D2 Case study Warpbreaks: solutions

1.
```
> str(warpbreaks)

# 'data.frame': 54 obs. of  3 variables:
#  $ breaks : num  26 30 54 25 70 52 51 26 67 18 ...
#  $ wool   : Factor w/ 2 levels "A","B": 1 1 1 1 1 1 1 1 1 1 ...
#  $ tension: Factor w/ 3 levels "L","M","H": 1 1 1 1 1 1 1 1 1 2 ...
```

   We have 54 observations on the three variables, wool and tension are coded as factors. The breaks are obviously count data.

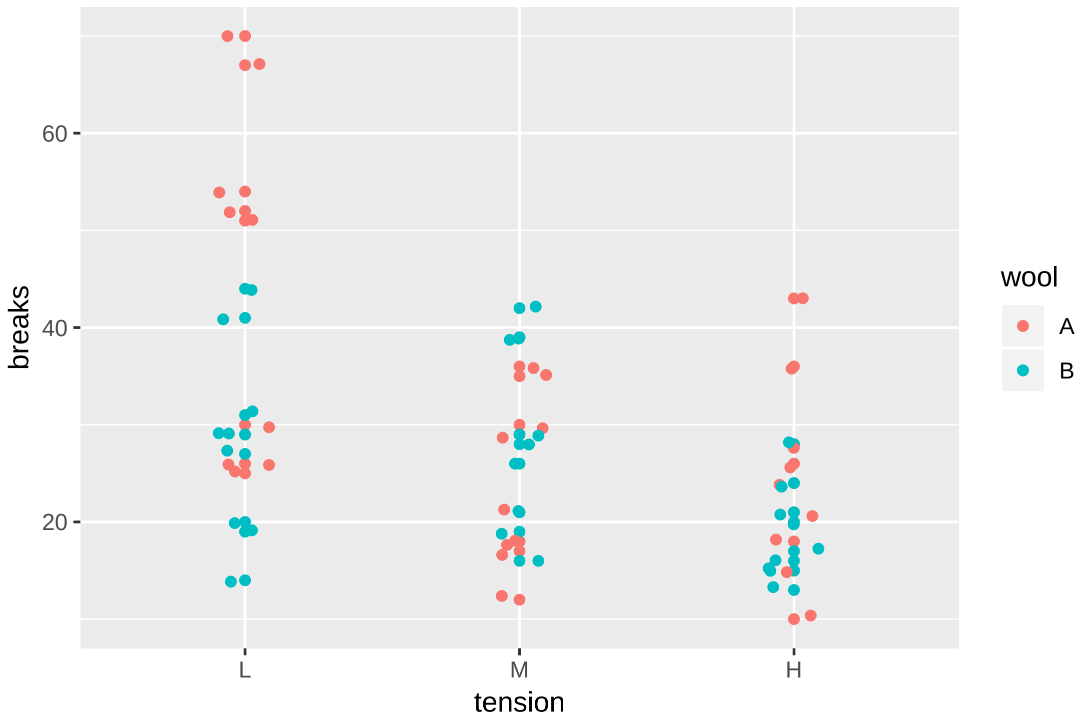
2. Note the use of `list` in the code:

```
> with(warpbreaks, tapply(breaks, list(tension, wool), mean))

#          A        B
# L 44.55556 28.22222
# M 24.00000 28.77778
# H 24.55556 18.77778

> with(warpbreaks, tapply(breaks, list(tension, wool), sd))

#           A        B
# L 18.097729 9.858724
# M  8.660254 9.431036
# H 10.272671 4.893306
```

   We find differences in the means. The `(B, H)` treatment seems to be the best one and the `(L, A)` combination seems particularly bad. This treatment also has sample standard deviation roughly twice that of the other treatments.

3. We use a (jittered) strip plot. Box plots would be possible too.

```
> library(ggplot2)
> ggplot(warpbreaks, aes(x = tension, y = breaks, color = wool)) +
+     geom_point() + geom_jitter(width = 0.1)
```



There are some clusters in the data ...

At the low tension, wool A clearly has more warpbreaks than wool B, but also a higher variability. At medium tension, wool A is perhaps a little better than wool B, but there is much overlap. At high tension, wool A has some observations with a few more warpbreaks. This means we might need to incorporate interaction effects.

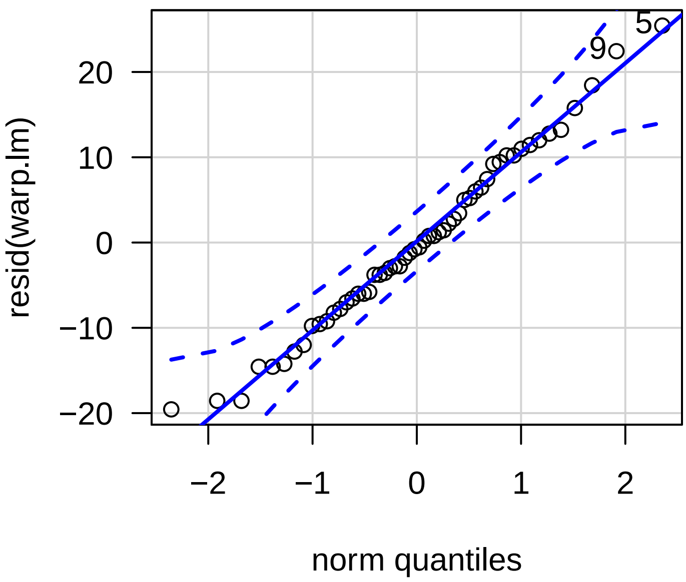
4. We use a two-way ANOVA with interaction effects.

```
> library(car)
> warp.lm <- lm(breaks ~ wool * tension, data = warpbreaks)
> Anova(warp.lm, type = 2)

# Anova Table (Type II tests)
#
# Response: breaks
#               Sum Sq Df F value    Pr(>F)
# wool           450.7  1  3.7653 0.0582130 .
# tension       2034.3  2  8.4980 0.0006926 ***
# wool:tension 1002.8  2  4.1891 0.0210442 *
# Residuals     5745.1 48
# ---
# Signif. codes:
# 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction of wool and tension is significant so that the model may not be further simplified. In other words, the tension effect depends on the wool used.

5. We begin with normality:

```
> library(car)
> qqPlot(resid(warp.lm), las = 1) ## also qqPlot(warp.lm) is okay
```




```
# [1] 5 9

> shapiro.test(resid(warp.lm))$p.value

# [1] 0.8161929
```

The QQ plot looks unsuspicious and the Shapiro-Wilk test does not reject the null hypothesis of normality. No problems are found regarding normality.

Let us check the homoskedasticity assumption now. From the unequal sample standard deviations calculated above, and from the plot of the data too, we can already guess that this is not going to end well.

```
> bartlett.test(breaks ~ interaction(wool, tension), data = warpbreaks)

#
#  Bartlett test of homogeneity of variances
#
# data:  breaks by interaction(wool, tension)
# Bartlett's K-squared = 12.977, df = 5, p-value =
# 0.0236

> leveneTest(breaks ~ interaction(wool, tension), data = warpbreaks)

# Levene's Test for Homogeneity of Variance (center = median)
#       Df F value  Pr(>F)
# group  5   2.891 0.02322 *
#       48
# ---
# Signif. codes:
# 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Both Bartlett's and Levene's test reject the null hypothesis of equal variances. We have found a problem with the equal variance assumption.

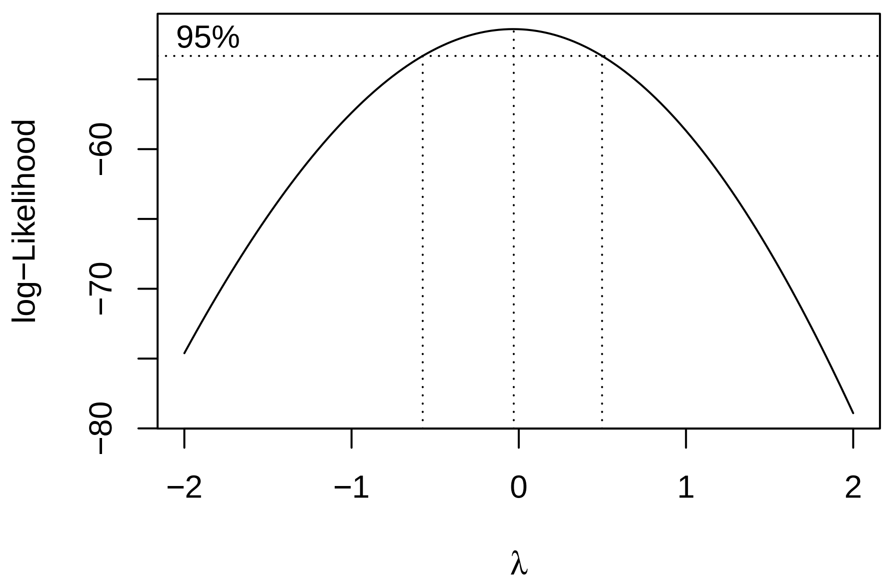
6. Normality is unproblematic, so one way to proceed now is to use heteroskedasticity-corrected methods to test the significance of the factors.

```
> Anova(warp.lm, type = 2, white.adjust = 'hc3')

# Coefficient covariances computed by hccm()

# Analysis of Deviance Table (Type II tests)
#
# Response: breaks
#              Df      F   Pr(>F)
# wool          1 1.4387 0.236230
# tension       2 6.4666 0.003261 **
# wool:tension  2 3.3814 0.042279 *
# Residuals    48
# ---
# Signif. codes:
# 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The resulting ANOVA table adjusts for the unequal variances and confirms that the interaction of wool and tension has a significant effect on the average number of warpbreaks.

7. The aim is to transform the data such that the transformed data have more homogeneous variances. One has to be careful not to destroy the normality of the residuals by transforming. We try the Box-Cox approach:

```
> library(MASS)
> b <- boxcox(breaks ~ wool * tension, data = warpbreaks,
+             lambda = seq(-2, 2, 1/100))
```

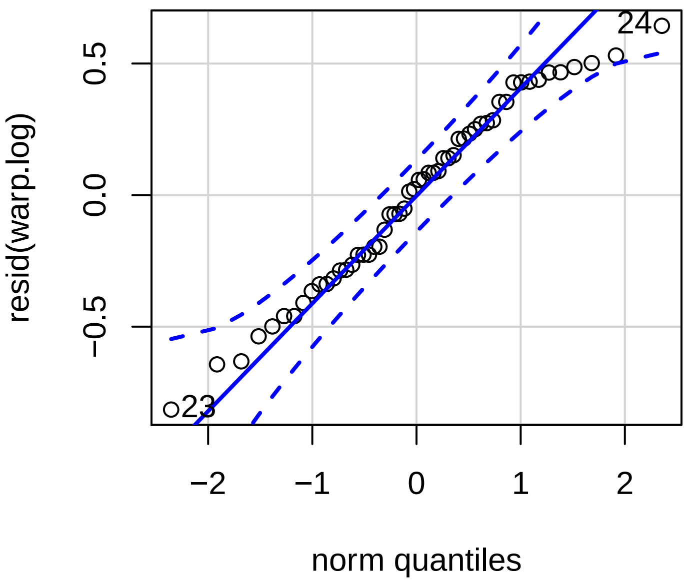



```
> b$x[which.max(b$y)]

# [1] -0.03
```

We see that values of λ near zero give high log-likelihoods; $\lambda = 0$ corresponds to the log function. We thus log-transform the number of warpbreaks. This is not problematic since the minimum number of warpbreaks is 10.

We fit the model and check the normality assumption first:

```
> warp.log <- lm(log(breaks) ~ wool * tension, data = warpbreaks)
> qqPlot(resid(warp.log))
```




```
# [1] 23 24

> shapiro.test(resid(warp.log))

#
#  Shapiro-Wilk normality test
#
# data:  resid(warp.log)
# W = 0.97292, p-value = 0.2583
```

Although the right tail of the normal QQ plot now looks a bit curved, the null hypothesis of normality is not rejected. Let us test whether the homoskedasticity is rejected with the residuals from the log-transformed data:

```
> bartlett.test(log(breaks) ~ interaction(wool, tension),
+               data = warpbreaks)

#
```

```
#  Bartlett test of homogeneity of variances
#
# data:  log(breaks) by interaction(wool, tension)
# Bartlett's K-squared = 2.8778, df = 5, p-value =
# 0.7188

> leveneTest(log(breaks) ~ interaction(wool, tension),
+            data = warpbreaks)

# Levene's Test for Homogeneity of Variance (center = median)
#       Df F value Pr(>F)
# group  5  0.5456 0.7408
#       48
```

There is no reason to reject the variance homogeneity assumption now.

Let us see what the log-transformed data yield as result regarding the significance.

```
> Anova(warp.log, type=2)

# Anova Table (Type II tests)
#
# Response: log(breaks)
#              Sum Sq Df F value   Pr(>F)
# wool         0.3125  1  2.2344 0.141511
# tension      2.1762  2  7.7792 0.001185 **
# wool:tension 0.9131  2  3.2642 0.046863 *
# Residuals    6.7138 48
# ---
# Signif. codes:
# 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The conclusion is the same as with the heteroskedasticity-corrected ANOVA: the interaction term remains significant.

8. We calculate the fitted means for each factor level combination and combine this with the sample means. For brevity, we create the confidence intervals already here.

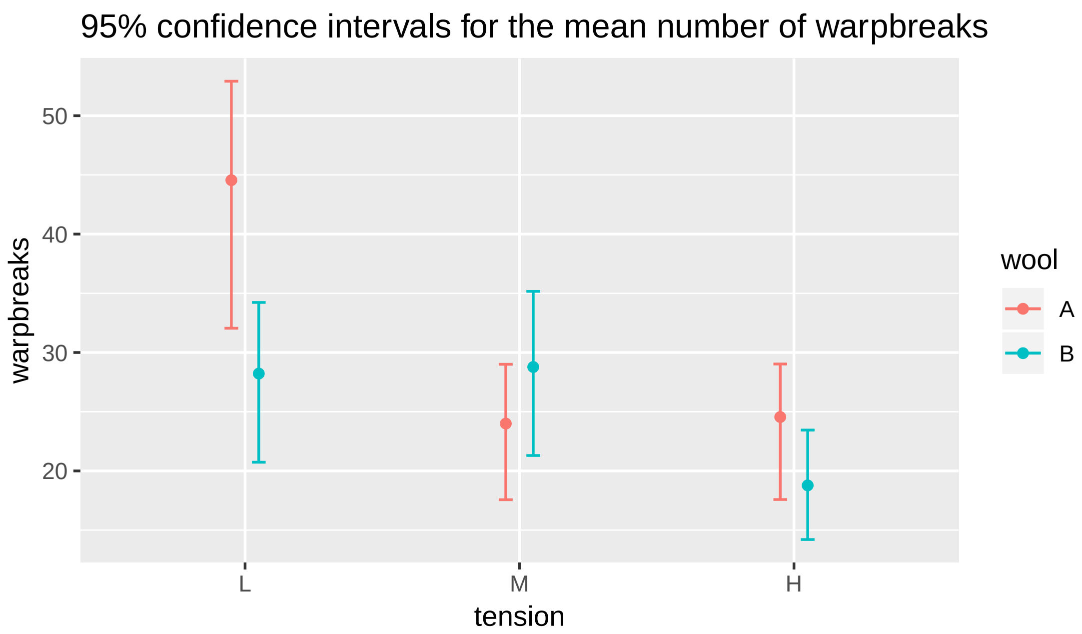
```
> warp.df <- expand.grid(tension = levels(warpbreaks$tension),
+                        wool = levels(warpbreaks$wool))
> log.pred <- predict(warp.log, warp.df, interval = "confidence")
```

```
> ## compute sample means, then bind everything together
> warp.means <- as.numeric(with(warpbreaks,
+                               tapply(breaks, list(tension, wool), mean)))
> warp.ci <- cbind(warp.df, warpbreaks = warp.means,
+                  model = exp(log.pred))
> warp.ci

#   tension wool warpbreaks model.fit model.lwr model.upr
# 1       L    A   44.55556  41.17969  32.04977  52.91043
# 2       M    A   24.00000  22.57289  17.56827  29.00316
# 3       H    A   24.55556  22.59260  17.58361  29.02849
# 4       L    B   28.22222  26.63906  20.73293  34.22765
# 5       M    B   28.77778  27.36669  21.29924  35.16256
# 6       H    B   18.77778  18.24975  14.20361  23.44851
```

We see that the model tends to underfit the sample means a bit (this is due to the log transformation). For example, the `(L,A)` combination has a sample mean of 44.6 warp breaks and a fitted value of 41.2.

9. The 95% confidence intervals on the scale of the data are given above. We plot the resulting confidence intervals, "dodging" the intervals a bit to the left and right so that they are not overplotted.

```
> pos <- position_dodge(0.2)
> ggplot(warp.ci, aes(x = tension, y = warpbreaks, colour = wool)) +
+     geom_errorbar(aes(ymin = model.lwr, ymax = model.upr),
+                   width = .1, position = pos) +
+     geom_line(position = pos) +
+     geom_point(position = pos) +
+     ggtitle("95% confidence intervals for the mean number of warpbreaks")

# geom_path: Each group consists of only one observation.
# Do you need to adjust the group aesthetic?
```



The points representing the sample means are not in the middle of the confidence intervals due to the logarithmic transformation.

10. We have to tell `glht` what comparisons to perform. There are several ways to do this, we show a simple way here.

```
> coef(summary(warp.log))

#                  Estimate Std. Error t value  Pr(>|t|)
# (Intercept)        3.7179     0.1247 29.8236 1.239e-32
# woolB             -0.4356     0.1763 -2.4706 1.709e-02
# tensionM          -0.6012     0.1763 -3.4100 1.326e-03
# tensionH          -0.6003     0.1763 -3.4051 1.345e-03
# woolB:tensionM     0.6281     0.2493  2.5193 1.514e-02
# woolB:tensionH     0.2221     0.2493  0.8908 3.775e-01
```

Comparing two levels can be done by defining a suitable contrast (i. e. by adding or subtracting the coefficients which correspond to the comparison under study).

As first example, for the A wool, what is the difference between the medium and the low tension level? According to the model,

$$\begin{aligned} \mu_{\text{medium,A}} - \mu_{\text{low,A}} &= (\mu + \beta_{\text{medium}}) - \mu \\ &= \beta_{\text{medium}} \,, \end{aligned}$$

which is the first contrast below.

As second example, for the B wool, what is the difference between the mean number of warpbreaks of the `high` tension and the `medium` tension? The difference of the means is

$$\begin{aligned}\mu_{\text{high,B}} - \mu_{\text{medium,B}} &= (\mu + \beta_{\text{high}} + \beta_{\text{B}} + \beta_{\text{high:B}}) - (\mu + \beta_{\text{medium}} + \beta_{\text{B}} + \beta_{\text{medium:B}}) \\ &= \beta_{\text{high}} + \beta_{\text{high:B}} - \beta_{\text{medium}} - \beta_{\text{medium:B}}) \,.\end{aligned}$$

This is the fourth contrast below. The other contrasts are found similarly.

```
> library(multcomp)
> K <-c("tensionM = 0",
+       "tensionH - tensionM = 0",
+       "tensionM + woolB:tensionM = 0",
+       "tensionH + woolB:tensionH - tensionM - woolB:tensionM = 0")
> S <- summary(glht(warp.log, linfct = K))
> names(S$test$coefficients) <- c("A: M - L", "A: H - M",
+                                  "B: M - L", "B: H - M")
> S


#
#   Simultaneous Tests for General Linear Hypotheses
#
# Fit: lm(formula = log(breaks) ~ wool * tension, data = warpbreaks)
#
# Linear Hypotheses:
#                 Estimate Std. Error t value Pr(>|t|)
# A: M - L == 0 -0.601196   0.176303  -3.410  0.00509 **
# A: H - M == 0  0.000873   0.176303   0.005  1.00000
# B: M - L == 0  0.026948   0.176303   0.153  0.99970
# B: H - M == 0 -0.405175   0.176303  -2.298  0.09260 .
# ---
# Signif. codes:
# 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# (Adjusted p values reported -- single-step method)
```

We change the names manually here to improve the display of the solution. A more systematic solution is found in Section 3 of `https://cran.r-project.org/web/packages/multcomp/vignettes/multcomp-examples.pdf`

For the A wool, only the difference between medium and low tension is significant; for the B wool, no significant difference is found.