Confidence Intervals:

1. C.I. width: confidence ↑, width ↑;
   sample size ↑, width ↓;
   population "noise" ↑, width ↑
2. C.I. width = (2)(Margin of Error)

Interval:

1. General: $\text{point estimator} \pm \left(\text{table value}\right)\left(\text{STD error}\right)$
2. Z-interval
   - $\bar{x} \pm Z\sigma_{\bar{x}}$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
3. t-interval
   - $\bar{x} \pm t s_{\bar{x}}$ $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

Test Statistic:

1. General: $\dfrac{\text{point estimator} - \text{null value}}{\text{STD error}}$
2. Z: $\dfrac{\bar{x} - \text{null value}}{\sigma_{\bar{x}}}$
3. t: $\dfrac{\bar{x} - \text{null value}}{s_{\bar{x}}}$

Sample Size for Desired Margin of Error:

1. $n = \left(\frac{(Z)(\sigma)}{E}\right)^2$ where E is the desired margin of error

Hypothesis Testing:

1. H_o = null hypothesis
2. H_a = alternative hypothesis
3. α = Type I Error Probability = P(Type I Error)
4. β = Type II Error Probability = P(Type II Error)
5. $Power = 1 - \beta$; Power = P[Reject H_o]
   a. sample size ↑, power ↑
   b. effect size ↑, power ↑
   c. α-level ↑, power ↑
   d. noise ↑, power ↓
   e. as n ↑, power curve becomes tighter and closer to the ideal
6. reject H_o if p-value is below α-level

	Reject H_o	*Fail to Reject H_o*
H_o True	Type I Error	Correct
H_o False	Power	Type II Error

Make a conclusion in terms of/in the context of the problem

t-Distribution (σ *unknown*):

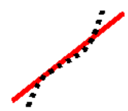
1. more spread than STD normal; symmetric at 0; as the degrees of freedom ↑, becomes more like STD Normal
2. degrees of freedom = $n - 1$

Normal Probability Plot:

Right Skew - If the plotted points appear to bend up and to the left of the normal line that indicates a long tail to the right.

Left Skew - If the plotted points bend down and to the right of the normal line that indicates a long tail to the left.

Short Tails - An S shaped-curve indicates shorter than normal tails, i.e. less variance than expected.



Long Tails - A curve which starts below the normal line, bends to follow it, and ends above it indicates long tails. That is, you are seeing more variance than you would expect in a normal distribution.

1. $x = \mu + Z\sigma$
2. Ordered data → normal score (Z)

Sign Test:

1. Nonparametric – use median in hypothesis
2. $x_i - \text{null value} = \#_i$ $n = i$
3. If $x_i = \text{null value}$ then drop that value and adjust n.
4. B = # of positive signs (test statistic)
5. Binomial (n,p=0.05) => Reference distribution
   ↳ H_o true
6. Binomial Distribution (n, p)
   a. n = # of trials; p = $P[success]$
   b. trials must be Independent of one another
   c. $p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & Otherwise \end{cases}$
   d. $E(X) = np$
   e. $V(X) = np(1-p)$

Comparing Means-Paired Data:

1. Normality: t-procedures
2. Nonparametric: Sign Test
   a. Dependent samples
   b. mean difference: $\mu_d = \mu_1 - \mu_2$
   c. sample size → n = # of pairs
   d. pt. estimator: $\bar{d}$
   e. STD error: $s_{\bar{d}} = \frac{s_e}{\sqrt{n}}$
   f. C.I.: $\bar{d} \pm t \cdot s_{\bar{d}}$
   g. T.S.: $t = \dfrac{\bar{d} - \text{null value}}{s_{\bar{d}}}$

3. Nonparametric: Signed-Rank Test
    a. Calculate the differences; delete any zeros
    b. absolute value each difference
    c. assign ranks; ties get the mean rank
    d. create sign ranks (bring back signs)
    e. total negative ranks: T^-
        total positive ranks: T^+
    - under H_o true : $T^- = T^+ = \frac{n(n+1)}{4}$

Comparing Means-Independent Data:

1. Normal: Two-sample t-inference
    a. parameter of interest: $\mu_1 - \mu_2$
    b. pt. estimate: $\bar{Y}_1 - \bar{Y}_2$
        $E(\bar{Y}_1 - \bar{Y}_2) = \mu_1 - \mu_2$
        $Var(\bar{Y}_1 - \bar{Y}_2) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$
    c. STD error: $s_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
    d. T.S.: $t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{s_{\bar{Y}_1 - \bar{Y}_2}}$
    e. C.I.: $(\bar{Y}_1 - \bar{Y}_2) \pm t \cdot s_{\bar{Y}_1 - \bar{Y}_2}$
2. Nonparametric: Sum Rank Test
    a. Arrange the data for each group in increasing order
    b. rank the observations as though they were all drawn from the same (assign tied data the average of the ranks)
    c. total the ranks for the 2 groups T_1, T_2
    d. check $T_1 + T_2 = \frac{(n_1+n_2)(n_1+n_2+1)}{2}$
    e. use T_1 as the test statistic
    f. make conclusion with table on page 1176

Correlation:

-one quantitative response-dependent (y)
-one quantitative predictor-independent/control (x)
-Association/Relationship/Co-related/Co-vary
- Scatter plot (must have correlation)

Questions of Interests:

Strength (tightness of trend)
Type (linear vs. nonlinear)
Direction (positive/negative association)

Linear Correlation Coefficient: $\frac{obs - mean}{stdev}$

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} = \frac{s_{xy}}{(n-1)\, s_x\, s_y}$$

population correlation: ρ (parameter)
sample correlation: r (statistic)

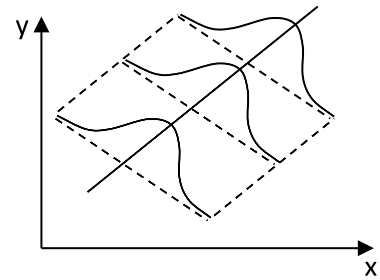
Linear Regression:

modeling relationship(s) among variables
modeling mean/location of Y as a linear function of X

$$E(X|Y) = \beta_0 + \beta_1 X$$
$$Y = \beta_0 + \beta_1 X + \varepsilon$$

3 parameters of interest:

1. intercept - β_0
2. slope - β_1
3. variance - σ^2



Sums of Squares:

$$\sum (observed - predicted)^2 = \sum (errors)^2$$
$$\sum (y - \bar{y})^2 = \sum (errors)^2$$

1. $\sum(y - \bar{y})^2 = \sum(y - \hat{y})^2 + \sum(\hat{y} - \bar{y})^2$
    $SS_{total} = SS_{error} + SS_{regression}$

d.f. - $n - 1 = \left(n - \frac{\#\,(beta)}{parameters}\right) + \#\, predictors$

2. Best Line
    a. minimizes the sum of squared errors
    b. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$
3. Estimates:
    a. $\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = r\frac{s_y}{s_x}$
    b. $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Mean Squares Error:

1. General: $MS = \frac{SS}{df}$
2. Regression: $MS_{reg} = \frac{SS_{reg}}{d.f._{reg}}$
3. Error: $MS_{err} = \sigma^2 = \frac{\sum(y-\bar{y})^2}{n - \frac{\#\,of\,(beta)}{parameters}} = \frac{SS_{err}}{d.f._{err}}$
4. $R^2 = \frac{explained}{unexplained} = \frac{SS_{regression}}{SS_{error}}$ $0 \le R^2 \le 1$
5. $SE_{\hat{\beta}_0} = \sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}} = \sqrt{MS_E} \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}}$
6. $SE_{\hat{\beta}_1} = \frac{\sigma}{\sqrt{s_{xx}}} = \frac{\sqrt{MS_E}}{\sqrt{s_{xx}}}$
    Note: to reduce $SE_{\hat{\beta}_1}$, increase s_{xx}
7. C.I. for $\hat{\beta}_0$, intercept: $\hat{\beta}_0 \pm t \cdot SE_{\hat{\beta}_0}$
    $d.f. = n - \frac{\#\,of\,(beta)}{parameters}$
8. C.I. for $\hat{\beta}_1$, slope: $\hat{\beta}_1 \pm t \cdot SE_{\hat{\beta}_1}$
    $d.f. = n - \frac{\#\,of\,(beta)}{parameters}$
9. $H_0: \beta_1 = \frac{null}{value} \rightarrow typically\ 0 \leftrightarrow no\ relationship$
    $H_a: \beta_1 <, \neq, > \frac{null}{value}$
10. T.S. $- t = \frac{estimate - \frac{null}{value}}{SE_{\hat{\beta}_1}}$

Analysis of Variance (ANOVA):

Source	*d.f.*	*SS*	*MS*	*F*	*p-value*
Regression	# predictors	SS_{reg}	MS_{reg}	$\frac{MS_{reg}}{MS_{err}}$	Table Value
Residual Error	$n - \#\ (beta)\ parameters$	SS_{err}	MS_{err}	-----	-----
Total	$n-1$	SS_{total}	-----	-----	-----

**error = residual $= y - \bar{y}$

Model Utility Test (F-test):

1. $H_0: not\ useful\ model$ $\quad y = \mu + \varepsilon$
   $H_a: model\ useful$ $\quad$ predictor coefficient $\neq 0$
2. F-distΔ (Table 8; pg 1181-1193)
   a. numdf = df_1 = df_{reg}
   b. dendf = df_2 = df_{res}
3. $SE_{\hat{Y}} = \sqrt{MS_E} \cdot \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{s_{xx}}}$
4. C.I. for the mean of $Y|X$: $\hat{Y} \pm t \cdot SE_{\hat{Y}}$
   -use df_{res}
5. P.I. for $Y|X$: $\hat{Y} \pm t \cdot \sqrt{MS_E} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{s_{xx}}}$
   -use df_{res}
6. Watch for:
   a. Large residuals
   b. non-random pattern
   c. unusual x-values/predictor values

Lack of Fit:

1. $SS_{residules} = SS_{\substack{pure\\error}} + SS_{\substack{lack\\of\ fit}}$
   d.f. - $n - \frac{\#\ of\ (beta)}{parameters} = (n-m) + \left(m - \frac{\#\ of\ (beta)}{parameters}\right)$
   $m = \#$ of distinct x values
2. $H_0: no\ lack\ of\ fit; appropriate\ model$
   $H_a: lack\ of\ fit; poor\ model$
3. T.S. $- F = \frac{MS_{LOF}}{MS_{PE}}$ $\quad df_1 = df_{LOF};\ df_2 = df_{PE}$

Multiple Linear Regression:

1. y = location + spread $= \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$
   k : predictors; k + 1 : β parameters
2. Estimate of $\sigma^2 = MS_E = \frac{SS_E}{n - \frac{\#\ of\ (beta)}{parameters}}$
3. Estimate of $\sigma = \sqrt{MS_E}$
4. $R^2 = \frac{SS_{reg}}{SS_{total}} = 1 - \frac{SS_{err}}{SS_{tot}}$
   Note: can increase R^2 by adding polynomial terms
5. $R^2_{adj} = 1 - \frac{\frac{SS_E}{df_E}}{\frac{SS_{tot}}{df_{tot}}}$
6. H_0: all predictor coefficients = 0; not useful model
   H_a: at least one predictor coefficient $\neq 0$; useful model
7. T.S. $- F = \frac{MS_{reg}}{MS_{err}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-(k+1)}}$
   $df_1 = df_{reg} = k;\ df_2 = df_{err} = n - (k+1)$
8. C.I. for β_i: $\hat{\beta}_i \pm t \cdot SE_{\hat{\beta}_i}$
   -use df_{res}
9. T.S. for β_i: $\frac{\hat{\beta}_i - \substack{null\\value}}{SE_{\hat{\beta}_i}}$
   -use df_{res}

Testing a Subset of Parameters:

1. Full Model vs. Reduced Model
2. H_0: insignificant; $\beta_i = 0$
   H_a: at least one $\beta_i \neq 0$
3. Estimation of the mean response at predictor values:
   a. C.I. : $\hat{Y} \pm t \cdot SE_{\hat{Y}}$ $\quad$ -use df_{res}
4. Prediction of the response at predictor values (wider):
   a. P.I.: $\hat{Y} \pm t \cdot \sqrt{MS_E + (SE_{\hat{Y}})^2}$ $\quad$ -use df_{res}
5. Partial F-Test: $F = \frac{SS_{reg,2} - SS_{reg,1}}{df_{reg,2} - df_{res,1}} \Big/ \frac{SS_{reg,1}}{df_{reg,1}}$
   where the model for 1 is larger than 2

Matrix Arithmetic:

1. $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$
   $\underline{Y} = n\ x\ 1; X = n\ x\ (k+1);\ \underline{\beta} = (k+1)x\ 1;$
   $\underline{\varepsilon} = n\ x\ 1$
2. Least Squares Estimation: $\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y}$
   $\underline{\hat{\beta}} = (k+1)x\ 1;\ X' = (k+1)x\ n;$
   $X = n\ x\ (k+1);\ \underline{Y} = n\ x\ 1$
   Note: each estimate of the coefficient is a linear combination of the y's
3. $X\underline{\hat{\beta}} = \underline{\hat{Y}}$
   $\underline{\hat{Y}} = nx\ 1$
4. Hat Matrix: $\lfloor X(X'X)^{-1}X' \rfloor \underline{Y} = H\underline{Y}$
5. Residual: $\underline{Y} - \underline{\hat{Y}} = [I\ \underline{Y} - H\ \underline{Y}] = [I - H]\ \underline{Y}$
6. Estimated: $Var\left(\underline{\hat{\beta}}\right) = MS_E[(X'X)^{-1}]$
   Small?
   -determinate small
   -trace small
   -maximum eigenvalue small

Things that could go wrong:

1. Wrong Model: nonlinear
2. non-constant variance
3. outliers

How to diagnose?:

1. graphical analysis
   a. why not just plot the y's (not the same mean, not the same distribution) => go to the residuals

Residuals:

1. Standardized
   a. $\frac{e_i}{\sigma\sqrt{1-h_{ii}}}$
   b. significant: ± 2
2. "Studentized"
   a. Internal – standardized (Minitab); student (SAS)
      i. $\frac{e_i}{\sqrt{MS_E}\sqrt{1-h_{ii}}}$
   b. External – deleted t residual (Minitab); rstudent (SAS)
      i. $\frac{e_i}{\sqrt{MS_{E_{(i)}}}\sqrt{1-h_{ii}}}$
      ii. $MS_{E_{(i)}}$: mean square error fitted with the i^th case deleted
   c. Plots
      i. histogram, boxplot, npp, …
      ii. scatterplots (vs): time, $\hat{y}$, x
3. Something wrong?
   a. change model: add predictors; go to non-linear model
   b. go away from least squares estimation: use a different method of model fitting
   c. if errors are normal with constant variance, consider transformation of x
   d. if errors are non-normal and/or constant variance, consider transformation of y

Potential/Leverage:

1. focus on the predictors; measure of how far the predictor values are away from the center of the x's
   a. General
      i. the h_{ii} can be interpreted as the distance the i^th case of predictors is from the center of all the predictors
         1) large values, potential/leverage to influence
         2) guidelines for "too big": $\frac{2(\#\, predictors)}{n}$

Influence:

1. Consideration of both the x (predictors) and y (response)

2. Methods of Comparison:
   a. DFBetas (SAS only)
      i. influence of the i^th case on estimating β_k
      ii. $\frac{\hat{\beta}_k-\hat{\beta}_{k(i)}}{\sqrt{MS_{E_{(i)}}\cdot c_{kk}}}$
      c_{kk} is the k^th diagonal element of $(X'X)^{-1}$
      iii. Guidelines:
         1) > 1, further investigation
         2) $> \frac{2}{\sqrt{n}}$, large data set; $n \geq 100$
   b. DFFits
      i. measure the influence the i^th obs has on the fitted value $\hat{y}_i$
      ii. $DFFits_i = \frac{\hat{y}_k-\hat{y}_{k(i)}}{\sqrt{MS_{E_{(i)}}\cdot h_{ii}}}$
      iii. Guidelines:
         1) $|DFFits| > 1$, small/medium data sets
         2) $|DFFits| > 2\sqrt{\frac{\#\, parameters}{n}}$, $n \geq 100$
   c. Cooks Distance
      i. consider the influence of the i^th case on all fitted values "aggregate influence measure"
      ii. $p =$ # parameters

$$D_i = \frac{\sum_{j=1}^{n}(\hat{y}_j - \hat{y}_{j(i)})}{p \cdot MS_E}$$

      iii. Guidelines:
         1) $D_i > 1$, unreasonable
   d. Cov Ratio
      i. $\frac{\left|MS_{E_{(i)}}\left(X_{(i)}^{i}X_{(i)}\right)^{-1}\right|}{|MS_E(X'X)^{-1}|}$
      ii. Guidelines:
         1) $|CovRatio - 1| \geq \frac{3p}{n}$

Model Building:

1. plot y vs. each X (matrix scatterplot, correlation matrix)
2. multicollinearity: relationship amongst predictors (independent variables)
   a. effects
      i. insignificant t-tests when you expect significance
      ii. signs for the betas opposite of what you expect
      iii. error significant in the computation; numerical analysis
   b. $SE_{\hat{\beta}_i} = \sqrt{MS_E\left(\frac{1}{1-R_i^2}\right)}$

      ↳ variance of inflation

      Guidelines: $VIF > 10$

      where R_i^2 is the R^2 when x_i is regressed on the other predictors
3. Compare Criteria
   a. R^2: high
   b. MS_E: low
   c. $R^2 - adj$: high
   d. $PRESS = \sum_{i=1}^{n}(y_i - \hat{y}_{(i)})^2$: low

      ↳ predicted residuals of SS
   e. Mallow's $C_p \approx p$
      i. $C_p = \frac{SS_{E_p}}{MS_E} - (n-2p)$
4. Model Selection:
   a. Backward Selection
      i. start big-remove predictors: fit the model with all predictors; pick out biggest p-value; refit model => continue until all predictors are significant
   b. Forward Selection
      i. start small-add predictors: consider all SLR models; add the predictor with the smallest p-value => continue until no longer significant
   c. Stepwise Selection
      i. combination of backward and forward selection: check overall model to make sure that all predictors are still useful; if not, it is removed from the model

One-Way ANOVA:

$y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \tau_i + \varepsilon_{ij}$

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

$H_0: \tau_1 = \tau_2 = \cdots = \tau_k = 0$

1. compare the means of three or more groups
2. Hypotheses:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_a: means\ are\ not\ all\ equal$$

3. Requirements:
   a. the k populations all have normal distributions
   b. the k populations all have the same standard deviations $\sigma_1 = \sigma_2 = \cdots = \sigma_k$
4. Notation:
   a. y_{ij} = observation for the j[th] unit in the i[th] group
   b. n_i = number of units in the i[th] group
   c. n = total number of units in the study
   d. $\bar{y}_{i\cdot}$ = sample mean for the i[th] group
   e. $\bar{y}_{\cdot\cdot}$ = sample mean for the entire set of data (grand mean)
   f. k = number of groups
5. Assessment of the within-group variation [error/noise]
   a. $SS_{within} = (n_1-1)(s_1^2) + (n_2-1)(s_2^2) + \cdots + (n_k-1)(s_k^2)$
   b. $df_{within} = n - k$
   c. $MS_{within} = \frac{SS_{within}}{df_{within}} = \frac{(n_1-1)(s_1^2)+(n_2-1)(s_2^2)+\cdots+(n_k-1)(s_k^2)}{n_1+n_2+\cdots n_k-k}$
6. Assessment of the between-group variation [treatment]
   a. $SS_{between} = \sum n_i\,(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$

      $SS_{between} = n\sum(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$, if all groups have the same number of units
   b. $df_{between} = k - 1$
   c. $MS_{between} = \frac{SS_{between}}{df_{between}}$
7. $\bar{y}_{\cdot\cdot} = \frac{n_1\bar{y}_{1\cdot}+n_2\bar{y}_{2\cdot}+\cdots n_k\bar{y}_{k\cdot}}{n_1+n_2+\cdots+n_k}$
8. F-test: $F = \frac{MS_{between}}{MS_{within}}$
9. $SS_{total} = \sum\sum(y_{ij} - \bar{y}_{\cdot\cdot})^2$
10. $s = \sqrt{\frac{SS_{total}}{n-1}}$
11. $R^2 = \frac{SS_{between}}{SS_{total}}$

Multiple Comparisons:

α_I = individual type I error rate

α_E = experiment wise type I error rate

1. Fisher's LSD (liberal)
   a. $t_{ij} = \frac{(\bar{y}_{i\cdot} - \bar{y}_{j\cdot})}{SE_{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}}$
      where $SE_{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}} = \sqrt{MS_{within}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$
   b. $df = df_{within}$
   c. CI: $\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm t \cdot SE_{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}$
2. Kruskal-Wallis
   a. assign ranks to all the data, averaging for repeated values
   b. determine the mean rank within each group $\bar{R}_i$
   c. If H_0 true, the $\bar{R}_i$ "equal"
   d. H test stat: $\frac{12}{n(n+1)} \sum n_i(\bar{R}_i - \bar{R})^2$
      reject large H (pg 428)
   e. total ranks = $\frac{n(n+1)}{2}$
   f. mean rank = $\frac{n+1}{2}$
3. Bonferroni Correction (conservative)
   a. want $\alpha_E = \alpha$ => choose $\alpha_I = \frac{\alpha}{m}$ for m independent tests
   b. LSD - $t \cdot SE_{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}$
      find t with $\frac{\alpha_I}{2}$ in each tail (t-tables)
   c. CI: $\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm t \cdot SE_{\bar{y}_{i\cdot} - \bar{y}_{j\cdot}}$
4. Tukey's W Procedure (HSD) (mod-conservative)
   a. Table 10
   b. common sample size: $w = q_\alpha(\#trt, df_{err})\sqrt{\frac{MS_E}{n}}$
   c. uncommon sample size: $w = \frac{q_{\alpha_E}(\#trt, df_{err})}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$
   d. $|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| \geq w$
   e. CI: $\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm w$
5. SNK (mod-liberal)
   a. $SNK = q_{\alpha_E}(\#\,means, df_{err})\sqrt{\frac{MS_E}{n}}$
   b. $|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| \geq SNK$
6. Dunnett's Compare to Control
   a. Are treatments different than the control?
   b. two sided test: $|\bar{y}_{i\cdot} - \bar{y}_{cont}| \geq d_{\alpha_E}\left(\frac{\#\,non}{control}, df_{err}\right)\sqrt{\frac{2MS_E}{n}}$

Blocking (Randomized Complete Block Design):

$y_{ij} = \mu_{ij} + \varepsilon_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$

$\sum \tau_t = 0 \qquad \sum \beta_b = 0$

total # obs = bt (with no repeats)

$SS_{total} = SS_{block} + SS_{trt}$

RCBD ANOVA Table:

Source	*d.f.*	*SS*	*MS*	*F*	*p-value*
Treatment	t-1	SS_{trt}	MS_{trt}	$\frac{MS_{trt}}{MS_{err}}$	Table Value
Block	b-1	SS_{blk}	MS_{blk}	$\frac{MS_{blk}}{MS_{err}}$	Table Value
Error	(b-1)(t-1)	SS_{err}	MS_{err}	-----	-----
Total	bt-1	SS_{tot}	-----	-----	-----

1. Treatment Effective?
   $H_0: \tau_1 = \tau_2 = \cdots = \tau_k = 0$
   $H_a: not\ all\ \tau = 0$
2. Blocking Effective?
   $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$
   $H_a: not\ all\ \beta = 0$
3. Comparison Tests:
   a. Tukey's Pairwise comparisons for RCBD
      i. $|\bar{y}_{i\cdot} - \bar{y}_{j*}| \geq q_\alpha(\#trt, df_{err})\sqrt{\frac{MS_E}{b}}$
   b. Bonferroni
      i. $|\bar{y}_{i\cdot} - \bar{y}_{j*}| \geq t\sqrt{MS_E\left(\frac{1}{b} + \frac{1}{b}\right)}$

Factorial Structure:

Factor A – a levels

Factor B – b levels

Balanced Design:

ab = # treatments

abn = # observations (equal replication)

$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$

ME ME INT

$SS_{tot} = SS_E + SS_A + SS_B + SS_{AB}$

ANOVA:

Source	*d.f.*	*SS*	*MS*	*F*	*p-value*
A	a-1	SS_A	MS_A	$\frac{MS_A}{MS_{err}}$	Table Value
B	b-1	SS_B	MS_B	$\frac{MS_B}{MS_{err}}$	Table Value
AB	(a-1)(b-1)	SS_{AB}	MS_{AB}	$\frac{MS_{AB}}{MS_{err}}$	Table Value
Error	ab(n-1)	SS_{err}	MS_{err}	-----	-----
Total	abn-1	SS_{total}	-----	-----	-----

1. Test: $H_0: \alpha\beta_{ij} = 0$ → no interaction
   a. No: Main Effects → no main effects
      i. $H_0: \alpha_i = 0$
      ii. $H_0: \beta_j = 0$
   b. Yes: M/C between trt means
2. Simple Effects: comparison in which all but one factor is fixed at a certain level
   a. # of simple effects = $\binom{a}{2}b + \binom{b}{2}a$
   b. use M/C defined in Blocking