

ANOVA

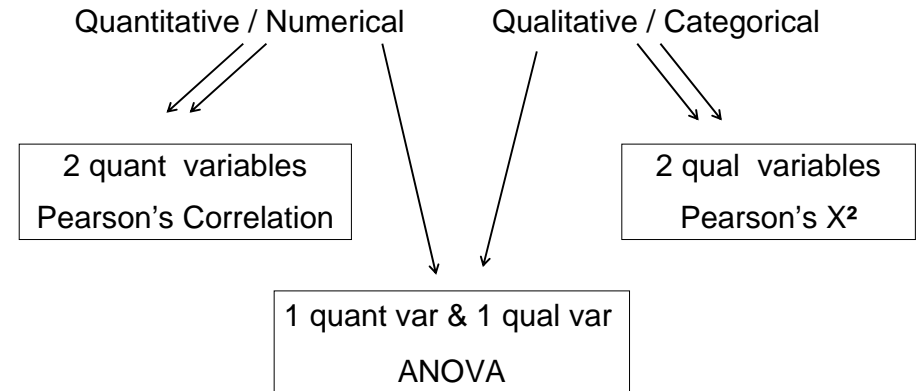
Analysis of Variance

- Picking the correct bivariate statistical test
- When, Why to Use ANOVA
- Summarizing/Displaying Data -- 1 qual & 1 quant var
- How ANOVA & F Work
- Research and Null Hypotheses for ANOVA
- Making decisions about H₀: and R_H:
- Causal Interpretation
- Between Groups and Within-Groups ANOVA

So... what are the bivariate null hypothesis significance tests (NHSTs) we'll be using ???

What are the two kinds of variables that we've discussed?

What are the possible bivariate combinations?



We have separate bivariate statistics for each of these three data situations...

There is lots to learn about each of the statistical tests, but right now I want you to be sure you can tell when to use which one...

the "secret" is to figure out whether each variable is qualitative or quantitative, then you'll know which of the 3 stats to use !!

We want to know whether there is a relationship between someone's IQ and their amount of political campaign contributions.

IQ is ... Contributions is ... Stat?

We want to know whether men and women make different amounts of political campaign contributions.

Gender is ... Contributions is ... Stat?

We want to know whether men or women are more likely to make a political contribution.

Gender is ... Contributions is ... Stat?

Here's a few more...

"relationship" expressions of hypotheses

- I expect there is a relationship between a person's height and their weight.
- I believe we'll find that there is a relationship between a person's gender and their weight.
- My hypothesis is that there is a relationship between a person's gender and whether or not they have a beard

"tend to..." expressions of hypotheses

- I expect that males tend to be heavier than females.
- My hypothesis is that taller folks also tend to be heavier
- I expect that folks with beards tend to be males.

"if ... then more likely..." expressions of hypotheses

- If you have a beard, then you are more likely to be male.
- If you are heavier, then you are more likely to be taller.
- If you are heavier, then you are more likely to be male.



When to use ANOVA ?

- 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 & Gender
- School Attended & Type of Job Obtained
- Favorite Sport & Age
- Age group & Performance Speed
- Income & Years of Education
- Income & Highest Educational Degree

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
Mean	12.5	17.0
Std	2.1	2.2

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

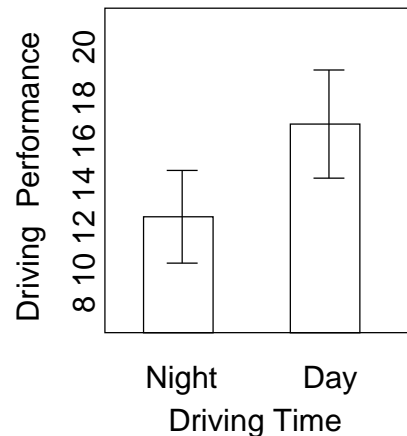
- GRE & Gender
- Favorite Sport & Age
- Failure Rate & Brand of Computer
- Age group & Performance Speed
- Income & Highest Educational Degree

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

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

$$\bar{X}_{G1} < \bar{X}_{G2}$$

$$\bar{X}_{G1} = \bar{X}_{G2}$$

$$\bar{X}_{G1} > \bar{X}_{G2}$$

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”.

Examples of Research Hypotheses for ANOVA

For a given pair of variables there are only three possible RH:s

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

- A population of psychiatric patients that receives group therapy will have lower average depression scores than a population of psychiatric patients that receives individual therapy.
- A population of psychiatric patients that receives group therapy will have statistically equivalent average depression scores as a population of psychiatric patients receiving individual therapy.
- A population of psychiatric patients that receives group therapy will have higher average depression scores than a population of psychiatric patients that receives individual therapy.

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 $\bar{X}_{G1} = \bar{X}_{G2}$

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:



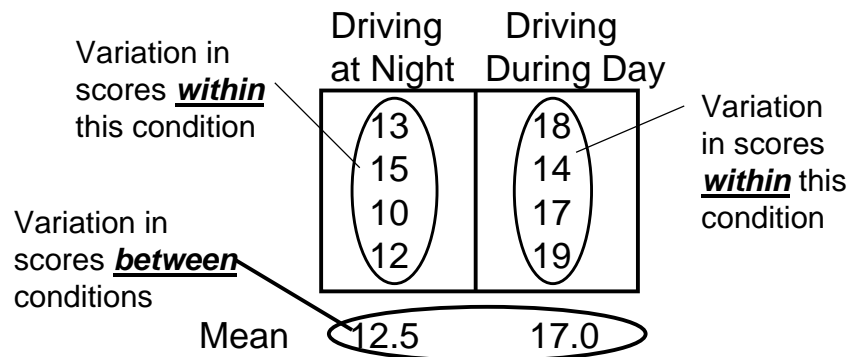
How ANOVA “Works”

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 can’t say the groups represent populations with different means
 - if the F is “large enough” then the groups probably don’t represent populations with the same mean 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
 $F = 5.21$ $p = .024$
- 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)
 $F\text{-obtained} = 5.21$
- Look up the “critical value” of F for the design on the “F-table”
 $F\text{-critical} = 4.41$
- 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 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 $F\text{-obtained} > F\text{-critical}$

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, p = .15

Mean for 30-practice group = 74.8%

Retain or Reject H0: ? Support for RH: ?

RH: Schizophrenics who receive psychotropic drugs will respond as well to "talk therapy" than 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: ? Support for RH: ?

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: ? Support for RH: ?

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

Statistical decisions & errors with ANOVA...

In the Population

Statistical Decision

$X_{G1} < X_{G2}$

$X_{G1} = X_{G2}$

$X_{G1} > X_{G2}$

$X_{G1} < X_{G2}$
(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!!



Practice with causal interpretation of ANOVA results ...

Which of the following might and can't be given a causal interpretation ?

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

Can't tell ??? Remember, you need the operational definitions & procedural info. to know if this is a True Exp, with RA, etc...



ANOVA (F) - - - 2 Types

- **Between Groups ANOVA**
 - we collect the quantitative variable from each participant, who is only ever in **one** condition of the qualitative variable
 - also called “between subjects”, “independent groups” or “independent subjects”
- **Within-Groups ANOVA**
 - we collect the quantitative variable from each participant, who is in **all** conditions of the qualitative variable (one at a time, of course)
 - also called “within-subjects”, “dependent groups”, “dependent subjects”, or “repeated measures” designs

Between Subjects (Between Groups)

Control		Experimental	
Kim	Kate	Will	Sue
John	Sam	Dani	Dom

Within-subjects (Within-groups, Repeated Measures)

Control		Experimental	
Kim	Kate	Kim	Kate
John	Sam	John	Sam

Comparison of BG & WG ANOVA

- Data collection is different for the two
 - BG - each participant is only in one condition
 - WG - each participant will be in all conditions
- Same H0: (No mean difference)
- Same kind of RH (expected mean difference)
- Computation is slightly different for BG & WG
- H0: testing is the same $p < .05$ or $F\text{-crit} < F\text{-obt}$
- Determining Causality is the Same
 - random assignment (initial equivalence)
 - manipulation of IV (temporal precedence and ongoing eq)
 - control of procedural variables (ongoing equivalence)

Practice With Determining Whether Design is BG or WG

Here's two different versions of a study to test that "more practice leads to better performance"-- which is BG and which WG

- Each person is introduced to the task and either performs immediately (score is # correct of 10), or is given 30 practices first and then performs (and is scored the same way).
- Each person is introduced to the task and immediately performs (score is # correct out of 10). Is then given 30 practices and performs again (getting another scores of #correct out of 10).



... people 70+ years old
... night or during daylight hours.
... day and then during the night
... (u

... e
... lat
... o e
... as
... ew



A researcher has decided that watching you demented. 10 people were in the another 10 people watched the Simpsons the end of the week, the participants filled out a dementia scale.

To test the notion that Tokay Geckos are more active feeders at night than during the day, the researcher followed each of 12 geckos around for 24 hours and recorded the number of bugs (pets, small children, whatever) they ate during lighted and dark hours.



Your turn – tell type of design & pick which 2 variables to specify in SPSS ...

We wanted to know whether females or males studied more hours per week for this class.

Design ? What's the IV? What's the DV?

Which two variables would go into SPSS (yes, pick 2 – only 2)

- gender
- # hours males study
- # hours study
- # hours females study

Draw the boxes!

--	--

We wanted to know whether students studied more hours per week for the lecture or the laboratory of this class.

Design ? What's the IV? What's the DV?

Which two variables would go into SPSS (yes, pick 2 – only 2)

- lecture vs. lab
- # hours study
- # hours study for lecture
- # hours study for lab

--	--



What about t-tests?

- Whenever you want to compare the means from a 2- Between Groups design, you can use either a BG ANOVA or a BG t-test
- Whenever you want to compare the means from a 2-Within-Groups design, you can use either a WG ANOVA or a WG t-test

The two procedures will produce exactly the same:

- group means
- p-value & NHST results
- $t^2 = F$
- ANOVA $df_{error} = t\text{-test } df$

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!)