Design and Analysis of Experiments

Homework #1 - Cable Strength Analysis

Morten Middelfart and Morten Goodwin Olsen

October 24, 2008

For the assignment, from which this homework is based upon, see `http://www.math.aau.dk/~slb/kurser/design-08/homework/homework-1.pdf`.

The data used for this analysis can be seen in table 1.

1	2	3	4	5	6	7	8	9
345	329	340	328	347	341	339	339	342
327	327	330	344	341	340	340	340	346
335	332	325	342	345	335	342	347	347
338	348	328	350	340	336	341	345	348
330	337	338	335	350	339	336	350	355
334	328	332	332	346	340	342	348	351
335	328	335	328	345	342	347	341	333
340	330	340	340	342	345	345	342	347
337	345	336	335	340	341	341	337	350
342	334	339	337	339	338	340	346	347
333	328	335	337	330	346	336	340	348
335	330	329	340	338	347	342	345	341

Table 1: Cable strength.

Initially, we need to load the data into R. Note that lots, which is used later in the assignment, is already loaded here.

R-commands:

$$\begin{aligned} y &= as.numeric(matrix(scan('data-1.txt'), nrow = 12, ncol = 9, byrow = TRUE)) \\ cables &= as.factor(gl(9, 12)) \\ lots &= as.factor(c(rep("A", 48), rep("B", 60))) \end{aligned}$$

a **Graphical display**

To display the data graphically, we show it both as a boxplot and dotplot as shown in figure 1.

R-commands:

$$
\begin{aligned}
par & \quad (mfcol = c(1,2)); \\
plot & \quad (as.numeric(cables), y, pch = 16, xlab = "Cables", ylab = "Strengths"); \\
boxplot & \quad (y \sim cables, xlab = "Cables", ylab = "Strengths")
\end{aligned}
$$

The graphical display as boxplot of the data is in figure 1.

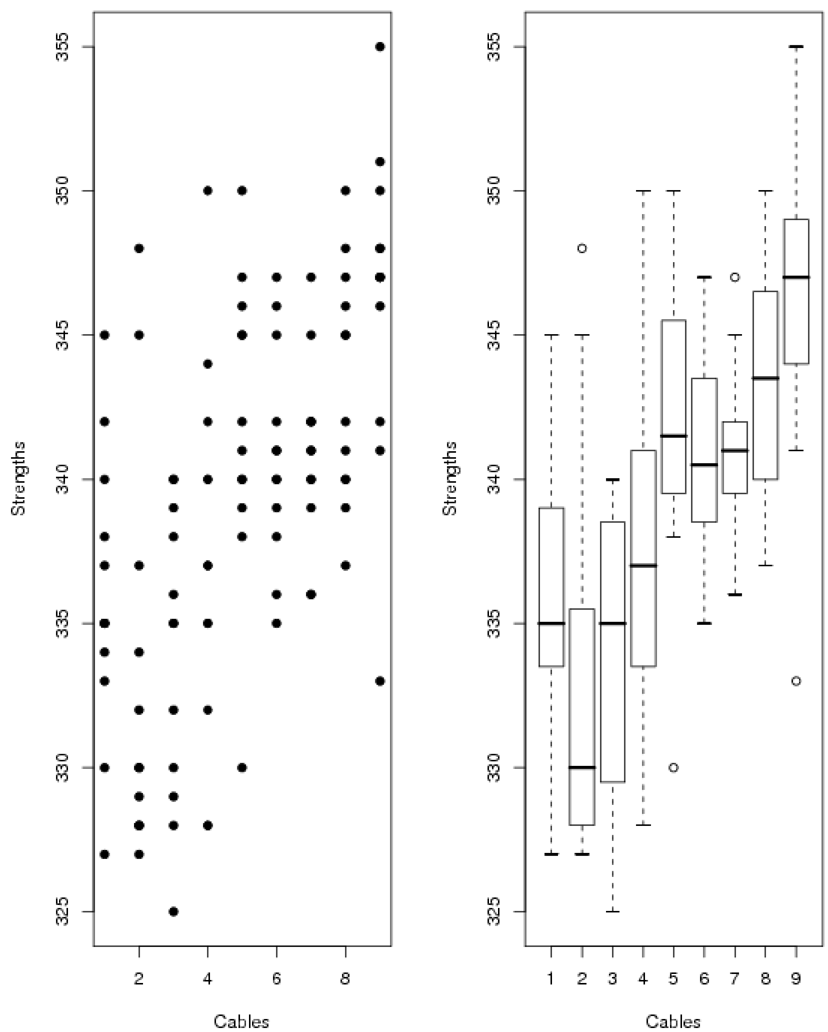

Figure 1: Boxplot of Cable and Strength

b **Mean and standard deviation**

For mean and standard deviation we use the built in *mean* and *sd* in R.

R-commands:

$$tapply(y, cables, mean) \qquad \#means$$
$$tapply(y, cables, sd) \qquad \#StandardDeviations$$

The mean and standard deviation for all cables are presented in table 2.

Cable Number	**Mean**	**Standard Deviation**
1	335.9167	4.962740
2	333.0000	6.954397
3	333.9167	5.053502
4	337.3333	6.429101
5	341.9167	5.230302
6	340.8333	3.737606
7	340.9167	3.175426
8	343.3333	4.052683
9	346.2500	5.577960

Table 2: Mean and Standard Deviation of the cables.

c **Difference in variance between the cables**

To test the difference variance, we use the Bartlett's test. For the Bartlett's test to be applicable, we need to first make sure all values are normally distributed. This is shown in figure 2.

R-Commands:

$$qqnorm(y); qqline(y) \qquad \# \textit{Check if is normally distributed.}$$
$$bartlett.test(y \sim cables) \qquad \# \textit{The Bartlett's test.}$$

This results in the following results: Bartlett's K-squared = 10.3054, df = 8, p-value = 0.2442.

Since the p-value,0.2442, is significantly larger than 0.05, it indicates that the result measurements are of equal variance.

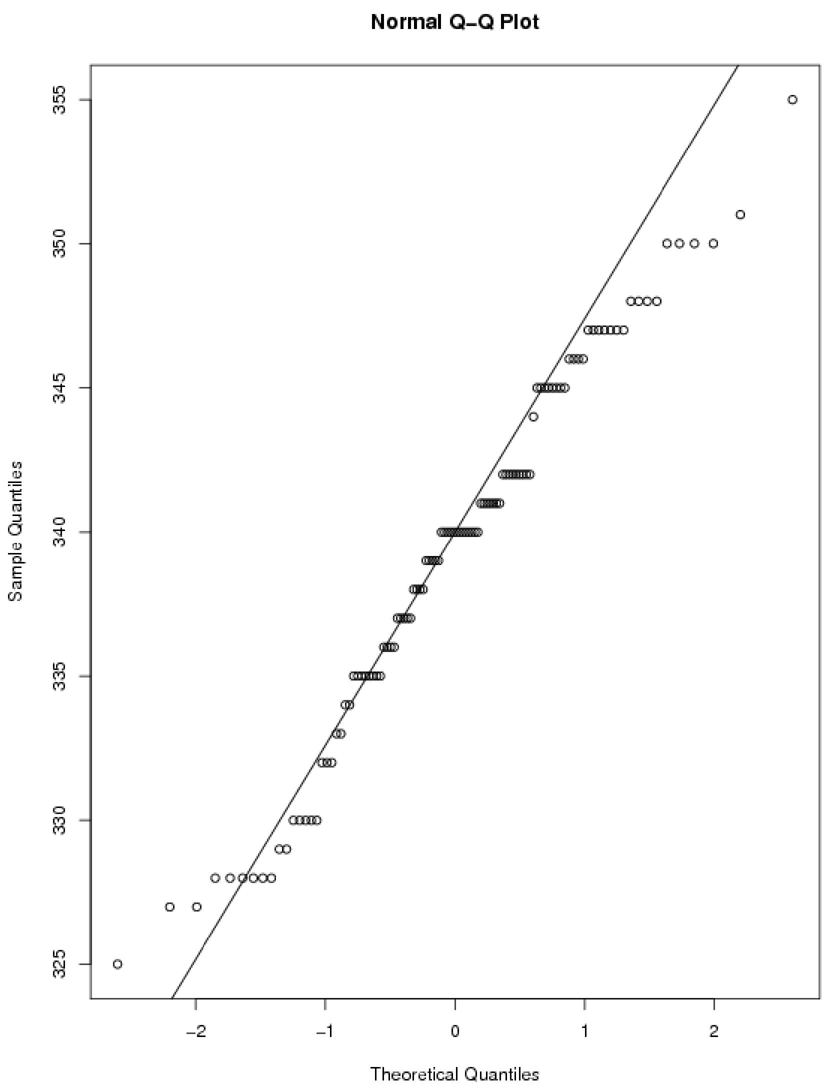

Figure 2: QQplot showing that all results from the cables are approximately normally distributed.

d **Difference between the cables**

To test if there is a significant difference between the cables, we use ANOVA.

R-commands:

$$summary(aov(y \sim cables))$$

The ANOVA table for the differences can be seen in table 3.

	Df	Sum Sq	Mean Sq	F value	Pr(F)
Cables	8	1924.30	240.54	9.0651	2.861e-09
Residuals	99	2626.92	26.53		

Table 3: ANOVA for all cables cables

This gives us a p-value $2.861e-09$, which means there is a significant difference between the cables.

e Significant difference between pairs of cables

For testing if there is a significant difference of pairs of cables, we use Tukey's Test. All comparisons can be ssen in table 4

R-commands:

$$TukeyHSD(aov(allvalues \sim cables))$$

R-commands for plots:

$$par(mfcol = c(2,2));$$
$$plot(TukeyHSD(aov(y[0:48] \sim cables[0:48]))); plot(y \sim lots);$$
$$plot(TukeyHSD(aov(y[49:108] \sim cables[49:108])));$$
$$plot(TukeyHSD(aov(y \sim cables), ordered = TRUE))$$

In figure 3, we show that there is overlap both within lot_A and within lot_B. Furthermore, we show that there is overlap also between lot_A and lot_A. At the end, we show the comparison between all cables.

From table 4 and figure 3 we can see that cables combinations 9-2, 9-3, 9-1, 8-2, 8-3, 5-2, 9-4, 5-3, 7-2, 6-2, 8-1, 7-3, and 6-3 all have significant differences given the defined significance level. In general, it seems that cable 9 and 8 have very significant strength compared to cable 2 and 3.

cable	diff	lwr	upr	p adj
2-1	-2.91666667	-9.5862033	3.752870	0.9001899
3-1	-2.00000000	-8.6695366	4.669537	0.9892109
4-1	1.41666667	-5.2528700	8.086203	0.9990062
5-1	6.00000000	-0.6695366	12.669537	0.1134042
6-1	4.91666667	-1.7528700	11.586203	0.3301469
7-1	5.00000000	-1.6695366	11.669537	0.3078757
8-1	7.41666667	0.7471300	14.086203	0.0177372
9-1	10.33333333	3.6637967	17.002870	0.0001200
3-2	0.91666667	-5.7528700	7.586203	0.9999617
4-2	4.33333333	-2.3362033	11.002870	0.5058630
5-2	8.91666667	2.2471300	15.586203	0.0015997
6-2	7.83333333	1.1637967	14.502870	0.0094665
7-2	7.91666667	1.2471300	14.586203	0.0083163
8-2	10.33333333	3.6637967	17.002870	0.0001200
9-2	13.25000000	6.5804634	19.919537	0.0000003
4-3	3.41666667	-3.2528700	10.086203	0.7890028
5-3	8.00000000	1.3304634	14.669537	0.0072965
6-3	6.91666667	0.2471300	13.586203	0.0359707
7-3	7.00000000	0.3304634	13.669537	0.0320909
8-3	9.41666667	2.7471300	16.086203	0.0006615
9-3	12.33333333	5.6637967	19.002870	0.0000021
5-4	4.58333333	-2.0862033	11.252870	0.4270474
6-4	3.50000000	-3.1695366	10.169537	0.7663385
7-4	3.58333333	-3.0862033	10.252870	0.7427032
8-4	6.00000000	-0.6695366	12.669537	0.1134042
9-4	8.91666667	2.2471300	15.586203	0.0015997
6-5	-1.08333333	-7.7528700	5.586203	0.9998635
7-5	-1.00000000	-7.6695366	5.669537	0.9999256
8-5	1.41666667	-5.2528700	8.086203	0.9990062
9-5	4.33333333	-2.3362033	11.002870	0.5058630
7-6	0.08333333	-6.5862033	6.752870	1.0000000
8-6	2.50000000	-4.1695366	9.169537	0.9571063
9-6	5.41666667	-1.2528700	12.086203	0.2103260
8-7	2.41666667	-4.2528700	9.086203	0.9648392
9-7	5.33333333	-1.3362033	12.002870	0.2279266
9-8	2.91666667	-3.7528700	9.586203	0.9001899

Table 4: Tukeys Test of Cables

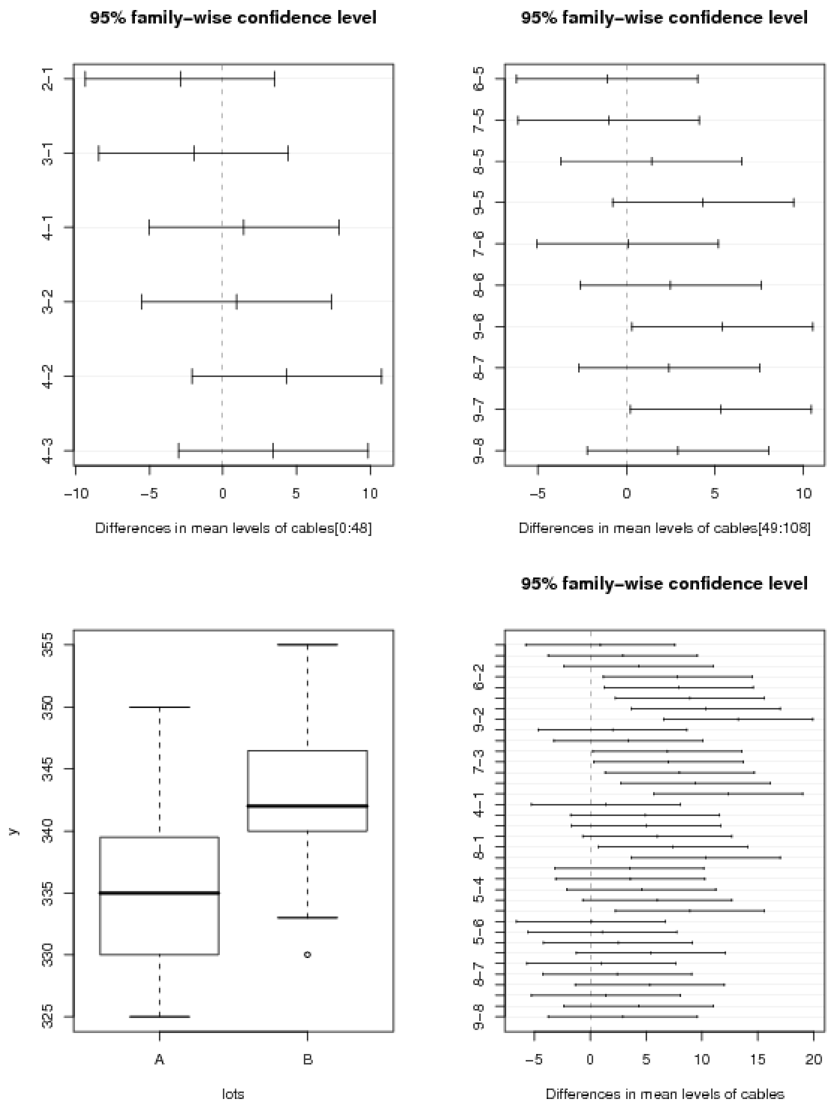

Figure 3: Tukeys Test of Cable Strength

f SS_A **and** SS_B **for cables in** lot_a **and** lot_b**, respectively.**

Lot_a is defined as the four first cables (cable $1-4$ - first 48 values), while lot_b is defined as the five last cables (cable $5-9$ - last 60 values).

To calculate the sum of squares for lot_a and lot_b we use ANOVA.

R-commands:

$$summary(aov(y[1:48] \sim cables[1:48]))\#lot_a$$
$$summary(aov(y[49:108] \sim cables[49:108]))\#lot_b$$

This gives the sum of squares for lot_a,$SS_A = 137.42$ and for lot_b,$SS_B =$ 243.23.

g **Comparing the Lot A and Lot B To be updated.**

Hypothesis:
H_0: $\mu_a = \mu_b$.
H_1: $\mu_a - \mu_b \neq 0$.

For comparison between lot_A and lot_B we again use ANOVA. This also gives us the sum of squares for the complete lot SS_{lots}. R-Commands:

$$summary(aov(y \sim lots))$$

This gives us a sum of squares of 1543.65 and a p-value of 3.804e-11, indicating that there is significant difference between lot_A and lot_B.

h **Individual sum of squares are equal to total sum of squares**

For checking that that the sums of squares match, we use the results previously calculated results for SS_A,SS_B,SS_{lots} and SS_{total}. SS_A,SS_B and SS_{lots} have already been calculated in assignment f and SS_{total} in assignment d.

$$
\begin{aligned}
SS_A + SS_B + SS_{lots} &= SS_{Total} \\
137.42 + 243.2 + 1543.65 &= 1924.30
\end{aligned}
$$

i **Comparing all cables**

Complete ANOVA table is available in table 5, including significance levels as makes as in R.

The F-value is calculated as $F = MS/MSE$ and the P-value as $P = 1 - pf(F, DF, DFE)$

Lot_A:

$$F_{lotb} = \frac{MS_A}{MSE_A} = \frac{45.82}{34.96} = 1.31$$
$$P_{lota} = 1 - pf(F_{lota}, DF_A, DFR_A) = 1 - pf(1.31, 3, 44) = 0.2831475$$

Lot_B:

$$F_{lotb} = \frac{MS_B}{MSE_B} = \frac{60.817}{19.79} = 3.073118$$
$$P_{lotb} = 1 - pf(F_{lotb}, DF_B, DFR_B) = 1 - pf(3.0728, 4, 55) = 0.02349302$$

Lots

$$F_{lots} = \frac{MS}{MSE} = \frac{240.54}{26.53} = 9.066717$$
$$P_{lots} = 1 - pf(F_{lotb}, DF_B, DFR_B) = 1 - pf(9.0651, 8, 99) = 2.860473e - 09$$

Summary of var	**SS**	**DF**	**MS**	**F**	**P**
lot_A	137.43	3	45.81	1.31	0.2831
lot_B	243.23	4	60.81	3.0728	0.02349*
$lots$	1 543.65	1	1 543.65	1 543.65	3.804e-11***
Residential	2 626.92	99	26.54		
Total	4 551.22	107			

Table 5: ANOVA Table

From the data in table 5 we can conclude the following:

- There is a significant difference between lot_a and lot_b.
- There is no significant difference within lot_a.
- There is however, some significant differences within lot_b.