BE7023 Homework 3

Mike Lape

September 26, 2018

```
#setwd("C:/Users/lapt3u/Box/UC/Fall_2018/BE7023_Adv_Biostats/adv_biostats/hw_3")
library(multcomp)
dat <- litter
```

1. Describe the data.

The litter dataset contains the dose response of litter weights in rats. It contains 4 different variables, 'dose' which represents the dosage at 1 of 4 discrete levels: 0, 5, 50, and 500 units (a unit is not given), 'gesttime' which represents the gestation time for the litter, 'number' representing the number of rats in a litter, and finally 'weight' representing the average post-birth weights of a litter.

```
summary(dat)
```

```
##   dose        weight          gesttime         number
##  0  :20   Min.   :19.22   Min.   :21.50   Min.   : 5.00
##  5  :19   1st Qu.:27.77   1st Qu.:21.50   1st Qu.:12.00
##  50 :18   Median :30.76   Median :22.00   Median :14.00
##  500:17   Mean   :30.33   Mean   :22.09   Mean   :13.43
##           3rd Qu.:33.30   3rd Qu.:22.50   3rd Qu.:15.00
##           Max.   :38.75   Max.   :23.00   Max.   :17.00
```

2. What is the dimension of the data.

```
dim(dat)
```

```
## [1] 74  4
```

```
# The data has 74 rows/observations and 4 columns/variables.
```

3. What is the nature of each variable in the data?

```
sapply(dat,class)
```

```
##      dose    weight  gesttime    number
##  "factor" "numeric" "numeric" "integer"
```

```
levels(dat$dose)
```

```
## [1] "0"   "5"   "50"  "500"
```

Using the class and levels functions we can see dose is a factor with 4 levels, 0, 5, 50, and 500, of an unspecified unit. The rest of the variables, gesttime, number, and weight are numeric. However, number is specifically an integer type meaning there won't be any decimals in that variable, which makes sense as there can't be 0.2 of a litter, litters are made up of a whole number of rats.

4. Get summary statistics of the variables

```
summary(dat)
```

```
##   dose        weight          gesttime         number
## 0  :20   Min.   :19.22   Min.   :21.50   Min.   : 5.00
## 5  :19   1st Qu.:27.77   1st Qu.:21.50   1st Qu.:12.00
## 50 :18   Median :30.76   Median :22.00   Median :14.00
## 500:17   Mean   :30.33   Mean   :22.09   Mean   :13.43
##          3rd Qu.:33.30   3rd Qu.:22.50   3rd Qu.:15.00
##          Max.   :38.75   Max.   :23.00   Max.   :17.00
```

Because dose is a factor, its summary statistics only include counts for each level, where as the rest of the numerical variables have the usual min, max, mean, etc. summary statistics.

5. Obtain the mean 'weight' by dose.

```
means <- aggregate(weight ~ dose, dat, mean)
means$weight <- round(means$weight, 3)
means
```

```
##    dose weight
## 1     0 32.309
## 2     5 29.308
## 3    50 29.866
## 4   500 29.646
```

6. Regress 'weight' by dose

```
mod <- lm(weight ~ dose, dat)
summary(mod)
```

```
##
## Call:
## lm(formula = weight ~ dose, data = dat)
```

```
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.4265  -2.7010   0.6787   3.2280   9.3216
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  32.3085     0.9683  33.365   <2e-16 ***
## dose5        -3.0001     1.3873  -2.162   0.0340 *
## dose50       -2.4424     1.4070  -1.736   0.0870 .
## dose500      -2.6620     1.4286  -1.863   0.0666 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 4.331 on 70 degrees of freedom
## Multiple R-squared:  0.07727,    Adjusted R-squared:  0.03773
## F-statistic: 1.954 on 3 and 70 DF,  p-value: 0.1289
```

7. What is the R2? Is it significant?
   The R2 for this regression is 0.038 with a p-value of 0.129, so it is not significant.

8. Do 'aov' of 'weight' by 'dose'.
   Do doses differ significantly in average weight? Report the p-value.

```
var_mod <- aov(weight ~ dose, dat)
summary(var_mod)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## dose         3  109.9   36.64   1.954  0.129
## Residuals   70 1312.8   18.75
```

```
# With a p-value of 0.129 we can say that the doses do not differ significantly
# in average litter wight.
```

9. Test normality and homoscedasticity

```
shap_test <- shapiro.test(mod$residuals)
shap_test
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  mod$residuals
## W = 0.98325, p-value = 0.4342
```

```
# We can see with a p-value of 0.434 which is greater than 0.05 that according
# to the Shapiro-Wilk test our data is both normal and homoscedastic.
```

10. Do pairwise comparisons by Tukey's HSD

```
tukey <- TukeyHSD(var_mod)
tukey
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = weight ~ dose, data = dat)
##
## $dose
##                diff       lwr       upr     p adj
## 5-0     -3.0000789 -6.651348 0.6511899 0.1439828
## 50-0    -2.4423889 -6.145305 1.2605270 0.3131533
## 500-0   -2.6620294 -6.421829 1.0977702 0.2531876
## 50-5     0.5576901 -3.191100 4.3064800 0.9794716
## 500-5    0.3380495 -3.466938 4.1430375 0.9954573
## 500-50  -0.2196405 -4.074216 3.6349354 0.9987864
```
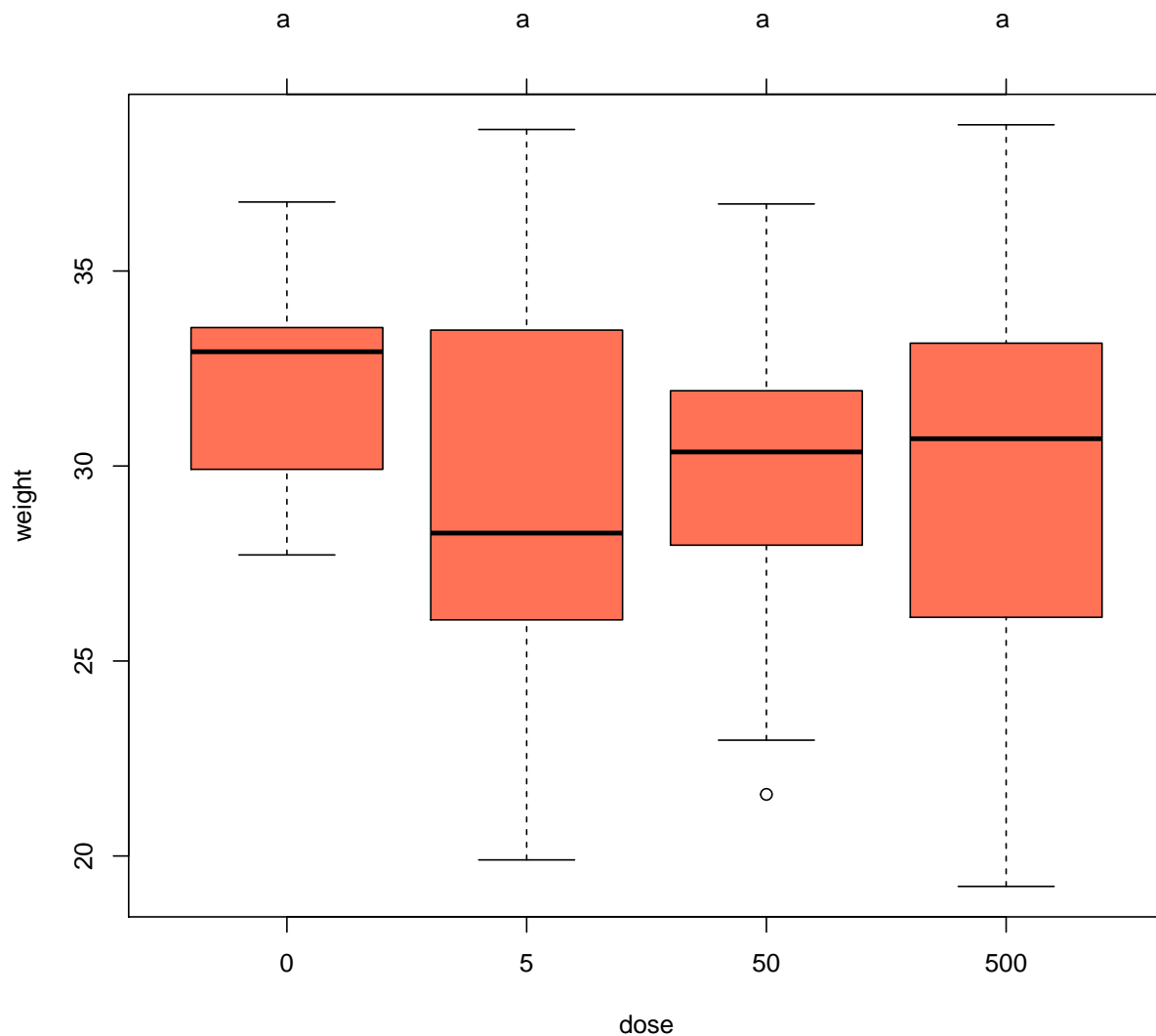
11. Get a graph of the pariwise comparisons

```
par(las=2)
par(mar = c(5,8,4,2))
plot(tukey)
```

95% family–wise confidence level

Differences in mean levels of dose
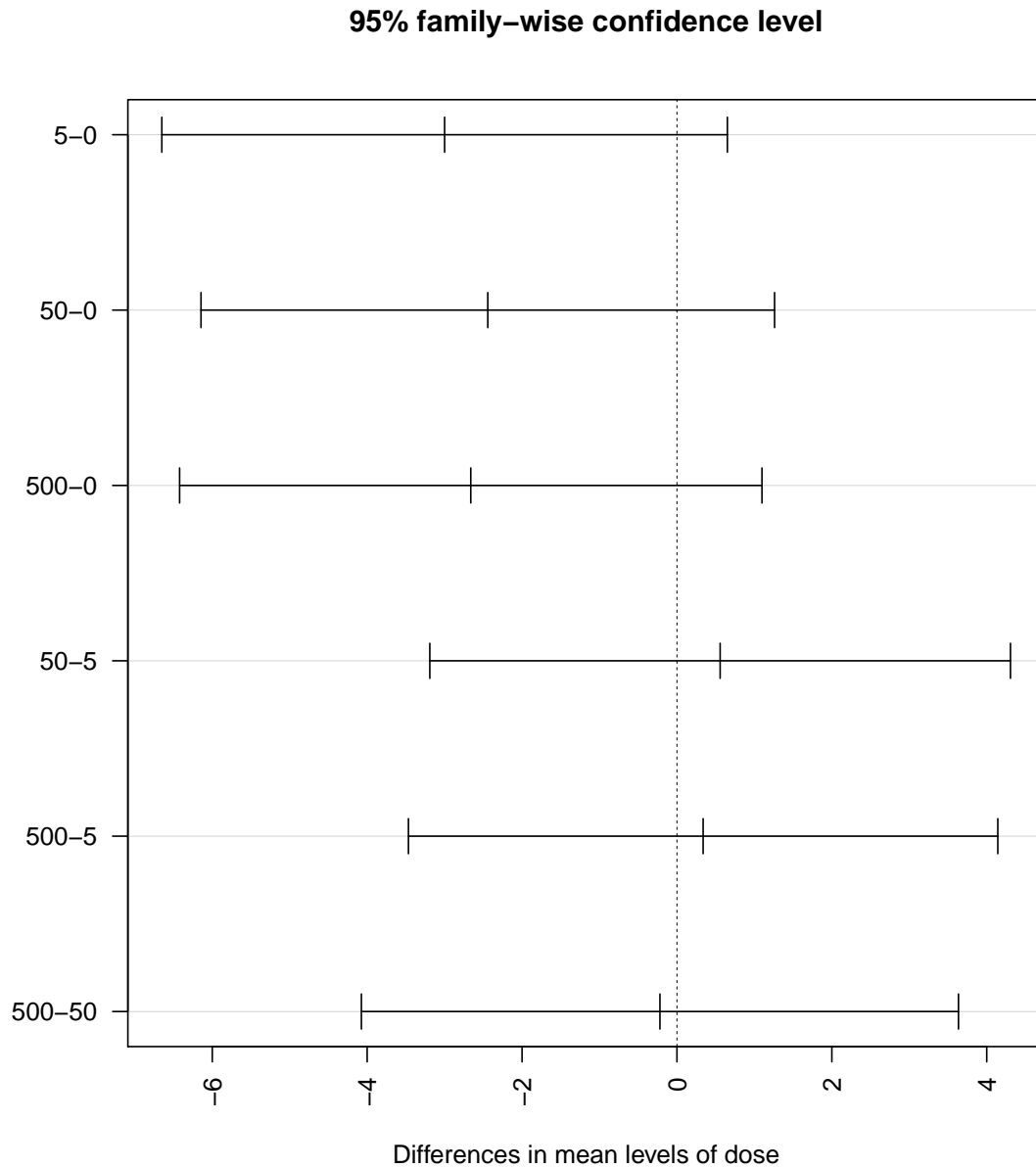
12. Get a graph of cld. Comment on the graph.

```
par(mar = c(5,4,8,2))
lin_hyp <- glht(var_mod, linfct = mcp(dose = "Tukey"))
plot(cld(lin_hyp, level = 0.05), col = "coral1")
```

We can see from the letters across the top, as well as all the boxes overlapping with each other that there is no significant difference in litter weight in response to different doses.

13. Regress weight by dose and gestation. Is gestation significant?

```
dbl_mod <- lm(weight ~ dose + gesttime, dat)
summary(dbl_mod)
```

```
##
## Call:
## lm(formula = weight ~ dose + gesttime, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -11.5649  -2.0068    0.1476    3.0755    7.3023
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -45.366     24.984  -1.816  0.07374 .
## dose5         -3.477      1.318  -2.639  0.01027 *
## dose50        -1.788      1.344  -1.330  0.18780
## dose500       -3.019      1.352  -2.232  0.02883 *
## gesttime       3.519      1.131   3.111  0.00271 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.085 on 69 degrees of freedom
## Multiple R-squared:  0.1908, Adjusted R-squared:  0.1439
## F-statistic: 4.067 on 4 and 69 DF,  p-value: 0.005105
```

We can see from the summary of our model that gesttime is significant with a p-value of 0.00271.

14. Do pairwise comparisons by dose.

```
dbl_tuk <- glht(dbl_mod, linfct = mcp(dose = "Tukey"))
summary(dbl_tuk)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = weight ~ dose + gesttime, data = dat)
##
## Linear Hypotheses:
##               Estimate Std. Error t value Pr(>|t|)
## 5 - 0 == 0     -3.4769     1.3175  -2.639   0.0491 *
## 50 - 0 == 0    -1.7875     1.3437  -1.330   0.5468
## 500 - 0 == 0   -3.0191     1.3524  -2.232   0.1247
## 50 - 5 == 0     1.6894     1.3919   1.214   0.6201
## 500 - 5 == 0    0.4579     1.3642   0.336   0.9868
## 500 - 50 == 0  -1.2315     1.4192  -0.868   0.8213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

15. Get a graph of the pairwise comparisons.

```
par(las=2)
par(mar = c(5,8,4,2))
plot(dbl_tuk)
```

95% family−wise confidence level

Linear Function