ANALYSIS OF VARIANCE

RESEARCH PROBLEM

Sometimes we want to know how means differ between ***more than two*** groups

RESEARCH PROBLEM

Want to see how racial categories (black, white, latinx, and other) differ in incomes.

- Examining mean differences b/w three (or more) groups

THE *T*-TEST WAY

Recall…

- *t*-Test examines *mean differences* between *two groups*
  - Compares group 1's mean to group 2's mean

THE *T*-TEST WAY

To compare means of 3 or more groups, can calculate series of *t*-tests

- each comparing two groups at a time…

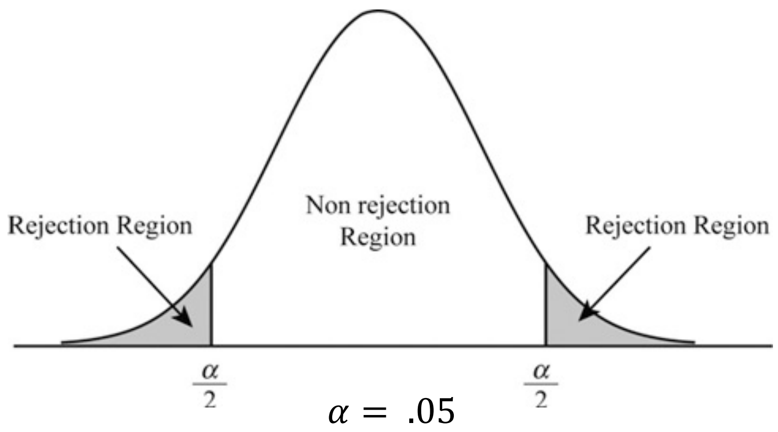
- Compare Incomes between…
  - Black vs White
  - Black vs Latinx
  - Black vs Other
  - White vs Latinx
  - White vs Other
  - Latinx vs Other

When comparing means for 4 groups, there are 6 total comparisons that need to be done

THE *T*-TEST WAY

Could calculate a series of *t*-tests… but…

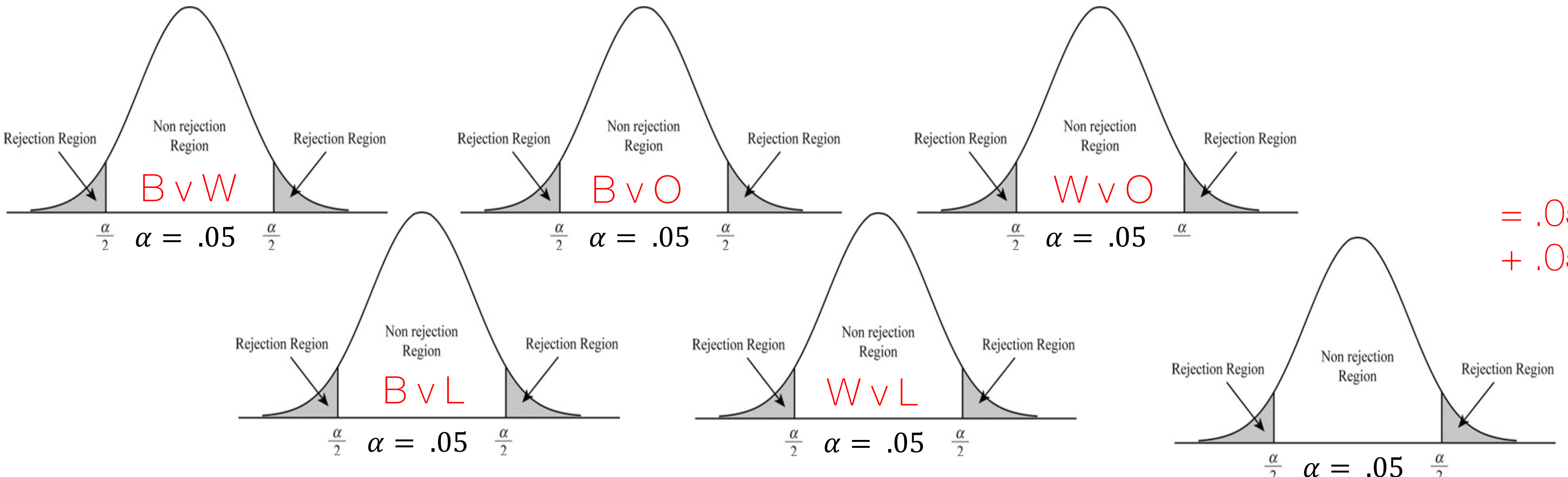
- <u>Each *t*-test</u> uses an alpha of $\alpha = .05$
  - Each comparison has a 5% chance of being wrong/due to sampling error



THE *T*-TEST WAY

Could calculate a series of *t*-tests… but…

- With each comparison, we'd be adding up all the alphas ($\alpha = .05$)
  - So with six comparisons, we'd have a combined alpha of 6*.05 = .30.



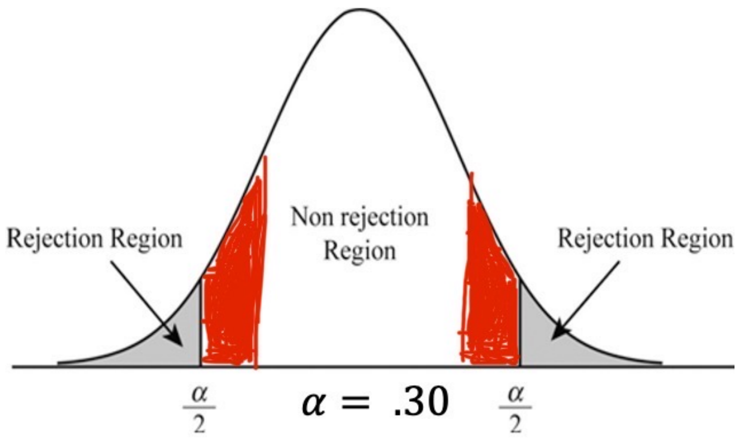
= .05 + .05 + .05 + .05 + .05 + .05 = <u>*.30*</u>

$\alpha = .30$

THE *T*-TEST WAY

Could calculate a series of *t*-tests… but…

- With each comparison, we'd be adding up all the alphas ($\alpha = .05$)
  - So with six comparisons, we'd have a combined alpha of 6*.05 = .30.



- More likely to find significance when it may not exist (Type I Error)
  - More likely to reject null for each *t*, than we would for ONE overall (omnibus) test that holds the sum of alpha at .05

A NEW HOPE

Test that holds alpha constant at $\alpha = .05$, and makes ONE overall comparison

- Divides $\alpha = .05$ by the number of comparisons you have

A NEW HOPE

Analysis of Variance

ANOVA (COMPARED TO *T*-TEST)

Independent Samples T-test (*t*)

- Examines mean differences b/w (compares means of) <u>two groups</u>

Analysis of Variance / ANOVA (*F*)

- Examines mean differences b/w (compares means of) <u>three or more groups</u>

LOGIC OF ANOVA

Total variation across all groups is divisible into two components

- Variation between groups
  - Deviation of group means from one another

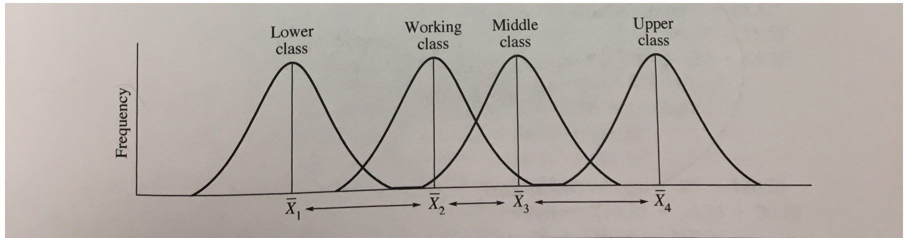


FIGURE 8.2 *Graphic Representation of Variation between Four Groups of Social Class*

- Variation within groups
  - Deviation of raw scores from their group mean

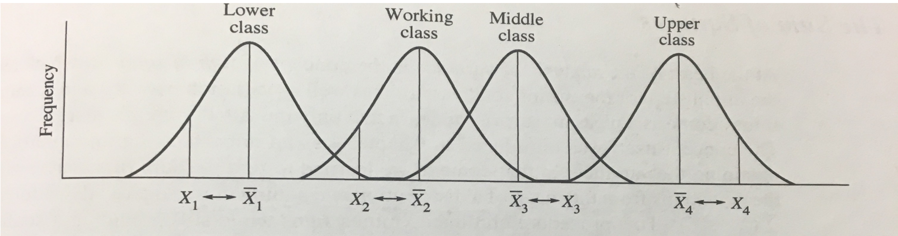


FIGURE 8.1 *Graphic Representation of Variation within Four Groups of Social Class*

LOGIC OF ANOVA

Just like t-test…

- Variation between groups
  - Deviation of group means from one another
- Variation within groups
  - Deviation of raw scores from their group mean

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{SD_{\bar{X}_1 - \bar{X}_2}}$$

- *t* indicates the size of the difference *between* groups relative to the size of the variation *within* each group
- Larger *t* means there is greater variation between groups, and increase the likelihood of rejecting the null hypothesis
- Larger *t* means extreme group differences, beyond what expected by the null hypothesis

LOGIC OF ANOVA

Much like t-ratio, we're interested in variation between groups

- Variation between groups
  - Deviation of group means from each other
- Variation within groups
  - Deviation of raw scores from their group mean

$$ANOVA = \frac{Between}{Within}$$

LOGIC OF ANOVA

Much like t-ratio, we're interested in variation between groups

- Variation between groups
  - Deviation of group means from each other
- Variation within groups
  - Deviation of raw scores from their group mean

$$F = \frac{Between}{Within}$$

- *F* indicates the size of the difference *between* groups relative to the size of the variation *within* each group
- Larger *F* means there is greater variation between groups, and increase the likelihood of rejecting the null hypothesis
- Larger *F* means extreme group differences, beyond what expected by the null hypothesis

ANOVA

One-way ANOVA (*F*)

- ANOVA with one dependent and one independent variable

Other Bivariate ANOVAs

- RMANOVA (Repeated Measures)
  - One dependent variable and one independent variable, but IV groups are timepoints

Multivariate ANOVAs

- Factorial ANOVA
  - One dependent variable and two or more independent variables (refer to the # of groups in each variable: e.g. 3x2 Factorial)
- MANOVA (Multivariate Analysis of Variance)
  - Two or more dependent variables and usually one independent variable (for one-way) or two or more IVs (for two-way, three-way, etc.)
- ANCOVA (Analysis of Covariance)
  - Like regression: One dependent variable, but two or more IVs, one of which is a covariate, which is interval-ratio, not grouping

ANOVA (RESEARCH QUESTION)

Is there a mean difference in Y by categories of X?

- Does the mean (Y) vary by group (X)?

ANOVA (VARIABLE TYPES)

IV: nominal, ordinal (e.g. categorical/discrete)

- Grouping variable
  - Three or more groups/samples compared

DV: interval-ratio (e.g. continuous)

ANOVA (ASSUMPTIONS)

1. Independence of Observations
   - Groups are not related or dependent upon each other. Case can't be in more than one group. No ties between observations

2. Equal Sample (Group) Sizes
   - The number of cases in each group should be relatively similar.

3. Homogeneity of Variance
   - Both groups have approximately equal variances (SD^2). The distributions (or spread) for the groups are approximately equal. Keppel & Zedeck (1989) suggest that variance comparison should not exceed 10:1 ratio.
     - *Examine variances/SD in summary table of output*

4. Normality of Distribution
   - Distribution must be relatively normal
     - *Visual inspection using*
       - *Histogram*
       - *Normality (Q-Q) plots*
       - *Box-and-Whiskers plots*
     - If violated, use "unequal variances assumed" formula, otherwise, use "equal variances assumed"

ANOVA (HYPOTHESES)

Null hypothesis (H_0):

- group means are equal
  - H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 \ldots$

Research hypothesis (H_1)

- group means are not equal
  - H_1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \ldots$

Rejecting null means:

- there is significant variation in group means
- *At least* one group mean is significantly different than the others

CALCULATING ANOVA (*F*)

To compute variation between and variation within, we use the sum of the squared deviations from the mean (e.g. "Sum of Squares")

$$F = \frac{Between}{Within}$$

CALCULATING ANOVA (F)

$$SS_{within} = \sum (X - \bar{X}_{group})^2$$

$$SS_{between} = \sum N_{group}(\bar{X}_{group} - \bar{X}_{total})^2$$

$$SS_{total} = SS_{between} + SS_{within}$$

CALCULATING ANOVA (*F*)

Within-Group Sum of Squares

- The mean difference in scores within a single sample, then sums them up
- Amount of unexplained variance

$$SS_{within} = \sum (X - \bar{X}_{group})^2$$

$$SS_{within} = SS_{total} - SS_{between}$$

CALCULATING ANOVA (*F*)

Between-Group Sum of Squares

- The mean difference between all groups

$$SS_{between} = \sum N_{group}(\bar{X}_{group} - \bar{X}_{total})^2$$

$$SS_{between} = SS_{total} - SS_{within}$$

CALCULATING ANOVA (F)

Total Sum of Squares

- Sum of the within and between sum of squares
- Total variation in scores

$$SS_{total} = \sum (X - \bar{X}_{total})^2$$

$$SS_{total} = SS_{between} + SS_{within}$$

CALCULATING ANOVA (F)

Because Sum of Squares is a measure of variation, you may think that, to calculate F, you may want to do the following: $\frac{SS_{between}}{SS_{within}}$

CALCULATING ANOVA (*F*)

But… Sum of Squares increases as sample size increases, so number will be HUGE with increased numbers of cases

To fix, need to adjust Sum of Squares by controlling for the number of scores included in the calculations

- Divide each Sum of Squares by the degrees of freedom
- This is known as the Mean Square

CALCULATING ANOVA (*F*)

To calculate <u>Mean Square Between</u> and <u>Mean Square Within</u>, we divide both by their respective degrees of freedom

$$MS_{between} = \frac{SS_{between}}{df_{between}}$$

$$MS_{within} = \frac{SS_{within}}{df_{within}}$$

CALCULATING ANOVA (*F*)

To calculate degrees of freedom Between and degrees of freedom Within, we use a similar logic (e.g. n-1) for each, in the following calculations…

$$df_{between} = k - 1$$

(k is number of groups)

$$df_{within} = N_{total} - k$$

CALCULATING ANOVA (*F*)

$$MS_{between} = \frac{SS_{between}}{df_{between}}$$

$$MS_{within} = \frac{SS_{within}}{df_{within}}$$

$df_{between} = k - 1$ where k is number of groups

$$df_{within} = N_{total} - k$$

CALCULATING ANOVA (*F*)

ANOVA statistic (*F*) is based on a ratio of mean square between divided by mean square within

$$F = \frac{MS_{between}}{MS_{within}} = \frac{\frac{SS_{between}}{df_{between}}}{\frac{SS_{within}}{df_{within}}}$$

HYPOTHESIS TESTING WITH ANOVA

Is the variation (difference) between group means “extremely” different from what is expected by chance/the null hypothesis?

F-TEST AND THE *F*-DISTRIBUTION

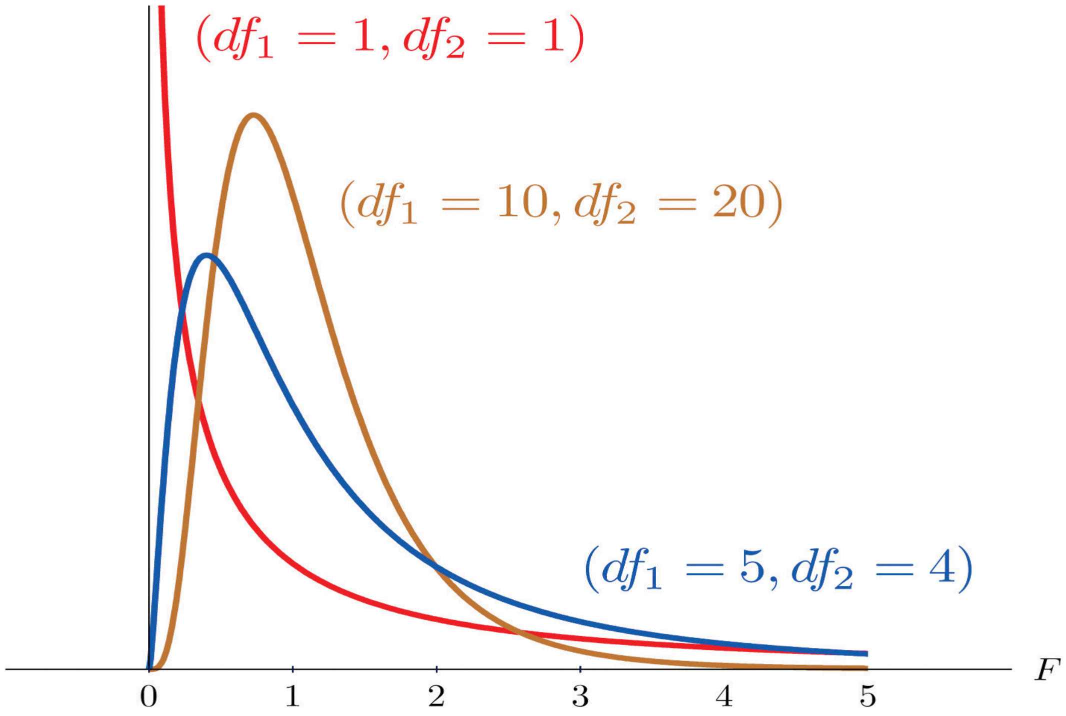
The F-distribution has multiple curves

- Each curve based on sample size or degrees of freedom

$$F = \frac{MS_{between}}{MS_{within}}$$

$$df_{between} = k - 1$$

$$df_{within} = N_{total} - k$$



IS THE *F* EXTREME?

Recall the *t*-Test:

- If $|t_{obtained}| \geq |t_{critical}|$, then…
  - Difference between mean of group 1 and mean of group 2 is so extreme that we can't blame it on sampling error, therefore… H_0 is probably not true, so…
  - $t_{obtained}$ is in rejection region
  - Reject H_0
  - $p \leq \alpha$
  - Statistically significant difference

IS THE *F* EXTREME?

Similar for *F*:

- If $|F_{obtained}| \geq |F_{critical}|$, then…
  - Overall differences between group means is so extreme that we can't blame it on sampling error, therefore… H_0 is probably not true, so…
  - $F_{obtained}$ is in rejection region
  - Reject H_0
  - $p \leq \alpha$
  - Statistically significant difference between group means

IS THE *F* EXTREME?

But because we square the differences, we only have positive values…

- If $F_{obtained} \geq F_{critical}$, then…
  - Overall differences between group means is so extreme that we can't blame it on sampling error, therefore… H_0 is probably not true, so…

  - $F_{obtained}$ is in rejection region
  - Reject H_0
  - $p \leq \alpha$
  - Statistically significant difference between group means

IS THE *F* EXTREME?

F has multiple distributions, based on two *df*s and α:

- Because of SS, F will always be positive
  - Thus, distribution only has one-tail

- We must refer to a table ([*F* Table]) to figure out what F_{critical} is.
  - Your $\underline{\alpha}$
    - Usually $\alpha = .05$
  - Your df_{between} and df_{within}
    - $df_{\text{between}} = K - 1$
    - $df_{\text{within}} = N_{total} - K$
  - Where they intersect is our F_{critical}

- Then, evaluate to see if $F_{\text{obtained}} \geq F_{\text{critical}}$

IS THE *F* EXTREME?

Hypothesis Testing:

- If $F_{obtained} \geq F_{critical}$, reject the null hypothesis
- If $F_{obtained} < F_{critical}$, fail to reject the null hypothesis

REPORTING ANOVA (*F*)

Report

- The test used
- If you reject or fail to reject the null hypothesis
- The variables used in the analysis
- The degrees of freedom, calculated value of the test, and p-value
  - $F(df_{between}, df_{within}) = F_{obtained}$, p-value

- "Using a one-way ANOVA, I reject/fail to reject the null hypothesis that there is no mean difference in dependent variable across groups of the independent variable, in the population F(?,?) = ?, *p* ? .05"

EXAMPLE: MONTHLY INCOME BY RACE

Does income (dollars per month) vary by race?

- Is there a significant difference between mean monthly income for people of different racial categories?
  - Black, White, Latinx, Other

EXAMPLE: MONTHLY INCOME BY RACE

Race	Income
black	3670
black	3132
black	3798
black	3929
black	3203
white	6628
white	4702
white	4264
white	4063
white	4847
latinx	2239
latinx	3143
latinx	4580
latinx	3158
latinx	3471
other	3131
other	4462
other	3901
other	3410
other	4927

EXAMPLE: MONTHLY INCOME BY RACE

Race	Income
black	3670
black	3132
black	3798
black	3929
black	3203
white	6628
white	4702
white	4264
white	4063
white	4847
latinx	2239
latinx	3143
latinx	4580
latinx	3158
latinx	3471
other	3131
other	4462
other	3901
other	3410
other	4927

	black	white	latinx	other
	3670	6628	2239	3131
	3132	4702	3143	4462
	3798	4264	4580	3901
	3929	4063	3158	3410
	3203	4847	3471	4927
n	5	5	5	5
group mean	3546.4	4900.8	3318.2	3966.2
grand mean	3932.9			

$$SS_{within} = \sum (X - \bar{X}_{group})^2$$

$$SS_{between} = \sum N_{group}(\bar{X}_{group} - \bar{X}_{total})^2$$

$$SS_{total} = SS_{between} + SS_{within}$$

EXAMPLE: MONTHLY INCOME BY RACE

	black	white	latinx	other
	3670	6628	2239	3131
	3132	4702	3143	4462
	3798	4264	4580	3901
	3929	4063	3158	3410
	3203	4847	3471	4927
n	5	5	5	5
group mean	3546.4	4900.8	3318.2	3966.2
grand mean	3932.9			

$$SS_{within} = \sum (X - \bar{X}_{group})^2$$

$$SS_{between} = \sum N_{group}(\bar{X}_{group} - \bar{X}_{total})^2$$

$$SS_{total} = SS_{between} + SS_{within}$$

	black (X-group mean)^2	white (X-group mean)^2	latinx (X-group mean)^2	other (X-group mean)^2
Within Group SS	15276.96	2983219.84	1164672.64	697559.04
	171727.36	39521.44	30695.04	245817.64
	63302.56	405514.24	1592139.24	4251.04
	146382.76	701908.84	25664.04	309358.44
	117923.56	2894.44	23347.84	923136.64
SS within	9664313.6			

	black (group mean - grand mean)^2	white (group mean - grand mean)^2	latinx (group mean - grand mean)^2	other (group mean - grand mean)^2
Between Group SS	746911.25	4684152.05	1889280.45	5544.45
SS between	7325888.2			

EXAMPLE: MONTHLY INCOME BY RACE

N_{total} = 20

k = 4

$SS_{between}$ = 7,325,888

SS_{within} = 9,664,314

EXAMPLE: MONTHLY INCOME BY RACE

$df_{between} = k - 1 = 4 – 1 = 3$

$df_{within} = N - k = 20 – 4 = 16$

EXAMPLE: MONTHLY INCOME BY RACE

$$\text{MS}_{\text{between}} = \frac{SS_{between}}{df_{between}} = \frac{7{,}325{,}888}{3} = 2{,}441{,}963$$

$$\text{MS}_{\text{within}} = \frac{SS_{within}}{df_{within}} = \frac{9{,}664{,}314}{16} = 604{,}020$$

$$F = \frac{MS_{between}}{MS_{within}} = \frac{2{,}441{,}963}{604{,}020} = 4.043$$

EXAMPLE: MONTHLY INCOME BY RACE

Using our two df values and our obtained F, we can determine whether our F, representing size of the mean differences between groups, is significant

$F_{obtained} = 4.043$

$df_{between} = 3$

$df_{within} = 16$

- In our Appendix F ($\alpha = .05$), we follow the $df_{between} = 3$ column and $df_{within} = 16$ row, we get:

$F_{critical} = 3.24$

EXAMPLE: MONTHLY INCOME BY RACE

$F_{\text{obtained}} = 4.043$

$df_{\text{between}} = 3$

$df_{\text{within}} = 16$

$F_{\text{critical}} = 3.24$

- Because $F_{\text{obtained}} > F_{\text{critical}}$, our mean differences between groups is extreme enough to fall in the rejection region. Therefore…
  - F_{obtained} is in rejection region
  - Reject H_0
  - $p < .05$
  - Statistically significant difference between group means

- (at least) one of our group means is significantly different than others

EXAMPLE: MONTHLY INCOME BY RACE

Report

- The test used
- If you reject or fail to reject the null hypothesis
- The variables used in the analysis
- The degrees of freedom, calculated value of the test, and p-value
  - $F(df_{between}, df_{within}) = F_{obtained}$, p-value

- "Using a one-way ANOVA, I reject the null hypothesis that there is no mean difference in income across racial categories, in the population $F(3,16) = 4.043$, $p < .05$"

HOW MUCH OF THE MEAN DIFFERENCE IS EXPLAINED BY GROUP MEMBERSHIP?

Much like correlation, where r-squared (r^2) gives us percent of variance in outcome (Y) that is explained by variance in the predictor (X), we can get the same with eta-squared (η^2). This is known as effect size.

$$\eta^2 = r^2 = \frac{SS_{effect}}{SS_{total}} = \frac{SS_{between}}{SS_{total}}$$

EXAMPLE: MONTHLY INCOME BY RACE

Eta-squared (η^2) gives us percent of variance in outcome (Y) that is explained by variance in the predictor (X)

$$\eta^2 = \frac{SS_{between}}{SS_{total}} = \frac{SS_{between}}{SS_{between} + SS_{within}} = \frac{7{,}325{,}888}{7{,}325{,}888 + 9{,}664{,}314} = 0.4311831$$

The effect size is $\eta^2 = .4312$. This means that 43.12% of the variation in monthly income can be explained by someone's race.

WHAT IS THE STRENGTH OF THE EFFECT?

Much like correlation, where r ($\boldsymbol{r}$), using the Cohen (1988) cutoffs, gives us strength or magnitude of the effect we can get the same with eta ($\boldsymbol{\eta}$).

Small

- *r* less than/equal to |.29| ($r \leq |.29|$)

Moderate

- *r* between |.30| and |.49| ($|.30| \leq r \leq |.49|$)

Large

- *r* greater than/equal to |.50| ($r \geq |.50|$)

$$\eta = r = \sqrt{\eta^2} = \sqrt{r^2}$$

EXAMPLE: MONTHLY INCOME BY RACE

Eta (η) gives us strength or magnitude of the effect.

$$\eta = r = \sqrt{.4311831} = .6566453$$

This indicates that there is a large effect between the variables

SIGNIFICANT DIFFERENCES, WHERE?

Remember, ANOVA is an omnibus (overall) test of differences

If we find significant *F*, at least ONE group mean is significantly different than others

- It does not mean that every group mean is significantly different from all others.

SIGNIFICANT DIFFERENCES, WHERE?

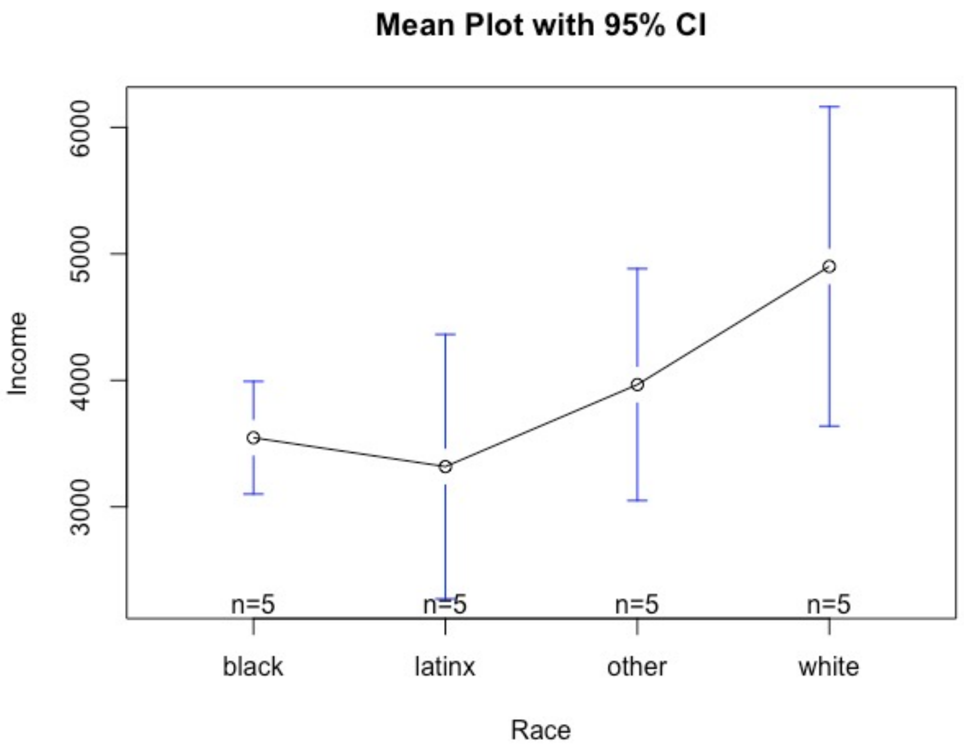
After finding significant F, next step is to identify ***which*** group means are significantly different others

SIGNIFICANT DIFFERENCES, WHERE?

Plotting means will help identify where differences lie

EXAMPLE: MONTHLY INCOME BY RACE

Plotting means will help identify where differences lie



SIGNIFICANT DIFFERENCES, WHERE?

Planned Comparisons/Multiple Comparison Procedures:

- <u>Conducting a series of t-tests (mean comparisons)</u> to see where the true mean differences lie

SIGNIFICANT DIFFERENCES, WHERE?

Can run many t-tests (two-group comparisons)… but run into the same problem… increasing alpha/Type I error.

SIGNIFICANT DIFFERENCES, WHERE?

Family-Wise Error Rate (FWER):

- Likelihood/probability of committing Type I error if conducted all possible two-group comparisons were considered after ANOVA.
- Slightly different (adjusted) from simply adding up alpha for each comparison

$$FWER = 1 - (1 - \alpha)^c$$

- Where α is your alpha level for the ANOVA, and c is the number of total comparisons across all groups.
  - 3 groups == 3 comparisons: $FWER = 1 - (1 - .05)^3$ = .143 = 14.3% chance of Type I error
  - 4 groups == 6 comparisons: $FWER = 1 - (1 - .05)^6$ = .265 = 26.5% chance of Type I error
  - 5 groups == 10 comparisons: $FWER = 1 - (1 - .05)^{10}$ = .401 = 40.1% chance of Type I error

SIGNIFICANT DIFFERENCES, WHERE?

Because every FWER is larger than accepted alpha ($\alpha = .05$), if conducting multiple comparisons, <u>must adjust the alpha</u> for all comparisons to sum to $\alpha = .05$

Bonferroni's Alpha Adjustment:

$$\alpha_{adj} = \frac{\alpha}{c}$$

- Where α is your desired alpha level ($\alpha = .05$), and c is the number of total comparisons across all groups.

SIGNIFICANT DIFFERENCES, WHERE?

Post Hoc Tests, in order of most to least stringent:

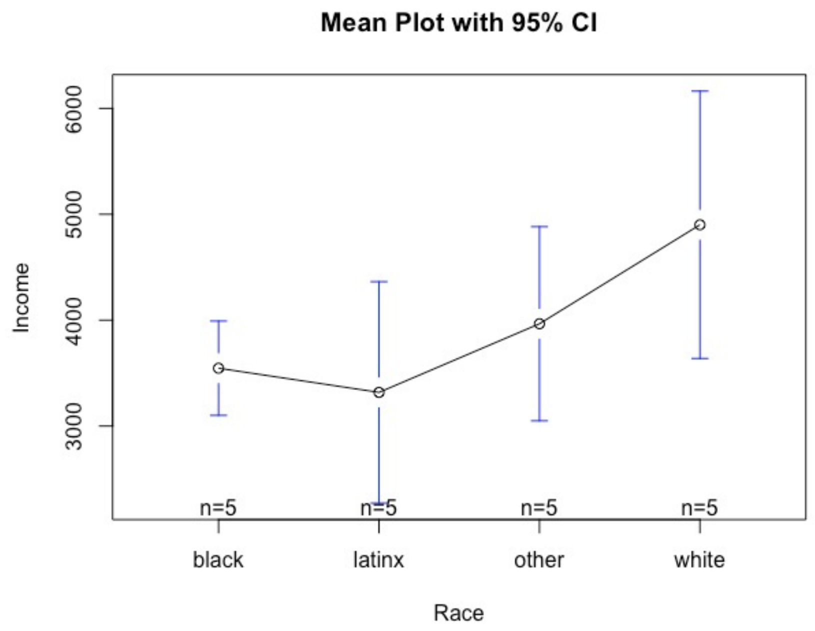
- Scheffé's Test
- Tukey's HSD (Honestly Significant Difference)
- Newman-Keuls Test
- Fisher's LSD (Least Significant Difference)
- (Dunnet's Test, rarely used)

- All incorporate the Bonferroni adjustment, therefore significance can be read in relation to our traditional alpha level ($\alpha = .05$).

SIGNIFICANT DIFFERENCES, WHERE?

Tukey's HSD

Group	Difference	*p* (adjusted)
latinx-black	-228.2	0.9657923
other-black	419.8	0.8279317
white-black	1354.4	0.0611291
other-latinx	648	0.5651327
white-latinx	**1582.6**	**0.0248001**
white-other	934.6	0.2662779



- We can see from the results of the Tukey HSD test that the only real difference (or the mean difference that is holding the significant ANOVA) is the comparison between white's mean income and latinx' mean income.