ANalysis Of VAriance

Dr. Tom Ilvento
FREC 408

ANOVA

- ANOVA provides a strategy to compare two or more population means associated with various treatments
- It is used when we have
  - A dependent variable (AKA Response Variable)
  - One or more categorical variables (ordinal) or continuous variables that are thought of as independent variables that influence the dependent variables
  - E.g. levels of fertilizer; group membership; different treatments

ANOVA

- ANOVA is used heavily in experimental designs in the biological sciences
  - Treatment versus control groups
  - Levels of treatment of a drug
  - Levels of applications of fertilizers or pesticides
- It is possible to have continuous and categorical independent variables
- Its origins were in agricultural studies

Elements of a Designed Experiment

- **Response Variable**: the variable of interest to be measured in the experiment. (Def11.3 p650)
  - Also known as the **dependent variable**.
- **Factors**: variables which are thought to influence the response variable (Def11.5 p650)
  - Quantitative
  - Qualitative
- **Factor Levels**: the levels of the factor that are experimentally manipulated (Def11.6 p650)
  - In a single factor experiment, the factor levels are the **treatments**

Elements of a Designed Experiment

- **Treatments:** when two or more factors are utilized, the treatments are the combinations of factor levels used in the experiment. (Def11.7 p650)
  - Factor 1: fertilizer (low; medium; high)
  - Factor 2: water (low; high)
    - Treatment 1: low fertilizer, low water
    - Treatment 2: low fertilizer, high water
    - Treatment 3: medium fertilizer, low water
    - And so forth.....

Elements of a Designed Experiment

- **Experimental Unit:** the objects on which the response variable and factors are observed (Def11.4 p650)
  - People
  - Plants
  - Animals
  - Schools

Designed versus observational

- **Designed Experiment**
  - The specification of the treatments
  - And the way experimental units are assigned to treatments is under the control of the researcher
- **Observational Study**
  - The researcher observes the treatments and the response on a sample of experimental units

Completely Randomized Design

- The treatments are **randomly assigned** to the experimental units
- **Or independent random samples** of experimental units are selected from target populations for each treatment (Def11.8 p655)
  - The book refers to both designed and observational studies as being randomized designs
  - Most think of ANOVA for designed experiments

What is ANOVA? HINT!!!!!!

It is all about the variance!

It is all about the variance!

It is all about the variance!

It is all about the variance!

It's the variance!

It is all about the variance!

Book Example of contrasts of two samples, p653

Obs	Sample 1	Sample 2
1	6	6
2	-1	1
3	0	3
4	4	7
5	1	6
Sum	10.0	25.0
Mean	2.0	5.0
Var	6.5	6.5
Std Dev	2.9	2.9

Difference of Means Test

t-test: Two-Sample Assuming Equal Variances

	Sample 2	Sample 1
Mean	5.00	2.00
Variance	0.50	0.50
Observations	5	5
Pooled Variance	0.5	
Hypothesized Mean Differenc	0	
df	8	
t Stat	6.708	
P(T<=t) one-tail	0.000	
t Critical one-tail	1.860	
P(T<=t) two-tail	0.000	
t Critical two-tail	2.306	

Based on a

$t^* = 6.708$ and an $\alpha = .05$,

I would reject

Ho: $\mu_2 - \mu_1 = 0$

Book Example of contrasts of two samples, p653

- *One way to determine whether a difference exists between the population means is to examine the difference between the sample means and compare it to a measure of variability within the samples.* P653
- The difference of the two means is only part of the story.
- The other part is the variability and separation of the two samples
- ANOVA uses this strategy to compare two or more means

ANOVA

- We will **decompose** the variance of our dependent variable
  - Part **due to the treatments** or independent variables – part that is explained
  - Part that is unexplained or **random "error"**
- I will adjust these variances from different sources by dividing by degrees of freedom to get an average deviation

ANOVA

- We will decompose
- **Total Sum of Squares =**
  - **Sum of Squares for Treatment +**
  - **Sum of Squares for Error**

$$SS(Total) = \sum_{i=1}^{n} (y_i - \bar{Y})^2$$

Let's Look at the data comparing males and female cholesterol level

Cholesterol Level

	Females	Males
Mean	200.32	196.09
Standard Error	0.88	0.97
Median	201	196
Mode	194	196
Standard Deviation	10.72	12.37
Sample Variance	114.94	153.07
Kurtosis	-0.49	0.02
Skewness	-0.11	0.09
Range	47	61
Minimum	176	166
Maximum	223	227
Sum	29647	32158
Count	148	164
Confidence Level(95.0%)	1.74	1.91

How might we approach this data?

- A large sample difference of means test
- ANOVA - randomized design for an observational study
  - **Response Variable:** cholesterol level
  - **Factor:** gender
    - Females
    - Males
  - **Experimental Units:** people