CHE 320 Spring 2017 Exam 3 Study Guide

Chapter 6

1. Given a multiple linear regression model and sample values, be able to predict the response (6-1)
2. Be able to complete an ANOVA table for significance of regression (simple/multiple) (Table 6-3, 6-3.2)
   (a) Calculate Sum of Squares, Mean Squares, Degrees of Freedom
   (b) Construct F-ratio test for significance of regression
   (c) Draw conclusions using P-value
3. Be able to calculate confidence intervals on model parameters, mean response, and future observations (simple/multiple) (6-2.3, 6-3.2)
4. Understand how to use residual plots to check for model adequacy, and ways to remedy inadequacies (simple/multiple) (6-2.5, 6-3.3)
5. Understand multicollinearity, its effects on the model, and ways to remedy (6-3.3)
6. Be familiar with variable selection techniques (Backward elimination, Forward Selection, All possible regressions) and how they operate (6-4.3)

Chapter 7

1. Understand importance and uses of factorial experiments (7-2)
2. Given a set of data or JMP/Minitab output, be able to: (7-3.1, 7-3.2, 7-3.4)
   (a) Calculate main effects and their standard errors
   (b) Determine a t ratio for coefficient significance
   (c) Test hypotheses with P-values
3. Use residuals to qualitatively evaluate model assumptions (independence, normality) (7-3.3)

Exam 3 Review:

ANOVA Multiple Regression:

	DF	SS	MS	F	P-value
Model	k	$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	SS_m/DF_m	$\frac{MS_m}{MS_E}$	
Error	$N-p$	$\sum(y_i-\hat{y}_i)^2$	SS_E/DF_E		
Total	$N-1$	SS_m+SS_E			

Calculating Main Effects / Std Error of effects:

1) From Data:

Main Effect: $\bar{Y}_{x^+} - \bar{Y}_{x^-}$

SE(effect): $\sqrt{\dfrac{\hat{\sigma}^2}{n\,2^{k-2}}}$

$$\hat{\sigma}^2 = \sum_{i=1}^{2^k} \frac{\hat{\sigma}_i^2}{2^k}$$ (average over i Variances)

	Effect	Effect Estimate	se(effect)	T-ratio	p-value
~~A~~	A			$\frac{A}{se(A)}$	
	B				
	AB				

$$DF = 2^k(n-1)$$

replicates

main effects

2) From Regression:

Variable	Coeff Estimate	se(coeff)	t-ratio	p-val
Int.	β_0	$se(\beta_0)$	$\frac{\beta_0}{se(\beta_0)}$	
X_1	β_1	$se(\beta_1)$		
X_2	β_2	$se(\beta_2)$		
⋮				

$$DF = 2^k(n-1)$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$\text{Effect Est.} = 2 \cdot \beta$$

$$se(\text{effect}) = 2 \cdot se(\beta)$$

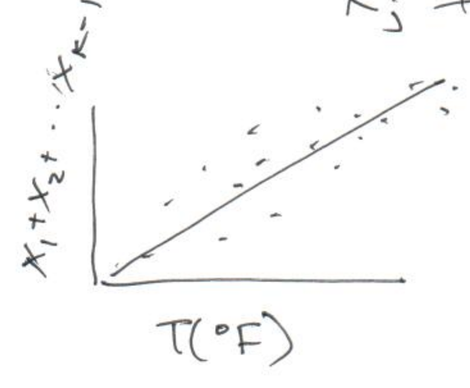
<u>Multicollinearity</u>: when >2 variables are highly correlated and can be linearly predicted from the others

Ex: °F vs °C

$$VIF(\beta_j) = \frac{1}{1 - R_j^2} \qquad j = 1, 2, \ldots k$$

$$\uparrow$$

$R_j^2 \neq R^2 \Rightarrow$ regressing X_j against $k-1$ regressors



VIF > 10 ⇒ multicollinearity

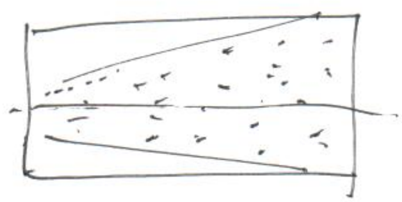
Model Adequacy:

Assumptions:

Normality: 1) Normal prob. plot of residuals

2) Residuals vs $\hat{y}$

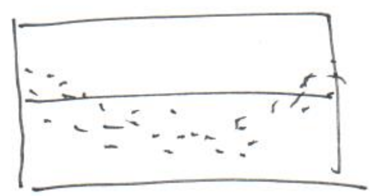
Independence: Residuals vs time/run order



variance ↑ over time

↳ transformations

$\frac{1}{y}$, $\ln(y)$ . . .



model inadequacy

↳ add higher order terms

Existence of Outliers:

- standardize residuals

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2}}$$ → 95% should be within $(-2, +2)$

- studentized residuals (more sensitive)

$$r_i = \frac{e_i}{se(e_i)}$$

- Cook's distance

$$D_i = \frac{r_i^2}{p} \frac{h_{ii}}{(1-h_{ii})}$$ >1 means point was influential