One-Way ANOVA

Cohen Chapter 12

EDUC/PSY 6600

“It is easy to lie with statistics.
It is hard to tell the truth
without statistics.”

-Andrejs Dunkels

Motivating examples

- Dr. Vito randomly assigns 30 individuals to 1 of 3 study groups to evaluate whether one of **2 new approaches** to therapy for adjustment disorders with mixed anxiety and depressed mood are more effective than the **standard approach**. Participants are matched on current levels of anxiety and depressed mood at baseline. Scores from the BAI and BDI are collected after 2 months of therapy.

- Dr. Creft wishes to assess differences in oral word fluency **among three groups of participants**: Right hemisphere stroke, left hemisphere stroke, and healthy controls. Scores on the COWAT are collected from 20 participants per group and the means of each group are compared.

Research Design Vocab

- ***Experimental design***
  - Participants are randomly **assigned** to levels and at least one factor is **manipulated**
  - Participants are randomly selected from multiple **preexisting (observed)** populations
- ***Fixed or random effects***
  - **Fixed** effects design: Levels of each factor systematically chosen by researcher
  - **Random** factors design: Levels of each factor are chosen randomly from a larger subset (rarer)
- ***Independent** (Between-Subjects) or **Repeated** (Within-Subjects) factors*
  - **Independent**: Participants randomly allocated to each level of a factor
  - **Repeated** measures design: Participants are paired or a dependency exists (multiple observations)

Research Design Vocab

- ***Experimental design***
  - Participants are randomly **assigned** to levels and at least one factor is **manipulated**
  - Participants are randomly selected from multiple **preexisting (observed)** populations

If the levels of the grouping variable are **highly ordinal or continuous** in nature, **regression** or a rank type test will be more powerful than ANOVA

- ANOVA is appropriate in cases where the groups are more nominal in nature.

Some variables can be construed as both!!! (e.g. Grade level)

- *probably want to analyze both ways*

Analysis of Variance (ANOVA)

- ANOVA designs can be used for…
  - Experimental research
  - Quasi-experimental studies
  - Field/observational research

- *Other names for 1-way ANOVA…*
  - Single factor ANOVA
  - *Univariate ANOVA*
  - Simple ANOVA
  - Independent-ANOVA
  - Between-subjects ANOVA

Omnibus test for group (MEAN) differences

ONE Dependent Variable (DV)
"outcome"

Continuous (interval/ratio)
&
normally distributed

ONE Independent Variable (IV)
"predictor"

Categorical (nominal)
$\geqslant$ 3 independent samples or groups
Factor with *k levels*

Example: noise & words memorized

Group		
A	B	C
8	7	4
10	8	8
9	5	7
10	8	5
9	5	7

- Study to determine if noise inhibits learning ($N = 15$)
- Students **randomized** to 1 of 3 groups (k = 3 & $n = 5$)
  - IV = grouping factor with 3 levels
    - Group A: **No** noise (no music, quiet room)
    - Group B: **Moderate** noise (classical music)
    - Group C: **Extreme** noise (rock music)
- Participants are given 1 minutes to memorize list of 15 nonsense words
  - DV = # of correct nonsense words recalled

Steps of a Hypothesis test

1) State the **Hypotheses** (Null & Alternative)
2) Select the **Statistical Test** & Significance Level
   - *Examples include: z, t, F,* χ^2
   - α *level (commonly use .05)*
   - *One vs. Two tails (usually prefer 2)*
3) Select random **samples** and **collect** data
4) Find the region of **Rejection**
   - *Based on* α *& # of tails*
5) Calculate the **Test Statistic**
   - *Select the appropriate formula*
   - *May need to find degrees of freedom*
6) Make the Statistical **Decision**

Hypotheses of ANOVA

- Means: $\mu_1, \mu_2, \mu_3, \ldots, \mu_k$
- Variances: $\sigma_1^2, \sigma_2^2, \sigma_3^2, \ldots, \sigma_k^2$

$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$
H_1: Not H_0

Many ways to reject H_0

NOT $H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_k$

Example: Noise & Words Memorized

Null Hypothesis:

The number of words recalled is the same regardless of the music/noise.

$H_0: \mu_{none} = \mu_{moderate} = \mu_{extreme}$

Alternative Hypothesis:

At least one music/noise level results in a **different** number of words recalled.

H_1: Not H_0

Example: noise & words memorized

Group		
A	B	C
8	7	4
10	8	8
9	5	7
10	8	5
9	5	7

1. Enter data into R

```
# A tibble: 15 x 2
     outcome group
       <dbl> <fct>
1         8.  A
2        10.  A
3         9.  A
4        10.  A
5         9.  A
6         7.  B
7         8.  B
8         5.  B
9         8.  B
10        5.  B
11        4.  C
12        8.  C
13        7.  C
14        5.  C
15        7.  C
```

2. Calculate the Group Means

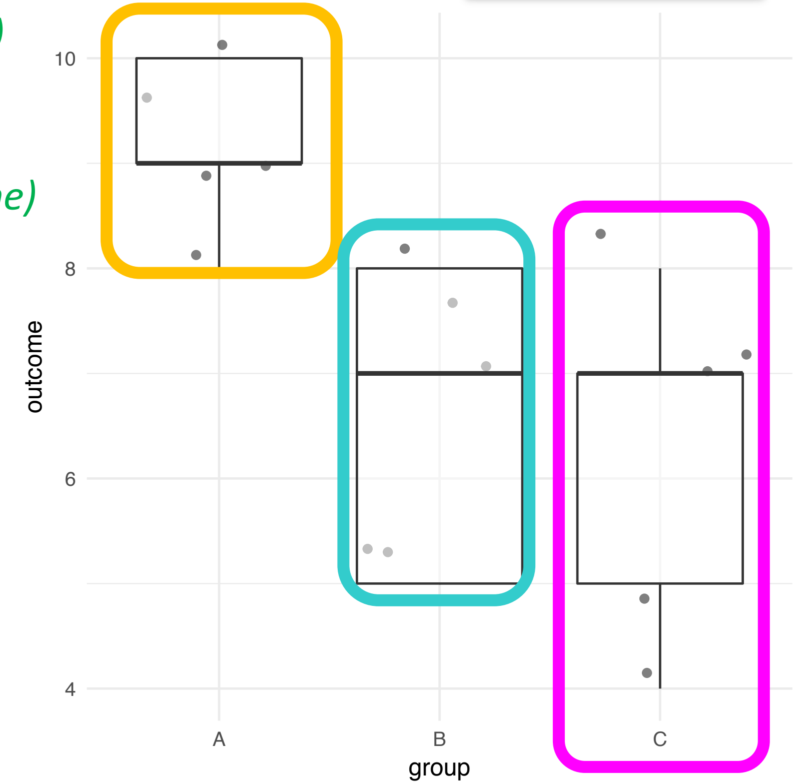
```
df %>%
  group_by(group) %>%
  furniture::table1(outcome)
```

IV (groups)

DV (outcome)

	group		
	A	B	C
	n = 5	n = 5	n = 5
outcome			
	9.2 (0.8)	6.6 (1.5)	6.2 (1.6)

3. Visualize the data



Link: Independent sample "t-test" & ANOVA

Same principle underlies many statistical tests

$$F = \frac{MS_B}{MS_W} = \frac{\text{Measure of effect (or treatment) assessed by examining variance (or differences) between groups}}{\text{Measure of random variation (or error) assessed by examining variance (or differences) within groups}}$$

$$\text{Stats} = \frac{\text{Stuff we can explain with our variables}}{\text{Stuff we CANNOT explain with our variables (random error)}}$$

<u>Numerator</u>

MS_B : Compute variance <u>between</u> (among) sample means, multiply by n_j

<u>Denominator</u>

MS_W : Compute average of sample variances

Link: Independent sample “t-test” & ANOVA

- Same question as before…
  - **Do group means significantly differ?**
  - Or…Do mean differences on DV ‘<u>between</u>’ groups **EXCEED** differences ‘<u>within</u>’ groups?
    - **Between**-groups differences
      - Differences in DV **due to IV (group)**
    - **Within**-groups differences
      - Differences in DV **due to pooled random error or variation**

- Same analysis approach as before…

Link: Independent sample "t-test" & ANOVA

- Same question as before…
  - **Do group means significantly differ?**
  - Or…Do mean differences on DV 'between' groups **EXCEED** differences 'within' groups?
    - **Between**-groups differences
      - Differences in DV **due to IV (group)**
    - **Within**-groups differences
      - Differences in DV **due to pooled random error or variation**
- Same analysis approach as before…

$$\boxed{\boldsymbol{F = t^2}}$$

$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Link: Independent sample "t-test" & ANOVA

- Same question as before…
  - **Do group means significantly differ?**
  - Or…Do mean differences on DV 'between' groups **EXCEED** differences 'within' groups?
    - **Between**-groups differences
      - Differences in DV **due to IV (group)**
    - **Within**-groups differences
      - Differences in DV **due to pooled random error or variation**
- Same analysis approach as before…

$$\boxed{F = t^2}$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Could rewrite as… $t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{s_p^2\left(\frac{2}{n_j}\right)}}$,

Where n_j = sample size for any group j. Then....

Link: Independent sample "t-test" & ANOVA

- Same question as before…
  - **Do group means significantly differ?**
  - Or…Do mean differences on DV '<u>between</u>' groups **EXCEED** differences '<u>within</u>' groups?
    - **Between**-groups differences
      - Differences in DV **due to IV (group)**
    - **Within**-groups differences
      - Differences in DV **due to pooled random error or variation**
- Same analysis approach as before…

$$\boldsymbol{F = t^2}$$

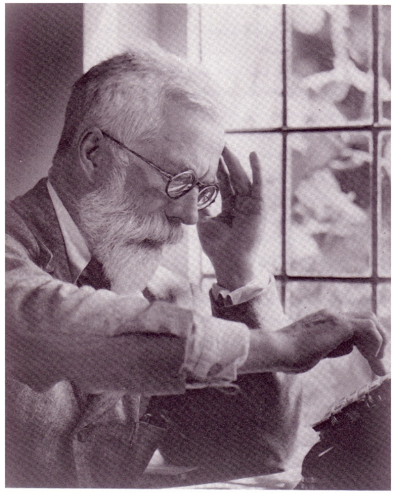
$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Could rewrite as… $t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{s_p^2\left(\frac{2}{n_j}\right)}}$,

Where n_j = sample size for any group j. Then….

$$t^2 = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{2s_p^2}{n_j}} = \frac{n_j(\bar{X}_1 - \bar{X}_2)^2}{2s_p^2} = \boxed{\frac{n_j\frac{(\bar{X}_1 - \bar{X}_2)^2}{2}}{s_p^2} = F}$$

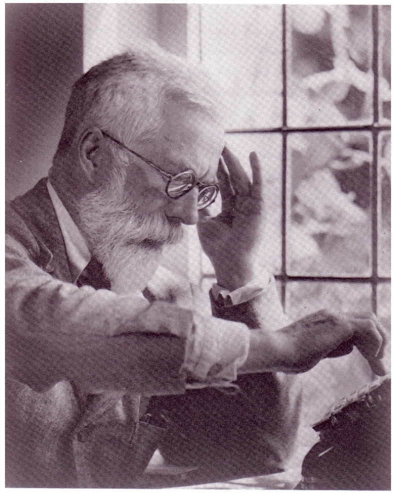
F-distribution


Fisher at his desk calculator at Whittingehame Lodge, 1952

Sir **Ronald A. Fisher** (1920-40's) & **agricultural** experiments…

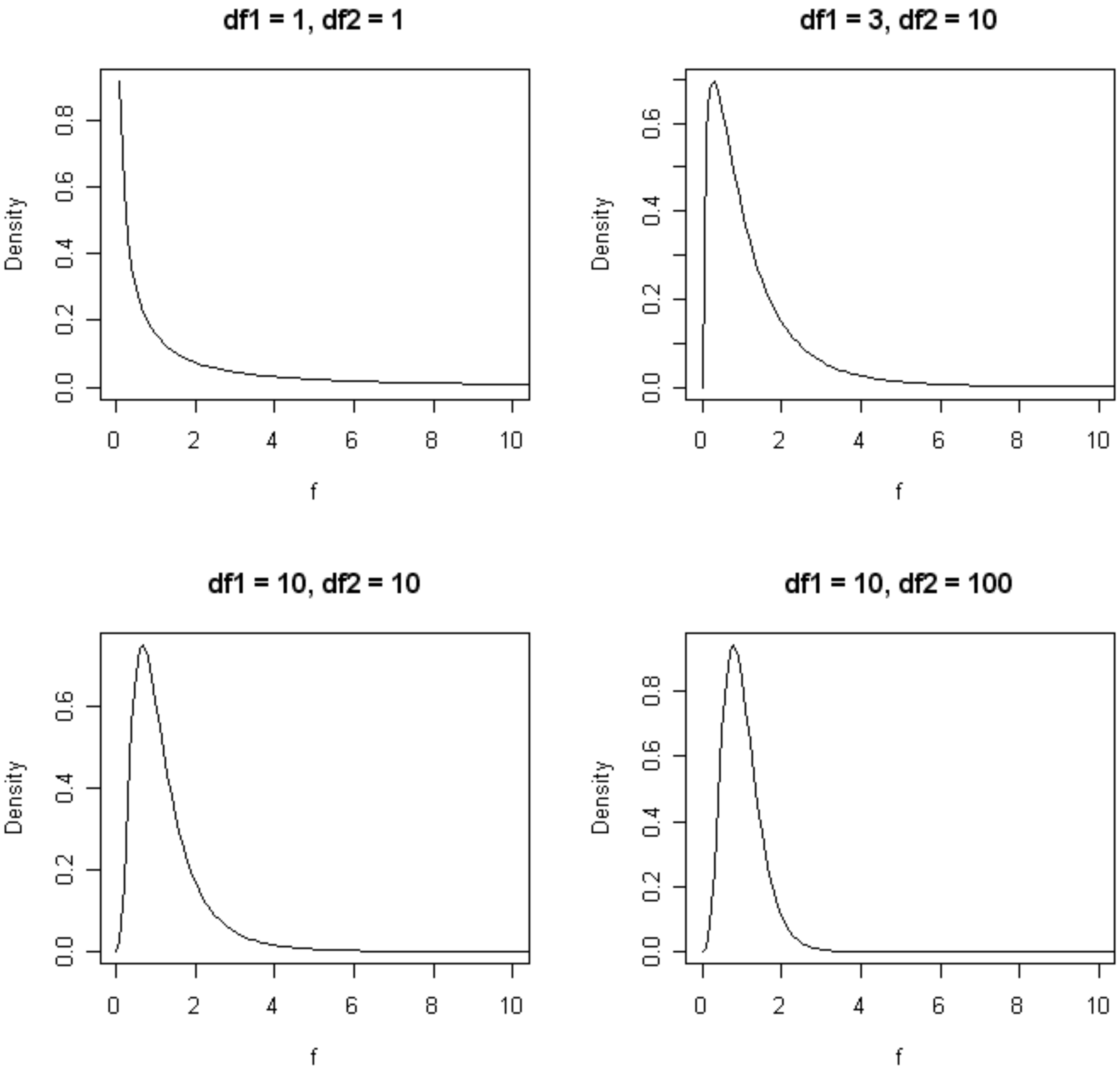
- *F*-distribution
  - Continuous theoretical probability distribution
  - Probability of **ratios** (fraction) of variance **between** groups to variance **within** groups
- Positively skewed
  - Range: 0 to ∞
  - one-tailed
  - More "normal" as $N \uparrow$
  - Mean ≈1… $M = \frac{df_W}{df_{W-2}}$
- Family of distributions
  - Need **2 *df*** and α to determine F_{crit}
    - df_{Within} and $df_{Between}$ (more later…)

$$\boxed{F = t^2}$$



Fisher at his desk calculator at Whittingehame Lodge, 1952

Sir **Ronald A. Fisher** (1920-40’s) & **agricultural** experiments…



Link: Independent sample "t-test" & ANOVA

Specific situation: **2 groups, when $n_1 = n_2$**

Numerator: Variation **between** (among) group **means**

'Variance' of 2 means multiplied by n_i

Mean Square Between (MS_B) or Mean Square Treatment (MS_T)

$$t^2 = \frac{n_j \dfrac{\left(\bar{X}_1 - \bar{X}_2\right)^2}{2}}{s_p^2} = F$$

Denominator: Pooled variation **within groups**

Pooled variance (s_p^2) = average of 2 variances when *n*s are equal

Mean Square Within (MS_W) or Mean Square Error (MS_E)

Link: Independent sample "t-test" & ANOVA

'Mean Square' or ***MS***
is another term for the **variance**

'Square': Refers to the sum of SQUARED (*SS*) deviations from the mean

Mean: AVERAGE of the *SS* deviations

SS is divided by *N* or *N* - 1 to yield variance
So, Mean of the sum of SQUARED deviations = Variance

All we want to know is whether variation among group means exceeds that variation within groups

Will create a **ratio of the *MS*s**, the *F*-statistic, to see if this ratio is significantly **different from 1**

Prior example

- Applying data from independent-samples *t*-test example
- (drug *v.* placebo and depression)
  - Recall, $t = 1.96$, $p = .085$

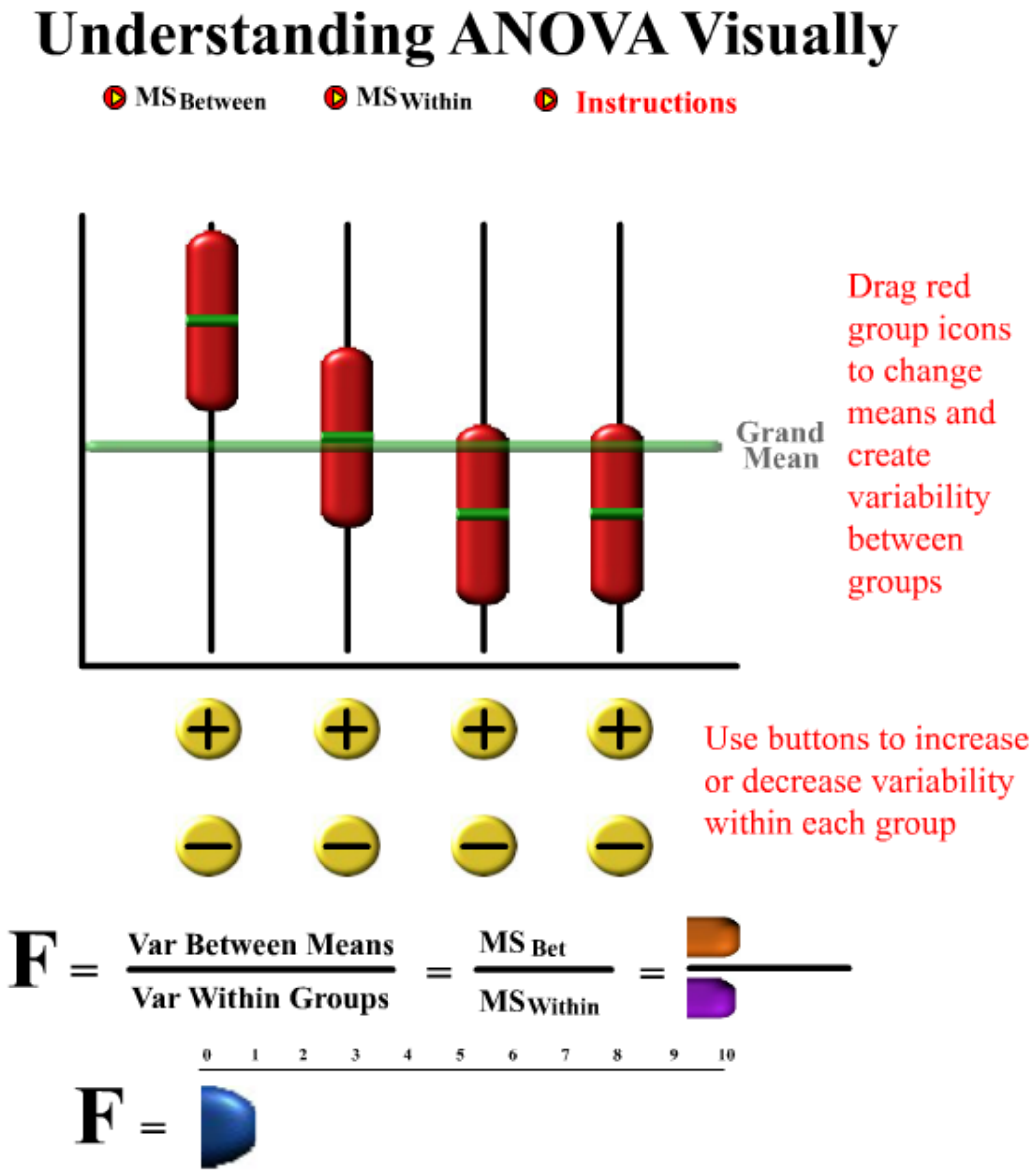
$1.96^2 = 3.84$

$$\boxed{t^2 = F}$$

Group 1 - Drug	Group 2 - Placebo
11	11
1	11
0	5
2	8
0	4

$$t^2 = \frac{n_j \dfrac{\left(\bar{X}_1 - \bar{X}_2\right)^2}{2}}{s_p^2}$$

Interactive Applet



- http://web.utah.edu/stat/introstats/anovaflash.html

Assumptions

Large or multiple violations will GREATLY increase risk of inaccurate *p*-values
Increased probability of Type I or II error

Independent, Random Sampling (for the IV) ← ensure by planning ahead!

- For **preexisting** (observed) populations: randomly select a sample from each population
- For **experimental** (assigned) conditions: randomly divide your sample *(of convenience)* for assignment to groups
- Ensure no connection between subjects in the different groups (no matching!) ← MUST!!!

Normally distributed (DV)

- Robust requirement…if samples are large, this isn't as important
- If not normal (or small samples)
  - alternatives: use the Krukal-Wallis H test

HOV: homogeneity of Variance (DV)

- Since an average variance is computed for denominator of *F*-statistic, variance should be similar for all groups: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots = \sigma_k^2$
- σ_e^2, pooled (averaged) variance, must be representative of each group so that MS_W is accurate
- Testing: Levene's Test
- All test for HOV are underpowered if samples are small, so you have to use judgement ;)
- If NOT HOV
  - alternatives: Welch, Brown-Forsythe, etc.

F-statistic: numerator = MS_B

Recall from CLT, relationship between variance of population (σ^2) & variance of *SDM* ($SE^2 = \sigma^2_{\bar{X}}$)

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n_j}} \rightarrow \sigma^2_{\bar{X}} = \frac{\sigma^2}{n_j} \rightarrow \sigma^2_{\bar{X}} \cdot n_j = \sigma_e^{\,2} = MS_B$$

One estimate of population variance ($\sigma_e^{\,2}$)

- Cannot compute population variance of all possible means as we only have a sample
  - Estimate population **variance** with sample means and **multiply** by sample size:

Equal Sample Sizes

$$MS_B = n \cdot s^2_{\bar{X}}$$

UN-equal Sample Sizes

$$MS_B = \frac{\sum n_j\left(\bar{X}_j - \bar{X}_G\right)^2}{k - 1}$$

If H_0 true, $MS_B = \sigma_e^{\,2}$

Have drawn *k* independent samples
From the SAME population
(i.e. group differences = 0)

If H_0 false, $MS_B \neq \sigma_e^{\,2}$

MS_B reflects BOTH
population variance
AND
group differences

Example: noise & words memorized

1. *Find grand mean:*

$$\overline{X_G} = \frac{9.2+6.6+6.2}{3} = \frac{22}{3} = 7.33$$

	group		
	A n = 5	B n = 5	C n = 5
outcome	9.2 (0.8)	6.6 (1.5)	6.2 (1.6)

2. *Find the SD of the means:*

$$s_{\bar{X}}^2 = \frac{(9.2-7.33)^2+(6.6-7.33)^2+(6.2-7.33)^2}{3-1} = \frac{5.3067}{2} = 2.65$$

3. *Multiply by n*

$$MS_B = 5 \cdot 2.65 = \mathbf{13.267}$$

<u>Equal</u>
Sample Sizes

$$MS_B = n \cdot s_{\bar{X}}^2$$

F-STATISTIC: DENOMINATOR = MS_W

Second estimate of population variance (σ_e^2)

- **Pooling** sample variances yields best estimate
  - $\sigma_1^2 = s_1^2$; $\sigma_2^2 = s_2^2$; … ; $\sigma_j^2 = s_j^2$
- Average subgroup (*j*) variance: $\sigma_e^2 = s_e^2$

<u>Equal</u>
Sample Sizes

$$MS_W = \sigma_e^2 = \frac{\sum s_j^2}{k}$$

<u>UN-equal</u>
Sample Sizes

$$MS_W = \sigma_e^2 = \frac{\sum (n_j - 1) s_j^2}{n_T - k}$$

Goal should be to obtain equal *ns*
BUT…
1 group > 50% larger other group: too much

k = # subgroups
j denotes the *j-th* subgroup

<u>Regardless of whether H_0 true:</u>

$$MS_W = \sigma_e^2$$

Not affected by group MEANS

Example: noise & words memorized

- *1. Average the* ***<u>VARIANCES's</u>:***

$$MS_W = \frac{0.8^2 + 1.5^2 + 1.6^2}{3} = \frac{5.5}{3} = \mathbf{1.9}$$

	group					
	A		B		C	
	n = 5		n = 5		n = 5	
outcome	9.2	(0.8)	6.6	(1.5)	6.2	(1.6)

<u>Equal</u>
Sample Sizes

$$MS_W = \sigma_e{}^2 = \frac{\sum s_j{}^2}{k}$$

Logic of "anova"

$$F_{Ratio} \text{ or } F_{Statistic} = \frac{MS_B}{MS_W}$$

- In ANOVA, 2 <u>independent</u> estimates of <u>same</u> population (error) variance are computed: σ^2, now called σ_e^2
  - MS_B : Variance <u>between</u> group means corrected by sample sizes (n_j)
  - MS_W : Average variance <u>within</u> groups
- Ratio of *2* estimates of population variance
- Hence the term *Analysis of Variance,* instead of something related to means comparisons (even though that is what we are interested in doing)
- Increased **variance among means** indicates means are **spread out** & likely differ from one another or come from different populations
- Large *F*-ratio indicates differences among means is <u>NOT</u> likely due to chance

ANOVA is

$$\frac{\text{Between-Group Measure of Variation Due to Estimate of Random Variation (Error)} + \text{Effect of IV (Group)}}{\text{Within-Group Estimate of Random Variation (Error)}}$$

Logic of "anova"

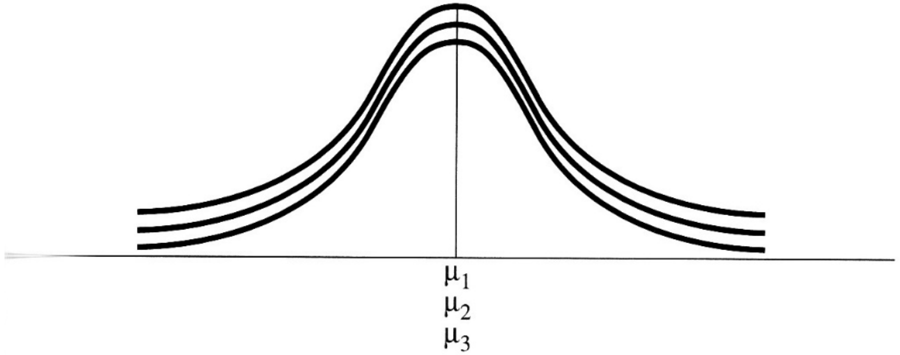
IF** all samples are the **same sizes…

$$\left.\begin{array}{l} MS_B = \sigma_e{}^2 = n_j \cdot s^2{}_{\bar{X}} \\ MS_W = \sigma_e{}^2 = \dfrac{\sum s_j{}^2}{k} \end{array}\right\} \xrightarrow{ratio} \boldsymbol{F} = \frac{MS_B}{MS_W}$$

When estimates of $\sigma_e{}^2$ (variances) are…

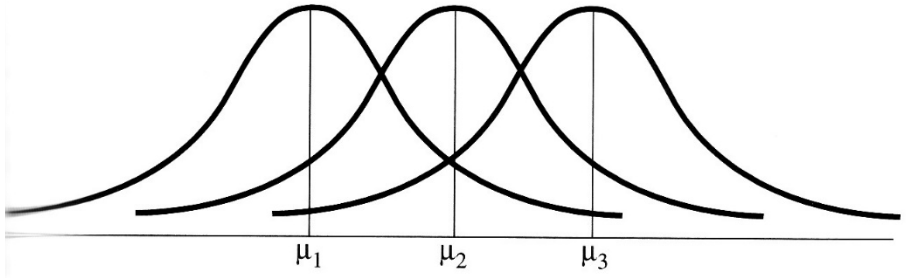
Equal: Fail to reject H_0

- All means come from same population
- Both are estimates of the same population variance $\sigma_e{}^2$
- F-ratio ≈ 1



Unequal: Reject H_0

- **Unlikely** that all means come from same population
- **Effect of IV surpasses random error/variation within groups**
- F-ratio significantly > 1, $MS_B > MS_W$



CALCULATIONS:

2 Approaches

SUMMARY STATS KNOWN
(shown on previous few slides)

SUM OF SQUARES (SS) APPROACH
(alternate formulas here)

$$SS = \sum_{i=1}^{n}(X_i - \bar{X})^2$$

Can '**partition**' total variation in DV due to group effects (IV) and error

$$SS_{Total} = SS_{Between} + SS_{Within}$$

Total

How different are ALL individuals from the "GRAND MEAN"

Inner Sum: individuals in each subgroup

Outter Sum: subgroups in the whole

$$SS_{Total} = \sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij} - \bar{X}_{GM})^2$$

$$\boxed{df_T = n_T - 1}$$

$$F_{Ratio} \text{ or } F_{Statistic} = \frac{MS_B}{MS_W}$$

Between

How different are "GROUP MEANS" from the "GRAND MEAN"

$$SS_{Between} = n_j \sum_{j=1}^{k}(\bar{X}_j - \bar{X}_{GM})^2$$

$$\boxed{df_B = k - 1}$$

$$MS_{Between} = \frac{SS_B}{df_B} = \frac{n_j \sum_{j=1}^{k}(\bar{X}_j - \bar{X}_{GM})^2}{k-1}$$

Within

How different are individuals from their "GROUP's MEAN"

Inner Sum: individuals in each subgroup

Outter Sum: subgroups in the whole

$$SS_{Within} = \sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2$$

$$\boxed{df_w = n_T - k}$$

$$MS_{Within} = \frac{SS_W}{df_W} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2}{N-k}$$

Can '**partition**' total variation in DV due to group effects (IV) and error

$$SS_{Total} = SS_{Between} + SS_{Within}$$

F-statistic

$$F_{Ratio} \text{ or } F_{Statistic} = \frac{MS_B}{MS_W}$$

- F_{crit} → F-distribution table
  - (different table per α)
    - *Across the top: find* df_B
    - *Down the side: find* df_W

- **If H_0 is true, $MS_B = MS_W$**

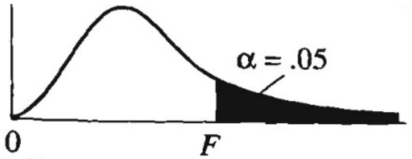
F-statistic ≈ 1

Both are estimates of variance of **same** population

- **If H_0 is false, $MS_B > MS_W$**

F-statistic exceeds F_{crit} by some amount

At least one mean significantly differs from another



df Denominator	df NUMERATOR 1	2	3	4	5	6	7	8	9	10	12	15
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07
27	4.21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Example: noise & words memorized

Test statistic: F-score observed

Critical Value: F-crit for α=.05

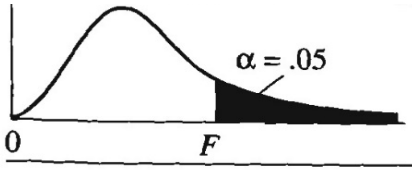
$$df = (3-1, 15-3) = \mathbf{(2, 12)}$$

$$MS_B = \mathbf{13.267}$$

$$MS_W = \mathbf{1.9}$$

$$F(2,12) = \frac{13.267}{1.90} = \mathbf{6.98}$$

Example: noise & words memorized



df Denominator	1	2	3	4	5	6	7	8	9	10	12	15
						df Numerator						
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
4	7.7	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
5	6.6	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
6	5.9	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
7	5.5	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
8	5.3	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
9	5.1	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
10	4.9	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
11	4.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12	4.7	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
13	4.6	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53
14	4.6	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46
15	4.5	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40
16	4.4	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35
17	4.4	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31
18	4.4	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27
19	4.3	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23
20	4.3	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20
21	4.3	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18
22	4.3	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15
23	4.2	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13
24	4.2	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11
25	4.2	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09
26	4.2	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07
27	4.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Critical Value: F-crit for α=.05

$$F_{crit}(2,12) = \mathbf{3.89}$$

$$F(2,12) = \frac{13.267}{1.90} = \mathbf{6.98}$$

Conclusion:

- AT LEAST ONE noise/music levels has a different mean # of words memorized.
- In fact it is the no noise/music condition that has the most words memorized.
- What type of music is playing doesn’t seem to make as much of a difference.

R Code: ANOVA

```
df %>%
  group_by(group) %>%
  furniture::table1(outcome)
```

```
──────────────────────────────────────
                      group
            A           B           C
            n = 5       n = 5       n = 5
outcome
            9.2 (0.8)   6.6 (1.5)   6.2 (1.6)
──────────────────────────────────────
```

Same as with t-tests

```
df %>%
  car::leveneTest(outcome ~ group,
                  data = .,
                  center = mean)
```

```
Levene's Test for Homogeneity of Variance (center = mean)
       Df F value  Pr(>F)
group  2  2.8213   0.09902 .
      12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What does this tell us?

R Code: ANOVA

```
df %>%
  group_by(group) %>%
  furniture::table1(outcome)
```

```
─────────────────────────────────────────────
                     group
             A         B         C
             n = 5     n = 5     n = 5
outcome
             9.2 (0.8) 6.6 (1.5) 6.2 (1.6)
─────────────────────────────────────────────
```

```
df %>%
  car::leveneTest(outcome ~ group,
                  data = .,
                  center = mean)
```

```
Levene's Test for Homogeneity of Variance (center = mean)
       Df F value  Pr(>F)
group  2  2.8213   0.09902 .
      12
---
Signif.                              ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

aov_4() function performs ANOVA

```
fit_anova <- df %>%
 afex::aov_4(outcome ~ group + (1|id),
       data = .)
```

(1|id) tells it the ID variable

R Code: ANOVA

```r
fit_anova <- df %>%
  afex::aov_4(outcome ~ group + (1|id),
      data = .)

fit_anova
fit_anova$anova
```

```
Anova Table (Type 3 tests)

Response: outcome
  Effect    df  MSE      F ges p.value
1  group 2, 12 1.90 6.98 ** .54    .010
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘+’ 0.1 ‘ ’ 1
```

R Code: ANOVA

```
fit_anova <- df %>%
  afex::aov_4(outcome ~ group + (1|id),
      data = .)

fit_anova
fit_anova$anova
```

```
Anova Table (Type 3 tests)

Response: outcome
  Effect    df  MSE       F ges p.value
1  group 2, 12 1.90 6.98 ** .54    .010
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘+’ 0.1 ‘ ’ 1
```

```
Anova Table (Type III tests)

Response: dv
            Sum Sq Df  F value    Pr(>F)
(Intercept) 806.67  1 424.5614 9.847e-11 ***
group        26.53  2   6.9825  0.009745 **
Residuals    22.80 12
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

R Code: ANOVA

```
fit_anova <- df %>%
  afex::aov_4(outcome ~ group + (1|id),
        data = .)

fit_anova
fit_anova$anova
```

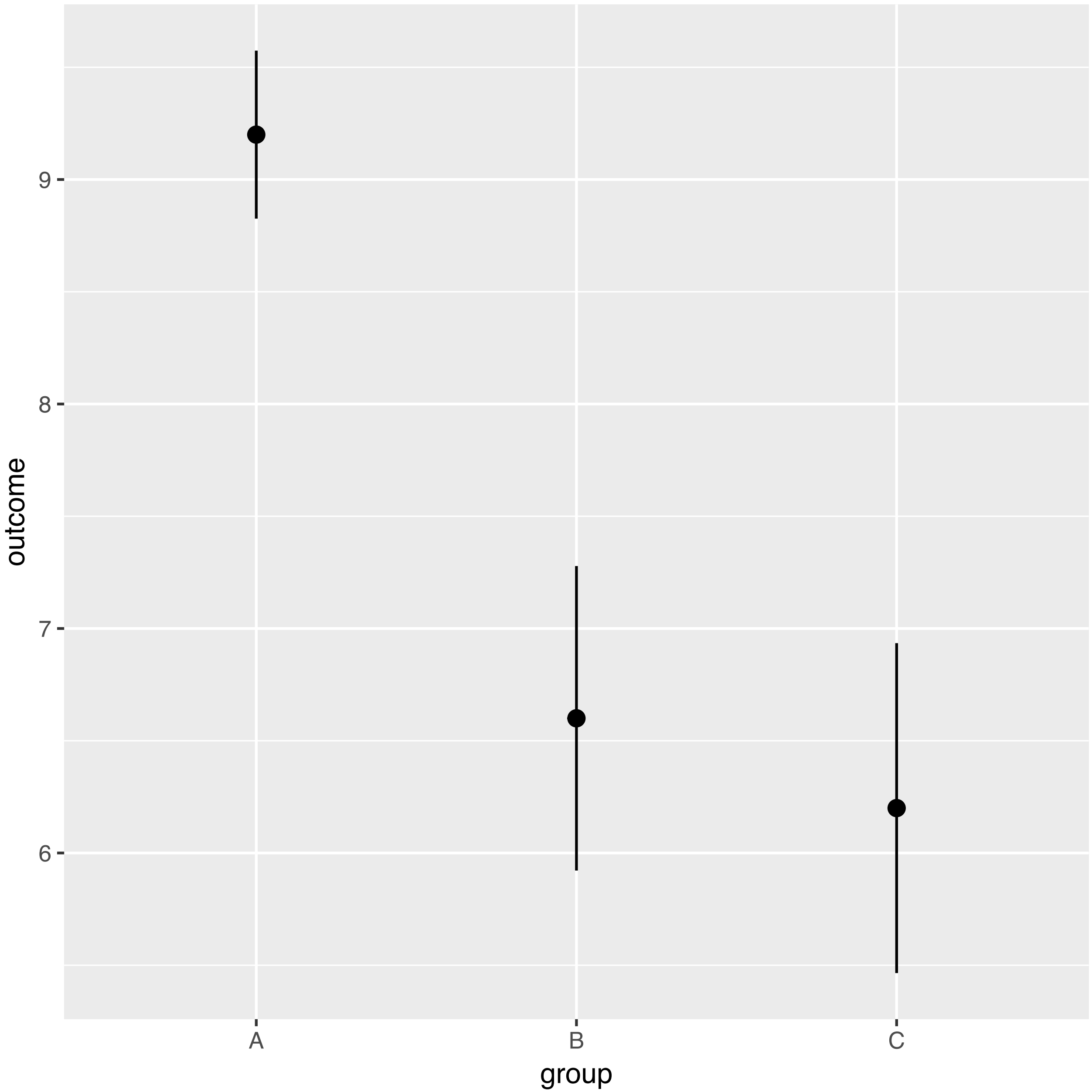
Reaches into the `fit_anova` object and grabs this more informative table

```
Anova Table (Type 3 tests)

Response: outcome
  Effect     df  MSE       F ges p.value
1  group 2, 12 1.90 6.98 ** .54    .010
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘+’ 0.1 ‘ ’ 1
```

```
Anova Table (Type III tests)

Response: dv
            Sum Sq Df  F value    Pr(>F)
(Intercept) 806.67  1 424.5614 9.847e-11 ***
group        26.53  2   6.9825  0.009745 **
Residuals    22.80 12
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

R Code: ANOVA

```
df %>%
  ggplot(aes(group, means)) +
    stat_summary()
```



Measures of Association

- **Term preferred over "Effect size" for ANOVA**
  - Amount or % of variation in DV explained/accounted for by knowledge of group membership (IV)
  - Correlation between grouping variable (IV) and outcome variable (DV)

- **4 measures:**
  - Eta-squared (η^2)
  - Omega-squared (ω^2)
  - Cohen's *f*
  - Intra-class Correlation Coefficients (ρ)

ω^2 is least biased, but unfamiliarity and 'difficulty' of computation have limited use

η^2 probably sufficient in many cases

Measures of Association: eta-squared

η^2 **: Measure of % reduction in error IN THIS DATA (SAMPLE)**

- SS_{Total} = Error in DV around grand mean
- SS_{Within} = Error around group means
- By knowing group membership we reduce error by $SS_{Between} = SS_{Total} - SS_{Within}$
- % reduction in error expressed as: $\eta^2 = \frac{SS_B}{SS_T} = \frac{df_B \cdot F}{df_B \cdot F + df_W}$
- η^2 can be biased with sample data

- Adjusted $\eta^2 = 1 - \frac{MS_W}{MS_T}$

- Compute using information from ANOVA summary table
  - $\eta^2 = SS_B / SS_T$
  - $\eta^2_{adj} = 1 - (MS_W / MS_T)$

Range: 0 to 1
Small: .01 to .06
Medium: .06 to .14
Large: > .14

Example: noise & words memorized

$df = (3-1, 15-3) = (\mathbf{2}, \mathbf{12})$

$F(2,12) = \dfrac{13.267}{1.90} = \mathbf{6.98}$

$MS_W = \mathbf{1.90} \xrightarrow{\text{"}SS = MS/df} SS_B = \mathbf{1.9} * (\mathbf{12}) = \mathbf{22.8}$

$MS_B = \mathbf{13.267} \xrightarrow{\text{"}SS = MS/df} SS_B = \mathbf{13.267} * (\mathbf{2}) = \mathbf{26.534}$

$$\eta^2 = \frac{SS_B}{SS_T} = \frac{df_B \cdot F}{df_B \cdot F + df_W}$$

Using SS

$$\eta^2 = \frac{26.534}{26.534 + 22.8} = \mathbf{0.5378}$$

Using F & df's

$$\eta^2 = \frac{2 \cdot 6.98}{2 \cdot 6.98 + 12} = \mathbf{0.5378}$$

Example: noise & words memorized

$df = (3-1, 15-3) = (\mathbf{2}, \mathbf{12})$

$MS_W = \mathbf{1.90} \xRightarrow{\text{"}SS=MS/df} SS_B = \mathbf{1.9} * (\mathbf{12}) = \mathbf{22.8}$

$F(2,12) = \dfrac{13.267}{1.90} = \mathbf{6.98}$

$MS_B = \mathbf{13.267} \xRightarrow{\text{"}SS=MS/df} SS_B = \mathbf{13.267} * (\mathbf{2}) = \mathbf{26.534}$

$$\eta^2 = \frac{SS_B}{SS_T} = \frac{df_B \cdot F}{df_B \cdot F + df_W}$$

Using SS

$$\eta^2 = \frac{26.534}{26.534 + 22.8} = \mathbf{0.5378}$$

Using F & df's

$$\eta^2 = \frac{2 \cdot 6.98}{2 \cdot 6.98 + 12} = \mathbf{0.5378}$$

Conclusion

The type of noise/music in the room accounts for 54% of the variation in the number of words each person was able to memorize

Measures of Association: OMEGA-squared

- **ω^2 : Measure of % reduction in error IN THIS POPULATION (ESTIMATE TRUTH)**
- Alternative for "fixed-effects" ANOVA
  - More conservative than η^2 (and less biased)
  - Range: 0 to 1 (can be negative when $F < 1$)
    - Same interpretation as η^2
  - Compute using information from ANOVA summary table
    - Equation for fixed effects ANOVA only

$$\omega^2 = \frac{SS_B - (k-1)MS_W}{SS_T + MS_W} = \frac{(k-1)(F-1)}{(k-1)(F-1) + n_j \cdot k}$$

Range: 0 to 1
Small: .01 to .06
Medium: .06 to .14
Large: > .14

Measures of association: Cohen's f

- **Traditional effect size index**
  - *Not a measure of association*
  - Generalization of Cohen's *d* to ANOVA
  - Compute using ANOVA summary information

$$f = \sqrt{\frac{\omega^2}{1-\omega^2}} = \sqrt{\frac{\frac{k-1}{n_j \cdot k}\left(MS_B - MS_W\right)}{MS_W}}$$

- Converting from f to ω^2 →

$$\omega^2 = \frac{f^2}{1+f^2}$$

Measures of Association: Intra-class correlation coefficient (ICC)

- Measure of association for random-effects ANOVA
- At least 6 *ICCs* available
  - Type selected depends on data structure
- Range: 0 to 1
  - Commonly used measure of agreement for continuous data
- Basic form:

$$\rho_{\text{intraclass}} = \frac{MS_B - MS_W}{MS_B + (n_j - 1)MS_W}$$

  - Measures extent to which observations within a treatment are similar to one another relative to observations in different treatments

APA Results

Methods

- Describe statistical and sample size analyses
- Describe factor and its levels
- Results of data screening

Results

- Reporting *F*-test:
  - $F(df_B, df_W) = F$-statistic, p = / <, measure of association and effect/effect size, power (optional)
- Don't need to include *MSE* (or MS_W) as Cohen suggests
- Discuss any follow-up tests, if any (next lecture)

Method

"A one-way ANOVA was used to test the hypothesis that the means of the three groups (Control, Moderate Noise, and Extreme Noise) were different following the experiment. A sample size analysis conducted prior to beginning the study indicated that five participants per group would be sufficient to reject the null hypothesis with at least 80% power if the effect size were moderate (Cohen's $f = .95$)."

Results

"Results indicated a significant difference among the group means, $F(2, 12) = 6.98$, $p < .01$, $\omega^2 = .44$"

ANOVA vs. multiple t-tests

- Why not run series of independent-samples *t*-tests?
- Could, and will usually get same results, but this approach becomes more difficult under 2 conditions:
  - Large *k*
    - $k(k-1) / 2$ different *t*-tests!
  - Factorial designs
- Danger of increased risk of Type I error when conducting multiple *t*-tests on same data set
  - *In next lecture we explain ways to potentially limit this risk*

Power: use G*Power