Jennifer Starling
jennstarling@gmail.com

DISTRIBUTIONAL RELATIONSHIPS

1. Linear combo of Normals. $X_i \sim N(\mu_i, \sigma_i^2)$. $\rightarrow$ $Y = b + a_1X_1 + \ldots + a_nX_n$, $Y \sim N(b + \Sigma a_i\mu_i, \Sigma a_i^2\sigma_i^2)$

2. Dist of Normal Sample Mean: $X_i \sim N(\mu, \sigma_x^2) \rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$ (X's frm same norm dist)

3. Chi-Squared: a) If $X_1 \ldots X_n \sim \chi^2$ w/ df $n_1 \ldots n_k \rightarrow X_1 + \ldots + X_n \sim \chi^2$ w/ df $= n_1 + \ldots + n_k$
   b) If $Z_1 \ldots Z_n \sim N(0,1) \rightarrow Z_1^2 + \ldots + Z_n^2 \sim \chi^2$ w/ df$=n$

4. T-dist: If $Z \sim N(0,1)$ & $U \sim \chi_n^2 \rightarrow T = Z/\sqrt{U/n}$ with df$=n$.

5. F-dist: If $U \sim \chi_m^2$ & $V \sim \chi_n^2 \rightarrow W = \frac{U/m}{V/n} = \frac{nU}{Vm} \sim F$ with df1$=m$, df2$=n$.

6. Cauchy: If X & $Y \sim N(0,1) \rightarrow X/Y \sim$ Cauchy, loc param$=0$, scale$=1$.

7. Difference of Two Normal Sample Means: $X_1 \ldots X_{20}$ = random sample frm $N(20,5)$. X & Y indep.
   $Y_1 \ldots Y_{25}$ = random sample frm $N(24,4)$. Then $W = \bar{X} - \bar{Y} \sim N(\mu_x - \mu_y, \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}) = N(-4, \frac{5}{20} + \frac{4}{25})$.

8. To standardize a normal distribution: $Z = (X-\mu)/\sigma$. (So standardize $\bar{X}$ via $(\bar{X}-\mu)/\sqrt{\frac{\sigma^2}{n}}$.)

9. Diff of Two Normal Non-Indep Sample Means: (extension of 7): $\bar{X} - \bar{Y} \sim N(\mu_x - \mu_y, \frac{\sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y\rho}{n})$

1a: If $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, with X & Y indep $X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

REVIEW of SAMPLING DISTRIBUTIONS

- Sampling Distribution - Dist of an estimator for a pop. parameter ($\bar{X}$, S, $\hat{\Theta}$, $\hat{\$}$, etc.)

- How good is an estimator?
  1. Bias: $E(\hat{\theta}) - \theta$, On average does $\hat{\theta} = \hat{\theta}$?
  2. MSE($\hat{\theta}$): Mean Squared Error. $E[(\theta - \hat{\theta})^2] = Var(\hat{\theta}) + Bias(\hat{\theta})^2$.
     Avg squared distance frm θ to $\hat{\theta}$. Want to minimize MSE.

- Standard Error: Std. dev. of the estimator, ie how spread out are possible values around sampling mean of estimator.

1) Sampling Dist of Sample Mean $\bar{X}$:
   1. If $F \sim N(\mu, \sigma^2)$ then $\bar{X} \sim N(\mu, \sigma^2/n)$.
   2. For large n, $\bar{X} \sim$ approx $N(\mu, \sigma^2/n)$. (B/c CLT: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$ for large n)
      ↳ If data corr, $\rho > 0 \rightarrow$ we are underestimating σ^2/n.

- Standard Error of $\bar{X}$: $SE(\bar{X}) = S/\sqrt{n}$. If corr, use $\frac{S}{\sqrt{n}}\sqrt{\frac{1+\hat{\rho}}{1-\hat{\rho}}}$.
  Can be viewed as std dev in error rel to true mean, b/c $\bar{X}$ is an unbiased estimator of μ.

2) Sampling Dist of Sample Std. Dev S: For large n, $S \sim$ approx $N(\sigma, \frac{\mu_4 - \sigma^4}{4n\sigma^2})$, $\mu_4 = 4^{th}$ cent. mom.

3) Sampling Dist of Sample Proportion $\hat{p}$: $\hat{p}$ has mean p, sd $\sqrt{\frac{p(1-p)}{n}}$
   1. For sampling with replacement, $n\hat{p} \sim$ Binomial
   2. For sampling w/o replacement, $n\hat{p} \sim$ Hypergeometric
   3. For large n, $\hat{p} \sim N(p, \frac{p(1-p)}{n})$. Only use if $\min[np, n(1-p)] \geq 5$.

4) Sampling Dist of MLE $\hat{\theta}$:
   1. Invariance Property of MLE's $\rightarrow$ MLE of $h(\theta)$ is $h(\hat{\theta})$.
   2. Asymptotic Property of MLE's $\rightarrow$ $h(\hat{\theta}) \sim N(h(\hat{\theta}), \frac{(h'(\hat{\theta}))^2}{I_n(\theta)})$
   3. For small n, use parametric bootstrap.

Hypothesis Testing Principles

· Reject H_0 (null hypothesis) when pval < chosen nominal α val (usually .01 or .05).
· Type I Error: "false alarm", rejecting H_0 when H_0 true. (Type A = 1 personality)
· Type II Error: "missed detection", fail to reject H_0 when H_0 false. (Type B = 2 personality)
· Size of a Test (α): chosen limit on P[Type I Error].
· Power Fctn ($\gamma(\theta)$): · If θ in null (H_0) space, gives P(Type I error).
· If θ in H_1 space, gives P(correct decision). Then $1-\gamma(\theta)$ = P(Type 2 error).
· P-val: Smallest val of α which would lead to rejecting H_0, ie prob of getting a test stat at least as extreme as the one observed.

* Always provide a Conf. Interval for parameter(s), to go with the test.

· Need to adjust chosen α if doing multiple tests at once.
  · If multiple indep events, ie orthog. contrasts, use $\alpha_{PC} = 1-(1-\alpha_0)^{1/M}$, M = # contrasts.
  · If not orthogonal, use Bonferroni adjustment, $\alpha_{PC} = \alpha_0/M$.
  · If only testing some level combos of trts, Tukey-adjusted pvals too conservative (ie if testing pairwise compares of Trt's B levels @ ea. level of A. Then use α_0/M, where M = (# B levels choose 2) · # A levels.

· Factors Influencing Power:
1) How big is diff btwn estimated mean(s) + true pop mean(s)? Bigger diff = easier to detect.
2) How much data do you have? More data = ↑ chance of detecting diffs.
3) How much variation is there? If ↑ var + small n, hard to detect diffs. Small var → easy to detect.

Confidence Intervals

1. CI for normal mean μ_i: $\hat{\mu}_i \pm t_{\alpha/2, n-t} \hat{SE}(\hat{\mu}_i)$ where $\hat{SE}(\hat{\mu}_i) = \hat{\sigma}_e / \sqrt{n_i}$
   - If only 1 trt w/ mean μ, just use $t_{\alpha/2}$ & $\hat{SE}(\hat{\mu}) = \hat{\sigma}/\sqrt{n}$.
   - Pivot $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim t$ dist w/ df = n-1 (for μ) or df = n-t (for μ_i).

2. CI for Diff in 2 Normal means ($\mu_i - \mu_k$): $(\hat{\mu}_i - \hat{\mu}_k) \pm t_{\alpha/2, n-t} \hat{SE}(\hat{\mu}_i - \hat{\mu}_k)$ where $\hat{SE}(\hat{\mu}_i - \hat{\mu}_k) = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}$.
   - If equal vars, $\hat{SE}(\hat{\mu}_i - \hat{\mu}_k)$ simplifies to $\hat{\sigma}_e \sqrt{1/n_i + 1/n_k}$.
   - If unequal vars, the pooled SE works b/c of Welch-Satterthwaite.

3. CI for Normal St. Dev $\hat{\sigma}_e$: $\left[\sqrt{\frac{SSE}{\chi^2_{\alpha/2, n-t}}}, \sqrt{\frac{SSE}{\chi^2_{1-\alpha/2, n-t}}}\right]$ where $SSE = (n-t)\hat{\sigma}_e^2$.
   - If only 1 trt, use $\hat{\sigma}$ instead of $\hat{\sigma}_e$ and (n-1) instead of (n-t).
   - To get CI for $\hat{\sigma}_e^2$, just square upper & lower bounds for $\hat{\sigma}_e$ CI above.
     - ↳ Note: Point estimator $\hat{\sigma}_e^2 = MSE$ is unbiased. But $\hat{\sigma}_e$ is biased.
   - This all works b/c $X_1 \ldots X_n \sim N(\mu, \sigma^2) \rightarrow \frac{(n-1)}{\sigma^2}\hat{\sigma}^2 \underset{(or)}{=} \frac{SSE}{\sigma_e^2} \sim \chi^2_{n-1}$ or χ^2_{n-t}

4. CI for Var. Component σ_A^2 for Random Effects: $\left[\frac{1}{rC_U} SS_{TRT}\left(1 - \frac{F_U}{F_0}\right), \frac{1}{rC_L} SS_{TRT}\left(1 - \frac{F_L}{F_0}\right)\right]$
   - Is a conservative (approx) CI, true coverage prob is higher. SAS gives this in PROC MIXED.
   - Requires equal sample sizes, $n_1 = \ldots = n_t = r$.
   - $F_0 = MS_{TRT}/MSE$. $F_U = F_{\alpha/4, t-1, n-t}$. $F_L = F_{1-\alpha/4, t-1, n-t}$.
   - $C_U = \chi^2_{\alpha/4, t-1}$. $C_L = \chi^2_{1-\alpha/4, t-1}$.

5. CI for β in Exponential Dist: $\left[\frac{2n\bar{Y}}{\chi^2_{1-\alpha/2, 2n}}, \frac{2n\bar{Y}}{\chi^2_{\alpha/2, 2n}}\right]$
   - $\hat{\beta} = \bar{Y}$ is MLE for β.

6. CI for β_1/β_2 to Compare Exponential Dists: $\left[\frac{\bar{X}}{\bar{Y}} F_{\alpha/2, n_1, n_2}, \frac{\bar{X}}{\bar{Y}} F_{1-\alpha/2, n_1, n_2}\right]$
   - Works b/c quotient of two chi-sq dists is F.

7. Dist-Free CI for Quantile Q(u), for non-Normal data: $[Y_{(r)}, Y_{(s)}]$
   - $Y_{(r)}$ & $Y_{(s)}$ order statistics → need R code to obtain r & s vals.
   - r & s so that $P[Y_{(r)} \le Q(u) \le Y_{(s)}] = 1 - \alpha$

8. Dist-Free CI for Median Q(.5) (or $\tilde{\mu}$): $[Y_{(r)}, Y_{(n-r+1)}]$
   - $Y_{(r)}$ & $Y_{(n-r+1)}$ order statistics.
   - Use "Table VII.3 - Conf. Intervals for Medians" frm CRC Handbook of Tables.
     - ↳ In table, k=r.

Confidence Intervals CTD.

9. CI for Pop. Proportion p (Wald): $\hat{p} \pm Z_{\alpha/2} \hat{SE}(\hat{p})$ where $\hat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
   - Not great for small n, Wilson's is better. Don't use if $Y = n\hat{p} < 5$.

10. CI for Diff in 2 Pop Proportions $(p_1 - p_2)$: $(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

12. CI for Pop. Proportion p (Wilson): $\tilde{p} \pm \left(\sqrt{n}/\tilde{n}\right) Z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p}) + \frac{1}{4n} Z^2_{\alpha/2}}$
   - $\tilde{Y} = Y + .5 Z^2_{\alpha/2}$ · $\tilde{n} = n + Z^2_{\alpha/2}$ · $\tilde{p} = \tilde{Y}/\tilde{n}$
   - If $n \geq 40$, Wald + Wilson not very close, use Wilson

13. CI for Pop. Proportion p (Agresti-Coull): $\tilde{p} \pm Z_{\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$
   - If $n \geq 40$, Wilson + AC close, AC is always a bit wider. AC is easier to compute.

* Rules for Choosing Which CI to Use:

1. The 'best' interval is the one with highest coverage prob.
2. If two CIs have same coverage prob, CI with shortest expected width is better.

* CI's give $(100-\alpha)\%$ certainty that an estimated parameter is in the interval. But are two other types of intervals: prediction + tolerance intervals.

PREDICTION INTERVALS

* Forecasting the value of the next realization of an RV, ie 95% certain that Y_{n+1} will be in the PI.

1. PI for Normal Pop. Dist: $\bar{Y} \pm t_{\alpha/2}\,\hat{\sigma}\sqrt{1+1/n}$ → (if params known, $\mu \pm Z_{\alpha/2}\sigma$)
   - For $Y_1 \dots Y_n \sim N(\mu, \sigma^2)$, actual next measured value is $Y_{n+1} = \mu + \sigma Z^{\alpha/2}$.
   - Point estimator for $Y_{n+1} = \bar{Y}$
   - Very similar to CI for μ, but add 1 in the sqrt → adds width b/c of uncertainty in predicting the next realization.

2. PI for Exponential (β) Dist: $\left[\bar{Y}F_{1-\alpha/2,2,2n},\ \bar{Y}F_{\alpha/2,2,2n}\right]$
   - $\bar{Y}$ is MLE for β

3. Dist-Free way to get PI for non-Normal data: Use Box-Cox to transform data ($X=g(Y)$).
   - Check if $X=g(Y)$ is increasing or decreasing fctn.
   - If g increasing: $\left[g^{-1}\left(\bar{X} - t_{\alpha/2}\,\hat{\sigma}_X\sqrt{1+1/n}\right),\ g^{-1}\left(\bar{X} + t_{\alpha/2}\,\hat{\sigma}_X\sqrt{1+1/n}\right)\right] = \left[g^{-1}(L),\ g^{-1}(U)\right]$
   - If g decreasing: $\left[g^{-1}(U),\ g^{-1}(L)\right]$

TOLERANCE INTERVALS

* Estimating a region in the pop. that contains 100P% of pop. values.

1. Engineering TI: (S_L, S_U) A specific set of limits, defining acceptable range of values.
   - Only stats question: what proportion of a process's output is acceptable?
   - If pdf f known: $P[\text{Acceptable}] = \int_{S_L}^{S_U} f(y)\,dy$
   - If pdf f unknown: Let $\hat{p}$ = proportion of acceptable units in sample data. Then form a CI on $\hat{p}$, such as Agresti-Coull.

2. Natural TI (no unknown dist or params): Set up $[T_L, T_U]$ containing 100P% of vals.
   - $T_L = Q(u_1)$, $T_U = Q(u_2)$ where $u_1 + 1 - u_2 = \alpha$.
   - For $N(\mu, \sigma^2)$ with μ & σ^2 known, is $\mu \pm \sigma Z_{\alpha/2}$.

3. Statistical TI (unknown dist or params): "100γ% conf that 100P% of vals in interval."
   a. 2-sided TI for $N(\mu, \sigma^2)$: $\hat{\mu} \pm \hat{\sigma} K_{P,\gamma}$
      - $K_{P,\gamma}$ from Table "Factors for Determining 2-Sided Tol Limits", top half.

   b. 1-sided TI for $N(\mu, \sigma^2)$: $\hat{\mu} \pm K^*_{P,\gamma}$ (− for lower, + for upper)
      - $K^*_{P,\gamma}$ from same table, bottom ½ of page.

   c. Lower TBound for $\text{Exp}(\beta)$:
      - If β known, $W_P = -\beta \log(P)$ where P = % of values wanted above bound.
      - If β unknown, $W_{P,\gamma} = -\hat{\beta}\left[\frac{2n}{\chi^2_{1-\gamma,2n}}\right]\log(P)$

   d. Dist-Free TI: $[Y_{(r)}, Y_{(n-s+1)}]$
      - $Y_{(1)} < \ldots < Y_{(n)}$ are order statistics
      - To find r & s, use m=r+s with Somerville's 'Annals of Mathematical Statistics' Tolerance Interval tables, top ½ of page.
      - Can use bottom ½ of pg to determine sample size. (Find n so are 100γ% confident that 100P% of pop. is btwn $Y_{(1)}$ & $Y_{(n)}$.)

   e. Dist-Free Way to get TI for non-Normal data: Use Box-Cox transform, $X = g(Y)$.
      - Check if $X = g(Y)$ is increasing or decreasing fctn.
      - If g increasing: $[g^{-1}(L_{P,\gamma}), g^{-1}(U_{P,\gamma})]$.
      - If g decreasing: $[g^{-1}(U_{P,\gamma}), g^{-1}(L_{P,\gamma})]$.

642 Process Summary:

Given description of an experiment:

1. Identify 3 components of experimental design.
2. Write a model relating response to factor effects (cell means + effects models)
   - Include descriptions of terms + constraints. Be able to express model in matrix form.
3. Provide AOV table, including df + Expected Mean Squares (EMS) for all Sources of Var (SV).
4. Provide approx. F-statistic, using Satterthwaite Approx if needed.
5. Estimate all variance components for random effects.
6. Provide estimated SE's for diffs in trt means.
7. Compute values of Tukey-Kramer HSD statistic, + Dunnett's.
8. Provide contrast to test specific linear combos of trt means.
9. Assess model assumptions + specify alt. approaches if violated.
10. Determine sample sizes, power curves + relative efficiency.

- If factorial is missing trts, specify contrasts in obs. trt. means to express main + interaction effects. Provide test statistics + state if contrasts are estimable.

- If experiment has mix of fixed/random effects, crossed/nested factor effects, determine appropriate analysis (mult. comparisons, contrasts, var components). Give formulas for HSD + Dunnett, and F-test for contrasts.

- Given generators for a fractional factorial design,
  - Specify selected trts.
  - Determine alias set.
  - Determine design resolution
  - Determine important effects (main and/or interactions) when $df_{ERROR}=0$.

- If >1 measurement per EU, specify measurement process:
  - Subsampling
  - Repeat Measures, longitudinal or spatial
  - Crossover

- Describe 4 types of SS + when each is appropriate

3 Components of Experimental Designs

(1) Randomization: How EU's are assigned to trts (or EU's selected frm trt pops).

- CRD - One randomization, all EU's homogeneous. t = # trts.
  - Assign 1...n to EUs, randomly permute. Assign 1st n, EU's to trt 1, and so forth.
- RCBD - Non-homog EU's, split into blocks of EU's alike wrt chars affecting response to trt.
  - A mini-CRD inside each block. r = # blocks, mt responses/block, T=trts, m=reps.
- BIBD - Like RCBD, but not all trts in all blocks.
  - t = trts, b = blocks, r = # blocks ea trt is in, k = EU's/block, λ = # blocks in which EU pair is together.
  - Conditions: 1) $n = tr = bk$ 2) $\lambda = \frac{r(k-1)}{t-1}$, integer 3) $\lambda < r < b$ 4) $b \geq t$
  - want to maximize k.
- LSD - Two blocking factors, row + col. t unstructured trts, t^2 EU's.
- Split-Plot - whole plot factor randomized over larger WP EU's. Split-Plot factor randomized on smaller SP EU's.
  - 2 sizes of EU.
  - Can have Split-Split Plot, w/ 3 sizes of EU w/ a factor randomized over ea. size EU.
  - Can have a block as highest level first, then diff-size EU's in ea. block.
    - ↳ RCBD Split-Plot or RCBD Split-Split Plot.
- Crossover Design - Multiple measurements on each EU. KEY - each EU gets each trt.
  - Diff frm subsampling + rep. measures b/c in those, ea. EU only gets 1 trt.

(2) Treatment Structure: How trts constructed - set of factors to be studied.

- Single factor: One factor w/ multiple levels. May be fixed or random.
- Crossed: 2+ factors w/ all levels combined (ie A=2 levels, B=3 levels → 2×3=6 trts). "Factorial"
- Fractional Factorial: All levels of 2+ trts combined, only a subset is selected for experiment.
- Nested: Values for levels of B depend on which level of A selected → B nested in A. "B(A)"

(3) Measurement Structure:

- Single measurement on each EU.
- Subsampling: Multiple meas. on same EU, at randomly selected locations on EU.
- Repeated Measures: Multiple meas. on same EU.
  - ↳ Longitudinally / temporally - multiple meas. @ specified points in time. (Same times for all EU's)
  - ↳ Spatially - multiple meas. at specified locs. (Same locs for all EU's.)

* greek letters = fixed. roman alphabet = random.

MODELS FOR EXPERIMENTS:

1. **CRD - Single fixed factor:**

   Cell Means: $y_{ij} = \mu_i + e_{ij}$

   Effects: $y_{ij} = \mu + \tau_i + e_{ij}$

   Matrix: $Y = X\beta + e$

   - μ = grand mean
   - τ_i = F trt effect, $i = 1 \ldots t$. $\tau_t = 0$.
   - e_{ij} = residuals, R. $j = 1 \ldots n_i$. $\sim$ iid $N(0, \sigma_e^2)$.

   · Variance components: none.

2. **CRD - Single random factor:**

   Cell Means: $y_{ij} = M_i + e_{ij}$

   (Conditional on M_i. $y_i | A \sim N(M_i, \sigma_e^2)$)

   Effects: $y_{ij} = \mu + A_i + e_{ij}$

   - μ = grand mean
   - A_i = R trt effect, $i = 1 \ldots t$. $\sim$ iid $N(0, \sigma_A^2)$.] indep
   - e_{ij} = residuals, R. $j = 1 \ldots n_i$. $\sim$ iid $N(0, \sigma_e^2)$.

   · Variance components: $\sigma_Y^2 = \sigma_A^2 + \sigma_e^2$

3. **CRD - Single fixed factor + subsampling:**

   $y_{ijk} = \mu + \tau_i + e_{ij} + d_{ijk}$

   - μ = grand mean
   - τ_i = F trt effect, $i = 1 \ldots t$. $\tau_t = 0$.
   - e_{ij} = R effect of EU, $j = 1 \ldots n_i$. $\sim$ iid $N(0, \sigma_e^2)$] indep
   - d_{ijk} = R effect of subsample, $k = 1 \ldots m_{ij}$. $\sim N(0, \sigma_d^2)$

   · Variance Components: $\sigma_Y^2 = \sigma_e^2 + \sigma_d^2$

4. **CRD - Single random factor + subsampling:**

   $y_{ijk} = \mu + A_i + e_{ij} + d_{ijk}$

   - μ = grand mean
   - A_i = R trt effect, $i = 1 \ldots t$. $\sim$ iid $N(0, \sigma_A^2)$
   - e_{ij} = R effect of EU, $j = 1 \ldots n_i$. $\sim$ iid $N(0, \sigma_e^2)$] indep
   - d_{ijk} = R effect of subsample, $k = 1 \ldots m_{ij}$. $\sim$ iid $N(0, \sigma_d^2)$

   · Var components: $\sigma_Y^2 = \sigma_A^2 + \sigma_e^2 + \sigma_d^2$

5. **CRD - 2 Fixed factors:** (A * B)

   Cell Means: $y_{ijk} = \mu_{ij} + e_{ijk}$

   Effects: $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$

   - μ = grand mean
   - τ_i = F effect of A. $i = 1 \ldots a$. $\tau_a = 0$
   - β_j = F effect of B. $j = 1 \ldots b$. $\beta_b = 0$.
   - $(\tau\beta)_{ij}$ = A * B int. effect. $(\tau\beta)_{ab} = 0$.
   - e_{ijk} = residuals, $\sim$ iid $N(0, \sigma_e^2)$.

   · Var Components: none

6. **CRD - 3 fixed factors:** (A, B, C)

Cell Means: $Y_{ijkl} = \mu_{ijk} + e_{ijkl}$

Effects: $Y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + e_{ijkl}$

- Var Components: none

- μ = grand mean
- τ_i = F effect of A, $i=1\ldots a$. $\tau_a = 0$.
- β_j = F effect of B, $j=1\ldots b$. $\beta_b = 0$.
- γ_k = F effect of C, $k=1\ldots c$. $\gamma_c = 0$.
- $(\tau\beta)_{ij}$ = A*B int effect, $(\tau\beta)_{ab} = 0$.
- $(\tau\gamma)_{ik}$ = A*C int effect, $(\tau\gamma)_{ac} = 0$.
- $(\beta\gamma)_{jk}$ = B*C int effect, $(\beta\gamma)_{bc} = 0$.
- $(\tau\beta\gamma)_{ijk}$ = A*B*C int effect, $(\tau\beta\gamma)_{abc} = 0$.
- e_{ijkl} = residuals, $l = 1\ldots$ # reps. $\sim$ iid $N(0, \sigma^2_e)$

7. **Augmented CRD:** 2 factors, A quant, B qual. A has a "0" control level $\rightarrow A_1B_1 = A_1B_2 = \ldots A_1B_b$.

- Only $t = (a-1)b+1$ unique trts.
- Use 3 separate models:
  1. Model for SV=A: $Y_{ij} = \mu + \tau_i + e_{ij}$ (Use all data)
  2. Model for SV = B + A*B: $Y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + e_{ijk}$ (Exclude control data).
  3. Model for Error: $Y_{ij} = \mu + \alpha_i + e_{ij}$ (Use all data, ignore trt structure - analyze as CRD.)

8. **CRD - 2 mixed factors:** (A fixed, B random)

$Y_{ijk} = \mu + \tau_i + b_j + (\tau b)_{ij} + e_{ijk}$

- Var Components: $\sigma^2_Y = \sigma^2_B + \sigma^2_{A\cdot B} + \sigma^2_e$

- μ = grand mean
- τ_i = F effect of A, $i=1\ldots a$. $\tau_a = 0$.
- b_j = R effect of B, $j=1\ldots b$. $\sim$ iid $N(0, \sigma^2_B)$
- $(\tau b)_{ij}$ = R effect of A*B. $\sim$ iid $N(0, \sigma^2_{A\cdot B})$
- e_{ijk} = residuals, R. $\sim$ iid $N(0, \sigma^2_e)$

$\}$ indep

9. **CRD - 3 Mixed factors:** (A f, B f, C r)

$Y_{ijkl} = \mu + \tau_i + b_j + c_k + (\tau b)_{ij} + (\tau c)_{ik} + (bc)_{jk} + (\tau bc)_{ijk} + e_{ijkl}$

- Var Components: $\sigma^2_Y = \sigma^2_B + \sigma^2_C + \sigma^2_{A\cdot B} + \sigma^2_{A\cdot C} + \sigma^2_{B\cdot C} + \sigma^2_{A\cdot B\cdot C} + \sigma^2_e$

- μ = grand mean
- τ_i = F effect of A, $i=1\ldots a$. $\tau_a = 0$.
- b_j = R effect of B, $j=1\ldots b$. $\sim$ iid $N(0, \sigma^2_B)$
- c_k = R effect of C, $k=1\ldots c$. $\sim$ iid $N(0, \sigma^2_C)$
- $(\tau b), (\tau c), (bc)$ = 2-way interactions, $\sim$ iid $N(0$, var comp.)
- $(\tau bc)_{ijk}$ = 3 way A·B·C effect. $\sim$ iid $N(0, \sigma^2_{A\cdot B\cdot C})$.
- e_{ijkl} = R residuals. $\sim$ iid $N(0, \sigma^2_e)$.

$\}$ indep

10. **CRD - Nested:** F2 in F1, F3 in F2 ("fully hierarchical"). Notation: F3(F1, F2).

- Say F1 fixed, F2 & F3 random. (If F2 random & F3(F2) → F3 also random)

$Y_{ijk} = \mu + \tau_i + b_{j(i)} + c_{k(ij)}$

- μ = grand mean
- τ_i = F effect of F1, $\tau_a = 0$. $i = 1 \ldots a$.
- $b_{j(i)}$ = R effect of F2 in F1. $j = 1 \ldots b$. $\sim iid\ N(0, \sigma^2_{B(A)})$
- $c_{k(ij)}$ = R effect of F3 in F1 & F2. $k = 1 \ldots c$. $\sim iid\ N(0, \sigma^2_{C(A,B)})$

} indep

- Var components: $\sigma^2_{B(A)}$, $\sigma^2_{C(A,B)}$

11. **CRD - Nested + Crossed:** (A fixed. B(A) random. C fixed. Fully crossed).

$Y_{ijkl} = \mu + \tau_i + b_{j(i)} + \gamma_k + (\tau\gamma)_{ik} + (b\gamma)_{j(i),k} + e_{ijkl}$

- μ = grand mean
- τ_i = F effect of A, $i = 1 \ldots a$. $\tau_a = 0$.
- $b_{j(i)}$ = R effect of B(A). $j = 1 \ldots b$. $\sim iid\ N(0, \sigma^2_{B(A)})$.
- γ_k = F effect of C. $k = 1 \ldots c$. $\gamma_c = 0$.
- $(\tau\gamma)_{ik}$ = F effect of A*C. $(\tau\gamma)_{ac} = 0$.
- $(b\gamma)_{j(i),k}$ = R effect of B(A)*C. $\sim iid\ N(0, \sigma^2_{B(A)\cdot C})$
- e_{ijkl} = residuals, $l = 1 \ldots$ # reps. $\sim iid\ N(0, \sigma^2_{e(A,B,C)})$

} indep

- Var Components:

$\sigma^2_Y = \sigma^2_{B(A)} + \sigma^2_{B(A)\cdot C} + \sigma^2_{e(A,B,C)}$

- Note: σ^2_e is really 3-way int - have no df left-over after allocating for this term.

$\sigma^2_e = (\tau b\gamma)_{j(i),kl}$

12. **RCBD:** Generalized, trt can be single or factorial.

$Y_{ijk} = \mu + \tau_i + b_j + (\tau b)_{ij} + e_{ijk}$

- μ = grand mean
- τ_i = trt effect, $i = 1 \ldots t$, $\tau_a = 0$. (Use A_i if random levels)
- b_j = R block effect, $j = 1 \ldots r$, $\sim iid\ N(0, \sigma^2_B)$
- $(\tau b)_{ij}$ = R trt*block effect, $\sim iid\ N(0, \sigma^2_{\tau B})$
- e_{ijk} = R residuals, $k = 1 \ldots m$, $\sim iid\ N(0, \sigma^2_e)$

- t = # trts. r = # blocks. m = reps of trt/block.
- Var components: $\sigma^2_Y = \sigma^2_B + \sigma^2_{\tau B} + \sigma^2_e$

13. **LSD:** (2 block factors A & B, trt factor C)

$Y_{ijkl} = \mu + a_i + b_j + \tau_k + e_{ijkl}$

- μ = grand mean
- a_i = R block effect of A, $\sim iid\ N(0, \sigma^2_A)$
- b_j = R block effect of B, $\sim iid\ N(0, \sigma^2_B)$
- τ_k = F trt effect of C, $k = 1 \ldots t$. $\tau_t = 0$.
- e_{ijk} = R residuals, $\sim iid\ N(0, \sigma^2_e)$.

} indep

- Var Components: $\sigma^2_Y = \sigma^2_A + \sigma^2_B + \sigma^2_e$

14. **LSD w/ >1 Rep:** (2 block factors A & B, trt factor C)

$Y_{ijkl} = \mu + a_i + b_j + c_l + \tau_k + e_{ijkl}$

- μ = grand mean
- a_i = R block effect of A, $\sim iid\ N(0, \sigma^2_A)$
- b_j = R block effect of B, $\sim iid\ N(0, \sigma^2_B)$
- c_l = R effect of rep, $l = 1 \ldots r$. $\sim iid\ N(0, \sigma^2_C)$
- τ_k = F effect of trt, $k = 1 \ldots t$. $\tau_t = 0$.
- e_{ijkl} = R residuals, $\sim iid\ N(0, \sigma^2_E)$

} indep

- Var Components: $\sigma^2_Y = \sigma^2_A + \sigma^2_B + \sigma^2_C + \sigma^2_E$

15. LSD - New levels of Blocking Factors for Ea. Rep:

$Y_{ijkl} = \mu + a_{i(l)} + b_{j(l)} + c_l + \tau_k + e_{ijkl}$

· Var Components: $\sigma_Y^2 = \sigma_{A(R)}^2 + \sigma_{B(R)}^2 + \sigma_R^2 + \sigma_e^2$

· Note: Both blocks are nested in Rep now.

· μ = grand mean

· $a_{i(l)}$ = R effect of block A in Rep, ~iid $N(0, \sigma_{A(R)}^2)$

· $b_{j(l)}$ = R effect of block B in Rep, ~iid $N(0, \sigma_{B(R)}^2)$

· c_l = R effect of Rep, ~iid $N(0, \sigma_R^2)$

· τ_k = F effect of trt, $k=1 \ldots t$. $\tau_t = 0$.

· e_{ijkl} = R effect of residuals, ~iid $N(0, \sigma_e^2)$

16. BIBD:

$Y_{ijg} = \mu + \tau_i + b_j + e_{ijg}$

· Var Components: $\sigma_Y^2 = \sigma_B^2 + \sigma_e^2$

· Note: g = indicator var. g=1 if ~~trt~~ trt i in block j, g=0 otherwise.

· μ = grand mean

· τ_i = F effect of trt, $i=1 \ldots t$. $\tau_t = 0$.

· b_j = R effect of block, ~iid $N(0, \sigma_B^2)$

· e_{ijg} = Residuals, ~iid $N(0, \sigma_e^2)$

17. RCBD Split-Plot: (R blocks, 2 fixed factors - A=WP, B=SP)

$Y_{ijk} = \mu + c_k + \tau_i + d_{ik} + \gamma_j + (\tau\gamma)_{ij} + e_{ijk}$

· Var Components: $\sigma_Y^2 = \sigma_c^2 + \sigma_D^2 + \sigma_e^2$

· μ = grand mean

· c_k = R block effect, ~iid $N(0, \sigma_c^2)$

· τ_i = F WP trt effect, $i=1 \ldots a$. $\tau_a = 0$.

· d_{ik} = WP error term, ~iid $N(0, \sigma_D^2)$

· γ_j = F SP trt effect, $j=1 \ldots b$. $\gamma_b = 0$.

· $(\tau\gamma)_{ij}$ = F A*B effect, $(\tau\gamma)_{ab} = 0$

· e_{ijk} = SP error term, ~iid $N(0, \sigma_e^2)$

indep

18. RCBD Split-Split Plot: (R blocks, 3 fixed factors, A, B, C)

$Y_{ijkl} = \mu + c_l + \tau_i + d_{il} + \gamma_j + (\tau\gamma)_{ij} + e_{ijl} + \beta_k + (\tau\beta)_{ik} + (\gamma\beta)_{jk} + (\tau\gamma\beta)_{ijk} + f_{ijkl}$

· Var components: $\sigma_Y^2 = \sigma_R^2 + \sigma_D^2 + \sigma_e^2 + \sigma_F^2$

· μ = grand mean

· c_l = R block effect, $l=1 \ldots r$. ~iid $N(0, \sigma_R^2)$

· τ_i = F WP trt effect, $i=1 \ldots a$. $\tau_a = 0$.

· d_{il} = R WP error term, ~iid $N(0, \sigma_D^2)$

· γ_j = F SP trt effect, $j=1 \ldots b$. $\gamma_b = 0$.

· $(\tau\gamma)_{ij}$ = F WP*SP trt int, $(\tau\gamma)_{ab} = 0$.

· e_{ijl} = R SP error term, ~iid $N(0, \sigma_e^2)$

· β_k = F SSP trt effect, $k=1 \ldots c$. $\beta_c = 0$.

· $(\tau\beta)_{ik}$ = F WP*SSP trt int. $(\tau\beta)_{ac} = 0$.

· $(\gamma\beta)_{jk}$ = F SP*SSP trt int. $(\gamma\beta)_{bc} = 0$.

· $(\tau\gamma\beta)_{ijk}$ = F WP*SP*SSP trt int. $(\tau\gamma\beta)_{abc} = 0$

· f_{ijkl} = R SSP error term, ~iid $N(0, \sigma_F^2)$.

19. RCBD Strip-Plot: (R blocks, Factors A + B fixed)

- 3 size EU's → A strip, B strip, AB intersection. 3 error terms.

$Y_{ijk} = \mu + C_k + \tau_i + d_{ik} + \gamma_j + e_{jk} + (\tau\gamma)_{ij} + f_{ijk}$

- Var components: $\sigma_Y^2 = \sigma_R^2 + \sigma_D^2 + \sigma_E^2 + \sigma_F^2$

- μ = grand mean
- C_k = R block effect, $k=1 \ldots r$. $\sim iid\ N(0, \sigma_R^2)$
- τ_i = F trt A effect, $i=1 \ldots a$. $\tau_a = 0$.
- d_{ik} = WP R error for A, $\sim iid\ N(0, \sigma_D^2)$
- γ_j = F trt A effect, $j=1 \ldots b$. $\gamma_b = 0$.
- e_{jk} = WP R error for B, $\sim iid\ N(0, \sigma_E^2)$
- $(\tau\gamma)_{ij}$ = A*B trt effect, $(\tau\gamma)_{ab} = 0$.
- f_{ijk} = A*B intersection error, $\sim iid\ N(0, \sigma_F^2)$

indep

20. Crossover: (EUs serve as blocks)

$Y_{ijkdc} = \mu + \alpha_i + b_{j(i)} + \gamma_k + \tau_{d(i,k)} + \lambda_{c(i,k-1)} + e_{ijk}$

- μ = grand mean
- α_i = F effect of i^{th} sequence
- $b_{j(i)}$ = R effect of j^{th} EU in seq i, ~
- γ_k = F effect of k^{th} time period
- $\tau_{d(i,k)}$ = direct effect of trt in period k, seq i. F.
- $\lambda_{c(i,k-1)}$ = F carryover effect of trt applied in period $(k-1)$ in seq i.

21. ANCOVA: (X = continuous covariate)

- Say 2 trts + control. $I_1 = 1$ if trt control, else 0. $I_2 = 1$ if trt 1, else 0.
- "extreme blocking", w/ 1 trt/block, essentially.
- Find best-fitting model:

1. Diff Slopes, Diff Intercepts: $Y_i = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \underbrace{\beta_3 X_i}_{\text{cov. impact}} + \underbrace{\beta_4 I_1 X_i}_{\text{differential impact of trt ctrl}} + \underbrace{\beta_5 I_2 X_i}_{\text{differential impact of trt 1.}} + e_i$
2. Same Slope, Diff Intercepts: $Y_i = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \beta_3 X_i + e_i$
3. Same Slope, Same Intercepts: $Y_i = \beta_0 + \beta_3 X_i + e_i$

- If start w/ (1) + find no sig ev. of diff slopes, switch to (2).

22. Repeat Measures: Several approaches.

1. Analyze as Split-Plot, rep meas = SP trt, EU = WP trt. $Y_{ijk} = \mu + \tau_i + d_{k(i)} + \beta_j + (\tau\beta)_{ij} + e_{ijk}$
   ↳ approx analysis, depends on Compound Symmetry (constant corr btwn obs on same EU, regardless of spacing.
2. Repeated Measures Model - depends on sphericity
3. MANOVA

EXPECTED MEAN SQUARES + DF

SV	# levels	Q_F	$Q_{(L)}$	$\sigma^2_{O(L)}$	$Q_{F \cdot L}$	$\sigma^2_{F \cdot O(L)}$	σ^2_e	df
F	$a=3$	$\frac{abcr}{a}=16$	0	0	0	$\frac{abcr}{abc}=2$	$\frac{abcr}{abcr}=1$	$(a-1)=3$
L	$b=2$	0	$\frac{abcr}{b}=24$	$\frac{abcr}{bc}=6$	0	2	1	$(b-1)=1$
R $O(L)$	$c=4$	0	0	6	0	2	1	$b(c-1)=6$
$F \cdot L$	$a \cdot b=6$	0	0	0	$\frac{abcr}{ab}=8$	2	1	$(a-1)(b-1)=2$
R $F \cdot O(L)$	$a \cdot c=12$	0	0	0	0	2	1	$(a-1) \cdot b(c-1)=12$
Error	(abcr)	0	0	0	0	0	1	$47-\Sigma$ others $=24$

- Say reps = 2. Total $df = rabc - 1 = 48 - 1 = 47$
- KEY: For nested terms, div by nested term and term in nest. For $O(L)$, div by (# $O(L)$ levels · # L levels)
- KEY: For $\sigma^2_{F \cdot O(L)}$, across top, the row for $F \cdot L$ gets a value b/c F + L are both part of $F \cdot O(L)$.

↓

Then multiply across to get EMS for each SV:

$EMS_F = \sigma^2_e + 2\sigma^2_{F \cdot O(L)} + 16Q_F$

$EMS_L = \sigma^2_e + 2\sigma^2_{F \cdot O(L)} + 6\sigma^2_{O(L)} + 24Q_L$

$EMS_{O(L)} = \sigma^2_e + 2\sigma^2_{F \cdot O(L)} + 6\sigma^2_{O(L)}$

$EMS_{F \cdot L} = \sigma^2_e + 2\sigma^2_{F \cdot O(L)} + 8Q_{F \cdot L}$

$EMS_{F \cdot O(L)} = \sigma^2_e + 2\sigma^2_{F \cdot O(L)}$

$EMS_{error} = \sigma^2_e$

- The yellow terms indicate what terms disappear under H_0 for the F-test for each SV's test of significance.

F-STATISTIC + F-TEST

1) Compare F to $F_\alpha, df1, df2$. $F = MS_{sv} / MS_{sv}$ under H_0. df1 = numerator df, df 2 = denom df.
   - If $F > F_\alpha$, sig ev. that H_0 doesn't hold, ie SV has sig effect. (b/c testing if F is "close" to 1. If H_0 holds, F will be closer to 1 than critical value.) → if $a/b \approx 1$ → $a \approx b$.

2) To calc F stat: $F = MS_{sv} / MS_{sv}$ under H_0.
   - For denom, look @ EMS for SV under H_0 (w/o yellow above). Find match lower in table. df2 = df for matching SV.
   - If no matching EMS lower in table, use "Approx F-test" + Satterthwaite Approx to get F denom.

↓

- M = linear combo of other EMS's to equal exact EMS you're looking for. (ie $EMS_A + EMS_B - EMS_{A \cdot B}$)
- F stat = MS_{sv} / M. df1 = df_{sv}. (M = Linear combo of MS's matching combo of EMS's)
- Satterthwaite to get df2: $$df2 = \frac{M^2}{\underbrace{\frac{MS_A^2}{df_A} + \frac{MS_B^2}{df_B} + \frac{MS_{AB}^2}{df_{A \cdot B}}}_{\text{one term for each component of } M}}$$

ESTIMATING VARIANCE Components

1. MOM Method, 2. REML. 3. MLE.

We don't do an F-test for random SV'p, as not relevant b/c of randomness of factor levels chosen for each run of experiment. Instead, for each random SV, we estimate the var component & see what proportion each component is of total var.

1. MOM Method: Using AOV tbl, set MS = EMS for ea. random SV. Solve & plug in, starting from bottom up.

Ex:

SV	MS	EMS
A (f)	MS_A	$\Leftrightarrow \sigma^2_e + r\sigma^2_{A\cdot B} + rbQ_A$
B (r)	MS_B	$\Leftrightarrow \sigma^2_e + r\sigma^2_{A\cdot B} + ra\sigma^2_B$
A·B (r)	$MS_{A\cdot B}$	$\Leftrightarrow \sigma^2_e + r\sigma^2_{A\cdot B}$
Error (r)	MSE	$\Leftrightarrow \sigma^2_e$

Want to estimate $\hat{\sigma}^2_e, \hat{\sigma}^2_B, \hat{\sigma}^2_{A\cdot B}$.

Think of each line as a $MS = \hat{EMS}$ eqn.

- Step 1: $MSE = \hat{\sigma}^2_e$.
- Step 2: $MS_{A\cdot B} = \hat{\sigma}^2_e + r\hat{\sigma}^2_{A\cdot B} \rightarrow MS_{A\cdot B} = MSE + r\hat{\sigma}^2_{A\cdot B} \rightarrow \hat{\sigma}^2_{A\cdot B} = (MS_{A\cdot B} - MSE)/r$
- Step 3: $MS_B = \hat{\sigma}^2_e + r\hat{\sigma}^2_{A\cdot B} + ra\hat{\sigma}^2_B$

$$\rightarrow MS_B = MSE + r\left(\frac{MS_{A\cdot B} - MSE}{r}\right) + ra\hat{\sigma}^2_B$$

$$\rightarrow \hat{\sigma}^2_B = (MS_B - MSE - MS_{A\cdot B} + MSE)/ra \rightarrow \hat{\sigma}^2_B = (MS_B - MS_{A\cdot B})/ra$$

Then proportionally allocate var. estimates: $\hat{\sigma}^2_Y = \hat{\sigma}^2_e + \hat{\sigma}^2_{A\cdot B} + \hat{\sigma}^2_B$

Var Component	Proportion of Tot. Var.
B (σ^2_B)	$\hat{\sigma}^2_B / \hat{\sigma}^2_Y$
A·B ($\sigma^2_{A\cdot B}$)	$\hat{\sigma}^2_{A\cdot B} / \hat{\sigma}^2_Y$
Error (σ^2_e)	$\hat{\sigma}^2_e / \hat{\sigma}^2_Y$

Problem: MOM ests could be negative. Only acceptable choices in this case:

a) Interpret it as indication of incorrect model and/or outliers.

b) Use MLE or REML estimates. Should really be using these anyways- MOMs are weak estimators, b/c 2 moments don't define a distribution.

2. MLE: Involves repeated iterations btwn obtaining estimates of β & V. MLE's are biased estimators, & size of bias can be serious. REML is an alternative, better procedure.

3. REML: Preferred var component estimator in most situations. Default provided by SAS in PROC MIXED. Never calc this by hand - use SAS. But here is the concept.

- Mixed model in Matrix form: $Y = X\beta + Zu + e$
  - $\downarrow$
  - $E[Y] = X\beta$, $var[Y] = V = ZGZ' + R$
  - u & e have var-covar matrices G & R
  - $E\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $var\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$

- Y = response vector of y's
- X = known Design matrix for Fixed effects
- β = vector of Fixed-effects variables (unknown)
- Z = known design matrix for Random effects
- u = unknown vector of random effects
- e = iid random errors

- To find REML, -2-log-likelihood fctn is partitioned into 2 components.
  - $\hookrightarrow (n-r)\log(2\pi) + \log(|V|) + \log(|X'V^{-1}X|) + (n-r)\log(P'V^{-1}P)$, r = rank of $P = Y - X(X'V^{-1}X)^{-1}X'V^{-1}Y$
- SAS maximizes this fctn to get $\hat{V}$, the REML estimators of the var components V.
- $\hat{V}$ is then used to get $\hat{\beta}$, an estimator of the fixed parameters, β, and $\hat{u}$, predicted vals of rand. effects.
  - $\hookrightarrow \hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y$ and $\hat{u} = \hat{G}Z'\hat{V}^{-1}(Y - X\hat{\beta})$

- REML Properties:
  - If G & R known, then $\hat{\beta}$ is BLUE of β, and $\hat{u}$ is BLUP of u.
  - BLUE = Best linear unbiased estimator
  - BLUP = Best linear unbiased predictor, where "best" = minimum MSE.
  - * G & R are generally unknown, so use asymptotic REML properties.

ESTIMATED STAND. ERRORS FOR MEANS + DIFFS in MEANS

For use in CI's, and anywhere else you need $\hat{SE}$.

Derive these - do not rely on a formula, b/c formula depends on model + EMS for SV. Need model!

1. $\hat{SE}(\hat{\mu})$ - Can be for cell means (indiv trts) or marginal means.

Say have RCBD w/ 2×3 factorial - N has 2 fixed levels, D has 3 fixed levels. R=3 blocks, so k=1..3.

$Y_{ijk} = \mu + b_k + \tau_i + \gamma_j + (\tau\gamma)_{ij} + e_{ijk}$, for $i=1,2$ and $j=1,2,3$ and $k=1,2,3$.

AOV:

SV	EMS
Block	$\sigma^2_e + 6\sigma^2_B$
N	$\sigma^2_e + 9Q_N$
D	$\sigma^2_e + 9Q_D$
N·D	$\sigma^2_e + 3Q_{N\cdot D}$
Error	σ^2_e

a) Marginal Mean of N: $\hat{SE}(\hat{\mu}_{i.})$

- Find $Var(\hat{\mu}_{i.}) = Var(\bar{Y}_{i..})$ then take square root.
- First, all fixed terms drop out of var. calc.

$$Var(\bar{Y}_{i..}) \Rightarrow \bar{b}_{.} + \bar{e}_{i..}$$

- Replace ea. random term w/ its var component, + divide by product of all terms averaged over (ie the dots).

· $Var(\bar{Y}_{i..}) = \frac{\sigma^2_B}{3} + \frac{\sigma^2_e}{3\cdot 3}$ → Get common denom → $= \frac{3\sigma^2_B + \sigma^2_e}{9} = \frac{6\sigma^2_B + 2\sigma^2_e}{18}$

· Then match to EMS. If can't find a match, use Satterthwaite + create Linear combo.

$= \frac{EMS_B + EMS_{ERR}}{18}$. Then plug in MS in place of EMS + take sqrt:

· $\hat{SE}(\hat{\mu}_{i.}) = \sqrt{(MS_B + MSE)/18}$

· If had REML estimators, can just plug them in instead of finding EMS equivalent of equation.

b) Cell Means: $\hat{SE}(\hat{\mu}_{ij})$ · Very similar. First find $Var(\hat{\mu}_{ij})$, then take square root.

$Var(\hat{\mu}_{ij}) = Var(\bar{Y}_{ij.}) \Rightarrow \bar{b}_{.} + \bar{e}_{ij.} = \frac{\sigma^2_B}{3} + \frac{\sigma^2_e}{3} = \frac{6\sigma^2_B + 6\sigma^2_e}{18} = \frac{EMS_B + 5EMS_{ERR}}{18}$

· Then plug in MS in place of EMS + take square root.

· $\hat{SE}(\hat{\mu}_{ij}) = \sqrt{(MS_B + 5MSE)/18}$

*Note: Can use EMS equations output from SAS, or derive manually.

2. $\hat{SE}(\hat{\mu}_1 - \hat{\mu}_2)$ - Can be for Cell Means or Marginal Means

· Similar to previous page. All fixed components drop out, & identical components cancel out.

· Start with $Var(\bar{Y}_{i...} - \bar{Y}_{n...})$, or difference of whatever means you are comparing.

$$Var(\bar{Y}_{i..} - \bar{Y}_{n..}) \Rightarrow \underbrace{Var[\bar{b}_{.j} - \bar{b}_{.j}]}_{\text{identical terms cancel}} + Var[(\bar{\tau b})_{ij'} - (\bar{\tau b})_{nj}] + Var[\bar{e}_{i..ijk} - \bar{e}_{n..ijk}]$$

each of these diffs are indep, so turn into $2 * \sigma^2_{\text{var component}}$, div by product of dot terms

$$= 0 + \frac{2\sigma^2_{AB}}{2} + \frac{2\sigma^2_e}{2 \cdot 3} \quad (\text{say } j=1,2 \;\&\; k=1,2,3)$$

Then get common denom & find corresponding EMS (or linear combo)

$$= 2 \cdot \frac{(9\sigma^2_{A \cdot B} + \sigma^2_e)}{18} = \frac{2(EMS_{A \cdot B})}{18} \rightarrow \text{Plug in MS \& take sqrt:}$$

$$\rightarrow \hat{SE}(\hat{\mu}_{1.} - \hat{\mu}_{2.}) = \sqrt{MS_{A \cdot B}/9}$$

* NOTE: If using linear combo/Satterthwaite, you'll need to calc the df if using the SE as part of a CI (for use in $t_{\alpha/2, df}$ critical value)

TUKEY HSD + TUKEY-KRAMER: PAIRWISE COMPARISON OF TRT MEANS

1. Tukey HSD:

$HSD = q_{\alpha,t,v}\sqrt{\frac{1}{2}\hat{SE}(\mu_i-\mu_j)^2}$
- $\alpha = .05$ or other chosen value
- t = # treatments
- v = df for $\hat{SE}$, obtained using mom.
- $q_{\alpha,t,v}$ = Studentized range stat (Tbl VII). $(\bar{y}_{max}-\bar{y}_{min})/\hat{SE}(\hat{\mu}_i)$

- This is for the simple case when SE is same for all means being compared.
- Most basic case, for $\mu_i-\mu_j$: $\sqrt{1/2\,\hat{SE}(\mu_i-\mu_j)} = \sqrt{\frac{1}{2}\cdot 2(\sigma_e^2/n)} = \sqrt{\sigma^2_e/n} = \sqrt{MSE/n}$
- Also works for marginal means (if 2 factors), if both factors have fixed levels.
  Just end up w/ different $\hat{SE}$'s, leading to $\sqrt{MSE/rb}$ for $\bar{\mu}_{i.}-\bar{\mu}_{k.}$, or $\sqrt{MSE/ra}$ for $\bar{\mu}_{.j}-\bar{\mu}_{.h}$.

Test: $\mu_i \neq \mu_j$ if $|\bar{y}_i-\bar{y}_j| > HSD$. Simult $100(1-\alpha)\%$ CI: $|\bar{y}_i-\bar{y}_j| \pm HSD$

- If have an A*B sig int, need to do pairwise A comparison @ ea level of B, or vice versa.
  ↓
- Bonferroni-Adj Tukey HSD: Use $q_{\alpha/b,a,df}$ for comparing A trt means @ ea. level of b.
  - b = # B levels
  - a = # A levels
  - df = same df for $\hat{SE}(\hat{\mu}_{i.}-\hat{\mu}_{j.})$
- Still have eq. reps + vars for ea. level of B.

2. Tukey-Kramer: If SE's/vars not same, then $\hat{SE}$'s + est. vars are used.
- CRD w/ unequal reps
- Complex structures, ie blocking
- 2 Factors, A fixed + B random
- Same formulas as above, but before were using actual SE vals - now using estimates.
- If using linear combo/Satterthwaite for $\hat{SE}$, use the Satterthwaite df also.

Ex: A fixed, B random, C random. $y_{ijkl} = \mu + \tau_i + b_j + c_k + (\tau b)_{ij} + (\tau c)_{ik} + (bc)_{jk} + (\tau bc)_{ijk} + e_{ijkl}$
Want Tukey for pairwise comparison of A trt means.

$Var(\bar{y}_{i...}-\bar{y}_{h...}) = Var[\bar{b}_. + \bar{c}_. + (\overline{\tau b})_{i.} + (\overline{\tau c})_{i.} + (\overline{bc})_{..} + (\overline{\tau bc})_{i..} + \bar{e}_{i...}] -$
$Var[\bar{b}_. + \bar{c}_. + (\overline{\tau b})_{h.} + (\overline{\tau c})_{h.} + (\overline{bc})_{..} + (\overline{\tau bc})_{h..} + \bar{e}_{h...}]$

$= 2Var((\overline{\tau b})_{i.}) + 2Var((\overline{\tau c})_{i.}) + 2Var((\overline{\tau bc})_{i..}) + 2Var(\bar{e}_{i...})$
$\quad\;\, \llcorner j \qquad\qquad \llcorner k \qquad\qquad \llcorner jk \qquad\qquad \llcorner jkl$

$= \frac{2\sigma^2_{A\cdot B}}{b} + \frac{2\sigma^2_{A\cdot C}}{c} + \frac{2\sigma^2_{A\cdot B\cdot C}}{bc} + \frac{2\sigma^2_e}{bcr}$

Ex ctd: $= 2\left[cr\sigma^2_{A\cdot B} + br\sigma^2_{A\cdot C} + r\sigma^2_{A\cdot B\cdot C} + \sigma^2_e\right] / bcr$

Then we would check for exact EMS in AOV table. Don't have it (HO9 p 28), so make linear combo:

$= \left(\frac{2}{bcr}\right)[EMS_{A\cdot B} + EMS_{A\cdot C} - EMS_{A\cdot B\cdot C}]$. This is $\hat{Var}$, so take sqrt to get $\hat{SE}(\hat{\mu}_{i..} - \hat{\mu}_{k..})$.

So $HSD = q_{\alpha, a, v}\sqrt{\frac{1}{2}\left(\frac{2}{bcr}\right)\underbrace{[MS_{A\cdot B} + MS_{A\cdot C} - MS_{A\cdot B\cdot C}]}_{\text{"M"}}}$

To get df v: $$v = \frac{M^2}{\frac{MS^2_{A\cdot B}}{df_{A\cdot B}} + \frac{MS^2_{A\cdot C}}{df_{A\cdot C}} + \frac{MS^2_{A\cdot B\cdot C}}{df_{A\cdot B\cdot C}}}$$

DUNNETT'S COMPARISON OF ALL TRT MEANS TO CONTROL

- Similar methods as Tukey. If complex w/ diff SE's + reps, use LSE ~~ests~~ for $\hat{\mu}_i$, $\hat{\mu}_c$ + $\hat{SE}$.

1. $D = d_{\alpha, t-1, v}\sqrt{2\,SE(\mu_i - \mu_c)^2}$
   - $\alpha = .05$ or other chosen val.
   - $t-1 = \#$ trts -1
   - v = df for $\hat{SE}$, obtained using mom
   - d = Dunnett crit. vals, Table III.

- **Test:** 1. Trt mean ≠ Ctrl ($H_0: \mu_i = \mu_c$): Reject H_0 if $|\bar{y}_i - \bar{y}_c| > D$.
  Simult $100(1-\alpha)\%$ 2-sided CI for diff in mean + ctrl: $(\bar{y}_i - \bar{y}_c) \pm D$

  2. Trt mean < ctrl ($H_0: \mu_i \le \mu_c$): Reject H_0 if $(\bar{y}_i - \bar{y}_c) > D$

  3. Trt mean > ctrl ($H_0: \mu_i \ge \mu_c$): Reject H_0 if $(\bar{y}_i - \bar{y}_c) < -D$

- If have sig A*B int, need to do A comparisons @ ea. level of B.

↓

Bonferroni-Adj. Dunnett's: Use $d_{\alpha/b, a-1, v}$
- $b = \#$ levels of B
- $a = \#$ levels of A
- v = df, same as above

- If don't have eq. $\hat{SE}$'s + reps, similar to Tukey-Kramer. Use LSE approx of $\hat{\mu}_i$, $\hat{\mu}_c$ + $\hat{SE}(\hat{\mu}_i - \hat{\mu}_c)$.
- See prev. pg for Satterthwaite example if need linear combo to estimate $\hat{SE}$.
- This page is using generalized $\hat{\mu}_i$ + $\hat{\mu}_c$. Can do this w/ cell means or marginal means, just be careful to adjust $t-1$, v, and derive proper SE.

HSU'S - FIND "BEST" (LG or SM) TRT MEANS

- Produces a group of trt means that contains the "best" (ie largest, smallest) mean w/ prob $100(1-\alpha)\%$.
- Uses Dunnett's D statistic from prev. pg.

- Let D = Dunnett's statistic, see prev. pg.

1. Calc a vector of t M_i's for "best" = max, or m_i's for "best" = min.
   - $M_i = \max(\bar{y}_h)$, $h \neq i$ for $i = 1 \ldots t$. Largest sample mean excl. i^{th} sample mean, for ea. trt.
   - $m_i = \min(\bar{y}_h)$, $h \neq i$ for $i = 1 \ldots t$. Smallest sample mean excl. i^{th} sample mean, for ea. trt.

2. For max, $K_i = M_i - D$
   For min, $k_i = m_i - D$

3. Test for Best = Max: i^{th} mean μ_i is in group w/ largest means if $\hat{\mu}_i \geq K_i$.
   Test for Best = min: i^{th} mean μ_i is in group w/ smallest means if $\hat{\mu}_i \leq k_i$.

- Use Dunnett's guidelines for unequal reps or complex situations. Derive $\hat{SE}$, adjust $t-1$ + v appropriately.
- If repeating HSU's test for group of "best" trt means for A, at each level of B, use Dunnett's Bonferroni adjustment.

CONTRASTS FOR LINEAR COMBOS OF TRT MEANS

- Tukey, Dunnett & Hsu are really contrasts. We can create our own contrasts to test questions of interest.
- Can test things like "is $\mu_1 = \mu_3$?" ($C = \mu_1 - \mu_3$) or "is avg of μ_1 + μ_2 equal to avg of $\mu_3, \mu_4 + \mu_5$?" ($C = (\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4 + \mu_5)/3$)
- Can use polynomial orthog. contrasts to test for trends in trt means. (Req's equally spaced quant. levels)

Orthogonality: C_1 & C_2 are mutually orthogonal if $\sum_{i=1}^{t} \frac{k_i d_i}{n_i} = 0$, where k_i = coeffs for C1, d_i = coeffs for C2.
- Is a convenient property b/c then SS_C's partition SS_{TRT}, using $(t-1)$ orthog contrasts.
- Is not mathematically necessary. Non-orthog contrasts may be ones researcher is interested in.

Testing: $H_0: C = 0$, ie no signif. evidence of trend, or property you are testing ($\mu_1 = \mu_3$) holds.
- If $\hat{C} < 0$, evidence of neg trend. $\hat{C} > 0$, evidence of pos. trend.

Adjusting P-Vals:
- If testing 2+ contrasts at once, adjust pvals for mult comparison. (SAS does NOT adjust.)
- Use Bonferroni Adjustment:
  1. If non-orthog, compare pvals to $\alpha_{PC} = \alpha / M$, M = # of contrasts.
  2. If orthog, compare pvals to $\alpha_{PC} = 1 - (1-\alpha)^{1/M}$.
- Only use Scheffe's adjustment if choosing contrasts after running experiment (which is cheating). Scheffe's adjusts pval for all possible contrasts.

Process for Testing Trends:
1. Pick t = # levels of trt being contrasted, & P_1 = linear, P_2 = Quadratic, etc.
2. Using chosen t & p_i pick coeffs from Table XI.
3. Apply coeffs to appropriate trt means:

Ex1: 2 factors. ~~Linear~~ Quadratic trend for levels of A. Then for levels of B. Then multiply for A * B interaction.
Say A & B each have 3 levels. So t=3, p=2, coeffs = 1 -2 1

Main A:

	1	1	1	-2	-2	-2	1	1	1
Main A:	μ_{11}	μ_{12}	μ_{13}	μ_{21}	μ_{22}	μ_{23}	μ_{31}	μ_{32}	μ_{33}
		A1			A2			A3	

	1	-2	1	1	-2	1	1	-2	1
Main B:	μ_{11}	μ_{12}	μ_{13}	μ_{21}	μ_{22}	μ_{23}	μ_{31}	μ_{32}	μ_{33}
	(B1	B2	B3)	(B1	B2	B3)	(B1	B2	B3)

	1	-2	1	-2	4	-2	1	-2	1
A*B Int: (multiply)	μ_{11}	μ_{12}	μ_{13}	μ_{21}	μ_{22}	μ_{23}	μ_{31}	μ_{32}	μ_{33}

Trends for 3 Factors: Use same coeffs on A & B, then double & flip signs for C, for interactions.

Ex 2: Linear trend for A, main effect. A=3 levels, B=2 levels, C=2 levels. $t=3, p=2$, coeffs = -1 0 1

$$C = \underbrace{\overset{-1}{\mu_{111}}\ \overset{-1}{\mu_{112}}\ \overset{-1}{\mu_{121}}\ \overset{-1}{\mu_{122}}}_{A=1}\quad \underbrace{\overset{0}{\mu_{211}}\ \overset{0}{\mu_{212}}\ \overset{0}{\mu_{221}}\ \overset{0}{\mu_{222}}}_{A=2}\quad \underbrace{\overset{1}{\mu_{311}}\ \overset{1}{\mu_{312}}\ \overset{1}{\mu_{321}}\ \overset{1}{\mu_{322}}}_{A=3}$$

Handling Trtmts Missing Obs:

- Have to modify contrast - a contrast is non-est if any included μ_{ij} has 0 obs.
- To avoid bias, can't have unbalanced contrasts:

Ex: 8 trts, following #'s for each: $n_{11}=0$, $n_{12}=5$, $n_{13}=4$, $n_{14}=5$, $n_{21}=4$, $n_{22}=0$, $n_{23}=7$, $n_{24}=6$

C ⇒ Compare mean responses of A (2 levels). No sig A·B interaction.

$$C = \underline{\mu_{12}} + \mu_{13} + \mu_{14} - \underline{\mu_{21}} - \mu_{23} - \mu_{24}$$

- This is not a valid contrast - is biased! Can't include μ_{21}, b/c $\mu_{11}=0$. Can't include μ_{12}, b/c $\mu_{22}=0$.
- Better approach: $C = \mu_{13} + \mu_{14} - \mu_{23} - \mu_{24}$

  (Use only B=3 & B=4, since obs are available for both levels of A.)

Ex: Means observed as follows, * = missing.

	B1	B2	B3	B4
A1	$\bar{y}_{11}$	*	*	$\bar{y}_{14}$
A2	$\bar{y}_{21}$	$\bar{y}_{22}$	$\bar{y}_{23}$	*
A3	$\bar{y}_{31}$	*	$\bar{y}_{33}$	$\bar{y}_{34}$

- $C = \mu_{11} - \mu_{14} - \mu_{31} + \mu_{34}$
  - Interaction (diff in F2 contrast @ A1 vs A3)
  - Estimable - none of these ij indices are missing

- $C = \mu_{11} + \mu_{21} + \mu_{31} - \mu_{13} - \mu_{23} - \mu_{33}$
  - Main effect of B
  - Not estimable, b/c μ_{13} is missing

- $C = \mu_{11} - \mu_{14} + \mu_{21} - \mu_{24} + \mu_{31} - \mu_{34}$
  - Main effect of B
  - Not estimable, b/c μ_{24} is missing.

F-Test Basics:

- $H_0: C=0$.
- F statistic: MS_C/MSE ($SS_{TOT} - SSE = SS_{TRT}$ → Want to decompose SS_{TRT} into $SS_{C1} \ldots SS_{Cm}$.)

SAMPLE SIZE, RELATIVE EFFICIENCY & POWER

1. **Detecting Diff Btwn At Least 1 Pair of Trt Means:** (fixed)
   - Want α-level test with power γ to detect difference D.
   - D and estimate of σ_e^2 supplied by researcher.
   - Use Keuhl's Table IX for power curve.

$$\phi = \sqrt{\lambda/t}, \quad \lambda = \frac{rD^2}{2\sigma_e^2}, \quad r = \#\text{ reps}, \quad v1 = t-1, \quad v2 = t(r-1). \quad t = \#\text{ trts}.$$

   - For eq. reps, $n = tr$, so find r to give desired power (plug in/iterate), then get $n = tr$.
   - For 2 factors, to detect diff btwn any 2 cell means μ_{ij} vs μ_{nk}, $v1 = (a-1)(b-1)$. $v2 = t(r-1)$.

2. **Detecting Diff Btwn Any 2 Marginal Means:** (fixed, main effects)
   - Say have factors A + B. Researcher specifies $D = \bar{\mu}_{i.} - \bar{\mu}_{n.}$ to detect, + σ_e^2.
   - Use Keuhl's Table IX for power curve.

$$\phi = \sqrt{\lambda/t}, \quad \lambda = \frac{rD^2}{2\sigma_e^2}, \quad r = \#\text{ reps}.$$

   - For A marginals, $v1 = (a-1)$. $v2 = t(r-1)$.
   - For B marginals, $v1 = (b-1)$. $v2 = t(r-1)$.

   - For eq. reps, # reps $= r/b$ for A marginals, or r/a for B marginals.

3. **Power for Var Components Test:** (Random, $H_0: \sigma_A^2 = 0$)
   - Use Keuhl's Table X (uses λ instead of ϕ)

$$v1 = t-1. \quad v2 = n-t. \quad \lambda = \sqrt{1 + r\,\sigma_A^2/\sigma_e^2}, \quad r = \#\text{ reps}.$$

4. **Power Fctn for Normal Mean, σ^2 known:** $\left(TS = (\bar{x}-\mu_0)/\sqrt{\sigma^2/n}\right)$

		POWER	PVAL
$H_0: \mu \le \mu_0$	$\gamma(u) =$	$\Phi\left[Z_\alpha + \frac{(\mu-\mu_0)}{\sigma/\sqrt{n}}\right]$	$1 - \Phi(TS)$
$H_0: \mu \ge \mu_0$	$\gamma(u) =$	$\Phi\left[-Z_\alpha + \frac{(\mu-\mu_0)}{\sigma/\sqrt{n}}\right]$	$\Phi(TS)$
$H_0: \mu \ne \mu_0$	$\gamma(u) =$	$\Phi\left[-Z_{\alpha/2} + \frac{(\mu_0-\mu)}{\sigma/\sqrt{n}}\right] + 1 - \left[Z_{\alpha/2} + \frac{(\mu_0-\mu)}{\sigma/\sqrt{n}}\right]$	$2\left[1 - \Phi(\lvert TS\rvert)\right]$

* If σ^2 unknown, depends on t dist and ncp Δ.

* For asymptotic normality, can use the above Z-based fctns.

1. Rel Efficiency of RCBD to CRD:

- The measure of # of obs needed to get same precision in estimating trt means as have in RCBD.
- $\hat{RE} > 1$ means RCBD more efficient. (1.25 = RCBD is 125% more efficient. Need $n \cdot 1.25$ for CRD to equal.)
- If CRD is more efficient, can't go back + rerun your data as a CRD. Live with it - your randomization has been done. This is useful knowledge for next time.

$$\hat{RE} = K + (1-K)H, \quad K = \frac{r(t-1)}{rt-1}, \quad H = MSB/MSE. \quad (t = \#\text{ trts},\ r = \#\text{ blocks})$$

2. Rel Efficiency of LSD:

- LSD to RCBD (LSD Rows = RCBD Blocks): $\left(\frac{t-1}{t}\right) + \left(1 - \frac{t-1}{t}\right) MS_{COL}/MSE$
- LSD to RCBD (LSD Cols = RCBD Blocks): $\left(\frac{t-1}{t}\right) + \left(1 - \frac{t-1}{t}\right) MS_{ROW}/MSE$
- LSD to CRD: $\left(\frac{t-1}{t}\right) + \left(1 - \frac{t-1}{t}\right)\left[\frac{MS_{ROW} + MS_{COL} + MS_{TRT}}{2MSE}\right]$

FRACTIONAL FACTORIAL EXPERIMENTS

- A 2^{n-p} design has 2^n possible trts, with 2^p selected trts.
- Given the generators (defining contrasts), specify selected trts, alias set, resolution & important effects.
- To choose included trts, randomly choose +1 or -1 for each generator. Keep only trts whose values match each generator's +1 or -1.
- This is for each trt having 2 levels (hi/lo). More complex otherwise.
- The # of generators (not including implicit) equals p.
- Find implicit generators by multiplying all possible combos of chosen generators.
- The contrasts w/in the Alias Set define which effects are confounded.

Ex: A model includes 7 factors, ABCDEFG, each with 2 levels. Use a 2^{7-3} design.
Selected generators are: ABDG = -1, ACDF = +1, BCDE = -1.

1. $2^{7-3} = 2^4 = 16$ trts are included in this study.

2. Would the trt (A B C D E F G) = (H L H H L H H) be included?
   - Take + and - (for H & L) vals for this trt, & multiply together selections for ea. generator.
     - ABDG = (+)(-)(+)(+) = -1 → matches generator 1.
     - ACDF = (+)(+)(+)(+) = 1 → matches generator 2.
     - BCDE = (-)(+)(+)(-) = 1 → Does Not match generator 3, so trt not included.

3. Find implicit generators.
   - (ABDG)(ACDF) = A^2BCD^2FG = BCFG (squared terms = 1)
   - (ABDG)(BCDE) = AB^2CD^2EG = ACEG
   - (ACDF)(BCDE) = ABC^2D^2EF = ABEF
   - (ABDG)(ACDF)(BCDE) = $A^2B^2C^2D^3EFG$ = DEFG

4. What is Resolution? Smallest # of terms in any generator, incl. implicit, is 4. So R$\overline{IV}$.

5. Which interactions do we need to assume not significant to estimate Factor A?
   - Set up Alias Set, just for Factor A:

I	ABDG	ACDF	BCDE	BCFG	ACEG	ABEF	DEFG
A	BDG	CDF	ABCDE	ABCFG	CEG	BEF	ADEFG

These are the effects A is confounded with, shown via Alias Set for A.
Must assume all these are not significant to estimate A.

· Main Effects:

· A has only 2 levels, so Main Effect of A is: $(\hat{\mu}_{2\ldots} - \hat{\mu}_{1\ldots}) = (\bar{Y}_{2\ldots} - \bar{Y}_{1\ldots})$

$\frac{2}{n} A' \times Y$ where · A = vector of (+1)-high, or (-1)-low levels for each Y response.

· Y = vector of responses

· 2-Way Interactions:

· A and B both have only 2 levels, so A*B int is: $(\hat{\mu}_{11} - \hat{\mu}_{12}) - (\hat{\mu}_{21} - \hat{\mu}_{22}) = (\bar{Y}_{11..} - \bar{Y}_{12..}) - (\bar{Y}_{21..} - \bar{Y}_{22})$

$\frac{4}{n} (A \circ B)' \times Y$ where · $(A \circ B)$ = vector of each +/- 1 element of A times ea. element of B, $[A_1B_1 \quad A_2B_2 \ldots]$

· Y = vector of responses.

· Calculating Resolution: A design is of Resolution R if all effects containing S or fewer factors are not confounded w/ any effects containing fewer than R-S factors.

Ex: Resolution 4

S=1: Mains not confounded with (4-1)=3 → 2 way or other mains.

S=2: 2 way ints not ~~confounded~~ confounded with (4-2)=2 → 1 way (mains). So some 2-factor ints confounded w/ other 2-factor ints.

ASSESSING MODEL CONDITIONS (AOV Testing)

3 conditions: normality, equal vars, independence.

1. Normality:
   - If each n_i large (>15-20), Normal Prob Plot + SW separately for each of t trt pops. Never occurs IRL.
   - Do this: Normal Prob Plot & SW test for residuals, all obs together at once.
     - Shapiro-Wilks (SW) Test: H_0, data is normal.

2. Equal vars:
   - Brown-Forsythe-Levene (BFL) Test of Homog of Vars
     - Need all $n_i \geq 3$ to do BFL.
     - Apply BFL to all Y_{ij} data vals together.
     - H_0: data has equal variances.

4. Independence:
   - To test for corr in residuals, must have a time sequence or spatial ordering index for data.
   - Normal Data w/ Index: Use Durbin-Watson test to check for AR(1) corr.
     - DW test statistic, $H_0: \rho = 0$.
     - for d_L & d_U vals, use Table for 'Significant Points for d_L & d_U.'
     - If $DW < d_L$, reject H_0. If $DW > d_U$, do not reject H_0. If $d_U < DW < d_L$, inconclusive.
   - Non-normal Data w/ Index: Use Runs Test.
     - Center all obs: $Y_t = X_t - \bar{X}$, count # runs (R), ie seqs of obs all pos or all neg.
     - Count # of pos Y_t's (n_1) & # of neg Y_t's (n_2).
     - when n_1 & $n_2 \leq 20$, use Table A30 to get R_L & R_U vals (Annals of Math. Stats.)
     - If $R < R_L$, pos corr. If $R > R_U$, neg corr. If $R_L < R < R_U$, inconclusive.
   - If no index, can only verify independence by examining how experiment is run.

- Troubleshooting:

1. Unequal Reps: Results minimally affected by BFL, but SW is more sensitive to this. Use studentized residuals to help w/ this.
2. Mixed Model: Use Pearson's Conditional Residuals. In this case, can't use BFL or SW - look @ graphs.
3. $n_i < 3$: BFL invalid here - use graphs to detect eq. vars.

* How to detect if a residual is an outlier:
  - Box Plot - see if residual is 'extreme outlier', beyond 3IQRs
  - Declare Y_{ij} an outlier if $|\hat{e}^*_{ij}| \geq t_{.0005, df_E}$ where $\hat{e}^*_{ij}$ = studentized resid $(\hat{e}_{ij} / \hat{SE}(\hat{e}_{ij}))$

· What to do when model conditions not met?

1. Transform Data:

· Box-Cox Transform (fix normality) or Power Transform (fix vars)

· If both conditions not met, transform usually fixes both. Try Box-Cox first.

· Be very careful with transformations esp ones like $Y_{ij} = 1/\sqrt{x_{ij}}$, where ordering of trt means is reversed. Careful with CI's, because $x = g(y) \nrightarrow \mu_y = g^{-1}(\mu_x)$.

· Suggested CI on transform data: $\mu_y \; CI = (e^{L_x + \hat{\sigma}^2/2}, e^{U_x + \hat{\sigma}^2/2})$

· L_x & U_x are CI limits on transformed data.

· $\hat{\sigma}^2$ is MSE from original data (ie est. variance)

2. GLIMMX in SAS:

· GENMOD & GLIMMX for non-normal analysis if data from Exponential family.

3. Non-Parametric Tests:

· Kruskal-Wallis (KW) Test:

· Requires data to be from same loc-scale family, w/ same scale (loc can be different).

· H_0: There is no trt diff, ie $\tau_1 = \ldots = \tau_t$. (Trt pops have identical dists.)

· Generalization of Wilcoxon Rank Sum proc. * $KW = \frac{SS_{TRT}}{SS_{TOT}}$ w/ ranks instead of Y's.

1. R_{ij} = rank for ea obs, ranking based on all obs together.

2. Compute total & mean Ranks for ea trt.

3. KW test statistic computed. Reject H_0 if $KW > h_\alpha$, h_α in Table A.12 (Hollander & Wolfe).

4. Large sample approx: Reject H_0 if $KW > \chi^2_{t-1,\alpha}$ (upper α %tile)

· Friedman's Test: (for blocked data)

· Like KW, except data is ranked separately in each block. Then total & mean ranks calced all together.

· F_r test statistic computed.

· H_0 rejected if $F_r > F_\alpha$ from Table A.22 (H & W), or if $F_r > \chi^2_{t-1,\alpha}$.

· Shortcut: transform data to ranks. Run AOV in SAS on transformed data. Obtain SS_{TRT} from AOV (don't use AOV results). Then $F_r = \frac{12}{t(t+1)} SS_{TRT}$.

4. Correlation:

· If data pos. corr, this inflates Type I error rate.

· Must build correlation into model. Crossover is one example where corr is built-in.

AOV + POST-AOV INFERENCES PROCESS

1. F-Test:
   - For fixed effects, use AOV F-test (PROC GLM ok).
     - Test if all means equal for each SV.
     - If reject H_0: All means equal, do post-AOV inference to investigate diffs.
     - Start at the bottom, work way up, stop when get to sig. ev. of an interaction. (See below.)
   - For random effects, no AOV-test, not interested in whether ea. var component = 0.
     - Just estimate random effects + calculate proportion of var $\hat{\sigma}_Y^2$ for ea. effect.
     - Do not use PROC GLM for AOV or LSMEANS, + remember to include /TEST in MODEL statement. Even with "/TEST," GLM underests the SE's (only considers σ_e^2 in denom). Use PROC MIXED.
     - If an SV involves a random effect (ie F*R), SV is random.
     - No post-AOV inferences, just estimate proportions of ea. var component.
     - Blocking effects are always random.

   It is not wrong to do an F-test, we just usually don't care if var component = 0, care more about relative proportions.

2. Post-AOV Inferences:
   - On fixed effects only.
   - Include contrasts, Tukey, Dunnett's, Hsu's.

3. Choosing What Factor Combos to Test:
   - If a factor is involved in an interaction w/ a fixed effect, must analyze factor at each level of interaction factor separately.
   - If factor interacts w/ a random effect, don't care - can avg over levels of 2nd factor as if no int.

Ex: 2 factors:
- A fixed, B fixed: If A*B sig, analyze A at ea level of B. If not, analyze A avg over B.
- A fixed, B rand: If A*B sig, or not, don't care b/c A*B rand. Analyze A avg'ed over B.
- A rand, B rand: No post-AOV inference, just allocate var component proportions.

4. Nested Factors:
   - Start with F-test for outside factor, work your way in.
   - If a factor is random, anything nested within that factor is random.
   - If a fixed factor is sig, post-AOV comparisons at each level of the factor that the sig factor is nested in. (F2(F1) → if F2 sig, analyze F2 at each level of F1.)

Ex: Nested Factors:

- Case 1: F1 fixed, F2 in F1 random, F3 in F2 random.
    1. Test F-test on F1, for diffs in means. If rejected, mult. comparisons across F1 levels.
    2. Test F-test on F2, to see if $\sigma^2_{F2(F1)}=0$. If rejected, no further comparisons.
- Case 2: F1 fixed, F2 in F1 fixed, F3 in F2 random.
    1. F-test for F1. If rejected, mult. comparisons across levels of F1.
    2. F-test for F2(F1). If rejected, mult comparisons of F2 separately for ea level of F1.
    3. Could also test $\sigma^2_{F3(F1,F2)}=0$ if H_0: rejected for F2's all equal.
- Case 3: F1 random, F2 in F1 random, F3 in F2 random.
    1. F-test for F1. If rejected, no further comparisons.
    2. F-test for F2. If rejected, no further comparisons. (Just one ~~test~~, avg'ed over F1 levels.)

Ex: 3 fixed factors - A, B, C.

1. If A*B*C sig, analyze A separately at ea B-C combo.
2. If A*B*C not sig, A*B sig, A*C & B*C not sig, analyze A separately at ea B level, avg'ed over C.
3. If A*B*C not sig, A*B & A*C sig, analyze A at ea level of B, and then at ea level of C.
4. If A*B*C not sig, all 2-way ints not sig, analyze A avg'ed over B & C levels.

Ex: 3 factors - A fixed, B fixed, C random.

1. Do not care if A*B*C sig, as this is random SV.
2. If A*B sig, A*C sig, B*C sig, analyze A at ea level of B. (Don't care about *C ints b/c C is random)
3. If A*C sig, A*B not sig, analyze A avg'ed over levels of B & C.

5. Mix of Nested & Crossed:

- Say A fixed, B(A) random, C fixed.
- Must consider all SV's: A*C, B(A)*C, A, B(A), C.
- Same methodology as before - if factor of interest interacts with a fixed factor, analyze separately for ea level of fixed factor. If interacts with a random factor, can avg over levels of random factor. Nested rules apply for analysis of main effects with nesting.

Expressing Models in Matrix Form

1. Matrix Form: $Y = X\beta + e$
   - Y = vector of responses $(1 \times n)$
   - X = design matrix $(n \times t)$
   - β = vector of parameters $(t \times 1)$
   - e = vector of residuals

2. Cell Means Model: $t = 4$, $n_1 = n_2 = 3$, $n_3 = n_4 = 2$.

$$\beta = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3. Basic Effects Model, no restrictions: $t = 4$, $n_1 = n_2 = 3$, $n_3 = n_4 = 2$ (now X is $n \times t+1$)

$$\beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad X = \begin{array}{c} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \\ \tau \;\; \mu_1 \;\; \mu_2 \;\; \mu_3 \;\; ? \end{array}$$

So now we are over-parameterized.

Over-parameterized.

4. Effects Model with Restrictions: $t = 4$, $n_1 = n_2 = 3$, $n_3 = n_4 = 2$. Set $\tau_4 = 0$.

$$\beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \quad X = \begin{array}{c} \mu \quad \tau_1 \quad \tau_2 \quad \tau_3 \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Leave any 0 parameters out of design matrix. This extends to more complex models, where you'd leave τ_4, γ_3 and $(\tau\gamma)_{43}$ out as all = 0.

* Usually we are doing Effects Model with restrictions.

5. Mixed Model Matrix Form: $Y = X\beta + Zu + e$
   - Y = vector of responses
   - X = design matrix for fixed effects
   - β = parameters vector for fixed effects
   - Z = design matrix for random effects
   - u = parameters vector for random effects
   - e = vector of residuals

· Mixed model illustration: Say $A=4$ fixed levels, $B=3$ random levels, # reps $=2$, $t=12$ trts.

· $y_{ijk} = \mu + \tau_i + b_j + (\tau b)_{ij} + e_{ijk}$ for $i=1\ldots4$, $j=1\ldots3$, $k=1\ldots2$.

· Restrictions: 1) $\tau_4 = 0$ 2) $b_j \sim iid\ N(0, \sigma_B^2)$. $(\tau b)_{ij} \sim N(0, \sigma^2_{A\cdot B})$. $e_{ijk} \sim N(0, \sigma_e^2)$. All are indep.

$$Y = \begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{131} \\ y_{132} \\ \vdots \\ y_{432} \end{bmatrix} \quad X = \begin{array}{c} \begin{matrix} \mu & \tau_1 & \tau_2 & \tau_3 \end{matrix} \\ \begin{bmatrix} 1_6 & 1_6 & 0_6 & 0_6 \\ 1_6 & 0_6 & 1_6 & 0_6 \\ 1_6 & 0_6 & 0_6 & 1_6 \\ 1_6 & 0_6 & 0_6 & 0_6 \end{bmatrix} \end{array} \quad \beta = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}$$

(24×1) (24×4) (4×1)

$$Z = \begin{array}{c} \begin{matrix} b_1 & b_2 & b_3 & (\tau b)_{11} & (\tau b)_{12} & \cdots & (\tau b)_{43} \end{matrix} \\ \begin{bmatrix} 1_2 & 0_2 & 0_2 & 1_2 & 0_2 & & 0_2 \\ 0_2 & 1_2 & 0_2 & 0_2 & 1_2 & & 0_2 \\ 0_2 & 0_2 & 1_2 & 0_2 & 0_2 & & 0_2 \\ 1_2 & 0_2 & 0_2 & 0_2 & 0_2 & & 0_2 \\ 0_2 & 1_2 & 0_2 & 0_2 & 0_2 & & 0_2 \\ 0_2 & 0_2 & 1_2 & 0_2 & 0_2 & & 0_2 \\ 1_2 & 0_2 & 0_2 & 0_2 & 0_2 & & 0_2 \\ 0_2 & 1_2 & 0_2 & 0_2 & 0_2 & & 0_2 \\ 0_2 & 0_2 & 1_2 & 0_2 & 0_2 & & 0_2 \\ 1_2 & 0_2 & 0_2 & 0_2 & 0_2 & & 0_2 \\ 0_2 & 1_2 & 0_2 & 0_2 & 0_2 & & 0_2 \\ 0_2 & 0_2 & 1_2 & 0_2 & 0_2 & & 1_2 \end{bmatrix} \end{array} \quad u = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ (\tau b)_{11} \\ (\tau b)_{12} \\ (\tau b)_{13} \\ (\tau b)_{21} \\ (\tau b)_{22} \\ \vdots \\ (\tau b)_{43} \end{bmatrix} \quad e = \begin{bmatrix} e_{111} \\ e_{112} \\ e_{113} \\ e_{121} \\ \vdots \\ e_{432} \end{bmatrix}$$

(24×15) (15×1) (24×1)

Generating Realizations from a Distribution

1. Process:

    1. $U \sim \text{Uniform}(0,1)$. Randomly generate a value of U, such as .38, .75, .62, etc.
    2. Derive (or look up) the quantile fctn corresponding to the dist you're generating from.
    3. Plug in the U value into the quantile fctn to get Y = realization.
        - Recall: $Q = F^{-1}$, ie $Q(u) = F^{-1}(u)$, F = cdf.

2. Examples: Let $U = .38$ be a realization from a $U(0,1)$ dist.

a) Weibull ($\alpha=4, \alpha=1.5$): $F(y) = 1 - e^{-(y/1.5)^4}$ $\rightarrow$ Find $Q(u)$: $u = 1 - e^{-(y/1.5)^4}$ $\rightarrow \ln(1-u) = -(y/1.5)^4 \rightarrow$

$y = 1.5\ln\left(\frac{1}{1-u}\right)^4$. Then plug in $u = .38$: $Q(.38) = 1.5\left(\frac{1}{1-.38}\right)^4 = \underline{1.247}$.

b) NegBinom ($r=8$, $p=.7$): Is a discrete fctn, so can't just take inverse as easily. Want to find

a value of y so $F(y-1) < .38 \le F(y)$. $F(y) = \sum_{i=1}^{k} \binom{i+r-1}{k}(1-p)^r p^i$, cdf is pnbinom (y, r, p) in R.

Recall Neg Binom is modeling number of failures, before r^{th} success.

Use R to generate vals, try $y = c(0,1,2,3)$: $F(y) = \begin{cases} .0576 & \text{for } y=0 \\ .196 & \text{for } y=1 \\ .383 & \text{for } y=2 \\ .569 & \text{for } y=3 \end{cases}$

So $\underline{y=3}$ is our realization.

c) Uniform ($a=.3$, $b=2.5$): $F(y) = \frac{y-a}{b-a}$ for $y \in [a,b)$. $\rightarrow$ Find $Q(u)$: $u = \frac{y-a}{b-a}$ $\rightarrow$

$y = a + u(b-a)$. Then plug in $u = .38$: $Q(.38) = .3 + .38(2.5 - .3) = \underline{1.136}$.

Hypothesis Testing - Means

1. **Testing Equality of a Group of Trt Means:**
   - F-test, see page 15. $H_0: \mu_1 = \mu_2 = \ldots = \mu_t$.
   - $F = MS_{sv}/MS_{sv}$ under HO, df1 = df MSsv, df2 = dfmssv under Ho.
   - Theory: If Ho: all means equal holds, then F ratio should be 1. $F_{\alpha, df1, df2}$ is critical value, and if $F > F_\alpha$, ratio is too far from 1 and so reject Ho.

2. **Single Normal Pop, σ^2 known:** (not very frequent IRL)
   - Testing if mean μ if >, < or = to a specific value, μ_0.
   - Test Statistic: $Z = \frac{(\hat{\mu} - \mu_0)}{SE(\hat{\mu})} = \frac{(\bar{Y} - \mu_0)}{\sigma/\sqrt{n}}$, b/c $SE(\hat{\mu})$ is known since σ known. $SE = \sqrt{\frac{\sigma^2}{n}} = \sigma/\sqrt{n}$.
   - $Z \sim N\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}}, 1\right)$ so val depends on μ.
   - Critical vals: $\mu_0 \pm Z_\alpha SE(\hat{\mu}) = \mu_0 \overset{(\text{or})}{\pm} Z_\alpha \frac{\sigma}{\sqrt{n}}$. (So $100(1-\alpha)\%$ CI $= \mu_0 \pm Z_{\frac{\alpha}{2}} \sigma/\sqrt{n}$)
   - Reject $H_0: \mu \leq \mu_0$ if $\hat{\mu} > \mu_0 + Z_\alpha \sigma/\sqrt{n}$, ie if $Z \geq Z_\alpha$.
   - Reject $H_0: \mu \geq \mu_0$ if $\hat{\mu} < \mu_0 - Z_\alpha \sigma/\sqrt{n}$, ie if $Z \leq -Z_\alpha$.
   - Reject $H_0: \mu \neq \mu_0$ if $\hat{\mu}$ outside $\mu_0 \pm Z_{\alpha/2} \sigma/\sqrt{n}$. (use $\alpha/2$ b/c splitting prob in 2 tails.) ie if $|Z| \geq Z_{\alpha/2}$.
   - Sample size: $n = \left[\frac{\sigma(Z_\alpha + Z_\beta)}{\delta}\right]^2$, δ = deviation from μ_0 to detect, want pwr $\gamma \geq 1-\beta$.

3. **Single Normal Pop, σ^2 unknown:** (t-test) - Also see pg 39.
   - Testing if mean μ >, < or = to a specific value, μ_0.
   - Test Statistic: $T = \frac{(\hat{\mu} - \mu_0)}{\hat{SE}(\hat{\mu})} = \frac{(\hat{\mu} - \mu_0)}{\hat{\sigma}/\sqrt{n}}$, using estimated SE (ie $\hat{SE}$) b/c σ^2 unknown.
   - $T \sim t$ dist, df = n-1. (ncp = 0 b/c ncp $= \frac{\hat{\mu} - \mu_0}{\hat{\sigma}/\sqrt{n}}$ under Ho, which = 0 → central t dist.)
   - 2-sided: Reject $H_0: \mu = \mu_0$ if $|T| > t_{\alpha/2, df}$, ie if $\hat{\mu}$ outside: CI $\hat{\mu}_0 \pm t_{\alpha/2, df} \hat{SE}(\hat{\mu})$.
     - $100(1-\alpha)\%$ CI is $\hat{\mu} \pm t_{\alpha/2, df} \hat{SE}(\hat{\mu}) = \hat{\mu} \pm t_{\alpha/2} \hat{\sigma}/\sqrt{n}$
   - 1-sided: Reject $H_0: \mu > \mu_0$ if $T < -t_\alpha$, ie $\hat{\mu} < \mu_0 - t_{\alpha, df} \hat{SE}(\hat{\mu})$.
     Reject $H_0: \mu < \mu_0$ if $T > t_\alpha$, ie $\hat{\mu} > \mu_0 + t_{\alpha, df} \hat{SE}(\hat{\mu})$.

- Sample size: Use Table A11, Mason Gunst + Hess. Eqn iterative, b/c depends on df = n-1.
- Robustness: Heavy tails + skewness + pos corr all inflate Type II error rate, and decrease power.

HYPOTHESIS TESTING - MEANS, CTD.

3. Single Non-Normal Pop:

a) For large n, can approximate using t-test.

- $T = \frac{(\bar{X} - \mu_0)}{\hat{\sigma}/\sqrt{n}}$ converges quickly to normal dist.

· How large n needs to be depends on skewness or heavy-tailed-ness of dist.

b) For small/medium n, use test for pop. median $\tilde{\mu}$ instead of mean.

- Transformations not reccommended.
- Sign Test: Very basic $\tilde{\mu}$ test, not exact b/c based on (discrete) binom dist.
- Wilcoxon Signed Rank Test:
  - Tests if median $\tilde{\mu}$ is $>$, $<$ or $=$ a specified value $\tilde{\mu}_0$.
  - $Y_i = X_i - \tilde{\mu}_0$. Discard $Y_i = 0$'s, let $n^* = \#$ nonzero Y_i's.
  - Rank $|Y_i|$ from sm to lg. If ties, assign avg. of ranks to ea member of tied grp.
  - $\Sigma\, Y_i\text{'s} > 0 = W+$, and $\Sigma\, Y_i\text{'s} < 0 = W-$.
  - Keuhl's Table A10 for $W_{n,\alpha}$ critical values
  - Reject $H_0: \tilde{\mu} \geq \tilde{\mu}_0$ if $W+ \leq W_{n,1-\alpha}$.
  - Reject $H_0: \tilde{\mu} \leq \tilde{\mu}_0$ if $W+ \geq W_{n,\alpha}$.
  - Reject $H_0: \tilde{\mu} = \tilde{\mu}_0$ if $W_{max} \geq W_{n,\alpha/2}$. $W_{max} = \max(W+, W-)$.

4. 2-Sample Test - Diff in 2 Pop Means: (Normal)

- t-tests can be used in this case also
- Called "Student's t-test" if pop vars assumed to be equal.
- Called "Welch's" or "Welch/Satherthwaite t-test" if no assumption of equal vars.
  - This case also called "unpaired" or "indep. samples" t-test.
- "Paired" or "Repeated Measures" t-test for testing diff btwn means taken on same EU, such as before/after measurements. Accounts for 2 pops not being indep.
- Paired t-test also called "Correlated Groups t-test".
  - Can be used in 2 situations:
    1) 2 measurements on same EU (ie before & after).
    2) 2 separate groups of EU's, but matched pairwise (ie X & Y both have 18 yo, 36 yo EUs)

5. 2-Sample Test - Diff in 2 Non-Normal Pop Means:

- Wilcoxon Rank Sum Test, or its large-sample normal approximation.
- See pg 40.

HYPOTHESIS TESTING - t-Tests (for means)

- One-sample t-test, does mean $=, <, >$ a specified μ_0 value? (See pg 37).
- Two-sample t-tests for diff in two pop. means, is diff $=, >, <$ specified θ_0 val? (See pg 38).

- **Test Statistic Form**: $T = \dfrac{\hat{\beta} - \beta_0}{\hat{SE}(\hat{\beta})}$ (point est $\hat{\mu}$ or $\hat{\theta} = \hat{\mu}_x - \hat{\mu}_y$ − specified value) (std error of point estimator)

  - T is ratio of diff in param from nominal value & std. error of param.
  - $T \sim t$ dist (central t-dist under H_0), with $df = n-k$ (n = # obs, k = # params estimated)

- **Assumptions**: Data from X & Y are normal and indep.

- **Mechanics of t-test**:
  - Reject $H_0: \beta = \beta_0$ if $|T| > t_{\alpha/2, df}$
  - Reject $H_0: \beta < \beta_0$ if $T > t_{\alpha, df}$
  - Reject $H_0: \beta > \beta_0$ if $|T| < -t_{\alpha, df}$

  CI: $\hat{\beta} \pm t_{\alpha/2, df}\, \hat{SE}(\hat{\beta})$

1. **One-Sample**: $T = \dfrac{\hat{\mu} - \mu_0}{\hat{SE}(\hat{\mu})} \Rightarrow \hat{SE}(\hat{\mu}) = \hat{\sigma}/\sqrt{n}$, $df = n-1$.

2. **2-Sample, Equal Vars**: $T = \dfrac{(\hat{\mu}_x - \hat{\mu}_y) - \theta_0}{\hat{SE}(\hat{\mu}_x - \hat{\mu}_y)} \Rightarrow \hat{SE}(\hat{\mu}_x - \hat{\mu}_y) = \sqrt{\dfrac{\hat{\sigma}_p^2}{n} + \dfrac{\hat{\sigma}_p^2}{m}}$, $\hat{\sigma}_p^2 = \dfrac{(n-1)\hat{\sigma}_x^2 + (m-1)\hat{\sigma}_y^2}{(n-1) + (m-1)}$
   - $df = (n-1) + (m-1)$
   - B/c assuming eq. vars, $\hat{\sigma}_p^2$ is "pooled var", ie est var based on WAVG of $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$.
   - Power fctn uses ncp, but is central t under H_0 (b/c under H_0, $ncp = 0$).

3. **2-Sample, Unequal Vars**: $T = \dfrac{(\hat{\mu}_x - \hat{\mu}_y) - \theta_0}{\hat{SE}(\hat{\mu}_x - \hat{\mu}_y)} \Rightarrow \hat{SE}(\hat{\mu}_x - \hat{\mu}_y) = \sqrt{\dfrac{\hat{\sigma}_x^2}{n} + \dfrac{\hat{\sigma}_y^2}{m}}$, $df = \dfrac{(c+1)^2 (n-1)(m-1)}{c^2(m-1) + (n-1)}$, $c = \dfrac{\hat{\sigma}_x^2/n}{\hat{\sigma}_y^2/m}$.
   - Called Welch's (or Welch-Satterthwaite) t-test. Is approx. test.

4. **2-Sample Paired**: $T = \dfrac{(\hat{\mu}_x - \hat{\mu}_y) - \theta_0}{\hat{SE}(\hat{\mu}_x - \hat{\mu}_y)} \Rightarrow \hat{SE}(\hat{\mu}_x - \hat{\mu}_y) = \sqrt{\dfrac{\hat{\sigma}_x^2 + \hat{\sigma}_y^2 - 2\sigma_x\sigma_y\rho}{n}}$, ρ = est. corr btwn samples.
   - $df = n/2 - 1$ since half of samples depend on other half. (n = total # obs.)
   - For small n, get large std error. Alternate approach: Let $D_i = X_i - Y_i$. If D normal, do one-sample t-test. If D non-normal, do Wilcoxon Signed Rank Test.

HYPOTHESIS TESTING - MEANS, CTD. 2

5 ctd: 2-Sample Test - Diff in 2 Non-normal Pop. Means:

· Wilcoxon Rank Sum Test:

· X + Y non-normal, but must be members of the same loc-scale family, w/ same scale.

· Want to compare diff in loc params to specified value θ_0.

· Is a rank-transform procedure, identical to Mann-Whitney Test.

· Steps: 1. Order the $N = n + m$ obs from smallest to largest.

2. Replace obs with ranks. For ties, assign avg. of tied ranks to tied vals.

3. $W_1 = \Sigma$ of ranks for X_i's. $W_2 = \Sigma$ of ranks for Y_i's.

4. Use table A.11 of 641 book to find $W_{\alpha,n,m}$ crit vals. (Tamhane + Dunlop)

· Test: · Reject $H_0: \theta_X < \theta_Y$ if $W_1 \geq W_{\alpha,n,m}$. (X + Y can be switched in steps to test $\theta_Y < \theta_X$.

· Reject $H_0: \theta_X = \theta_Y$ if $W_1 \geq W_{\alpha/2,n,m}$ OR $W_1 \leq n(n+m-1) - W_{\alpha/2,n,m}$.

· Lg Sample Normal Approximation:

· Table works for up to $n = 12$ and $m = 12$.

· For n and m ≥ 10, can use normal approx. (wilcoxon converges quickly to N.)

· W_1 approx $\sim N(\mu_W, \sigma_W^2)$ with $\mu_W = \frac{n(n+m+1)}{2}$, $\sigma_W^2 = nm(n+m+1)/12$.

· Then $Z = \frac{W_1 - \mu_W}{\sigma_W} \sim N(0,1)$. Now use Z tables to set critical vals:

· $W_{\alpha,n,m} \approx \mu_W + Z_\alpha \sigma_W$

6. One-Sample Normal t-test w/ Correlated Data:

· Adjust $\hat{SE}(\hat{\mu})$ for corr, otherwise we underestimate SE, T too large, + CI narrow, + ↑ P[Type 1].

· If equicorrelated, $\hat{SE}(\hat{\mu}) = \sqrt{\frac{\hat{\sigma}^2}{n}[1 + (n-1)\rho]}$ with ρ = est. corr.

· If autocorr AR1, $\hat{SE}(\hat{\mu}) = \sqrt{\frac{\hat{\sigma}^2}{n}\left[\frac{1+\rho}{1-\rho}\right]}$

HYPOTHESIS TESTING - MEANS, CTD. 3

7. 2-Sample Normal, X & Y indep, but X_i's corr and Y_i's corr:
   - Just adjust both variances for corr, like prev. pg.
   - Equicorrelated: $\hat{SE}(\hat{\mu}_x - \hat{\mu}_y) = \sqrt{\frac{\hat{\sigma}_x^2}{n}(1+(n-1)\rho_x) + \frac{\hat{\sigma}_y^2}{m}(1+(n-1)\rho)}$
   - Autocorr AR1: $\hat{SE}(\hat{\mu}_x - \hat{\mu}_y) = \sqrt{\frac{\hat{\sigma}_x^2}{n}\left[\frac{1+\rho_x}{1-\rho_x}\right] + \frac{\hat{\sigma}_y^2}{m}\left[\frac{1+\rho_y}{1-\rho_y}\right]}$

(*) SAMPLE SIZES FOR 2-POPULATION NORMAL MEANS TESTS:

1. 2 Pops, Require $n=m$: $n = 2\left[\frac{\sigma(Z_\alpha + Z_\beta)}{\delta}\right]^2$
   - σ is estimate supplied by researcher.
   - δ is delta in pop. means → size α test has power $1-\beta$ whenever $|\mu^* - \mu_0| \geq \delta$.
   - β is P[Type II error]
   - For a 2-sided test, use $Z_{\alpha/2}$ instead of Z_α.

2. 2 Pops, $n \neq m$, still require sample sizes to be multiples ($kn=m$): $n = \frac{k+1}{k}\left[\frac{\sigma(Z_\alpha + Z_\beta)}{\delta}\right]^2$
   - σ, δ and β same as above.
   - For a 2-sided test, use $Z_{\alpha/2}$ instead of Z_α.

3. 2 Pops, Paired: Use $D_i = X_i - Y_i$, 1-pop formula: $n = \left[\frac{\sigma_D(Z_\alpha + Z_\beta)}{\delta}\right]^2$
   - δ and β same as above. σ_D for D_i, supplied guess from researcher.
   - For a 2-sided test, use $Z_{\alpha/2}$ instead of Z_α.

HYPOTHESIS TESTING - PROPORTIONS

- **Basics:** Y is an r.v. describing one of the following:
  - # of Type A units selected with replacement from a finite pop.
  - # of Type A units selected with or w/o replacement from an essentially infinite pop.
  - # of successes in n iid Bernoulli trials.
- $Y \sim$ Binomial (n,p) with point estimator for p is $\hat{p} = Y/n$.

1) Single Pop, small n: ("small n" if $\min[np_0, n(1-p_0)] < 5$)
   - Testing if p is $>$, $<$ or $=$ a specified value p_0.
   - Reject $H_0: p \le p_0$ if $Y > B_{\alpha,p_0}$ where B_{α,p_0} is upper α %tile of Binom (n, p_0).
   - Reject $H_0: p \ge p_0$ if $Y \le B_{1-\alpha,p_0}$ where $B_{1-\alpha,p_0}$ is lower α %tile of Binom (n,p_0).
   - Reject $H_0: p = p_0$ if $Y \le B_{\alpha/2,p_0}$ OR $Y > B_{1-\alpha/2,p_0}$.

2) Single Pop, Large n Approximation:
   - Test Statistic: $TS = \dfrac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ with $TS \approx N(0,1)$ as $n \to \infty$ under H_0.

   - Reject $H_0: p \le p_0$ if $TS > Z_\alpha$.
   - Reject $H_0: p \ge p_0$ if $TS \le -Z_\alpha$.
   - Reject $H_0: p = p_0$ if $|TS| \ge Z_{\alpha/2}$.

   - Note: TS uses a different SE than the Wald CI pivot (pg 3). Wald uses $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
   This can be used as an alternate test, but is very hard to get the power fctn.

- Single Pop Sample Size:
  - Find n so an α-sized test has power $\gamma(p^*) = 1-\beta$ whenever $|p^* - p_0| \ge \delta$.
  - $n = \left[\dfrac{Z_\alpha \sqrt{p_0(1-p_0)} + Z_\beta \sqrt{p^*(1-p^*)}}{\delta}\right]^2$ · For 2-sided test, use $Z_{\alpha/2}$ instead of Z_α.

HYPOTHESIS TESTING - PROPORTIONS, CTD.

3) Multiple Pops - Testing If All Equal:

- **Pearson Chi-Sq Test of Homogeneity of Proportions:**
  - Testing if all t proportions equal, like ANOVA for proportions. $H_0: p_1 = \ldots = p_t$, vs H_1: not all $=$.
  - Have t pops, with indep. random samples of $n_1 \ldots n_t$. All units Type A $(i=1)$ or Type 2 $(i=2)$.

- Test Statistic: $\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{t} [O_{ij} - \hat{E}_{ij}]^2 / \hat{E}_{ij}$ with $\chi^2 \sim \chi^2_{t-1}$ under H_0.

  - O_{ij} = Total # obs for Type i (A or B) in Pop. j.
  - $\hat{E}_{ij} = n_j R_i / N$ where R_i = total # A or B obs across all pops.

  (We use $\hat{E}_{ij}$ b/c under H_0, all pops have same proportion, p_0. We estimate p_0 as $\hat{p}_0 = R_1/N$, ie total # Type A obs / total # obs. Then $\hat{E}_{1j} = n_j \hat{p}_0$, and $\hat{E}_{2j} = n_j(1-\hat{p}_0)$.

- Test: Reject $H_0: p_1 = \ldots = p_t$ if $\chi^2 \geq \chi^2_{\alpha, t-1}$.
  - If reject H_0, will want multiple comparisons to analyze how p_i's are different.
  - Condition: If any $\hat{E}_{ij} < 1$ or over 20% of $\hat{E}_{ij}$'s < 5, this test is invalid. Use Fisher Exact Test instead. (Happens for small n.)

- **Pearson Chi-Sq Special Case: $t=2$:**
  - Must meet same conditions as above to be valid.

  - Test Statistic: $Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $\hat{p}$ = WAVG of $\hat{p}_1$ & $\hat{p}_2$: $\hat{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$. $Z \sim N(0,1)$ for H_0.

  - Test:
    - Reject $H_0: p_1 \leq p_2$ if $Z > Z_\alpha$. Pval $= 1 - \Phi(Z)$.
    - Reject $H_0: p_1 \geq p_2$ if $Z \leq -Z_\alpha$. Pval $= \Phi(Z)$.
    - Reject $H_0: p_1 = p_2$ if $|Z| \geq Z_{\alpha/2}$. Pval $= 2\Phi(Z)$.

    * $\Phi(Z)$ = pnorm(z,0,1) where pnorm is cdf eval @ Z.

- **Sample Size for 2 Pops (Limited to $n_1 = n_2$):**
  - Want an α-size test with P[Type II error] $= \beta$ for given vals p_1 & p_2, with $D = p_1 - p_2$.

  - $$n = \left[\frac{Z_\alpha \sqrt{\frac{(p_1+p_2)(2-p_1-p_2)}{2}} + Z_\beta \sqrt{p_1(1-p_1) + p_2(1-p_2)}}{D} \right]^2$$

HYPOTHESIS TESTING - PROPORTIONS, CTD 2.

4) Multiple comparisons of t Pops:

- If Pearson Chi-Sq Homog of Props test rejects, want to do further analysis to see the nature of the diffs for $p_1 \dots p_t$.
- Calc the 2-pop Z statistic (pg 43) for each pair of proportions $(p_i, p_{i'})$. There will be $\binom{t}{2}$ pairs.
- Compare the p-val for each pair test to $\alpha_{pc} = \alpha_0 / M$, $M = \#$ of tests, to check for sig. ev. of each pair being different.

5) Fisher Exact Test - 2 pops when Pearson conditions not met (n small):

- Test $H_0: p_1 \le p_2$, or $H_0: p_1 = p_2$ using Hypergeo dist.
- Think of 2-pop data as a 2×2 table - really a possible set of 2×2 tables, all with same fixed row/col totals. (We observe one of many possibilities.)

	A (Succ)	B (Fail)	Total
Pop 1	X	$n_1 - X$	n_1
Pop 2	Y	$n_2 - Y$	n_2
Total	m	$n - m$	n

- Under H_1, find which other tables have stronger support for H_1 than observed table.

- Test:
  - $H_0: p_1 \le p_2$, $\text{p-val} = 1 - \text{phyper}(x_0 - 1, n_1, n_2, m) = \displaystyle\sum_{k=x_0}^{\min(n_1,m)} \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n}{m}}$, x_0 = observed X val.
  - $H_0: p_1 = p_2$, for pval, sum all $k = x_0$'s where $P(k) \le P(x_0)$.
  - Use software to do all of this.

HYPOTHESIS TESTING - PROPORTIONS, 2×2 TABLES

- On pg 43-44, we looked at Pearson Chi-Sq for $t=1$, and Fisher Exact Test.
- There are several other ways to analyze 2 populations with 2 possible outcomes, ie a 2×2 table.

1) Odds Ratio:

- Compares the odds of 2 populations
- OR = [Odds for pop1 / Odds for pop 2] with Odds = $p_i/(1-p_i)$
- $\hat{OR}$ point estimator = $\dfrac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \dfrac{n_{11}n_{22}}{n_{12}n_{21}}$ (product of diagonals)
- Table setup:

	Outcome A	Outcome B	Total	Proportion of A's
pop1	n_{11}	n_{12}	$n_{1.}$	$\hat{p}_1 = n_{11}/n_{1.}$
pop2	n_{21}	n_{22}	$n_{2.}$	$\hat{p}_2 = n_{21}/n_{2.}$
Total	$n_{.1}$	$n_{.2}$	n	

- OR ranges from 0 to ∞.
- If OR=1, both pops are equally likely to have Outcome A.
- If OR<1, $p_1 < p_2$, so pop 1 is less likely than pop 2 to have Outcome A.
- If OR>1, $p_1 > p_2$, so pop1 is more likely than pop 2 to have Outcome A.

- There is an alternative expression for OR in logistic regression:
  - $\text{Logit}(p) = \text{Log}(\text{Odds})$.
  - $L = \log(OR) = \text{logit}(p_1) - \text{logit}(p_2)$.

2) Relative Risk:

- more meaningful than OR for prospective or retrospective studies.
- $RR = p_1/p_2$, but p_1 & p_2 are unknown.
- Point estimator $\hat{RR} = \hat{OR}\left[\dfrac{1+\frac{n_{21}}{n_{22}}}{1+\frac{n_{11}}{n_{12}}}\right]$

- OR & RR nearly equal if Outcome A is rare.
- RR also called "prevalence ratio" in cross-sectional data.

HYPOTHESIS TESTING - PROPORTIONS, 2×2 TABLES, CTD. 1

3) Sensitivity & Specificity:

- Used in evaluating effectiveness of screening tests.
- $t=2$ pops, with p_1 = condition present, p_2 = condition absent. There are 2 response values, test pos (+) or test neg (−).

	Test (+)	Test (−)	Total
Condition Present (pop1)	n_{11}	n_{12}	n_1
Condition Absent (pop2)	n_{21}	n_{22}	n_2

- Prevalence: Proportion of pop. with condition present
- Sensitivity: Prob of a correct (+) test. Estimated by $\frac{n_{11}}{n_1}$.
- Specificity: Prob of a correct (−) test. Estimated by $\frac{n_{22}}{n_2}$.

- The most important thing to know is, what is prob that condition present given (+) test?
- Bayes Theorem:
  - $P[\text{condition present} \mid k \text{ indep tests all } (+)] = \dfrac{(\text{sens})^k\,(\text{prev})}{(\text{sens})^k(\text{prev}) + (1-\text{spec})^k\,(1-\text{prev})}$
  - The tricky part is estimating prevalence.
  - Run k indep tests from same batch bc if trait is rare (prevalence small), most (+) tests are false (+)'s. So this improves performance of test.

4) McNemar's Test of Matched Pairs:

- Used for testing when data is from n pairs of correlated EU's, with 2 possibilities:
  1) Responses taken from same EU twice (before/after).
  2) One response each from pair of related EU's (hub/wife, L eye/R eye).

Table:

		EU response 2: Yes	No	Total
EU Response 1	Yes	n_{11}	n_{12}	$n_{1.}$
	No	n_{21}	n_{22}	$n_{2.}$
	Total	$n_{.1}$	$n_{.2}$	n (pairs)

- Let $m = n_{12} + n_{21}$
- Let $B \sim \text{Binom}(m, .5)$.

- Test: Is proportion of Yes pairs same for Responses 1 & 2. (Is $p_{.1} >, <, = p_{1.}$)
  - For H_0: $p_{1.} < p_{.1}$, pval $= P[B \geq n_{12}]$. Reject if pval $\leq \alpha$.
  - For H_0: $p_{1.} > p_{.1}$, pval $= P[B \leq n_{12}]$. Reject if pval $\leq \alpha$.
  - For H_0: $p_{1.} = p_{.1}$, pval $= 2 \cdot \min(P[B \geq n_{12}], P[B \leq n_{12}])$.

4 (td) McNemar's Test of matched Pairs:

- If $m > 15$, can use asymptotic result.
- Qmn Test Statistic $= \frac{(n_{12}-n_{21})^2}{(n_{12}+n_{21})}$, with $Qmn \approx \chi^2_1$ under $H_0 \rightarrow \sqrt{Qmn} \sim N(0,1)$ under H_0.
- $100(1-\alpha)\%$ CI for $(\hat{p}_{1.}-\hat{p}_{.1})$: $(\hat{p}_{1.}-\hat{p}_{.1}) \pm Z_{\alpha/2}\hat{SE}(\hat{p}_{1.}-\hat{p}_{.1})$, $\hat{SE} = \sqrt{\frac{\hat{p}_{1.}(1-\hat{p}_{1.})+\hat{p}_{.1}(1-\hat{p}_{.1})-2(\hat{p}_{11}\hat{p}_{22}-\hat{p}_{12}\hat{p}_{21})}{n}}$
  - The $\hat{SE}$ contains "$-2(\hat{p}_{11}\hat{p}_{22}-\hat{p}_{12}\hat{p}_{21})$ because data is correlated from pairs.

5) CMH Test + Breslow-Day Test:

- For K 2x2 tables, where data is blocked by some covariate Z, so there is one 2x2 tbl for each Z val.
- Want to test the indep of the pop (X) and response (Y) for each val of covariate Z, ie conditional on Z.
- CMH loses its power to detect over-all X-Y association when associations vary frm + to − across K tbs.
  - ↳ So we only do CMH if OR is same for all K tables, or the tables have a "common odds ratio".

- Step 1: Breslow-Day Test for Common Odds Ratio:
  - $H_0: OR_1 = \ldots = OR_K$ vs H_1: not all OR's equal.
  - Test statistic is complex. Obtain TS + pval from SAS.

- Step 2a: CMH Test: (Only if Breslow-Day Ho holds)
  - H_0: X + Y indep given val of Z, ie $p_{ijk} = p_{i.k}\,p_{.jk}$ for all $k = 1..Z$, ie $OR = 1$ for all 2x2 tables. ($H_1: OR \neq 1$)
  - Test Stat: $CMH = \frac{\left[\sum_{k=1}^{z}(n_{11k}-\mu_{11k})\right]^2}{\left[\sum_{k=1}^{z} Var(n_{11k})\right]}$ where $\mu_{11k} = \frac{n_{1.k}\,n_{.1k}}{n_{..k}}$ and $var(n_{11k}) = \frac{n_{1.k}\,n_{2.k}\,n_{.1k}\,n_{.2k}}{n^2_{..k}(n_{..k}-1)}$
    - $CMH \sim \chi^2_1$ under Ho, so use $\chi^2_{\alpha/2,1}$ critical vals, or test $\sqrt{CMH}$ against $Z_{\alpha/2}$ crit vals
    - If $|\sqrt{CMH}| > Z_{\alpha/2}$

*Simpson's Paradox: You can get contradicting results by analyzing data as a single 2x2 table vs. breaking data into K 2x2 tables based off of K vals of an explanatory variable.

Example: In smokers vs non-smokers mortality study, nonsmokers had higher death rates in single 2x2 table. But when blocked by age grp into K 2x2 tables, smokers had higher death rates. (Younger women ↓ likely to die during study, but ↑ likely to smoke.)

This is why it is so important to carefully consider confounding variables.

HYPOTHESIS TESTING - STANDARD DEVIATIONS

1) **Single Normal Pop:** (Assumption of normality = CRITICAL, test not robust, no CLT for lg n)

- Testing if std dev is >, <, = a specific value, σ_0. if nonnormal, use bootstrap.
- Test statistic $TS = \frac{(n-1)\hat{\sigma}^2}{\sigma_0^2}$ where $TS \sim \chi^2_{n-1}$ under H_0.
- Test:
  - For $H_0: \sigma \le \sigma_0$, reject H_0 if $TS \ge \chi^2_{n-1,\alpha}$.
  - For $H_0: \sigma \ge \sigma_0$, reject H_0 if $TS \le \chi^2_{n-1,1-\alpha}$.
  - For $H_0: \sigma = \sigma_0$, reject H_0 if $TS \ge \chi^2_{n-1,\alpha/2}$ or $TS \le \chi^2_{n-1,1-\alpha/2}$.

2) **Multiple Normal Pops - Hartley's Test:**

- Testing if $\sigma_1 = \ldots = \sigma_t$ for t pops.
- Data must be normal with equal sample sizes ($n_1 = \ldots = n_t$)
- Test stat $F_{max} = \frac{\hat{\sigma}^2_{max}}{\hat{\sigma}^2_{min}}$ where $\hat{\sigma}^2_{max}$ & $\hat{\sigma}^2_{min}$ are min & max $\hat{\sigma}_i^2$'s.
- $H_0: \sigma_1^2 = \ldots = \sigma_t^2$, all vars equal.
- Reject H_0 if $F_{max} > F_{a=\alpha, t, df2=n-1}$.
  - This is not a normal F crit val. See Hartley Fmax table; is based on Fmax sampling dist.

3) **Multiple Pops - BFL:** (Normality & eq. sample sizes not required.)

- Used to test equality of variances for ANOVA.
- Testing if $\sigma_1 = \ldots = \sigma_t$
- Test stat L is complex, obtain frm SAS. $\left[z_{ij} = |y_{ij} - \tilde{y}_{i}|, \; L = \frac{\sum_{i=1}^{t} n_i (\bar{z}_{i.} - \bar{z}_{..})^2/(t-1)}{\sum_{i=1}^{t}\sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{i.})^2/(N-t)} \right]$
- Reject H_0 if $L \ge F_{\alpha, t-1, N-t}$ where $N = \sum_{i=1}^{t} n_i$. Or get pval from SAS.

HYPOTHESIS TESTING - AUTOCORRELATION

1) AR(1) with Normal Residuals - Von Neumann Test:

- Testing if data is first-order autocorrelated.
- Test statistic $Q = \dfrac{\frac{1}{n-1}\sum_{t=2}^{n}(X_t - X_{t-1})^2}{\frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2}$ with large Q = neg corr, small Q = pos corr.

- Testing $H_0: \rho = 0$ vs two possible alternatives: $H_1: \rho > 0$ and $H_1: \rho < 0$.
- Test:
  - For $H_0: \rho = 0$ vs $H_1: \rho > 0$, reject H_0 if $Q < Q_{P,\alpha}$.
  - For $H_0: \rho = 0$ vs $H_1: \rho < 0$, reject H_0 if $Q > Q_{N,\alpha}$.

  - If $n \le 60$, use table to find crit vals $Q_{P,\alpha}$ and $Q_{N,\alpha}$.
  - If $n > 60$, $Q_{P,\alpha} \approx \frac{2n}{n-1} - Z_\alpha \frac{2}{\sqrt{n}}$ and $Q_{N,\alpha} \approx \frac{2n}{n-1} + Z_\alpha \frac{2}{\sqrt{n}}$

2) Distribution-Free Runs Test:

- Testing if data is correlated.
- $X_1 \ldots X_t$ must be equally spaced obs.
- Test:
  1. Center obs using $Y_t = X_t - \bar{X}$.
  2. Count R, the # of runs (sequences of all pos or all neg Y_i values).
  3. Count # of pos Y_i's (n_1) and neg Y_i's (n_2).
  4. Data is corr if $R \le R_L$ or $R \ge R_U$. Get R_L & R_U from table.
  5. If n_1 and $n_2 > 20$, use asymptotic result:
     - Data correlated if $Z > Z_{\alpha/2}$.
     - $Z = \dfrac{|R-\mu| - .5}{\sigma}$ with $\mu = 1 + \dfrac{2n_1n_2}{n_1+n_2}$ and $\sigma^2 = \dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$

DISTRIBUTIONS - DISCRETE

1) Binomial:

- # of successes in n iid indep Bernoulli trials, with equal prob of success (p) for all trials. # of Type A units selected with replacement.
- Possible values: 1... n; each trial has possible outcomes 1 or 0.
- + skewed if $p<.5$, - (L) skewed if $p>.5$.

2) Neg Binomial:

- Hold # of successes (r) at a fixed value, & estimate # of trials needed to obtain r successes.
- Possible values: r, r+1, r+2...

3) Poisson:

- Estimates # of events occurring during a specific time period or space.
- The key parameter λ is the avg # of events occurring in the time or space.
- Possible values: 0, 1, 2...
- Similar graph as binomial, but R skewed. Skewness decreases as interval widens.
- If $X \sim Pois(\lambda)$ on 1 unit interval, $Y \sim Pois(k\lambda)$ on k unit intervals.

4) Geometric:

- Estimates # of trials until first success.
- Special case of neg binom, with r=1.

5) Hypergeometric:

- Estimates # of Type A units in a pop containing only A & B units, where sampling without replacement is used.
- Possible values: integers btwn max$[0, M-(N-n)]$ and min$[n, M]$.
    - M= pop Type A units, N= total pop units, n= total sample units.

6) Bernoulli:

- A single random trial with possible outcomes 1 (success, event) & 0 (fail, no event).
- Prob of success is p, prob of failure is 1-p.

7) Uniform Discrete:

- n finite outcomes, all equally likely to be observed, with $p = 1/n$.
- Example is throwing a single die, with 6 equally likely outcomes.

DISTRIBUTIONS - CONTINUOUS

1) Uniform - Continuous:

- modelling events which are all equally likely.
- Probability of the events changes based on length of interval.

2) Normal:

- Used to model symmetric RVs with med-weight tails + few extreme values.
- Used in many stats applications bc of Central Limit Theorem.

3) Lognormal:

- Used in reliability modeling for time to failure of a device
- Used in modeling growth of plants, tumors + living tissue
- Right skewed
- If $X \sim$ Normal, then $Y = e^X$ is lognormal.

4) Beta:

- Used to model behavior of RVs limited to finite intervals.
- If we want # of successes random w/ prob of each success fixed, use binomial. If we want to think of # of successes as observed w/ prob of each success random/unknown, use Beta.
- Example: Let p = proportion of defectives coming from a plant, where a company has many plants. Model p using Beta.

5) Exponential:

- Used to model the time btwn 2 event occurrences in a Poisson process. ($T \sim Exp(1/\lambda)$)
- Used in reliability/survival analysis to model time to failure with a constant failure rate.
- Key property is memorylessness, ie constant failure rate.

6) Double Exponential ("Laplace"):

- Symmetric with sharper peak + heavier tails than normal, but lighter than Cauchy.
- Two back-to-back exponential dists.

7) Weibull:

- Used to model time to failure when the failure rate is proportional to a power of time.
- Used to model extreme obs (min or max) in a pop. (Special case of Extreme Value Dist class.)
- Is a generalization of the exponential dist.

DISTRIBUTIONS - CONTINUOUS, CTD.

8) Gamma:

- A 2-parameter family of dists with many applications.
- Parameters: α=shape, $\beta = 1/\theta$ = rate parameter = inverse scale parameter.
- If T is time btwn k events in a Pois(λ) process, $T \sim Gamma(\alpha = k, \beta = 1/\lambda)$.
- The Exponential (β) dist is Gamma($\alpha=1, \beta$).
- The Chi-sq(df=v) dist is Gamma($\alpha = v/2, \beta = 2$).

9) ChiSquared:

- R skewed, becomes more normal as df=v becomes large. (df is a shape parameter)
- Used in many tests, such as Goodness of Fit tests and hypothesis tests for variances.
- Special case of Gamma & Weibull.
- Sum of squared indep normal RVs.

10) t Dist:

- Used to model pops in which extreme values occur more frequently than in normal.
- Symmetric with heavier than normal tails, converges to normal as $df = v \to \infty$.
- Parameter df=v is a shape parameter.

11) F Dist (Fisher):

- Often used as a critical value in ANOVA & Regression testing.
- R skewed, related to Chisq.
- If $T \sim t(v)$, then $T^2 \sim F(df1=1, df2=v)$.

12) Cauchy:

- Symmetric dist with such heavy tails that mean & var do not exist.
- Used when pop has very large vals compared to the point of symmetry (finance).
- If $T \sim t(v=1)$, then $T \sim$ standard Cauchy with $\theta_1 = 0, \theta_2 = 1$.

13) Logistic:

- Symmetric with tail weights btwn double exponential & Cauchy.

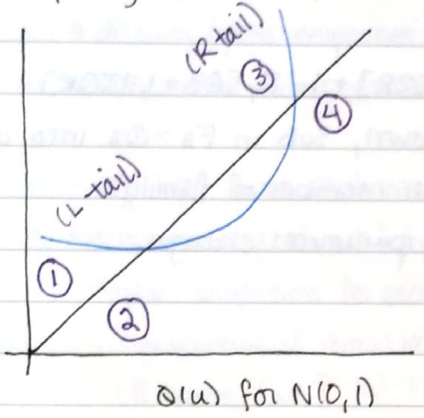
14) Normal Mixture Dist:

- May be bimodal, with multiple peaks.
- If two pops have same mean & diff vars, will have one peak that is sharper than normal, with heavier tails than normal.

GOODNESS OF FIT TESTING - GRAPHICAL

1) Quantile Ref Dist Plots:

- Plot Q_0 on x-axis vs $\hat{Q}$ on y-axis, where Q_0 is quantile fctn for std. member of family for the dist you are checking to see if data fits.
- If plot is close to a 45° line ($y=x$), data is good fit for chosen dist.
- For Weibull, remember that $Q_Z(u) = \log(-\log(1-u))$ on x-axis, $\log(y_i)$ on y-axis.
- Example, comparing χ^2_4 vs $N(0,1)$:



- If χ^2_4 data followed 45° line, normal would be good fit.
- Not symmetric - would look like S if symmetric.

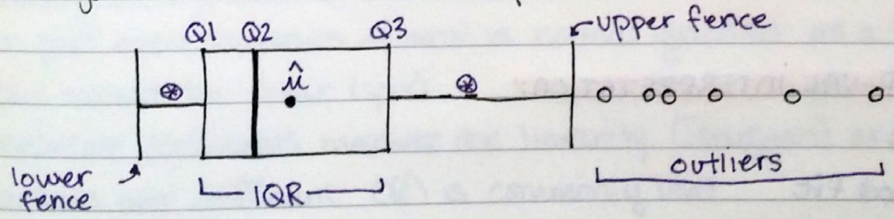
① If tail here, L tail shorter than normal*. (or "lighter")
② If tail here, L tail longer than normal*. (or "heavier")
③ If tail here, R tail longer than normal*.
④ If tail here, R tail shorter than normal*.
*or whatever ref dist is on x-axis.

2) Q-Q Plots:

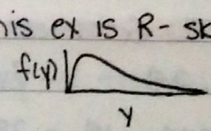
- Plot the quantile fctns of 2 data sets ($\hat{Q}$'s) against ea other to see if data from same dist.
- A graphical method of comparing 2 data sets.
- If data from same dist, points will fall on/close to line $y=x$.
- If points fall on a line other than $y=x$, 2 dists are linearly related.
- If points fall on a non-linear line, dists related by some fctn $Y=h(X)$.
- Use plot to compare shape, skew, loc, scale of 2 dists.
    - Ex: If plot arc or S-shaped, one dist is more skewed than the other.
- If points not close to straight line, could still be frm same dist but with different shape params.

3) Box Plots:

- Used to get basic info about the shape of a dist.



(This ex is R-skewed)



- $\hat{Q}1 = \hat{Q}(.25)$, $\hat{Q}2 = \hat{Q}(.5) = \hat{\mu}$, $\hat{Q}3 = \hat{Q}(.75)$
- $IQR = \hat{Q}3 - \hat{Q}1$
- Lower Fence $= \hat{Q}1 - 1.5\,IQR$
- Upper Fence $= \hat{Q}3 + 1.5\,IQR$

GOODNESS OF FIT TESTING - GRAPHICAL, CTD.

3) Box Plots ctd:

- Outlier Classification:
  - Obs are outliers if $Y <$ Lower Fence or $Y >$ Upper Fence.
  - Obs are extreme outliers if $Y < \hat{Q}1 - 3IQR$ or $Y > \hat{Q}3 + 3IQR$.
  - For loc-scale families Prob(Outlier) same for all members, regardless of loc or scale.

- Probability of an Outlier:
  - If cdf + all params known, $P[\text{Outlier}] = F_Y[Q1 - 1.5IQR] + 1 - F_Y[Q3 + 1.5IQR]$
  - If loc-scale family known, but parameters unknown, sub in F_Z + Q_Z into above formula. F_Z + Q_Z are cdf + quantile fctn for std. member of family.
  - Only works for loc-scale. P(outlier) changes if shape parameter changes.

- Expected # of Outliers: $E_n = nP[\text{Outlier}]$

GUIDE TO GOODNESS OF FIT P-VAL INTERPRETATION:

- pval $\geq .25$ - Excellent fit
- pval btwn $.15 - .25$ - Good fit
- pval btwn $.05 - .15$ - Moderately good fit
- pval btwn $.01 - .05$ - Poor fit
- pval $< .01$ - Unacceptable fit.

*on TI89: binomcdf(n,p,k) = dbinom(k,n,p)

GOODNESS OF FIT TESTS

1) Chi-Sq GOF Test (Discrete Dists):

* Must have all $E_i > 1$, & at most 20% $E_i < 5$. Can merge bins, but affects df of test.

- Tests if data is a good fit for the selected discrete dist.
- H_0: Data is a good fit.
- Test Statistic $Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ with $Q \approx \chi^2$, df = k-1-w under H_0.
  - k = # of bins / poss. responses (Ex: # diseased kids in family can be 1, 2, 3, ≥4 → k=4)
  - w = # of params being estimated (Ex: Testing fit to Poisson(3) → w=0, Pois(λ) → w=1)
  - O_i = observed count (# of events) in each of the k bins. Observed frequency.
  - $E_i = np_i$ (or $n\hat{p}_i$ if not fully specified). Expected/theoretical frequency.
    - ↳ p_i = proportion for each y_i if known. If $\hat{p}_i$, $\hat{p}_i = Y_i/n$ for each bin, ie proportion of total obs falling in each bin.
  - If known, $E_i = \frac{F(Y_u) - F(Y_L)}{N}$ where $F(Y_u)$ & $F(Y_L)$ = cdf eval at min & max Y_i in each of the k bins.

p_i = dbinom(k,n,p) with p = overall known or est prob

If combine bins, add 2 vals together, ie dbinom(4,n,p) + dbinom(5,n,p)

- Test: Reject H_0 if $Q \geq \chi^2_{\alpha, k-1-w}$. (Large Q → O_i's & E_i's do not match.)

2) Shapiro-Wilks (SW) GOF to Normal:

- Most powerful test for normality. Use this any time testing if data normal.
- Do not need to specify μ or σ^2.
- H_0: data is normal.
- Use SAS to get Test Statistic & p-val.
- For large n, test is sensitive to many small deviations from normality. Always pair with Normal Ref Dist Plot to cross-check results.

3) Normality Testing w/ Correlation Coefficients:

- Can plot order statistics of data vs normal quantiles as a Ref Dist Plot. If data normal, plot should be linear (y=x).
- Correlation coefficients measure the linearity (Pearson's) or monotonicity of the relationship.
- Pearson's corr coefficient (R) is commonly used.
- If sample corr exceeds a critical val, data normal. Special tables for crit vals needed.
- Not as powerful as SW for normality, & Pearsons is sensitive to outliers.

GOODNESS OF FIT TESTS, CTD.

3) Anderson-Darling (AD) Test:

- For testing GOF to dists other than normal. (Normal works, but SW is better.)
- Need all parameters known.
    - If params unknown, Exponential & Weibull have Modified AD tables.
- H_0: Data fits chosen dist.
- Get Test Statistic & p-val from SAS or R.
- More sensitive than Cramer von Mises (CvM), which nobody uses now.

4) Kolmogorov-Smirnov (K-S) Test:

- Also for testing GOF to dists other than normal. (Normal works, but SW is better.)
- Has less power than AD, but is nice bc AD requires different sets of crit. val. tables for each dist. KS only needs one crit value table, which works for any dist.
- Works best when data deviates from dist in a global fashion, near center of dist, b/c is less sensitive to deviations in tails.
- If some parameters unknown, there are modifications to KS for exponential & weibull.

5) Cramer von Mises Test:

- Like AD, but unable to detect differences in tails.
- Nobody uses anymore - just be aware that it exists.

SURVIVAL/RELIABILITY ANALYSIS

- Examining time to occurrence of an event.
- Examples: death, device failure, cancer-free test.

1) Survival Fctn:
   - T is an RV for time at which event occurs, with cdf $F(t)$ and pdf $f(t)$.
   - $S(t) = 1 - F(t)$
   - $S(t)$ describes prob[event occurs after time t]. (ie prob device works longer than time t.)

2) Hazard Fctn:
   - $h(t) = f(t)/S(t)$
   - Prob [event occurring at time t given event hasn't occurred at time t]
   - Instantaneous failure rate.
   - Often reported as # of failures per unit of time.

3) Cumulative Hazard Fctn:
   - $H(t) = \int_0^t h(\tau)\, d\tau$
   - Integral of the Hazard Fctn.
   - The accumulated hazard over time

BASIC PROPERTIES

1) Expected Value (Mean):

- $E(X)=\int_{-\infty}^{\infty} x f(x)\,dx$ where $f(x)$ is the pdf of X.
- The mean, and 1^{st} moment.
- Expected val of a fctn of X: $E[g(X)]=\int_{-\infty}^{\infty} g(x) f(x)\,dx$
- Linear combo of Expected Vals: $Y=aX+b$. $E[Y]=aE(X)+b$.
- Product of indep RVs: $E[f(x)g(x)] = E[f(x)]\cdot E[g(x)]$.

2) Variance:

- $Var(X)=\int_{-\infty}^{\infty} x^2 f(x)\,dx$

* If X & Y not indep, $Var(X-Y+c) = Var(X-Y)+0$ $= Var(X)+Var(Y)-2Cov(X,Y)$

- Second moment.
- $Var(X)=E[(X-\mu_X)^2] = E(X^2)-E(X)^2$
- $Var(aX+b)= a^2 Var(X)+0$
  - $Var(\text{constant})=0$; shifting a dist does not change its var.
- Only if $X_1 \dots X_n$ indep: $Var(X_1+\dots+X_n)=\Sigma Var(X_i)$, var of sums = sum of vars.

3) Standard Deviation:

- $sd(X)=\sqrt{Var(X)}$
- In same units as X, whereas $Var(X)$ is in units^2.

4) Covariance:

- Measures how two vars change together, ie measures the tendency of X & Y to be on same or opposite sides of their respective means.
- $Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)] = E[XY]-E(X)E(Y)$
- If X & Y indep, $Cov(X,Y)=0$. Reverse does not hold; $Cov(X,Y)=0 \not\rightarrow X,Y$ indep.
- If $Y=aX+b$, $Cov(X,Y)=1$ if $a>0$, or -1 if $a<0$.

5) Correlation:

- Scaled version of Covariance.
- $\rho(X,Y)=\dfrac{Cov(X,Y)}{\sigma_X\sigma_Y}$
- $\rho(X,Y)$ values can range from -1 to 1.

BASIC PROPERTIES, CTD

6) Moment-Generating Fctn:

- $M_x(s) = E[e^{sx}]$
- Then generate moment k: $E[X^k] = \left. \frac{\partial^k M_x(s)}{\partial s^k} \right|_{s=0}$, ie $E(X^2) = M''_x(0)$.
- Is sometimes easier to calculate moments using the mgf.

7) Chebyshev's Inequality:

- Was used to develop WLLN, which says as $n \to \infty$, $\hat{\mu}_x \to \mu_x$. (Weak Law of Lg #'s)
- Used to determine how much of a dist falls so many sdevs outside mean.
- $P(|X - \mu_x| \geq a) \leq \frac{\sigma_x^2}{a^2}$. To find what % of a dist falls outside y sdevs of mean, let $a = y\sigma_x$.

8) Bias:

- Bias of an estimator is difference btwn expected value of parameter & true value.
- $\text{Bias}(\theta) = E(\theta) - \theta$

9) Mean Squared Error (MSE):

- Used to evaluate how well parameter estimates its true value, ie how $\hat{\theta}$ estimates θ.
- If choosing btwn $\hat{\theta}_1$ & $\hat{\theta}_2$ to estimate θ, choose the one with smallest MSE.
- $MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2(\hat{\theta})$
- Measures how close, on avg, is estimator to true value of parameter.

10) Standard Error: (SE):

- The std dev of the sampling dist of a parameter or statistic.
- See pg 18-19.
- Is a measure of the accuracy/reliability of the parameter (ie $\hat{\theta}$, $\bar{X}$, etc.)
- Difference from std dev → SD tells us about shape of our dist, ie how close individual data values are from mean. SE tells us how close our sample mean is to the true pop. mean (or some other parameter).

BASIC PROPERTIES, CTD 2.

11) Calculating an MLE by Hand:

- Find Likelihood fctn, $L(\theta|s) = \prod_{i=1}^{n}$ pdf
- Take $\log(L(\theta|s))$ to get log-likelihood fctn.
- Take 1st deriv of $\ell(\theta)$ wrt θ. Set = to 0, solve for θ.
- Take 2nd deriv of $\ell(\theta)$ and verify $\ell''(\theta) < 0$, so we have found a max (not a min).

- Invariance Property of MLEs: If $\hat{\theta}$ is MLE of θ, then $g(\hat{\theta})$ is MLE of $g(\theta)$.

12) How to use the Φ Fctn to Represent $N(0,1)$ Probs:

- $\Phi(x_0)$ indicates the cdf of the $N(0,1)$ dist, evaluated at x: $P(X < x_0)$ for $X \sim N(0,1)$.
- Ex 1: $P(\bar{X} \leq x_0) = P\left[Z \leq \frac{x_0 - \mu_0}{\sigma/\sqrt{n}}\right] = \Phi\left[\frac{x_0 - \mu_0}{\sigma/\sqrt{n}}\right]$

13) Independence check:

- To check if X & Y indep, verify $E(XY) = E(X)E(Y)$

14) Cov of a Linear Combo (Bilinearity):

Form of WEIBULL PARAMETERIZATION

1) SAS uses Weibull cdf form $F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\gamma}}$

· where α = scale & γ = shape

2) Other form: $F(x) = 1 - e^{-x^{\gamma}/\beta}$

· where $\beta = \alpha^{\gamma}$

CENSORED DATA

Types of Censoring: (we are measuring time to occurrence of an event) - could also be amount, ie amount of stress to fracture.

1) No censoring - complete data set:

- We observe $T_1 \ldots T_n$ on all n units.
- Use standard methods of estimation & inference.

2) Right Censoring:

- Observe T_i for $i = 1 \ldots m$, but only know the lower bound for the remaining n-m units.
- Just know remaining n-m units have value > than last recorded value.
- Examples: pt leaves study, study ends, ~~device break~~
- Type I & Type II Censoring are special cases of Right Censoring.

3) Type 1 Censoring (Special Case of R Censoring):

- Study terminates at a pre-selected time, t_c.
- All EUs (n-m) that haven't failed yet, as of t_c, are Type 1 censored.
- For n-m censored units, only know that $T_i > t_c$.

4) Type 2 Censoring (Special case of R Censoring):

- Stop experiment when m^{th} unit fails. (m is pre-selected)
- The remaining n-m units are Type 2 censored.
- For n-m censored units, only know that $T_{(i)} > T_{(m)}$.

5) Left Censoring:

- The n-m units missing measurements are Left Censored, where we only know val of T_i is < recorded val.
- Only have upper bound on value of T_i for censored units.
- Units inspected at certain times - some units fail prior to 1st inspection.
- A device is not sensitive to measure obs below a known threshold.

6) Random Censoring:

- Individual EUs fail to have times recorded due to some external event.
- We use min $\{T_i, C_i\}$ where C_i is time EU leaves study.
- Type 1 is a special case of Random, where C_i's are all the same.
- Pt stops coming, lab equipment fails while recording vals for a given EU, pt is removed due to side effects.

CENSORED DATA, CTD.

- Note: Times $T_1 \dots T_n$ are iid, so failure of 1 unit has no impact on failure of other units.

- **Q-Q Plots w/ Censoring:** Plot only the uncensored values.

- **SAS Fitting & Analysis:**
  - Data set → include an indicator var for censored/not censored for each obs, call it c.
  -
```
PROC LIFEREG DATA = dataset;
    model V * c(1) = / dist = weibull covb;
RUN;
```
    - V = var containing responses
    - weibull is dist you want to fit.
    - covb = print cov matrix of param ests.
  - ↳ Then plug param ests into Survival fctn, 1 - cdf, to get P(V > specified value).

- **Parametric Estimation - Known Dist:**
  - Use MLE techniques, but modify likelihood fctn based on type of censoring.
  - Type II ends up using Windsorized Mean to est mean.

- **Parametric Estimation - Dist-Free (Unknown Dist):**
  - Kaplan Meier Prod Lim Est is undefined for $t > T_{(m)}$.
  - Can use Kaplan Meier PLE to obtain ests for mean & median.
  - Used modified version if have any tied values.

STAT 608 - Regression

HANDOUT 1 - OVERVIEW/FORMULAS

1. **Expectation & Variance of Matrices:**
   - $E(AXB + C) = AE(X)B + C$
   - $Var(AX) = A\,VAR(X)A'$
   - $Cov(AX, BY) = A\,Cov(X,Y)B'$

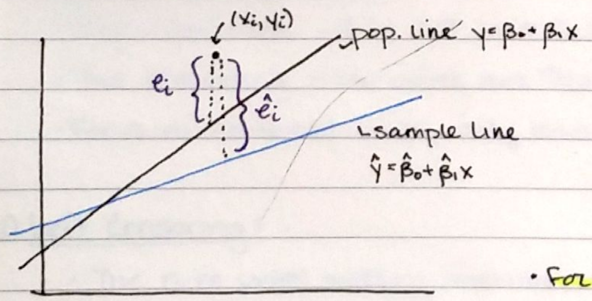
   where
   - X & Y are matrices of R.V.s, or
   - X & Y are vectors of RVs
   - A, B, C are matrices of constants

2. **Quadratic Form:** If X = vector of RVs, & A = n-dimensional symmetric matrix, then $X'AX$ is "quadratic form of X".
   - The expectation of the quadratic form: $E(X'AX) = tr(\Sigma A') + \mu' A \mu$ where μ & $\Sigma A'$ are expected value & covar matrix of X, and tr = trace (sum diag)

HANDOUT 2 - SIMPLE LINEAR REGRESSION

1. **Simple Linear Regression (SLR):** Modeling relationship btwn 2 vars.
   - Data is collected as n response-covariate (y, x) pairs: $(x_1, y_1) \ldots (x_n, y_n)$
   - x is the predictor/explanatory var.
   - y is the response/dependent var.

Model in Vector Form:



1) $y = \beta_0 + \beta_1 x$ is population least squares reg. line.
2) $y_i = \beta_0 + \beta_1 x_i + e_i$ is each population point $(1 \ldots n)$
3) $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ is LS regression line for sample
4) $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{e}_i$ is value at each obs from sample

- For each point: $y_i - \hat{y}_i = \hat{e}_i$, estimated residuals
- Are no e_i or $\hat{e}_i$ for (1) or (3) above, bc these are the regression lines.
- There are e_i residuals for y_i, & $\hat{e}_i$ sample residuals for $\hat{y}_i$.
- Each measures distance of a point to the line.

Goal: Estimate β_0 & β_1 to minimize the sum of squared residuals (SSR):

$$R(\beta_0, \beta_1) = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

Point: Testing if there is an x-y relationship. Is it linear? What is slope?
- Slope: As x increases by 1 unit, y changes by $\hat{\beta}_1$ units.

2. Least Squares Line / LS Estimation of β:

- Matrix form of problem: $Y = X\beta + e$ with $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, X = n×n design matrix, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$, $e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$

- Design Matrix: Depends if fitting an intercept, using dummy variables, etc. For no dummy vars, with intercept, $X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$

- Want to find the LSE of $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ so that the slope & intercept minimize $\Sigma_i \hat{e}_i^2$

$$\boxed{\hat{\beta} = (X'X)^{-1}X'Y}$$

In non-matrix notation:
- $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ (This is why LS line goes thru $(\bar{X}, \bar{Y})$)
- $\hat{\beta}_1 = \dfrac{\Sigma_i (Y_i - \bar{Y})(X_i - \bar{X})}{\Sigma_i (X_i - \bar{X})^2} = \dfrac{SS_{XY}}{SS_X}$

LSE $\hat{\beta}$ Derivation: Want to minimize $R(\beta_0, \beta_1) = R(\beta) = \Sigma_i \hat{e}_i^2 = e'e$

1. Put $R(\beta)$ in matrix form: $R(\beta) = e'e = (Y - X\beta)'(Y - X\beta) = (Y' - X'\beta')(Y - X\beta)$, then FOIL:
   $= Y'Y - Y'X\beta - \beta'X'Y + \beta X'X\beta$. Then take $\partial/\partial\beta$ and set = to 0:

$$\frac{\partial}{\partial\beta} = 0 - \underbrace{2X'Y}_{\text{b/c wrt }\beta\text{, so scalar }X} + \{X'X + (X'X)\beta\} = -2X'Y + 2(X'X)\beta$$

- Set = 0, solve for $\hat{\beta}$: $-2X'Y + 2(X'X)\beta = 0 \rightarrow (X'X)\hat{\beta} = X'Y \rightarrow \hat{\beta} = (X'X)^{-1}X'Y$

3. Expected Val of a Matrix & Cov Matrix:

- $E(X)$ is just the expected value of each indiv element in the matrix
- Cov matrix:

$$\Sigma_{ij} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots \\ & \sigma_{22} & \vdots \\ \cdots & & \sigma_{ij} \end{bmatrix}$$

- Diags are $\sigma_{11} = Var(X_1)$, $\sigma_{22}^2 = var(X_2)$, etc.
- $\sigma_{12} = Cov(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)]$
- If $X_1 \perp X_2$ (they are indep), all off-diag elements are 0.

- $Var(y_i) = Var(e_i)$
- Projected $\hat{y}$ val is: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

4. Assumptions Required to use S-L Regression Model to Make Inferences:

1. X & Y are linearly related.
2. Errors are iid $N(0,\sigma^2)$ → indep of ea other, normal, mean=0, equal vars (σ^2).

5. Inferences:

Want to make 4 types of inferences.

1) CI & hypothesis test for slope (β_1)
2) CI & hypothesis test for intercept (β_0)
3) CI for pop. regression line @ specific x val, ie mean response at x^* (for $\hat{y}^*=\hat{\beta}_0+\hat{\beta}_1 x^*$)
4) PI for projected value of y at a given x, x^*.

(A CI is always reported for a parameter. PI reported for the value of an RV.)

5a) Inferences for Slope (β_1):

- for a given val of x, $\hat{\beta}_1 \sim N(\beta_1, \sigma^2/SS_x) \to T = \frac{\text{RV-mean}}{\text{std. err}} = \frac{\hat{\beta}_1-\beta_1^\circ}{\hat{\sigma}/\sqrt{SS_x}} \sim t_{n-2}$ under H_0.
- Want to test if slope=0, ie $H_0: \beta_1^\circ=0$.
- Reject H_0 if $|T| > t_{\alpha/2, n-2}$
- 100(1-α)% CI for β_1: $\hat{\beta}_1 \pm t_{\alpha/2,n-2}\,\hat{SE}(\hat{\beta}_1)$

5b) Inferences for Intercept (β_0):

- for a given val of x, $\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left[\frac{1}{n}+\frac{\bar{x}^2}{SS_x}\right]\right) \to T = \frac{\hat{\beta}_0-\beta_0^\circ}{\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{SS_x}}} \sim t_{n-2}$ under H_0.
- Reject $H_0: \beta_0^\circ = \beta_0^\circ$ if $|t| > t_{\alpha/2,n-2}$
- 100(1-α)% CI for β_0: $\hat{\beta}_0 \pm t_{\alpha/2,n-2}\,\hat{SE}(\hat{\beta}_0)$

5c) Inferences (CI) for Pop. Regression Line:

- Find a CI for the unknown pop. regression line at a given val of x.
- "For a given val of x, what is the mean Y val?" (Mean val is on regression line.)

- For a given x^*, $T = \frac{\hat{y}^* - (\beta_0+\beta_1 x^*)}{\left[\sqrt{\frac{\hat{\sigma}^2}{n}+\frac{(x^*-\bar{x})^2}{SS_x}}\right]}$ denom is $\hat{SE}(\hat{\beta}_0+\hat{\beta}_1 x^*)$

- 100(1-α)% CI for $E(Y|X=x^*) = \beta_0+\beta_1 x^*$ is:

$$(\hat{\beta}_0+\hat{\beta}_1 x^*) \pm t_{\alpha/2,n-2}\,\hat{SE}(\hat{\beta}_0+\hat{\beta}_1 x^*)$$

$* \hat{\sigma}^2 \approx \frac{RSS}{n-k}$, k = # params being estimated

5d) Prediction Intervals for Actual Y values:

- Finding a PI for a predicted val of Y at a given x.
- CLT does NOT apply. Must have errors ~ Normal.
- Chosen x must be btwn $x_{(1)}$ & $x_{(n)}$ - interpolation is ok, extrapolation is not.

- Predicted val of Y given $x = x^*$: $\hat{y}^* = \hat{\beta}_0 + \hat{\beta}_1 x^*$

- $T = \dfrac{y^* - \hat{y}^*}{\sqrt{\hat{\sigma}^2 \left(1 + 1/n + \frac{(x^* - \bar{y})^2}{SS_x}\right)}}$ ← denom is $\hat{SE}(\hat{y}^*)$

note: in matrix form, $\hat{SE}(\hat{y}^*) = \sqrt{\hat{\sigma}^2 [\vec{1} + (\vec{1}x^*)(x'x)^{-1}[1 \, x^*]']}$

- same as CI for regression line, but SE has a 1 added for extra width.

- $100(1-\alpha)\%$ PI for y at $x = x^*$: $(\hat{\beta}_0 + \hat{\beta}_1 x^*) \pm t_{\alpha/2, n-2} \, \hat{SE}(\hat{y}^*)$

6) Dummy Variable Regression:

- Can use a regression model w/ x as a dummy indicator var to do ANOVA & compare the means of two pops. (Or more, if multivar regression.)

- Model: $y_i = \alpha_0 + \alpha_1 x_i + e_i$ where:
  - y_i = response var
  - α_0 = mean for pop 1 (μ_1)
  - $\alpha_1 = \mu_2 - \mu_1 \Rightarrow$ mean for pop 2 is $\mu_1 + \alpha_1$
  - $x_i = \begin{cases} 0 & \text{if obs in pop 1} \\ 1 & \text{if obs in pop 2} \end{cases}$

- Test $H_0: \alpha_1 = 0$ to see if $\mu_1 = \mu_2$ (same as $H_0: \mu_1 = \mu_2$)

- LSE for $\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 - \bar{y}_1 \end{bmatrix}$

- Design matrix: have to update x with restrictions, so model is not over-parameterized.

$$X = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ \vdots & 1 \\ \vdots & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix} \begin{array}{l} \left.\right] \text{pop 1 indicator} \\ \\ \left.\right] \text{pop 0 indicator} \end{array}$$

intercept

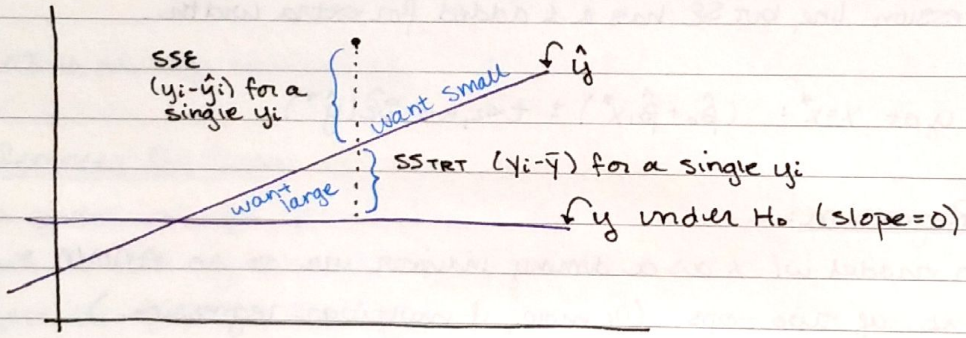
drop last col → don't use $\begin{matrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \end{matrix}$

$* \; h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SS_x}$

7) Geometric Explanation of $\hat{y}$, the Hat Matrix, + ANOVA:

- $\hat{y} = Hy = X\hat{\beta}$, where H = 'hat matrix'.
- $H = X(X'X)^{-1}X'$, an $(n \times n)$ matrix. Each h_{ii} is a 'leverage score'.
  - H is the projection matrix; it projects y onto $\hat{y}$.
  - H maps the vector of obs vals ($\hat{y}$) to the vec of fitted vals (y). ie, H measures the influence of each $\hat{y}_i$ on each y_i.

- AOV breaks up SS_{TOT} into $SS_{TRT} + SSE$ (or SST = SSReg + SSR)
  - SS_{TRT} = amount of variance attributed to the model.
  - SSE = amount of variance attributed to error; unexplained var. Want small.



- $SS_{TOT} = SS_{TRT} + SSE \Rightarrow \underbrace{\sum_i (y_i - \bar{y})^2}_{SS_{TOT}} = \underbrace{\sum_i (\hat{y}_i - \bar{y})^2}_{SS_{TRT}} + \underbrace{\sum_i (y_i - \hat{y}_i)^2}_{SSE}$

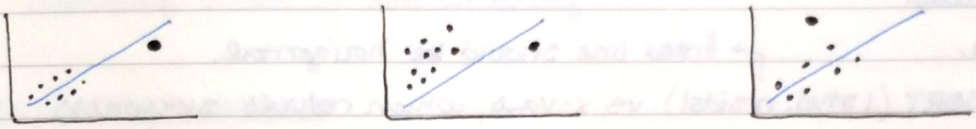
Also: SST = SSReg + RSS

Handout 3 - Checking Model Assumptions (Simple Linear Regression)

1) Tools & Concepts Needed:

- Leverage Points: Points with unusually large influence on the model.
  - A point whose x-val is far away from other x-vals.
  - We classify leverage points as 'good' or 'bad'.
    - Good: A leverage point that doesn't change slope, ie not an outlier.
    - Bad: A leverage point that changes the slope; is also an outlier.

- Classify a point as a leverage point if $h_{ii} > 2 * avg(h_{ii}) = 2\left(\frac{p+1}{n}\right)$, $p = \#$ predictors.
  - For simple linear regression, is equiv to $h_{ii} > 4/n$.



good L point   bad L point   outlier but not L point

- Cook's Distance:
  - A way of looking at leverage & outlier points by assessing X & Y together. (Leverage alone only looks at X.)
  - $D_i = \frac{\sum_{j=1}^{\hat{n}} [\hat{Y}_{j(i)} - \hat{Y}_i]^2}{(p+1)\hat{\sigma}^2} = \frac{r_i^2 h_{ii}}{(p+1)(1-h_{ii})}$ with $p = \#$predictor vars, r_i = std. residuals

  - D_i noteworthy (large) if $D_i > \frac{4}{n-2}$. In practice, look for big gaps.
  - Cook's cutoff lines in SAS & R plots identify bad leverage points.

- Standardized Residuals: $r_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$
  - Resids div by their std. errors
  - Vals $> |t_{\alpha/2, n-p}|$ are outliers

- Studentized Residuals: $\hat{e}^*_{ij} = \frac{\hat{e}_{ij}}{\hat{\sigma}_e \sqrt{1 - 1/n_i}}$ where $\hat{\sigma}_e = MSE$
  - Classify point as outlier if $|\hat{e}^*_{ij}| > 2$, roughly.

2) Regression Diagnostic Steps:

1. Determine if proposed regression model is valid.
   - Plot standardized residuals, with x-vals on x-axis. Should be a blob.
   - Other patterns (parabola) indicate non-linear relationship or other trends.
   - Also look for unequal variances, see (3). Use same plot.

2. Q-Q Normal Ref Dist Plot of Residuals:
   - Verify if errors normally distributed. Use std. resids in Q-Q plot.
   - Can also use SW test; careful if have large n vals. Always use with Q-Q.

3. Equal vars of Residuals:
   - Use plot above (1). fitted line should be horizontal.
   - Even better, plot SQRT (|stud. resids|) vs x-vals, which reduces skewness in abs vals.
   - Can also use BFL test to go w/ plots, IF x is an indicator var.

4. If data has an index (spatial or temporal), check for indep errors.
   - Durbin-Watson Test
   - If no index, must validate based on how experiment was run.

5. Analyze leverage points + outliers:
   - Look at plot of std. resids; outliers fall outside $(-2,2)$.
   - Look at leverage points + cook's distance to identify bad leverage points.
   - Bad leverage points mean could be fitting wrong model, or these points are different somehow - remove them, or add another predictor var. May point to an important model feature not being considered.

Note: High R^2 does not mean a good fit! Have to check assumptions.
- R^2 takes on diff vals when diff portions of the full possible range of x vals are observed.
- R^2 is not invariant under transformation, nor under location changes (shifts).

3) What to do when assumptions violated?

1. Consider another model (polynomial regression, multiple regression, etc.)
2. Transform data: can overcome a number of issues.
3. Weighted Least Sqs, for unequal regression.

   *(1) + (3) later, transforms first:

4) Transformations:

Can be used for the following:

- Overcome probs w/ constant variance (can fix normality, too)
- Estimate percentage effects
- Overcome probs w/ non-linearity

4a) Non-constant variance:

- Can sometimes transform one or more vars to get constant variance.
- $f(y) = \sqrt{Y}$, "Square Root Transform", good for count data like Poisson, doing transform on x + y. (Makes sense if X + Y in same units, such as counts.)
- $Var[f(y)] = [f'(E(Y))]^2 Var(Y)$ using a Taylor Series expansion.

4b) Estimate Percentage Effects:

- log(y) transform used to estimate % effects (transform x + y).
- Without transform, "increasing x by 1 unit means approx $\hat{\beta}_1$ change in y." But now, "a 1% chg in x means a $\hat{\beta}_1$% change in y."
- End up with $\beta_1 = \Delta \log(y) / \Delta \log(x) = \% \Delta y / \% \Delta x$. (Works as long as β_1 fairly small.)

4c) Overcoming Linearity Issues:

- Can transform x, y or both.
- 2 types of transforms:
  1) Box-Cox: $y_{new} = y^{\lambda}$, + if $\lambda = 0$, use $y_{new} = \log(y)$.
     - Usually choose λ as $\{-1, -1/2, -1/3, -1/4, 0, 1/4, 1/3, 1/2, 1\}$
     - Often do x and y, bc then both are normal + thus linearly related.
  2) Inverse Response Plots:
     - Plot y on the x-axis. Plot $g^{-1}(y)$ on y-axis to discover $g^{-1}(y)$'s shape.
     - Plot shows several lines, each w/ a different coefficient k.
     - Pick one that visually fits best, + use transform $y_{new} = y^k$.

4) Transformations, ctd:

4d) 2 approaches for transforming both x & y:

call the transform fctn $\psi(x,\lambda)$

1. Approach 1: Transform x to as normal as possible. (Try Box-Cox) Then use model $y = g(\beta_0 + \beta_1 \psi(x,\lambda) + e)$. Use inverse response plot to find transformation for y.
2. Approach 2: Transform X & Y simult. to joint normality using multivariate Box-Cox.

* Sometimes approaches yield same answer. (1) is more robust, often works when (2) fails.

4e) Transformations and CI/PI's:

- when transforming endpoints back to original vars, must add a correction factor. Only for CI; PI doesn't need this.
- w/o correction factor, back-transform leads to biased endpoints where the coverage prob is wrong. (B/c relationship btwn X & $g(Y)$ not linear - if it were, we wouldn't have had to do a transform.)
- Correction factors:

T	Trans Model	Back Transform (Correction Factor in pink)
log	$\log(y) = \beta_0 + \beta_1 x + e$	$\hat{E}(Y) = \exp(\hat\beta_0 + \hat\beta_1 x + \hat\sigma^2/2)$
Sqrt	$\sqrt{y} = \beta_0 + \beta_1 x + e$	$\hat{E}(Y) = (\hat\beta_0 + \hat\beta_1 x)^2 + \hat\sigma^2$
Inverse	$1/y = \beta_0 + \beta_1 x + e$	$\hat{E}(Y) = \frac{1}{\hat\beta_0 + \hat\beta_1 x}\left(1 + \hat\sigma^2/(\hat\beta_0 + \hat\beta_1 x)^2\right)$
Inverse Sqrt	$1/\sqrt{y} = \beta_0 + \beta_1 x + e$	$\hat{E}(Y) = \frac{1}{(\hat\beta_0 + \hat\beta_1 x)^2 + \hat\sigma^2}\left(\frac{1 + 2\hat\sigma^4 + 4(\hat\beta_0 + \hat\beta_1 x)^2 \hat\sigma^2}{[(\hat\beta_0 + \hat\beta_1 x)^2 + \hat\sigma^2]^2}\right)$

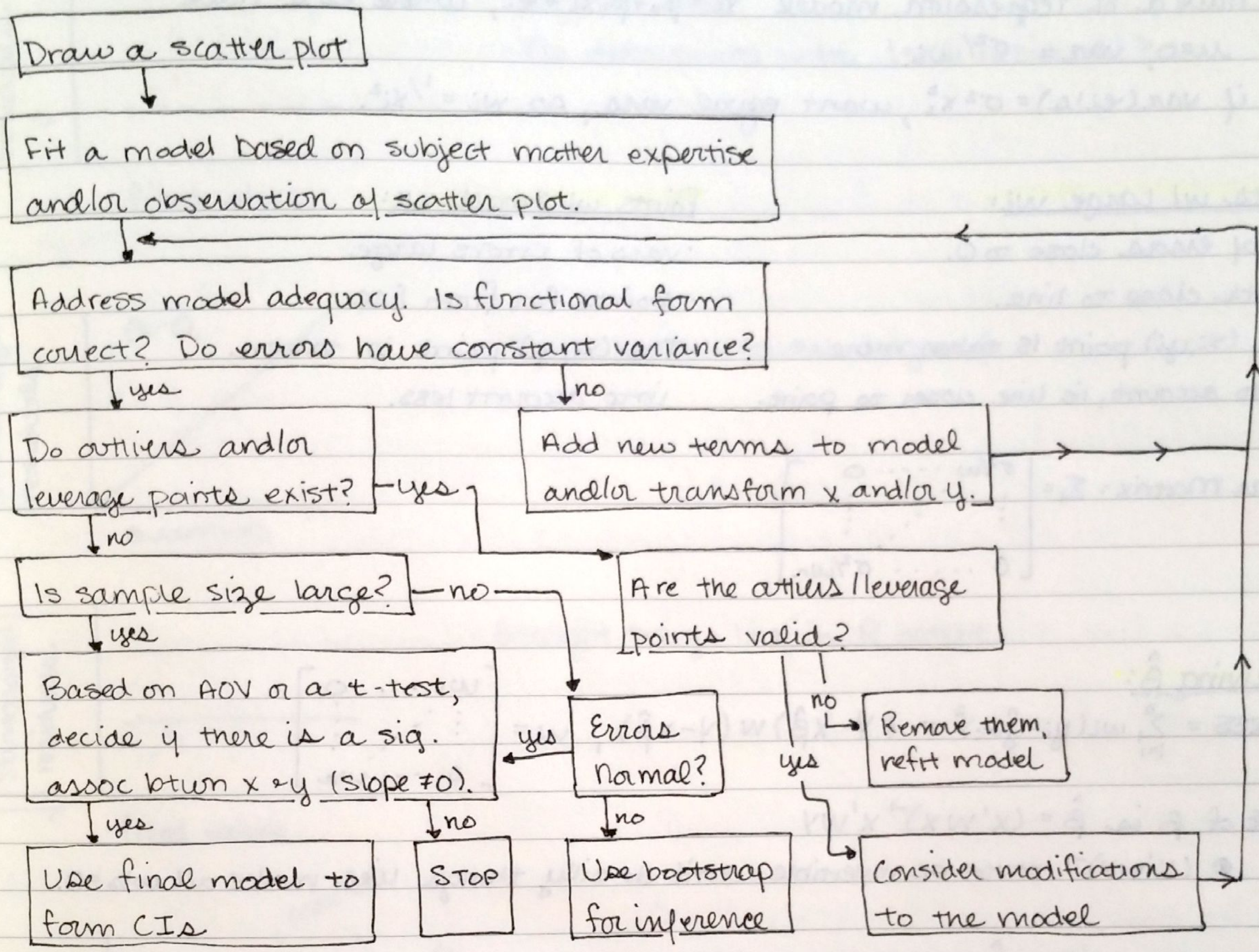
* $\hat\sigma^2 = \hat{MSE}$

Example of how to back-transform endpoints:

- Log transform. $Z = \log(Y)$. CI for Z: (L_Z, U_Z). CI for μ_Y:
  $\exp\left[L_Z + \frac{\hat\sigma_Y^2}{2}, U_Z + \frac{\hat\sigma_Y^2}{2}\right]$

* If $g(Y)$ monotone increasing, CI is $[g^{-1}(L_Z), g^{-1}(U_Z)]$
  If $g(Y)$ monotone decreasing, CI is $[g^{-1}(U_Z), g^{-1}(L_Z)]$

5) Regression Diagnostics Flow Chart:



CHAPTER 4 - HANDOUT 4, WEIGHTED LEAST SQUARES

- WLS is another method to handle unequal variances.
- We have a SL regression model $Y_i = \beta_0 + \beta_1 x_i + e_i$, where e_i's have $\mu = 0$, var $= \sigma^2 / w_i$.
- So if $var(e_i\text{'s}) = \sigma^2 x_i^2$, want equal vars, so $w_i = 1/x_i^2$.

Points w/ Large w_i:

- Var of errors close to 0.
- Points close to line.
- The (x_i, y_i) point is taken more into account, ie line closer to point.

Points w/ Small w_i:

- vars of errors large
- Points far from line
- The (x_i, y_i) point is taken into account less.

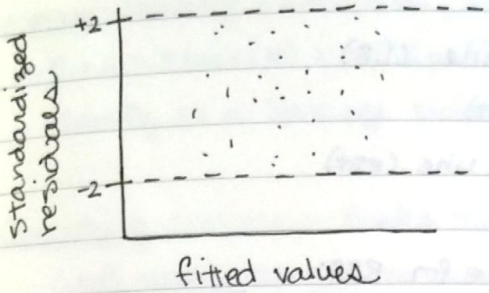
- Covar Matrix: $\Sigma_1 = \begin{bmatrix} \sigma^2/w & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2/w_n \end{bmatrix}$

Deriving $\hat{\beta}$:

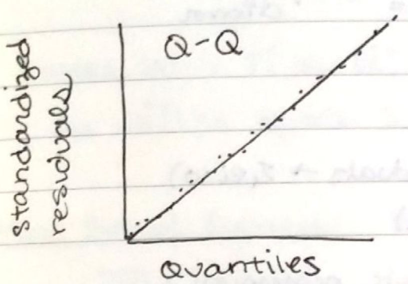
- $WRSS = \sum_{i=1}^{n} w_i (y_i - \hat{y}_{w_i})^2 = (Y - X\hat{\beta}) W (Y - X\hat{\beta})$, $W = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{bmatrix}$
- LSE of β is $\hat{\beta} = (X'WX)^{-1} X'WY$
  - ⊛ $(X'WX)^{-1}$ must be invertible - don't do silly things like make all w's $= 0$.

- Can also derive $\hat{\beta}$ by saying $e_i = x_i \varepsilon_i$ if $var(e_i) = x_i \sigma_i^2$, for example. Then sub into $Y_i = \beta_0 + \beta_1 x_i + e_i$ and solve for e_i. Then minimize $R(\beta_0, \beta_1) = \sum_1 \varepsilon_i^2$ by taking 1^{st} derivs wrt β_0 & β_1. (608 HW2, prob 5)

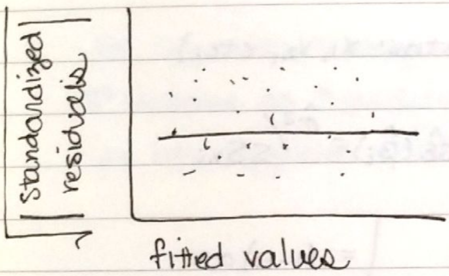
Quick Guide to Plots: What We Want to See



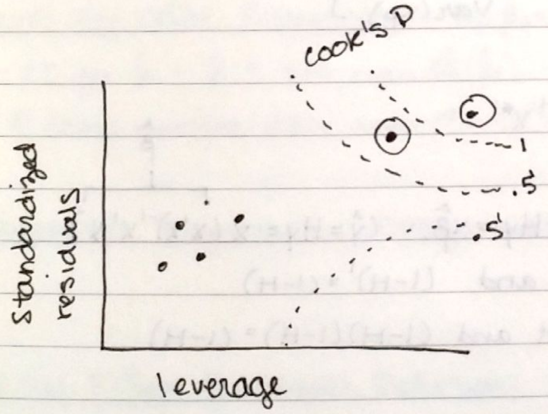
- random blob
- No data points with |std. residual| > 2



- Data points follow diagonal



- Straight horiz. line in R output.



- watch for 'bad leverage points' on wrong side of Cook's D lines.

Exam 2 Equations:

Review:
1. $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$ — pop. point
2. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ — pop. regression line (LS)
3. $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{e}_i$ — sample point (est)
4. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ — sample regression line (est)

Matrix Form: $Y = X\beta + e$ & $\hat{Y} = X\hat{\beta} + \hat{e}$

Can solve for RSS:
Resid Std err $= \sqrt{MSE} = \sqrt{RSS/df_{error}}$

- $\hat{e}_i = y_i - \hat{y}_i$
- $Var(e_i) = \sigma^2$
- $Var(y) = \sigma^2 I$
- $Var(\hat{e}_i) = \sigma^2 (I - H) = MSE = \hat{\sigma}^2 = RSS/df_{error}$
- $Var(\hat{y}) = \sigma^2 H$

- $\hat{e}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2})$
- $\sum y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_i x_{i1} + \hat{\beta}_2 \sum_i x_{i2}$ (residuals $\rightarrow \sum_i e_i = 0$)
- $\sum_i y_i x_i = \hat{\beta}_0 \sum_i x_i + \hat{\beta}_1 \sum_i x_i^2$ (SLR)
- $\sum_i y_i x_{i1} = \hat{\beta}_0 \sum x_{i1} + \hat{\beta}_1 \sum_i x_{i1}^2 + \hat{\beta}_2 \sum x_{i2}^2$ (mult. regression)

- $\sum_i \hat{e}_i = \sum x_{i1} \hat{e}_i = 0$ (In mult. regression, can use any col from X matrix: x_1, x_2, etc.)

Slope Stuff: $\hat{\beta} = (X'X)^{-1}X'Y$ · $E(\hat{\beta}|X) = \beta$ · $Var(\hat{\beta}|X) = \sigma^2 (X'X)^{-1}$ · $\hat{SE}(\hat{\beta}_i) = \hat{\sigma}^2 / SSx_i$

Covariance: $\Sigma_{ij} = Cov(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] = \begin{bmatrix} Var(x_1) & & 0 \\ & \ddots & \\ 0 & & Var(x_p) \end{bmatrix} = (I - H)\sigma^2$

- Pop. mean at a given x^* is $E(Y|X = x^*) = x^*\beta$
- Var of pop mean at a given x^* is $V(\hat{y}^*) = \sigma^2 x^* (X'X)^{-1} x^{*\prime}$

Hat Matrix: Projects y onto $\hat{y}$. · $H = X(X'X)^{-1}X$ · Then $\hat{y} = Hy = X\hat{\beta}$. $(\hat{y} = Hy = X\underbrace{(X'X)^{-1}X'Y}_{\hat{\beta}})$

- H and (I−H) are symmetric & idempotent.
  - $H' = H$ and $(I-H)' = (I-H)$
  - $HH = H$ and $(I-H)(I-H) = (I-H)$

Orthogonality: A & B (both n×n) are orthog. if $AB = BA = 0$.
1) H and (I−H) are orthog: $H(I-H) = H - HH = H - H = 0$, & $(I-H)H = H - HH = H - H = 0$.
2) $\hat{e} = y - \hat{y}$ and $\hat{y} = Hy$ are orthog: $\hat{y}'\hat{e} = \hat{y}'(y - \hat{y}) = (Hy)'(y - Hy) = y'H'(I-H)y = y'H(I-H)y = y'0y = 0$.

Studentized Residuals: $r_i = \hat{e}_i / \hat{\sigma}\sqrt{1 - h_{ii}}$ (also written as $\hat{e}_i^*$)

Corr. Btwn Residuals: There is a small amount of corr btwn residuals. $Cov(e) = \sigma^2$.
$Cov(\hat{e}_i, \hat{e}_j) = -h_{ij}\sigma^2$ for $i \neq j$. $\rightarrow corr(\hat{e}_i, \hat{e}_j) = \dfrac{-h_{ij}}{\sqrt{(1-h_{jj})(1-h_{ii})}}$ for $i \neq j$.

⊛ $\bar{y} = \bar{\hat{y}}$

Leverage: leverage points = h_{ii}, diagonals of matrix H. $\cdot\ h_{ii} = 1/n + (x_i-\bar{x})^2/SS_x$

- $\sum_i h_{ii} = \text{trace}(H) = tr(I_{p+1}) = p+1$
- Classify as a leverage point if $h_{ii} > 2 * \text{avg}(h_{ii}\text{'s}) \approx \frac{2(p+1)}{n}$ ($= 4/n$ for SLR)

Classifying Outliers: Books typically say to classify as outlier if $|\hat{e}_i^*| > 2$.

- Cutoff really depends on acceptable prob of misclassifying a point as an outlier.
- $S = 1-(2\Phi(c)-1)^n \rightarrow c = \Phi^{-1}\left(\frac{1+(1-s)^{1/n}}{2}\right)$ · c = cutoff value to use instead of 2.
  · S = .05, .1, or any specified threshold.
- Works b/c $1 - P[\text{all } |\hat{e}_i^*| < c] = 1 - P[|\hat{e}_i^*| < 2]^n$ → So we are finding prob of at least one $|\hat{e}_i^*| > c$ equals S, the specified α.

More Sum of Squares: $R^2 = \frac{SSReg}{SST} = 1 - \frac{RSS}{SST}$ (SST = SSReg + RSS) ie $SS_{TOT} = SS_{TRT} + SSE$

- $SST = \sum_i (y_i-\bar{y})^2$ · $SSReg = \sum_i(\hat{y}_i-\bar{y})^2 = \sum_i(\hat{y}_i-\bar{\hat{y}})(y_i-\bar{y})$ RSS (SSE) $= \sum(y_i-\hat{y}_i)^2$

- For regression through the origin, $R^2 = 1 - SSReg/\sum_i y_i^2$
- R^2 increases as # predictors increase. Solution: $R^2_{adj} = 1 - \frac{RSS/(n-p-1)}{SST/(n-1)}$
- ρ = corr coefficient $= \sqrt{R^2} = \frac{\sum_i(y_i-\bar{y})(\hat{y}_i-\bar{y})}{\sqrt{\sum_i(y_i-\bar{y})^2\sum_i(\hat{y}_i-\bar{y})^2}}$

Anova for Mult. Regression: $H_0: \beta_1 = \ldots = \beta_p = 0$ (all slopes 0). $F = \frac{SSReg/p}{RSS/(n-p-1)}$ p = # slopes/coeffs, excludes β_0.

- CI for $\hat{\beta}_i$: $\hat{\beta}_i \pm t_{\alpha/2, n-p-1}\, \hat{SE}(\hat{\beta}_i)$
- If doing multiple CIs, use $\alpha_{pc} = \alpha_0/m$, m = # slopes, to adjust for mult. comparisons.

Polynomial Regression: Design matrix $X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$

Partial F-Test for Model Reduction: $H_0: \beta_1 = \ldots \beta_K = 0$, ie some subset of slopes = 0.

- If fail to reject H_0, can remove this subset of predictors from model.
- $F = \frac{(RSS_{reduced} - RSS_{full})/(df_{reduced} - df_{full})}{RSS_{full}/df_{full}}$ * The df's are the df_{error} for each model.

- Is just one tool in arsenal of model reduction. Is not wise to eliminate all predictors with insig t-vals, b/c pvals change as ea. predictor is removed, if have multicollinearity. Use as a tool to get a sense of what is going on when have a small # of total predictors.

EXAM 2 EQNS, CTD.

Model Checking: A model is valid if $E(Y|X=x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$ and $Var(Y|X=x) = \sigma^2$.

- If model valid, plot of resids against any predictor or linear combo of predictors...
  1) Has a random scatter of points (a blob).
  2) Has constant variability as horiz. axis increases.
  ↳ Any patterns imply invalid model.
- But unlike SLR, resid plots only help diagnose what is wrong if these both hold:
  1) $E(Y|X=x) = g(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)$, ie some fcth $g^{-1}(y)$ exists so $g^{-1}(y)$ has linear rel w/ x's.
  2) $E(x_i|x_j) = \alpha_0 + \alpha_1 x_j$, ie predictors themselves are linearly related.

Model Selection Methods:

- Overfitting - Too many predictors. Doesn't generalize well to new data.
- Underfitting - Too few predictors

2 General Strategies:

1) Select a model by looking @ subsets of predictors: Fwd, Backwards, Stepwise.
2) Look @ all possible combos + select based on R^2_{adj}, AIC, AICC, or BIC.
   - If have m vars, there are 2^m possible regression eqns. If m small, run all.
   - If m large, not feasible to run all, so use fwd/backward/stepwise.

1) Backward: Start w/ all m vars. Delete predictor w/ largest pval (as long as it isn't sig). Rerun model with p-1 predictors. Keep deleting vars until everything left is sig.
2) Forward: Run all m models w/ only 1 var. Add var w/ smallest pval (as long as it is sig.) Then run all models w/ x_1 + one other predictor. Add var w/ smallest pval. Repeat until no more sig vars to add in.
3) Stepwise: 2 fwd, 1 back, 1 fwd, 1 back, and so forth.

* If Cov of predictors = 0, all 3 methods yield same model.

Selection Criteria Looking @ All Subsets:

1) R^2_{adj} - Choose subset w/ highest. Prone to overfitting; not often used alone.
2) AIC - $n\log(RSS/n) + 2p$ → choose model w/ smallest AIC, ie closest to $-\infty$.
3) AICC - $AIC + 2(p+2)(p+3)/(n-p-3)$ → choose model w/ smallest AICC.
   - Converges to AIC as $n \to \infty$. Corrects for AIC's bias.
4) BIC - $n\log(RSS/n) + (p+2)\log(n)$ → Choose model w/ smallest. Favors simpler models than AIC.

⊛ AIC overfits when n large. BIC underfits when n small.

More on Model Selection:

- If polynomial regression, may need to do selection by hand. SAS doesn't know to keep x in the model just b/c x^2 is in model.
- LASSO - Very efficient calculation. Does var selection + parameter estimation simultaneously. Shrinkage method.

How to Assess Predictive Ability of Models?

- If you end up with a few viable models, compare by splitting ~~test~~ data set, use 60-70% to train model, + 30-40% to assess model error.
- Split randomly. If corr structure exists, must preserve it.
- Want 2 data sets similar in terms of dists, means, vars, outliers. Doesn't always happen, esp if n small.
- Cross-validation.

Var Inflation Factors (VIF): Probs w/ multicollinearity if any VIF > 5.

- $VIF_j = \frac{1}{1-R_j^2}$ → R_j^2 denotes R^2 val from regressing x_j onto other predictors, ie using x_j as the response variable. This is R^2, not R^2_{adj}.

- Note: $Var(\hat{\beta}_j) = \frac{1}{1-R_j^2} \cdot \frac{\sigma^2}{(n-1)\sigma^2_{x_j}}$ for $j = 1 \ldots p$.

- VIF = 5 means $Var(\hat{\beta}_j)$ is 5x as large as it would've been otherwise.

PROOFS NOTES (VELLEMAN)

CH 1 - SENTENTIAL LOGIC

· Basic connective symbols: ($\equiv$ means equivalent)

SYMBOL	MEANING	EX	TERM
$\vee$	OR (inclusive)*	$P \vee Q \equiv$ P or Q	The disjunction of P & Q.
$\wedge$	AND	$P \wedge Q \equiv$ P and Q	The conjunction of P & Q.
$\neg$	NOT	$P \wedge \neg Q \equiv$ P and not Q	The negation of P.

Parenthesis convention: $\neg$ applies to statement immediately following it.

$\neg P \wedge Q$ means $(\neg P) \wedge Q$, not $\neg(P \wedge Q)$.

* Inclusive OR used in math. XOR common in computer science.

TRUTH TABLES:

P	Q	$P \vee Q$
0	0	0
1	0	1
0	1	1
1	1	1

P	Q	$P \wedge Q$
0	0	0
1	0	0
0	1	0
1	1	1

P	$\neg P$
1	0
0	1

P	Q	P XOR Q
0	0	0
1	0	1
0	1	1
1	1	0

· Arguments Notation:
  · Symbol $\therefore$ is 'therefore'.
  · Can represent an argument in this form:

$$\begin{array}{l} P \vee Q \\ \underline{\neg Q} \\ \therefore P \end{array}$$

· Can make truth tables for arguments & conclusions
· Ex: Is this argument valid?

$$\begin{array}{l} (\neg S \wedge L) \vee S \\ \underline{S} \\ \therefore \neg L \end{array}$$

vars in premises		premises		conclusion
S	L	$(\neg S \wedge L) \vee S$	S	$\neg L$
0	0	0	0	1
1	0	1	1	1
0	1	1	0	0
1	1	1	1	0

Argument invalid! Why? Bc it is possible for both premises to be true and the conclusion to be false.

⊛ Arguments are equivalent if they have identical truth tables.

How to confirm validity of an argument via truth table?

1) Argument is invalid if there exists any row in the table where ALL premises = 1 AND conclusion = 0.

2) Argument is valid if, for all rows where all premises = 1, conclusion = 1.

(Don't care about rows where some or all premises ≠ 1.

LOGICAL EQUIVALENCE LAWS (pg 21)

De Morgan's Law	$\neg(P \wedge Q) \equiv \neg P \vee \neg Q$
	$\neg(P \vee Q) \equiv \neg P \wedge \neg Q$
Commutative Law	$P \wedge Q \equiv Q \wedge P$
	$P \vee Q \equiv Q \vee P$
Associative Law	$P \wedge (Q \wedge R) \equiv (P \wedge Q) \wedge R$
	$P \vee (Q \vee R) \equiv (P \vee Q) \vee R$
Idempotent Law	$P \wedge P \equiv P$
	$Q \vee Q \equiv Q$
Distributive Law	$P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge R)$
	$P \vee (Q \wedge R) \equiv (P \vee Q) \wedge (P \vee R)$
Absorption Law	$P \vee (P \wedge Q) \equiv P$
	$P \wedge (P \vee Q) \equiv P$
Double Negation Law	$\neg\neg P \equiv P$

TAUTOLOGY LAWS: (Tautology = a statement that is always true, such as $P \vee \neg P$.)

1. $P \wedge$ (a tautology) $\equiv P$
2. $P \vee$ (a tautology) $\equiv$ a tautology
3. $\neg$(a tautology) $\equiv$ a contradiction

CONTRADICTION LAWS: (Contradiction is a statement which is always false, ie $P \wedge \neg P$.)

1. $P \wedge$ (contractiction) $\equiv$ contradiction
2. $P \vee$ (contradiction) $\equiv P$
3. $\neg$(a contradiction) $\equiv$ a tautology

· $\forall$ = 'For every' / 'For all'. $\forall x \in \mathbb{R}$ = 'For all x in the reals...'

· $\exists$ = 'There exists' / 'There is at least one'

SET THEORY NOTATION:

· Set - A collection of objects; order unimportant. $A = \{1, 2, 3\}$

· $\in$ - An object is an element of a set. $2 \in A$. $7 \notin A$. ($\notin$ = not an element of)

· Truth Set: If a statement contains variables, it is not just T or F - it is dependent on the value of the variables. Truth set is the set of vars for which a statement is true. Truth set of $P(x) = \{x \mid P(x)\}$

* Elements can appear in a set > once. $A = \{1, 2, 3\}$. $B = \{1, 2, 2, 2, 3, 3\}$. $A \equiv B$.
* A set is completely determined once all of its elements are specified.

* Sets often too big to list all elements → notated by specifying a pattern or an elementhood test:
  · $A = \{x \mid x$ is a prime number$\}$
  · $A = \{x \mid x^2 < 20\}$

Free vs Bound Variables:

· Take $y \in \{x \mid x^2 < 20\}$. y is ~~free~~ free, x is bound.

· x is a dummy or placeholder var, representing the set elements.

· The notation $\{x \mid ...\}$ 'binds' the x variable.

* Statements can be written in complex set notation, without any free variables.

Ex: $2 \in \{w \mid 6 \notin \{x \mid x$ is ~~an even number~~ divisible by $w\}\}$. How to break it down?

  · Inner part: $6 \notin \{x \mid x$ is divisible by $w\}$ = "6 is not divisible by w. Sub this in:
  · $2 \in \{w \mid 6$ is not divisible by $w\}$.
  · This means '6 is not divisible by 2', so this statement is false.

Universe of Discourse: when a statement contains free vars, statement often implies a context for the type of vars.

  · Best to explicitly define, so no room for confusion. $\{x \in U \mid P(x)\}$ = 'set of all x in U so that P(x).'

  We say variables 'range over' the universe of discourse.

Ex: 1. $x^2 < 20$ might be over all real numbers.

  2. 'x is male' might range over all humans.

* The Empty Set contains no elements. Notated by $\emptyset$.

  · $\{\} = \emptyset$. Ex: $\{x \in \mathbb{Z} \mid x \neq x\} = \emptyset$.

Commonly Used Universes of Discourse:

- $\mathbb{R} = \{x \mid x$ is a real number$\}$ (Reals: any number on the number line.)
- $\mathbb{Q} = \{x \mid x$ is a rational number$\}$ (Rationals = any number which can be written as p/q, where p & q are integers.)
- $\mathbb{Z} = \{x \mid x$ is an integer$\}$
- $\mathbb{N} = \{x \mid x$ is a natural number$\} = \{0, 1, 2, 3 \ldots\}$ (0 is sometimes not included in $\mathbb{N}$)

* $\mathbb{R}$, $\mathbb{Q}$ & $\mathbb{Z}$ can be followed by a + or - superscript to indicate only pos or neg values.

- 0 is neither positive nor negative.
- $\{x \mid x$ is a negative integer$\} = \mathbb{Z}^-$
- $\{x \in \mathbb{R} \mid x^2 < 20\}$ = The set of x in real #s, so that $x^2 < 20$.

SET OPERATIONS:

1. Intersection: $A \cap B \equiv \{x \mid x \in A$ and $x \in B\} \equiv x \in A \wedge x \in B$
2. Union: $A \cup B \equiv \{x \mid x \in A$ or $x \in B\}$ (inclusive OR) $\equiv x \in A \vee x \in B$
3. Difference: $A \setminus B \equiv \{x \mid x \in A$ and $x \notin B\} \equiv x \in A \wedge \neg x \in B$
4. Symmetric Difference: $A \Delta B \equiv (A \cup B) \setminus (A \cap B) \equiv (A \setminus B) \cup (B \setminus A)$. Elements in A or B, but not both.
   - Same logical laws apply here (DeMorgan's, tautologies, etc.)
   - Can form truth tables & Venn Diagrams to check equivalences.

4. Disjoint: A & B disjoint if $A \cap B = \emptyset$. (Share no common elements.)
5. Subset: $A \subseteq B \equiv$ 'A is a subset of B'. All elements of A are also in B. (Includes $A = B$.)
   $A \subseteq B \equiv \forall x (x \in A \rightarrow x \in B)$.

Proof Example: Theorem: For any sets A & B, $(A \cup B) \setminus B \subseteq A$.

(Must show: if $x \in (A \cup B) \setminus B$, x is also in A.)

Suppose $x \in (A \cup B) \setminus B$. So $x \in (A \cup B) \wedge x \notin B$. Rewrite as $(x \in A \vee x \in B) \wedge x \notin B$.

Distributive property yields $(x \in A \wedge x \notin B) \vee (x \in B \wedge x \notin B)$.

The second piece, $(x \in B \wedge x \notin B)$ is a contradiction, so by contradiction law, we simplify the statement to $(x \in A \wedge x \notin B)$. So $x \in A$ is true, thus any element in $(A \cup B) \setminus B$ is also in A. □

* Even if a conclusion seems intuitive, use the laws & BE RIGOROUS! This is how to avoid jumping to false conclusions.

Conditional & Biconditional Connectives:

1) $P \to Q \equiv$ 'If P then Q'. P is the 'antecedent', Q is the 'consequence'.

- Also written as: $P \to Q$, P, $\therefore Q$

P	Q	$P \to Q$
0	0	1
1	0	0
0	1	1
1	1	1

Conditional Law: $P \to Q \equiv \neg P \vee Q$ (and $\neg P \vee Q \equiv \neg(P \wedge \neg Q)$)

- Converse: $Q \to P$. These are NOT equivalent! $P \to Q \equiv Q \to P$ is a logical fallacy.
- Contrapositive: Contrapos of $P \to Q$ is $\neg Q \to \neg P$.
  - Logically equivalent! $P \to Q \equiv \neg Q \to \neg P$.

Other ways to say $P \to Q$ in english:

- If P, then Q.
- Q, if P.
- P implies Q.
- P only if Q.
- P is a sufficient condition for Q.
- Q is a necessary condition for P.

Simplification Example with 2 Conditionals:

$(R \to C) \wedge (S \to C) \equiv (\neg R \vee C) \wedge (\neg S \vee C)$ by conditional law

$\equiv (\neg R \wedge \neg S) \vee C$ by distributive law $\equiv \neg(R \vee S) \vee C$ by DeMorgan's

$\equiv (R \vee S) \to C$ by conditional law.

* 'Necessary and sufficient' $\equiv$ biconditional $P \leftrightarrow Q \equiv$ P IFF Q.

2) $P \leftrightarrow Q \equiv$ 'P if and only if Q'. This is the biconditional statement.

- Biconditional Law: $P \leftrightarrow Q \equiv (P \to Q) \wedge (Q \to P)$
  - Also, $P \leftrightarrow Q \equiv (P \to Q) \wedge (\neg P \to \neg Q)$ using contrapositive.

- Truth Table:

P	Q	$P \leftrightarrow Q$
0	0	1
1	0	0
0	1	0
1	1	1

CH 2 - Quantificational Logic

Two quantifiers express how many values of x make $P(x)$ true.

1. $\forall$ = universal quantifier. $\forall x\, P(x)$ = 'For all x, $P(x)$.'
2. $\exists$ = existential quantifier. $\exists x\, P(x)$ = 'There exists an x so that $P(x)$.' (At least one.)

↓

2a. $\exists! x\, P(x)$ = $\exists$ exactly 1 unique x so that $P(x)$. $\equiv \exists x\,(P(x) \wedge \neg \exists y\,(P(y) \wedge x \neq y))$

- Quantifiers bind a variable; for some $P(x)$, $\forall x P(x) \equiv \forall w P(w)$.
- Parenthesis convention: $\forall x\, P(x) \rightarrow Q(x) \equiv (\forall x P(x)) \rightarrow Q(x)$, same as $\neg$ convention.
- Words indicating need for quantifier: everyone, someone, everything, something, nobody, nothing.

Multiple Quantifiers:

- Statements can have 2+ quantifiers. Read them from L to R. Order matters!*
  - Ex: Let $M(x,y)$ = 'x is married to y'. $S(x)$ = 'x is a student.'
    Then 'some students are married' $\equiv \exists(x)\,(S(x) \wedge \exists y\, M(x,y))$.

* Order matters for $\forall x \exists y$ (for all x, there is a y...) + $\exists x \forall y$ (there is an x, for which all y...)
Order does not matter for $\forall x \forall y$ (for all x + y) + for $\exists x \exists y$ (for some x + y combo...)
↳ For any x + y, these statements do not prevent x+y from being the same value.

Ex: 1) $\forall x \exists y\,(x+y=5)$. True → $y = 5-x$ fulfils this for all x values.
2) $\exists y \forall x\,(x+y=5)$. False → for each y, only one x satisfies $x+y=5$. No val of y where it is there for all x.

Quantifier Negation Laws:	Bounded Quantifier Negation Laws:
1) $\neg \forall x P(x) \equiv \exists x \neg P(x)$	1) $\neg \forall x \in \mathcal{U}\, P(x) \equiv \exists x \in \mathcal{U}\, \neg P(x)$
2) $\neg \exists x P(x) \equiv \forall x \neg P(x)$	2) $\neg \exists x \in \mathcal{U}\, P(x) \equiv \forall x \in \mathcal{U}\, \neg P(x)$

- Key: Re-express neg statements as pos using these + other laws - easier!
- Can apply 2+ times in a row: $\neg \forall x \exists y\, P(x,y) \equiv \exists x \neg \exists y\, P(x,y) \equiv \exists x \forall y \neg P(x,y)$.

Clarifying Universe of Discourse: (Notation) - Called 'bounded quantifiers'.

- $\forall x \in \mathcal{U}\, P(x) \equiv$ 'For all x in $\mathcal{U}$, $P(x)$.' Also written as $\forall x \in \mathcal{U}\,(P(x))$.
- $\exists x \in \mathcal{U}\, P(x) \equiv$ 'There exists at least one x in $\mathcal{U}$ so that $P(x)$'. Also $\exists x \in \mathcal{U}\,(P(x))$.

↓

These are really abbreviations:
- $\exists x \in \mathcal{U}\, P(x) \equiv \exists x\,(x \in \mathcal{U} \wedge P(x))$.
- $\forall x \in A\, P(x) \equiv \forall x\,(x \in A \rightarrow P(x))$.

Truth Set Notation Note:

- $\forall x P(x) \equiv$ '$P(x)$ is true for every value in some universe of x's.' Also, $\forall x \in \mathcal{U} P(x)$.
- Similar set notation: $\{x \in \mathcal{U} \mid P(x)\} \equiv$ truth set; all x's in $\mathcal{U}$ where $P(x)$ true.

Empty Set Notes: Let $\mathcal{U} = \emptyset$.

1. $\exists x \in \mathcal{U} P(x)$ always false if $\mathcal{U} = \emptyset$.
2. $\forall x \in \mathcal{U} P(x)$ always (vacuously) true if $\mathcal{U} = \emptyset$. See pg 69 for explanation.
3. $\emptyset \subseteq A$ regardless of A. Also vacuous. (Empty set is a subset of all sets.)

Distributive Laws for Quantifiers:

1) $\forall$ distributes over $\wedge$: $\forall x (P(x) \wedge Q(x)) \equiv \forall x P(x) \wedge \forall x Q(x)$.
2) $\exists$ distributes over $\vee$ (not $\wedge$): $\exists x (P(x) \vee Q(x)) \equiv \exists x P(x) \vee \exists x Q(x)$.

Summary so far:

- Have introduced 7 logical symbols: $\wedge, \vee, \neg, \rightarrow, \leftrightarrow, \forall, \exists$.
- All math statements can be represented & analyzed using these.

Ex: Simplification of negative statements (2.2.1)

1) $\neg A \subseteq B \equiv \neg \forall x (x \in A \rightarrow x \in B)$ by definition of a subset
$\equiv \exists x \neg (x \in A \rightarrow x \in B)$ by quantifier negation law
$\equiv \exists x \neg (x \notin A \vee x \in B)$ by conditional law
$\equiv \exists x (x \in A \wedge x \notin B)$ by DeMorgan's law. $\square$

Ex: Writing some math statements in logical form. (2.2.3)

1. Universe of discourse is $\mathbb{N}$:
   - x is a perfect square $\equiv$ x = square of some natural # $\equiv \exists y (x = y^2)$
   - x is a multiple of y $\equiv$ x = y · some natural # $\equiv \exists z (x = yz)$
   - x is prime $\equiv$ x > 1 and x can't be written as product of 2 smaller natural #s $\equiv$
     $(x>1) \wedge \neg \exists y \exists z (x = yz \wedge (y<x) \wedge (z<x))$
   - x is the smallest number that is a multiple of both y & z.
     $\equiv$ x is a multiple of y and z, and there is no smaller number that is a multiple of y & z
     $\equiv \exists a (x = ya) \wedge \exists b (x = zb) \wedge \exists w \big((w<x) \wedge \exists c (w = yc) \wedge \exists d (w = zd) \big)$
2. Universe of discourse is $\mathbb{R}$:
   - The identity element for addition is 0 $\equiv \forall x (x + 0 = x)$
   - Every real # has an additive inverse $\equiv \forall x \exists y (x + y = 0)$
   - Neg numbers don't have square roots $\forall x (x < 0 \rightarrow \neg \exists y (y^2 = x))$

· Every pos number has exactly 2 square roots (do this in several steps)

$\equiv \forall x\,(x>0 \rightarrow y$ & z are square roots of x, and $y \neq z$, an no other sq. roots of x)

$\equiv \forall x(x>0 \rightarrow \exists y \exists z((y^2=x) \wedge (z^2=x) \wedge (y \neq x) \wedge \neg \exists w((w^2=x) \wedge (w \neq y) \wedge (w \neq z)))$.

Advanced Set Notation:

· So far, we define sets using the 'elementhood test notation': $\{x \mid P(x)\}$

· Can modify notation to allow a more complex expression in place of x, before $\mid$.

Ex: Let S be set of all perfect squares. $\equiv$ S is set of all numbers of form n^2, where $n \in \mathbb{N}$. $\equiv$ $S = \{n^2 \mid n \in \mathbb{N}\}$.

· The following equivalences apply:

1) $\{n^2 \mid n \in \mathbb{N}\} \equiv \{x \mid \exists n \in \mathbb{N}\,(x = n^2)\}$ 'Set of x where there is some natural n so $x = n^2$.'

2) $x \in \{n^2 \mid n \in \mathbb{N}\} \equiv \exists n \in \mathbb{N}\,(x = n^2)$

Indexed Family:

· Used to describe sets where the order of elements matters.

Ex: P is the set of the first 100 prime numbers. Elements are $p_1, p_2, \ldots, p_{99}, p_{100}$.

↳ I is the 'index set' → $I = \{1, 2, \ldots, 100\} = \{i \in \mathbb{N} \mid 1 \leq i \leq 100\}$

↳ P called the 'indexed family'.

↳ Notation: $P = \{p_i \mid i \in I\}$

Generally: · Any indexed family $A = \{x_i \mid i \in I\} \equiv A = \{x \mid \exists i \in I\,(x = x_i)\}$

· Then it follows that $x \in \{x_i \mid i \in I\} \equiv \exists i \in I\,(x = x_i)$

Logical Forms of Set Theory Notation: (Ex 2.3.1)

· Helpful to simplify set theory notation into logical notation.

· $y \in \{\sqrt[3]{x} \mid x \in \mathbb{Q}\} \equiv \exists x \in \mathbb{Q}\,(y = \sqrt[3]{x})$

· $\{x_i \mid i \in I\} \subseteq A \equiv \forall x\,(x \in \{x_i \mid i \in I\} \rightarrow x \in A)$ by definition of a subset, $A \subseteq B \equiv \forall x(x \in A \rightarrow x \in B)$.

$\equiv \forall x\,(\exists i \in I\,(x = x_i) \rightarrow x \in A)$ by filling in $x \in \{x_i \mid i \in I\} \equiv \exists i \in I(x = x_i)$.

· $\{n^2 \mid n \in \mathbb{N}\}$ and $\{n^3 \mid n \in \mathbb{N}\}$ are not disjoint.

$\equiv \exists n \in \mathbb{N}\; \exists m \in \mathbb{N}\,(n^2 = m^3)$

Families of Sets:

· $\mathcal{F}$ is a family of sets, ie a set whose elements are all sets.

· Ex: Let $A = \{1, 2, 3\}$, $B = \emptyset$, and $C = \{4\}$. Then $\mathcal{F} = \{A, B, C\}$.

Note: $A \in \mathcal{F}$, but $4 \notin \mathcal{F}$.

Indexed Families of Sets:

- can extend concept of indexed families to include sets of sets in a given order.
- $\mathcal{F} = \{C_s \mid s \in S\}$

Power Set:

- If A is some set, power set $\mathscr{P}(A)$ is the set whose elements are all the subsets of A.

1) $\mathscr{P}(A) = \{x \mid x \subseteq A\}$
2) Thus, $x \in \mathscr{P}(A) \equiv x \subseteq A \equiv \forall y (y \in x \rightarrow y \in A)$

- For some set X, all subsets of X are elements of $\mathscr{P}(X)$, by definition.
- Ex: Let $A = \{7, 12\}$. Then $\mathscr{P}(A) = \{\phi, \{7\}, \{12\}, \{7, 12\}\}$
- $\mathscr{P}(\phi) = \{\phi\}$, which $\not\equiv \phi$ without braces.

Logical Forms of Power Set Notation: (2.3.3)

START w/ outer symbol & work in.

- $x \in \mathscr{P}(A) \equiv x \subseteq A \equiv \forall y (y \in x \rightarrow y \in A)$ by definition, & by definition of a subset.
- $\mathscr{P}(A) \subseteq \mathscr{P}(B) \equiv \forall x (x \in \mathscr{P}(A) \rightarrow x \in \mathscr{P}(B)) \equiv \forall x [\forall y (y \in x \rightarrow y \in A) \rightarrow \forall y (y \in x \rightarrow y \in B)]$
- $B \in \{\mathscr{P}(A) \mid A \in \mathcal{F}\} \equiv \exists A \in \mathcal{F} (B = \mathscr{P}(A))$. Then $B = \mathscr{P}(A) \equiv \forall x (x \in B \leftrightarrow x \in \mathscr{P}(A))$ bc precisely equal.
  - $\equiv$ Sub in to get $\exists A \in \mathcal{F} \forall x (x \in B \leftrightarrow x \subseteq A)$
  - $\equiv \exists A \in \mathcal{F} \forall x (x \in B \leftrightarrow \forall y (y \in x \rightarrow y \in A))$ by definition of subset.
- $x \in \mathscr{P}(A \cap B) \equiv \forall y (y \in x \rightarrow y \in A \cap B) \equiv \forall y (y \in x \rightarrow (y \in A \land y \in B))$.
- $x \in \mathscr{P}(A) \cap \mathscr{P}(B) \equiv x \in \mathscr{P}(A) \land x \in \mathscr{P}(B) \equiv \forall y (y \in x \rightarrow y \in A) \land \forall y (y \in x \rightarrow y \in B)$.

Power Set Distribution:

1) $\mathscr{P}(A \cap B) \equiv \mathscr{P}(A) \cap \mathscr{P}(B)$ Illustrate these using $A = \{1, 2\}$ & $B = \{2, 3\}$.
2) $\mathscr{P}(A \cup B) \not\equiv \mathscr{P}(A) \cup \mathscr{P}(B)$

Intersection & Union of Families of Sets:

Let $\mathcal{F}$ be a family of sets.

1) $\cap\mathcal{F} = \{x \mid \forall A \in \mathcal{F} (x \in A)\} = \{x \mid \forall A (A \in \mathcal{F} \rightarrow x \in A)\}$
   'Intersection of $\mathcal{F}$', ie elements common to all sets in $\mathcal{F}$.
2) $\cup\mathcal{F} = \{x \mid \exists A \in \mathcal{F} (x \in A)\} = \{x \mid \exists A (A \in \mathcal{F} \land x \in A)\}$
   'Union of $\mathcal{F}$', ie elements in at least one set in $\mathcal{F}$.

- $\cap\mathcal{F}$ is undefined for $\mathcal{F} = \phi$.
- For 2 sets A & B, if $\mathcal{F} = \{A, B\}$, $\cap\mathcal{F} = A \cap B$ and $\cup\mathcal{F} = A \cup B$.

(*) $x \in \cup\mathcal{F} \equiv \exists A \in \mathcal{F} (x \in A)$. $x \in \cap\mathcal{F} \equiv \forall A \in \mathcal{F} (x \in A)$. Also same for indexed, ie $\cup_{i \in I} A_i = \exists i \in I (y \in A_i)$

$\cap_{i \in I} A_i = \forall i \in I (y \in A_i)$