12/5/2023

1. Office Hours
   12/6(Wed): 10AM-12PM
2. Hw10 due today but will allow 12-hour extension without penalty.
3. Final Exam
   a. Normal (positive part), t, chi-square, and f tables will also be provided.
   b. Stand-alone calculator. Don't use your calculator to find critical values and p-values. Use Tables!!
   c. Final exam: 2:40PM – 5:30PM on 12/7 (Thursday). NO EXCUSE, PLEASE!!
   d. Location: Same classroom
4. Coverage

Coverage for Final:

Chap 6 – Descriptive Statistics

a. Summary Statistics
b. Stem-and-Leaf
c. Histogram
d. Box plots
e. Probability plots ( ad.test() )

Chap 7 – Point Estimation

a. Unbiased estimator
b. Mean squared Error of an estimator
c. Relative Efficiency
d. MLE

Chap 8 – Confidence Interval

Ch8-1 ~ Ch8-4 except Choice of sample size

Chap 9 – One-sample hypothesis testing

Ch9-1~Ch9-5 except choice of sample size

Chap 10 – Two-Sample hypothesis testing

Ch10-1, Ch10-2

[Note] You should be familiar with the **R outputs** for Hypothesis testing (ie. t.test()). Some questions will be based on the R outputs.

Reminders:

1. Still need to know all basic definitions in statistics or probability, such as mean, variance, standard deviation, etc.
2. Sample variance $s^2$ vs Standard deviation $s$ → most common mistakes in the exams.
3. For the hypothesis testing, first identify if it's for one-sample, two-sample, or paired.
4. When you compute the testing statistics, check your hypothesis again. If your hypothesis is $H_0: \mu = \Delta_0, where\ \Delta_0 \neq 0$, don't forget to subtract it.
5. Then, identify if it's a two-sided or one-sided testing. If two-sided, don't forget to use $\frac{\alpha}{2}$!!
6. Compute the CI accordingly.
7. Use the correct degrees of freedom
8. Please clearly write down your hypotheses and be careful about two-sided or one-sided testing.
9. Please show all the work. Don't just give me $t_0 = 1.72$. You will lose points for this!!
10. Find the p-value or the range of p-value from tables.
11. Get familiar with your formula sheets. Know where you can find the right formulas to use.

**<u>CI – Ch8</u>**

| Parameter | $x_1, x_2, \dots, x_n$ are from | CI |
|---|---|---|
| $\mu$ | Normal with known variance $\sigma^2$ | $\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ |
| $\mu$ | Any distribution with large sample size $n \geq 40$, unknown $\sigma^2$ | $\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$ |
| $\mu$ | Normal with unknown variance, small sample size, $n < 40$ | $\bar{x} - t_{\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}}$ |
| $\sigma^2$ | Normal distribution | $\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$ |
| $p$ | Proportion | $\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ |

## <u>One Sample – Ch9</u>

1. Large sample ($n \geq 40$): $z_0 = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ $or$ $z_0 = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$

2. Small Sample: Normally distributed

   a. Known population variance: $z_0 = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

   b. Unknown population variance: $t_0 = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$

3. HT on variance: $\chi_0^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$

4. HT on proportion: similar to large sample: $\hat{p} = \frac{x}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)$

$$z_0 = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{n\hat{p} - np}{\sqrt{np(1-p)}}$$

**Two Samples – Ch10:**

1. Large sample ($n_1, n_2 \geq 40$): $z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $or\ z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

2. Small Sample: Normally distributed

a. Known population variances: $z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

b. Unknown variance:

I. Equal Variance Assumption: $t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$

where $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

II. Unequal Variance Assumption: $t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_v$

where $v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = 13.783 \approx 13$