

# BTRY 6020 Homework VII

---

**NAME:** student name

**NETID:** student NetID

**DUE DATE:** May 1, 2017, by 8:40 am

---

## Question 1.

A study was designed to investigate the effects of certain toxic agents. Four medications (hypothesized poison antidotes) and 3 poisons were used in a completely randomized factorial experiment consisting of 12 treatment combinations (each poison with each antidote). The design was balanced with 4 animals randomly assigned to each of the 12 treatments. The two factors were fixed in this study as they were the only ones about which statistical inference was needed. For each animal, the survival time (in hours) was recorded following administration of the treatment combination. Data appears in file Hwk7Q1DatSp17.

A) What is the experimental unit in this experiment?

```
library(readxl)
Q1Dat <- read_excel("~/BTRY6020Sp17/Homework/Hwk7/Hwk7Q1DatSp17.xlsx")
Q1Dat$Surv=Q1Dat$survtime
Q1Dat$Med=as.factor(Q1Dat$medication)
Q1Dat$Pois=as.factor(Q1Dat$poison)
```

In this case, the animals were the experimental units-they were randomly assigned the treatments (Poison-medication combinations)

B) Give the general linear model for this analysis and define each term. What constraints are imposed upon this model for the analysis?

$Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \sigma^2)$ , so

$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$ , where

$Y_{ijk}$  = kth observation at level j of factor B and level i of factor A.

$\alpha_i$  = effect of ith level of factor A

$\beta_j$  = effect of jth level of factor B

$(\alpha\beta)_{ij}$  = interaction effect at being at the ith level of factor A and jth level of factor B

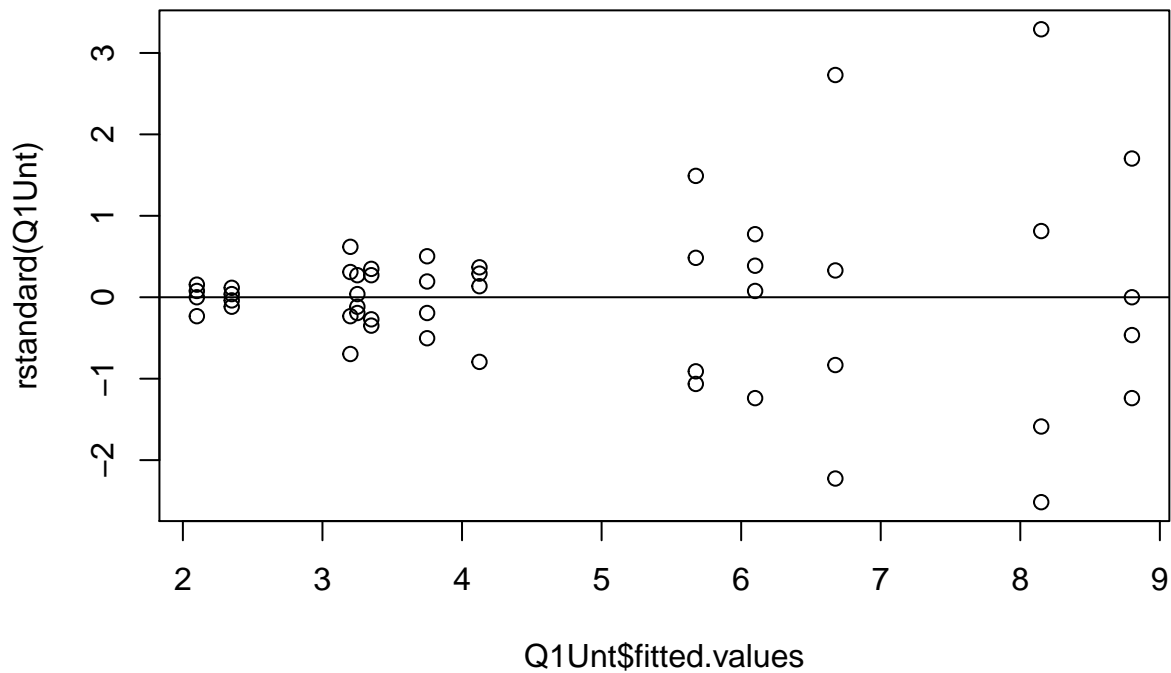
$\epsilon_{ijk}$  = error term of the kth observation at the jth level of factor B and the ith level of factor A.

C) Without any data transformation, analyze the data and report the ANOVA table for the fixed effects model.

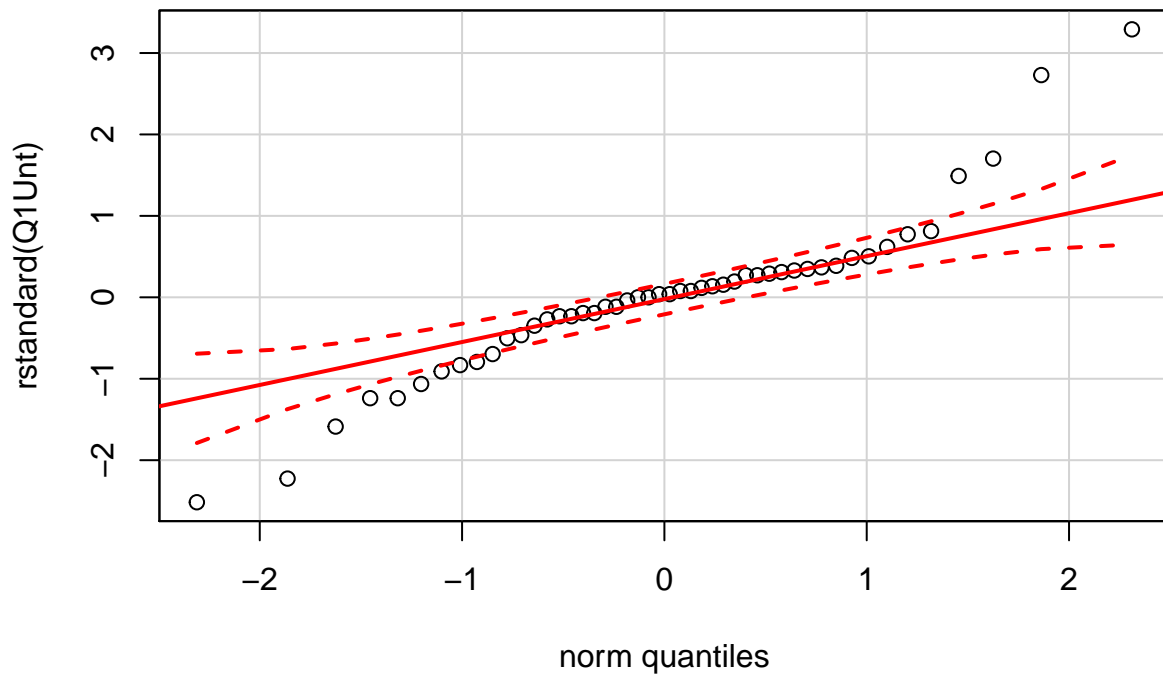
```
Q1Unt <- aov(Surv ~ Pois * Med, data=Q1Dat)
anova(Q1Unt)
```

```
## Analysis of Variance Table
##
## Response: Surv
##      Df Sum Sq Mean Sq F value    Pr(>F)
## Pois      2 103.301   51.651  23.2217 3.331e-07 ***
## Med       3  92.121   30.707  13.8056 3.777e-06 ***
## Pois:Med   6  25.014    4.169   1.8743  0.1123
## Residuals 36  80.072    2.224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(Q1Unt$fitted.values, rstandard(Q1Unt))
abline(h = 0)
library(car)
```



```
qqPlot(rstandard(Q1Unt))
```

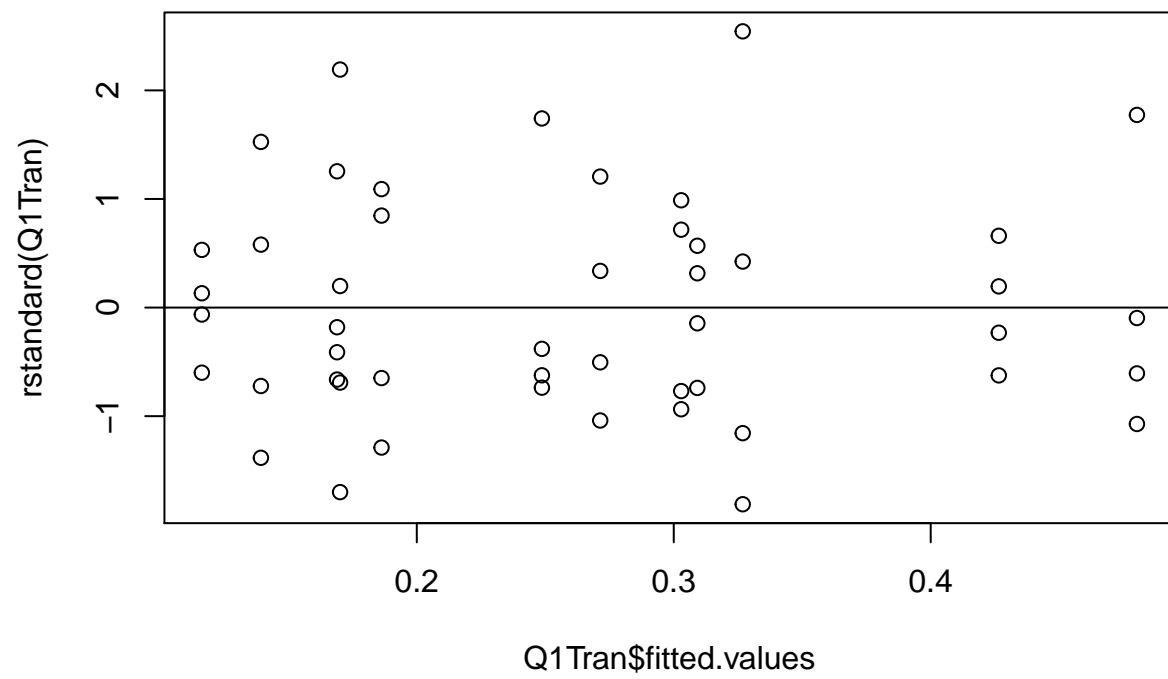


D) Make a plot of standardized residuals against predicted values and assess the validity of the equal variance assumption. Is the assumption of equal variance satisfied? If this assumption is not satisfied, make an appropriate transformation. Which transformation stabilizes the variance of residuals (append your plot)?

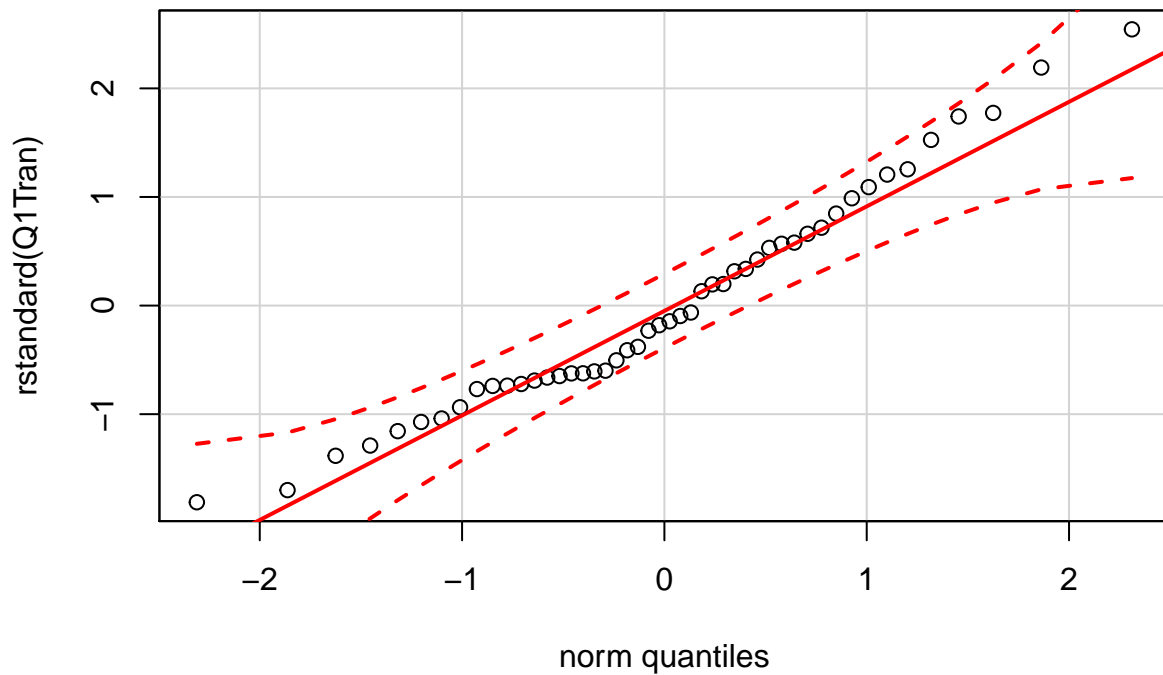
```
Q1Dat$InvSurv=(1/Q1Dat$Surv)
Q1Tran <- aov(InvSurv ~ Pois * Med, data=Q1Dat)
anova(Q1Tran)
```

```
## Analysis of Variance Table
##
## Response: InvSurv
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Pois      2  0.34877  0.174386  72.6347 2.310e-13 ***
## Med       3  0.20414  0.068048  28.3431 1.376e-09 ***
## Pois:Med   6  0.01571  0.002618   1.0904  0.3867
## Residuals 36  0.08643  0.002401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(Q1Tran$fitted.values, rstandard(Q1Tran))
abline(h = 0)
```



```
library(car)
qqPlot(rstandard(Q1Tran))
```

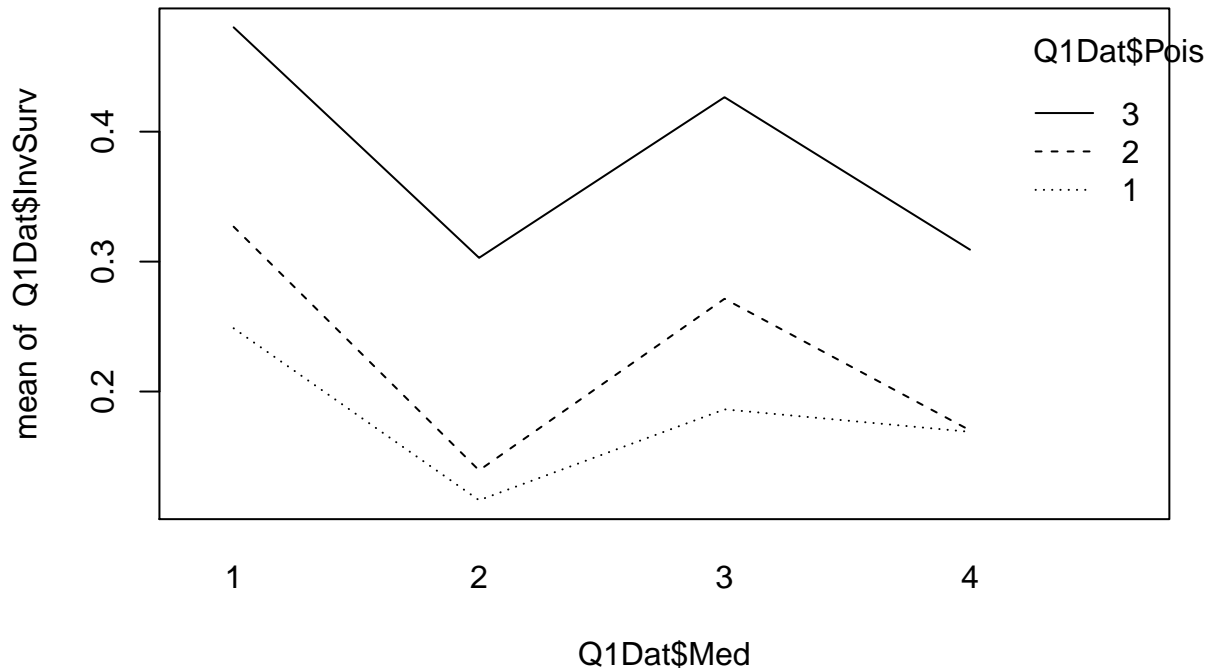


The residual plot using the transformed data shows approximate equal variance, and so this assumption is satisfied. In this case we used the inverse transformation,  $y^t = 1/y = 1/\text{survTime}$ . Also, the normality of residuals is attained with this transformation.

Answer questions (E) through (H) based on the transformed data from (D):

E) Make an interaction plot on your transformed data. What does this tell you, and why (be brief).

```
interaction.plot(Q1Dat$Med, trace.factor=Q1Dat$Pois, response=Q1Dat$InvSurv)
```



From the interaction plot, we see that both factors seem to have an effect, since the curves bounce up and down and are widely separated. However, they are approximately parallel, indicating interactions may be unimportant.

- F) Perform a hypothesis test of the null hypothesis that the interaction between the two factors is not significant. Show the null and alternative hypotheses, the test statistic, the p-value, and state your conclusions. Use  $\alpha = .05$ .

Here we test

$$H_o : (\alpha\beta)_{ij} = 0 \text{ for all } i, j \text{ vs } H_a : \text{Not } H_o$$

TS = MS(Pois:Med)/(MSE) = 1.0904; P =  $P(F_{6,36} > 1.0904) = .387$ , so fail to reject  $H_o$  at any reasonable alpha level, and conclude we have no evidence to say the interactions are not 0.

- G) Perform a hypothesis test of the null hypothesis that the mean survival time is equal across all types of poisons. Show the null and alternative hypotheses, the test statistic, the p-value, and state your conclusions. Use  $\alpha = .05$ .

Here we test:

$$H_o : \alpha_i = 0 \text{ for all } i, \text{ vs } H_a : \text{Not } H_o$$

TS = MS(Pois)/(MSE) = 72.63; P =  $P(F_{2,36} > 72.63) = 2.31 \times 10^{-13}$ , so reject  $H_o$  at any reasonable alpha level, and conclude that the different poisons have an effect on the different survival times.

- H) Perform a hypothesis test of the null hypothesis that the mean survival time is equal across all medications. Show the null and alternative hypotheses, the test statistic, the p-value, and state your conclusions. Use  $\alpha = .05$ .

Here we test:

$$H_o : \beta_j = 0 \text{ for all } j, \text{ vs } H_a : \text{Not } H_o$$

$TS = MS(\text{Med})/(MSE) = 28.34$ ;  $P = P(F_{3,36} > 28.34) = 1.38 \times 10^{-9}$ , so reject  $H_o$  at any reasonable alpha level, and conclude that the different medications have an effect on the different survival times.

- I) What medication would you recommend for each of the poisons? (Be careful; does your transformation effect the interpretation of results from the output?)

```
TukeyHSD(Q1Tran, "Med")
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = InvSurv ~ Pois * Med, data = Q1Dat)
##
## $Med
##      diff      lwr      upr    p adj
## 2-1 -0.16574024 -0.21961439 -0.111866085 0.0000000
## 3-1 -0.05721354 -0.11108770 -0.003339391 0.0338163
## 4-1 -0.13583383 -0.18970798 -0.081959676 0.0000004
## 3-2  0.10852669  0.05465254  0.162400848 0.0000235
## 4-2  0.02990641 -0.02396774  0.083780562 0.4509177
## 4-3 -0.07862029 -0.13249444 -0.024746132 0.0020068
```

Since taking the inverse of numbers, as we did in this transformation, changes the ordering, we look for the Med that produces the least InvSurv. From Tukey's table, we see that no medication is smaller than all the rest. We'll have to look at them individually for each poison.

```
TukeyHSD(Q1Tran)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = InvSurv ~ Pois * Med, data = Q1Dat)
##
## $Pois
##      diff      lwr      upr    p adj
## 2-1 0.04686413 0.004520105 0.08920815 0.0273208
## 3-1 0.19964249 0.157298471 0.24198651 0.0000000
## 3-2 0.15277837 0.110434345 0.19512239 0.0000000
##
## $Med
##      diff      lwr      upr    p adj
## 2-1 -0.16574024 -0.21961439 -0.111866085 0.0000000
## 3-1 -0.05721354 -0.11108770 -0.003339391 0.0338163
## 4-1 -0.13583383 -0.18970798 -0.081959676 0.0000004
## 3-2  0.10852669  0.05465254  0.162400848 0.0000235
## 4-2  0.02990641 -0.02396774  0.083780562 0.4509177
## 4-3 -0.07862029 -0.13249444 -0.024746132 0.0020068
##
## $`Pois:Med`
##      diff      lwr      upr    p adj
## 2:1-1:1 0.078158915 -0.042770940 0.199088771 0.5233013
## 3:1-1:1 0.231580446 0.110650591 0.352510301 0.0000051
## 1:2-1:1 -0.132341687 -0.253271542 -0.011411832 0.0219001
## 2:2-1:1 -0.109348860 -0.230278715 0.011580996 0.1079253
## 3:2-1:1 0.054209193 -0.066720662 0.175139049 0.9105830
## 1:3-1:1 -0.062415711 -0.183345566 0.058514144 0.8066803
```

```

## 2:3-1:1 0.022703837 -0.098226019 0.143633692 0.9999358
## 3:3-1:1 0.177810605 0.056880749 0.298740460 0.0005454
## 1:4-1:1 -0.079719886 -0.200649741 0.041209969 0.4940208
## 2:4-1:1 -0.078534675 -0.199464531 0.042395180 0.5162194
## 3:4-1:1 0.060492437 -0.060437418 0.181422292 0.8349361
## 3:1-2:1 0.153421531 0.032491675 0.274351386 0.0041972
## 1:2-2:1 -0.210500602 -0.331430458 -0.089570747 0.0000322
## 2:2-2:1 -0.187507775 -0.308437630 -0.066577920 0.0002371
## 3:2-2:1 -0.023949722 -0.144879577 0.096980133 0.9998915
## 1:3-2:1 -0.140574626 -0.261504482 -0.019644771 0.0116809
## 2:3-2:1 -0.055455079 -0.176384934 0.065474777 0.8977199
## 3:3-2:1 0.099651689 -0.021278166 0.220581545 0.1923664
## 1:4-2:1 -0.157878801 -0.278808657 -0.036948946 0.0029139
## 2:4-2:1 -0.156693591 -0.277623446 -0.035763735 0.0032122
## 3:4-2:1 -0.017666478 -0.138596334 0.103263377 0.9999949
## 1:2-3:1 -0.363922133 -0.484851988 -0.242992278 0.0000000
## 2:2-3:1 -0.340929306 -0.461859161 -0.219999450 0.0000000
## 3:2-3:1 -0.177371253 -0.298301108 -0.056441397 0.0005663
## 1:3-3:1 -0.293996157 -0.414926012 -0.173066301 0.0000000
## 2:3-3:1 -0.208876609 -0.329806465 -0.087946754 0.0000371
## 3:3-3:1 -0.053769841 -0.174699697 0.067160014 0.9148628
## 1:4-3:1 -0.311300332 -0.432230187 -0.190370477 0.0000000
## 2:4-3:1 -0.310115121 -0.431044977 -0.189185266 0.0000000
## 3:4-3:1 -0.171088009 -0.292017864 -0.050158154 0.0009662
## 2:2-1:2 0.022992827 -0.097937028 0.143922683 0.9999272
## 3:2-1:2 0.186550880 0.065621025 0.307480736 0.0002575
## 1:3-1:2 0.069925976 -0.051003879 0.190855832 0.6788409
## 2:3-1:2 0.155045524 0.034115668 0.275975379 0.0036765
## 3:3-1:2 0.310152292 0.189222436 0.431082147 0.0000000
## 1:4-1:2 0.052621801 -0.068308054 0.173551656 0.9254184
## 2:4-1:2 0.053807012 -0.067122844 0.174736867 0.9145059
## 3:4-1:2 0.192834124 0.071904269 0.313763979 0.0001496
## 3:2-2:2 0.163558053 0.042628198 0.284487908 0.0018195
## 1:3-2:2 0.046933149 -0.073996706 0.167863004 0.9649446
## 2:3-2:2 0.132052696 0.011122841 0.252982552 0.0223783
## 3:3-2:2 0.287159464 0.166229609 0.408089320 0.0000000
## 1:4-2:2 0.029628974 -0.091300882 0.150558829 0.9991857
## 2:4-2:2 0.030814184 -0.090115671 0.151744040 0.9988369
## 3:4-2:2 0.169841297 0.048911441 0.290771152 0.0010736
## 1:3-3:2 -0.116624904 -0.237554760 0.004304951 0.0671541
## 2:3-3:2 -0.031505357 -0.152435212 0.089424499 0.9985805
## 3:3-3:2 0.123601411 0.002671556 0.244531267 0.0414362
## 1:4-3:2 -0.133929079 -0.254858935 -0.012999224 0.0194382
## 2:4-3:2 -0.132743869 -0.253673724 -0.011814013 0.0212503
## 3:4-3:2 0.006283244 -0.114646612 0.127213099 1.0000000
## 2:3-1:3 0.085119548 -0.035810308 0.206049403 0.3970248
## 3:3-1:3 0.240226316 0.119296460 0.361156171 0.0000024
## 1:4-1:3 -0.017304175 -0.138234031 0.103625680 0.9999959
## 2:4-1:3 -0.016118965 -0.137048820 0.104810891 0.9999980
## 3:4-1:3 0.122908148 0.001978293 0.243838003 0.0435210
## 3:3-2:3 0.155106768 0.034176913 0.276036623 0.0036581
## 1:4-2:3 -0.102423723 -0.223353578 0.018506132 0.1642067
## 2:4-2:3 -0.101238512 -0.222168368 0.019691343 0.1758293
## 3:4-2:3 0.037788600 -0.083141255 0.158718456 0.9932491

```



```

## 1:4-3:3 -0.257530491 -0.378460346 -0.136600635 0.0000005
## 2:4-3:3 -0.256345280 -0.377275135 -0.135415425 0.0000006
## 3:4-3:3 -0.117318168 -0.238248023 0.003611688 0.0640821
## 2:4-1:4 0.001185211 -0.119744645 0.122115066 1.0000000
## 3:4-1:4 0.140212323 0.019282468 0.261142178 0.0120145
## 3:4-2:4 0.139027112 0.018097257 0.259956968 0.0131698

```

Examining all the combinations, we can conclude that for poison 1, medications 2, 3, or 4 could be best; for poison 2, medications 2 or 4 could be best, and for poison 3, medications 2 or 4 could be best. In each case, medications 2 or 4 could be best, so to find the best all-around medication, more data need to be collected to see if medication 2 or 4 is overall the best.

## Question 2.

An experiment was conducted to investigate the effect of four diets (1 = diet #1, 2 = diet #2, etc.) on weight gains of two breeds of cows. Twelve three-week-old cows, having the same birth date, were randomly selected from each of the two breeds. So, there were a total of 24 animals in this experiment. Three animals from each breed were randomly assigned to each of the four diets. After a period of five weeks, weight gain (pounds) was recorded for each animal. Data can be found in the file Hwk7Q2DatSp17.

The herdsman wants to know: Which diets maximize the weight gain for these two breeds of cows? (Note that he needs all comparisons on mean weight gains to have a *combined* error rate of no more than 5%.)

Include the following parts in your answer:

- A) Formulation of the research question and choice of the appropriate statistical technique used to answer this question.

Here we want to answer the question: Which diet produces maximum weight gain for each breed of cow? We will use a twoway ANOVA using Diet and Breed as factors, and use Tukey's multiple comparison method for finding the best diet for each breed.

- B) Notation for the random variable(s) and parameter(s) of interest; define these explicitly. Give the distributional assumptions for your random variable(s) and state all assumptions necessary for the statistical application you intend to use.

$Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \sigma^2)$ , so  $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$ , where

$Y_{ijk}$  = kth observation at level j of Breed and level i of Diet.  $\alpha_i$  = effect of ith level of Diet  $\beta_j$  = effect of jth level of Breed  $(\alpha\beta)_{ij}$  = interaction effect at being at the ith level of Diet and jth level of Breed  $\epsilon_{ijk}$  = error term of the kth observation at the jth level of factor B and the ith level of factor A.

Constraints:  $\sum_i \alpha_i = 0$ ,  $\sum_j \beta_j = 0$ ,  $\sum_i (\alpha\beta)_{ij} = 0$ ,  $\sum_j (\alpha\beta)_{ij} = 0$  Distribution of errors:  $\epsilon_{ijk} \sim N(0, \sigma^2)$

Additionally, observations must be independent, and outliers can't be driving our conclusions.

- C) Calculations for the analysis. For hypothesis and significance tests, formulate the null and the alternative hypotheses, calculate the value of your test statistic, and then calculate your p-value. For confidence intervals, show and apply the appropriate formula. Use  $\alpha = 0.05$  if not otherwise specified.

```
library(readxl)
Q2Dat <- read_excel("~/BTRY6020Sp17/Homework/Hwk7/Hwk7Q2DatSp17.xlsx")
Diet.f <- as.factor(Q2Dat$Diet)
Breed.f <- as.factor(Q2Dat$Breed)
Cow1 <- aov(Weight ~ Breed.f * Diet.f, data=Q2Dat)
anova(Cow1)
```

```
## Analysis of Variance Table
##
## Response: Weight
##          Df Sum Sq Mean Sq  F value    Pr(>F)
## Breed.f    1  377.63   377.63  233.8246 5.724e-11 ***
## Diet.f      3  298.31    99.44   61.5717 5.247e-09 ***
## Breed.f:Diet.f  3   28.46     9.49    5.8734 0.006667 **
## Residuals   16   25.84     1.62
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see from the ANOVA table above that there exists a significant Breed-Diet interaction, so we must determine the optimal Diet for each Breed.

## TukeyHSD(Cow1)

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Weight ~ Breed.f * Diet.f, data = Q2Dat)
##
## $Breed.f
##      diff      lwr      upr p adj
## 2-1 7.933333 6.833499 9.033167 0
##
## $Diet.f
##      diff      lwr      upr      p adj
## 2-1 -3.916667 -6.0158313 -1.817502 0.0003507
## 3-1 5.383333 3.2841687 7.482498 0.0000091
## 4-1 -2.366667 -4.4658313 -0.267502 0.0245108
## 3-2 9.300000 7.2008354 11.399165 0.0000000
## 4-2 1.550000 -0.5491646 3.649165 0.1911665
## 4-3 -7.750000 -9.8491646 -5.650835 0.0000001
##
## $`Breed.f:Diet.f`
##      diff      lwr      upr      p adj
## 2:1-1:1 10.000000 6.40758418 13.5924158 0.0000010
## 1:2-1:1 -4.000000 -7.59241582 -0.4075842 0.0236751
## 2:2-1:1 6.166667 2.57425084 9.7590825 0.0004294
## 1:3-1:1 7.733333 4.14091751 11.3257492 0.0000302
## 2:3-1:1 13.033333 9.44091751 16.6257492 0.0000000
## 1:4-1:1 -0.500000 -4.09241582 3.0924158 0.9996011
## 2:4-1:1 5.766667 2.17425084 9.3590825 0.0008816
## 1:2-2:1 -14.000000 -17.59241582 -10.4075842 0.0000000
## 2:2-2:1 -3.833333 -7.42574916 -0.2409175 0.0322235
## 1:3-2:1 -2.266667 -5.85908249 1.3257492 0.4083864
## 2:3-2:1 3.033333 -0.55908249 6.6257492 0.1321370
## 1:4-2:1 -10.500000 -14.09241582 -6.9075842 0.0000005
## 2:4-2:1 -4.233333 -7.82574916 -0.6409175 0.0153167
## 2:2-1:2 10.166667 6.57425084 13.7590825 0.0000008
## 1:3-1:2 11.733333 8.14091751 15.3257492 0.0000001
## 2:3-1:2 17.033333 13.44091751 20.6257492 0.0000000
## 1:4-1:2 3.500000 -0.09241582 7.0924158 0.0590248
## 2:4-1:2 9.766667 6.17425084 13.3590825 0.0000014
## 1:3-2:2 1.566667 -2.02574916 5.1590825 0.7919741
## 2:3-2:2 6.866667 3.27425084 10.4590825 0.0001268
## 1:4-2:2 -6.666667 -10.25908249 -3.0742508 0.0001787
## 2:4-2:2 -0.400000 -3.99241582 3.1924158 0.9999090
## 2:3-1:3 5.300000 1.70758418 8.8924158 0.0020770
## 1:4-1:3 -8.233333 -11.82574916 -4.6409175 0.0000137
## 2:4-1:3 -1.966667 -5.55908249 1.6257492 0.5721051
## 1:4-2:3 -13.533333 -17.12574916 -9.9409175 0.0000000
## 2:4-2:3 -7.266667 -10.85908249 -3.6742508 0.0000647
## 2:4-1:4 6.266667 2.67425084 9.8590825 0.0003596
```

From Tukey's output, we see that for Breed 1, Diet 3 produces greater weight gain over all other three diets.

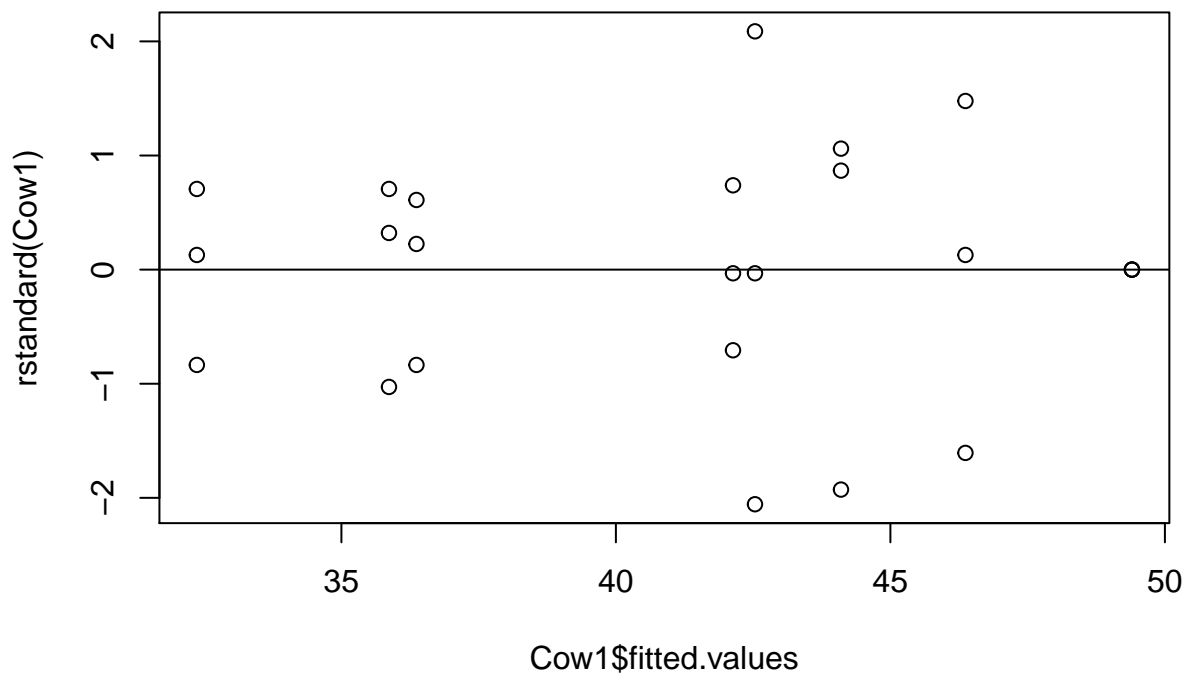
For Breed 2, it's not so straight forward> Both Diets 1 and 3 are better than either 2 or 4, but there is no significant difference between Diets 1 and 3. Therefore, the best diet for Breed 2 could be either Diet 1 or

Diet 3.

- D) Discuss whether the assumptions stated in Part B) above are met sufficiently for the validity of the statistical inferences; use graphs and other tools where applicable.

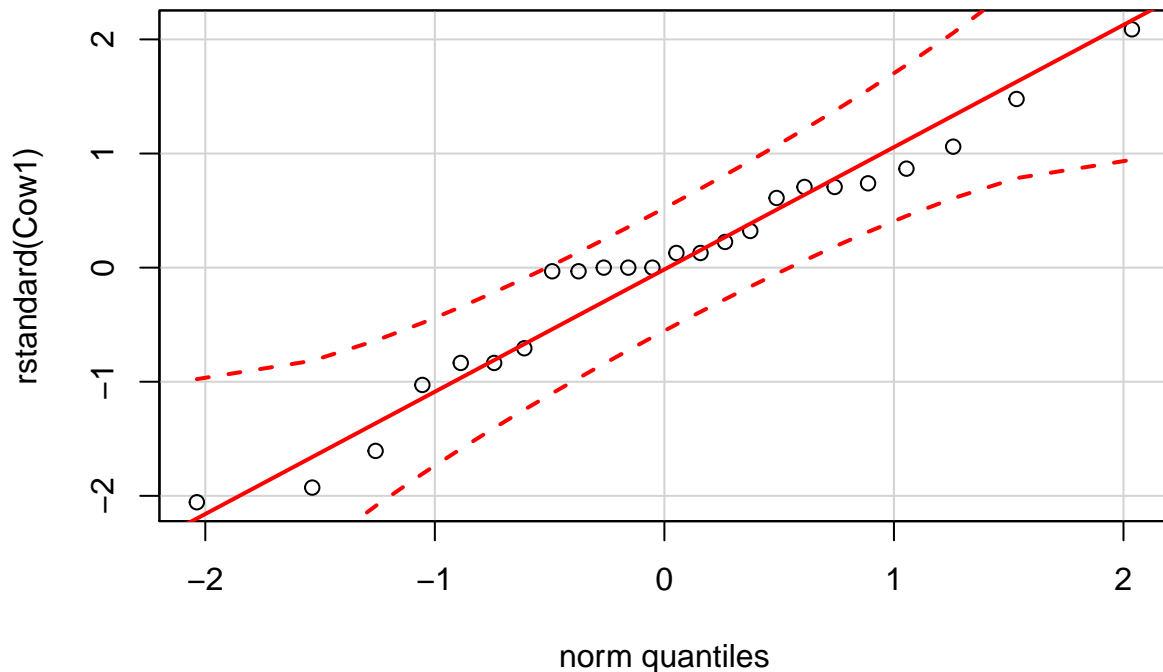
The independence assumption is met by the random selection of cows and randomized assignment of cows to diets.

```
plot(Cow1$fitted.values, rstandard(Cow1))  
abline(h = 0)
```



The constant variance assumption is met, from the residual plot, though there is some slight indication that variance increases as weight increases. However, this view is not pronounced and is based solely on a few data points, so this assumption is not violated.

```
library(car)  
qqPlot(rstandard(Cow1))
```



The normality assumption is met as none of the points in our qqPlot deviate outside the 95% confidence bounds.

Also, there are no outliers driving our conclusions, as none are apparent in either plot.

- E) Discuss the sampling scheme and whether or not it is sufficient to meet the objective of the study. Be sure to include whether or not subjective inference is necessary and if so, defend whether or not you believe it is valid.

With the random selection of cows, and the random assignment of cows to diets, the sampling scheme here is sufficient. However, with only 3 animals from each breed assigned to each diet, power may be low, and with increasing sample size we may have been able to find a truly optimal diet for Breed 2 from the two diets we determined could be the best.

- F) State the conclusions of the analysis. These should be practical conclusions from the context of the problem, but should also be backed up with statistical criteria (like a p-value, etc.). Include any considerations such as limitations of the sampling scheme, impact of outliers, etc., that you feel must be considered when you state your conclusions.

A two factor factorial experiment, with a completely randomized design, was done to determine which of four diets produced the greatest weight gain in two breeds of cows. We determined that for Breed 1, Diet 3 produced greater weight gains than all the other diets. However, for Breed 2, we could only determine that Diets 1 and 3 were better than Diets 2 and 4. We could not determine if Diet 1 or Diet 3 produced better weight gains for Breed 2, with only 3 animals per diet. We recommend that the experiment be redone with Breed 2 only and Diets 1 and 3 only, with more animals per diet, so as to be able to detect which of these two diets is optimal.

However, there were no other issues with this experiment, and the conclusions we have drawn are sound.