# Hypothesis Testing

Clayton W. Schupp

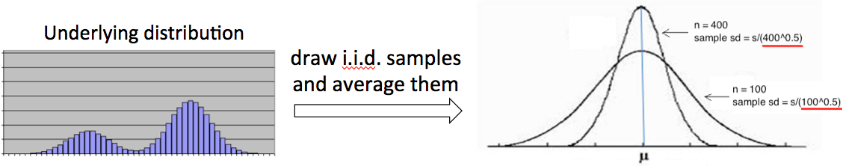
Galvanize

## Morning Lecture Objectives

- The general steps of a statistical hypothesis test
- One-tail vs. Two-tailed tests
- Type-I and Type-II Error
- One-sample and Two-sample tests of mean
- One-sample and Two-sample tests of proportion

# Central Limit Theorem

The CLT states that given certain conditions, the mean of a sufficiently large number of *i.i.d* random variables will be approximately normal, regardless of the underlying distribution



# Central Limit Theorem

- Not only is the sample mean normally distributed, but the variance of the sample mean is smaller

$$\bar{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$$

- As with any normal variable, we can derive a standard normal Z-score

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

# Hypothesis Testing

- Estimation
  - Parameter value is unknown
  - Goal is to find a point estimate and a confidence interval for likely values
- Hypothesis Testing
  - Parameter value is stated
  - Goal is to see if the parameter has changed

# General Steps of Hypothesis Testing

1. State the null hypothesis ($H_0$) and alternative hypothesis ($H_A$)
2. Choose the significance level, $\alpha$
3. Compute the appropriate test statistic
4. Compute the p-value under the assumption that $H_0$ is true
   - If p-value $\leq \alpha \longrightarrow$ Reject $H_0$ in favor of $H_A$
   - if p-value $> \alpha \longrightarrow$ Fail to Reject $H_0$

# Null Hypothesis vs. Alternative Hypothesis

- Null Hypothesis ($H_0$)
  - Typically a measure of the status quo such as no effect
  - In terms of the parameter: $H_0 : \mu = 0$
- Alternative Hypothesis ($H_A$)
  - Usually states the effect the researcher hopes to detect
  - Example: Advertising causes 1% lift
  - In terms of the parameter: $H_A : \mu \neq 0$

## Two-sided vs One-sided tests

- By default, we should compute a two-sided test which is more conservative:
  - For example: $H_0 : \mu = \mu_0$ vs $H_A : \mu \neq \mu_0$
  - Reject if test statistic in upper or lower tail
  - One half of p-value in each tail
- However, if we expect the effect to be in a specific direction, we can use a one-sided test:
  - Example: $H_0 : \mu \leq \mu_0$ vs $H_A : \mu > \mu_0$
  - Reject $H_0$ if test statistic in tail designated by $H_A$
  - P-value calculated based on direction specified in $H_A$

# Two-sided vs One-sided tests

| Direction | $H_0$ | $H_A$ | P-value |
|---|---|---|---|
| 2-sided Test | $=$ | $\neq$ | One half of P-value in each tail |
| Left-Tail | $\geq$ | $<$ | All of P-value in left tail |
| Right-Tail | $\leq$ | $>$ | All of P-value in right tail |

# Type I and Type II Errors

- Type I error: Rejecting $H_0$ when it is true
- Type II error: Failing to reject $H_0$ when it is false

| | $H_0$ is True | $H_0$ is False |
|---|---|---|
| Fail to Reject $H_0$ | Correct Decision $(1-\alpha)$ | Type II Error $(\beta)$ |
| Reject $H_0$ | Type I Error $(\alpha)$ | Correct Decision $(1-\beta)$ |

Type I Error is also the Level of Significance

## Multiple Comparisons

- If a researcher wants to conduct multiple tests (i.e. make multiple comparisons), we need to adjust the individual $\alpha_I$ rate so that the overall experimental $\alpha_E$ rate remains at the desired level.
- Bonferroni correction to the individual rate is straightforward, easy to apply, but overly conservative

$$\alpha_I = \frac{\alpha_E}{m} \text{ where } m \text{ is number of comparisons}$$

## Multiple Comparisons Example:

- Lets say you have 5 categories and you want to make all possible pairwise comparisons of the mean

$$\mu_i = \mu_j \quad \forall i \neq j$$

- This gives us 10 possible comparisons. If we use $\alpha_I = 0.05$ then the overall error rate would be

$$\alpha_E = 1 - (1 - \alpha_I)^{10} = (1 - 0.95^{10}) = 0.401$$

So we would have a 40% chance of falsely rejecting at least one of the null hypotheses

# One Sample Test of Population Mean

Z-test used when $\sigma^2$ is known or $n \geq 30$

1. Default choice should be a two-sided test: $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$
2. Level of significance, $\alpha$
3. Calculating standardized test statistic:

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu_0, \sigma^2) \xrightarrow{CLT} \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

$$z_{ts} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

4. P-value for a 2-sided test:

$$\text{P-value} = 2P(Z > |z_{ts}|)$$

# One Sample Test of Population Mean

T-test used when $\sigma^2$ is unknown and $n < 30$

1. Example of a right-tail test: $H_0 : \mu \leq \mu_0$ vs. $H_A : \mu > \mu_0$
2. Level of significance, $\alpha$
3. Calculating standardized test statistic:

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu_0, \sigma^2) \xrightarrow{CLT} \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

$$t_{ts} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t_{df} \text{ where } df = n - 1$$

4. P-value for a right-sided test:

$$\text{P-value} = P(T > t_{ts})$$

# One Sample Test of Population Proportion

1. Example of a left-sided test: $H_0 : p \geq p_0$ vs. $H_A : p < p_0$
2. Level of significance, $\alpha$
3. Calculating standardized test statistic:

$$X \sim Bin(n, p_0) \xrightarrow{CLT} \hat{p} = \frac{x}{n} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$$

$$z_{ts} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

4. P-value for a left-sided test:

$$\text{P-value} = P(Z < z_{ts})$$

# Two Sample Test of Difference in Population Means

1. $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_A : \mu_1 - \mu_2 \neq 0$
2. Level of significance, $\alpha$
3. Calculating standardized test statistic:

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ and } \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\longrightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$t_{ts} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{df}$$

4. P-value for a 2-sided test:

$$\text{P-value} = 2P(T > |t_{ts}|)$$

# Two Sample Test of Difference in Population Means

- With no assumptions about the population variances we utilize the formula on the previous slide with

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$$

For simplicity, you can always choose the more conservative

$$df = min(n_1 - 1, n_2 - 1)$$

- Assumption of equal variance ($\sigma_1^2 = \sigma_2^2$), use the pooled variance estimator in the test statistic:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t_{ts} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{df} \text{ where } df = n_1 + n_2 - 2$$

# Two Sample Test of Difference in Population Proportions

1. $H_0 : p_1 - p_2 = 0$ vs. $H_A : p_1 - p_2 \neq 0$
2. Level of significance, $\alpha$
3. Calculating standardized test statistic:

$$\hat{p}_1 = \frac{x_1}{n_1} \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right) \quad \hat{p}_2 = \frac{x_2}{n_2} \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

$$z_{ts} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \text{ where } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

4. P-value for a two-sided test:

$$\text{P-value} = 2P(Z > |z_{ts}|)$$

# Two Sample Test of Difference in Population Proportions

1. $H_0 : p_1 - p_2 = D$ vs. $H_A : p_1 - p_2 \neq D$
2. Level of significance, $\alpha$
3. Calculating standardized test statistic:

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right) \text{ and } \hat{p}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

$$z_{ts} = \frac{\hat{p}_1 - \hat{p}_2 - D}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

4. P-value for a two-sided test:

$$\text{P-value} = 2P(Z > |z_{ts}|)$$