Math 4720: Statistical Methods

10^{th} and 11^{th} Week Summary (04/03/25)

- **ANOVA**
- 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 $F > F_\alpha(df_B, df_E)$.

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

  R: aov(Wt sport + gender, data=ais2)

- 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:

  Homogeneity of variances: $\sigma_1 = \sigma_2 = \cdots = \sigma_t$.

  Data are generated from normal distributions for each treatment.

- What if normality fails?

  We use the non-parametric test: "Kruskal-Wallis Test"

- What if the equality of variances fails?

  We "transform" the data

- What are the common transformations?

  If $\sigma^2 \propto \mu$, the use $Y_T = \sqrt{Y}$ or $\sqrt{Y + 0.375}$

  If $\sigma^2 \propto \mu^2$, the use $Y_T = \ln(Y)$ or $\ln(Y + 1)$

  If $\sigma^2 \propto \mu(1 - \mu)$, the use $Y_T = \sin^{-1}\sqrt{Y}$

- Assumptions for Oneway ANOVA model ($y_{ij} = \mu + \tau_i + \epsilon_{ij}$, where $i = 1, \cdots, t$ and $j = 1, \cdots, n_i$):

  To test $H_0 : \tau_i = 0$ vs $H_a : \tau_i \neq 0$ for some i:

  (1) The ϵ_{ij}'s are independent and normally distributed

  (2) $\text{Var}(\epsilon_{ij}) = \sigma^2$ (a constant value)

- Checking the Assumptions:

  Obtain the residuals ($r_{ij} = y_{ij} - \hat{\mu} - \hat{\tau}_i$) and fitted values ($\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i$), then

  (1) The QQ-plot of r_{ij}'s should be linear

  (2) The scatterplot of r_{ij}'s versus $\hat{y}_{ij}$'s should follow a random pattern.