STAT 500

ANOVA - Model Diagnostics and
Non-parametric Test for One-Way ANOVA

Model Assumptions

- ANOVA F-test, contrasts and pairwise comparisons assume:

$$\epsilon_{ij} \text{ are i.i.d. } N(0, \sigma^2)$$

  - Independence of samples and observations
  - Homogeneous variance - $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_r^2 = \sigma^2$
  - Normal Distribution - Random error terms have a normal distribution

Model Diagnostics

- Many results from two-sample model diagnostics apply.
  - Independence - critical aspect
  - Equal Variances - very important
  - Normality - only a concern for small sample sizes, or very skewed distributions
  - Outliers - results not robust

- Use residuals to assess model assumptions

$$e_{ij} = Y_{ij} - \overline{Y}_{i\cdot}.$$

Not independent of each other because all e_ij depend on bar(Y_i)

Independence Assumption

- Data Collection
  - Random Sample(s) from multiple populations
  - Observations from multiple treatment groups
- Study designed to produce independent responses.

Homogeneous Variances Assumption

- Construct histograms or boxplots of residuals for each sample
- Plot residuals versus predicted values, and there should be no trend.
  - Beware of interpretation if n_i's are very unequal
  - Expect larger range of ϵ_{ij} if n_i is larger
- Study ratio of sample standard deviations

$$\frac{\max S_i}{\min S_i}$$

Homogeneous Variances Assumption

- Tests for equality of variances
  - Brown-Forsythe test
  - Levene's test
  - etc.
- Consequences of unequal variances on F test:
  - Minor if sample sizes are the same.
  - Large distortion of α level if very unequal sample sizes
  - Decreased power

If unequal variance, we can use power transformation, just like t-test.

Normality

- Histogram of **residuals**
- Normal probability plot of residuals
- Numerical summaries - skewness and kurtosis
- Tests for Normality
  - Kolmogorov-Smirnov
  - Cramer-von Mises
  - Anderson-Darling

One-Way ANOVA

- Assumptions
  - Independence Assumption
  - Homogeneous Variances Assumption
  - Normal Distribution Assumption
- What if the homogeneous variances and/or normal distribution assumptions are violated to the point where p-values and confidence levels cannot be trusted?
- Transform data and check whether the homogeneous variances and normal distribution assumptions are appropriate for transformed data.
- Non-parametric Tests

One way ANOVA is called "one way" because there is only one independent variable or factor.

Kruskal-Wallis Test
One-way ANOVA on Ranks

- Assumptions
  - Independence
- Null hypothesis: r populations have the same distribution
  - Distribution is not required to be normal.
  - Implies equal medians, percentiles, means and variances

Kruskal-Wallis Test

- Combine the data into a single data set
- Order the N observations from smallest to largest
- Assign ranks R_{ij}
  - The smallest observation gets rank=1, the second smallest gets rank=2, etc...
  - For tied observations, average the ranks

Kruskal-Wallis Test

- Calculate $\bar{R}_{i.}$ = the mean rank of observations in group i.
- The test statistic is:

$$H = (N-1)\frac{\sum_{i=1}^{r} n_i(\bar{R}_{i.} - \bar{R})^2}{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(R_{ij} - \bar{R})^2}$$

where $\overline{R} = (N+1)/2$ = the average of all ranks 1 through N

Kruskal-Wallis Test

- If H_0 is true, H will have an approximate χ^2 distribution with $r-1$ degrees of freedom.
- Approximation is best when $n_i \geq 5$ for all i.
- P-value $= P(\chi^2_{r-1} > H)$