

PSY 3307

Two-Way Analysis of Variance

Jonathan A. Pedroza PhD

Cal Poly Pomona

2021-10-26

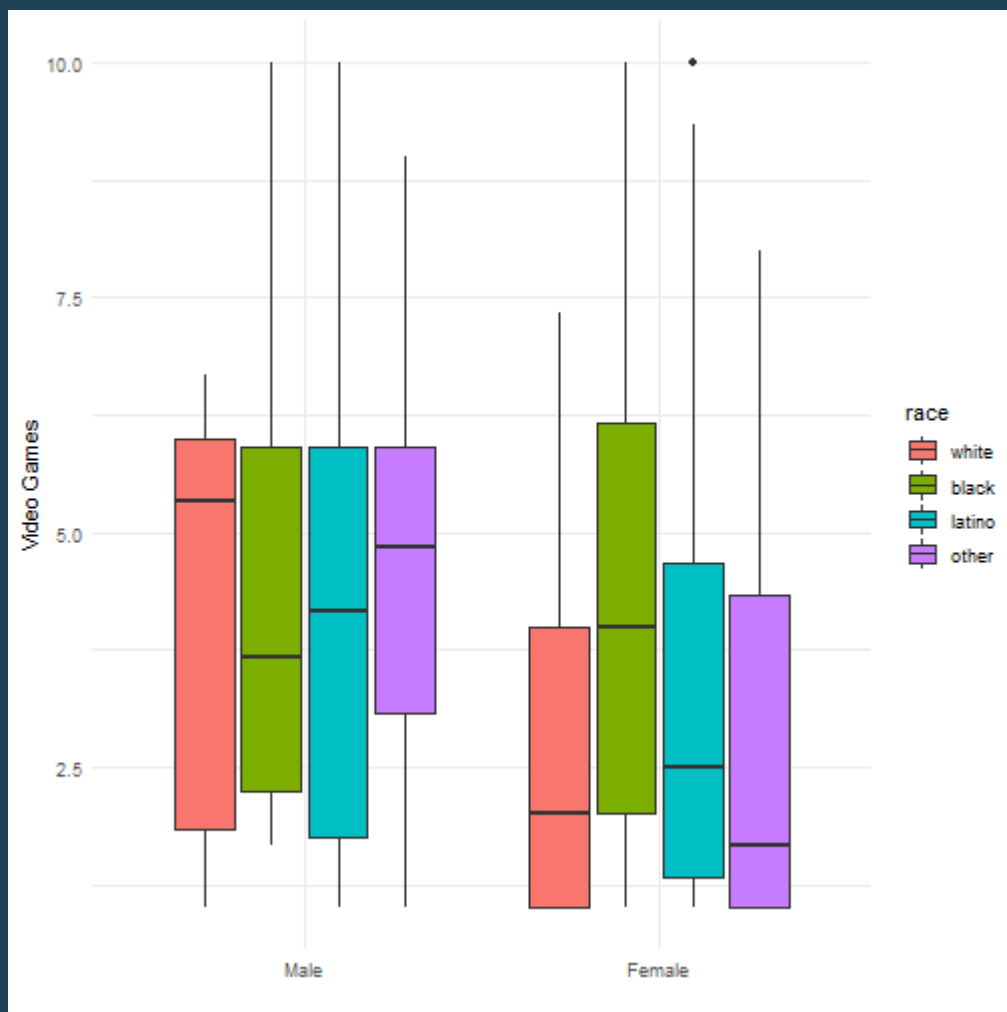
Agenda

- Go over the problem set
- Talk about two-way ANOVA
- Talk about how to do one-way ANOVA in SPSS
 - small introduction into two-way ANOVA

```

jp <- rio::import(here::here("jp_thesis_1.sav")) %>%
  janitor::clean_names() %>%
  rowid_to_column() %>%
  mutate(sex = ccc_gender,
         sex = recode(sex, "1" = "Male",
                       "2" = "Female"),
         sex = factor(sex),
         race = recode(ccc_ethnicity, "0" = "other",
                          "1" = "other",
                          "2" = "black",
                          "3" = "white",
                          "4" = "latino",
                          "5" = "other",
                          "6" = "other"),
         race = factor(race),
         bmi = ccc_bmi,
         games_avg = (mtuas_videogaming_q1 + mtuas_videogaming_q2 + mtuas_videogaming_q3) / 3,
         tv_avg = (mtuas_tv_q1 + mtuas_tv_q2) / 2,
         race = relevel(race, ref = "white"),
         sex = relevel(sex, ref = "Male"))

```



```
games_aov <- aov(games_avg ~ sex*race, data = jp)
summary(games_aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## sex              1   67.9    67.91  12.458 0.00047 ***
## race             3   33.2    11.06   2.029 0.10940
## sex:race         3   13.9     4.62   0.847 0.46885
## Residuals      364 1984.2     5.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

TukeyHSD(games_aov)

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = games_avg ~ sex * race, data = jp)
##
## $sex
##           diff          lwr          upr      p adj
## Female-Male -0.922324 -1.436203 -0.4084449 0.0004697
##
## $race
##           diff          lwr          upr      p adj
## black-white   1.0951037 -0.1055662  2.2957736 0.0880413
## latino-white  0.4365617 -0.5218704  1.3949937 0.6426557
## other-white   0.2870969 -0.9494590  1.5236527 0.9322516
## latino-black -0.6585420 -1.5789904  0.2619063 0.2533249
## other-black  -0.8080068 -2.0153608  0.3993472 0.3109927
## other-latino  -0.1494648 -1.1162571  0.8173276 0.9784633
##
## $`sex:race`
##           diff          lwr          upr      p adj
## Female:white-Male:white -1.42105263 -3.95082181  1.1087165 0.6791225
## Male:black-Male:white    0.33333333 -2.47400554  3.1406722 0.9999610
## Female:black-Male:white  -0.03333333 -2.58558979  2.5189231 1.0000000
## Male:latino-Male:white   -0.07657658 -2.47472390  2.3215707 1.0000000
## Female:latino-Male:white -0.82222222 -3.14691967  1.5024752 0.9609956 / 32
```

```
effectsize::eta_squared(games_aov, alternative = "two.sided", ci = .95)
```

```
## # Effect Size for ANOVA (Type I)
```

```
##
```

```
## Parameter | Eta2 (partial) |          95% CI
```

```
## -----
```

```
## sex       |           0.03 | [0.01, 0.08]
```

```
## race      |           0.02 | [0.00, 0.04]
```

```
## sex:race  |        6.93e-03 | [0.00, 0.03]
```

```
games_lm <- lm(games_avg ~ sex*race, data = jp)
summary(games_lm)
```

```
##
## Call:
## lm(formula = games_avg ~ sex * race, data = jp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.643 -1.889 -0.445  1.577  6.656
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.16667    0.73832   5.643 0.0000000336 ***
## sexFemale     -1.42105    0.82980  -1.713   0.0877 .
## raceblack      0.33333    0.92085   0.362   0.7176
## racelatin     -0.07658    0.78662  -0.097   0.9225
## raceother      0.47619    0.96669   0.493   0.6226
## sexFemale:raceblack  1.05439    1.07105   0.984   0.3256
## sexFemale:racelatin  0.67541    0.89363   0.756   0.4503
## sexFemale:raceother -0.33292    1.11495  -0.299   0.7654
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.335 on 364 degrees of freedom
## Multiple R-squared:  0.05476,    Adjusted R-squared:  0.03658
## F-statistic: 3.012 on 7 and 364 DF,  p-value: 0.004344
```



```
effectsize::eta_squared(games_lm, alternative = "two.sided", ci = .95)
```

```
## # Effect Size for ANOVA (Type I)
```

```
##
```

```
## Parameter | Eta2 (partial) |          95% CI
```

```
## -----
```

```
## sex       |           0.03 | [0.01, 0.08]
```

```
## race      |           0.02 | [0.00, 0.04]
```

```
## sex:race   |        6.93e-03 | [0.00, 0.03]
```

```
effectsize::eta_squared(games_aov, alternative = "two.sided", ci = .95)
```

```
## # Effect Size for ANOVA (Type I)
```

```
##
```

```
## Parameter | Eta2 (partial) |          95% CI
```

```
## -----
```

```
## sex       |           0.03 | [0.01, 0.08]
```

```
## race      |           0.02 | [0.00, 0.04]
```

```
## sex:race  |        6.93e-03 | [0.00, 0.03]
```

```
.03 + .02 + 6.93e-03
```

```
## [1] 0.05693
```

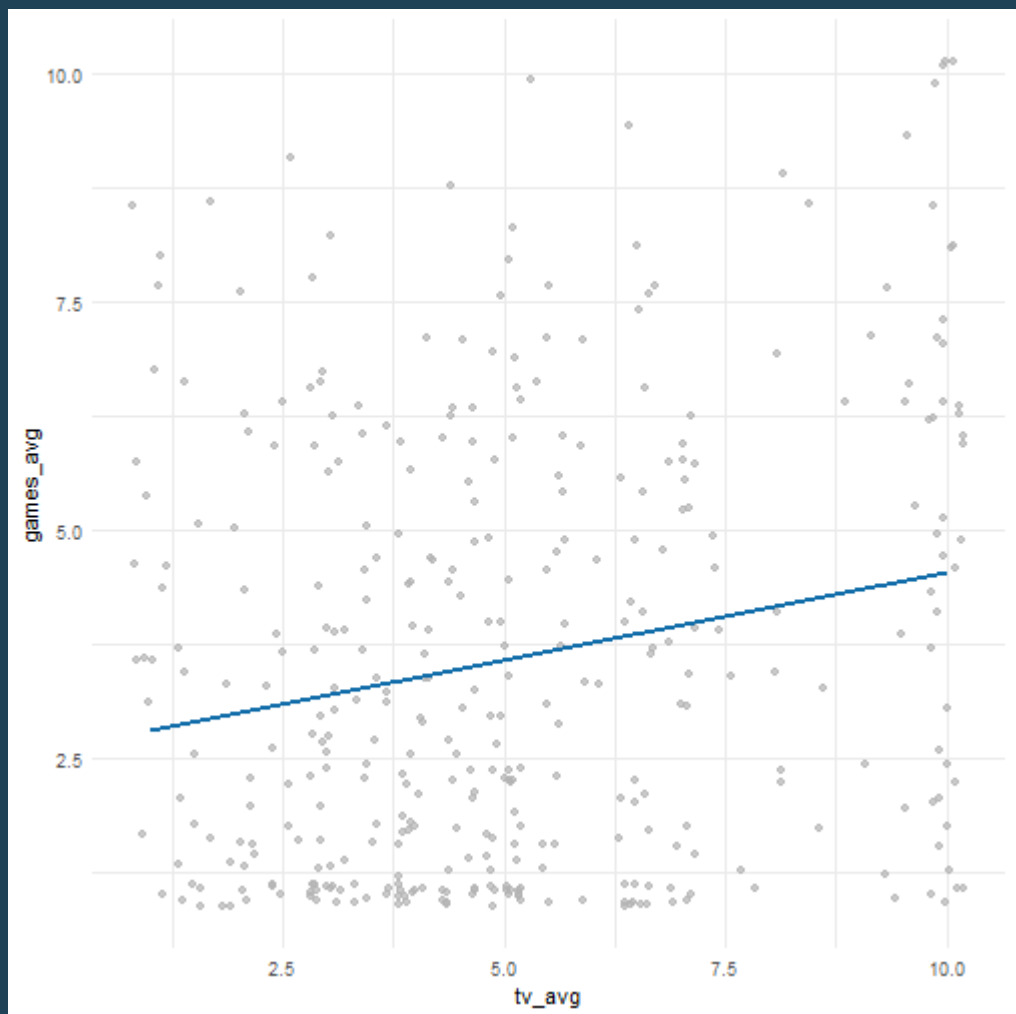
```
summary(games_lm)$r.squared
```

```
## [1] 0.05475845
```

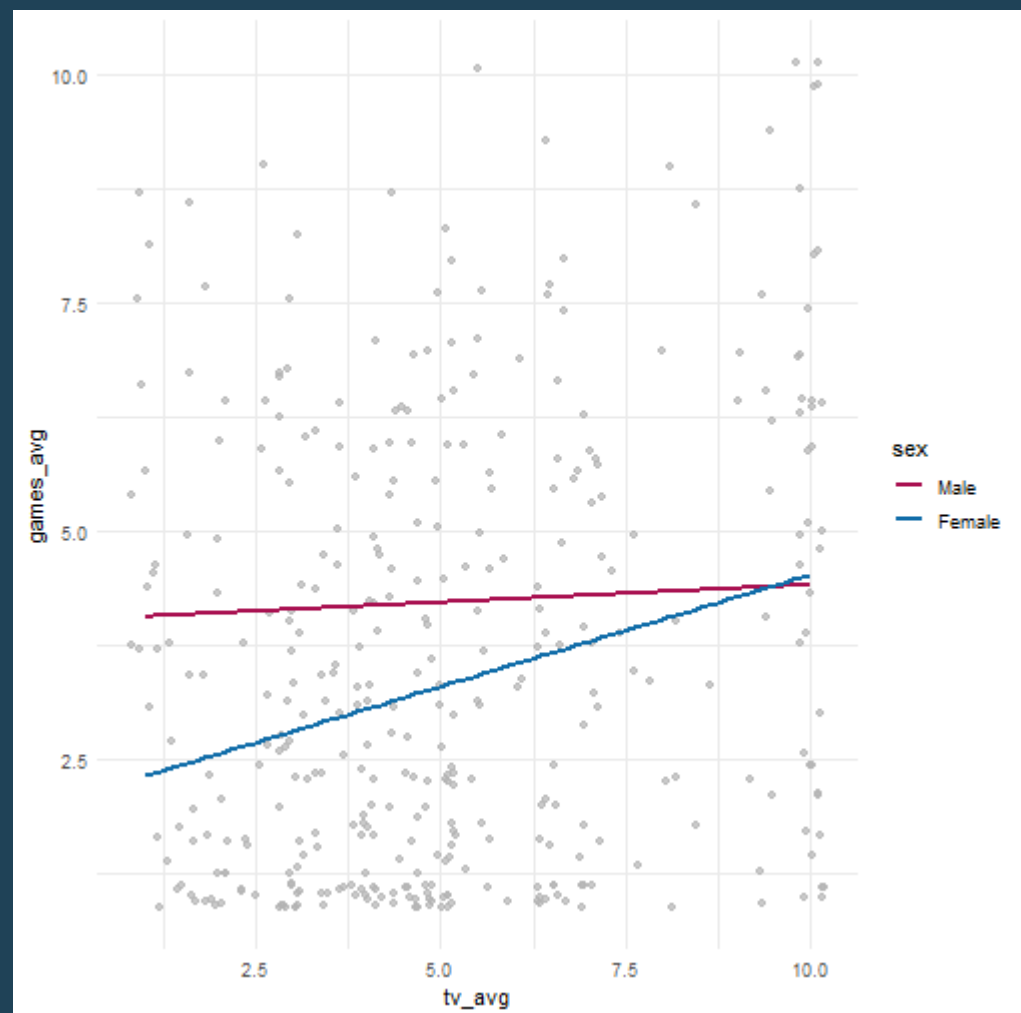
```
games_lm2 <- lm(games_avg ~ tv_avg*sex, data = jp)
summary(games_lm2)
```

```
##
## Call:
## lm(formula = games_avg ~ tv_avg * sex, data = jp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5225 -1.8583 -0.4204  1.4894  6.5796
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.02427    0.50918   7.903 0.00000000000000318 ***
## tv_avg         0.03881    0.08857   0.438  0.66151
## sexFemale     -1.95077    0.59839  -3.260  0.00122 **
## tv_avg:sexFemale  0.20609    0.10461   1.970  0.04958 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.289 on 368 degrees of freedom
## Multiple R-squared:  0.08115,    Adjusted R-squared:  0.07366
## F-statistic: 10.83 on 3 and 368 DF,  p-value: 0.0000007761
```

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



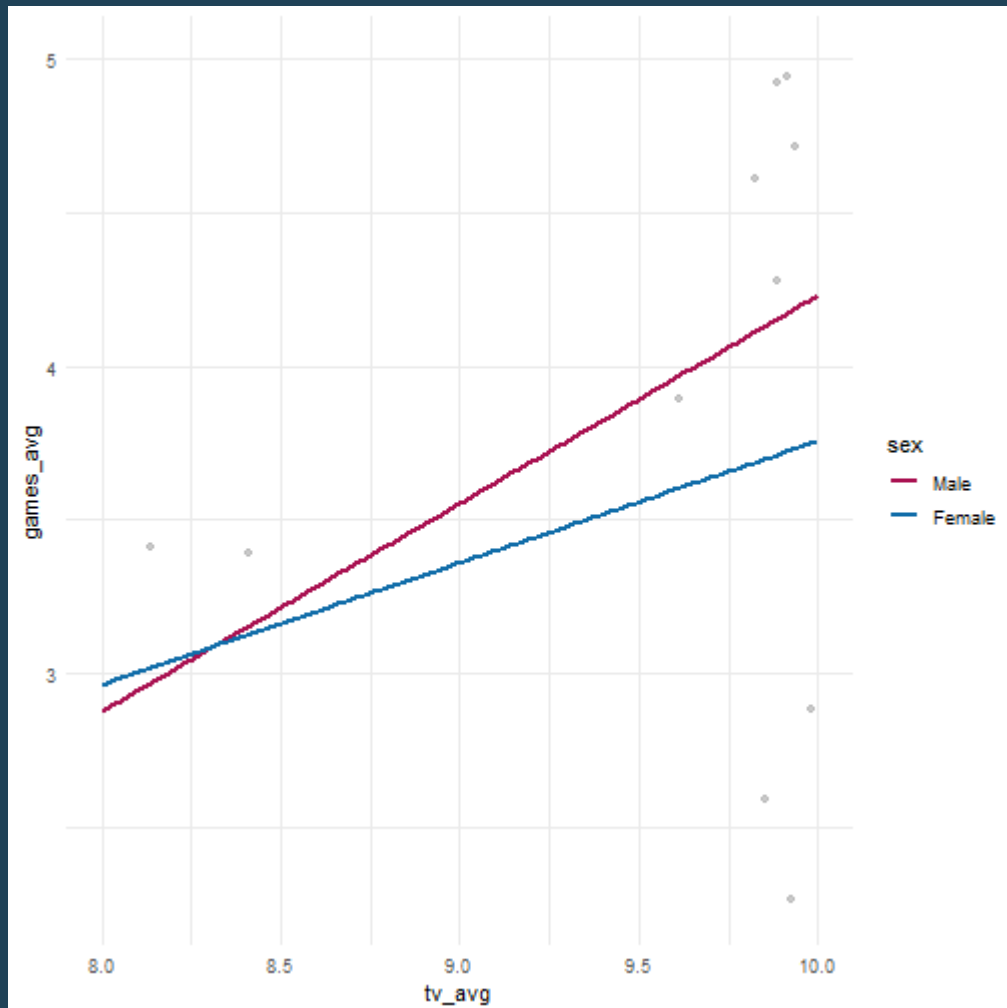
```
## `geom_smooth()` using formula 'y ~ x'
```



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

```
## Warning: Removed 353 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 361 rows containing missing values (geom_point).
```



```
interactions::probe_interaction(games_lm2, pred = tv_avg, modx = sex)
```

```
## SIMPLE SLOPES ANALYSIS
```

```
##
```

```
## Slope of tv_avg when sex = Female:
```

```
##
```

| ## | Est. | S.E. | t val. | p |
|----|-------|-------|--------|-------|
| ## | ----- | ----- | ----- | ----- |
| ## | 0.24 | 0.06 | 4.40 | 0.00 |

```
##
```

```
## Slope of tv_avg when sex = Male:
```

```
##
```

| ## | Est. | S.E. | t val. | p |
|----|-------|-------|--------|-------|
| ## | ----- | ----- | ----- | ----- |
| ## | 0.04 | 0.09 | 0.44 | 0.66 |

| | | | | |
|---|--|--|---|--|
| $\begin{bmatrix} - \\ \text{Male} \\ \text{Female} \end{bmatrix}$ | $\begin{bmatrix} \text{White} \\ \text{Male White} \\ \text{Female White} \end{bmatrix}$ | $\begin{bmatrix} \text{Black} \\ \text{Male Black} \\ \text{Female Black} \end{bmatrix}$ | $\begin{bmatrix} \text{Latino} \\ \text{Male Latino} \\ \text{Female Latina} \end{bmatrix}$ | $\begin{bmatrix} \text{Other} \\ \text{Male O} \\ \text{Female O} \end{bmatrix}$ |
|---|--|--|---|--|


```
set.seed(102621)

table <- data.frame(sex = c("male", "male", "male", "female", "female",
                             white = rnorm(6, 10, 1.5),
                             black = rnorm(6, 5, 2.4),
                             latino = rnorm(6, 15, 3.1),
                             other = rnorm(6, 12, 5))
```

```
table
```

| ## | sex | white | black | latino | other |
|------|--------|-----------|----------|----------|-----------|
| ## 1 | male | 9.826393 | 6.659524 | 17.89335 | 2.092671 |
| ## 2 | male | 12.819043 | 3.846659 | 14.55356 | 5.404460 |
| ## 3 | male | 7.326992 | 7.134491 | 18.51053 | 11.091112 |
| ## 4 | female | 8.978196 | 3.108109 | 14.12090 | 10.223289 |
| ## 5 | female | 13.758694 | 5.488555 | 12.02529 | 14.294188 |
| ## 6 | female | 8.770219 | 5.401871 | 15.96283 | 13.253053 |

Two-way ANOVA

- **two-way ANOVA** is a parametric ANOVA test that involves two IVs
 - one-way ANOVA is with one IV
- we will be covering a **two-way between-subjects ANOVA**, or two IVs that involve independent samples
- we will come back to within-subjects ANOVA
 - repeated measures ANOVA
 - **two-way within-subjects ANOVA** is when every participant receives the same two IVs
- lastly, we will cover **two-way mixed-design ANOVA**, which is when one IV is between-subjects and the other IV is within-subjects

Two-way ANOVA

- to understand two-way ANOVAs we will refer to the amount of levels each variable has
 - Ex: Variable 1: Sex (Male and Female) & Variable 2: Generation (baby boomers, generation X, millennials)
 - The design is a 3 x 2 (3 by 2) ANOVA, is also a 2 x 3 ANOVA (just know which variable has each number of levels)
- our two-way ANOVAs are not the same thing as an ANCOVA because we are interested in the **interaction** of our two IVs
- in a two-way ANOVA, we need to think about the **cells**, or the combination of the two factors/IVs' levels
 - the columns are one IV
 - the rows are the other IV

Two-way ANOVA

- when each level of both IVs are combined, this is seen as a factorial design
- assumptions of a two-way ANOVA
 - pretty much exactly the same assumptions as a one-way ANOVA
 - homogenous variance
 - interval/ratio DV
 - normally distributed

Main Effects

- JP Note: Just because it states it is a main effect, it is not automatically cause --> effect
 - this is just a term
- the main effect is the association between one IV and the DV
 - we have been doing this with one-one ANOVA
- In a two-way ANOVA, we get the mean for each IV
 - this is referred to as **collapsing**, where we average the scores from each level in one factor to calculate the main effect mean for the other factor

Main Effects

- a **main effect mean** is the mean of the level of one factor after collapsing the other factor
- *to see a main effect, look at how the main effect means change as the levels of the factor change*
- simply, when we are interested in one IV and the association with the DV, we either look at our data horizontally or vertically

Interactions

- the difference between a one-way ANCOVA and a two-way ANOVA is the inclusion of an interaction
- two IVs combined and the influence of both on the DV is referred to as a two-way interaction
 - less common is a three-way interaction, which is three variables together
- rather than collapsing, we look at the cell means
 - **cell means** are the means of the scores from one cell

```
set.seed(102621)

table <- data.frame(sex = c("male", "male", "male", "female", "female",
                             white = rnorm(6, 10, 1.5),
                             black = rnorm(6, 5, 2.4),
                             latino = rnorm(6, 15, 3.1))
```

```
table
```

| ## | sex | white | black | latino |
|------|--------|-----------|----------|----------|
| ## 1 | male | 9.826393 | 6.659524 | 17.89335 |
| ## 2 | male | 12.819043 | 3.846659 | 14.55356 |
| ## 3 | male | 7.326992 | 7.134491 | 18.51053 |
| ## 4 | female | 8.978196 | 3.108109 | 14.12090 |
| ## 5 | female | 13.758694 | 5.488555 | 12.02529 |
| ## 6 | female | 8.770219 | 5.401871 | 15.96283 |

Interactions

- Ex: we would look at the average for white males, black males, latino males and white females, black females, latina females
- For our main effects, we compare the level means, for the interaction, we look at the cell means
- **two-way interaction effect** is when a IV-DV relationship is dependent on the other IV values
- the book has a mention of an interaction with one value going up as the other goes down
 - **but** there are times where the visual looks like an interaction but the interaction is cancelled out (I'll show an example)
- when including an interaction, the order of caring is:
 - is the interaction is significant, good
 - if not, then look at the main effects

Recap

- In a two-way between-subjects ANOVA we have
 - the first main effect (IV 1 and its association with DV)
 - the second main effect (IV 2 and its association with DV)
 - the interaction effect (IV 1 x IV2 and this association with DV)
- this means that we are essentially running three different F tests
 - we'll show this in SPSS
 - one for the first main effect
 - another for the second main effect
 - a third for the interaction
 - then whatever is leftover in the error/within
- look at the interaction first, then the main effects

Looking at Main Effects

$$HSD = (q_k) \left(\sqrt{\frac{MS_{wn}}{n}} \right)$$

- we need the k , df_{wn} , MS_{wn} , and the n
 - make sure to remember that the n may be different for groups

$$\eta^2 = \frac{SS_{bn}}{SS_{total}}$$

Looking at the Interaction

- along with working out the calculations, we will also visualize the interaction
- you can visualize it several different ways
 - line graph
 - boxplots -bar graphs
- I'll show all of these in SPSS
- parallel lines in a visualization indicates there is no interaction

Tukey for Interaction

- can be used to see what specific cell means are significantly different
- it is slightly different for SPSS
- we don't compare every single cell mean though
- **counfounded comparisons** are comparisons between two cells that differ along more than one factor
- **unconfounded comparisons** are comparisons that differ along only one factor
 - this is the preferred way of looking at comparisons
 - we'll only look at vertical or horizontal comparisons
- because we're not looking at all comparisons, we'll examine adjusted k

```
set.seed(102621)

table <- data.frame(sex = c("male", "male", "male", "female", "female",
                             white = rnorm(6, 10, 1.5),
                             black = rnorm(6, 5, 2.4),
                             latino = rnorm(6, 15, 3.1))
```

```
table
```

| ## | sex | white | black | latino |
|------|--------|-----------|----------|----------|
| ## 1 | male | 9.826393 | 6.659524 | 17.89335 |
| ## 2 | male | 12.819043 | 3.846659 | 14.55356 |
| ## 3 | male | 7.326992 | 7.134491 | 18.51053 |
| ## 4 | female | 8.978196 | 3.108109 | 14.12090 |
| ## 5 | female | 13.758694 | 5.488555 | 12.02529 |
| ## 6 | female | 8.770219 | 5.401871 | 15.96283 |

Interpreting Two-way ANOVA

- We'll look at the interaction first
 - Then look at each main effect if there is a nonsignificant interaction
1. look at the F tests
 2. look at post-hoc comparisons
 3. look at effect size of each component of the model (main effects and interaction)

Final Note

Because we're talking about generalized linear models, we will begin to talk about parameter estimates and what they mean

- Something super important to note is how to interpret single associations
- we will have more exposure to this in regression, but it is super helpful for interpretation

For example, if we see a significant association between sex and our outcome, we would interpret it as:

For a one point increase in X, there is a __ increase/decrease in Y (continuous)

For a one point difference Between Males and Females, there is a __ increase/decrease in Y (categorical)