

## Example Q      Strength of cotton yarn

### Description of data

An experiment was done with the objects of estimating:

- (i)      The difference in mean strength of two worsted yarns produced by slightly different processes, and
- (ii)     The variation of strength between and within bobbins for yarns of this type.

For each yarn a considerable number of bobbins were produced and 6 bobbins selected at random. From each of these, 4 short lengths were chosen at random for strength testing. The breaking loads are given in table Q.1.

Table Q.1		Breaking loads (oz)					
Bobbin	1	2	3	4	5	6	
Yarn A	15.0	15.7	14.8	14.9	13.0	15.9	
	17.0	15.6	15.8	14.2	16.2	15.6	
	13.8	17.6	18.2	15.0	16.4	15.0	
	15.5	17.1	16.0	12.8	14.8	15.5	
Yarn B	18.2	17.2	15.2	15.6	19.2	16.2	
	16.8	18.5	15.9	16.0	18.0	15.9	
	18.1	15.0	14.5	15.2	17.0	14.9	
	17.0	16.2	14.2	14.9	16.9	15.5	

### General considerations

This problem illustrates in fairly simple form the analysis of data in which the pattern of random variation has some nontrivial structure. Because of the nature of the response variable, it is reasonable to concentrate on means and variances, this being totally appropriate if the variation is normal. It would be possible to use broadly similar, although more complicated, techniques based on detailed models involving particular non-normal distributions, Poisson, binomial, exponential, etc., were that appropriate.

The numbering of bobbins within each yarn is random, as too is the numbering of test lengths within bobbins. Thus the analysis-of-variance table has the doubly nested form shown in Table Q.4. Note that the form of the table is settled by the way the data were obtained. It would be possible to examine differences between individual bobbins, although this would normally be fruitful only if more information were available to characterize the bobbins as individuals. Because the bobbins are chosen from a large population, it is helpful to regard between-bobbin variation as a form of random variation, described by a component of variance between bobbins, i.e. the variance over the population of the ‘true’ bobbin means.

That is, we characterize the random variation for each yarn by two components of variance, one between lengths within bobbins and the other between bobbins. These are estimated from the analysis-of-variance table, either separately for each yarn or pooled over yarns.

More broadly, there are four possibilities that could have arisen:

- (i)      The bobbins are identifiable individuals;
- (ii)     Only the variance of the ‘true’ means, calculated over the bobbins individually observed, is of interest;

- (iii) The bobbins are a sample from a finite population of known size and only the variance of the ‘true’ means, calculated over the finite population, is of interest;
- (iv) The population of bobbins is effectively infinite and again the variance of the ‘true’ means is required.

The analysis-of-variance table is the same in all cases; it is the interpretation of the bobbin means that is different. More than one of (i) – (iv) might be relevant for different purposes, although in fact (iv) is the most useful here.

The components of variance have at least three uses. Firstly, they provide summary descriptions of important aspects of the problem under study and so may be regarded as primary parameters. Next, they clarify the fact that to compare the means of the two yarns, the mean square between bobbins within yarns provides an appropriate estimate of error. Finally, via a process of synthesis of variance, it is possible to estimate the consequences of some different scheme of sampling in which the number of repeat observations per bobbin is changed.

To see that the analysis of variance does not overlook some major aspect of importance, it is desirable to check that:

- (a) No single observation or small number of ‘wild’ observations has inflated the mean square within bobbins;
- (b) No single bobbin has inflated the mean square between bobbins;
- (c) There is no relation between the bobbin means and standard deviations that might suggest a transformation;
- (d) The between-bobbin variation is approximately normal.

It is possible also to examine for each yarn the variation from bobbin to bobbin in the mean square within bobbins. Because bobbins are randomly numbered, excessive variation in these mean squares would imply either some non-independence in sampling or a long-tailed distribution within bobbins.

In principle it would be possible to make similar analyses of higher cumulants.

### The analysis

Mean breaking loads for each bobbin and yarn are summarized in Table Q.2. The overall mean for yarn B is 0.86 oz higher than for yarn A.

Table Q.2 Mean breaking loads (oz)

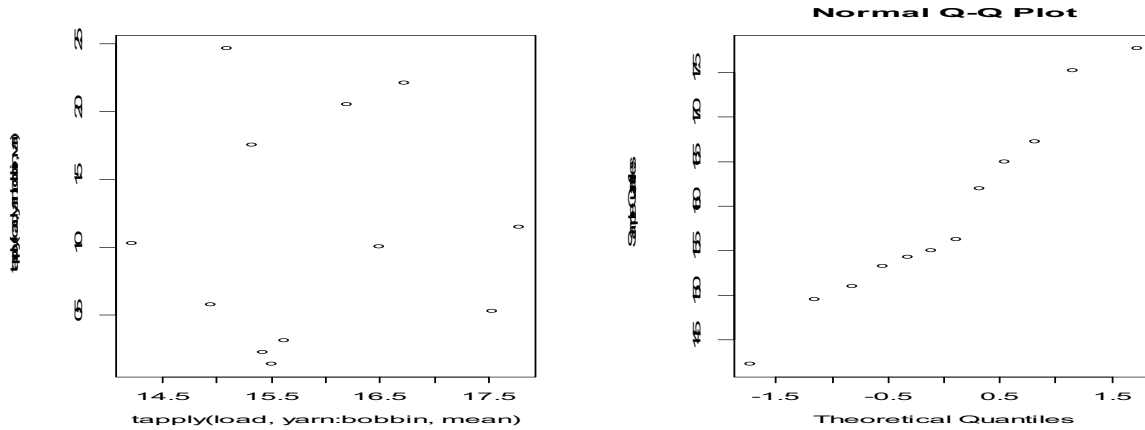
Bobbin	1	2	3	4	5	6	mean
Yarn A	15.325	16.500	16.20	14.225	15.100	15.500	15.475
Yarn B	17.525	16.725	14.95	15.425	17.775	15.625	16.338

Table Q.3 gives the mean squares within bobbins.

Table Q.3 Mean squares within bobbins (3 df)

Bobbin	1	2	3	4	5	6
Yarn A	1.756	1.01	2.053	1.029	2.467	0.140
Yarn B	0.529	2.21	0.577	0.229	1.149	0.316

There is no evidence of any systematic relationship between these and the bobbin means (see figure below), nor is there evidence of heterogeneity between the mean squares within either yarn (see the solution to Exercise (i)b). Approximate normality of the variation between bobbins is confirmed by plotting bobbin means against expected normal order statistics (Pearson and Hartly, 1966, Table 28) (see figure below).



The analysis of variance is given in Table Q.4 (slightly rearranged).

Table Q.4 Analysis of Variance Table

	Df	Sum Sq	Mean Sq	E [m. s.]
Between yarns	1	8.927	8.927	
Between bobbins	10	40.779	4.078	$\sigma_0^2 + 4\sigma_1^2$
Within yarn A	5	13.210	2.642	
Within yarn B	5	27.569	5.514	
Within bobbins	36	40.382	1.122	
Within yarn A	18	25.355	1.409	
Within yarn B	18	15.028	0.835	
Total	47	90.088		

There is no significant difference in variation according to yarn, either within bobbins ( $F=1.409/0.835=1.69$  with  $(18, 18)$  df) or between bobbins ( $F=5.514/2.642 = 2.09$  with  $(5, 5)$  df). It is convenient to pool mean squares within yarns.

Now the standard error of the observed difference of 0.86 oz between yarns is equal to  $\sqrt{(\sigma_0^2 + 4\sigma_1^2)/12}$  (see Exercise (iii) for derivation details), and can be estimated directly from the mean square between bobbin as  $\sqrt{4.078/12} = 0.58$  oz with 10 degrees of freedom. Thus no evidence of any systematic difference in strength between yarns A and B can be established. Confidence limits at any desired level are readily found.

Estimates of the components of variation  $\sigma_0^2$ , within bobbins, and  $\sigma_1^2$ , between bobbins, are required if, for example, recommendations are to be made for further experimentation. Suppose  $m$  bobbins are selected for each yarn, and  $l$  lengths per bobbin tested and that it takes  $k$ , a known value, times as long to sample a bobbin as it does to perform a single test. The total time taken is proportional to

$$T = k(m + l)$$

And the variance of the observed difference in mean strength will be proportional to (see Exercise (iv) for derivation details)

$$V = \frac{\sigma_0^2}{ml} + \frac{\sigma_1^2}{m}$$

The recommended optimum choice of  $l$  is obtained by minimizing  $V$  subject to given  $T$  ( or minimizing  $T$  subject to given  $V$ ), i.e. (see Exercise (iv) for derivation details)

$$l_{opt} = \sqrt{k} \frac{\sigma_0}{\sigma_1}$$

Equating observed mean squares in Table Q.4 to their expected values gives estimates for the components of variance,

$$\tilde{\sigma}_0^2 = 1.122 \quad \text{and} \quad \tilde{\sigma}_1^2 = (4.078 - 1.122)/4 = 0.739$$

Thus if, say,  $k=4$  the estimated value of  $l_{opt}$  is  $\sqrt{4(1.122/0.739)} = 2.463$ ; we would take  $l = 3$ .

### Further points and exercises

(i) (a) Show that for the normal-theory estimates of variance,  $s^2$  say, based on  $n$  observations,

$$VAR[\log(s^2)] \approx \frac{1}{n-1}$$

Solution:

Assume that  $X_1, \dots, X_n$  are independent, identically distributed random variables from a  $N(\mu; \sigma^2)$  distribution. Let  $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$ .

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

$$VAR[\chi^2(n-1)] = 2(n-1)$$

$$\Rightarrow VAR \left[ \frac{(n-1)s^2}{\sigma^2} \right] = \frac{(n-1)^2}{\sigma^4} VAR[s^2] = 2(n-1)$$

$$\Rightarrow VAR[s^2] = \frac{2\sigma^4}{n-1}$$

Applying  $\Delta$  - method,

$$VAR[\log(s^2)] \approx \frac{1}{(E[s^2])^2} VAR[s^2] = \frac{1}{(\sigma^2)^2} \frac{2\sigma^4}{n-1} = \frac{2}{n-1}$$

(b) Use the result from (a) to check the homogeneity within yarns of the mean squares given in Table Q.3

Solution:

Homogeneity of (group) variance (also called *homoscedasticity*) is an assumption of ANOVA that refers to the situation that variances of the observations in individual groups are equal. Tests of homogeneity of variances concern the question of whether several populations have the same variance.

The following are two methods that can be employed to check the homogeneity within yarns of the mean squares.

- Confidence Interval Method

Check if observed  $\log(s^2)$ 's fall in the 95% confidence interval..

Yarn A	$s^2$	1.756	1.007	2.053	1.029	2.467	0.140
	$\log(s^2)$	0.565	0.010	0.719	0.030	0.904	-1.966

The 95% confidence intervals for the mean squares yarn A is

$$\overline{\log(s^2)} \pm 1.96 \sqrt{\text{VAR}[\log(s^2)]} = 0.0435 \pm 1.96\sqrt{2/(4-1)} = (-1.567, 1.644)$$

All the observed values of  $\log(s^2)$  except the last one fall in the 95% CI, indicating the homogeneity of variance within Yarn A.

Yarn B	$s^2$	0.529	2.209	0.577	0.229	1.149	0.316
	$\log(s^2)$	-0.635	0.793	-0.545	-1.470	0.140	-1.139

The 95% confidence intervals for the mean squares yarn B is

$$\overline{\log(s^2)} \pm 1.96 \sqrt{\text{VAR}[\log(s^2)]} = -0.476 \pm 1.96\sqrt{2/(4-1)} = (-2.076, 1.124)$$

All the values of  $\log(s^2)$  falls in the 95% CI, indicating the homogeneity of variance within Yarn B.

- Bartlett's *M*-test Method

When comparing the variances of *two* random samples from a normal distribution we use the F test but keep in mind that the F test is quite sensitive to departures from normality.

Bartlett's *M*-test or the Neyman–Pearson–Bartlett test of homogeneity of variances is the standard procedure when we have  $k > 2$  populations and if the populations are approximately normally distributed (Armitage and Berry, 1994). Applications arise in the analysis-of variance context, principally in testing the equal “within-group” variances assumption.

The test of the null hypothesis of equality of group variances is based on comparing (the logarithm) of a pooled estimate of variance (across all of the groups) with the sum of the logarithms of the variances of individual groups.

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

$$H_a: \sigma_t^2 \neq \sigma_u^2$$

To test null hypothesis  $H_0$ , Bartlett introduces the statistic *M*:

$$M = \left( \sum_{j=1}^k (n_j - 1) \right) \log (s^2) - \sum_{j=1}^k (n_j - 1) \log(s_j^2)$$

where

$$s_j^2 = \frac{1}{(n_j - 1)} \sum_{i=1}^{n_j} (X_{ji} - \bar{x}_j)^2$$

$$s^2 = \frac{1}{\sum_{j=1}^k (n_j - 1)} \sum_{j=1}^k (n_j - 1) s_j^2$$

Under  $H_0$ , as each  $(n_j - 1)$  gets arbitrarily large, M has a limiting  $\chi^2(k-1)$  distribution, i.e.

$$M \xrightarrow{d} \chi^2(k-1) \quad \text{as } (n_j - 1) \rightarrow \infty, j = 1, \dots, k$$

To achieve closeness to  $\chi^2$  for finite  $(n_j - 1)$ , Bartlett suggests a modified statistic T:

$$T = \frac{M}{C}$$

where

$$C = 1 + \frac{1}{3(k-1)} \left[ \sum_{j=1}^k \frac{1}{(n_j - 1)} - \frac{1}{\sum_{j=1}^k (n_j - 1)} \right]$$

A test with approximate size  $\alpha$  is based on referring T to a  $\chi^2(k-1)$  distribution. So we have the decision rule:

$$\begin{aligned} &\text{Reject } H_0 \text{ if } T > \chi^2(\alpha, k-1) \\ &\text{Accept } H_0 \text{ if } T \leq \chi^2(\alpha, k-1) \end{aligned}$$

Substituting the numbers in Table Q.3 into the formulas, we calculate and get  $T = 4.783$  for yarn A, and  $T = 4.775$  for yarn B. They both are less than the tabulated value of  $\chi^2(5)$  at significant level, say  $\alpha = 10\%$ , which is 9.24. So the data is consistent with the assumption of homoscedasticity.

But Bartlett's test is not reliable with moderate departures from normality and Levene's test is routinely used as an alternative.

**(ii)** Show that the expected value of the mean square between bobbins is  $\sigma_0^2 + 4\sigma_1^2$

Solution:

Example Q regards the two-stage nested design that involves two factors A and B having  $a$  and  $b$  levels, respectively, with  $b$  levels of factor B being nested under each level of factor A. The linear model for the design is

$$\begin{aligned} y_{ijr} &= \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{r(ij)} \\ i &= 1, \dots, a; j = 1, \dots, b; r = 1, \dots, n \end{aligned}$$

where

$y_{ijr}$ : the response variable corresponding to  $i$ th level of factor A,  $j$ th level of factor B nested within the  $i$ th level of factor A and the  $r$ th replicate within the  $j$ th level of B

$\mu$ : the overall mean

$\alpha_i$ : the effect due to the  $i$ th level of factor A

$\beta_{j(i)}$ : the effect due to the  $j$ th level of factor B nested within the  $i$ th level of factor A

$\varepsilon_{r(ij)}$ : the error term that takes into account the random variation within a particular cell

$a$ : the number of levels of factor A  
 $b$ : the number of levels of factor B nested within the  $i$ th level of factor A  
 $n$ : the number of replicates in each level of B

For the error term  $\varepsilon_{r(ij)}$  we assume that  $\varepsilon_{r(ij)} \sim NID(0, \sigma^2)$

The assumptions concerning  $\alpha_i, \beta_{j(i)}$  depend on whether the levels of A and B are fixed or random. Here we consider only the mixed model with A fixed and B random, we may assume

$$\sum_{i=1}^a \alpha_i = 0 \quad \text{and} \quad \beta_{j(i)} \sim NID(0, \sigma_B^2)$$

The total corrected sum of squares is

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{r=1}^n (y_{ijr} - \bar{y}_{...})^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{r=1}^n (\bar{y}_{i..} - \bar{y}_{...} + \bar{y}_{ij.} - \bar{y}_{i..} + y_{ijr} - \bar{y}_{ij.})^2 \\ &= bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 + n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{r=1}^n (y_{ijr} - \bar{y}_{ij.})^2 \end{aligned}$$

since the three cross-product terms is zero. Symbolically, the equation can be written as

$$SS_T = SS_A + SS_{B(A)} + SS_E$$

where

$SS_T$ : total sum of squares with  $df_T = abn - 1$   
 $SS_A$ : sum of squares due to factor A with  $df_A = a - 1$   
 $SS_{B(A)}$ : sum of squares due to factor B with  $df_{B(A)} = a(b - 1)$   
 $SS_E$ : sum of squares due to error with  $df_E = ab(n - 1)$

The expected value of  $MS_{B(A)}$  is

$$E[MS_{B(A)}] = E\left[\frac{SS_{B(A)}}{df_{B(A)}}\right] = \frac{n}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b E[(\bar{y}_{ij.} - \bar{y}_{i..})^2]$$

$$\bar{y}_{ij.} - \bar{y}_{i..} = (\mu + \alpha_i + \beta_{j(i)} + \bar{\varepsilon}_{(ij)}) - (\mu + \alpha_i + \bar{\beta}_{(i)} + \bar{\varepsilon}_{(i)}) = \beta_{j(i)} - \bar{\beta}_{(i)} + \bar{\varepsilon}_{(ij)} - \bar{\varepsilon}_{(i)}$$

$$\begin{aligned} E[(\bar{y}_{ij.} - \bar{y}_{i..})^2] &= VAR[\bar{y}_{ij.} - \bar{y}_{i..}] = VAR[\beta_{j(i)} - \bar{\beta}_{(i)} + \bar{\varepsilon}_{(ij)} - \bar{\varepsilon}_{(i)}] \\ &= VAR[\beta_{j(i)}] + VAR[\bar{\beta}_{(i)}] - 2COV(\beta_{j(i)}, \bar{\beta}_{(i)}) + VAR[\bar{\varepsilon}_{(ij)}] + VAR[\bar{\varepsilon}_{(i)}] - 2COV(\bar{\varepsilon}_{(ij)}, \bar{\varepsilon}_{(i)}) \\ &= \sigma_B^2 + \frac{1}{b} \sigma_B^2 - \frac{2}{b} \sigma_B^2 + \frac{1}{n} \sigma^2 + \frac{1}{bn} \sigma^2 - \frac{2}{bn} \sigma^2 = \frac{b-1}{bn} (n\sigma_B^2 + \sigma^2) \end{aligned}$$

$$\Rightarrow E[MS_{B(A)}] = \frac{n}{a(b-1)} \sum_{i=1}^a \sum_{j=1}^b \frac{b-1}{bn} (n\sigma_B^2 + \sigma^2) = \sigma^2 + n\sigma_B^2$$

In the same way we can get

$$E[MS_{B(A)}] = \sigma^2 + n\sigma_B^2 + \frac{bn}{a-1} \sum_{i=1}^a \alpha_i^2$$

In Example Q, factor A is 'yarn', which is fixed, and factor B is 'bobbin', which is random. The number of levels of A is  $a = 2$ ; the number of levels of factor B is  $b = 6$ ; the number of replicates within each level of B is  $n = 4$ . The mean square between bobbins is  $MS_{B(A)}$ ; the variation within bobbins is denoted as  $\sigma_0^2$ ; the variation between bobbins is denoted as  $\sigma_1^2$ . So we have

$$E[MS_{B(A)}] = \sigma_0^2 + 4\sigma_1^2$$

(iii) Find the variance of  $\bar{y}_{1..} - \bar{y}_{2..}$ .

Solution:

The difference between the observed means of yarn A and yarn B, i.e.  $\bar{y}_{1..} - \bar{y}_{2..}$  estimates the difference between the effects due to the two level of factor A, i.e.  $\alpha_1 - \alpha_2$ . The variance of this estimator is

$$\begin{aligned} VAR[\bar{y}_{1..} - \bar{y}_{2..}] &= VAR[(\mu + \alpha_1 + \bar{\beta}_{.(1)} + \bar{\varepsilon}_{.(1)}) - (\mu + \alpha_2 + \bar{\beta}_{.(2)} + \bar{\varepsilon}_{.(2)})] \\ &= VAR[\bar{\beta}_{.(1)}] + VAR[\bar{\beta}_{.(2)}] + VAR[\bar{\varepsilon}_{.(1)}] + VAR[\bar{\varepsilon}_{.(2)}] \\ &= \frac{1}{6}\sigma_B^2 + \frac{1}{6}\sigma_B^2 + \frac{1}{4 \times 6}\sigma^2 + \frac{1}{4 \times 6}\sigma^2 \\ &= \frac{1}{12}\sigma^2 + \frac{1}{3}\sigma_B^2 \end{aligned}$$

Or in the other notation,

$$VAR[\bar{y}_{1..} - \bar{y}_{2..}] = \frac{1}{12}(\sigma_0^2 + 4\sigma_1^2)$$

(iv) Show that  $l_{opt} = \sqrt{k} \frac{\sigma_0}{\sigma_1}$

Solution:

Let  $m$  be the number of bobbins to be selected for each yarn,  $l$  the number of samples per bobbin. And suppose it takes  $k$  (known) times as long to sample a bobbin as it does to perform a single test. The total time taken is proportional to  $T = m(k + l)$ . Rearranging it yields

$$m = \frac{T}{k + l}$$

$$\begin{aligned} VAR[\bar{y}_{1..} - \bar{y}_{2..}] &= VAR[(\mu + \alpha_1 + \bar{\beta}_{.(1)} + \bar{\varepsilon}_{.(1)}) - (\mu + \alpha_2 + \bar{\beta}_{.(2)} + \bar{\varepsilon}_{.(2)})] \\ &= VAR[\bar{\beta}_{.(1)}] + VAR[\bar{\beta}_{.(2)}] + VAR[\bar{\varepsilon}_{.(1)}] + VAR[\bar{\varepsilon}_{.(2)}] \\ &= \frac{1}{m}\sigma_1^2 + \frac{1}{m}\sigma_1^2 + \frac{1}{lm}\sigma_0^2 + \frac{1}{lm}\sigma_0^2 \\ &= 2\left(\frac{1}{m}\sigma_1^2 + \frac{1}{lm}\sigma_0^2\right) \end{aligned}$$

The optimum choice of  $l$  is to minimizing

$$V = \frac{1}{m}\sigma_1^2 + \frac{1}{lm}\sigma_0^2 = \frac{k+l}{T}\sigma_1^2 + \frac{k+l}{lT}\sigma_0^2$$

with respect to  $l$ .

$$\begin{aligned} \frac{\partial V}{\partial l} &= \frac{\sigma_1^2}{T} - \frac{k\sigma_0^2}{l_{opt}^2 T} = 0 \\ \Rightarrow \quad l_{opt} &= \sqrt{k} \frac{\sigma_0}{\sigma_1} \end{aligned}$$



Appendix: R Code

```
#####
# R code for Example Q
#####

#-----
# input Table Q.1 and create data frame
#-----

load = c( 150,157,148,149,130,159, 170,156,158,142,162,156,
          138,176,182,150,164,150, 155,171,160,128,148,155,
          182,172,152,156,192,162, 168,185,159,160,180,159,
          181,150,145,152,170,149, 170,162,142,149,169,155) / 10
yarn = factor(rep(c('A', 'B'), each=24))
bobbin = factor(rep(1:6,times=8))
table = data.frame(yarn, bobbin, load)
```

```
> table
  yarn bobbin load
 1     A      1 15.0
 2     A      2 15.7
 3     A      3 14.8
 4     A      4 14.9
 5     A      5 13.0
 6     A      6 15.9
 7     A      1 17.0
 8     A      2 15.6
 9     A      3 15.8
10    A      4 14.2
11    A      5 16.2
12    A      6 15.6
13    A      1 13.8
14    A      2 17.6
15    A      3 18.2
16    A      4 15.0
17    A      5 16.4
18    A      6 15.0
19    A      1 15.5
20    A      2 17.1
21    A      3 16.0
22    A      4 12.8
23    A      5 14.8
24    A      6 15.5
25    B      1 18.2
26    B      2 17.2
27    B      3 15.2
28    B      4 15.6
29    B      5 19.2
30    B      6 16.2
31    B      1 16.8
32    B      2 18.5
33    B      3 15.9
34    B      4 16.0
35    B      5 18.0
36    B      6 15.9
37    B      1 18.1
38    B      2 15.0
39    B      3 14.5
40    B      4 15.2
41    B      5 17.0
42    B      6 14.9
43    B      1 17.0
44    B      2 16.2
45    B      3 14.2
46    B      4 14.9
47    B      5 16.9
48    B      6 15.5
```

```
#-----
# reproduce Table Q.2
#-----

tapply(load, list(yarn,bobbin), mean)
meanA = mean(load[yarn=='A'])
```

```
meanB = mean(load[yarn=='B'])
meanA; meanB; meanA-meanB
```

```
> tapply(load, list(yarn,bobbin), mean)
  |      1      2      3      4      5      6
  | A 15.325 16.500 16.20 14.225 15.100 15.500
  | B 17.525 16.725 14.95 15.425 17.775 15.625
```

```
> meanA; meanB; meanA-meanB
  | [1] 15.475
  | [1] 16.3375
  | [1] -0.8625
```

```
#-----
# reproduce Table Q.3
#-----
```

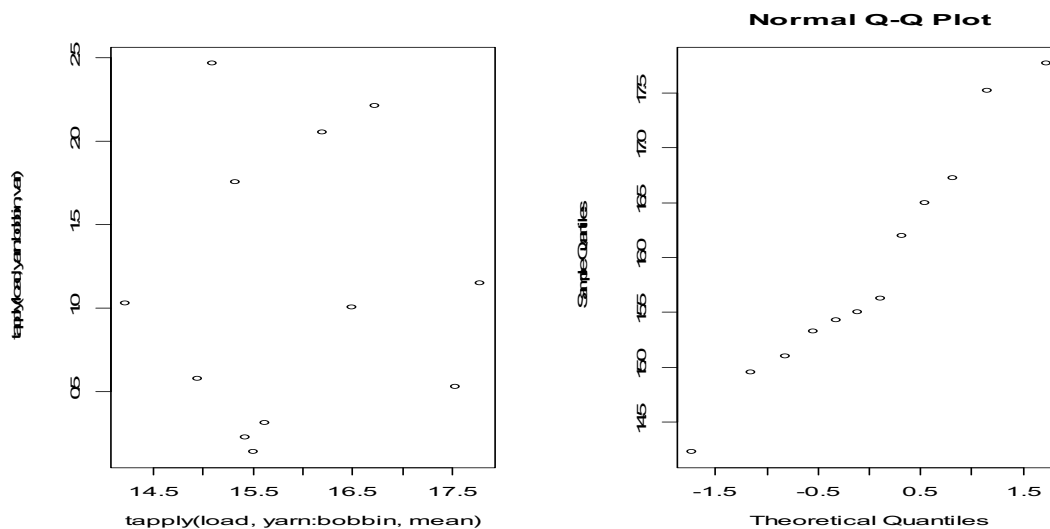
```
tapply(load, list(yarn,bobbin), var)
```

```
> options(digits=4)
> tapply(load, list(yarn,bobbin), var)
  |      1      2      3      4      5      6
  | A 1.7558 1.007 2.0533 1.0292 2.467 0.1400
  | B 0.5292 2.209 0.5767 0.2292 1.149 0.3158
```

```
#-----
# visualize the points concerning Table Q.2 and Q.3
#-----
```

```
# It is claimed that there is no evidence of any systematic
# relationship between mean squares and the bobbin means, and
# approximate normality of the variation between bobbins is confirmed
# by plotting bobbin means against expected normal order statistics
```

```
op = par(mfrow=c(1,2))
plot(tapply(load,yarn:bobbin,mean), tapply(load,yarn:bobbin,var))
qqnorm(tapply(load,yarn:bobbin,mean))
par(op)
```



```

#-----
# reproduce Table Q.4
#-----

# factor 'bobbin' is nested in factor 'yarn'
Q = aov(load ~ yarn + bobbin %in% yarn)
anova(Q)
SST = sum(anova(Q)[2])

  > anova(Q)
  Analysis of Variance Table

  Response: load
    Df Sum Sq Mean Sq F value    Pr(>F)
yarn   1  8.927   8.927   7.9581 0.007741 **
yarn:bobbin 10 40.779   4.078   3.6353 0.002023 **
Residuals  36 40.382   1.122
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  > SST
  [1] 90.08812

## in the output listed above,
## 'yarn' corresponds to 'between yarns' in Table Q.4
## 'yarn:bobbin' corresponds to 'between bobbins within yarns' in Table Q.4
## 'Residuals' corresponds to 'within bobbins' in Table Q.4

## Now partition 'between bobbins within yarns' into two components:
## 'between bobbins:within yarn A' and 'between bobbins:within yarn B'
## And partition 'within bobbins' into two components:
## 'within bobbins: within yarn A' and 'within bobbins: within yarn B'

H1 = aov(load ~ bobbin, subset = yarn == 'A' )
H2 = aov(load ~ bobbin, subset = yarn == 'B' )
summary(H1)
summary(H2)

  > summary(H1)
    Df Sum Sq Mean Sq F value Pr(>F)
bobbin   5 13.2100  2.6420  1.8756 0.1488
Residuals 18 25.3550  1.4086

## in the output listed above,
## 'bobbin' corresponds to 'between bobbins: within yarn A'
## 'residual' corresponds to 'within bobbins: within yarn A'

  > summary(H2)
    Df Sum Sq Mean Sq F value    Pr(>F)
bobbin   5 27.5688  5.5138  6.6044 0.001178 **
Residuals 18 15.0275  0.8349
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## in the output listed above,
## 'bobbin' corresponds to 'between bobbins: within yarn B'
## 'residual' corresponds to 'within bobbins: within yarn B'

```

```

#-----
# exercise (i) (b): Barlett's test
#-----

## Barlett statistic T (yarn A)
k = 6
ssja = c(1.756,1.007,2.053,1.029,2.467,0.140) # mean square for yarn A
logssja = log(ssja)
nj = c(4, 4, 4, 4, 4, 4)
njr = 1/(nj - 1)
nn = sum(nj - 1)
ssa = sum((nj - 1)*ssja) / nn
MA = nn * log(ssa) - sum((nj-1)*logssja)
CA = 1 + (1/3/(k-1)) * (sum(njr) - 1/nn)
TA = MA/CA
cat('Barlett test statistic T for yarn A is', TA,'\n')

## Bartlett statistic T ( yarn B )
k = 6
ssjb = c(0.529,2.209,0.577,0.229,1.149,0.316) # mean square for yarn B
logssjb = log(ssjb)
nj = c(4, 4, 4, 4, 4, 4)
nn = sum(nj - 1)
njr = 1/(nj - 1)
ssb = sum((nj-1)*ssjb) / nn

MB = nn * log(ssb) - sum((nj-1)*logssjb)
CB = 1 + (1/3/(k-1)) * (sum(njr) - 1/nn)
TB = MB/CB
cat('Barlett test statistic T for yarn B is', TB,'\n')

    > cat('Barlett test statistic T for yarn A is', TA,'\n')
| Barlett test statistic T for yarn A is 4.783164

    > cat('Barlett test statistic T for yarn B is', TB,'\n')
| Barlett test statistic T for yarn B is 4.775429

## Barlett's test using R function bartlett.test,
## which we expect to be consistent with the above results

bartlett.test(load ~ bobbin, data=table, subset=yarn=='A')
bartlett.test(load ~ bobbin, data=table, subset=yarn=='B')

    > bartlett.test(load ~ bobbin, data=table, subset=yarn=='A')

|
|      Bartlett test of homogeneity of variances
|
| data: load by bobbin
| Bartlett's K-squared = 4.7832, df = 5, p-value = 0.4429
|
    > bartlett.test(load ~ bobbin, data=table, subset=yarn=='B')

|
|      Bartlett test of homogeneity of variances
|
| data: load by bobbin
| Bartlett's K-squared = 4.7755, df = 5, p-value = 0.4439
|

```