# BG ANOVA

## Analysis of Variance for between groups designs

- When, Why to Use ANOVA
- BG ANOVA & BG t-tests
- Summarizing/Displaying Data -- 1 qual & 1 quant var
- How ANOVA & F Work
- Research and Null Hypotheses for ANOVA
- Making decisions about H0: and RH:
- Causal Interpretation
- Computation & notation stuff

When to use it ?

– Whenever you want to compare the means on a quantitative variable for two different groups or conditions.

Said Statistically ?

– When you are testing for a relationship between one quantitative variable and one qualitative variable -- by comparing the means of the quantitative variable for the categories of the qualitative variable

- Note: "groups" is often used when the qualitative variable is a subject variable, and "conditions" often used when it is a manipulated variable, but "categories", "situations", "subpopulations", "treatments" are all used.

To which of the following variable pairs would you apply ANOVA?

| | |
|---|---|
| • GRE & & Public/Private school | ANOVA |
| • School Attended & Type of Job Obtained | nope |
| • Favorite Sport & Age | ANOVA |
| • Age group & Performance Speed | ANOVA |
| • Income & Years of Education | nope |
| • Income & Highest Educational Degree | ANOVA |

What about between groups t-tests?

– Whenever you want to compare the means on a quantitative variable for two different groups or conditions, you can use either a BG ANOVA or a BG t-test

– The two procedures will produce exactly the same
  - group means
  - p-value & NHST results
  - $t^2 = F$
  - ANOVA dferror = t-test df

– We will emphasize ANOVA in this class, because it is used somewhat more often, and because, unlike t-tests, can be used for larger designs (with more IV conditions – later!)

## Summarizing & Reporting ANOVA Data

Qual var -- When tested

Quant -- Performance Score

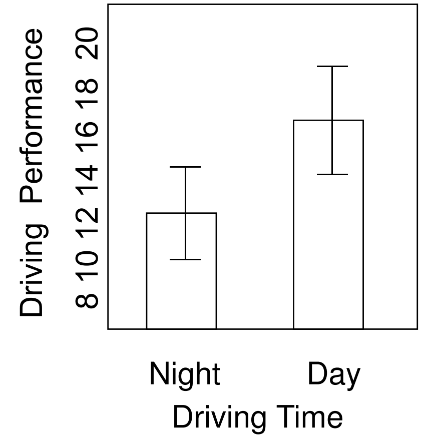
Report the mean and Std of the quant variable for each condition of the qual variable.

| | Driving at Night | Driving During Day |
|---|---|---|
| | 13 | 18 |
| | 15 | 14 |
| | 10 | 17 |
| | 12 | 19 |
| Mean | 12.5 | 17.0 |
| Std | 2.1 | 2.2 |

We'd report the mean of which variable in each pair?

| | |
|---|---|
| • GRE & & Public/Private school | GRE |
| • Favorite Sport & Age | Age |
| • Failure Rate & Brand of Computer | Failure rate |
| •Age group & Performance Speed | Speed |
| • Income & Highest Educational Degree | Income |

## Displaying ANOVA Data




A Bar Graph can be used to display the data.

- The height of each bar is the mean of the quant var for that condition of the qual var.
- The "whiskers" show variability around each mean. Might be ...
  - +/-1 std
  - +/-1 SEM (later…)

Table give more complete data, while graphs make it easier to "see" the means comparison quickly.

## Research Hypotheses for ANOVA

ANOVA RH: are …

- Always about mean differences
- Always about the populations represented by the groups or conditions, not the groups or conditions themselves (remember, this is about inference)
- For 2-group designs, have only three possible RH: patterns

$\overline{X}_{G1} < \overline{X}_{G2}$ $\overline{X}_{G1} = \overline{X}_{G2}$ $\overline{X}_{G1} > \overline{X}_{G2}$

Note: We'll use the symbol "=" but we'll use phrases like "equivalent", "nearly equal", "not significantly different", "statistically equivalent", etc.

It is unlikely that the average of two populations is exactly the same. What we mean is that the mean difference "isn't large enough to be "meaningful", "practical" or "important" and any difference we think we see is due to "chance". (Sampling Variability Happens !!!)

## Null Hypotheses for ANOVA

ANOVA H0: are …

- Always about mean differences
- Always about the populations represented by the groups or conditions, not the groups or conditions themselves (remember, this is about inference)
- Always that $\overline{X}_{G1} = \overline{X}_{G2}$

Using the Type of Therapy (group or individual) and Depression (measures on a 20-point scale) example from before...

• A population of psychiatric patients that receives group therapy will have the same average depression scores as a population of psychiatric patients receiving individual therapy.

The H0: is a mathematical expression, so "equal" is appropriate -- unlike for the RH:

## "Draw the boxes" to show designs and hypotheses

Psychiatric patients that receive group therapy have lower depression scores than those that receive individual therapy.

| Group | Individual |
|---|---|
| | |

H0: =
RH: <

Snapping turtles eat more crickets than do painted turtles.

| Snapping | Painted |
|---|---|
| | |

H0: =
RH: >

Graduate and undergraduate students will perform about the same on the next exam.

| Grad | UGrad |
|---|---|
| | |

H0: =
RH: =

## "Draw the boxes" to show designs and hypotheses

First-born children have better manners than only children.

| 1st Born | Only |
|---|---|
| | |

H0: =
RH: >

Practice alone works better than practice with feedback.

| Pract & FB | Pract |
|---|---|
| | |

H0: =
RH: <

Patients treated by Psychologists and Psychiatrists will have the same level of hysteria.

| Psychologists | Psychiatrists |
|---|---|
| | |

H0: =
RH: =

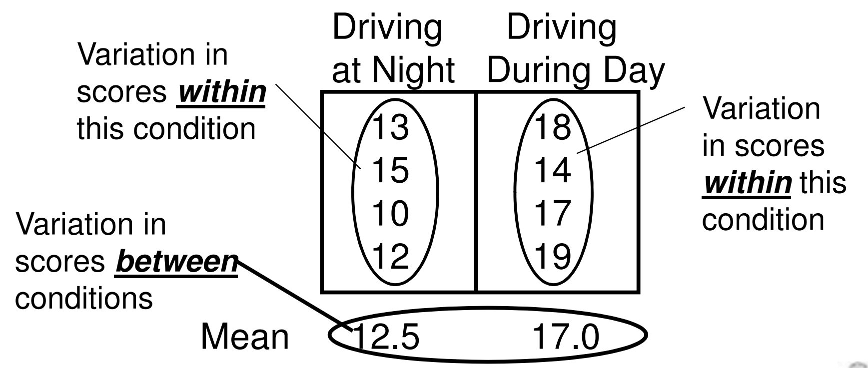
## How ANOVA "Works"

ANOVA is from -- ANalysis Of VAriance

- Variance is a statistical term for variation or variability
- In ANOVA, variation among the scores on the quantitative variable is divided into
  - variation *between* the groups / conditions
  - variation *within* the groups / conditions
- These two types of variation are then combined into the ANOVA summary statistic -- **F**
- F has a range of 0 to ∞
- We use regular "H0: testing logic"
  - if the F is small, then we have to conclude that the groups represent populations with equivalent means
  - if the F is "large enough" then we can conclude that the groups represent populations with different means on the quantitative variable

## Example of How ANOVA "Works"

Consider the following data set , with variation in the quantitative scores between & within each condition of the qualitative variable




## What "retaining H0:" and "Rejecting H0:" means...

- When you retain H0: you're concluding…
  - The mean difference between these groups/conditions in the sample ***is not*** large enough to allow me to conclude there is a mean difference between the populations represented by the groups/conditions.
- When you reject H0: you're concluding…
  - The mean difference between these groups/conditions in the sample ***is*** large enough to allow me to conclude there is a mean difference between the populations represented by the groups/conditions.

## Mechanics of H0: testing with ANOVA

There are two different ways of making this decision, depending upon whether you are doing the analysis on the computer or performing the computations by hand. You must be familiar with each procedure.

## On the computer:

- Obtain the summary statistic and p-value

  $F = 5.21 \quad p = .024$

- Decide whether to retain or reject H0:
  - if $p < .05$, reject H0: (decide the variables are related)
  - if $p > .05$, retain H0: (decide there is no relationship)
  - for the example, since $p < .05$ . . . reject H0:

Remember: p tells you the probability of a Type I error (False Alarm) if you reject H0: -- we're only willing to take a 5% risk

## Computing By Hand:

- Compute the "obtained value" of the summary statistic (based on the sample data -- sometimes called F-computed or -calculated)

  F-obtained = 5.21

- Look up the "critical value" of F for the design on the "F-table"

  F-critical = 4.41

- Decide whether to retain or reject H0:
  - If the "obtained value" is larger than the "critical value", reject H0:
  - If the "obtained value" is smaller than or equal to the "critical value", retain H0:
  - for the example $5.21 > 4.41$, so reject H0:

By-hand and computer analysis of the same data will always produce the same result, because if $p < .05$, then F-obtained > F-critical

A little H0: and RH: practice ...

RH: 1st graders given more practice would have better performance on their math test.

Mean for 10-practice group = 72.4% F = 1.23, F-critical = 3.21

Mean for 30-practice group = 74.8%

Retain or Reject H0: ? retain Support for RH: ? No - no effect

RH: People with a Psychiatric Dx who receive psychotropic drugs will respond as well to "talk therapy" as those who don't.

Mean for drug group = 18.2 F = 6.24, F-critical = 3.21

Mean for no-drug group = 24.8

Retain or Reject H0: ? reject Support for RH: ?

No - RH: = H0:, but got an effect

RH: Doing the WebEx will improve your scores on the computational homework.

Mean for WebEx group = 85.6% F = 8.98, p = .001

Mean for no-WebEx group = 92.4%

Retain or Reject H0: ? reject Support for RH: ? Nope ! Backwards !

Statistical decisions & errors with ANOVA...

| Statistical Decision | In the Population: $X_{G1} < X_{G2}$ | $X_{G1} = X_{G2}$ | $X_{G1} > X_{G2}$ |
|---|---|---|---|
| $X_{G1} < X_{G2}$ ($p < .05$) | Correct H0: Rejection & Direction | Type I "False Alarm" | Type III "Mis-specification" |
| $X_{G1} = X_{G2}$ ($p > .05$) | Type II "Miss" | Correct H0: Retention | Type II "Miss" |
| $X_{G1} > X_{G2}$ ($p < .05$) | Type III "Mis-specification" | Type I "False Alarm" | Correct H0: Rejection & Direction |

Remember that "in the population" is "in the majority of the literature" in practice!

## About causal interpretation of ANOVA results ...

Like the other bivariate statistics we have studied, we can only give a causal interpretation of the results if the data were collected using a true experiment

- random assignment of subjects to conditions of the qualitative variable ( IV ) - gives initial eq.
- manipulation of the IV by the experimenter - gives temporal precedence
- control of procedural variables - gives ongoing eq.

Only then can differences between the condition means be taken as evidence that the IV causes the DV

## Practice with causal interpretation of ANOVA results ...

Which of the following can't possibly be given a causal interpretation and for which is it possible (only if the RA, IV manip & confound control are done properly) ?

- The 10 year olds spelled more words correctly than the 5 year olds. — Nope!
- Those taught spelling using the computer did better than those taught by lecture. — Possible
- Democrats donate more to environmental causes than Republicans — Nope!
- After instruction about the importance of our environment, school children increased the amount they wanted to donate. — Possible

A bit about computational notation for BG ANOVA…

Start by sorting the DV data (X) from the study into two columns – one for each condition.

Then make a column of squared values ($X^2$) for each condition

Then sum each column -- making a $\Sigma X$ & $\Sigma X^2$ for each group

| Group Therapy (k1) | | Individual Therapy (k2) | |
|---|---|---|---|
| X | $X^2$ | X | $X^2$ |
| 3 | 9 | 5 | 25 |
| 5 | 25 | 6 | 36 |
| 4 | 16 | 8 | 64 |

$\Sigma X_{k1} = 12$ $\Sigma X_{k1}^2 = 50$ $\Sigma X_{k2} = 19$ $\Sigma X_{k2}^2 = 125$

A bit about computational notation for BG ANOVA, continued …

All the various calculations will use combinations of these four terms – be sure you are using the correct one !

$\Sigma X_{k1}$ $\Sigma X_{k1}^2$ $\Sigma X_{k2}$ $\Sigma X_{k2}^2$

Other symbols you'll need to know are…

- N = total number of participants in the whole study
- n = number of participants in a particular condition of the study
- k = number of conditions in the study