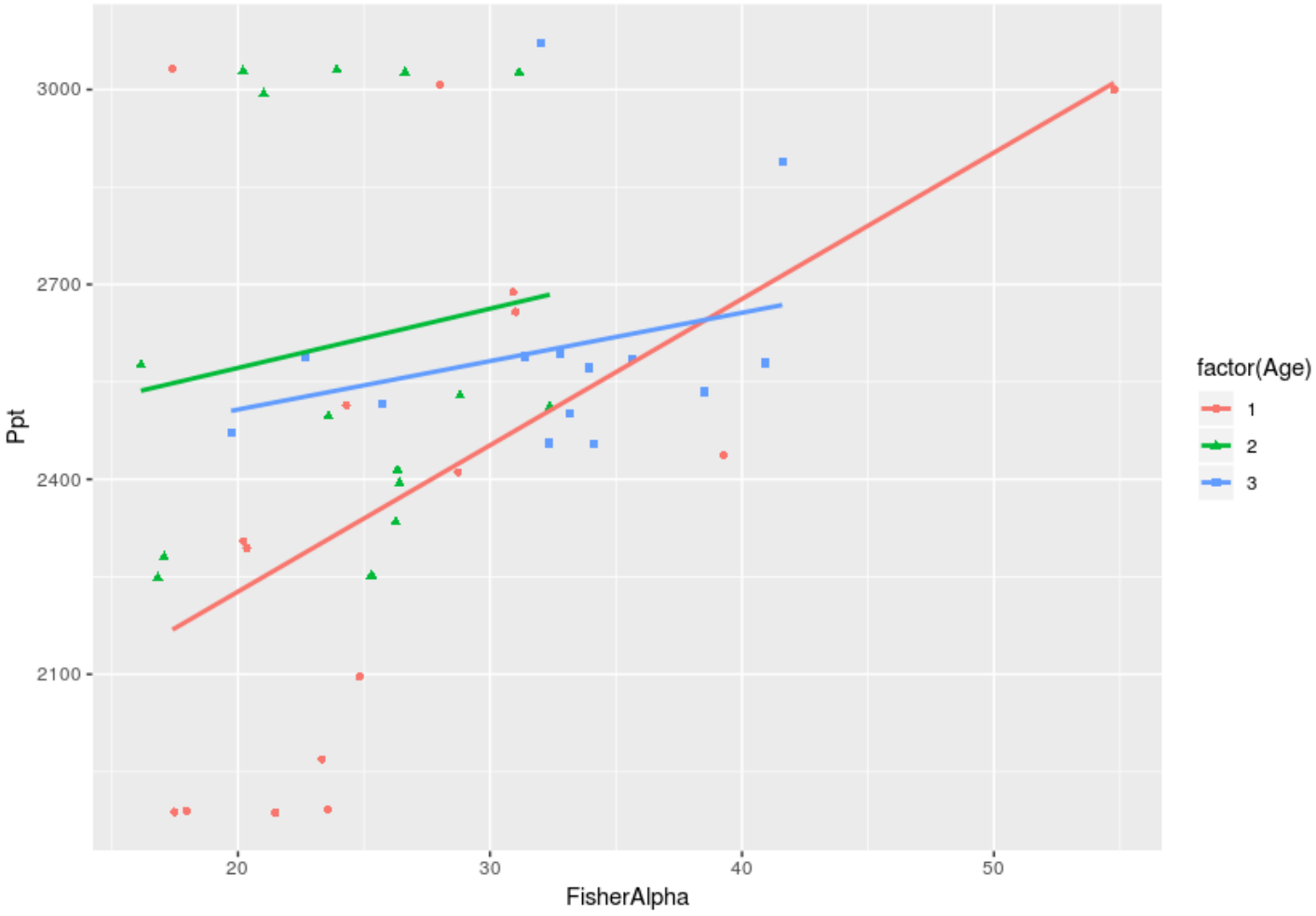
STAT 512 – Assignment 4

Vignesh M. Pagadala
Vignesh.Pagadala@ColoState.Edu

Mar 4, 2019

1. Construct a scatterplot of FisherAlpha (Y) vs Ppt (X) for all Age groups on the same plot. Overlay a fitted regression line for each Age group. (2 pts)

ANSWER:



2. Fit the ANCOVA model WITH interaction. Include the Type 3 ANOVA table in your assignment. What can we conclude about differences between the slopes for the Age groups? Briefly justify your response.

ANSWER:

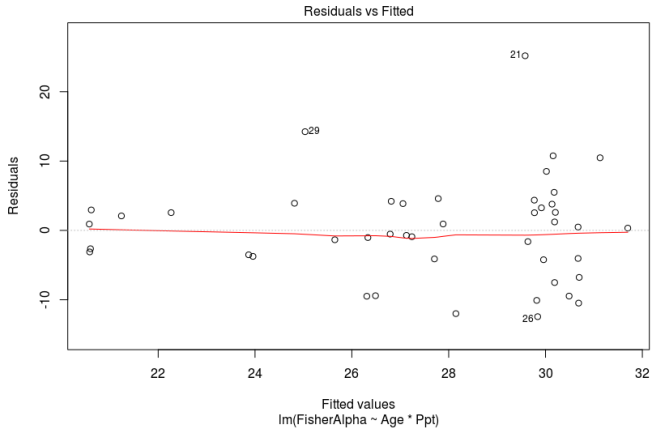
Anova Table (Type III tests)

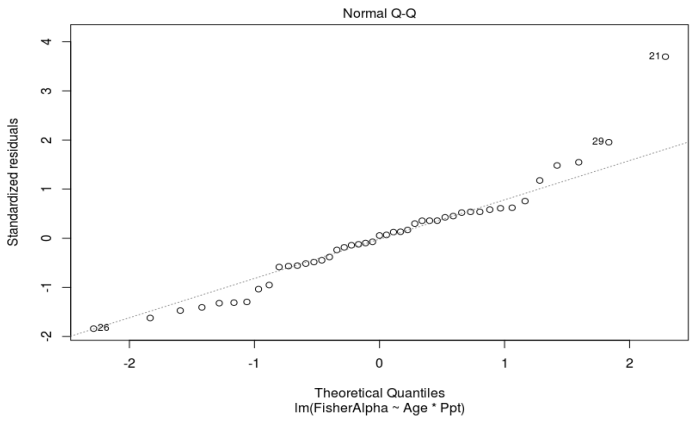
```
Response: FisherAlpha
               Sum Sq Df F value Pr(>F)
(Intercept)    1.11  1  0.0198 0.8887
Age           20.45  1  0.3651 0.5490
Ppt           81.54  1  1.4558 0.2345
Age:Ppt       11.70  1  0.2089 0.6500
Residuals   2296.41 41
```

We can see that the p values are not significant at all. It appears that the slopes increase as age decreases.

3. Consider the diagnostic plots (Resids vs Fitted and QQplot of residuals). You do not need to
include these plots in your assignment, but briefly discuss your findings.

ANSWER:





The resids vs fitted plot does not look good, since we can observe a clear megafone shape. The same is with the Q-Q plot, as there are outliers in the plot.

4. For each Age group, provide the estimated intercept, slope (corresponding to Ppt) and p-value
corresponding to a test of the slope. (6 pts)

ANSWER:

Coefficients:

```
               Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.548882  10.183315  -0.152  0.87989
Age.f2      19.688559  18.287184   1.077  0.28826
Age.f3       8.415358  30.434859   0.277  0.78362
Ppt          0.011810   0.004230   2.792  0.00807 **
Age.f2:Ppt  -0.009512   0.007164  -1.328  0.19200
Age.f3:Ppt  -0.001963   0.011792  -0.166  0.86867
```

5. Calculate Tukey adjusted pairwise comparisons of the slopes. What can we conclude about differences between the slopes for the Age groups? Briefly justify your response.

ANSWER:

$emmeans

```
 Age.f emmean   SE df lower.CL upper.CL
 1       28.2 1.83 39     24.5     31.9
 2       23.9 1.84 39     20.2     27.7
 3       31.7 2.02 39     27.6     35.8

Confidence level used: 0.95

$contrasts
 contrast estimate   SE df t.ratio p.value
 1 - 2        4.30 2.59 39  1.662  0.2327
 1 - 3       -3.46 2.72 39 -1.272  0.4192
 2 - 3       -7.77 2.73 39 -2.843  0.0189

P value adjustment: tukey method for comparing a family of 3 estimates
```

The p values are not significant, which implies that there is not much variation amongst the three Age groups.

6. Calculate emmeans for the Age groups at (A) Ppt = 2500 and (B) Ppt = 3000.

ANSWER:

(A)

```
$emmeans
 Age.f emmean   SE df lower.CL upper.CL
 1       28.2 1.83 39     24.5     31.9
 2       23.9 1.84 39     20.2     27.7
 3       31.7 2.02 39     27.6     35.8

Confidence level used: 0.95

$contrasts
 contrast estimate   SE df t.ratio p.value
 1 - 2        4.30 2.59 39  1.662  0.2327
 1 - 3       -3.46 2.72 39 -1.272  0.4192
 2 - 3       -7.77 2.73 39 -2.843  0.0189

P value adjustment: tukey method for comparing a family of 3 estimates
```

(B)

```
$emmeans
 Age.f emmean   SE df lower.CL upper.CL
 1       28.2 1.83 39     24.5     31.9
 2       23.9 1.84 39     20.2     27.7
 3       31.7 2.02 39     27.6     35.8

Confidence level used: 0.95

$contrasts
 contrast estimate   SE df t.ratio p.value
 1 - 2        4.30 2.59 39  1.662  0.2327
 1 - 3       -3.46 2.72 39 -1.272  0.4192
 2 - 3       -7.77 2.73 39 -2.843  0.0189

P value adjustment: tukey method for comparing a family of 3 estimates
```

7. Based on a backwards elimination approach, which model is preferred? Briefly justify your response. Use α = 0.05.

ANSWER:

From observations, the model with the interaction term does not appear to have significant p values. The model is still not significant even with Age, therefore, we shall choose the model with Ppt only.

8. Based on AIC criteria, which model is preferred? Briefly justify your response.
Hint: Use dredge() from MuMIn. Use code something like this:

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

ANSWER:

We clearly prefer the model with the lowest AIC. Looking at the output from the dredge function, the model with both predictors age and factor, without the interaction term appears to be the best.

9. Choose a model using "backwards elimination" (hypothesis testing) approach. Use α = 0.05. No need to discuss, just state your final model.

ANSWER:

The final model is: FisherAlpha ~ Ppt

10. Choose a model using "forward selection" (hypothesis testing) approach. Use α = 0.05. No need to discuss, just state your final model.

ANSWER:

The final model is: FisherAlpha ~ Age + Ppt (no interaction)

11. Choose a model using AICC. Hint: Use dredge() from MuMIn. No need to discuss, just state your final model.
Use code something like this:

```
library(MuMIn)
```

```
FullModel <- lm(BodyFat ~ ., data = InData)
options(na.action = "na.fail")
dredge(FullModel)
```

ANSWER:

The final model is: FisherAlpha ~ Age + Ppt (no interaction).