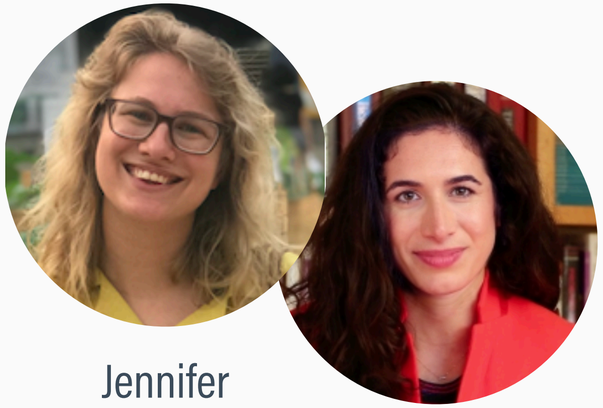
LAB WEEK 8

ANOVA

RACIAL DISPARITIES IN PTSD

People of color in the United States disproportionately bear the burden of trauma and posttraumatic stressdisorder (PTSD). Pregnant women of color are at particular risk, as perinatal PTSD is associated with adverse maternal and child health. However, PTSD is a heterogeneous disorder comprising discrete symptom dimensions. In a large sample of Latina, Black, and non-Hispanic White postpartum women in the United States (N = 1663), we examined racial and ethnic differences in the factors of the dysphoric arousal model—a leading dimensional model of PTSD.



Jennifer Thomas

Jennifer Sumner

Thomas, J.L., Carter, S.E., Dunkel-Schetter, Sumner, J.A. (2021). Racial and ethnic disparities in posttraumatic psychopathology among postpartum women. *Journal of Psychiatric Research*, *137*, 36-40.

KEY VARIABLES

Racial Ethnic Identity

The study classified woman into three racial groups: Black, Latina, and non-Hispanic White

Anxious Arousal (Hyperarousal)

Anxious arousal (hyperarousal) is the abnormally heightened state of anxiety that occurs whenever you think about a traumatic event (e.g., hypervigilance, exaggerated startle).

RESEARCH QUESTION

- Question: Do women of color (Black and Latina) differ from White women in their levels of anxious arousal?
- The research goal is to compare three different subgroups
- ANOVA is appropriate for between-subjects designs with two or more groups

LOAD PACKAGES AND IMPORT DATA

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

```
# LOAD R PACKAGES ----

# load R packages
library(ggplot2)
library(performance)
library(psych)
library(rstatix)
library(summarytools)

# READ DATA ----

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

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

SUMMARIZING DATA

= data frame name

= variable name

```
# INSPECT DATA ----

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

R OUTPUT

```
Data Frame Summary
PTSD
Dimensions: 1663 x 2
Duplicates: 1633

-------------------------------------------------------------------------------------------------------
No   Variable          Stats / Values          Freqs (% of Valid)   Graph          Valid      Missing
---- ---------------- ----------------------- -------------------- -------------- ---------- ---------
1    Race             1. Black                901 (54.2%)          IIIIIIIIII     1663       0
     [factor]         2. Latina               373 (22.4%)          IIII           (100.0%)   (0.0%)
                      3. White                389 (23.4%)          IIII

2    AnxiousArousal   Mean (sd) : 3.4 (1.8)   11 distinct values   :              1663       0
     [integer]        min < med < max:                             :              (100.0%)   (0.0%)
                      1 < 3 < 11                                   : .
                      IQR (CV) : 2 (0.5)                           : : .
                                                                   : : : : : .
-------------------------------------------------------------------------------------------------------
```

DESCRIPTIVE STATISTICS BY GROUP

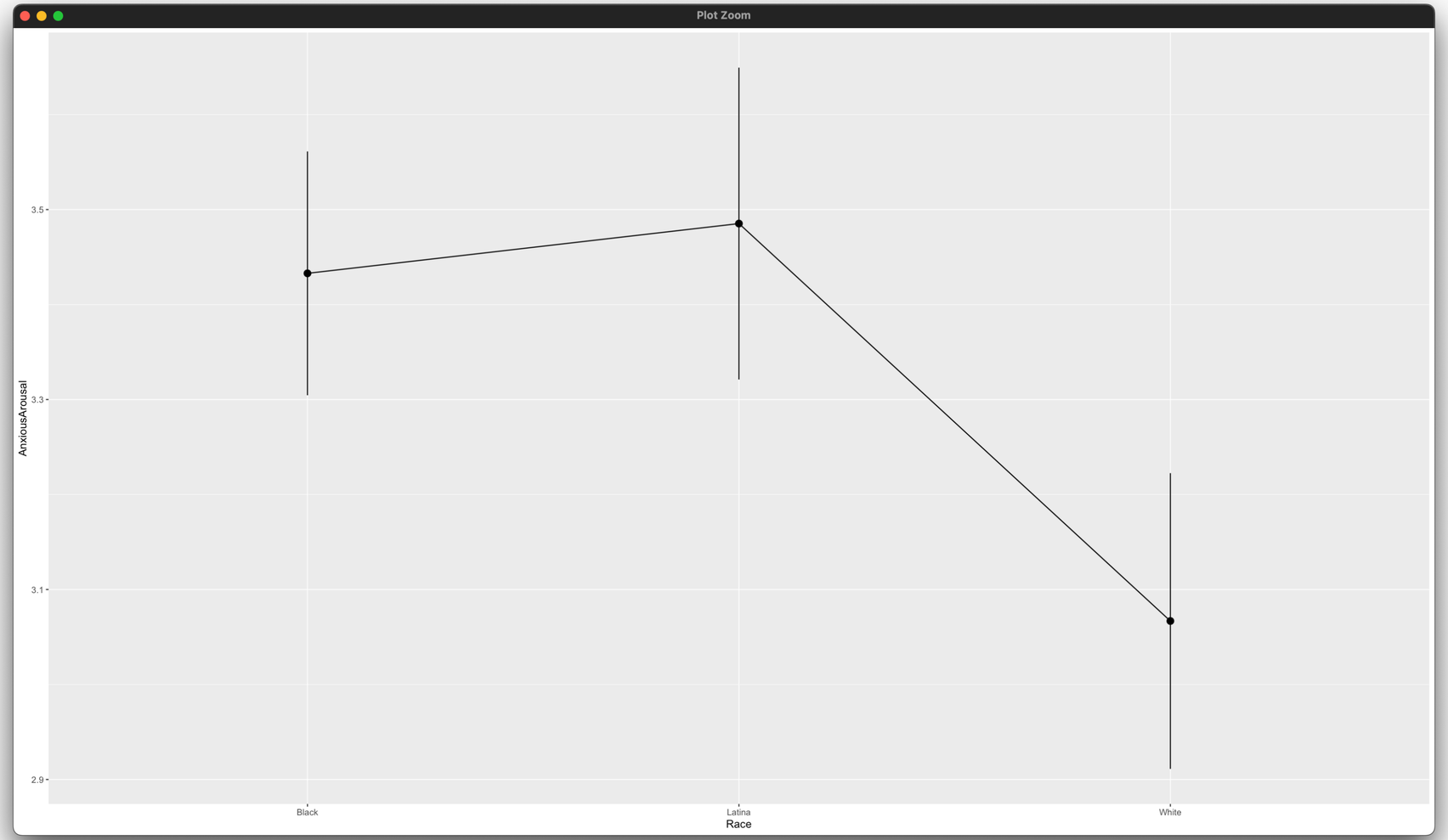
= data frame name
= variable name
= grouping variable

```
# DESCRIPTIVE STATISTICS AND 95% CONFIDENCE INTERVALS BY GROUP ----

# descriptive statistics separately by group (psych package)
describeBy(AnxiousArousal ~ Race, data = PTSD)

# means and 95% confidence interval error bars by group (ggplot2 package)
ggplot(data = PTSD, aes(x = Race, y = AnxiousArousal, group = 1)) +
  stat_summary(fun = mean, geom = 'line') +
  stat_summary(fun = mean, geom = 'point', size = 3) +
  stat_summary(fun.data = mean_cl_normal, geom = 'pointrange')
```

R OUTPUT

```
 Descriptive statistics by group
Race: Black
                vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
AnxiousArousal     1 901 3.43 1.96      3    3.19 1.48   1  11    10 1.14     0.88 0.07
------------------------------------------------------------------------------------
Race: Latina
                vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
AnxiousArousal     1 373 3.49 1.61      3    3.32 1.48   1  10     9 0.88     0.47 0.08
------------------------------------------------------------------------------------
Race: White
                vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
AnxiousArousal     1 389 3.07 1.56      3    2.85 1.48   1  10     9 1.44     2.56 0.08
```

R OUTPUT


Plot Zoom
3.5
3.3
3.1
2.9
AnxiousArousal
Black
Latina
White
Race

ANOVA

= data frame name
= variable name
= grouping variable

```r
# ANOVA ANALYSIS ----

# anova table with classic f statistic (base R)
anovaresults <- aov(AnxiousArousal ~ Race, data = PTSD)
summary(anovaresults)

# welch's f statistic for heterogeneous variation (base R)
oneway.test(AnxiousArousal ~ Race, data = PTSD)
```

R OUTPUT

```
> summary(anovaresults)

              Df Sum Sq Mean Sq F value  Pr(>F)
Race           2     44  22.032   6.797 0.00115 **
Residuals   1660   5381   3.241
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> oneway.test(AnxiousArousal ~ Race, data = PTSD)

   One-way analysis of means (not assuming equal variances)

data:  AnxiousArousal and Race
F = 8.4651, num df = 2.00, denom df = 868.45, p-value = 0.0002286
```

R-SQUARE EFFECT SIZE

```
# R-SQUARE (PROPORTION VARIANCE EXPLAINED) EFFECT SIZE ----

# r-square effect size (performance package)
r2(anovaresults)
```

R OUTPUT

```
# R2 for Anova Regression
       R2: 0.008
  adj. R2: 0.007
```

PAIRWISE (TUKEY) COMPARISONS

```
# PAIRWISE COMPARISONS (TUKEY TESTS) -----

# pairwise Tukey mean comparisons (base R)
TukeyHSD(anovaresults)
```

R OUTPUT

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = AnxiousArousal ~ Race, data = PTSD)

$Race
                     diff        lwr        upr     p adj
Latina-Black  0.05240231 -0.2076272  0.3124318 0.8841360
White-Black  -0.36601434 -0.6222339 -0.1097947 0.0023719
White-Latina -0.41841665 -0.7244741 -0.1123592 0.0039054
```

STANDARDIZED MEAN DIFFERENCE

= data frame name
= variable name
= grouping variable

```
# PAIRWISE STANDARDIZED MEAN DIFFERENCE EFFECT SIZES ----

# standardized mean difference effect size for all comparisons (rstatix package)
cohens_d(AnxiousArousal ~ Race, data = PTSD)
```

R OUTPUT

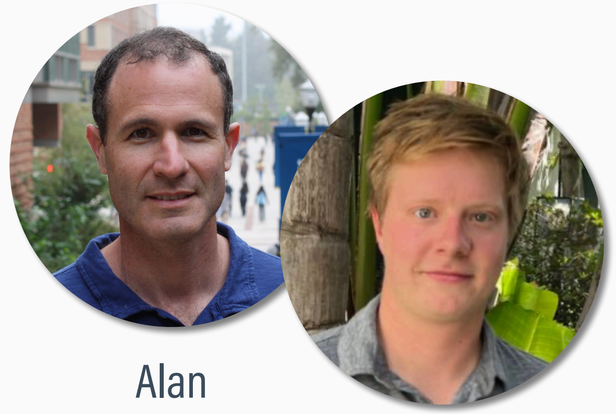
```
  .y.             group1 group2 effsize    n1    n2 magnitude
* <chr>           <chr>  <chr>    <dbl> <int> <int> <ord>
1 AnxiousArousal Black  Latina -0.0292   901   373 negligible
2 AnxiousArousal Black  White   0.206    901   389 small
3 AnxiousArousal Latina White   0.264    373   389 small
```

SMALL GROUP EXERCISE

Download two files from Bruin Learn: “Week 8 Lab. ANOVA.R” and “Week 8 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, VideoSpeedData.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 VideoSpeedData.csv file for this exercise contains data from study investigating the impact of video playback speed on later recall of learned information.

VIDEO SPEED AND MEMORY RECALL

We presented participants with lecture videos at different speeds and tested immediate and delayed (1 week) comprehension. Results revealed minimal costs incurred by increasing video speed from 1x to 1.5x, or 2x speed, but performance declined beyond 2x speed. We also compared learning outcomes after watching videos once at 1x, 1.5x, or 2.5x speeds.



Alan Castel

Dillon Murphy

Murphy, D.H., Hoover, K.M., Agadzhanyan, K., Kuehn, J.C., Castel, A.D. (2022). Learning in double time: The effect of lecture video speed on immediate and delayed comprehension. *Applied Cognitive Psychology*, *36*, 69–82.

KEY VARIABLES

Video Playback Speed

The independent variable had three conditions. Participants were assigned to watch lecture videos at three speeds: 1x, 1.5x, and 2.5x.

Recall Performance

The dependent variable, recall performance, was measured as the % correct on a test administered after a one-week delay.

SMALL GROUP EXERCISE TASK 1

- Use the provided URL to import the VideoSpeedData.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 describeBy function to get descriptive statistics for the DelayedRecall dependent variable within each of the three VideoSpeed groups (1x, 1.5x, and 2.5x speeds).
- Use the ggplot2 package to obtain a graph of the means and 95% confidence intervals.
- Based on the “inference by eye” rule of thumb, which groups appear to differ and which are similar?

SMALL GROUP EXERCISE TASK 3

- State and justify the hypotheses. Clearly write out the null hypothesis (H_0) and the alternative (research) hypothesis (H_1) in both statistical notation and plain language.
- Perform an ANOVA to determine whether delayed recall of learned information differs according to video playback speed (1x, 1.5x, and 2.5x speeds).

SMALL GROUP EXERCISE TASK 4

- Explain the conceptual meaning of the between-group and within-group (residual) mean square values. What pattern of large/small mean square values would provide strong evidence against the null hypothesis.

- Explain what the magnitude of the F-value tells you about how far the observed mean difference is from the null hypothesis.

- Interpret the probability value. I am NOT asking you to say whether the difference is statistically significant.

SMALL GROUP EXERCISE TASK 5

- In addition to determining whether treatment group membership has a statistically significant effect on pain interference, 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.

- Use post hoc tests to determine which specific group differences contributed to the overall F statistic. Which groups differed?