# Comparing 3 Conditions- ANOVA

Evaluation Methods & Statistics- Lecture 9

Benjamin Cowan & Andrew Howes

# Research Example (Lecture 7)

- Consequences of a secondary task on driving

- Does using a mobile phone to text cause driving quality to deteriorate?

# Research Example (This Week)

- Consequences of a secondary task on driving
  - Texting
  - Talking on phone

- Compared to just driving (control)

# How would we design this experiment?

# How would we design this experiment?

- IV- Secondary Driving Task
  - Level 1- Control Group (No secondary task)
  - Level 2- Texting
  - Level 3- Talking

- DV-Driving score

# How would we analyse the data?

- We could do 3 t-tests

  - Control to Texting
  - Control to Talking
  - Talking to Texting

- This would inflate our *Type I error rate*

# Type I error

- When we believe there is genuine effect in our population…..but actually there isn't (a false positive)

# Type II error

- When we believe there is **no** effect in the population…….but there is

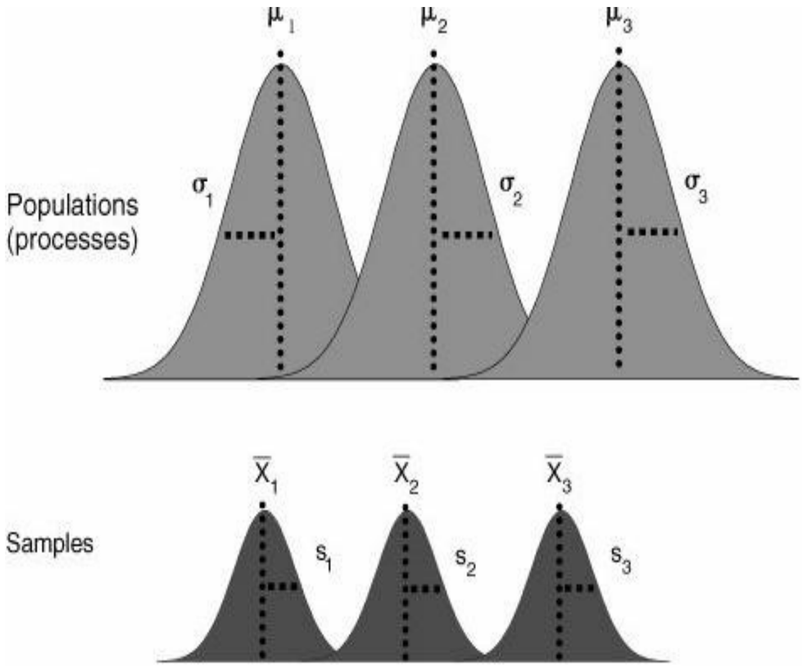
- Lot of natural variation between samples, too stringent controls for Type I error, low power of stats to find effects

# Familywise error rate

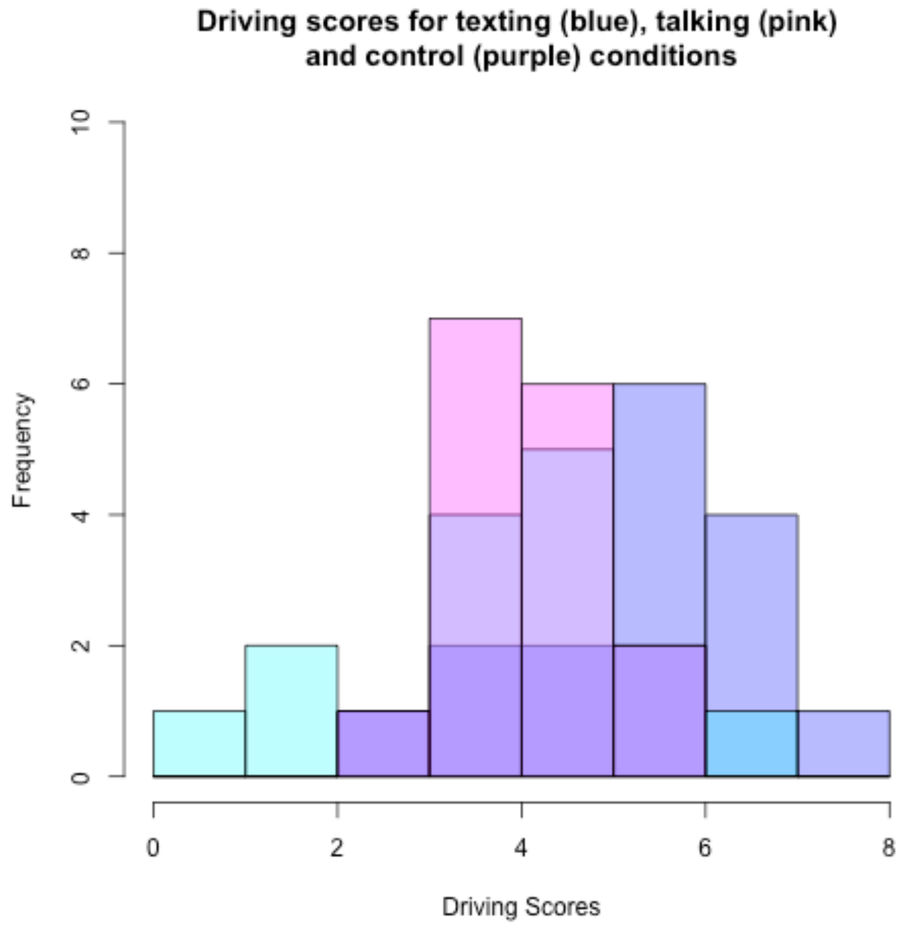
- If we have 3 tests in a *family of tests* and assume each is independent
- If we use Fishers level of 0.05 as our level of significance…
- The probability of a false positive (Type 1 error) in all of these tests
  - $0.95 \times 0.95 \times 0.95 = 0.857$
  - => Probability of Type 1 error is $1 - 0.857 = 0.143$
  - That is far greater than the Type I error for each test separately (0.05)
  - We therefore use ANOVA rather than lots of t-tests

# ANOVA- The Idea

- Compare 3 (or more) means to identify whether they are significantly different
  - i.e. whether they come from different populations
- Or more accurately…..we are testing the null hypothesis that the samples come from the same population.
- It is what we call an *omnibus test*
  - It tells us there is a significant difference, not where it is.



# Our Data

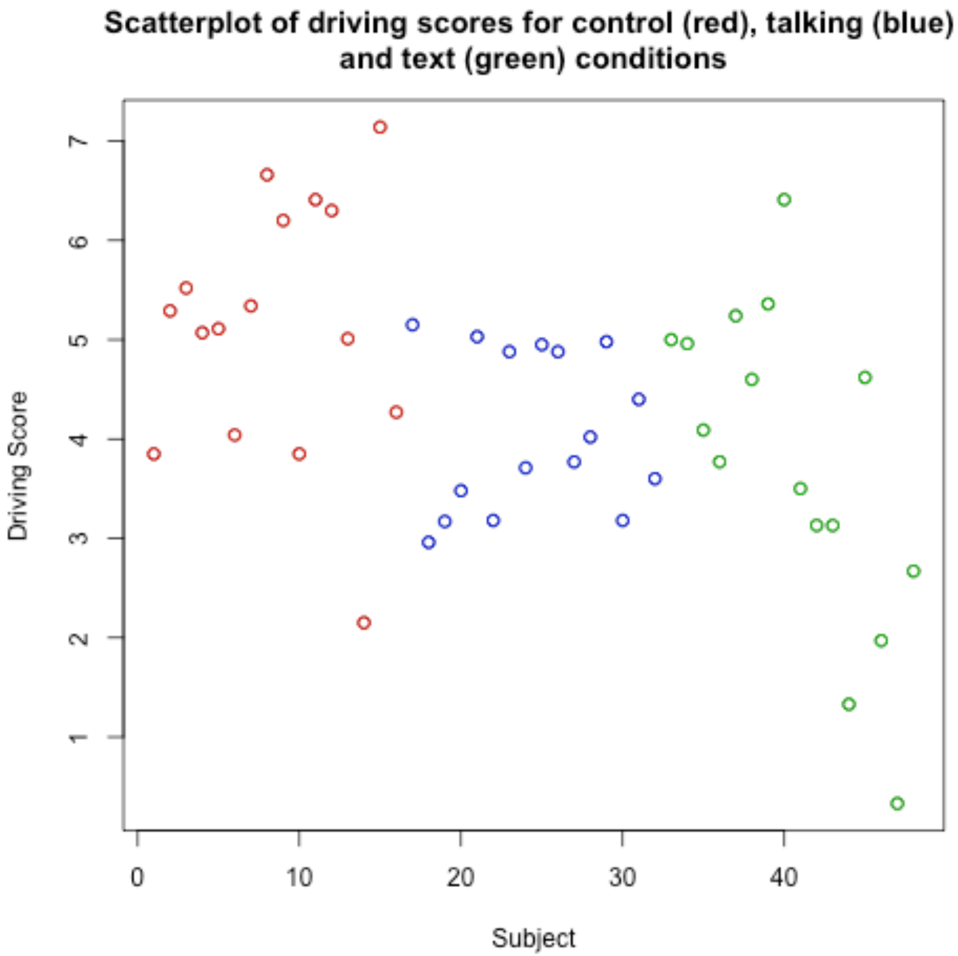


Driving scores for texting (blue), talking (pink)
and control (purple) conditions
Frequency
Driving Scores
0
2
4
6
8
10

# The Key: ANOVA & F Ratio

- F ratio is the ratio of **explained (that accounted for by the model we are proposing)** to **unexplained** variation
- This is calculated using the **Mean Squares**

# Our Data

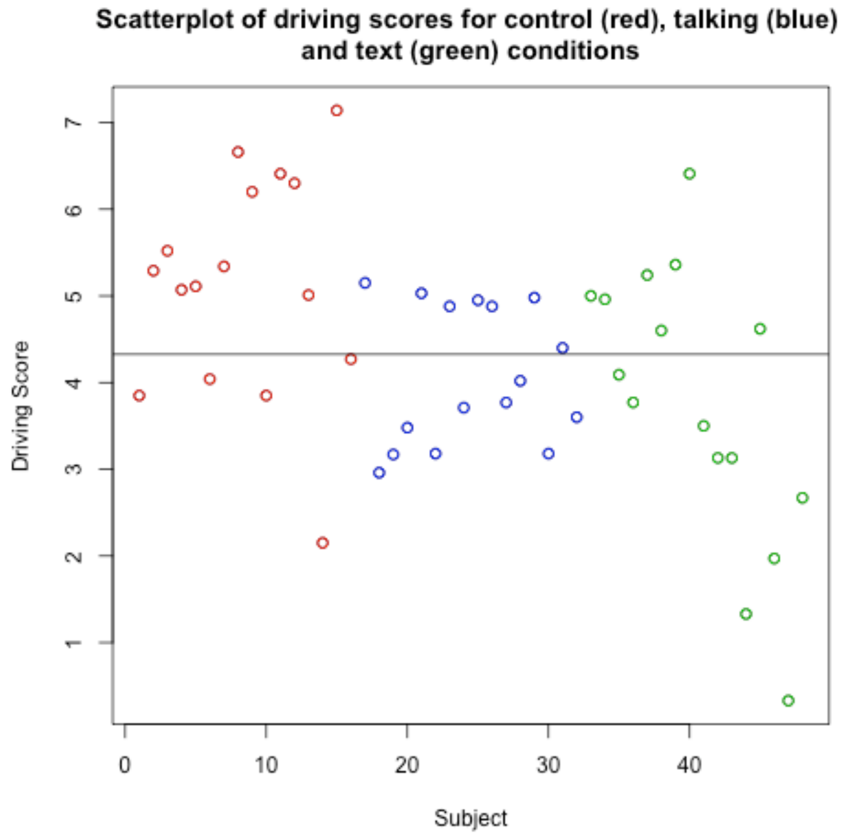


Scatterplot of driving scores for control (red), talking (blue)
and text (green) conditions
Driving Score
1
2
3
4
5
6
7
0
10
20
30
40
Subject

# The mindset of “models”

the mean is a statistical model, just sometimes not a very good one…….

Does the statistical model we have proposed explains the variation better?



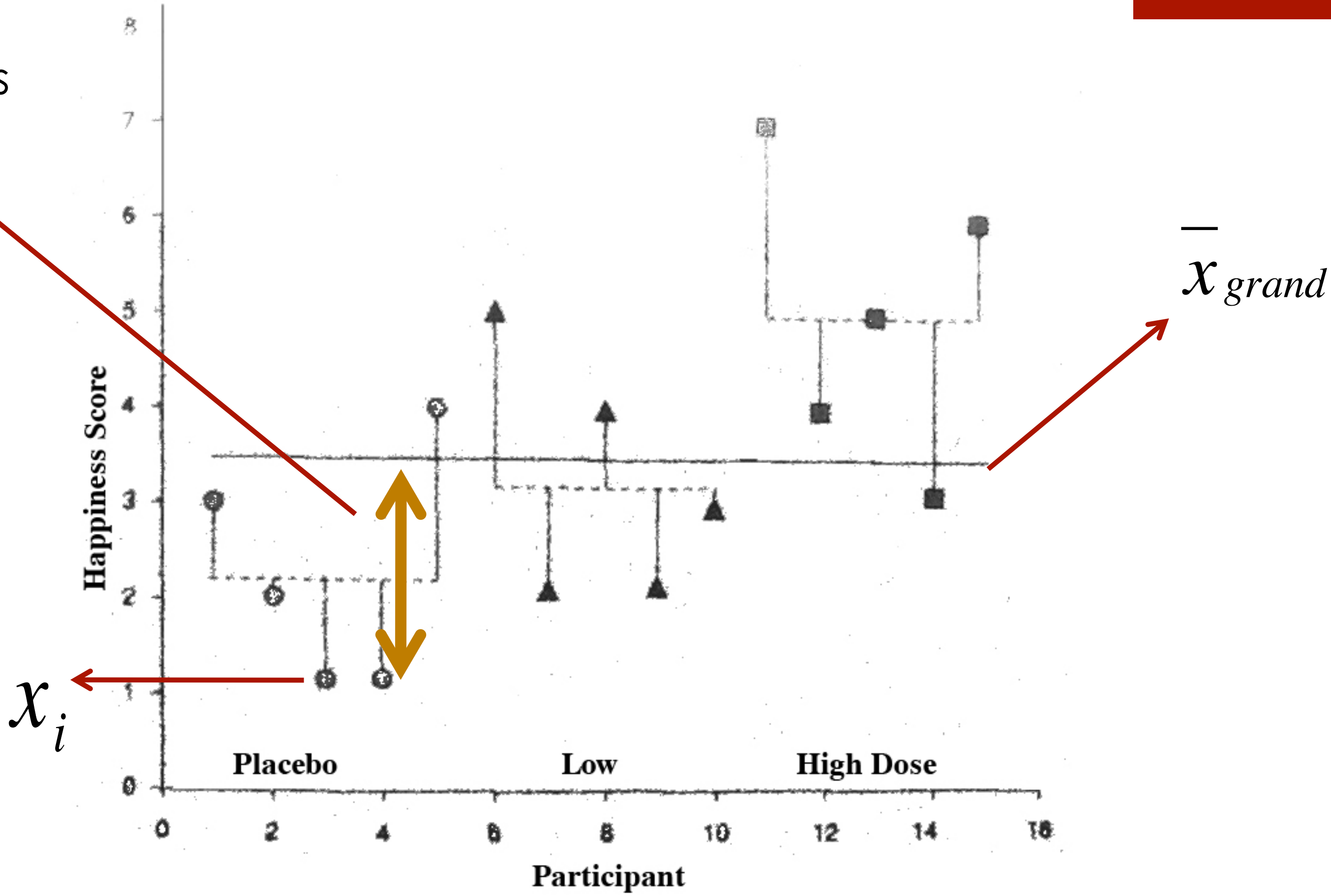
# Step 1- Total Sum of Squares

- The total amount of variation in our data
- This should look familiar

$$SS_T = \sum \left( x_i - \bar{x}_{grand} \right)^2$$

# Step 1- Graphically

What the equation is doing



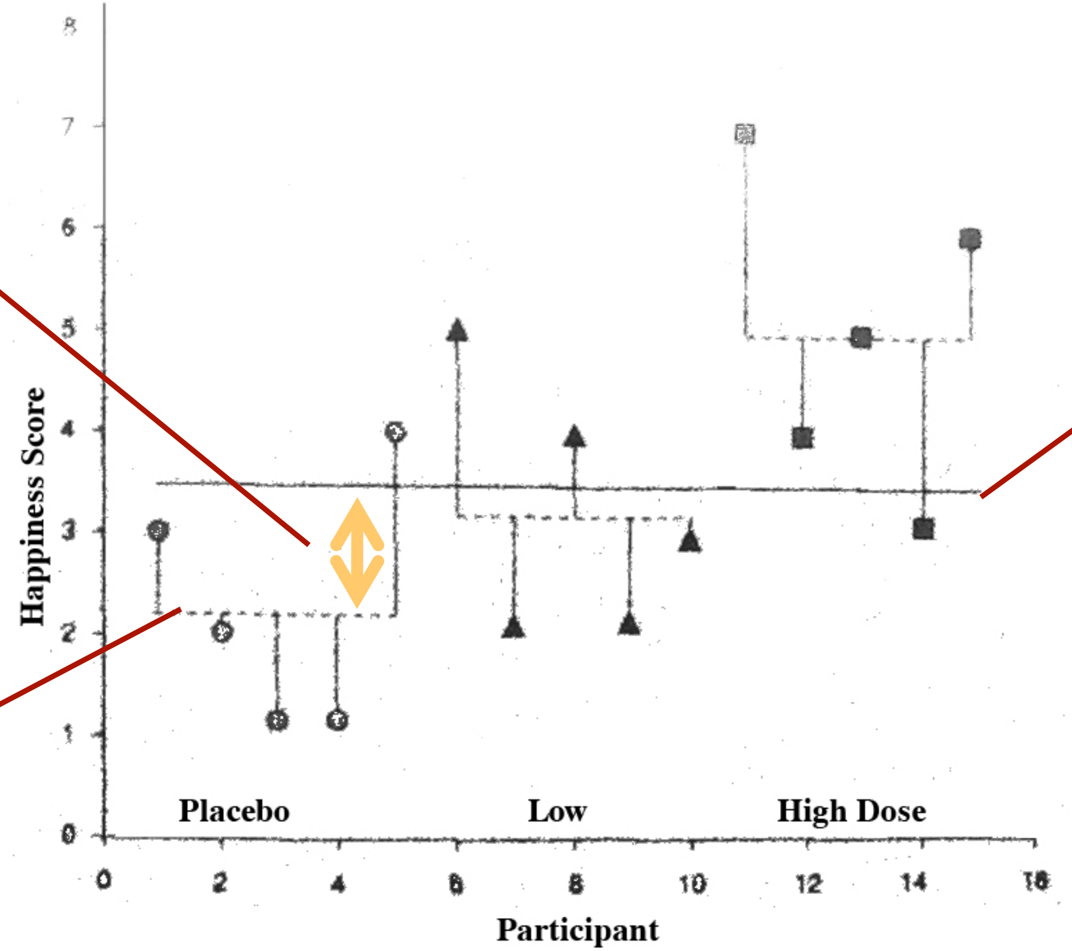
# Step 2- Model Sum of Squares

- We now need to know how much variation our model can explain
- How much the total variation can be explained due to data points coming from different groups in "the perfect model"
- $n_k$ is the amount of people in that condition

$$SS_M = \sum n_k(\bar{x}_k - \bar{x}_{grand})^2$$

# Step 2- Graphically

What the equation is doing




$\bar{x}_{grand}$

$\bar{x}_k$

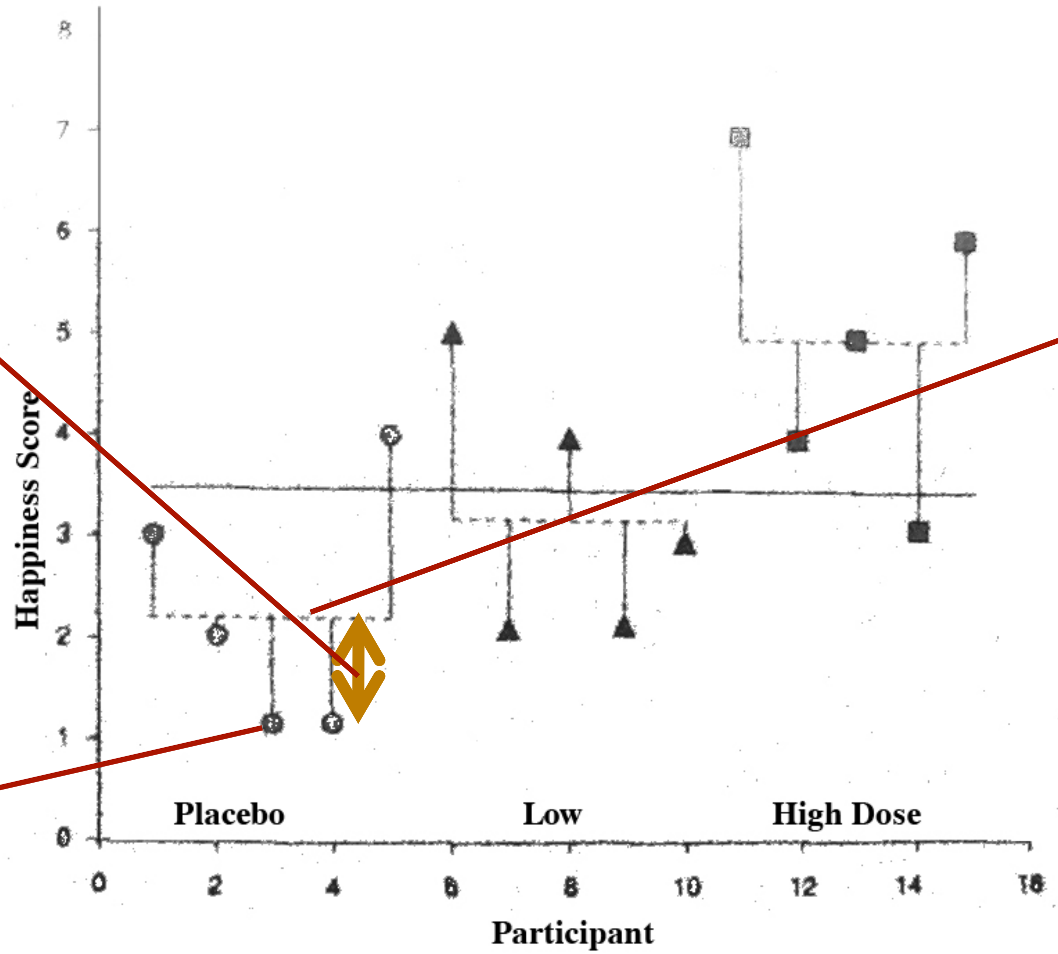
# Step 3- Residual Sum of Squares

- How much of the variation cannot be explained by the model i.e. what error is there in the model prediction?
- Easy way to calculate: $SS_R = SS_T - SS_M$
- But here is the real formula

$$SS_R = \sum \left( x_{ik} - \overline{x}_k \right)^2$$

# Step 3- Graphically

What the equation is doing




$\bar{x}_k$

$x_{ik}$

# Sum of Squares & Mean Squares

- These are summed values
  - Therefore impacted by the number of scores in the sum (remember variance in Lecture 3)

- We can get around this by dividing by the respective degrees of freedom for each SS

# Degrees of Freedom for each SS

- Degrees of Freedom for $SS_T$ (dfT):
  - N-1
- Degrees of Freedom for $SS_M$(dfM):
  - Number of Conditions (k) -1
- Degrees of Freedom for $SS_R$(dfR):
  - N-k

# F Ratio

- The F Ratio is calculated using the:

- Mean Squares model ($MS_M$):
  - $SS_M/df_M$

- Mean Squares residual (error) ($MS_R$):
  - $SS_R/df_R$

# F Ratio

$$\frac{\text{Variation explained by our model}}{\text{Variation unexplained by our model}}$$

# F Ratio

$$\frac{\text{Mean Square Model } (MS_M)}{\text{Mean Square Residual } (MS_R)}$$

# F Distribution

- F Distribution for specific pair of degrees of freedom

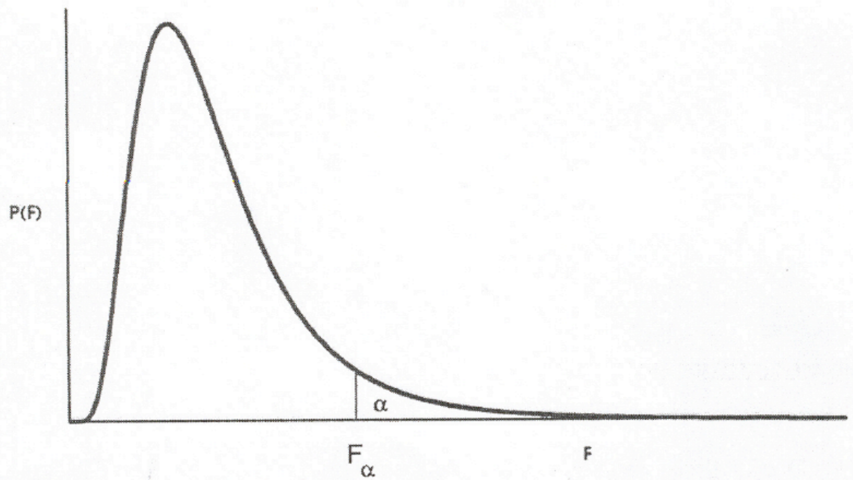



Figure K.1: The F distribution

- Table of Critical Values

***Critical values of F for the 0.05 significance level:***

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| **1** | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 |
| **2** | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 |
| **3** | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 |
| **4** | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 |
| **5** | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 |
| **6** | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 |
| **7** | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 |
| **8** | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 |
| **9** | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 |
| **10** | 4.97 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 |
| **11** | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.10 | 3.01 |
| **12** | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 |

# One Way Independent ANOVA- Assumptions

- Normally distributed data (what test?)
- Equality of Variance (what test?)
- Interval or ratio data
- Independent data

# One Way Independent ANOVA- Assumptions

- Normally distributed data (Shapiro-Wilk)
- Equality of Variance (Levene's)
- Interval or ratio data
- Independent data

# Repeated Measures ANOVA- Sphericity

- independence of data doesn't hold
  - data is from the same participants
- Instead we look for sphericity
  - variance of the differences between scores in each treatment are equal
  - Calculate difference between pairs of scores in all possible combination of treatment levels ,then calculate the variance of these differences

# Mauchly's test of sphericity

- It tests the hypothesis that the variances of the differences are equal (H0)

- If I got $p<0.05$ for this test would it be good or bad?

# Mauchly's test of sphericity

- It tests the hypothesis that the variances of the differences are equal (H0)

- If I got $p<0.05$ for this test would it be good or bad?
- It would be bad as it states there is a significant difference between variance of differences

- Corrections exist if this is the case, usually Greenhouse-Geisser correction is used

# One Way Repeated Measures ANOVA- Assumptions

- Normally distributed data for each condition (Shapiro-Wilkes test)
- Sphericity
- Interval or ratio data

# Running Independent ANOVA in R

Code:

```
model <-aov(score~condition, data=data)

summary(model)
```

Output:

```
> summary(model)
            Df Sum Sq Mean Sq F value Pr(>F)
condition    2  16.67   8.337   5.097 0.0101 *
Residuals   45  73.61   1.636
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

# Running Repeated Measures ANOVA in R

Code:

Install.packages ('ez'); library ('ez')

analysis <-ezANOVA(data=wai.Tot.Anova, dv=.(wai.score), wid=.(userid), within=.(wai.time), between=.(condition), type=3)

analysis

Output:

```
> analysis
$ANOVA
              Effect DFn DFd         F          p p<.05          ges
2          condition   1  42 0.1435817 0.70665485       0.0029651789
3           wai.time   2  84 3.7235569 0.02822214     * 0.0113989211
4 condition:wai.time   2  84 0.2371799 0.78937583       0.0007339116

$`Mauchly's Test for Sphericity`
              Effect         W            p p<.05
3           wai.time 0.7000419 0.0006684095     *
4 condition:wai.time 0.7000419 0.0006684095     *

$`Sphericity Corrections`
              Effect       GGe      p[GG] p[GG]<.05      HFe      p[HF]
3           wai.time 0.7692556 0.04021494         * 0.792803 0.03878871
4 condition:wai.time 0.7692556 0.73058235           0.792803 0.73751657
  p[HF]<.05
3         *
4
```

# Reporting ANOVA

- F ratio
- Degrees of Freedom ($dof_M$, $dof_R$)
- P value
- Back to the example we saw earlier:
  - **$F(2, 45) = 5.097$, $p<0.05$**

- We can therefore state that there is a significant effect of secondary task on driving score

# Omnibus test & Post Hoc

- Main effect of secondary driving task: $F(2, 45) = 5.097$, $p<0.05$
- Significant effect of our experiment conditions on driving score
- But how does this break down?
  - Control >Text?
  - Control > Text?
  - Text > Talk?
- We need ***post hoc tests***

# Post Hoc Tests

- Used when no specific a priori predictions about the data we have
- They are used for exploratory data analysis
- Pairwise comparisons
  - Like performing t-tests on all the pairs of mean in our data
  - There are many to choose from……..

# A selection of common post hoc tests

- LSD (Least Significant Difference)
  - Analogous to multiple t-tests

- Bonferroni
  - Uses Bonferroni correction to control for Type I
  - With multiple comparisons this may be too conservative (increase chance of Type II error)

- Tukey's test
  - Control Type I and better when testing large number of means

# Which one to choose?

- Trade off between:
  - Type I error rate likelihood
  - Statistical power (ability to find an effect if there is one)
  - Whether assumptions of ANOVA have been violated, although most are robust to minor variations

# Running Post Hoc tests in R

Code:

pairwise.t.test (data$score, data$condition, paired=FALSE, p.adjust.method="bonferroni")

Output:

```
    Pairwise comparisons using t tests with pooled SD

data:  data$score and data$condition

     control talk
talk 0.073   -
text 0.011   1.000

P value adjustment method: bonferroni
```

# Lecture Readings and Further concepts to consider

- Core:- Field (2009) Chapters 8 & 11 (pages 427-454)

- Other concepts to consider:
  - **Statistical Power:** Cohen (1992). A power primer. Psychological Bulletin
  - **Planned Contrasts**: Field (2009), Chapter 8, p.325-339