STAT 511: HW #3 Q:3 & 4

Rumil Legaspi

22 February 2021

Contents

Workspace Setup

```
setwd("C:/Users/RUMIL/Desktop/APU/STAT 511 - Millie Mao (Applied Regression Analysis)/week 4/Week 4")
#loading gpa data
gpa_data = read.table(file = "GPA.txt", header = FALSE, sep = "")

#Adding headers
names(gpa_data) <- c("GPA", "ACT")

#Defining dependent and independent vars
ACT = gpa_data$ACT #X
GPA = gpa_data$GPA #Y

gpa_lm = lm(GPA ~ ACT, data = gpa_data)
```

```r
#loading muscle data
muscle_data = read.table(file = "Muscle.txt", header = FALSE, sep = "")

#Adding headers
names(muscle_data) <- c("Muscle", "Age")

#Defining dependent and independent vars
Age = muscle_data$Age #X
Muscle = muscle_data$Muscle #Y

#creating our linear model
muscle_lm = lm(Muscle ~ Age, data = muscle_data)
```

3. Refer to the GPA problem in HW #1

(a). Compute the Pearson correlation coefficient and attach the appropriate sign.

```r
#Computing the sample correlation of two variables
#Choose any continuous variable, order does not matter
cor.test(ACT, GPA, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  ACT and GPA
## t = 3.0398, df = 118, p-value = 0.002917
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.09482051 0.42804747
## sample estimates:
##       cor
## 0.2694818
```

The result of our Pearson correlation coefficient is a positive 0.269.

$r = +0.27$

(b). Obtain Spearman rank correlation coefficient.

```r
#spearman correlation
cor.test(ACT, GPA, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  ACT and GPA
## S = 197904, p-value = 0.0005048
```

```
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.3127847
```

The result of our Spearman correlation coefficient is 0.31.

$r_s = 0.31$

(c). Which correlation coefficient is stronger? Why?

The Spearman correlation coefficient is stronger compared to the Pearson correlation between the variables ACT(X) and GPA(Y). This is because Spearman measures the data **points by ranks**. In other words, ACT and GPA are better suited because they are both **ordinal** types of data indicating a kind of ranking system, whereas Pearson does not.

---

4. Refer to the Muscle Mass problem in HW#1.

(a). Compute the Pearson product-moment correlation coefficient.

```r
#Age is x, Muscle is Y. But still placement order does not matter
cor.test(Age, Muscle, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  Age and Muscle
## t = -13.193, df = 58, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9180874 -0.7847085
## sample estimates:
##       cor
## -0.866064
```

The result of our Pearson correlation coefficient is a negative 0.86 indicating a fairly strong negative linear relationship between our variables, age and muscle.

$r = -0.86$

(b). Based on Part (a), test whether muscle mass and age are significantly correlated in the population at alpha = 0.05. State the null and alternative hypotheses, decision rule, and conclusion.

- 1. **Pearson Correlation Hypothesis Testing Stating Our Hypotheses:**
  - Null hypothesis: H_0: $\rho = 0$ (There is no statistically significant linear correlation between Age and Muscle)

- Alternative hypothesis: H_1: $\rho \neq 0$ (There exists a statistically significant linear correlation between Age and Muscle)

```r
summary(muscle_lm)
```

```
##
## Call:
## lm(formula = Muscle ~ Age, data = muscle_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.1368  -6.1968  -0.5969   6.7607  23.4731
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 156.3466     5.5123   28.36   <2e-16 ***
## Age          -1.1900     0.0902  -13.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.173 on 58 degrees of freedom
## Multiple R-squared:  0.7501, Adjusted R-squared:  0.7458
## F-statistic: 174.1 on 1 and 58 DF,  p-value: < 2.2e-16
```

```r
#Helps us find our T value
cor.test(Age, Muscle, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  Age and Muscle
## t = -13.193, df = 58, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9180874 -0.7847085
## sample estimates:
##       cor
## -0.866064
```

```r
#Gives us our critical value so we can compare with T statistic
qt(0.975, 58)
```

```
## [1] 2.001717
```

- 2. Calculating t statistic:
  - $t = -13.193$
  - Critical value = 2.00
- 3. Comparing with critical value and p value
  - $|t = -13.193| > 2.00$

    - p value 2.2e-16 < α (0.05)

- 4. Draw conclusion
    - Also, since the p value is very small and we can reject the NULL hypothesis that there is no linear correlation between age and muscle.
    - Therefore we can conclude with the alternative hypothesis that age and muscle are linearly correlated.

(c). Compute the Spearman rank correlation coefficient.

```r
#Age is x, Muscle is Y. But still placement order does not matter
cor.test(Age, Muscle, method = "spearman")
```

```
## Warning in cor.test.default(Age, Muscle, method = "spearman"): Cannot compute
## exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
##
## data:  Age and Muscle
## S = 67147, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.8657217
```

(d). Repeat the test in Part (b) using the Spearman rank correlation from Part (c).

- 1. **Spearman Correlation Hypothesis Testing Stating our Hypotheses:**
    - Null hypothesis: H_0: $\rho = 0$ (There is no statistically significant rank correlation between Age and Muscle)
    - Alternative hypothesis: H_1: $\rho \neq 0$ (There exists a statistically significant rank correlation between Age and Muscle)

```r
#Helps us find our T value
cor.test(Age, Muscle, method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  Age and Muscle
## S = 67147, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.8657217
```

```r
#Gives us our critical value so we can compare with T statistic
qt(0.975, 58)
```

```
## [1] 2.001717
```

- 2. Calculating t statistic:
  - $t = -13.193$
  - Critical value = 2.00
- 3. Comparing with critical value and p value
  - $|t = -13.193| > 2.00$
  - p value 2.2e-16 $< \alpha$ (0.05)
- 4. Draw conclusion
  - Similarly, since the p value is very small we can reject the NULL hypothesis that there is no significant rank correlation between age and muscle.
  - Therefore we can conclude with the alternative hypothesis that there is a significant rank correlation between age and muscle.

(e). How do your correlation coefficient estimates and test conclusions in Parts (a) and (b) compare to those obtained in Parts (c) and (d)?

Based on the results from our hypothesis tests on both Pearson and Spearman, both age and muscle hold a significant rank and linear correlation relationship. These two hypothesis help verify our correlation coefficient estimates and allow us to safely conclude that a correlation does exist.