Alexandrea Stylianou
Stats 543
Final

Grade sheet for Final Exam Problem Two Part 1:
Pearson's Correlation Coefficient

Part 1: Ignoring exposure status, is there a relationship between the Day 1 and Day 15 body weight measures? What is the strength and direction of the relationship? Is it possible to illustrate the relationship graphically? If yes, show how

---

<u>Topic:</u>
Previous studies have shown a link between pesticide exposure and thyroid disease, which can lead to increased weight gain, a symptom of thyroid disease.

<u>Population:</u>
Rats exposed to a certain pesticide

<u>Research Question:</u>
Whether or not there is a strong association between body weights measurements of day 1 and day 15.

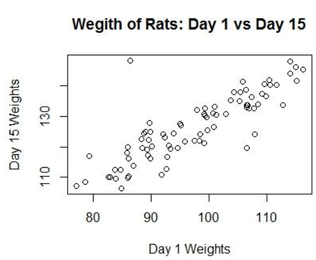
Methods

<u>Description of Outcome:</u> The weights of the rats day on 1 and day 15

<u>Description of Data Summary for Each Variable:</u>
110 rats were a part of a research study to test the effects of a certain pesticide on weight gain. The rats were randomly selected, and divide among two groups: rats that either receive low exposure of this pesticide or high exposure of this pesticide. The trial lasted for 15 days and the weights of the rats were recorded on day 1 and day 15.

<u>Verification of normality:</u>
Per the central limit theorem, if the sample size is above 20 the sample mean will follow a normal distribution.

<u>Verification of Linear Association:</u>
We will verify that the variables have a linear association by looking at the scatter plot of the weights on day 1 and day 15. Scatter plot below:



<u>Statement of Null Hypothesis:</u>
There is no association between day 1 of trial and day 15 of trial of body weight measurements (H_o: p=0)

<u>Statement of Alternative Hypothesis:</u>
There is an association between day 1 of trial and day 15 of trial of body weight measurements (H_o: p≠0)

Statistical Method for testing Association:
Pearson's correlation coefficient

Decision Rule:
Reject H_O in favor of H_A if p-value is less than alpha (p-value<α) otherwise fail to reject the null H_o

Method of Computation:
R statistical software version 2.11.1

Significance Level:
α=0.05

Results

Data Summary for Day 1 weights and Day 15 Weights included in table 1.0
Table 1.0 Summary table of Weights of Rats on Day 1 and Day 15

Group	*n*	mean	SD	95% CI
Day 1 Weights	110	97.54	10.22	95.6088, 99.4721
Day 15 Weights	110	127.25	10.51	125.2620, 129.2353

Normality is assumed per the CLT based on adequate sample size for each group

Assumptions:
Sample is representative
Sample is large enough
Sample measurements for each subject are independent of one another
Measurements of each group are independent of each other
Linear relationship
Heteroskedasticity

Test Results:
r: 0.76 | p-value: <0.0001

P-value Results
Since p-value < 0.05, we reject the H_O in favor of H_A

Description of Association:
The correlation coefficient for weights is 0.76, thus the relationship of body weight measurements on day 1 and day 15 of the sample is significant and positively associated.

Discussion

Evidence suggests, that on day 1 and day 15 of the trial there is a significant and positive association between body weight measurements of the rats.

Implication of Results

Moving forward, researchers should closely monitor the weight of the rats before, during, and after testing to ensure a high quality of results.

R-code

```
T4_Problem2<- read.csv(file.choose(),header=TRUE)
Et<-T4_Problem2[T4_Problem2$Exposure=='low' & T4_Problem2$BWDay1,1]
Ft<-T4_Problem2[T4_Problem2$Exposure=='low' & T4_Problem2$BWDay15,2]
Gt<-T4_Problem2[T4_Problem2$Exposure=='high' & T4_Problem2$BWDay1,1]
Ht<-T4_Problem2[T4_Problem2$Exposure=='high' & T4_Problem2$BWDay15,2]

Day1_lo<-c(Et) #49
Day15_lo<-c(Ft) #49
Day1_hi<-c(Gt) #61
Day15_hi<-c(Ht)#61

just_day_1<-T4_Problem2[T4_Problem2$BWDay1,1]
mean_just_day_1<-mean(just_day_1)

qqPlot((Day1_lo))
qqPlot(Day15_lo)
#low
MDay1_lo<-mean(Day1_lo) #98.01
MDay15_lo<-mean(Day15_lo) #126.57
#hi
MDay15_hi<-mean(Day15_hi) #127.78
MDay1_hi<-mean(Day1_hi) #97.15

#question one correlation. #Ignoring amount of radiation, is there a relationship between
#weight of day one, and day 15.
#what is strength and direction?
#per scatter plot, there is a positive association in weights on day 1 vs day 2
#null hypothesis, there is no relationship(H0: p=0). alternative hypothesis, there is a relationship(H0: p/=0)
#assumptions:
#sample is represenative.subjects are indepedntly measured, relationship is linear, sample size is above 20
#variabiity of one measyre to be more/less contast throughout the values of the other measure and vice/versa
(heteroskedasticity)
##################################scatter plot question 1##########################
plot(T4_Problem2$BWDay1, T4_Problem2$BWDay15,
    xlab="Day 1 Weights", ylab="Day 15 Weights",
    main = "Wegith of Rats: Day 1 vs Day 15")
######################################standard deviation##########################
#summerize the two measurements with mean sizes, means, SD 95CI
ddply(T4_Problem2, .(Exposure), summarize, BWDay1=sd(BWDay1))
#Exposure    BWDay1
#     high 10.777890
#      low  9.573986
ddply(T4_Problem2, .(Exposure), summarize, BWDay15=sd(BWDay15))
#Exposure  BWDay15
#     high 10.71949
#    low 10.32063
#######################################CI and Mean#################################
CI(Day1_lo, ci=0.95)
#upper      mean      lower
#100.76221  98.01224  95.26228
```

```
CI(Day15_lo, ci=0.95)
#upper      mean    lower
#129.5434 126.5790 123.6145
CI(Day15_hi, ci=0.95)
#upper      mean    lower
#130.5319 127.7866 125.0412
CI(Day1_hi, ci=0.95)
#upper      mean    lower
#99.92182 97.16148 94.40113

CI(T4_Problem2$BWDay1, ci=0.95)
#upper      mean    lower
#99.47213 97.54045 95.60878

CI(T4_Problem2$BWDay15, ci=0.95)
#upper      mean    lower
#129.2353 127.2486 125.2620

sd(T4_Problem2$BWDay15)
sd(T4_Problem2$BWDay1)

####################################pearsons's test###########################
cor.test(T4_Problem2$BWDay1, T4_Problem2$BWDay15,
         alternative = "two.sided",
         method = "pearson",
         conf.level = 0.95)

#Pearson's product-moment correlation

#data:  T4_Problem2$BWDay1 and T4_Problem2$BWDay15
#t = 12.25, df = 108, p-value < 2.2e-16
#alternative hypothesis: true correlation is not equal to 0
#95 percent confidence interval:
#0.6711525 0.8311335
#sample estimates:
#cor
#0.7625644

#r value is positive number close to 1, therefore there there is a strong relationship
```