CLPS900 FINAL

B01537135

2022-12-09

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
study <- read.csv("https://raw.githubusercontent.com/SCosta352/CLPS0900/main/workstudy.csv")
```

QUESTION 1

```
study <- study%>%
mutate(Apology=ifelse(apology==2,0,1))
```

```
study <- study%>%
mutate(Comp1=ifelse(comp1==2,0,1))
```

```
study <- study%>%
mutate(Excuse=ifelse(excuse==2,0,1))
```

```
study <- study%>%
mutate(Comp2=ifelse(comp2==2,0,1))
```

```
study <- study%>%
mutate(Comp3=ifelse(comp3==2,0,1))
```

dataframe with recoded observations (removed previously coded columns with new ones)

```
study1 <- study%>%
select(X, id, sex, age, rel.status1, intimacy, r.value, other.responsibility, Apology, Comp1, Excuse, C
```

```
study1 <- study1%>%
  mutate(Apology_Score=Apology+Comp1+Excuse+Comp2+Comp3)
```

QUESTION 2

```
study2 <- study1 %>%
  filter(!is.na(rel.status1))
```

Is there a significant difference in apology scores by relationship status?

if mu is mean apology score,

H0: mu(higher status) = mu(equal status) = mu(lower status)

H1: mu(i) != mu(k), for at least one pair of i, j incl. (1, 2, 3), i != k

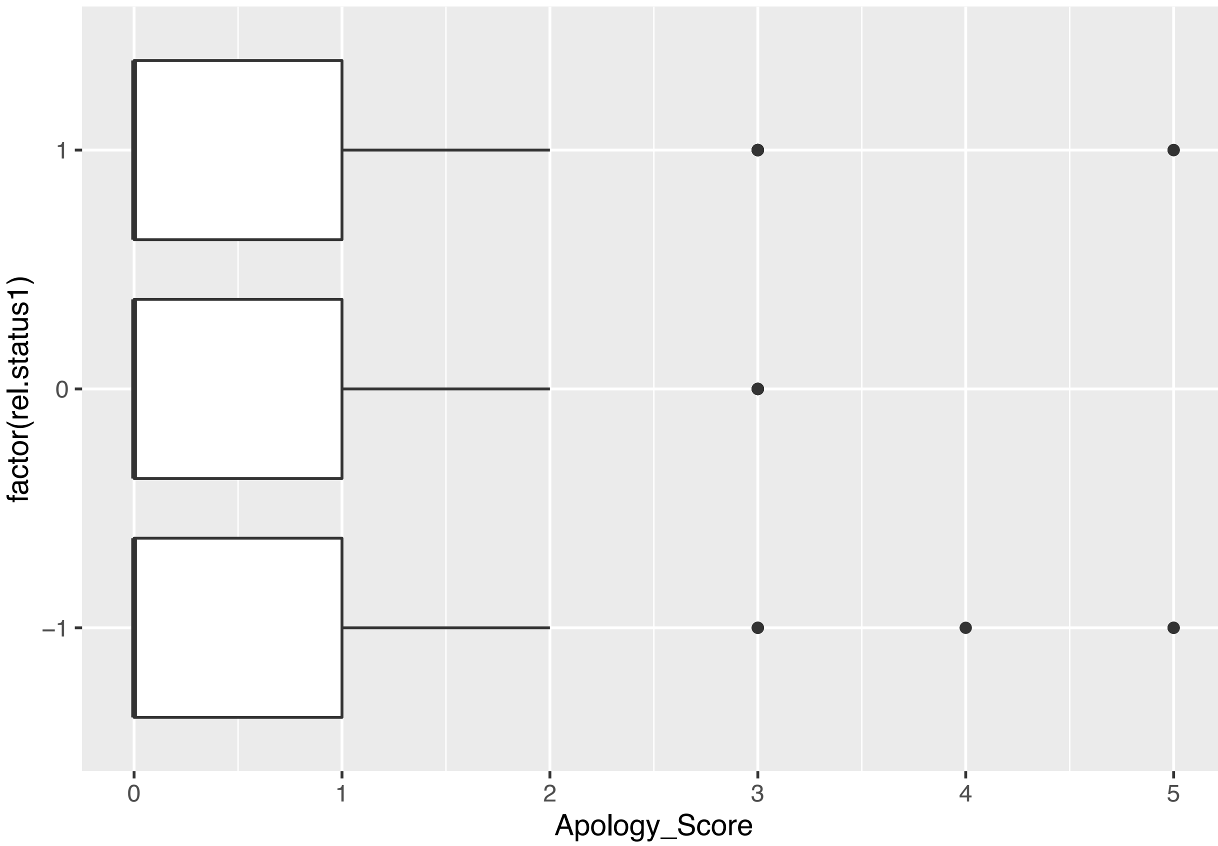
n > 30, can assume normality

```
sum_study2 <- study2 %>%
  group_by(rel.status1) %>%
  summarise(mean=mean(Apology_Score), variance = var(Apology_Score), count=n())
sum_study2
```

```
## # A tibble: 3 x 4
##   rel.status1  mean variance count
##         <int> <dbl>    <dbl> <int>
## 1          -1 0.701    1.10     87
## 2           0 0.646    0.692   144
## 3           1 0.788    0.959   226
```

conditions for variance and n>30 met.

```
ggplot(study2)+
geom_boxplot(aes(Apology_Score,factor(rel.status1)))
```



not very helpful because such small apology score values.

```
study2_model <- lm(Apology_Score ~ factor(rel.status1), study2)
anova(study2_model)
```

```
## Analysis of Variance Table
##
## Response: Apology_Score
##                      Df Sum Sq Mean Sq F value Pr(>F)
## factor(rel.status1)   2   1.84 0.91847  1.0196 0.3616
## Residuals           454 408.97 0.90082
```

since p = 0.3616 > alpha = 0.05, we fail to reject the null hypothesis. At alpha = 0.05, there is not enough evidence to conclude that the mean apology score differs significantly based on relationship status.

QUESTION 3

```
study3 <- study2%>%
select(age, intimacy, r.value, other.responsibility, Apology_Score)
```

3a

```
cor(study3)
```

```
##                                 age    intimacy     r.value other.responsibility
## age                      1.00000000  0.04986217 -0.07501437           0.04010566
## intimacy                 0.04986217  1.00000000  0.15257892          -0.05044917
## r.value                 -0.07501437  0.15257892  1.00000000          -0.14039335
## other.responsibility     0.04010566 -0.05044917 -0.14039335           1.00000000
## Apology_Score           -0.08963830  0.14133076  0.20927744          -0.29645681
##                         Apology_Score
## age                        -0.0896383
## intimacy                    0.1413308
## r.value                     0.2092774
## other.responsibility       -0.2964568
## Apology_Score               1.0000000
```

apology score is most strongly correlated with other.responsibility and r.value. These are a measure of the opponent's responsibility and, a measure of the individual's relationship with the opponent.

3b

other.responsibility and Apology_Score have a value of -0.29, which is higher than all the others.

```
ggplot(study3)+
  geom_point(aes(other.responsibility, Apology_Score))+
  labs(title="Association Between Apology Score and Other Responsibility", x= "Other Responsibility (ou
  geom_smooth(aes(other.responsibility,Apology_Score), method="lm",se=FALSE)
```

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


```r
study3model <- lm(other.responsibility~Apology_Score, study3)
study3 <- study3%>%mutate(residual=residuals(study3model))
summary(study3model)
```

```
##
## Call:
## lm(formula = other.responsibility ~ Apology_Score, data = study3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.0088 -1.4845  0.5155  1.9912  5.8020
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     9.0088     0.1375  65.523  < 2e-16 ***
## Apology_Score  -0.7622     0.1151  -6.621 1.01e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.333 on 455 degrees of freedom
## Multiple R-squared:  0.08789,    Adjusted R-squared:  0.08588
## F-statistic: 43.84 on 1 and 455 DF,  p-value: 1.006e-10
```

y(hat) = -0.7622x + 9.0088

or more simply y(hat) = -0.8x + 9

3c

the line of best fit is modeled so that it minimizes the sum of the squares of the residuals in this dataset.

The intercept value of 9.008 represents the fact that when the apology score is 0, it is expected that the allocation of responsibility score is around 9.

the slope means that if the apology score increases by one unit, the other.responsibility measure would go down by about 0.8.

3d

H0: beta(1) = 0

H1: beta(1) != 0

t statistic from table = -6.621

```r
anova(study3model)
```

```
## Analysis of Variance Table
##
## Response: other.responsibility
##                Df  Sum Sq Mean Sq F value    Pr(>F)
## Apology_Score   1  238.64 238.641  43.842 1.006e-10 ***
## Residuals     455 2476.69   5.443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

also checking with f to make sure:

```r
f <- 238.641 / 5.443
```

```r
sqrt(f)
```

```
## [1] 6.621454
```

```r
pf(f, 1, 455, lower.tail = FALSE)
```

```
## [1] 1.004723e-10
```

since p < alpha at 0.05 in this test, we can reject the null and conclude that there is an association between the apology score and the responsibility attributed to the opponent in a conflict.

3e

```
r <- cor(study3$Apology_Score,study3$other.responsibility)
r
```

```
## [1] -0.2964568
```

```
r^2
```

```
## [1] 0.08788664
```

With an R value of 0.08, it can be said that 8% of the variability in the apology scores can be explained by variability in the allocation of other responsibility. This is not very high but is helpful in explaining the data and how the patterns in it arise. Further analysis into the remainder of the varibility would be useful.

QUESTION 4

```
devalue <- read.csv("https://raw.githubusercontent.com/SCosta352/CLPS0900/main/devalue.csv")
```

4a

```
devtable <- devalue %>%
  group_by(Attractiveness, Commitment) %>%
  summarize(Mean=mean(rating), SD=sd(rating))
```

```
## `summarise()` has grouped output by 'Attractiveness'. You can override using
## the `.groups` argument.
```

```
devtable
```

```
## # A tibble: 4 x 4
## # Groups:   Attractiveness [2]
##   Attractiveness  Commitment   Mean    SD
##   <chr>           <chr>       <dbl> <dbl>
## 1 High Attractive High Commit  3.95  2.38
## 2 High Attractive Low Commit   6.92  2.86
## 3 Low Attractive  High Commit  4.73  2.63
## 4 Low Attractive  Low Commit   4.40  2.40
```

4b

```
devalue %>% group_by(Attractiveness) %>%
  summarise(Mean=mean(rating))
```

```
## # A tibble: 2 x 2
##   Attractiveness   Mean
##   <chr>           <dbl>
## 1 High Attractive  5.43
## 2 Low Attractive   4.56
```

```
devalue %>% group_by(Commitment) %>%
  summarise(Mean=mean(rating))
```

```
## # A tibble: 2 x 2
##   Commitment   Mean
##   <chr>       <dbl>
## 1 High Commit  4.34
## 2 Low Commit   5.66
```

```
devmodel <- lm(rating~Attractiveness+Commitment+Attractiveness*Commitment, devalue)
anova(devmodel)
```

```
## Analysis of Variance Table
##
## Response: rating
##                            Df  Sum Sq Mean Sq F value    Pr(>F)
## Attractiveness              1   38.03  38.030  5.7366 0.0175563 *
## Commitment                  1   87.42  87.423 13.1872 0.0003598 ***
## Attractiveness:Commitment   1  136.76 136.762 20.6296 9.723e-06 ***
## Residuals                 196 1299.36   6.629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Main effect of Attractiveness:

if mu is the mean rating,

H0: The mean rating does not differ significantly by attractiveness.

H0: mu(low attractiveness) = mu(high attractiveness)

H1: The mean rating differs signficantly by attractiveness.

H1: mu(low attractiveness) != mu(high attractiveness)

```
F = 38.030 / 6.629
```

```
1 - pf(F, 1, 196)
```

```
## [1] 0.01755335
```

Since p = 0.0176 < alpha at 0.05, we can reject the null hypothesis. This means that mean ratings of -, differ signficantly based on attractiveness. Those in the high attractiveness condition had higher mean ratings than those in the low attractiveness condition.

Main effect of Commitment:

if mu is the mean rating,

H0: The mean mean rating does not differ significantly by commitment status.

H0: mu(low commitment) = mu(high commitment)

H1: The mean mean rating differs significanrly by commitment status.

H1: mu(low commitment) != mu(high commitment)

```
Ff <- 87.423 / 6.629
```

```
1 - pf(Ff, 1, 196)
```

```
## [1] 0.0003596652
```

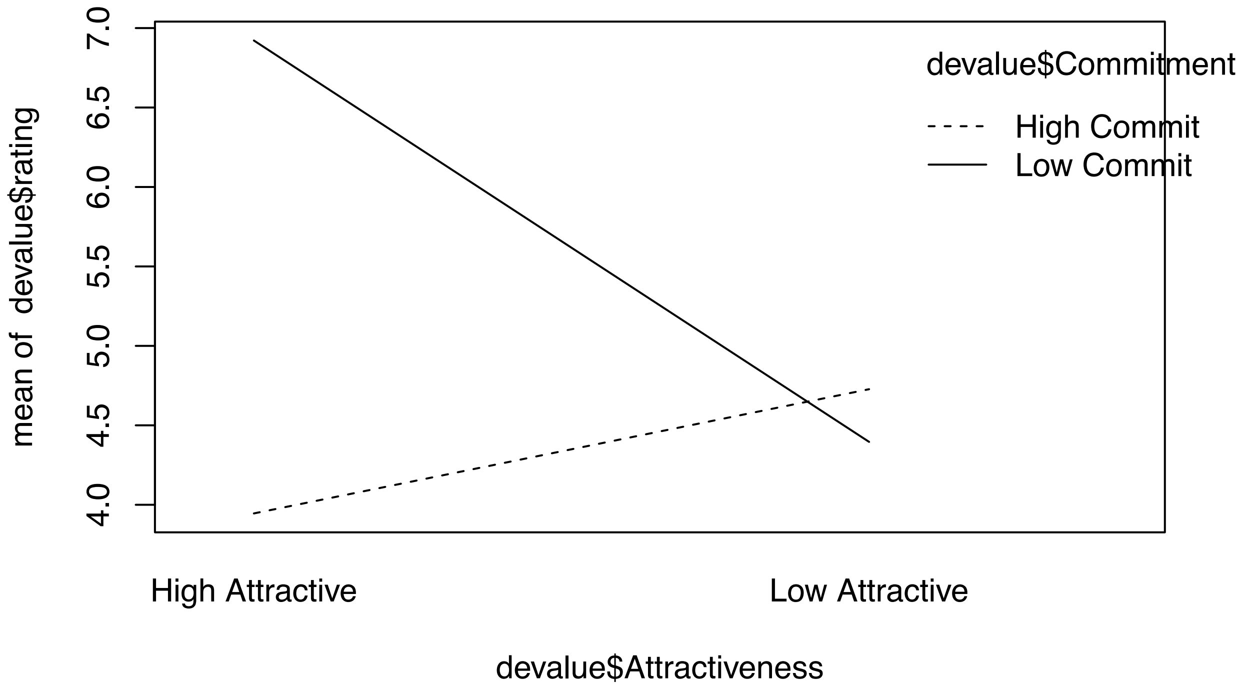
Since p = 0.0003596 < alpha at 0.05, we can reject the null hypothesis. This means that ratings of - differ significantly based on commitment. Those in the high commitment condition had lower mean ratings compared to those in the low commitment condition.

Interaction Effect

H0: The effect of attractiveness on mean rating does not depend on commitment status. There is no interaction.

H1: The effect of attractiveness on mean rating depends on commitment status. There is interaction.

```
interaction.plot(devalue$Attractiveness, devalue$Commitment, devalue$rating, fun=mean)
```



simply looking at the ANOVA table, f(1,28) = 20.62, with a p value of 9.723e-06. This means we can reject the null and there is an interaction, further tests would help determine where this interaction is.

```
hiattract <- devalue %>% filter(Attractiveness == "High Attractive")
model_hiattract <- lm(rating~Commitment,hiattract)
anova(model_hiattract)
```

```
## Analysis of Variance Table
##
## Response: rating
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Commitment  1 221.44 221.437  32.005 1.525e-07 ***
## Residuals  98 678.04   6.919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Fff <- 221.437 / 6.629
```

```
1-pf(Fff, 1, 196)
```

```
## [1] 2.909129e-08
```

Since $p <$ alpha at 0.05, we can reject the null hypothesis. This means that mean rating for those in the high attractiveness condition differ significantly by commitment.

4c

if mu represents the mean rating of those in the low commitment condition

H0: mu(high attractiveness) = mu(low attractiveness)

H1: mu(high attractiveness) != mu(low attractiveness)

```
lowcommit <- devalue %>% filter(Commitment == "Low Commit")
model_lowcommit <- lm(rating~Attractiveness, lowcommit)
anova(model_lowcommit)
```

```
## Analysis of Variance Table
##
## Response: rating
##                Df Sum Sq Mean Sq F value    Pr(>F)
## Attractiveness  1 159.51 159.515  22.894 6.052e-06 ***
## Residuals      98 682.82   6.968
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Flc <- 159.515 / 6.629
```

```
1 - pf(Flc, 1, 196)
```

```
## [1] 1.954189e-06
```

since $p < 0.05$, we can conclude that when commitment is low, the mean rating for high attractiveness participants is higher than that for low attractiveness participants.

if mu represents the mean rating of those in the high commitment condition

H0: mu(high attractiveness) = mu(low attractiveness)

H1: mu(high attractiveness) != mu(low attractiveness)

```
hicommit <- devalue %>% filter(Commitment == "High Commit")
model_hicommit <- lm(rating~Attractiveness, hicommit)
anova(model_hicommit)
```

```
## Analysis of Variance Table
##
```

```
## Response: rating
##                Df Sum Sq Mean Sq F value Pr(>F)
## Attractiveness  1  15.28 15.2775  2.4284 0.1224
## Residuals      98 616.54  6.2912
```

```
Fhc <- 15.2775 / 6.629
```

```
1 - pf(Fhc, 1, 196)
```

```
## [1] 0.130599
```

since p > 0.05, we can conclude that when commitment is high, there is no effect on mean rating based on attractiveness.

if mu represents the mean rating of those in the low attraction condition

H0: mu(high commitment) = mu(low commitment)

H1: mu(high commitment) != mu(low commitment)

```
lowattract <- devalue %>% filter(Attractiveness == "Low Attractive")
model_lowattract <- lm(rating~Commitment, lowattract)
anova(model_lowattract)
```

```
## Analysis of Variance Table
##
## Response: rating
##            Df Sum Sq Mean Sq F value Pr(>F)
## Commitment  1   2.75  2.7483  0.4335 0.5118
## Residuals  98 621.32  6.3400
```

```
Fla <- 2.7483 / 6.629
```

```
1 - pf(Fla, 1, 196)
```

```
## [1] 0.520403
```

since p > 0.05, we can conclude that when attraction is low, there is no efefct on mean rating based on commitment.

if mu represents the mean rating of those in the high attraction condition

H0: mu(high commitment) = mu(low commitment)

H1: mu(high commitment) != mu(low commitment)

```
highattract <- devalue %>% filter(Attractiveness == "High Attractive")
model_highattract <- lm(rating~Commitment, highattract)
```

```r
anova(model_highattract)
```

```
## Analysis of Variance Table
##
## Response: rating
##            Df Sum Sq Mean Sq F value    Pr(>F)
## Commitment  1 221.44 221.437  32.005 1.525e-07 ***
## Residuals  98 678.04   6.919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
Fha <- 221.437 / 6.629
```

```r
1 - pf(Fha, 1, 196)
```

```
## [1] 2.909129e-08
```

since p $<$ 0.05, we can conclude that when attractiveness is high, the mean rating for low commitment participants is higher than that for high commitment participants.

QUESTION 5

```r
first <- c(95,105,101,92,115,103,97,91,96,110,106,93,102,108,95)
second <- c(110,112,91,99,108,112,95,98,91,95,109,87,97,110,107)
guesses <- data.frame(first,second)
```

```r
mean(guesses$first)
```

```
## [1] 100.6
```

```r
mean(guesses$second)
```

```
## [1] 101.4
```

H0: The average first guess was off as much or more than the average second guess

H0: mu(abs(100 - guess2)) - mu(abs(100 - guess 1)) = 0

H1: The average first guess was less off than the average second guess

mu(abs(100 - guess2)) - mu(abs(100 - guess 1)) != 0

```r
guesses <- guesses %>%
  mutate(absD_first = abs(first - 100))
```

```r
guesses <- guesses %>%
  mutate(absD_second = abs(second - 100))
```

```r
dbar_first <- mean(guesses$absD_first)
dbar_second <- mean(guesses$absD_second)
dbar_first
```

```
## [1] 6.066667
```

```r
dbar_second
```

```
## [1] 7.666667
```

```r
sd_first <- sd(guesses$absD_first)
sd_second <- sd(guesses$absD_second)
sd_first
```

```
## [1] 3.59497
```

```r
sd_second
```

```
## [1] 3.735289
```

```r
guesses <- guesses %>%
  mutate(D = absD_second - absD_first)
```

```r
Dbar <- mean(guesses$D)
SD <- sd(guesses$D)
Dbar
```

```
## [1] 1.6
```

```r
SD
```

```
## [1] 5.578018
```

```r
Tguesses <- (Dbar - 0)/(SD/sqrt(15))
Tguesses
```

```
## [1] 1.110927
```

```r
2*(1-pt(Tguesses,15))
```

```
## [1] 0.2840904
```

p > alpha at 0.05, so we fail to reject the null hypothesis. This means that there is no significant difference between the two guesses. The first guess was not significantly better than the second.

checking:

```r
t.test(guesses$first, guesses$second, mu=0, alternative="two.sided", paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  guesses$first and guesses$second
## t = -0.35534, df = 14, p-value = 0.7276
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -5.628663  4.028663
```

```
## sample estimates:
## mean difference
##            -0.8
```