Lecture 7: Hypothesis Testing and ANOVA

Goals

- Overview of key elements of hypothesis testing

- Review of common one and two sample tests

- Introduction to ANOVA

Hypothesis Testing

- The intent of hypothesis testing is formally examine two opposing conjectures (hypotheses), H_0 and H_A

- These two hypotheses are mutually exclusive and exhaustive so that one is true to the exclusion of the other

- We accumulate evidence - collect and analyze sample information - for the purpose of determining which of the two hypotheses is true and which of the two hypotheses is false

The Null and Alternative Hypothesis

The null hypothesis, H$_0$:

- States the assumption (numerical) to be tested

- Begin with the assumption that the null hypothesis is TRUE

- Always contains the '=' sign

The alternative hypothesis, H$_a$:

- Is the opposite of the null hypothesis

- Challenges the status quo

- Never contains just the '=' sign

- Is generally the hypothesis that is believed to be true by the researcher

One and Two Sided Tests

- Hypothesis tests can be one or two sided (tailed)

- One tailed tests are directional:

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_A: \mu_1 - \mu_2 > 0$$

- Two tailed tests are not directional:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

P-values

- Calculate a test statistic in the sample data that is relevant to the hypothesis being tested

- After calculating a test statistic we convert this to a P-value by comparing its value to distribution of test statistic's under the null hypothesis

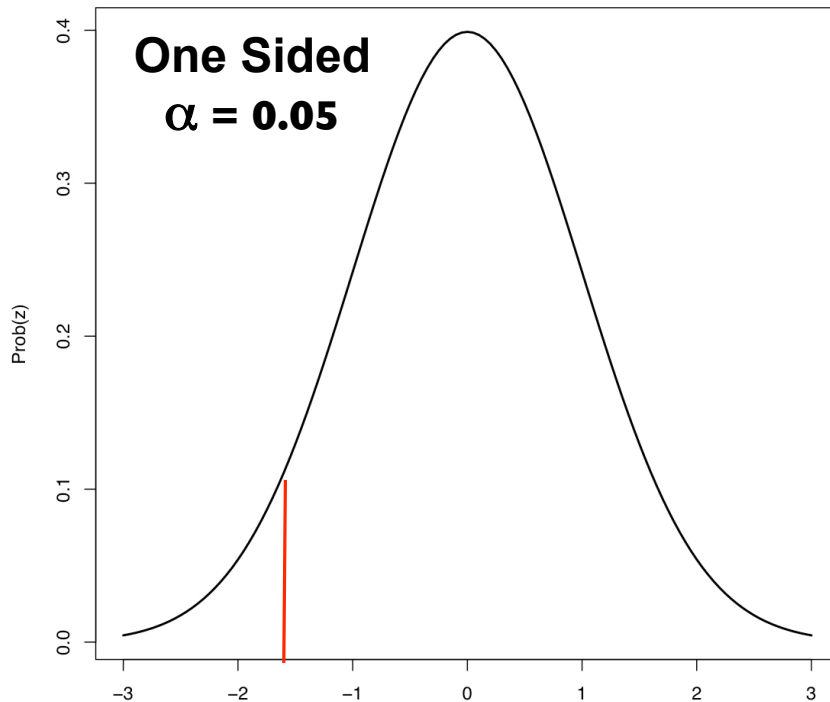- Measure of how likely the test statistic value is under the null hypothesis

$$\text{P-value} \leq \alpha \Rightarrow \text{Reject } H_0 \text{ at level } \alpha$$

$$\text{P-value} > \alpha \Rightarrow \text{Do not reject } H_0 \text{ at level } \alpha$$
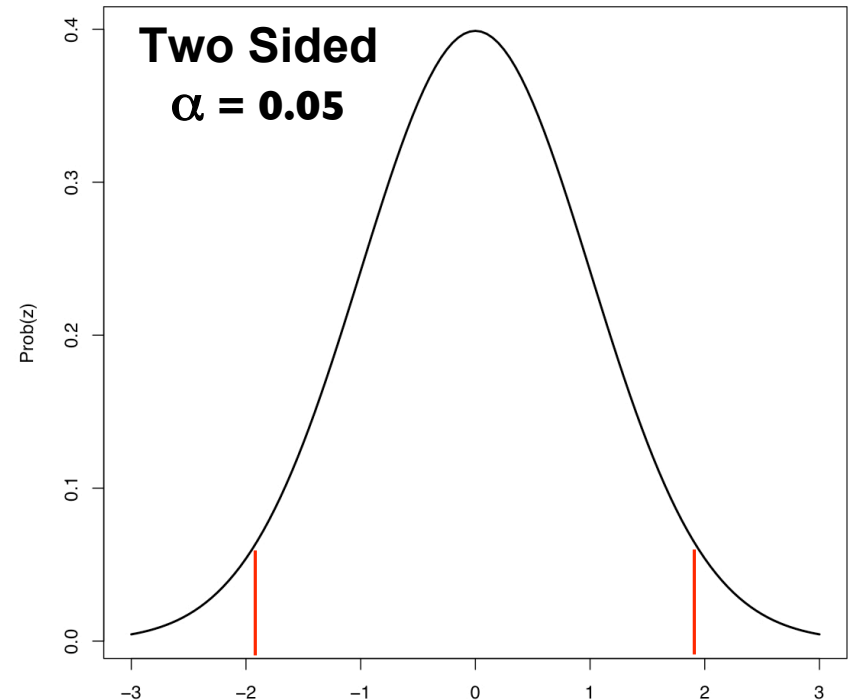
When To Reject H$_0$

Level of significance, α: Specified before an experiment to define rejection region

Rejection region: set of all test statistic values for which H$_0$ will be rejected

Critical Value = -1.64

Critical Values = -1.96 and +1.96

Some Notation

- In general, critical values for an α level test denoted as:

$$\text{One sided test}: \quad X_{\alpha}$$

$$\text{Two sided test}: \quad X_{\alpha/2}$$

where X depends on the distribution of the test statistic

- For example, if X ~ N(0,1):

$$\text{One sided test}: \quad z_{\alpha} \quad (\text{i.e., } z_{0.05} = 1.64)$$

$$\text{Two sided test}: \quad z_{\alpha/2} \quad (\text{i.e., } z_{0.05/2} = z_{0.025} = \pm 1.96)$$

Errors in Hypothesis Testing

Actual Situation "Truth"

Decision	H_0 True	H_0 False
Do Not Reject H_0		
Rejct H_0		

Errors in Hypothesis Testing

Actual Situation "Truth"

Decision	H_0 True	H_0 False
Do Not Reject H_0	Correct Decision $1 - \alpha$	Incorrect Decision β
Rejct H_0	Incorrect Decision α	Correct Decision $1 - \beta$

Type I and II Errors

Actual Situation "Truth"

Decision	H_0 True	H_0 False
Do Not Reject H_0	Correct Decision $1 - \alpha$	Incorrect Decision Type II Error β
Rejct H_0	Incorrect Decision Type I Error α	Correct Decision $1 - \beta$

$$\alpha = P(Type\ I\ Error) \quad \beta = P(Type\ II\ Error)$$

$$Power = 1 - \beta$$

Parametric and Non-Parametric Tests

- **Parametric Tests**: Relies on theoretical distributions of the test statistic under the null hypothesis and assumptions about the distribution of the sample data (i.e., normality)

- **Non-Parametric Tests**: Referred to as "Distribution Free" as they do not assume that data are drawn from any particular distribution

Whirlwind Tour of One and Two Sample Tests

Goal	Type of Data		
	Gaussian	Non-Gaussian	Binomial
Compare one group to a hypothetical value	One sample t-test	Wilcoxon Test	Binomial Test
Compare two paired groups	Paired t-test	Wilcoxon Test	McNemar's Test
Compare two unpaired groups	Two sample t-test	Wilcoxon-Mann-Whitney Test	Chi-Square or Fisher's Exact Test

General Form of a t-test

	One Sample	Two Sample
Statistic	$T = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$	$T = \dfrac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{s_p \sqrt{\dfrac{1}{m} + \dfrac{1}{n}}}$
df	$t_{\alpha, n-1}$	$t_{\alpha, m+n-2}$

Non-Parametric Alternatives

- **Wilcoxon Test**: non-parametric analog of one sample t-test

- **Wilcoxon-Mann-Whitney test**: non-parametric analog of two sample t-test

Hypothesis Tests of a Proportion

- **Large sample test (prop.test)**

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

- **Small sample test (binom.test)**

  - Calculated directly from binomial distribution

Confidence Intervals

- **Confidence interval:** an interval of plausible values for the parameter being estimated, where degree of plausibility specifided by a "confidence level"
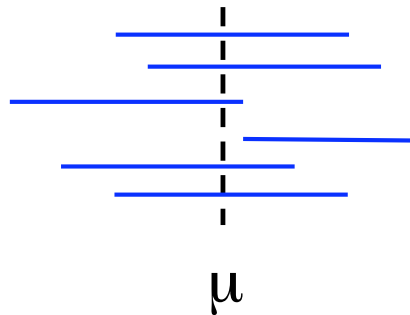
- **General form:**

$$\hat{x} \ \pm \ \text{critical value}_\alpha \ \bullet \ \text{se}$$

$$\bar{x} - \bar{y} \ \pm \ t_{\alpha, m+n-2} \ \bullet \ s_p \sqrt{\frac{1}{m} + \frac{1}{n}}$$

Interpreting a 95% CI

- We calculate a 95% CI for a hypothetical sample mean to be between 20.6 and 35.4. Does this mean there is a 95% probability the true population mean is between 20.6 and 35.4?

- **NO!** Correct interpretation relies on the long-rang frequency interpretation of probability

μ

- Why is this so?

Hypothesis Tests of 3 or More Means

- Suppose we measure a quantitative trait in a group of N individuals and also genotype a SNP in our favorite candidate gene. We then divide these N individuals into the three genotype categories to test whether the average trait value differs among genotypes.

- What statistical framework is appropriate here?

- Why not perform all pair-wise t-tests?

Basic Framework of ANOVA

- Want to study the effect of one or more *qualitative* variables on a *quantitative* outcome variable

- Qualitative variables are referred to as *factors* (i.e., SNP)

- Characteristics that differentiates factors are referred to as *levels* (i.e., three genotypes of a SNP

One-Way ANOVA

- **Simplest case is for One-Way (Single Factor) ANOVA**

  ➢ The *outcome* variable is the variable you're comparing

  ➢ The *factor* variable is the categorical variable being used to define the groups
    - We will assume k samples (groups)

  ➢ The *one-way* is because each value is classified in exactly one way

- **ANOVA easily generalizes to more factors**

Assumptions of ANOVA

- **Independence**

- **Normality**

- **Homogeneity of variances** (aka, Homoscedasticity)

One-Way ANOVA: Null Hypothesis

- The null hypothesis is that the means are all equal

$$H_0 \; : \; \mu_1 = \mu_2 = \ldots = \mu_k$$

- The alternative hypothesis is that at least one of the means is different
  - Think about the Sesame Street® game where three of these things are kind of the same, but one of these things is not like the other. They don't all have to be different, just one of them.

Motivating ANOVA

- A random sample of some quantitative trait was measured in individuals randomly sampled from population

- Genotyping of a single SNP
  - AA: 82, 83, 97
  - AG: 83, 78, 68
  - GG: 38, 59, 55

Rational of ANOVA

- Basic idea is to partition total variation of the data into two sources

    1. Variation within levels (groups)

    2. Variation between levels (groups)

- If H_0 is true the *standardized* variances are equal to one another

The Details

Our Data:

AA:	82, 83, 97	$\bar{x}_{1.} = (82 + 83 + 97)/3 = 87.3$
AG:	83, 78, 68	$\bar{x}_{2.} = (83 + 78 + 68)/3 = 76.3$
GG:	38, 59, 55	$\bar{x}_{3.} = (38 + 59 + 55)/3 = 50.6$

- Let X_{ij} denote the data from the ith level and j^{th} observation

- Overall, or **grand mean**, is:

$$\bar{x}_{..} = \sum_{i=1}^{K}\sum_{j=1}^{J}\frac{x_{ij}}{N}$$

$$\bar{x}_{..} = \frac{82 + 83 + 97 + 83 + 78 + 68 + 38 + 59 + 55}{9} = 71.4$$

Partitioning Total Variation

- Recall, variation is simply average squared deviations from the mean

$$SST \quad = \quad SST_G \quad + \quad SST_E$$

$$\sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij}-\bar{x}_{..})^2 \qquad \sum_{i=1}^{K} n_i \bullet (\bar{x}_{i.}-\bar{x}_{..})^2 \qquad \sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij}-\bar{x}_{i.})^2$$

Sum of squared deviations about the grand mean across all N observations

Sum of squared deviations for each group mean about the grand mean

Sum of squared deviations for all observations within each group from that group mean, summed across all groups

In Our Example

$$SST \quad = \quad SST_G \quad + \quad SST_E$$

$$\sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{..})^2 \qquad \sum_{i=1}^{K} n_i \bullet (\bar{x}_{i.} - \bar{x}_{..})^2 \qquad \sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{i.})^2$$

$(82-71.4)^2 + (83-71.4)^2 + (97-71.4)^2 +$

$(83-71.4)^2 + (78-71.4)^2 + (68-71.4)^2 +$

$(38-71.4)^2 + (59-71.4)^2 + (55-71.4)^2 =$

$3 \bullet (87.3-71.4)^2 +$

$3 \bullet (76.3-71.4)^2 +$

$3 \bullet (50.6-71.4)^2 =$

$(82-87.3)^2 + (83-87.3)^2 + (97-87.3)^2 +$

$(83-76.3)^2 + (78-76.3)^2 + (68-76.3)^2 +$

$(38-50.6)^2 + (59-50.6)^2 + (55-50.6)^2 =$

2630.2 $\qquad\qquad$ 2124.2 $\qquad\qquad$ 506

In Our Example

$$SST \quad = \quad SST_G \quad + \quad SST_E$$

$$\sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij}-\bar{x}_{..})^2 \qquad \sum_{i=1}^{K}n_i \bullet (\bar{x}_{i.}-\bar{x}_{..})^2 \qquad \sum_{i=1}^{K}\sum_{j=1}^{J}(x_{ij}-\bar{x}_{i.})^2$$
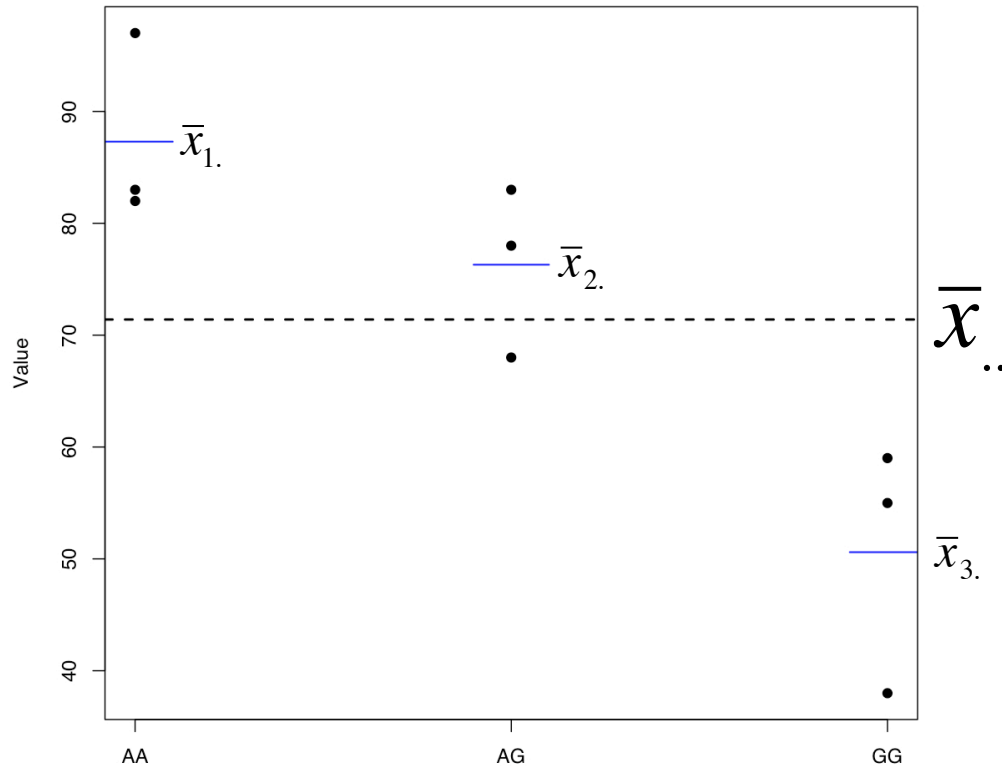
Calculating Mean Squares

- To make the sum of squares comparable, we divide each one by their associated degrees of freedom

  - $SST_G = k - 1$ $(3 - 1 = 2)$

  - $SST_E = N - k$ $(9 - 3 = 6)$

  - $SST_T = N - 1$ $(9 - 1 = 8)$

- $MST_G = $ 2124.2 / 2 = 1062.1

- $MST_E = $ 506 / 6 = 84.3

Almost There… Calculating F Statistic

- The test statistic is the ratio of group and error mean squares

$$F = \frac{MST_G}{MST_E} = \frac{1062.2}{84.3} = 12.59$$

- If H_0 is true MST_G and MST_E are equal

- Critical value for rejection region is $F_{\alpha, k-1, N-k}$

- If we define $\alpha = 0.05$, then $F_{0.05, 2, 6} = 5.14$

ANOVA Table

Source of Variation	df	Sum of Squares	MS	F
Group	k-1	SST_G	$\dfrac{SST_G}{k-1}$	$\dfrac{SST_G}{k-1} \Big/ \dfrac{SST_E}{N-k}$
Error	N-k	SST_E	$\dfrac{SST_E}{N-k}$	
Total	N-1	SST		

Non-Parametric Alternative

- Kruskal-Wallis Rank Sum Test: non-parametric analog to ANOVA

- In R, kruskal.test()