

# Homework 8 - Stats 230 (Two-way ANOVA with Interaction and Randomization of F tests : (Happy Faces, Swahili & Fruit Flies)

Dhyey Mavani

date

Due to exam proximity, no late submissions are allowed.

**PROBLEMS TO TURN IN: #7.32, #7.34ab, #8.34 (interpret output with code given),**

## Exercise 7.32

7.32 part a:

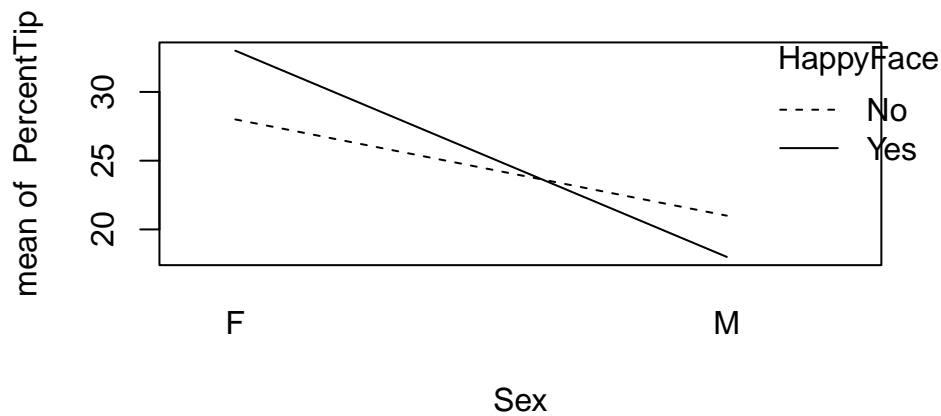
The two factors of interest are Face and Gender. I think Gender is Observational factor and Face is experimental factor. There are two levels to Gender, namely “M” and “F” signifying Male and Female respectively. There are two levels to Face, namely “Yes” and “No” signifying whether or not a happy face was drawn at the back of customers’ checks respectively.

7.32 part b:

SOLUTION: I think there is evidence of interaction between these two variables because we can see that the levels are different, or the so-called slopes of the connecting lines are different in the case of HappyFace variable being “Yes” and “No”. The interaction we can see here is the difference in differences form. In words, we can say that if a Happy face is drawn on the back of customers’ checks, then the mean percent tip difference between Female and Male is more than the case when the happy face is not drawn on the back of customers’ checks.

There is evidence of interaction because the slopes are very different. This means that the differences in mean percent tip between the levels of HappyFace for each level of sex are different. Moreover, the differences in mean percent tip between the levels of sex for each level of HappyFace are different. Also, the lines cross over one another, indicating that the interaction is disordinal, meaning that percent tip is completely dependent on HappyFace.

```
# Code for the plot is provided.  
# This is not ggformula syntax because of how the data was provided - just as summary counts  
Sex <- c("M", "M", "F", "F")  
HappyFace <- c("Yes", "No", "Yes", "No")  
PercentTip <- c(18, 21, 33, 28)  
interaction.plot(Sex, HappyFace, PercentTip)
```



```
data(Swahili)
```

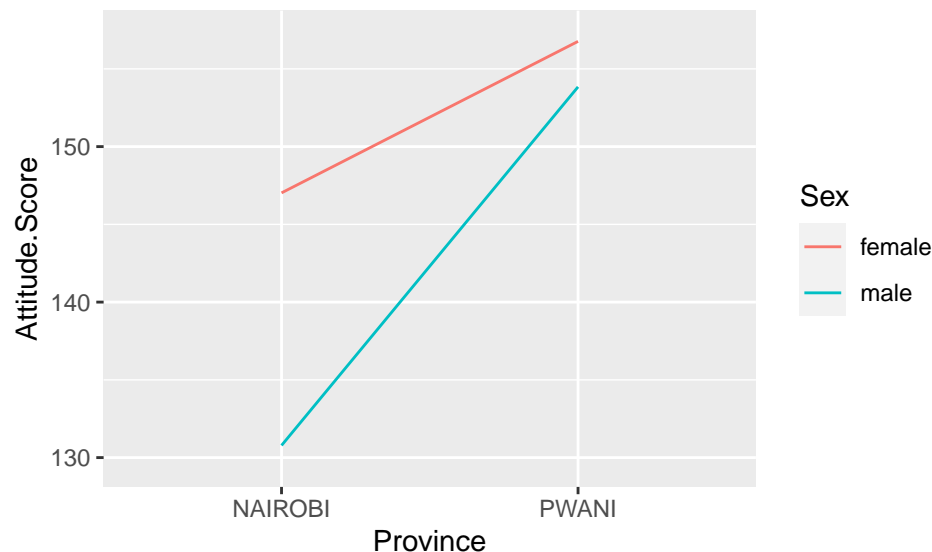
### Exercise 7.34ab

7.34 part a:

SOLUTION: There is evidence of interaction because the slopes are different. The interaction term between province and sex is significant, with a p-value of the F-test being  $1.66e-13 < 0.05$ .

From the figure it is apparent that females in PWANI have higher Attitude.Score than females in NAIROBI. Also, we can see that males in PWANI have higher Attitude.Score than males in NAIROBI. Moreover we can say that the difference in Attitude Scores of males between Provinces PWANI and NAIROBI is significantly more than the difference in Attitude Scores of females between Provinces PWANI and NAIROBI. This tells us that there is interaction between province and sex.

```
gf_line(Attitude.Score ~ Province, color = ~ Sex, data = Swahili, group = ~ Sex,
        stat = "summary", fun.y = "mean")
```



7.34 part b:

SOLUTION:

We have equal variance of errors because we can see the mostly equal spread of points around the x-axis in the residuals vs fitted plot.

Normality of errors is almost satisfied because we can see the qq plot with some of the points not falling along the line, but I don't see that alarming of a concern.

We can see that there is slight heteroskedasticity as the spread of points on left is slightly greater than that in right.

Finally, by the construction of the experiment, the samples are independent, and we can say that the treatment groups all have the same sample size so there is nothing to worry about there.

Hence, all conditions for using the ANOVA model are satisfied. I would, however, consider transforming the response variable to try to eliminate the pattern in the points in the normal qq plot.

```
mod2 <- lm(Attitude.Score ~ Province*Sex, data = Swahili)      #added interaction term by using *
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: Attitude.Score
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Province   1  32275   32275   349.82 < 2e-16 ***
## Sex         1  11021   11021   119.45 < 2e-16 ***
## Province:Sex 1    5320     5320    57.66 1.66e-13 ***
## Residuals 476  43917         92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aovmod2 <- aov(Attitude.Score ~ Province*Sex, data = Swahili)
summary(aovmod2)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Province      1  32275    32275    349.8 < 2e-16 ***
## Sex            1  11021    11021    119.5 < 2e-16 ***
## Province:Sex   1   5320     5320     57.7 1.7e-13 ***
## Residuals    476  43917         92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.tables(aovmod2)
```

```
## Tables of effects
##
## Province
## Province
## NAIROBI    PWANI
##    -8.2      8.2
##
## Sex
## Sex
## female    male
##  4.792 -4.792
##
## Province:Sex
##           Sex
## Province female male
##   NAIROBI  3.329 -3.329
##   PWANI   -3.329  3.329
```

```
mplot(mod2, which = 1)
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 130.65
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 23.188
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 1.4845e-15
```

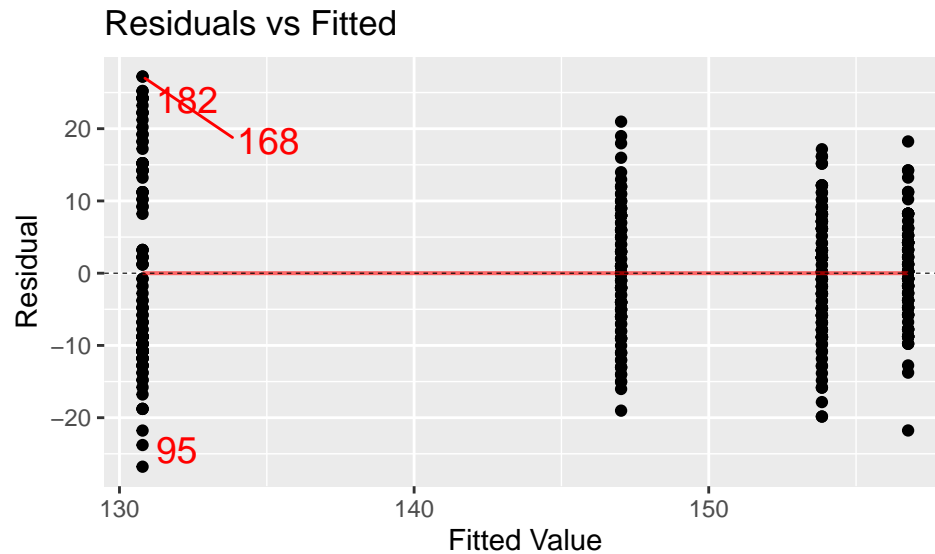
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 97.448
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object))), : pseudoinverse used at
## 130.65
```

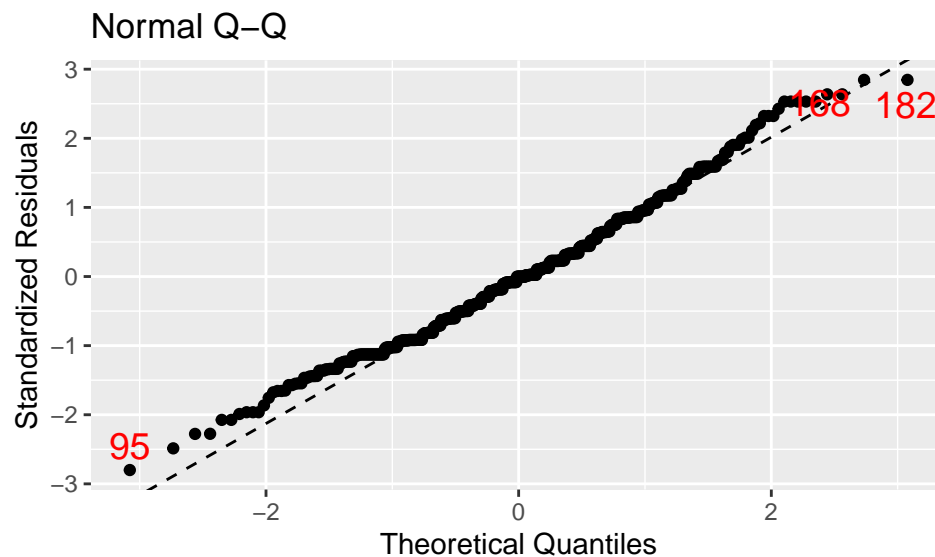
```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object))), : neighborhood radius
## 23.188
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object))), : reciprocal condition
## number 1.4845e-15
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object))), : There are other near
## singularities as well. 97.448
```



```
mplot(mod2, which = 2)
```



```
data(FruitFlies)
```

**Exercise 8.34** This next part is not required but to make sure we perform the correct permutation/randomization test, we want to be sure we've got the correct original statistic from the original model. This isn't strictly necessary, as the problem gave us the F value and we could just use that instead.

Note there IS a typo there, where they say 13.61 first and then 13.62 later. It's really 13.612.

```
anova(lm(Longevity ~ Treatment, data = FruitFlies))$"F value"
```

```
## [1] 13.612      NA
```

This is the same 13.61 described in the text, so we have the correct starting model.

The text requests doing the randomization 10000 times, so we set that up in the next code chunk.

```
set.seed(230) #make the results reproducible
t <- do(10000) * (anova(lm(shuffle(Longevity) ~ Treatment, data = FruitFlies))$"F value")
```

Now that we've performed the procedure, we can answer the questions that the text asks. We don't need a plot of the randomization distribution for this, but it's included it for reference.

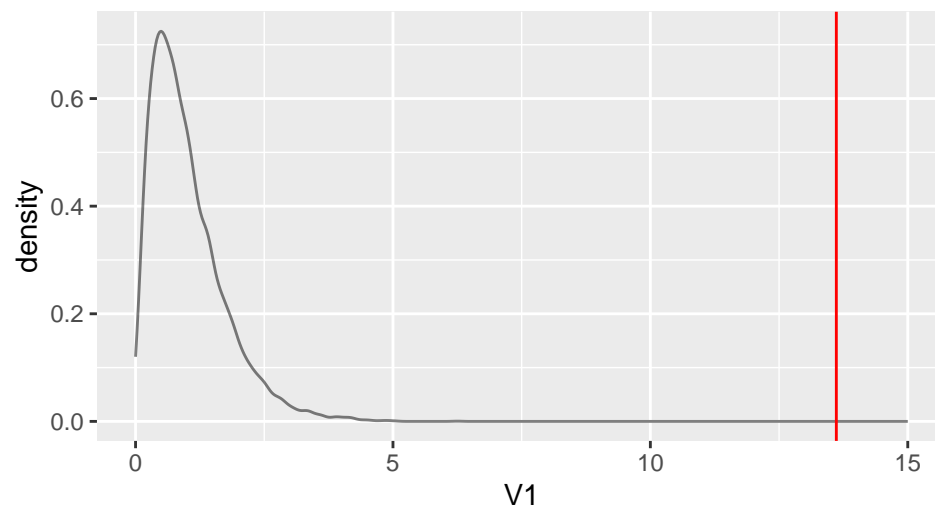
```
t <- as.data.frame(t)
favstats(~ V1, data = t)
```

```
##           min           Q1    median           Q3           max           mean           sd           n missing
## 0.00633813 0.476547 0.836466 1.36564 6.25308 1.00616 0.717658 10000              0
```

```
gf_dens(~ V1, data = t) %>%
  gf_lims(x = c(0, 15)) %>%
  gf_vline(xintercept = 13.612, color = "red") %>%
  gf_labs(title = "Randomization Distribution for F from One-Way ANOVA")
```

```
## Warning: geom_vline(): Ignoring 'mapping' because 'xintercept' was provided.
```

## Randomization Distribution for F from One-Way ANOVA



```
pdata(~ V1, 13.612, data = t, lower.tail = FALSE)
```

```
## [1] 0
```

SOLUTION:

None of my randomization F-statistics were greater than 13.62, and the biggest value I got for the F-statistic was 6.25308.