ps11.R

benjamin

2020-12-01

```
# Ben Reichert
# Problem Set 11
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.0     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

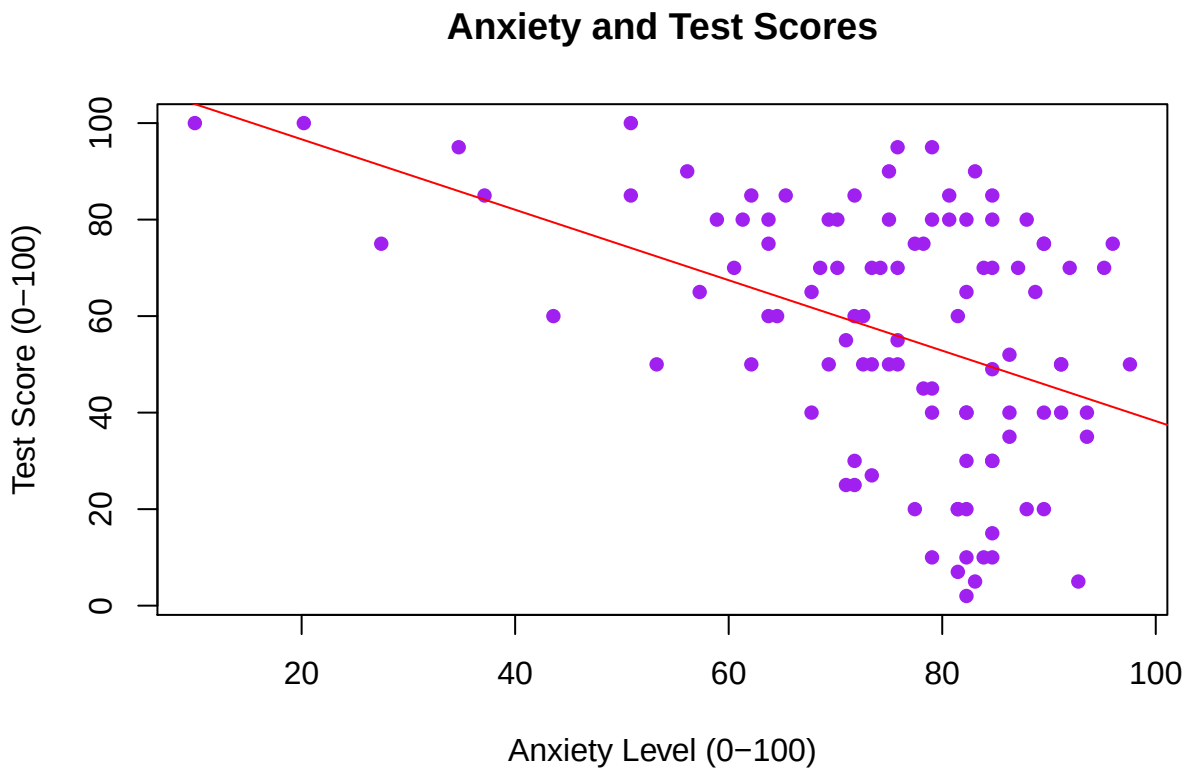
```
library(moderndive)
### 1
angst <- read.table("examanxiety.txt", header = TRUE, sep = "", dec = ".")

# a)
wom <- subset(angst, Gender == 'Female')
men <- subset(angst, Gender == 'Male')
print(t.test(wom[,4],men[,4]))
```

```
##
##  Welch Two Sample t-test
##
## data:  wom[, 4] and men[, 4]
## t = 0.32961, df = 100.41, p-value = 0.7424
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.110827  7.147444
## sample estimates:
## mean of x mean of y
##  75.40204  74.38373
```

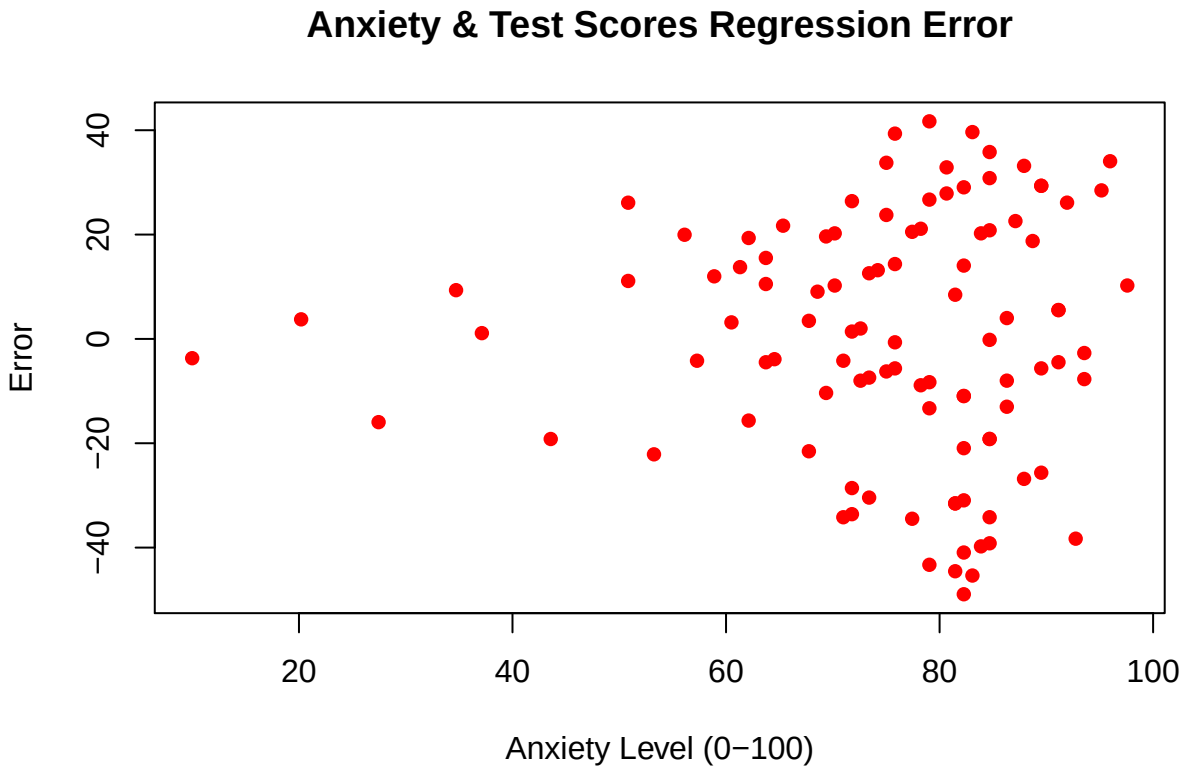
```
# Null: gender has no impact on anxiety levels.
# Alternative: gender has an impact on anxiety levels.
# Because our p-value: 0.7424 > 0.05, we fail to reject the null hypothesis that gender has no impact o

# b)
x <- angst[,4]
y <- angst[,3]
anxiety <- lm(y ~ x, data = angst)
print(get_regression_table(anxiety))
```

```
## # A tibble: 2 x 7
##   term      estimate std_error statistic p_value lower_ci upper_ci
##   <chr>        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept   111.      11.4        9.80       0    88.7   134.
## 2 x            -0.73     0.148     -4.92       0    -1.02   -0.436
```

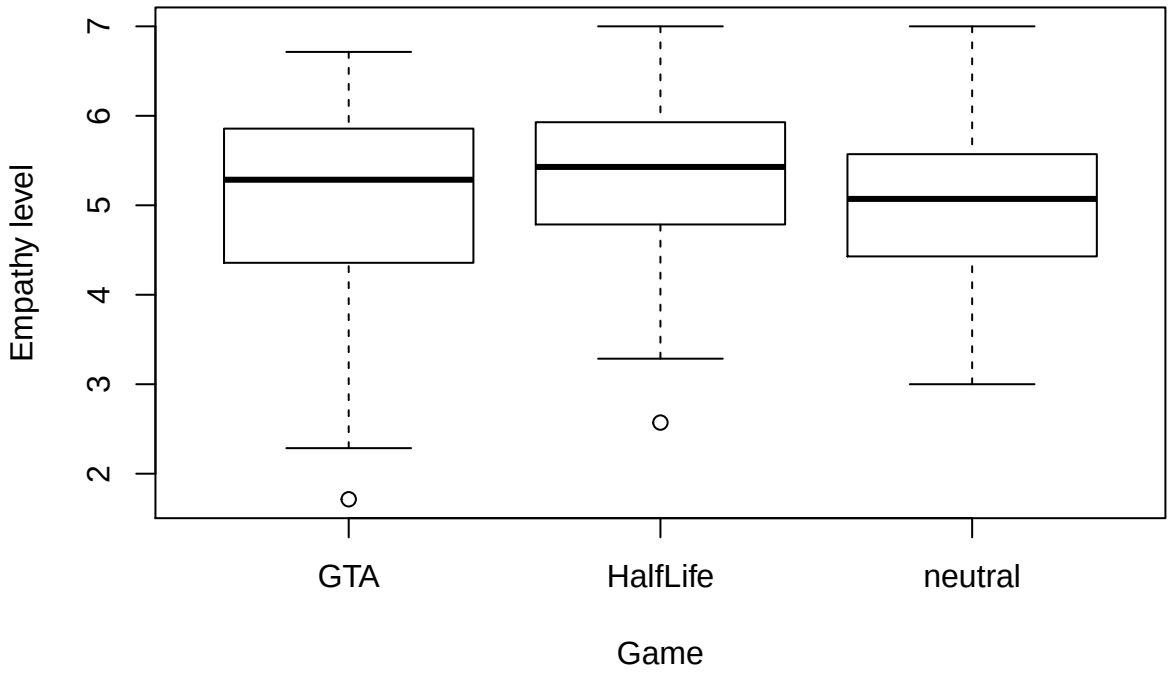
```
core <- cor(x,y)
b <- core * sd(y) / sd(x)
a <- mean(y) - b * mean(x)
# The regression formula is y = -0.73x + 111, where x is anxiety level and y is exam score.
# This means that for every anxiety level your exam score goes down 0.73.

# c)
plot(x, y, pch = 16, cex = 1, col = "purple", main = "Anxiety and Test Scores", xlab = "Anxiety Level (
abline(a, b, col="red")
```




```
# i) Linearity is met because the trend of the graph follows a linear pattern, which is that as anxiety
# ii) Independence is not met because it seems that the two variables impact the outcome of the other.
err <- rep(0,length(x))
for(i in 1:length(x)) {
  err[i] <- y[i] - (-.73*x[i]+111)
}
plot(x, err, pch = 16, cex = 1, col = "red", main = "Anxiety & Test Scores Regression Error", xlab = "A
```




```r
# iii) Equal variance of errors looks like it is met according to the graph of error along the regressi
# equal, but the variances trend in the same direction.
# iiii) Normality of errors is not met because the error along the regression line graph does not look

### 2
emp <- read.table("GameEmpathy.txt", header = TRUE, sep = "", dec = ".")

neut <- subset(emp, game.type == "neutral")
gta <- subset(emp, game.type == "GTA")
half <- subset(emp, game.type == "HalfLife")

# i) neutral
m_neut <- ave(neut[,4])
mn <- m_neut[1]

# ii) half-life
m_half <- ave(half[,4])
mh <- m_half[1]

# iii) GTA
m_gta <- ave(gta[,4])
mg <- m_gta[1]

gt <- emp[,2]
em <- emp[,4]
boxplot(em~gt,data=emp, main="Game and Empathy Levels",
        xlab="Game", ylab="Empathy level")
```




```
# Referencing the boxplot, it appears that the GTA game data has the highest variance yet a higher aver
# the neutral game. The half-life game data had the highest mean and lowest variance.
anova = aov(em~gt,data=emp) # analysis of variance
summary(anova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## gt            2   2.25   1.125   1.092  0.338
## Residuals   150 154.47   1.030
```

```
# The p-value of the variance analysis: 0 < 0.05 so we conclude that the means of the variances are sta
# leading us to conclude that the type of game has an impact on average empathy.

# b)
# i) neutral
print(t.test(neut[,4],neut[,3]))
```

```
##
##  Welch Two Sample t-test
##
## data:  neut[, 4] and neut[, 3]
## t = 4.5068, df = 80.322, p-value = 2.212e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.6592428 1.7017096
## sample estimates:
## mean of x mean of y
##  5.053810  3.873333
```

```
# Null: there is no significant difference between
# Alternative: gender has an impact on anxiety levels.
# Because our p-value: 0.7424 > 0.05, we fail to reject the null hypothesis that gender has no impact o
```

```r
# ii) half-life
print(t.test(half[,4],half[,3]))
```

```
##
##  Welch Two Sample t-test
##
## data:  half[, 4] and half[, 3]
## t = 2.5966, df = 100.86, p-value = 0.01082
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1258764 0.9407903
## sample estimates:
## mean of x mean of y
##  5.293939  4.760606
```

```r
# Null: there is no significant impact of gender on anxiety levels
# Alternative: gender has an impact on anxiety levels
# Because our p-value: 0.01 < 0.05, we reject the null hypothesis and assert the alternative that, for
# gender has an impact on anxiety levels.

# iii) GTA
print(t.test(gta[,4],gta[,3]))
```

```
##
##  Welch Two Sample t-test
##
## data:  gta[, 4] and gta[, 3]
## t = 3.7807, df = 93.967, p-value = 0.0002747
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4312498 1.3852185
## sample estimates:
## mean of x mean of y
##  5.029762  4.121528
```

```r
# Null: there is no significant impact that gender has on anxiety levels
# Alternative: gender has an impact on anxiety levels
# Because p-value: 0.0003 < 0.05, we reject the null hypothesis and take the alternative that, for GTA
# gender definitively has an impact on anxiety levels.
```