Math 4720: Statistical Methods

8^{th} Week Summary (3/6/25)

Case 1: One Variable

	Numerical Variable, σ known	**Numerical Variable,** σ unknown
Parameter of Interest:	Mean, μ	Mean, μ
Confidence Interval Formula:	$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ with df = $n-1$
Name of Hypothesis Test, H_0	One Sample Z Test, $H_0: \mu = \mu_0$	One Sample T Test, $H_0: \mu = \mu_0$
Test Statistic Formula:	$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ with df = $n-1$
***p*-value:**	$H_a: \mu \neq \mu_0$, p-value = $2P(Z \geq \lvert z \rvert)$ $H_a: \mu > \mu_0$, p-value = $P(Z \geq z)$ $H_a: \mu < \mu_0$, p-value = $P(Z \leq z)$	$H_a: \mu \neq \mu_0$, p-value = $2P(T \geq \lvert t \rvert)$ $H_a: \mu > \mu_0$, p-value = $P(T \geq t)$ $H_a: \mu < \mu_0$, p-value = $P(T \leq t)$

Case 2: Two Numerical Variables – Population Standard Deviations unknown

	Independent Samples		**Paired Samples**
	$\boldsymbol{\sigma_1 = \sigma_2}$	$\boldsymbol{\sigma_1 \neq \sigma_2}$	
Parameters of Interest:	Mean Difference, $\mu_1 - \mu_2$	Mean Difference, $\mu_1 - \mu_2$	Mean Difference, $\mu_1 - \mu_2$
Confidence Interval Formula:	$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $df = n_1 + n_2 - 2$, where $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$	$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $df = \frac{(n_1-1)(n_2-1)}{(1-c)^2(n_1-1) + c^2(n_2-1)}$, where $c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$ with df = $n-1$ where the subscript "d" denotes the calculation is performed on the differences "Sample1" – "Sample2"
Name of Hypothesis Test, H_0	Two-sample T Test, $H_0: \mu_1 = \mu_2$ or $H_0: \mu_1 - \mu_2 = 0$	Two-sample T Test, $H_0: \mu_1 = \mu_2$ or $H_0: \mu_1 - \mu_2 = 0$	Paired T Test, $H_0: \mu_1 = \mu_2$ or $H_0: \mu_1 - \mu_2 = 0$
Test Statistic Formula:	$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$t' = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$ with df = $n-1$
P-value:	$H_a: \mu_1 \neq \mu_2$, p-value = $2P(T \geq \lvert t \rvert)$ $H_a: \mu_1 > \mu_2$, p-value = $P(T \geq t)$ $H_a: \mu_1 < \mu_2$, p-value = $P(T \leq t)$		$H_a: \mu_1 \neq \mu_2$, p-value = $2P(T \geq \lvert t \rvert)$ $H_a: \mu_1 > \mu_2$, p-value = $P(T \geq t)$ $H_a: \mu_1 < \mu_2$, p-value = $P(T \leq t)$

- **Power analysis** (Sample size determination):

  Independent Samples:

  For one sided alternative test ($H_a: \mu_1 > \mu_2$ or $H_a: \mu_1 < \mu_2$): $n = \frac{2\sigma^2(z_\alpha + z_\beta)^2}{\Delta^2}$

  For two sided alternative test ($H_a: \mu_1 \neq \mu_2$): $n = \frac{2\sigma^2(z_{\alpha/2} + z_\beta)^2}{\Delta^2}$

  Paired Data:

  For one sided alternative test ($H_a: \mu_d > 0$ or $H_a: \mu_d < 0$): $n = \frac{\sigma_d^2(z_\alpha + z_\beta)^2}{\Delta^2}$

  For two sided alternative test ($H_a: \mu_d \neq 0$): $n = \frac{\sigma_d^2(z_{\alpha/2} + z_\beta)^2}{\Delta^2}$

- **NON-PARAMETRIC TESTS** . . . .
- Test for the population median M:

  SIGN TEST:

  Test Statistics: B = # of data values $> m_0$

  Minitab: Stat → Nonparametrics → 1-Sample Sign

  R: library("BSDA"); SIGN.test(dataset)

- Non-parametric TWO Independent Samples Test:

  WILCOXON RANK-SUM TEST(MANN–WHITNEY U TEST)

  Combine both the samples. Rank all values of the combined sample from lowest to the highest.

  T= sum of the ranks insample 1

  Minitab: Stat → Nonparametrics → Mann-Whitney

  R: wilcox.test(group1, group2)

- Non-parametric TWO Dependent Samples Test:

  WILCOXON SIGNED-RANK TEST

  Test Statistics: Rank the absolute values of the differences.

  T= sum of the ranks of negative(positive) differences

  Minitab: For differences, Minitab: Stat → Nonparametrics → 1-Sample Wilcoxon

  R: wilcox.test(before, after, paired = T)
