1: **Scatterplot:** form/shape, direction (positive/negative), strength (points follow recognizable form), unusual features (do not fit general trend). **Association:** look at scales (might be misleading). **PC Coeff:** strength of linear association; -1 to 1; 0 means no linear; r does not change if variables are rescaled. ***2:*** **Residual:** observed - predicted; $e_i = y_i - \hat{y}_i$. **Least Squares:** minimize SSE: $\sum_{i=1}^n e_i^2$. **Coefficients:** determined by taking derivative of the SSE and solving for the coeffs when setting to 0. **Interp Coeff:** slope coeff as the pred change in Y associated with a one -unit change in X_i, holding all other predictors constant; intercept as the predicted Y value when all preds are 0 (could be nonsensical). R^2**:** coeff of determination; $= \frac{SSE(\bar{y})-SSE(\hat{y})}{SSE(\bar{y})}$; unexplained variation in y / total var in y (look at points distribution). R^2 **Interp:** $R^2 * 100$ is the percent reduction in SSE by taking into account the predictors; percentage of variation in Y explained by the regression function with predictors; range is 0 to 1 and 1 if perfect predictions; $R^2 = r^2$ for simple linear regression. ***3:*** **Least Squares Estimate of** β**:** $SSE = \sum_{i=1}^n e_i^2 = e^T e$; $\hat{\beta} = (X^TX)^{-1}X^Ty$; $\hat{y} = X\hat{\beta} = Hy$. **Hat Matrix:** $H = X(X^TX)^{-1}X^T$; symmetric and h_{ij} describes the weight each of the values in the ith row of H have on the predicted value of y_i. **Contributions:** $h_{ij}y_j$ is the actual contribution of j makes to the value $\hat{y}_i$. ***4:*** **MSE:** $s^2 = \frac{SSE}{n-p}$; s or RMSE is the typical prediction error; expect 95% of observed y values to lie roughly within $2s$ of predicted values; called **residual standard error in R**. **s vs** R^2**:** both small s and large R^2 is the goal. ***5:*** **Perm. Test:** is there a relation between y and x find test stat that measures association for all possible perms of the resp var and compute the prop. of times an observed test stat as extreme as the one from original sample; p-value from the graph by counting. **p-value:** probability of obtaining a result (test stat) at least as extreme as the one observed, if the null were true; < 0.1 is some, < 0.05 is fairly strong, < 0.01 is very strong, < 0.001 is extremely strong; small p-value means result is unlikely to have occurred by change alone, if the null were true making it statistically significant. **Inference on** β_j**:** $t = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)}$; $1-\alpha$ confidence limits: $\hat{\beta}_j \pm t_{n-p,\alpha/2} SE(\hat{\beta}_j)$. **Standard Error of** β**:** $SE(\hat{\beta}_j) = \sqrt{MSE(C_{j+1}C_{j+1})}$ where C is the jth column of $(X^TX)^{-1}$; `summary(fit)` gives null $= 0$ and alternative $\neq 0$; `model.matrix(fit)` gives design matrix X. **One Sided H:** `pt()` with lower tail false gives area to the right (t > 0) and true gives left tail area. **68/95/99.7 Rule:** values like within 1/2/3 std dev of mean. **t-value:** t > 2 or $<$ -2 means results are significant. **Critical Value:** `qt(p, df)` gives t value for area p to the left of it; 90/95/99 percent CI is p = 0.95/0.975/0.995. **CI Slope Interp:** We are 95% confident that the expected *change* in the *response* for each one *unit* increase in the X_i falls between l and u, (after adjusting for other predictors in the model) **use only when interpreting about the effect of one predictor with multiple predictors, say for conclusions too in inference tests**. **Bonferroni:** $1 - \frac{1-C}{g}$ where C is the confidence level and g is the number of coefficients being tested; instead of $\alpha/2$ in each tail use $\frac{\alpha}{2g}$ in each tail; want each individual confidence to be above C in order for the joint level to be above C; **only for intervals**. **Joint CI Interp:** We are (at least) 95% confident that all intervals correctly capture the population parameters. ***6:*** **FINE Assumptions:** Form, Independence, Normality, Equal Variance; Form: expect linear form, Independence: errors are independent (in data description), Normality: errors follow a normal dist, Equal Variance: variance of errors is the same. **Plots:** Form: residuals vs fitted (no trend/curve) or residuals vs each X_i **for multiple predictors**; Equal Variance: residuals vs fitted (no fan shape); Normality: qq plot (no big departure from straight line) or histogram (no big skew); Independence: look at residuals vs observations number (want to be random). **Formal tests:** wilks $\Rrightarrow$ normality, pagan $\Rrightarrow$ equal variance, low p means violated. ***7:*** **Interp of Slope:** only log transforms can be restated in terms of the original vars, preds can always be restated in terms of original vars. **Transforming** Y: non-linearity, non-constant variance, and non-normality; X: non-linearity, high leverage, influence; changing space = correct problems. **Ladder of Powers:** p = 2, 1, 0.5, 0, -0.5, -1, -2; y* = $y^2, y, \sqrt{y}, 1/\sqrt{y}, 1/y, 1/y^2$; right to become better; log and sqrt not defined for zero or negative values, so transform $y/x + c$, where c makes all values ≥ 1. **Strategies:** skewed residuals: right skew is y down, left skew is y up; residual var inc. as x incr: y down, decr as x incr: y up; non-linear: correct non-normal and unequal var then y, only non-linear then x. **Non-linearity Bulges Point:** up and left is y up or x down; up and right is y up or x up; down and left is y down or x down; down and right is y down or x up. **Interp of X Transform:** If we multiply x_i by b (chosen log base) we predict a change of $\hat{\beta}_i$ in the mean value of y after adjusting for the other vars in model. **Interp of Y Transform:** Each one unit change in x_i changes the predicted median value of y by a factor of $b^{\hat{\beta}_i}$ after .... **Interp of Both:** A c-fold change in x_i changes the predicted median value of y by a factor of $c^{\hat{\beta}_i}$ after .... **Non-log Transform:** Each increase of one in $1/x_i$ is associated with an increase of $\hat{\beta}_i$ in predicted $\sqrt{y}$. **Median:** median instead of mean because $E[log(Y)] \neq log[E(Y)]$ but for median it is true. **Box-Cox:** round lambda to nearest 0.5. **Matrix Scatterplot:** don't reflect preds act jointly.

8A:

- R-sq equation: partitioning variability (sequential sum of squares) SSR, SSE, SSTO, which is R-sq
- Adding vars moves SSE to SSR
- Type 1: order matters, so far R-sq interp include after adjusting for var if after 1st var
- Type 1 SS block of info: anova(fit)
- percent of variation in y explained by adding var to model already containing the vars ...
- Type 2 SS: Anova(fit, type = 2); var entered into model last; additional var in y explained by var after adjusting for all other vars; not sequential

8B:

- degrees of freedom for total, error, reg model; SS divided by df is MS
- adding additional vars can never decrease R-sq
- R-adj-sq (no interp) equation
- General F test properties: p-values found and degrees of freedom
- Partial F-Test: anova(fit.reduced, fit.full) where reduced is the model without vars; null: r of betas is 0, alt: at least one beta is not 0; F stat equation; explain equation; are found
- Model Utility Test: summary(fit) where fit is full model; null: all betas are 0, alt: at least one beta is not 0; F stat equation and explain equation; reduced model is the mean model (no predictors)
- Single Coefficient Test: summary(fit) or Anova(fit, type='II') where fit is full model; null: one beta is 0, alt: not 0; F stat equation; reduced model is full model without 1 var; bold: equivalent to t-test t-sq = F
- Bold: Type 1 ANOVA table contains SST if you sum the SS for all; Type 2 ANOVA does not contain SST because it is not sequential
- Model Utility Test Decision: reject null: sufficient evidence...; conclude that at least one of vars ... is significantly useful in predicting y
- Single Coef Test Decision: large p-val means we do not have enough evidence to to believe that var is any diff from zero after adj...., that is var does not significantly improve the model containing other vars
- Bold: Cannot just drop all variables if not significant, since taking them out would affect significance of other vars, need to come out one at a time or perform partial F test
- Partial F Test Decision: reject null: adding r vars to model that already contains other vars does not significantly improve the model. Do not have enough evidence that the full model containing all vars is better than reduced model. So keep reduced model because of parsimony; larger model achieves a significant reduction in SSE, boost in SSR and R-sq and improves pred of y
- Partial F Null Hypoths: adding vars to smaller to form larger does not significantly improve prediction of y, pop reg coefs for vars in larger not in smaller model are 0, value of pop R-sq for larger model is not greater than value of param for smaller model

10:

11:

12:

Midterm 2 Notes

- hard question is about R^2 because given a bunch of output and need to know where to look; handout 8 anova tables
- 2 of the 3 F tests are on the Midterm
- Nothing about handout 12 really on the midterm
- No VIF problems but on that handout talked about added variable plots which is important
- transforming is back on the midterm (interpretations)
- more plots that we have seen before (harder plots) residuals vs fitted can have more than 1 assumption can be met
- stuff on the sequential sum of squres (handout 8) - focus on this

Midterm 2 Notes: Krish's Class

- lots of tables
- lots of stuff from Handout 8
- stuff on multicollinearity, nothing about cook's distance, not much on handout 12, nothing about leverage
- lots of questions from HW 4, on anova type 1 and type 2, tables, sequential, etc.
- hard question about Anova tables
- know the SS for Anova and know true differnce between type 1 and type 2 and whento use each specifically
- one question from H10 about the intervals and R output (detailed questions about the output), know interpretation of this too
- question on transformation similar to previous Midterm
- hard question on R^2 specifically