for the significant difference in independent or dependent population proportions

# Appendix

## Statistical Tables

**Table A1** Areas under the normal curve (z-scores)

Second decimal place in z										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4373	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4826	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									

**Table A2** Distribution of *t* for given probability levels

<i>df</i>	Level of significance for one-tailed test					
	.10	.05	.025	.01	.005	.0005
	Level of significance for two-tailed test					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.992
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
$\infty$	1.282	1.645	1.960	2.326	2.576	3.291

**Table A3** Distribution of r for given probability levels

<i>df</i>	Level of significance for one-tailed test			
	.05	.025	.01	.005
	Level of significance for two-tailed test			
	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.540	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.576	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641
14	.426	.497	.574	.623
15	.412	.482	.558	.606
16	.400	.468	.542	.590
17	.389	.456	.528	.575
18	.378	.444	.516	.561
19	.369	.433	.503	.549
20	.360	.423	.492	.537
21	.352	.413	.482	.526
22	.344	.404	.472	.515
23	.337	.396	.462	.505
24	.330	.388	.453	.496
25	.323	.381	.445	.487
26	.317	.374	.437	.479
27	.311	.367	.430	.471
28	.306	.361	.423	.463
29	.301	.355	.416	.486
30	.296	.349	.409	.449
35	.275	.325	.381	.418
40	.257	.304	.358	.393
45	.243	.288	.338	.372
50	.231	.273	.322	.354
60	.211	.250	.295	.325
70	.195	.232	.274	.303
80	.183	.217	.256	.283
90	.173	.205	.242	.267
100	.164	.195	.230	.254

**Table A4** Distribution of Chi-square for given probability levels

df	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.00016	.00663	.00393	.0158	.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.345	16.266
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.78	5.989	7.779	9.488	11.668	13.277	18.467
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.515
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620

26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.302
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703
32	16.362	17.783	20.072	22.271	25.148	27.373	31.336	35.665	38.466	42.585	46.194	50.487	53.486	62.487
34	17.789	19.275	21.664	23.952	26.938	29.242	33.336	37.795	40.676	44.903	48.602	52.995	56.061	65.247
36	19.233	20.783	23.269	25.643	28.735	31.115	35.336	39.922	42.879	47.212	50.999	55.489	58.619	67.985
38	20.691	22.304	24.884	27.343	30.537	32.992	37.335	42.045	45.076	49.513	53.384	57.969	61.162	70.703
40	22.164	23.838	26.509	29.051	32.345	34.872	39.335	44.165	47.269	51.805	55.759	60.436	63.691	73.402
42	23.650	25.383	28.144	30.765	34.147	36.755	41.335	46.282	49.456	54.090	58.124	62.892	66.206	76.084
44	25.148	26.939	29.787	32.487	35.974	38.641	43.335	48.396	51.639	56.369	60.481	65.337	68.710	78.750
46	26.657	28.504	31.439	34.215	37.795	40.529	45.335	50.507	53.818	58.641	62.830	67.771	71.201	81.400
48	28.177	30.080	33.098	35.949	39.621	42.420	47.335	52.616	55.993	60.907	65.171	70.197	73.683	84.037
50	29.707	31.664	34.764	37.689	41.449	44.313	49.335	54.723	58.164	63.167	67.505	72.613	76.154	86.661
52	31.246	33.256	36.437	39.433	43.281	46.209	51.335	56.827	60.332	65.422	69.832	75.021	78.616	89.272
54	32.793	34.856	38.116	41.183	45.117	48.106	53.335	58.930	62.496	67.673	72.153	77.422	81.069	91.872
56	34.350	36.464	39.801	42.937	46.955	50.005	55.335	61.031	64.658	69.919	74.468	79.815	83.513	94.461
58	35.913	38.078	41.492	44.696	48.797	51.906	57.335	63.129	66.816	72.160	76.778	82.201	85.950	97.039
60	37.485	39.699	43.188	46.459	50.641	53.809	59.335	65.227	68.972	74.397	79.082	84.580	88.379	99.607
62	39.063	41.327	44.889	48.226	52.487	55.714	61.335	67.322	71.125	76.630	81.381	86.953	90.802	102.166
64	40.649	42.960	46.595	49.996	54.336	57.620	63.335	69.416	73.276	78.860	83.675	89.320	93.217	104.716
66	42.240	44.599	48.305	51.770	56.188	59.527	65.335	71.508	75.424	81.085	85.965	91.681	95.626	107.258
68	43.838	46.244	50.020	53.548	58.042	61.436	67.335	73.600	77.571	83.308	88.250	94.037	98.028	109.791
70	45.442	47.893	51.739	55.329	59.898	63.346	69.335	75.689	79.715	85.527	90.531	96.388	100.425	112.317

Note: For larger values of df, the expression  $\sqrt{\frac{\chi^2}{df}} - \sqrt{2df - 1}$  may be used as a normal deviate with unit variance, remembering that the probability for  $\chi^2$  corresponds with that of a single tail of the normal curve

**Table A5** The F-distribution for given probability levels (.05 Level)

df <sub>1</sub>	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	$\infty$
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.8	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.49	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.15	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.1	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

**Table A6** The distribution of F for given probability Levels (.01 Level)

df <sub>1</sub>	df <sub>2</sub>	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366	6366
2	98.5	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.25	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.66	2.66
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.76	2.67	2.58	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.10	1.10



# Author Index

## B

Bargmann, Rolf, E., 217  
Bashaw, W.L., 217  
Bernoulli, James, 110  
Bock, Darrell, R., 217  
Bottenberg, Robert A., 217

## C

Cohen, Jacob, 246, 247, 249  
Cureton, Edward, E., 246

## D

Darwin, Charles, 207  
DeMoivre, Abraham, 110, 265  
Derenzo, Stephen, E., 112  
Draper, Norman, R., 217

## F

Findley, Warren, B., 217  
Fisher, Ronald, A., Sir, 126, 128, 136, 185,  
195, 246  
Fiske, Donald, W., 246

## G

Galton, Francis, Sir, 207  
Gauss, Carl Fredrich, 110, 265  
Glass, Gene, 246, 247, 259  
Gordon, Mordecai, H., 246, 249, 251, 252  
Gossett, William, S., 120, 185, 186, 195  
Graybill, Franklin, F.A., 217

## H

Hartley, H.O., 129  
Hedges, L., 250

## J

Jones, Lyle, V., 246

## L

Loveland, Edward, H., 246

## P

Pearson, Egon, S., 246  
Pearson, Karl, 114, 167, 207, 210,  
211, 246

## R

Rand, McNally, 48  
Rosenthal, R., 250  
Rubin, D., 250

## S

Scheffe, Henry, 198  
Smith, Harry, Jr., 217  
Student, 120, 185

## T

Thurstone, L.L., 207

**U**

UNIVAC, 48

**W**

Ward, Joe, H., 217

Winer, Ben, J., 217

Winkler, Henry, 129

# Subject Index

## A

- Addition law of probability
  - ADDITION R program (*see* ADDITION R program)
  - description, 23
  - exercises, 25–26
  - law of complements, 23–24
  - odds of events, 23
  - program output, 24–25
  - theory of probability, 23
  - true/false questions, 40
- ADDITION R program
  - definition, 24
  - DiceFreq object, 24
  - DiceSum, 24
  - exercises, 25–26
  - multiplication law of probability, 26
  - odd and even numbers, 25–26
  - output, 24–25
- Alpha level, 261
- Alternative hypothesis, 261
- Analysis of variance (ANOVA)
  - technique
    - definition, 261
    - description, 195
    - F tables, 196
    - heterogeneous, 195
    - homogeneous, 195
    - one-way ANOVA (*see* One-way ANOVA)
    - sample mean differences, 195
    - SET A and SET B, 195
- ANOVA technique. *See* Analysis of variance (ANOVA) technique
- Areas, normal curve (z-scores), 269–270

## B

- Bell shaped curve, 261
- BELL-SHAPED CURVE R program
  - description, 90
  - Kvals, 90
  - normal distribution exercises, 91–93
  - output, 91
  - population types, 90
- Binomial distribution, 261
- BINOMIAL R program
  - exercises, 108–109
  - numReplications* and *numTrials* sample, 107
  - output, 108
- Binomial test, 261
- Bootstrap procedure
  - definition, 261
  - description, 237–238
  - estimate, 239
  - program, 240–241
  - “pseudo” population, 239

## C

- Central limit theorem
  - CENTRAL R program (*see* CENTRAL R program)
  - definition, 261
  - description, 93–94
  - exercises, 99–101
  - homogeneous and heterogeneous population, 94
  - mathematical function, 94
  - non-normal populations, 94
  - output, CENTRAL R program, 95–99
  - sampling distribution, 94–95

- Central limit theorem (*cont.*)
  - statistical analysis, 94
  - true/false questions, 102
- CENTRAL R program
  - exercises, 99–101
  - infinite number, 95
  - output, 95–99
  - user-defined variables, 95
- Central tendency
  - constant data set, 74
  - data transformations, 75
  - definition, 262
  - description, 73
  - mean and median score, 73
  - MEAN-MEDIAN exercises, 75–77
  - MEAN-MEDIAN R program (*see* MEAN-MEDIAN R program)
  - output, MEAN-MEDIAN program, 74–75
- Chi-square distribution
  - CHISQUARE R program (*see* CHISQUARE R program)
  - definition, 262
  - numerator and denominator, 114–115
  - sample size, 114
  - statistic, 115
  - z-score, 114
- CHISQUARE R program
  - barplot function, 115
  - exercises, 117–120
  - output, 116–117
  - sample sizes, 115
- Chi-square statistics, 262
- Chi-square test
  - categorical variables, 167
  - computation, 168–169
  - CROSSTAB R program, 170–173
  - cross-tabulation, data, 167, 168
  - expected cell frequencies, 168
  - independence, 170
  - interpretation, 169
  - men and women, expected cell values, 168
  - statistical tests, 167
  - superintendent decides, 168
  - two-by-two table, 169, 170
- COMBINATION R program
  - description, 63
  - exercises, 65–66
  - factor and table functions, 63
  - FreqPopTable, 63
  - FreqPop vector, 63
  - hist function, 64
  - output, 64–65
  - round and invisible function, 63
- Combinations and permutations
  - definition, 262
  - description, 34
  - exercises, 38–39
  - factorial notation, 35
  - formula, 36
  - fundamental rules, 34–35
  - permutations, 35
  - possible combinations, 36
  - program output, 38
  - random sample, 34
  - relative frequencies, 36–37
  - R program, 37–38
  - true/false questions, 41
- Comprehensive R Archive Network (CRAN), 1
- Conditional probability
  - CONDITIONAL R program (*see* CONDITIONAL R program)
  - definition, 262
  - description, 29
  - equally likely events, 30
  - event A and B, 29
  - exercises, 32–34
  - marginal probabilities, 30
  - program output, 32
  - true/false questions, 41
- CONDITIONAL R program
  - description, 31
  - dimnames set, 31
  - exercises, 32–34
  - output, 32
  - print.char.matrix, 31
  - rep function, 31
- Confidence interval, 262
- Confidence level, 262
- CONFIDENCE R Program
  - output, 142–143
  - population mean, 142
  - sample sizes and confidence levels, 142
- Correlation
  - definition, 262
  - Pearson correlation (*see* Pearson correlation)
- CORRELATION R Program
  - cor function, 212
  - exercises, 213–216
  - output, 212–213
  - rho value, 212
- CRAN. *See* Comprehensive R Archive Network (CRAN)
- CROSSTAB R program, 170–173
- Cross-validation
  - bootstrap procedure, 237–241
  - definition, 228, 262
  - programs, 228–230

- regression weight and R-square value, 230–232
- Cumulative frequency distribution, 262
- D**
- Degrees of freedom, 262
- Dependent t-Test
  - calculation, 189
  - definition, 262
  - denominator, 188
  - description, 188
  - motion picture, 188
  - psychologist, 188
  - repeated measure sample data, 189
  - STUDENT R program, 190–193
  - students' attitudes, 188
  - t-values, 189
- DEVIATION R program
  - exercises, 139–140
  - output, 138
  - replication, 137, 138
  - sample standard deviation, 138
- Dichotomous population, 262
- Directional hypothesis, 262
- Dispersion
  - description, 77
  - DISPERSION R program (*see* DISPERSION R program)
  - normal distribution, 77
  - range, variance, 77
  - SD, 77–78
  - SS, 77
  - true/false questions, 102
- DISPERSION R program
  - description, 78
  - exercises, 79–81
  - output, 78–79
- E**
- Effect size, 262–263
- Equally likely events, 263
- Exponential function, 263
- F**
- Factorial notation, 263
- Factoring, 263
- F-curve, 263
- F-distribution
  - chi-square distribution and t-distribution, 127–128
  - definition, 263
  - degree of freedom pairs, F-curve, 127
  - exercises, 131–132
  - F-curve and F-ratio, 126
  - F-DISTRIBUTION R programs (*see* F-DISTRIBUTION R programs)
  - independent variances, 128–129
- F-DISTRIBUTION R programs
  - F-curve program output, 130
  - F-ratio program output, 130, 131
- Finite and infinite probability
  - description, 11
  - exercises, 14–17
  - output, 13–14
  - PROBABILITY R program, 12–13
  - relative frequency definition, 12
  - relative frequency, event, 12
  - stabilization, 12
  - trials, 11
  - true/false questions, 39–40
- Finite distribution, 263
- Fisher vs. Gordon chi-square approaches, 246
- Frequency distributions
  - definition, 263
  - histograms and ogives (*see* Histograms and ogives)
  - population (*see* Population distributions)
  - stem and leaf graph (*see* Stem and leaf graph)
- FREQUENCY R program
  - description, 56
  - histogram and ogive exercises, 58–62
  - output, 57–58
  - PlotHeight, 57
  - RelFreq and CumRelFreq vector, 56
  - TableData matrix, 56–57
- F-test
  - ANOVA technique (*see* Analysis of variance (ANOVA) technique)
  - definition, 263
  - multiple comparison tests, 198–199
  - ONEWAY program, 204
  - repeated measures analysis, variance, 199–200, 205
  - REPEATED program, 204
  - SCHEFFE program, 204
  - variance R programs (*see* Variance R programs)
- G**
- Graduate record exam (GRE) score, 223
- GRE score. *See* Graduate record exam (GRE) score

**H**

- Hartley F-max test, 263
- Histograms and ogives
  - cumulative frequency distribution, 55
  - description, 55
  - exercises, 58–62
  - FREQUENCY R program (*see* FREQUENCY R program)
  - output, FREQUENCY program, 57–58
  - S-shaped, 55–56
  - true/false questions, 71
- Hypothesis testing
  - confidence intervals
    - CONFIDENCE R program, 142–144
    - description, 140
    - population mean, 141
    - population standard deviation, 141
    - range of values, 141
    - standard error of a statistic, 140
  - DEVIATION R program (*see* DEVIATION R program)
  - mean and median, 136
  - population standard deviation, 136–137
  - properties, estimators, 136
  - public library, 135
  - sample standard deviation, 137
  - sample statistics, 135
  - sampling distributions, 135
  - standard error of the mean, 136
  - statement, 135
  - statistical (*see* Statistical hypothesis)
  - TYPE I Error (*see* TYPE I Error)
  - TYPE II Error (*see* TYPE II Error)
- HYPOTHESIS TEST R program
  - exercises, 150–152
  - numSamples, 149
  - output, 149
  - pnorm function, 148
  - p Value function, 148
  - tails, 149
  - z-tests/t-tests, 148

**I**

- Independent t-Test, 187–188, 263
- Infinite population, 263
- Intercept, 263
- Interquartile range, 263

**J**

- Jackknife procedure
  - calculation, standard error of mean, 235–237
  - definition, 232, 263

- descriptive information, 234
- detection, influential data value, 233
- R program, 234–235
- Joint probability
  - definition, 263
  - definition, 18
  - exercises, 21–22
  - frequency, 18
  - JOINT R program (*see* JOINT R program)
  - multiplication and addition laws, 18
  - true/false questions, 40
  - unbiased coin, 18
- JOINT R program
  - JOINT PROBABILITY exercises, 21–22
  - for loop calculation, 19–20
  - numEvents vector, 20
  - R Program, 19
  - table function, 19
  - values, 20–21
  - vectors, 19

**K**

- Karl Pearson made. *See* Pearson correlation
- Kurtosis, 264

**L**

- Law of complements, 264
- Leaves, 264
- Level of significance, 264
- Linear function, 264
- Linear regression
  - definition, 217–218, 264
  - equations, 218, 225
  - line and prediction errors
    - computation, correlation coefficient, 219–220
    - intercept and slope values, 220
    - standard deviation, 221–222
  - standard scores
    - applied research, 223
    - R program, 223–224
    - z-score formula, 222
- Line of best fit, 264

**M**

- Mean, 264
- MEAN-MEDIAN R program
  - description, 74
  - exercises, 75–77
  - output, 74–75
- Mean square, 264

- Median, 264
- Meta-analysis technique  
 calculation, unbiased estimate, 250–251  
 comparison and interpretation, effect size  
 measure, 248–250  
 defined, 245–246  
 Fisher vs. Gordon chi-square  
 approaches, 246  
 PRACTICAL R program, 257–258  
 practical significance vs. statistical tests  
 confidence interval, 256  
 effect size measures, 256–257  
 one and two-tailed tests, 255  
 TAAS, 255  
 t-values, 255–256, 258–259  
 R program, 251–252  
 statistics, common metric  
 conversion, 247  
 statistics, effect size measures  
 conversion, 247–248
- Mode, 264
- Monte Carlo, 264
- MS between groups, 264
- MS within groups, 264
- Multiplication law of probability  
 definition, 264  
 description, 26  
 events occurrence, dice, 26  
 exercises, 28–29  
 MULTIPLICATON R program (*see*  
 MULTIPLICATON R program)  
 program output, 27–28  
 relative frequency, 26  
 true/false questions, 41
- MULTIPLICATON R program  
 description, 27  
 exercises, 28–29  
 output, 27–28  
 print.char.matrix function, 27  
 SampleSizes, 27
- N**
- Non-directional hypothesis, 264
- Normal curve, 265
- Normal distribution  
 BELL-SHAPED CURVE R program (*see*  
 BELL-SHAPED CURVE R  
 program)  
 BELL-SHAPED CURVE, sample size, 91  
 definition, 265  
*DensityHeight*, 112  
 exercises, 113–114  
 normal bell-shaped curve, 89  
 NORMAL program output, 112  
 normal, symmetrical and unimodal, 90  
 population parameters, 90  
 probabilities, 112  
 program output, 91  
 sample size,  $n = 20$ , 93  
 sample size of 1,000, 92–93  
 true/false questions, 102  
 uniform, exponential and bimodal  
 populations, 93
- Null hypothesis, 265
- O**
- Ogive, 265
- One-sample t-test, 265
- One-way ANOVA  
 calculation, 197  
 customer standard deviation, 198  
 definition, 265  
 F-value, 198  
 number of customers, store mean, 197  
 sales, different clothing stores, 196  
 squares, stores ( $SS_B$ ) indication, 197  
 statistics, store, 197
- ONEWAY program, 201–202, 204
- Outlier, 265
- P**
- Parameter(s), 265
- Parametric statistics, 265
- Pearson correlation  
 coefficient formula, 208  
 computational version, 208  
 covariance, 207  
 interpretation  
 bivariate normal distribution, 211  
 CORRELATION R Program (*see*  
 CORRELATION R Program)  
 exam scores, 210  
 Karl Pearson's correlation  
 coefficient, 211  
 linear relationship, 211  
 nominal and ordinal data, 211  
 rho approaches, 212  
 scatter plot, 210  
 standard hypothesis testing  
 approach, 209  
 Karl Pearson made, 207  
 Royal Military College, 207  
 time spent and exam scores, 208  
 values, 209  
 z-score, 207

- Permutations, 265
- Pie chart, 265
- Population distributions
  - COMBINATION R program (*see* COMBINATION R program)
  - description, 62
  - exercises, 65–66
  - mean and mode score, 62
  - output, COMBINATION program, 64–65
  - true/false questions, 71–72
  - unimodal and bimodal, 62
- Power, 265
- Probability
  - addition law of probability, 23–26
  - combinations and permutations (*see* Combinations and permutations)
  - conditional probability (*see* Conditional probability)
  - definition, 265
  - finite and infinite (*see* Finite and infinite probability)
  - joint probability (*see* Joint probability)
  - multiplication law of probability (*see* Multiplication law of probability)
- Probability levels
  - distribution of chi-square, 273–274
  - distribution of Hartley F, 279
  - distribution of r, 271–272
  - distribution of t, 270–271
  - F-distribution, 275–278
- PROBABILITY R program
  - dimnames function, 13
  - finite and infinite exercises, 14–17
  - output, 13–14
  - paste and print function, 13
  - rbind, 13
  - SampleFreqs, 12–13
  - SampleSizes, 12
- Properties of estimators, 265
- Pseudo random numbers, 265
  
- Q**
- Quartile, 266
  
- R**
- Random assignment, 266
- Random number generator, 266
- Random numbers
  - definition, 266
  - description, 48
  - population distributions, 48
  - properties, 49
  - “pseudo random numbers”, 48
  - RANDOM R program (*see* RANDOM R program)
- RANDOM R program
  - cat and print function, 50
  - description, 49
  - exercises, 51–52
  - factor and table combination, 49–50
  - GroupedData matrix, 49
  - output, 50–51
- Random sampling, 266
- Range, 266
- Region of rejection, 266
- Regression R program, 223–224
- Regression weight, 266
- Repeated measures ANOVA, 266
- REPEATED program, 202, 203, 204
- Replication, research
  - Bootstrap procedure (*see* Bootstrap procedure)
  - cross-validation, 228–232
  - defined, 227
  - Jackknife procedure (*see* Jackknife procedure)
- Resampling technique, 237
- Research synthesis. *See* Meta-analysis technique
- R Fundamentals
  - accessing data and script programs, 8–9
  - exercises, 10
  - help.start(), 4
  - installation
    - CRAN, 1
    - description, 1
    - download and, 1
    - R-2.15.1, Windows, 2
  - load packages
    - description, 5
    - library(help=“stats”), 6
    - library(stats), 6
    - “stats”, 6
  - RGui window, 4
  - running programs, 7–8
  - simulation operations, 9
  - studio interface installation, 3
  - true/false questions, 10
  - warning, 10
- S**
- Sample, 266
- Sample error, 266
- SAMPLE R program
  - description, 81

- exercises, 82–84
  - for loop, 82
  - output, 82
  - range and standard deviation, 81
  - SampleSizes vector, 81
- Sample size effects
  - description, 81
  - SAMPLE R program (*see* SAMPLE R program)
  - standard deviation, 81
  - true/false questions, 102
- Sample vs. population
  - description, 43
  - exercises, 46–48
  - Gallop Poll, 44
  - output, STATISTICS program, 45–46
  - population characteristics/parameters, 43
  - sample error, 43–44
  - STATISTICS R program (*see* STATISTICS R program)
- Sampling distribution, 266
- Sampling error, 266
- Sampling without replacement, 266
- Sampling with replacement, 266
- Scheffe post-hoc test, 266
- SCHEFFE program
  - F-value, 202
  - output, 203
  - paired group means, 204
  - post-hoc test, 198, 199
- Skewness, 267
- Slope, 267
- Standard deviation, 267
- Standard errors of statistic, 267
- Standard score, 267
- Statistical distributions
  - binomial distribution
    - BINOMIAL R program (*see* BINOMIAL R program)
  - chi-square distribution
    - definition, 105
    - dichotomous populations, 105
    - individual “coefficient”, 106
    - probability combinations, 106
    - sample probabilities to theoretical probabilities, 106
    - z-scores, 107
  - chi-square distribution (*see* Chi-square distribution)
  - F-distribution (*see* F-distribution)
  - normal distribution, 112–114
  - t-distribution, 120–121
- Statistical hypothesis
  - data, 145
  - HYPOTHESIS TEST R program (*see* HYPOTHESIS TEST R program)
  - level of significance, 147–148
  - null and alternative, 146
  - numerous academic disciplines, 144
  - one-sample z-test, 147
  - population standard deviation, 147
  - possible outcomes, 145
  - probability area, 145–146
  - psychologist, 147
  - rejection, 148
  - sociologist, 147
  - TYPE I and TYPE II errors, 145
  - types, 145
  - U.S. SAT mean, 147, 148
  - z-value, 146
- Statistical theory
  - random numbers generation (*see* Random numbers)
  - sample vs. population (*see* Sample vs. population)
  - true and false questions
    - random numbers generation, 53
    - sample vs. population, 53
- Statistics, 267
- STATISTICS R program
  - description, 44
  - exercises, 46–48
  - output, 45–46
  - pseudo-random number generator, 44
  - rnorm, rbinom, and rexp, 45
  - SampleSizes and NumSamples, 44–45
- Stem and leaf graph
  - adult heights, 67
  - definition, 267
  - description, 66
  - exercises, 701–71
  - inter-quartile range, 66
  - output, STEM-LEAF program, 68–69
  - STEM-LEAF R program (*see* STEM-LEAF R program)
  - subdivision, 67
  - true/false questions, 72
  - type of histogram, 67
  - UPS parcel post packages, 66
- STEM-LEAF R program
  - cat command, 68
  - description, 67
  - exercises, 70–71
  - Grades1 and Grades, 68
  - output, 68–69
- STUDENT R program, 190–193
- Sum of squared deviations, 267
- Symmetric distribution, 267

**T**

- TAAS test. *See* Texas Assessment of Academic Skills (TAAS) test
- Tchebysheff inequality theorem  
 data distribution shapes, 85  
 definition, 267  
 description, 84  
 equations and functions, 86  
 exercises, 88–89  
 lower and upper limit, 85  
 $\mu - \sigma$  to  $\mu + \sigma$ , 85  
 range and normal distribution, 84  
 sample mean and standard deviations, 86  
 shape, data distribution, 85  
 standard deviation, 84  
 TCHEBYSHEFF program output, 87  
 TCHEBYSHEFF R program (*see* TCHEBYSHEFF R program)  
 “test norms”, 86  
 true/false questions, 102
- TCHEBYSHEFF R program  
 description, 86–87  
 exercises, 88–89  
 output, 87
- t-distribution  
 definition, 267  
 disciplines, 121  
 population standard deviation, 121  
 probability and inference, statistics, 120  
 probability areas, 120  
 students, 120  
 t-value, 121  
 z-score, 121
- t-DISTRIBUTION R program  
 description, 122  
 exercises, 123–126  
 output, 122–123
- Texas Assessment of Academic Skills (TAAS) test, 255, 256

## t-Test

- dependent (*see* Dependent t-Test)  
 independent, 187–188  
 mean statistical value, 185  
 one sample, 185–186  
 population parameter, 185  
 sample sizes, 185

## TYPE I Error

- areas, 153  
 confidence interval, 155  
 confidence level and region, rejection, 154  
 definition, 267  
 level of significance, 154  
 Null Hypothesis, 154  
 probability, 153

- research process, 153  
 sample mean, 153  
 sample size and sample statistic, 154  
 science, 152–153  
 scientific principles, 153  
 statement, 154  
 statistical values, 154  
 “tabled statistic”, 155  
 traditional math program, 153  
 TYPE I ERROR R program  
 computation, 155  
 exercises, 156–157  
 output, 156

## TYPE II Error

- decision outcomes, 160  
 definition, 267  
 directional and non-directional, 158  
 distributions, 160  
 effect size, 161  
 mathematics achievement, 160  
 Neyman-Pearson hypothesis-testing theory, 161  
 null and alternative hypothesis, 158  
 planning, 159  
 power curves, 160  
 probability, 161  
 product design, 159  
 and TYPE I error, 159  
 TYPE II ERROR R program  
 description, 161  
 exercises, 162–164  
 output, 162  
 p-value function, 161  
 qnorm, 162

**U**

- Uniform distribution, 267  
 Unimodal distribution, 267

**V**

- Variance, 267  
 Variance R programs  
 ONEWAY program output, 201–202  
 pf function, 202  
 REPEATED program output, 202, 203  
 SCHEFFE program output, 202, 203

**Z**

- Z-score, 267–268  
 z-Test  
 definition, 268

- dependent groups
  - confidence
    - interval, 180–181
  - interpretation, 181
  - null hypothesis, 179
  - research design, 178
  - sample data and computation,
    - collection, 179–180
  - statistical hypothesis, 178–179
- independent groups
  - confidence interval, 177
  - interpretation, 178
  - null hypothesis, 176
  - sample data and computation,
    - collection, 176–177
    - statistical hypothesis, 175–176
  - statistical packages, 175
  - testing differences, population
    - proportions, 175
- ZTEST R programs, 182–183