LAB WEEK 10

CORRELATION AND REGRESSION

DAILY DIARY STUDY OF MOOD

This study delivered a daily diary to 61 HIV infected men who have sex with men (MSM) between 16 and 24 years old for 66 days to measure HIV-risk behaviors and other psychosocial variables. The study examined the association between daily life stressors and daily negative mood. The study also examined the person-level association between the average number of daily life stressors across the 66 days and the average level of negative mood across that period.

Manuel Ramirez

Patrick Wilson

Ramirez, M, Wilson, P., Mitchell, R., Enders, C., & Woller, M. (in progress). Daily variability in depressed mood among gay and bisexual youth living with HIV. *Manuscript in preparation*.

KEY VARIABLES

Daily Life Stressors

Respondents were presented with list of stressful events (e.g., fights with family or friends, work stress, financial stress), and they checked how many they experienced each day.

Depression

Depression was measured using the Profile of Mood States (POMS). The POMS Depression-Dejection scale is computed by summing responses to five questionnaire items, where higher ratings reflect more negative mood.

RESEARCH QUESTION

- Question: Is there an association or trend between one's average number of life stressors and their average depressive mood?
- As the number of stressors increases or decreases, does depression change?
- Correlation and regression are appropriate for identifying trends between two numeric variables

LOAD PACKAGES AND IMPORT DATA

= data frame name
= variable name
= raw data file name

```
# LOAD R PACKAGES ----

# load R packages
library(GGally)
library(ggplot2)
library(psych)
library(summarytools)

# READ DATA ----

# github url for raw data
filepath <- 'https://raw.githubusercontent.com/craigenders/psych250a/main/data/DiaryData.csv'

# create data frame called ClinicalTrial from github data
Diary <- read.csv(filepath, stringsAsFactors = T)
```

SUMMARIZING DATA

= data frame name

= variable name

```
# INSPECT DATA ----

# summarize entire data frame (summarytools package)
dfSummary(Diary)
```

R OUTPUT

```
Data Frame Summary
Diary
Dimensions: 61 x 2
Duplicates: 0

-------------------------------------------------------------------------------------------------
No   Variable       Stats / Values          Freqs (% of Valid)   Graph                 Valid      Missing
---- ------------- ---------------------- -------------------- --------------------- ---------- ---------
1    NumStressors  Mean (sd) : 1.2 (1.1)   60 distinct values   :                     61         0
     [numeric]     min < med < max:                             :                     (100.0%)   (0.0%)
                   0 < 0.9 < 4.5                                : : .
                   IQR (CV) : 1.6 (0.9)                         : : :   . .
                                                                : : : : : : :   . .

2    Depress       Mean (sd) : 6.2 (2.1)   61 distinct values   :                     61         0
     [numeric]     min < med < max:                             :                     (100.0%)   (0.0%)
                   4 < 5.6 < 12.9                               : .
                   IQR (CV) : 2.6 (0.3)                         : : .
                                                                : : : : : . :
-------------------------------------------------------------------------------------------------
```

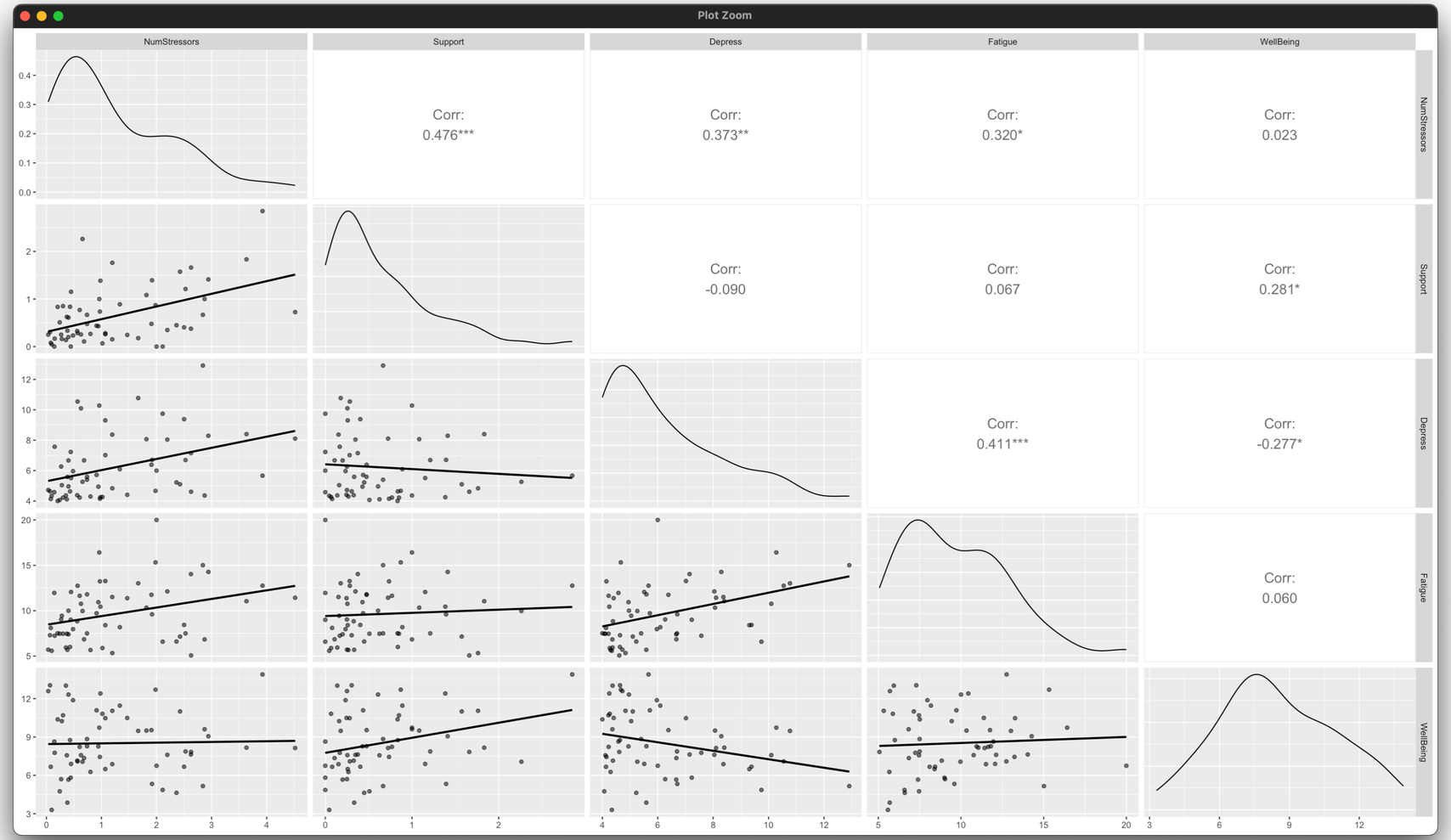
SCATTERPLOTS AND CORRELATIONS

= data frame name

= variable name

```
# SUBSET DATA: SELECT VARIABLES TO CORRELATE ----

# select variables to correlate
vars2corr <- Diary[, c('NumStressors', 'Support', 'Depress', 'Fatigue', 'WellBeing')]

# SCATTERPLOT MATRIX ----

# create the scatterplot matrix (GGalley package)
ggpairs(vars2corr,
        upper = list(continuous = wrap('cor', size = 5)),
        lower = list(continuous = wrap('smooth', method = 'lm', se = F, alpha = 0.6)),
        diag = list(continuous = wrap('densityDiag')))

# CORRELATION MATRIX ----

# correlation matrix, confidence intervals, and p-values (psych package)
print(corr.test(vars2corr), short = FALSE)
```

R OUTPUT


Plot Zoom
NumStressors
Support
Depress
Fatigue
WellBeing
Corr:
0.476***
Corr:
0.373**
Corr:
0.320*
Corr:
0.023
Corr:
-0.090
Corr:
0.067
Corr:
0.281*
Corr:
0.411***
Corr:
-0.277*
Corr:
0.060

R OUTPUT

```
Correlation matrix
             NumStressors Support Depress Fatigue WellBeing
NumStressors         1.00    0.48    0.37    0.32      0.02
Support              0.48    1.00   -0.09    0.07      0.28
Depress              0.37   -0.09    1.00    0.41     -0.28
Fatigue              0.32    0.07    0.41    1.00      0.06
WellBeing            0.02    0.28   -0.28    0.06      1.00

Sample Size
[1] 61

Probability values (Entries above the diagonal are adjusted for multiple tests.)
             NumStressors Support Depress Fatigue WellBeing
NumStressors         0.00    0.00    0.02    0.08      1.00
Support              0.00    0.00    1.00    1.00      0.17
Depress              0.00    0.49    0.00    0.01      0.17
Fatigue              0.01    0.61    0.00    0.00      1.00
WellBeing            0.86    0.03    0.03    0.65      0.00
```

R OUTPUT

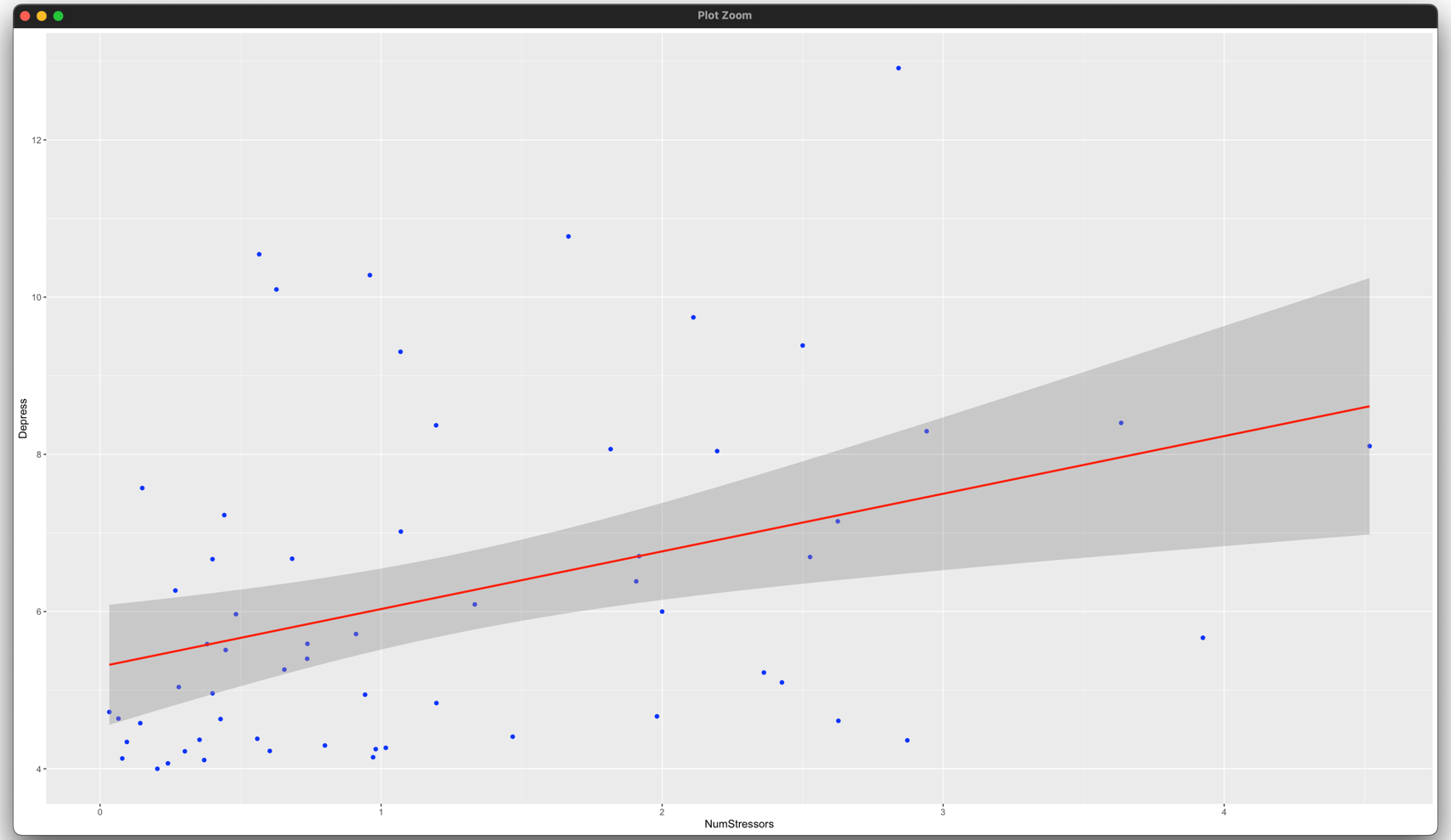
```
 Confidence intervals based upon normal theory.  To get bootstrapped values, try
cor.ci
            raw.lower raw.r raw.upper raw.p lower.adj upper.adj
NmStr-Spprt      0.25  0.48       0.65  0.00      0.15      0.71
NmStr-Dprss      0.13  0.37       0.57  0.00      0.03      0.64
NmStr-Fatig      0.07  0.32       0.53  0.01     -0.02      0.59
NmStr-WllBn     -0.23  0.02       0.27  0.86     -0.23      0.27
Spprt-Dprss     -0.33 -0.09       0.17  0.49     -0.40      0.23
Spprt-Fatig     -0.19  0.07       0.31  0.61     -0.24      0.36
Spprt-WllBn      0.03  0.28       0.50  0.03     -0.06      0.56
Dprss-Fatig      0.18  0.41       0.60  0.00      0.07      0.66
Dprss-WllBn     -0.49 -0.28      -0.03  0.03     -0.55      0.05
Fatig-WllBn     -0.20  0.06       0.31  0.65     -0.23      0.34
```

SCATTERPLOT AND REGRESSION LINE

= data frame name
= variable name
= predictor variable

```
# SCATTERPLOT WITH REGRESSION LINE ----

# scatterplot with regression line (ggplot2 package)
ggplot(data = Diary, aes(x = NumStressors, y = Depress)) +
  geom_point(color = 'blue') +
  geom_smooth(method = 'lm', color = 'red')
```

R OUTPUT



LINEAR REGRESSION

= data frame name
= variable name
= predictor variable

```
# LINEAR REGRESSION WITH 95% CONFIDENCE INTERVALS ----

# linear regression (base R)
results <- lm(Depress ~ NumStressors, data = Diary)
summary(results)

# 95% confidence intervals (base R)
confint(results)
```

R OUTPUT

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.0436 -1.3404 -0.4641  0.8722  5.5312
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.2988     0.3877  13.666  < 2e-16 ***
NumStressors   0.7334     0.2371   3.093  0.00303 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Intercept and slope

```
Residual standard error: 1.96 on 59 degrees of freedom
Multiple R-squared:  0.1395,   Adjusted R-squared:  0.1249
F-statistic: 9.565 on 1 and 59 DF,  p-value: 0.003028
```

F and R-square

```
                2.5 %   97.5 %
(Intercept)  4.5229278 6.074675
NumStressors 0.2588819 1.207889
```

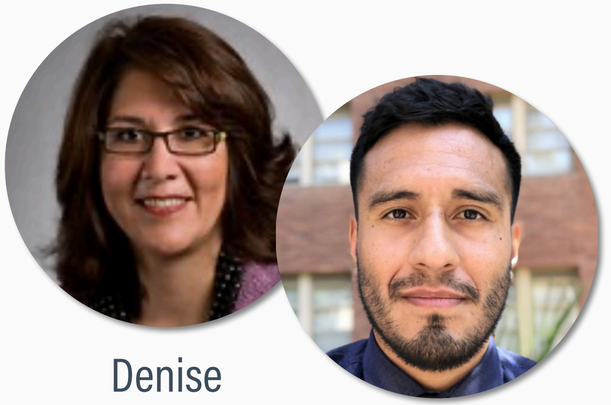
95% confidence intervals

SMALL GROUP EXERCISE

Download two files from Bruin Learn: “Week 10 Lab. Correlation and Regression.R” and “Week 10 Small Group Exercise.R”. The Lab script contains the R code we just discussed. The Exercise script contains only the URL for a different data set, Discrimination.csv. In groups of two or three, you will complete a series of R tasks that provide practice for the next assignment. There is no need to write code from scratch; instead, you can copy and paste code chunks from the Lab file into your Exercise script, modifying the data and variable names as needed. The Discrimination.csv file for this exercise contains data from study investigating the association between perceived discrimination and internalizing behaviors.

DISCRIMINATION AND INTERNALIZING BEHAVIOR

There is a well-documented relationship between perceived discrimination and internalizing symptoms among Latinx adolescents. However, few studies have examined how this psychosocial stressor relates to multiple domains of functioning in rural Latinx adolescents simultaneously. This study tested a spillover model of internalizing symptom development, where the negative effects of perceived discrimination are experienced through peer and family relationships.



Denise Chavira

Giovanni Ramos

Ramos, G., Delgadillo, D., Fossum, J., Montoya, A.K., Thamrin, H., Rapp, A., Escovar, E., Chavira, D.A. (2021). Discrimination and internalizing symptoms in rural Latinx adolescents: An ecological model of etiology. *Children and Youth Services Review*, *130*, 1-10.

KEY VARIABLES

Perceived Discrimination

Perceived discrimination refers to individuals' perception of negative attitude, judgment, or unfair treatment due to their specific characteristics such as gender, race, ethnicity, and social status.

Internalizing Behavior

Internalizing behaviors are behaviors directed inwardly toward oneself. Examples include anxious and depressive symptoms, social withdrawal, and somatic (physical. complaints.

SMALL GROUP EXERCISE TASK 1

- Use the provided URL to import theDiscrimination.csv file into an R data frame (import method #3 from the Week 0 lab script).
- Use the dfSummary function to get numeric and visual summaries of the data frame's variables.

SMALL GROUP EXERCISE TASK 2

- Use the ggpairs function in the Ggally package to obtain scatterplots among the following variables: Age, Discrimination, and Internalizing.
- Use the corr.test function from the psych package to obtain a correlation matrix.
- Consider the correlation between Discrimination and Internalizing. Provide an interpretation of the correlation and discuss its direction and strength.

SMALL GROUP EXERCISE TASK 3

- Use the ggplot2 package to obtain a scatterplot and regression line from an analysis where Discrimination is the predictor and Internalizing is the outcome.
- Use the lm function to perform a regression analysis where Discrimination is the predictor and Internalizing is the outcome.
- Provide an interpretation of the intercept and slope coefficients from the regression analysis.

SMALL GROUP EXERCISE TASK 4

- Decide on the null hypothesis that perceived discrimination and internalizing symptoms are unrelated. Based on the p-value, state whether you reject or fail to reject H_0. Justify your decision with both statistical reasoning and a short statement about the practical meaning.

- In addition to determining whether perceived discrimination has a statistically significant effect on internalizing symptoms, evaluate the strength of the relationship by interpreting the R^2 effect size. Explain what the numeric value means and compare it to Cohen's commonly used benchmarks.