Introduction to statistics and business analytics- BUAD 1560

8th week summary

Case 2: Two Numerical Variables – Population Standard Deviations unknown

	Independent Samples $\sigma_1 = \sigma_2$	Independent Samples $\sigma_1 \neq \sigma_2$	Paired Samples
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 = $P(T \leq t)$

- **INFERENCE ABOUT POPULATION STANDARD DEVIATION** ....
- Test for the population standard deviation (σ):

  Given the data is coming from normal distribution we can use the chi-squared test to test $H_0 : \sigma = \sigma_0$:

  Test Statistics: $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ follows $\chi^2(\nu = n-1)$

  $100(1-\alpha)\%$ CI for σ: $\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}$

  R: EnvStats::varTest(data, sigma.squared = 1)

- Test for the equality of the standard deviations ($H_0 : \sigma_1 = \sigma_2$):

  Given the data-sets are coming from normal distribution we can use the F test to test $H_0 : \sigma_1 = \sigma_2$

  Test Statistics: $F = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)}$ follows $F(\nu_1 = n_{num} - 1, \nu_2 = n_{den} - 1)$

  R: var.test(x, y, ratio = 1)