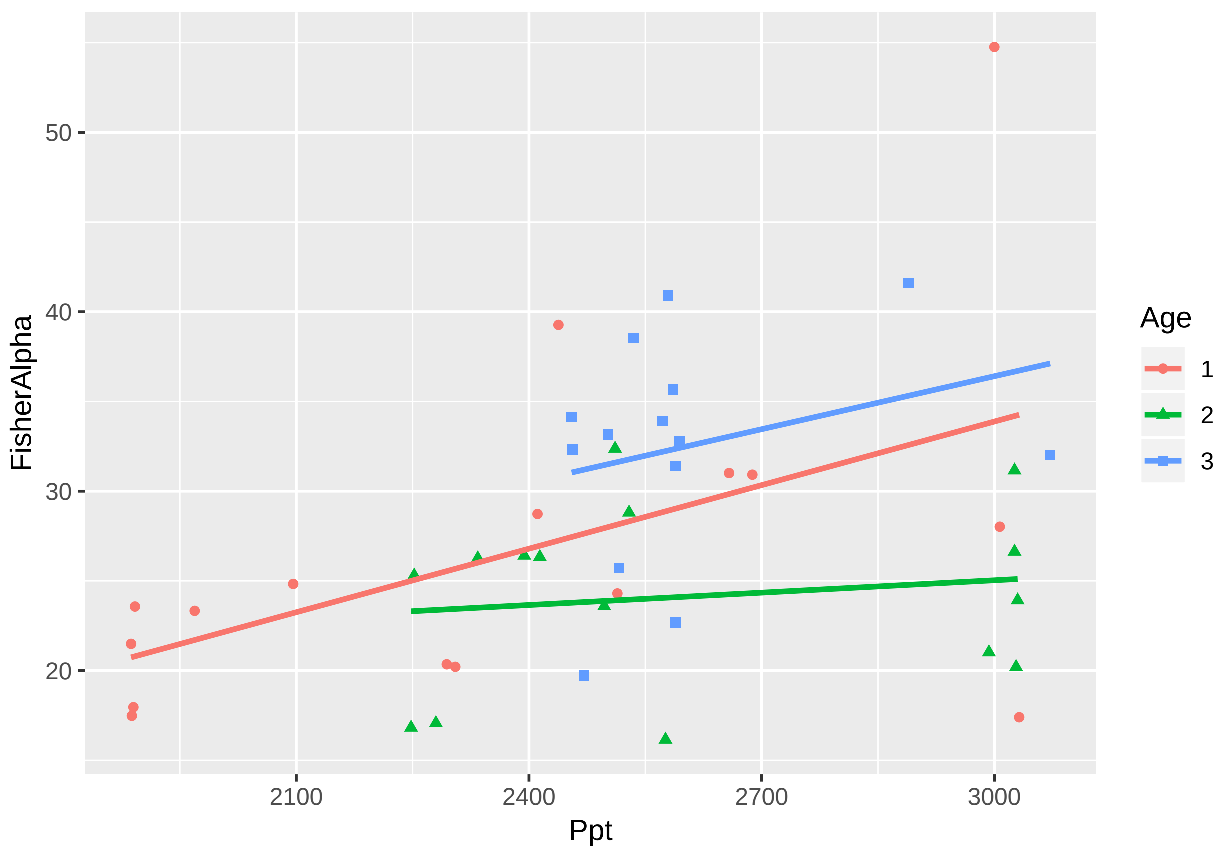
HW4 KEY

44 points total, 4 points per problem part unless otherwise noted.

FisherAlpha Part1

1. Scatterplot (**2pts**)

```
library(ggplot2)
AlphaData <- read.csv("C:/hess/STAT511_FA11/ASCII-comma/CH16/ex16-23.txt", quote = " ' ")
AlphaData$Age <- as.factor(AlphaData$Age)
qplot(Ppt, FisherAlpha, shape = Age, color = Age, data = AlphaData) +
  geom_smooth(method = "lm", se = FALSE)
```



2. ANCOVA WITH interaction (ANOVA table)
   Based on the F-test for interaction (p = 0.4195) we cannot conclude that there are differences between the slopes for the Age groups.

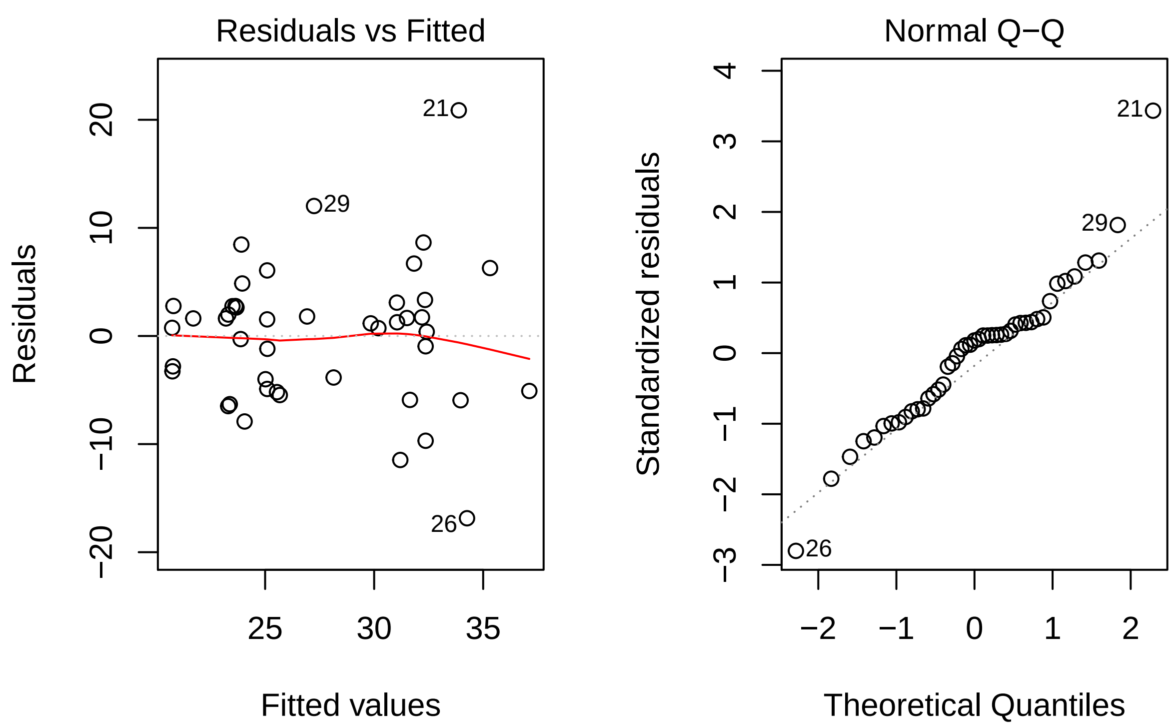
```
library(car)
Model1 <- lm(FisherAlpha ~ Age*Ppt, data = AlphaData)
Anova(Model1, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: FisherAlpha
```

```
##                  Sum Sq Df  F value   Pr(>F)
## (Intercept)     1.09  1  0.0231 0.879892
## Age            54.67  2  0.5826 0.563257
## Ppt           365.74  1  7.7946 0.008074 **
## Age:Ppt        83.36  2  0.8883 0.419514
## Residuals    1829.95 39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3. Interpretation of Diagnostic plots
   Full credit for anything that looks reasonable. Plots not required, but shown here. Based on the plot of Residuals vs Fitted values, there is some evidence of unequal variance. But this may be driven by observations 21 and 26. These observations appear as outliers in the QQplot.

```
par(mfrow=c(1,2))
plot(Model1, which = c(1,2))
```



4. Intercepts and Slopes (**6pts**)

Age	Intercept	Slope	p-value
1	-1.548	0.011	0.008
2	18.139	0.002	0.693
3	6.866	0.009	0.376

```
Model2 <- lm(FisherAlpha ~ Age + Age:Ppt -1, data = AlphaData)
summary(Model2)
```

```
##
```

```
## Call:
## lm(formula = FisherAlpha ~ Age + Age:Ppt - 1, data = AlphaData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.859  -4.899   1.168   2.769  20.879
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## Age1     -1.548882  10.183315  -0.152  0.87989
## Age2     18.139678  15.189510   1.194  0.23960
## Age3      6.866477  28.680668   0.239  0.81204
## Age1:Ppt  0.011810   0.004230   2.792  0.00807 **
## Age2:Ppt  0.002298   0.005782   0.397  0.69319
## Age3:Ppt  0.009847   0.011007   0.895  0.37647
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.85 on 39 degrees of freedom
## Multiple R-squared:  0.9505, Adjusted R-squared:  0.9429
## F-statistic: 124.8 on 6 and 39 DF,  p-value: < 2.2e-16
```

5. Pairwise comparison of Slopes
   Based on the Tukey adjusted pairise comparisons (all p-value > 0.05), we cannot conclude that there are differences between the slopes for the Age groups.

```r
library(emmeans)
emtrends(Model1, pairwise ~ Age, var = "Ppt")
```

```
## $emtrends
##  Age   Ppt.trend          SE df     lower.CL   upper.CL
##  1   0.011809953 0.004230100 39  0.003253769 0.02036614
##  2   0.002298146 0.005781921 39 -0.009396893 0.01399319
##  3   0.009847255 0.011007019 39 -0.012416542 0.03211105
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast     estimate          SE df t.ratio p.value
##  1 - 2     0.009511807 0.007164102 39   1.328  0.3887
##  1 - 3     0.001962698 0.011791870 39   0.166  0.9848
##  2 - 3    -0.007549109 0.012433225 39  -0.607  0.8171
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

6. emmeans at Ppt = 2500 and 3000

```r
emmeans(Model1, ~ Age, at = list(Ppt = 2500))
```

```
## NOTE: Results may be misleading due to involvement in interactions

##  Age   emmean       SE df lower.CL upper.CL
##  1   27.97600 1.794697 39 24.34588 31.60612
##  2   23.88504 1.877969 39 20.08649 27.68359
##  3   31.48462 2.138170 39 27.15976 35.80947
##
## Confidence level used: 0.95
```

```r
emmeans(Model1, ~ Age, at = list(Ppt = 3000))
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
##  Age   emmean       SE df lower.CL upper.CL
##  1   33.88098 3.156859 39 27.49563 40.26633
##  2   25.03412 2.869460 39 19.23008 30.83815
##  3   36.40824 4.764627 39 26.77087 46.04561
##
## Confidence level used: 0.95
```

FisherAlpha Part2

7. Using "backward elimination" we start with ANCOVA WITH interaction (most complicated model considered) and check significance of the Age:Ppt interaction (highest order term). Since the interaction is not significant (p = 0.419) we drop the interaction and reduce to the ANCOVA NO interaction model. Using the ANCOVA NO interaction model, both Age (p = 0.0076) and Ppt (p = 0.0114) are significant, so we use this model. Hence, the preferred model is **ANCOVA NO interaction**.

```r
Anova(Model1, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: FisherAlpha
##              Sum Sq Df F value   Pr(>F)
## (Intercept)    1.09  1  0.0231 0.879892
## Age           54.67  2  0.5826 0.563257
## Ppt          365.74  1  7.7946 0.008074 **
## Age:Ppt       83.36  2  0.8883 0.419514
## Residuals   1829.95 39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
Model2 <- lm(FisherAlpha ~ Age + Ppt, data = AlphaData)
Anova(Model2, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: FisherAlpha
##              Sum Sq Df F value   Pr(>F)
## (Intercept)   27.24  1  0.5838 0.449206
## Age          513.21  2  5.4988 0.007663 **
## Ppt          327.34  1  7.0146 0.011426 *
## Residuals   1913.31 41
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

8. The preferred model is **ANCOVA NO interaction**, since this model has the smallest AIC.

```r
library(MuMIn)
options(na.action = "na.fail")
dredge(Model1, rank = "AIC")
```

```
## Fixed term is "(Intercept)"
```

```
## Global model call: lm(formula = FisherAlpha ~ Age * Ppt, data = AlphaData)
```

```
## ---
## Model selection table
##    (Int) Age      Ppt Age:Ppt df   logLik   AIC delta weight
## 4  6.039   + 0.008613          5 -148.226 306.5  0.00  0.673
## 8 -1.549   + 0.011810       +  7 -147.223 308.4  2.00  0.248
## 2 26.480   +                   4 -151.779 311.6  5.11  0.052
## 3  6.491     0.008353          3 -153.572 313.1  6.69  0.024
## 1 27.560                       2 -156.577 317.2 10.70  0.003
## Models ranked by AIC(x)
```

BodyFat

9. Using **backwards elimination**, the model with **Triceps and Midarm** is selected.

```
BodyFat <- read.csv("C:/hess/STAT512/HW_2019/HW2/BodyFat.csv")
FullModel <- lm(BodyFat ~ ., data = BodyFat)
summary(FullModel)
ModelA <- lm(BodyFat ~ Triceps + Midarm, data = BodyFat)
summary(ModelA)
```

10. With **forward selection**, the model with **Thigh** is selected.

```
NullModel <- lm(BodyFat ~ 1, data = BodyFat)
ModelB <- NullModel
add1(ModelB, scope = FullModel, test = "F")
ModelB <- update(ModelB, ~ . + Thigh)
add1(ModelB, scope = FullModel, test = "F")
summary(ModelB)
```

11. Using **AICc**, the model with **Thigh** is selected.

```
options(na.action = "na.fail")
dredge(FullModel)
```