Chapter 22
Variability Analysis for Three or More Samples

1 General Purpose

In some clinical studies, the spread of the data may be more relevant than the average of the data. E.g., when we assess how a drug reaches various organs, variability of drug concentrations is important, as in some cases too little and in other cases dangerously high levels get through. Also, variabilities in drug response may be important. For example, the spread of glucose levels of a slow-release-insulin is important. In Chap. 21, the chi-square test for one sample and the F-test for two samples have been explained. In this chapter we will explain the Bartlett's test which is suitable for comparing three or more samples.

2 Schematic Overview of Type of Data File

Outcome	Treatment group (1, 2, 3, ...)
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

3 Primary Scientific Question

Is the difference in variabilities between three or more samples statistically significant.

4 Data Example (Bartlett's Test)

The Bartlett's test, appropriate for comparing multiple samples for differences in variabilities, uses the underneath equation (χ^2 = chi-square value).

$$\chi^2 = (n_1 + n_2 + n_3 - 3)\ln s^2 - \left[(n_1 - 1)\ln s_1^2 + (n_2 - 1)\ln s_2^2 + (n_3 - 1)\ln s_3^2\right]$$

where

n_1 = size sample 1
$s_1{}^2$ = variance sample 1
s^2 = pooled variance = $\frac{(n_1-1)s_1^2+(n_2-1)s_2^2+(n_3-1)s_3^2}{n_1+n_2+n_3-3}$ =
$\ln$ = natural logarithm

As an example, blood glucose variabilities are assessed in a parallel-group study of three insulin treatment regimens. For that purpose three different groups of patients are treated with different insulin regimens. Variabilities of blood glucose levels are estimated by group-variances ($\ln$ = natural logarithm):

	Group size (n)	Variance [(mmol/l)^2]
Group 1	100	8.0
Group 2	100	14.0
Group 3	100	18.0

$$\text{Pooled variance} = \frac{99 \times 8.0 + 99 \times 14.0 + 99 \times 18.0}{297} = 13.333$$

$$\begin{aligned}
\chi^2 &= 297 \times \ln 13.333 - 99 \times \ln 8.0 - 99 \times \ln 14.0 - 99 \times \ln 18.0 = \\
&\quad 297 \times 2.58776 - 99 \times 2.079 - 99 \times 2.639 - 99 \times 2.890 = \\
&\quad 768.58 - 753.19 = \\
&\quad 15.37
\end{aligned}$$

We have three separate groups, and, so, 3–1 = 2 degrees of freedom. The underneath chi-square table has an upper row with areas under the curve, a left-end column with degrees of freedom, and a whole lot of chi-square values. It shows that with a chi-square value of 15.37 a very significant difference between the three variances is demonstrated at $p < 0.001$. If the three groups are representative

comparable samples, then we may conclude, that these three insulin regimens do not produce the same spread of glucose levels.

Chi-squared distribution

Two-tailed P-value

df	0.10	0.05	0.01	0.001
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.851	11.345	16.266
4	7.779	9.488	13.277	18.466
5	9.236	11.070	15.086	20.515
6	10.645	12.592	16.812	22.457
7	12.017	14.067	18.475	24.321
8	13.362	15.507	20.090	26.124
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	27.688	34.527
14	21.064	23.685	29.141	36.124
15	22.307	24.996	30.578	37.698
16	23.542	26.296	32.000	39.252
17	24.769	27.587	33.409	40.791
18	25.989	28.869	34.805	42.312
19	27.204	30.144	36.191	43.819
20	28.412	31.410	37.566	45.314
21	29.615	32.671	38.932	46.796
22	30.813	33.924	40.289	48.268
23	32.007	35.172	41.638	49.728
24	33.196	36.415	42.980	51.179
25	34.382	37.652	44.314	52.619
26	35.536	38.885	45.642	54.051
27	36.741	40.113	46.963	55.475
28	37.916	41.337	48.278	56.892
29	39.087	42.557	49.588	58.301
30	40.256	43.773	50.892	59.702
40	51.805	55.758	63.691	73.403
50	63.167	67.505	76.154	86.660
60	74.397	79.082	88.379	99.608
70	85.527	90.531	100.43	112.32
80	96.578	101.88	112.33	124.84
90	107.57	113.15	124.12	137.21
100	118.50	124.34	135.81	149.45

5   Conclusion

An alternative to the Bartlett's test is the Levene's test. The Levene's test is less sensitive than the Bartlett's test to departures from normality. If there is a strong evidence that the data do in fact come from a normal, or nearly normal, distribution, then Bartlett's test has a better performance. Levene's test requires a lot of arithmetic, and is usually performed using statistical software. E.g., it is routinely used by SPSS when performing an unpaired t-test or one-way ANOVA (analysis of variance) (see also Cleophas, Zwinderman, SPSS for Starters, part 1, Springer New York, 2010, Chaps. 4 and 8).

We should add that assessing significance of differences between 3 or more variances does not answer which of the samples produced the best outcome. Just like with analysis of variance (Chap. 19), separate post hoc one by one analyses are required.

6   Note

More background, theoretical and mathematical information of variability assessments is given in Statistics applied to clinical studies 5th edition, Chap. 44, Springer Heidelberg Germany, 2012, from the same authors.