Math 4720: Statistical Methods

9^{th} Week Summary (03/22/25)

- **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: $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$ follows $\chi^2(\nu = n-1)$

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

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

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

- Test for equality of the standard deviations for more than two populations ($H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_t$):

  F test can be extended to more than two populations (Hartley's $F_{\max}$ test), but it is very sensitive to departures from normality.

  We prefer to use Brown-Forsythe-Levene(BFL) test:

  R: `lawstat::levene.test(data, group)`

- **ANOVA** ....

- ANalysis Of VAriance (ANOVA) is a popular statistical tool to analyze the effect of a categorical variable with different levels of treatments (factor) on a numerical variable (response).

- In general, we can write:

$$\text{Total Variation} = \text{Variation Between Treatments} + \text{Variation Within Treatments.}$$

- **Grand Mean:** $\bar{y}_{..} = \frac{n_1\bar{y}_{1.} + n_2\bar{y}_{2.} + \cdots + n_t\bar{y}_{t.}}{n_1 + n_2 + \cdots + n_t}$
- **Total Variability:** $\sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{..}\right)^2$
- **Variability Between Samples:** $\sum_{i=1}^{t} n_i(\bar{y}_{i.} - \bar{y}_{..})^2$
- **Variability Within Samples:** $\sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{i.}\right)^2$

	$SS(Total)$	=	SSB	+	SSE
df's:	$\sum n_i - 1$		$t-1$		$\sum n_i - t$

Treatment Levels

	1	2	3	. . .	t
	y_{11}	y_{21}	y_{31}		y_{t1}
	y_{12}	y_{22}	y_{32}		y_{t2}
	.	.	.		.
	.	.	.		.
	y_{1n_1}	y_{2n_2}	y_{3n_3}		y_{tn_t}
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$		$\bar{y}_{t.}$
St.dev.	s_1	s_2	s_3		s_t

- Hypothesis Testing

  $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$

  $H_a : \mu_i \neq \mu_j$ for some pair (i, j).

  Test Statistic: $F = \dfrac{SSB/df_B}{SSE/df_E}$

  Decision Rule: Reject H_0 in favor of H_a if $F > F_\alpha(df_B, df_E)$.