**Chapter 6: Multiple Linear Regression**

predictors coefficients: . describes *mean response* per unit increase of while all other variables are held constant. There is no interaction in this model; each term is additive. An int. model is . Models are linear if they are linear with respect to the parameters . Assumes linear rel btwn response & params. Matrix form is distributed .

A number and number symbols

Description automatically generated with medium confidence

Where is , is , is , and is . The least sq estimate seeks to minimize ; fitted values are a hyperplane where is a response surface. ANOVA -test tests if the predictors collectively help explain variation in ; cannot draw conclusions about individual pred; reject ; all coef and at least one coef not .

A white background with black text

Description automatically generatedA group of black text

Description automatically generated

Residuals . These residuals are usually correlated, unlike simple LR where they are uncorrelated.

*Regarding residual plots*: check to see model is correct, normality, constant variance, and independence. Plot against each predictor variable to check adequacy (i.e. if curvature effect is needed); plot against omitted variables to see if they impact the response; plot against interaction terms for potential interaction effects; and plot against fitted values to examine constancy of variance. *Scatterplots*: bivariate scatterplot in a matrix, helps detect nature and strength of bivar relationships, outliers, and range of . *Est Mean Response :* Individual CI : ; simul. CI for vectors : ; W-H conf band for whole line is . Be careful to only work in range of . Pred new obs: same but replace with and Scheffe pred lim for for vectors : . *Adj*  can only go down with more predictors, but is constant. Hence adjusts for add predictors: but often insufficient.

**Chapter 7: General Linear F-test and Multicollinearity**

*General Linear Test*: Consider a full model with predictors and reduced model with predictors (), then where dof is # of extra var. and interchangeable. *Example*: For , (these are the extra var) and At least one var . Then , ; , reject null: at least one of the extra var contains additional useful info to predict in a model w and . If you’re only testing one (), then . *Types of SS*: Type I is stepwise or sequential ss, dependent on var order; Type II doesn’t depend on order:

A close up of a text

Description automatically generated with medium confidenceA black text on a white background

Description automatically generatedA black text on a white background

Description automatically generated

*The coef of partial det* is also given by ; tells us proportionate reduction in variation after other variables are fit. Take sq root to get coef of partial corr ; tells us strength of linear relationship between 2 variables after accounting for all other variables. *Standardized Reg*: Makes all entries in fall btwn [-1, 1] common units, so calculating the inverse becomes less affected by roundoff errors when orders of magnitude are different. *Multicollinearity*: Happens when is almost singular (column lin dep), so taking inverse is difficult and it’s hard to isolate separate influence of each predictor on the response; estimates of reg coefs become unstable and have large standard errors, hard to say whether each predictor is significant or not. However, combined effect of all predictors (even if they’re collinear) can still yield accurate response predictions if testing samples follow the same multicollinear trend. When predictor variables are correlated, the regression coefficient of anyone variable depends on which other predictor variables are included in the model and which ones are left out. Thus, a regression coefficient does not reflect any inherent effect of the particular pred on the response, only a marginal or partial effect. *Extreme cases*: and are uncorr: estimator doesn’t depend on , Type I, II SS the same; each pred contribution is same regardless if other is present. If is a lin comb of : , estimator dne, Type II SS are zero; no contribution of pred if other is present.

**Chapter 8: Quantitative and Qualitative Predictors**

*Poly Reg*: Use when higher order terms are a good approx of true curvilinear response: centralize preds before raising to a power. Tradeoff is potential multicollinearity. *Example:* where and is a *second order model w/ two pred vars* and is the interaction effect coef. *Interaction Models:* an example and this model is not additive because there’s an interaction term. Let to identify 2 groups and is cont: when (group 1) and when (group 2).is group 1 slope and is group 2 slope. Three hypotheses of interest: (reg lines are same), (ints are same), and (slopes are the same). A predictor (race) with (4) classes requires (3) dummy variables:

A screenshot of a calendar

Description automatically generated

*and are cont:* The coef of one explanatory variable depends on other’s value; cannot discuss each predictor individually. represents the change in effect for on for a one unit increase in and vise-versa. If is significant, then there is interaction between and ; if not significant, then no interaction effect (relationship between and does not depend on level of and vise versa).

**Chapter 9: Model Selection**

Given predictors, models may be constructed. Let represent number of potential variables so the reg func containing all potential has parameters and . *Mallow’s Cp:* assumes is an unbiased est of model has no bias. Compares total MSE with variance. If you use all variables, then . Find subsets of variables such that is small and near . Biased models will fall above line ; unbiased will fall near . *Adj R2:* pick model with highest or equivalently, highest . *AIC/BIC:* Find models with small Akaike Info Crit or Bayesian Info Crit; . If is very large, tends to overselect the model but BIC favors smaller models. *PRESSp:* Prediction SS; look for model with small PRESS values because they also have small pred errors; . For each case , omit it and predict observed response using model generated from other cases. *Forward Select:* sig level for entry, begin with null model, for each predictor not in the model, fit a model adding that predictor and choose pred based on best val (only if val ). Add next pred given all previous pred in model (only if val ). Stop based if no var can be added with val . Related to Type II SS bc you're choosing the pred that explains the most additional variability (i.e., largest Type II SS) that isn't alr explained by other pred in the model. *Backward Elim:* sig level for staying in, fit model with all variables, delete var with smallest extra ss (only if val ), delete var with next smallest extra ss (only if val ), stop when all var has val . Related to Type II SS bc you’re assessing marginal contribution (contribution given all other preds) of preds and deleting the ones with the smallest that does not meet a sig level.

**Chapter 10: Model Diagnoses**

*Partial Reg Plots:* For , plot all with vs. all with ; shows strength of marginal relationship given other var in model. Helps detect nonlinearity, heterogenous var, and unusual obs; directly compares and after adjusting for all other var, which partial resid plots don’t do. For , and :

A graph of a slope

Description automatically generated

Fig (a) says contains no add info for pred beyond what’s alr in , so adding is not helpful. Fig (b) suggests adding a linear term to the model with and could be useful. Fig (c) suggests that adding could explain curve relationship. *More residuals:* Residuals may have very different variances in mult lin reg. Deleting the th case from fittings gives us the deleted resid , , and the studentized deleted resid is , which can help us detect outlying observations, but it is not indep. If then obs is an outlier (adjust for tests). is the MSE when the th case is omitted in fitting the reg func. The test is and . *Leverage*: is the th diag of the hat matrix and measures how much is contributing to pred of . . As goes up, goes down, so larger values of force to be closer to . Leverage values indicate value outlying cases. *DFFITS:* Measures influence (impact on reg slope) of case on , . Point is influential if abs val is for small data sets, for large data sets. *Cook’s Dist:* Measures the influence of the th case on all , and if is above the 50th percentile of , then the point associated with it is influential. *DFBETAS:* Measures influence of case on reg coef , where is the th diag element of . Large val = large impact of case on th reg coef. Point is influential if DFBETAS for small data sets, for large data. Outlier = unusual or vals; Influential = large impact on prediction or estimation. *Variance Inflation Factor:* measures how much est reg coef’s variances are inflated as compared to when the pred var’s are not linearly related. when ( is not lin related to other ). If then intercorr among variables so inflated variance for . suggests strong multicollinearity; where is tolerance. If mean is much larger than one, then serious multicollinearity present.

**Chapter 11: Model Remedies**

*Weighted Least Squares:* weight and is a diag matrix of ’s. WLS and max likelihood est is a matrix ; unbiased, consistent, and have minimum var. Use when appropriate reg relationship has been found but variances of error terms are unequal. Generalizes OLS by replacing equal weights of 1 with ; seeks to minimize . *Estimating Weights:* , so squared resid is an estimator of ; We can estimate the variance function describing the relation of to relevant pred vars by first fitting the reg model using unweighted ls and then regressing the squared residuals against the appropriate pred vars. Alternatively, we can estimate the st dev function describing the relation of to relevant pred vars by regressing the absolute resids obtained from fitting the regression model using unweighted ls against the appropriate pred vars. A second method of obtaining estimates of the error term variances can be utilized in designed experiments where replicate obs

arc made at each combination of levels of the pred vars. *Ridge Regression:* multicolin does not affect est mean responses or predictions PROVIDED the values of the predictor variables for which inferences are to be made follow the same multicollinearity pattern as the data on which the regression model is based. Ridge reg minimizes multicolin by modifying ols to allow biased est of reg coef; When an estimator has only a small bias and is substantially more precise than an unbiased estimator, it may well be preferred since it will have a larger probfability of being close to true param. Introduces biasing constant to make inversion less unstable/easier; the trick is selecting the best . Can do this by reading a ridge trace and VIFs: simultaneous plot of param ests for different values of . Curves may fluctuate widely when close to zero but eventually stabilize and slowly converge to 0. VIF's tend to fall quickly as moves away from zero and then change only moderately after that. Choose where reg coefs become stable and when VIF becomes suff small.

A graph of a function

Description automatically generatedA table of numbers with black text

Description automatically generated

For ex, above seems reasonable. *Bootstrap:* est precision of sample ests for nonstandard cases (ie nonconstant error var are est by iteratively reweighted ls); samples with replacement so duplicates and omissions from original sample possible; calculates est reg coef from bootstrap following same procedure as from original sample; repeated large number of times. If is fitted coef and is bootstrap est, then is an est of variability of sampling dist of . Can stop bootstrap when stabilizes fairly reasonably. Use *Fixed X Sampling* when reg func is good fit for data, error terms have constant var, and pred vars can be treated as fixed: resids from original fitting is the sample data. When doubting the above, use *Random X Sampling*: samples () pairs with replacement times.