Math 4720: Statistical Methods

10^{th} and 11^{th} Week Summary (11/19/18)

- **ANOVA** ....

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

- In general we can write:

  Total Variation = Variation Between Treatments + 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: $\sum n_i - 1 \qquad t-1 \qquad \sum n_i - t$

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 pairs (i, j)

  Test Statistics: TS: $F = \dfrac{SS_B/df_B}{SS_E/df_E}$

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

Source of Variation	df	Sum of Squares	Mean Square	F	p-value
Group (Between)	$t-1$	$\sum n_i(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2 = SS_B$	$\frac{SS_B}{df_B} = MS_B$	$\frac{MS_B}{MS_E} = F_{\text{calc}}$	$\Pr(F > F_{\text{calc}})$
Error (Within)	$N-t$	$\sum(n_i-1)s_i^2 = SS_E$	$\frac{SS_E}{df_E} = MS_E$		
Total	$N-1$	$\sum(y_{ij}-\bar{y}_{\bullet\bullet})^2 = SS_T$			

- For the above ANOVA table:

  $N = \sum_i n_i$

  $SS_T = SS_B + SS_E$

  MS_E is the pooled sample variance, an estimator for σ^2

- Assumptions:

  $\sigma_1 = \sigma_2 = \cdots = \sigma_t$

  Data is generated from normal distribution for each treatment

- What if normality fails?

  We use the Non-parametric test: "The Kruskal-Wallis Test"

- What if equality of variances fails?

  We "transform" the data: "see slides for Chapter 8(B)"