STAT 704 --- Chapter 6: Multiple Regression

• We now consider the situation when several predictors have a linear relationship with the response.

Example: (Two predictors, X_1 and X_2)

If $E(\varepsilon_i) = 0$, then

• This "response surface" is actually a ____________ as a function of X_1 and X_2, not a

• Generally, for k (= p – 1) predictors $X_1, \ldots, X_k$, our model is

where if $E(\varepsilon_i) = 0$,

Interpretations of regression coefficients:

• Again we assume ε_i (i = 1, …, *n*) are independent N(0, σ^2) random variables.

<u>Example</u> (Sec. 6.9) (Portrait Studio company analyzing sales based on data from 21 cities):
***Y* = sales (in thousands of dollars) for a city**
X_1 = number of people (in thousands) age 16 or younger
X_2 = per capita disposable income (in thousands of dollars) of city

• Assuming a linear model is appropriate (should check/verify with data):

Here, does β_0 have a reasonable interpretation?

β_2 is

Situations that the General Linear Model Encompasses

<u>Qualitative Predictors</u>: Example: *Y* = length of hospital stay, X_1 = age of patient, X_2 = gender of patient (X_2 = 1 for females, X_2 = 0 for males).

Note

<u>Polynomial regression</u>: Often appropriate to model a curvilinear relationship between response and predictor(s):

<u>Transformed Variables</u>:

$$\ln(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

<u>Interaction Effects</u>:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

Key: All these models are

Example of nonlinear model:

General Linear Model in Matrix Terms

Let

Then the general linear model can be written in matrix notation as:

Our assumptions about the random error vector $\underline{\varepsilon}$ are:

• Looks complicated, but it makes writing formulas for our least-squares estimates simple.

<u>Fitting the MLR Model</u> (Estimating $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$):
Recall least squares method: Choose estimates of $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ to minimize:

Vector calculus can show that the least-squares estimates are

• This will generally be found using a computer package.

• We can also write the <u>fitted values</u> and the <u>residuals</u> for all observations as vectors:

• **Note that in our matrix notation,**

<u>Example</u>: (Portrait Studio Data)

• **So fitted regression equation is:**

• **Interpretation of b_1:**

• **Interpretation of b_2:**

Analysis of Variance

- **Again, in multiple regression, we can decompose the total sum of squares into SSR and SSE.**
- **Formulas for SSTO, SSR, SSE given in book.**

Degrees of Freedom

- **Still n – 1 d.f. for SSTO**
- **Now, SSE has**

Leaves

ANOVA Table (Multiple Regression)

(Global) F-test for a Regression Relationship

- **In multiple regression, our F-test based on**

tests whether the entire set of predictors $X_1, \ldots, X_k$ explains a significant amount of the variation in Y.

- **If MSR ≈ MSE,**

- **If MSR >> MSE,**

• **Formally, we test:**

If   **we reject H_0 and conclude there is some regression relationship between *Y* and the predictors.**

• **The coefficient of multiple determination**

measures the proportion of sample variation in *Y* explained by its linear relationship with the <u>entire set</u> of predictors $X_1, \ldots, X_k$.

• **Again, $0 \le R^2 \le 1$.**

• **If we keep adding more predictors to our model, R^2 can only**

• **An adjusted R^2**

accounts for the number of predictors in the model.

• **It may decrease when we add useless predictors to the model.**

Note: R^2_a is not always between 0 and 1.

<u>Inferences about Individual Regression Parameters</u>

• The F-test concerns the <u>entire set</u> of predictors.
• If the F-test is "significant" (if we reject H_0), we may want to determine <u>which</u> of the individual predictors contribute significantly to the model.

Expected Values and Variances of Vectors

• If $\underline{Y}$ is a vector, then

• If $\underline{Y}$ is a vector, then

Examples in MLR model:

• **Note: If A is a constant matrix and $\underline{Y}$ is a random vector, then:**

• **So for the j-th estimated coefficient b_j in our model:**

• Then a 100(1 – α)% CI for β_j is

• To test whether X_j is a "significant predictor" in the presence of the other predictors in the model, we test:

using the test statistic

We reject H_0 if

<u>Note</u>: The results of the tests about individual coefficients depend on <u>which other predictors</u> are in the model.
• They therefore determine whether X_j has a significant <u>marginal</u> effect on the response, given that the other predictors are in the model (i.e., above and beyond the effect of the other predictors).

• Each t-test has the correct significance level, assuming it is the only t-test about a coefficient being done.

• If, in an exploratory model, we conduct multiple t-tests about several coefficients, then P[at least one Type I error] will be greater than the nominal α of each test, unless we adjust for multiple tests.

Simplest way: **Bonferroni method:**

More powerful way: **Holm method:**

Example **(Studio data):**

<u>CI for Mean Response and PI for Individual Response in MLR</u>

• We may construct a CI for the mean response corresponding to a <u>set</u> of values of the predictor variables: X_{hi}, …, X_{hk}.

Define

• We wish to estimate

• A point estimator is

• This estimator has expected value

and variance

• Therefore a 100(1 – α)% CI for E(Y_h) is

• The 100(1 – α)% prediction interval for a new response $Y_{h(new)}$ corresponding to $\underline{X}_h$ is

• In practice, we use software to find these intervals.

<u>Example 1</u> (Studio data): We wish to estimate, with a 95% CI, the mean sales in cities with 65.4 thousand people aged 16 or younger and per capita disposable income of 17.6 thousand dollars.

<u>Example 2</u> (Studio data): We wish to predict, with a 95% PI, the sales for a new city with 65.4 thousand people aged 16 or younger and per capita disposable income of 17.6 thousand dollars.