Lecture 28
Review

STAT 8010 Statistical Methods I
October 28, 2019

Whitney Huang
Clemson University

Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Agenda

1. Inferences for One Population Mean
2. Inferences for Two Population Means
3. Inferences for Matched Pairs
4. ANOVA
5. Multiple Comparisons & Linear Contrasts

Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Inferences for One Population Mean

Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Inferences for One Population Mean μ

Goal: To infer $\mu = \mathbb{E}(X)$ from a random sample $\{X_1, X_2, \cdots, X_n\}$

- Point estimation:

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

- Interval Estimation: $100 \times (1-\alpha)\%$ Confidence Interval (CI)
  - $\sigma = \sqrt{\text{Var}(X)}$ is known:

$$\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

  - σ is unknown:

$$\left(\bar{X}_n - t_{\alpha/2, df=n-1}\frac{s}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2, df=n-1}\frac{s}{\sqrt{n}}\right)$$

Notes

Margin of error & Sample Size Calculation

- Margin of error:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{if } \sigma \text{ known}$$

$$t_{\alpha/2, df=n-1}\frac{s}{\sqrt{n}} \quad \text{if } \sigma \text{ unknown}$$

$\Rightarrow$ CI for $\mu = \bar{X}_n \pm$ margin of error

- Sample size determination:

$$n = \left(\frac{z_{\alpha/2} \times \sigma}{\text{margin of error}}\right)^2,$$

if σ is given

Notes

Hypothesis Testing for μ

1. State the null and alternative hypotheses:

$$H_0 : \mu = \mu_0 \text{ vs. } H_a : \mu > \text{ or } \neq \text{ or } < \mu_0$$

2. Compute the test statistic:

$$z_{obs} = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}, \quad \sigma \text{ known; } t_{obs} = \frac{\bar{X}_n - \mu_0}{s/\sqrt{n}}, \quad \sigma \text{ unknown}$$

3. Make the decision of the test:

   Rejection Region/ P-Value Methods

4. Draw the conclusion of the test:
   We (do/do not) have enough statistical evidence to conclude that (H_a in words) at α% significant level.

Notes

Type I, II Error & Power

True State	Decision	
	Reject H_0	Fail to reject H_0
H_0 is true	Type I error	Correct
H_0 is false	Correct	Type II error

- Type I error: $\mathbb{P}(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha$
- Type II error: $\mathbb{P}(\text{Fail to reject } H_0 | H_0 \text{ is false}) = \beta$
- The power (PWR): $\mathbb{P}(\text{Reject } H_0 | H_0 \text{ is false}) = 1 - \beta$.

$$\Rightarrow \text{PWR}(\mu_a) = 1 - \beta(\mu_a) = 1 - \mathbb{P}(z^* \leq z_\alpha - \frac{|\mu_0 - \mu_a|}{\sigma/\sqrt{n}})$$

(see the figure in page 5, Lecture 20)

Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Duality of Hypothesis Test with Confidence Interval

There is an interesting relationship between CIs and hypothesis tests. If H_0 is rejected with significance level α then the corresponding confidence interval does not contain the value μ_0 targeted in the hypotheses with the confidence level $(1 - \alpha)$, and vice versa

Hypothesis testing at α level	$(1 - \alpha)$-level CI
$H_0 : \mu = \mu_0$ vs. $H_a : \mu \neq \mu_0$	$\bar{X} \pm t(\alpha/2, n-1)s/\sqrt{n}$
$H_0 : \mu = \mu_0$ vs. $H_a : \mu > \mu_0$	$(\bar{X} - t(\alpha/2, n-1)s/\sqrt{n}, \infty)$
$H_0 : \mu = \mu_0$ vs. $H_a : \mu < \mu_0$	$(-\infty, \bar{X} + t(\alpha/2, n-1)s/\sqrt{n})$

Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Inferences for Two Population Means

Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Statistical Inference for $\mu_1 - \mu_2$

- Point estimation: $\bar{X}_1 - \bar{X}_2$
- Interval estimation:

$$\bar{X}_1 - \bar{X}_2 \pm \text{ margin of error},$$

where margin of error $=$

$$t_{\alpha/2, df^*}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad df^* = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

When s_1 and s_2 "similar enoug", we replace $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ by $s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, where $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

Notes

Hypothesis Testing for $\mu_1 - \mu_2$

- State the null and alternative hypotheses:
  - Upper-tailed test: $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 > 0$
  - Lower-tailed test: $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 < 0$
  - Two-tailed test: $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 \neq 0$
- Compute the test statistic:

$$t_{obs} = \begin{cases} \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, & \sigma_1 = \sigma_2 \\ \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}, & \sigma_1 \neq \sigma_2 \end{cases}$$

- Make the decision of the test:
  Rejection Region/ P-Value Methods
- Draw the conclusion of the test

Review
CLEMSON UNIVERSITY
Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Inferences for Matched Pairs

Review
CLEMSON UNIVERSITY
Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Paired T-Tests

- **When to use**: before/after study, pairing subjects, study on twins, etc
- $H_0 : \mu_{diff} = 0$ vs. $H_a : \mu_{diff} > 0$ or $\mu_{diff} < 0$ or $\mu_{diff} \neq 0$, where μ_{diff} is the population mean of the paired difference
- Test statistic: $t_{obs} = \frac{\bar{X}_{diff} - 0}{\frac{s_{diff}}{\sqrt{n}}}$

Notes

ANOVA

Notes

ANOVA and Overall F Test

Overall F-Test

- $H_0 : \mu_1 = \mu_2 = \cdots = \mu_J$
  H_a : At least one mean is different
- ANOVA Table:

Source	df	SS	MS	F statistic
Treatment	$J-1$	SSTr	$\text{MSTr} = \frac{\text{SSTr}}{J-1}$	$F = \frac{\text{MSTr}}{\text{MSE}}$
Error	$N-J$	SSE	$\text{MSE} = \frac{\text{SSE}}{N-J}$	
Total	$N-1$	SSTo		

- Test Statistic: $F_{obs} = \frac{\text{MSTr}}{\text{MSE}}$. Under H_0, $F^* \sim F_{df_1=J-1, df_2=N-J}$

Notes

Multiple Comparisons & Linear Contrasts

Notes

Family-Wise Error Rate (FWER) and Mulitple Comparisons

- Family-Wise Error Rate (FWER) $\bar{\alpha}$: the probability of making 1 or more type I errors in a set of hypothesis tests
- Bonferroni Correction: Adjust the significant level for each of the m tests to be $\frac{\alpha}{m}$ to control the **FWER**
- Fisher's LSD and Tukey's HSD

Review
CLEMSON UNIVERSITY
Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes

Linear Contrasts

- **Definition**: Let $c_1, c_2, \cdots, c_J$ are constants where $\sum_{j=1}^{J} c_j = 0$, then $L = \sum_{j=1}^{J} c_j\mu_j$ is called a **linear contrast** of the population means.
- Point Estimation:

$$\hat{L} = \sum_{j=1}^{J} c_j \bar{X}_j$$

- Interval Estimation:

$$(\hat{L} - t_{(\alpha/2, df=N-J)}\hat{se}_{\hat{L}}, \hat{L} + t_{(\alpha/2, df=N-J)}\hat{se}_{\hat{L}}),$$

  where $\hat{se}_{\hat{L}} = \sqrt{\text{MSE}\left(\frac{c_1^2}{n_1} + \cdots + \frac{c_J^2}{n_J}\right)}$
- Hypothesis Testing for linear contrasts

Review
CLEMSON UNIVERSITY
Inferences for One Population Mean
Inferences for Two Population Means
Inferences for Matched Pairs
ANOVA
Multiple Comparisons & Linear Contrasts

Notes