Analysis of Variance

Post-hoc tests & contrasts

Jenine Harris
Brown School

Importing and cleaning the data

```
# load GSS rda file
load(file = "/Users/harrisj/Box/teaching/Teaching/Fall2020/data/gss2018.

# assign GSS to gss.2018
gss.2018 <- GSS
# remove GSS
rm(GSS)

# recode variables of interest to valid ranges
gss.2018.cleaned <- gss.2018 %>%
  select(HAPPY, SEX, DEGREE, USETECH, AGE) %>%
  mutate(USETECH = na_if(x = USETECH, y = -1)) %>%
  mutate(USETECH = na_if(x = USETECH, y = 999)) %>%
  mutate(USETECH = na_if(x = USETECH, y = 998)) %>%
  mutate(AGE = na_if(x = AGE, y = 98)) %>%
  mutate(AGE = na_if(x = AGE, y = 99)) %>%
  mutate(DEGREE = na_if(x = DEGREE, y = 8)) %>%
  mutate(DEGREE = na_if(x = DEGREE, y = 9)) %>%
  mutate(HAPPY = na_if(x = HAPPY, y = 8)) %>%
  mutate(HAPPY = na_if(x = HAPPY, y = 9)) %>%
  mutate(HAPPY = na_if(x = HAPPY, y = 0)) %>%
  mutate(SEX = factor(x = SEX, labels = c("male","female"))) %>%
  mutate(DEGREE = factor(x = DEGREE, labels = c("< high school",
                                                 "high school", "junior c
                                                 "college", "grad school"
  mutate(HAPPY = factor(x = HAPPY, labels = c("very happy",
                                               "pretty happy",
```

Visualizing the groups

```
# graph usetech
gss.2018.cleaned %>%
  drop_na(USETECH) %>%
  ggplot(aes(y = USETECH, x = DEGREE)) +
  geom_jitter(aes(color = DEGREE), alpha = .6) +
  geom_boxplot(aes(fill = DEGREE), alpha = .4) +
  scale_fill_brewer(palette = "Spectral", guide = FALSE) +
  scale_color_brewer(palette = "Spectral", guide = FALSE) +
  theme_minimal() +
  labs(x = "Highest educational attainment",
       y = "Percent of time spent using technology",
       title = "Distribution of time spent using technology\nuse by educ
```

Group means

```
# mean and sd of age by group
use.stats <- gss.2018.cleaned %>%
  drop_na(USETECH) %>%
  group_by(DEGREE) %>%
  summarize(m.techuse = mean(USETECH),
            sd.techuse = sd(USETECH))
use.stats
```

```
## # A tibble: 5 x 3
##   DEGREE         m.techuse sd.techuse
##   <fct>              <dbl>      <dbl>
## 1 < high school       24.8       36.2
## 2 high school         49.6       38.6
## 3 junior college      62.4       35.2
## 4 college             67.9       32.1
## 5 grad school         68.7       30.2
```

ANOVA results

```r
# conduct ANOVA for technology use by degree category
techuse.by.deg <- oneway.test(formula = USETECH ~ DEGREE,
                              data = gss.2018.cleaned,
                              var.equal = TRUE)
techuse.by.deg
```

```
##
##      One-way analysis of means
##
## data:  USETECH and DEGREE
## F = 43.304, num df = 4, denom df = 1404, p-value < 2.2e-16
```

Choosing and using post-hoc tests and contrasts

- ANOVA is an *omnibus* test, which means it identifies whether there are any differences, but doesn't give any information about what is driving the significant results.

- There are two main ways to determine where significant differences among groups are following a significant ANOVA:

  - Post-hoc tests
  - Planned contrasts

- The two methods are both useful for examining differences among means.

- The difference between post-hoc tests and planned contrasts is that post-hoc tests examine *each pair of means* to determine which ones are the most different from each other and are therefore driving the statistically significant results.

- On the other hand, planned contrasts compare specified subsets of means or groups of means.

Post-hoc tests: Bonferroni

- There are several different types of post-hoc tests, one of the more commonly used is the *Bonferroni* post-hoc test.
- The Bonferroni post-hoc test is a **pairwise** test that conducts a t-test for each pair of means but adjusts the threshold for statistical significance to ensure that there is a small enough risk of Type I error.

```
# find differences in mean tech use by degree groups
bonf.tech.by.deg <- pairwise.t.test(x = gss.2018.cleaned$USETECH,
                                    g = gss.2018.cleaned$DEGREE,
                                    p.adj = "bonf")
bonf.tech.by.deg
```

```
##
##      Pairwise comparisons using t tests with pooled SD
##
## data:  gss.2018.cleaned$USETECH and gss.2018.cleaned$DEGREE
##
##                < high school high school junior college college
## high school    3.8e-11       -           -              -
## junior college 2.8e-15       0.0022      -              -
## college        < 2e-16       8.0e-13     1.0000         -
## grad school    < 2e-16       7.3e-09     1.0000         1.0000
##
## P value adjustment method: bonferroni
```

Interpreting Bonferroni results

- Instead of a test statistic like t or F, the output is a matrix of p-values.
- While the calculation of the t-statistic for each pair of groups is the same as other independent samples t-tests, the corresponding p-value is *adjusted* to correct for the multiple statistical tests that could lead to an inflated Type I error.
- Specifically, the Bonferroni adjustment multiplies each p-value from each t-test by the overall number of t-tests conducted.
- There were 10 pairwise comparisons, so these p-values have been multiplied by 10.
- Higher p-values will not reach the threshold for statistical significance.
- A few of the resulting p-values were 1.00000; for p-values that are over 1 when adjusted by the multiplication, they are rounded to exactly 1.
- The adjusted p-values for seven of the t-tests fall below .05 and so indicate that the difference in mean time using technology between two groups is *statistically significant*.
    - There are significant differences in mean time between < *high school* and all of the other groups ($p < .05$).
    - There are significant differences in mean time using technology between *high school* and all other groups.
    - There are no differences among the means of the three college groups.

Reporting Bonferroni results

- Get the means for reporting:

```
# mean age by groups
use.stats
```

```
## # A tibble: 5 x 3
##   DEGREE         m.techuse sd.techuse
##   <fct>              <dbl>      <dbl>
## 1 < high school       24.8       36.2
## 2 high school         49.6       38.6
## 3 junior college      62.4       35.2
## 4 college             67.9       32.1
## 5 grad school         68.7       30.2
```

- Summarize the statistically significant differences identified in the Bonferroni post-hoc test:
    - The mean percentage of time using technology was statistically significantly ($p < .05$) *lower* for people with less educational attainment than a high school diploma ($m = 24.8$) or a high school diploma compared to each of the other groups where the mean percentage of time using technology ranged from 49.6 to 68.7. Likewise, the mean percentage of time using technology was statistically significantly ($p < .05$) lower for people with a high school diploma compared to people with a higher educational attainment and statistically significantly higher than people with less than a high school diploma ($p < .05$).

Post-hoc test: Tukey's Honestly Significant Difference

- The Bonferroni post-hoc test is generally considered a very conservative post-hoc test that only identifies the largest differences between means as statistically significant.
- This might be useful sometimes, but not always.
- When a less strict test will work, Tukey's Honestly Significant Difference (HSD) test is useful.
- Tukey's HSD post-hoc test is a modified t-test with the test statistic, q.

$$q = \frac{m_1 - m_2}{se}$$

Tukey's HSD test statistic

- This is the same exact formula as the t-statistic for the independent samples t-test.
- The q test statistic is the same as the t, but the q-distribution is different from the t-distribution, raising the critical value necessary for the q-statistic to reach statistic significance.
- Even with the same test statistic, it is more difficult to reach statistical significance with a Tukey's HSD q-statistic compared to a t-test.
- Unfortunately, the `TukeyHSD()` function does not work well with the `oneway.test()` output from earlier, so the entire ANOVA model has to be re-estimated.
- The `aov()` function works and takes similar arguments to the `oneway.test()` function, so nesting the `aov()` is one way to go.

```
# Tukey's post-hoc test for tech.by.deg
tukey.tech.by.deg <- TukeyHSD(aov(formula = USETECH ~ DEGREE,
                                  data = gss.2018.cleaned))
tukey.tech.by.deg
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = USETECH ~ DEGREE, data = gss.2018.cleaned)
##
## $DEGREE
##                                          diff        lwr        upr     p adj
```

Comparing Bonferroni and Tukey

- Comparing these results to the Bonferroni test results above, the means that were statistically significantly different with Bonferroni were also statistically significantly different with Tukey's HSD.

- However, the p-values were lower with Tukey's HSD.

- For example, the p-value for the comparison of bachelor to junior college was 1 with the Bonferroni post-hoc test but 0.58 with Tukey's HSD.

- Combine the F-test results and post-hoc results and wrote a single interpretation:

  - The mean time spent on technology use was significantly different across degree groups [$F(4,1404) = 43.3$; $p < .05$] indicating these groups likely came from a population with different mean time spent on technology use depending on educational attainment. The highest mean was 68.6982249% of time used for technology for those with graduate degrees. The lowest mean was 24.7875% of the time for those with less than a high school diploma. Mean percentage of time using technology was statistically significantly ($p < .05$) *lower* for people with less than a high school diploma ($m = 24.8$) compared to each of the other groups where the mean percentage of time using technology ranged from 49.6 to 68.7. Likewise, the mean percentage of time using technology was statistically significantly ($p < .05$) lower for people with a high school diploma compared to people with a higher

Planned comparisons

- Rather than comparing every group to every other group, it might be more interesting to compare all the college groups as a whole to the other groups or the two lowest groups to the two highest groups.

- Bonferroni and Tukey's HSD are not designed to group the groups together and compare these means, but this can be accomplished using planned comparisons.

- **Planned comparisons** could be used to compare one mean to another mean, two means to one mean, or really any subset of means to any other subset of means.

- Planned comparisons are computed by developing **contrasts** that specify which means to compare to which other means.

- For example, to compare all the college groups to high school group, the contrast would omit the less than high school group and compare the mean for everyone in the high school group to the mean of the combined three college groups.

```
# look at the levels of education variable
levels(x = gss.2018.cleaned$DEGREE)
```

```
## [1] "< high school"  "high school"    "junior college" "college"
## [5] "grad school"
```

Writing contrasts

- Rules for writing contrasts:
    - A contrast is a group of numbers used to group categories.
    - The categories grouped together should all be represented by the same number in the contrast.
    - The numbers in the contrast should all add to zero.
    - Any category not included in the contrast should be represented by a zero.
- Comparing the second level of the factor, high school, with the third, fourth, and fifth levels combined could be written as:
    - 0 (do not include)
    - 3 (high school)
    - -1 (junior college)
    - -1 (bachelor)
    - -1 (graduate)
- The three categories represented by -1 will be grouped together because they are all represented by the same number. Adding 0 + 3 + -1 + -1 + -1 is equal to zero.

Entering contrasts into R

- The first step is to enter the contrast into R as a vector.

```
# put the contrast in a vector
contrast1 <- c(0, 3, -1, -1, -1)
```

- The next step is to assign `contrast1` to the `DEGREE` variable using the `contrasts()` function and the `<-` to assign.

```
# link the contrast to the categorical variable using contrasts()
contrasts(x = gss.2018.cleaned$DEGREE) <- contrast1
```

- Use `str()` to examine the *structure* of `DEGREE` and see the contrast:

```
# view the structure of the DEGREE variable with contrast
str(object = gss.2018.cleaned$DEGREE)
```

```
##  Factor w/ 5 levels "< high school",..: 3 2 4 4 5 4 2 2 1 2 ...
##  - attr(*, "contrasts")= num [1:5, 1:4] 0 3 -1 -1 -1 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:5] "< high school" "high school" "junior college" "college
##   .. ..$ : NULL
```

- The second row of the `str()` output showed the contrast, which was one of the attributes of the `DEGREE` variable now.

Run the model and add the contrast

- Once the model has been estimated, the model object can be entered into `summary.aov()` along with information about the contrast.

```
# re-run the model using aov()
techuse.by.deg.aov <- aov(formula = USETECH ~ DEGREE,
            data = gss.2018.cleaned)

# apply the contrasts to the anova object techuse.by.deg.aov
# give the contrast a good descriptive name of "high school vs. college"
tech.by.deg.contr <- summary.aov(object = techuse.by.deg.aov,
                                 split = list(DEGREE = list("high school
tech.by.deg.contr
```

```
##                                   Df  Sum Sq Mean Sq F value    Pr(>F)
## DEGREE                             4  221301   55325   43.30  < 2e-16 ***
##   DEGREE: high school vs. college  1   64411   64411   50.41 1.97e-12 ***
## Residuals                       1404 1793757    1278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 936 observations deleted due to missingness
```

Interpreting the contrast output

- The output showed that mean technology use for those who finished high school was statistically significantly different from mean technology use for the three college groups combined [F(1, 1404) = 50.415; p < .001].

- Create a graph showing the high school group to the three college groups combined.

- To create the graph, use `mutate()` on the `DEGREE` variable so it groups the three college groups into a single group by recoding all three categories into one category called "college."

- Combine categories and examine the means:

```
# recode and compute the means for high school and college groups
gss.2018.cleaned %>%
  mutate(DEGREE = factor(x = DEGREE, labels = c("< high school",
                                                "high school", "college"
                                                "college", "college"))
  group_by(DEGREE) %>%
  summarize(m.techuse = mean(x = USETECH, na.rm = T),
            sd.techuse = sd(x = USETECH, na.rm = T))
```

```
## # A tibble: 3 x 3
##   DEGREE        m.techuse sd.techuse
##   <fct>             <dbl>      <dbl>
## 1 < high school      24.8       36.2
## 2 high school        49.6       38.6
## 3 college            67.0       32.3
```

Examining the combined category means visually

- The difference between the mean technology use time for high school (m = 49.61) compared to college (m = 67.87) is pretty large,
- A graph might help to add more context.

Add another contrast

- The less than high school group is probably also different from the college groups.
- Test both these ideas together using planned comparisons by writing some fancy code.

```
# less than HS v. all college contrast
contrast2 <- c(3, 0, -1, -1, -1)

# bind the two contrasts together into a matrix
cons <- cbind(contrast1, contrast2)

# connect the matrix with the factor variable
contrasts(x = gss.2018.cleaned$DEGREE) <- cons

# estimate the ANOVA with contrasts
tech.by.deg.contr <- summary.aov(object = techuse.by.deg.aov,
                                 split = list(DEGREE = list("high school
                                                            "< high scho
tech.by.deg.contr
```

```
##                                    Df  Sum Sq Mean Sq F value   Pr(>F)
## DEGREE                              4  221301   55325   43.30  < 2e-16 **
##   DEGREE: high school vs. college   1   64411   64411   50.41 1.97e-12 **
##   DEGREE: < high school vs. college 1   20188   20188   15.80 7.39e-05 **
## Residuals                        1404 1793757    1278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 936 observations deleted due to missingness
```

More rules for contrasts

- How many contrasts can be done and how are all these statistical comparisons inflating the Type I error.

- In addition to each comparison comparing two things and each comparison adding to zero, the planned comparisons as a group should isolate each group (e.g., the high school group) *only one time.*

- This ensures that the contrasts are independent of each other since the variance for each group is only used by itself in a statistical comparison one time.

- Because each group is isolated one time, the total maximum number of contrasts allowable is one less than the number of groups.

- Altogether:

    - Contrast values add to zero
    - Each contrast compares two groups
    - Each category should be isolated one time
    - The maximum number of contrasts is one less than the number of categories

How to ensure you are following the rules

(1) Add up each contrast to make sure it adds to zero

(2) Multiply each value in each contrast with the corresponding values in the other contrasts and add up the products; this should also add to zero

- If each contrast adds to zero and the sum of the products across contrasts adds to zero, then the set of contrasts follows the rules for independence and can be used to understand differences among means and groups of means.
- With five total degree groups, there should be four contrasts in the list, like this:
    - < high school v. high school and junior college
    - high school v. all three college groups
    - junior college v. bachelor's and graduate degrees
    - bachelor's v. graduate degree

```
# contrasts for ANOVA of tech time by degree
c1 <- c(2, -1, -1,  0,  0)
c2 <- c(0,  3, -1, -1, -1)
c3 <- c(0,  0,  2, -1, -1)
c4 <- c(0,  0,  0,  -1, 1)

# bind the contrasts into a matrix
```

Check the rules for the contrasts

- In the matrix, the vectors are columns rather than rows.
- The first column was the first contrast that added up (2 + -1 + -1 + 0 + 0) to zero.
- The other three columns all add to zero so the first requirement is met.
- The second requirement was more complicated.
    - Multiply each value across the four contrasts.
    - The first value for the first contrast is 2, the first value of the second contrast is 0, the first value of the third contrast is 0, and the first value of the fourth contrast is also 0.
    - Multiplying $2 \cdot 0 \cdot 0 \cdot 0 = 0$.
    - The product of the second values across the four contrasts is $-1 \cdot 3 \cdot 0 \cdot 0 = 0$.
    - The product of the third values is $-1 \cdot -1 \cdot 2 \cdot 0 = 0$.
    - The fourth values multiply to $0 \cdot -1 \cdot -1 \cdot -1 = 0$.
    - The product across the fifth values in the contrasts is $0 \cdot -1 \cdot -1 \cdot 1 = 0$.
- Adding all these products together results in $0 \cdot 0 \cdot 0 \cdot 0 \cdot 0 = 0$ and the second requirement is met.

Using the set of contrasts

- The set of contrasts was independent and ready to use:

```
# connect the matrix with the factor variable
contrasts(x = gss.2018.cleaned$DEGREE) <- conts

# estimate the ANOVA with 4 independent contrasts
tech.by.deg.contr <- summary.aov(object = techuse.by.deg.aov,
                                 split = list(DEGREE = list("< high scho
                                                            "high school
                                                            "jr college 
                                                            "bachelor's 

tech.by.deg.contr
```

```
##                                                    Df  Sum Sq Mean Sq
## DEGREE                                              4  221301   55325
##   DEGREE: < high school vs. high school & jr college  1   64411   64411
##   DEGREE: high school vs. college                   1   20188   20188
##   DEGREE: jr college vs. more college               1   63902   63902
##   DEGREE: bachelor's vs. grad degree                1   72800   72800
## Residuals                                        1404 1793757    1278
##                                                  F value    Pr(>F)
## DEGREE                                             43.30  < 2e-16 ***
##   DEGREE: < high school vs. high school & jr college  50.41 1.97e-12 ***
##   DEGREE: high school vs. college                  15.80 7.39e-05 ***
##   DEGREE: jr college vs. more college              50.02 2.40e-12 ***
##   DEGREE: bachelor's vs. grad degree               56.98 7.88e-14 ***
## Residuals
```

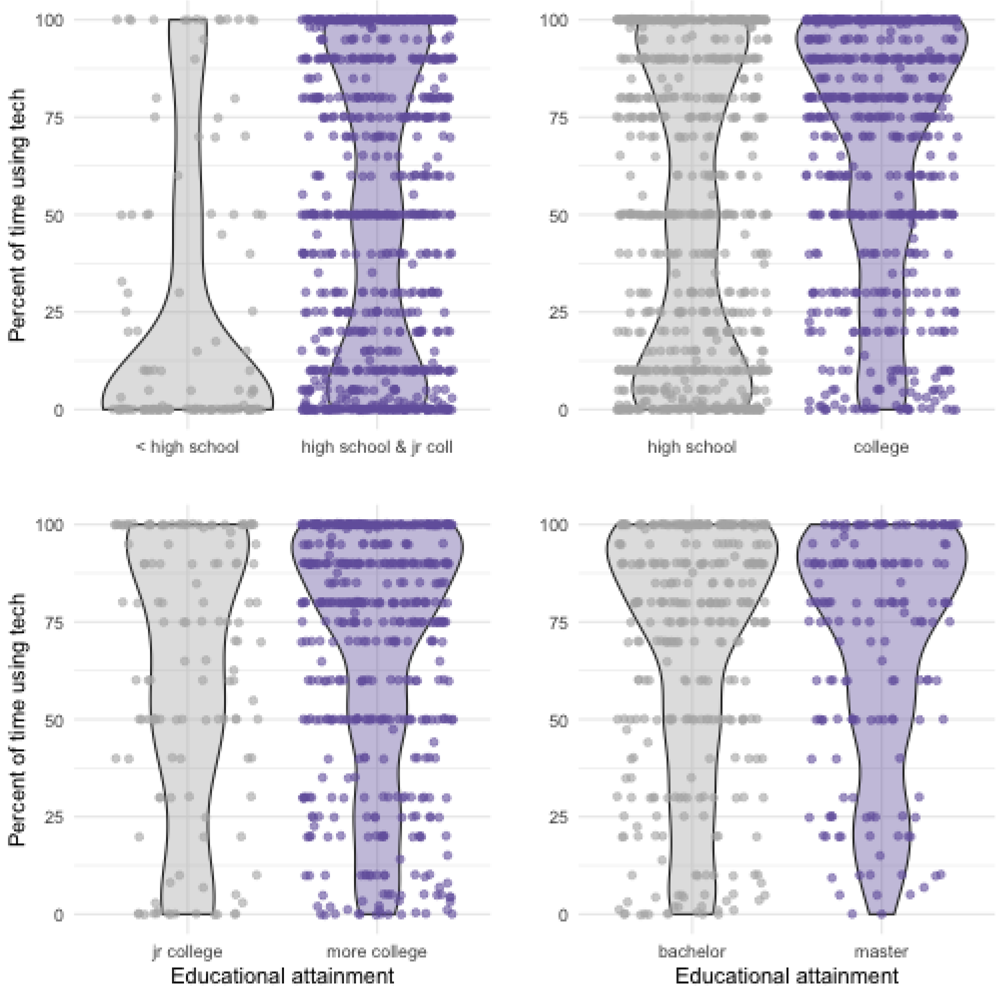
Adjusting the p-values for multiple comparisons

- Since multiple statistical tests inflate the probability of a Type I error, it might be a good idea to apply some sort of correction.
- While this was not an option available in the `aov()` or `summary.aov()` function, there is a `p.adjust()` function that adjusts p-values using one of several types of adjustments.
- The first argument in the `p.adjust()` command is `p =`, which takes a p-value or a vector of p-values.
- The second argument is `method =`, which is where to specify "bonferroni."
- To try it, enter the vector `Pr(>F)` from the `tech.by.deg.contr` object since this is the vector which held all the p-values:

```
# adjust p-values for multiple comparisons
adj.p.vals <- p.adjust(p = tech.by.deg.contr[[1]]$`Pr(>F)`, method = "bo
adj.p.vals
```

```
##
##                                1.329994e-33
## < high school vs. high school & jr college
##                                9.850341e-12
##                     high school vs. college
```

Visualizing all the comparison groups



When to use post-hoc and when to use planned

- When you have hypotheses ahead of time about which groups are different from one another, use planned comparisons.
- When you do not have hypotheses ahead of time about which means are different from each other, use post-hoc tests if the ANOVA has a statistically significant F-statistic.
- Good research practices suggest that having hypotheses ahead of time is a stronger strategy unless the research is truly exploratory.

Getting combined group descriptives to report

```
# contrast 1 statistics
gss.2018.cleaned %>%
  mutate(DEGREE = factor(DEGREE, labels = c("< high school",
                         "high school & jr coll", "high school & jr coll",
                         NA, NA))) %>%
  group_by(DEGREE) %>%
  summarise(m.techuse = mean(x = USETECH, na.rm = T),
            sd.techuse = sd(x = USETECH, na.rm = T))
```

```
## # A tibble: 3 x 3
##   DEGREE                  m.techuse sd.techuse
##   <fct>                       <dbl>      <dbl>
## 1 < high school                24.8       36.2
## 2 high school & jr coll        51.7       38.4
## 3 <NA>                         68.2       31.4
```

```
# contrast 2 statistics
gss.2018.cleaned %>%
  mutate(DEGREE = factor(DEGREE, labels = c(NA,
                         "high school", "any college",
                         "any college", "any college"))) %>%
  group_by(DEGREE) %>%
  summarise(m.techuse = mean(x = USETECH, na.rm = T),
```

Getting combined group descriptives to report

```
# contrast 3 statistics
gss.2018.cleaned %>%
  mutate(DEGREE = factor(DEGREE, labels = c(NA,
                         NA, "jr college",
                         "more than jr college", "more than jr college"))) 
  group_by(DEGREE) %>%
  summarise(m.techuse = mean(x = USETECH, na.rm = T),
            sd.techuse = sd(x = USETECH, na.rm = T))
```

```
## # A tibble: 3 x 3
##   DEGREE               m.techuse sd.techuse
##   <fct>                    <dbl>      <dbl>
## 1 <NA>                      45.8       39.3
## 2 jr college                62.4       35.2
## 3 more than jr college      68.2       31.4
```

```
# contrast 4 statistics
gss.2018.cleaned %>%
  mutate(DEGREE = factor(DEGREE, labels = c(NA,
                         NA, NA,
                         "bachelor", "master"))) %>%
  group_by(DEGREE) %>%
  summarise(m.techuse = mean(x = USETECH, na.rm = T),
```

Writing the final interpretation

- The mean time spent on technology use was significantly different across educational attainment groups [$F(4,1404) = 43$; $p < .05$] indicating these groups likely came from populations with different mean time spent on technology use. The highest mean was grad school% of time used for technology for those with graduate degrees. The lowest mean was < high school% of the time for those with less than a high school diploma. A set of planned comparisons found that the mean time spent using technology was statistically significantly ($p < .05$) lower for (1) those with < *high school* education ($m = 24.8$) compared to those with *high school or junior college* ($m = 51.7$), (2) those with a high school education ($m = 49.61$) compared to those with any college ($m = 67.0$), (3) those with a junior college degree ($m = 62.4$) compared to those with more college than that ($m = 68.2$), and (4) those with a bachelor's ($m = 67.9$) compared to those with a master's degree ($m = 68.7$). Overall the patterns show statistically significant increases in time spent using technology for those with more education.