$\underline{Model}$: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$

ANOVA Assumptions

- the t populations are normally distributed
- the t populations are independent
- the t populations have equal variance
- we have a SRS from each population (treatments randomly assigned to EUs)

A test is robust with respect to an assumption if violating the assumption does not greatly affect the properties of the test.

$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2) \rightarrow \hat{\varepsilon}_{ij} = e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \overline{Y}_{i\bullet}$

Assessing ANOVA Assumptions

- SRS: the F-test is not robust to violations of independence within groups
  - positive correlation → Standard Errors are under estimated
  - randomization helps guard against correlated observations
- Independence: the F-test is not robust to violations of independence between groups
  - violations can seriously affect Type I & Type II Errors rates
- Normality: the F-test is robust to departures from normality
  Tests:
  - Shapiro-Wilk Test
  - Anderson-Darling Test
  - Kolmogorov-Smirnov Test

  - Q–Q plot: plot observed quantiles vs. expected quantiles from a Normal Dist
- HOV: the F-test is not robust to violations (Standard Errors are inflated)
  Tests:
  - $F_{\max,s} = s_{\max}^2 / s_{\min}^2$ $df = n^* - 1$ where $n^* = \max(n_1, n_2)$
    * Quick and easy, but not robust to non-normal data
  - Brown & Forsythe's Test (a.k.a. "Levene's median")
    * Robust test of HOV
    * Equivalent to an ANOVA test on recoded data $z_{ij} = | y_{ij} - \tilde{y}_i |$
  - Residual – Predicted Plots
  - Standardized Residuals: $w_{ij} = e_{ij} / \sqrt{MSE}$
  - Studentized Residuals: $w_{ij}^* = e_{ij} / \sqrt{MSE(1 - 1/r_i)}$

Variance Stabilizing Transformations

Several scenarios lead to data where the SD is related to the mean

$$Y \sim \chi^2_k \quad \rightarrow \quad E(Y)=k \quad Var(Y)=2k$$
$$Y \sim Bin(n,p) \quad \rightarrow \quad E(Y)=np \quad Var(Y)=np(1-p)$$
$$Y \sim Poisson(\lambda) \quad \rightarrow \quad E(Y)=\lambda \quad Var(Y)=\lambda$$

Goal: Find a tranformation $W=g(Y)$ so that $Var(W)$ is constant across groups

Claim: If Y is a RV with $E(Y)=\mu$ and $SD(Y)=\sigma_y$ and $W=g(Y)$, where g is differentiable
Then $\sigma_w = |g'(\mu)|\sigma_y$

Box-Cox Power Transformation for a Completely Randomized Design (CRD)

Suppose the data, Yij , satisfy the following conditions: $E(Y)=\mu \quad and \quad \sigma_y = c\mu^m$

Let $W=Y^p$ and choose p so that σ_w does not depend on μ.

P	*Transformation*
2	y^2
1	none
½	√y
0	log(y)
-½	1/√y
-1	1/y
-2	$1/y^2$

Recall from Calculus:

Taylor Series approximation for a differentiable function $g(x)$ about $x=a$

- Approximates $g(x)$ with the polynomial $T_n(x)$
- $T_n(x) = g(a) + g'(a)(x-a) + \frac{g''(a)}{2!}(x-a)^2 + \cdots + \frac{g^{(n)}}{n!}(a)(x-a)^n$

Claim: If Y is a RV with $E(Y)=\mu$ and $SD(Y)=\sigma_y$ and $W=g(Y)$, where g is differentiable
Then $\sigma_w = |g'(\mu)|\sigma_y$

Let Y be a RV with a finite mean and SD

$$W=g(y) \overset{(T)}{=} g(\mu) + (Y-\mu)\,g'(\mu) + HigherOrderTerms$$

$$E(W) = g(\mu) + E(Y-\mu)\,g'(\mu) + E[H.O.T.]$$
$$\approx g(\mu)$$

$$Var(W) = E[W-E(W)]^2$$