__*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**:** $SSE = \sum_{i=1}^{n} e_i^2 = e^T e$; $\hat{\beta} = (X^T X)^{-1} X^T y$; $\hat{y} = X\hat{\beta} = Hy$. **Hat Matrix:** $H = X(X^T X)^{-1} X^T$; 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 $(X^T X)^{-1}$; `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.