**Multivariate Statistical Analysis, Exercise 8, Fall 2011, Prof. S.K. Jeng**

December 09, 2011 TA: H.C. Cheng

(Linear Regression with Least Squares)

The technique of regression, in particular linear regression, is one of the most popular statistical tool. A major purpose of regression is to explore the dependence of one variable on others. There are all forms of regression: linear, nonlinear, simple, multiple, parametric, nonparametric, etc. The word *regression* is used in statistics to signify a relationship of a target variable (say, $y_i$) between other variables (say, $x_{i,p}$). More generally, we have a relationship of the form

$$y_{n\times 1} = \mathbf{f}(X_{p\times n}) + \epsilon_{n\times 1},$$

where $n$ is the realization of the observable random variables, and the $\mathbf{f}$ is an arbitrary function; $\epsilon_{1\times n}$ is the random vector decribes the deviations or errors. It is also common to suppose it has zeros mean $\mathbb{E}(\epsilon) = 0$; that is

$$\mathbb{E}(y|X) = \mathbf{f}(X).$$

When we refer to *regression that is linear* (more precisely, affine), we mean that the conditional expectation of $Y$ given $X$ is a linear function of $X$ (*i.e.* the term linear regression refers to a specification that is *linear in the parameters*). Hence we have the linear model

$$y_{n\times 1} = X_{n\times p+1}\beta_{p+1\times 1} + \epsilon_{n\times 1} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1,p} \\ 1 & x_{21} & x_{22} & \dots & x_{2,p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{n,p} \end{pmatrix} \beta_{p+1\times 1} + \epsilon_{n\times 1},$$

where the all-ones column represents the intercept. Now we shall derive the estimator for $\beta$, as well as accurately describe what a "good" estimator is. A very common way is to use the least squares criterion; that is

$$\hat{\beta} = \arg\min_{\beta}(Y - X\beta)^T(Y - X\beta) = \arg\min_{\beta}(\epsilon^T\epsilon).$$

We can differentiate $\epsilon^T\epsilon$ to get

$$\hat{\beta} = (X^TX)^{-1}X^Ty,$$

such that

$$\hat{y} = X\hat{\beta},$$

where it can be shown that $X^T X$ is nonsingular if and only if the rank of $X$ equals $p+1$. Otherwire, we can take the pseudo-inverse of $X^T X$.

Now we define the minimum deviations as the *residual sum of squares (RSS)* or the *unexplained variation* by

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

In addition, we define *total variation* as $\sum_{i=1}^{n} (y_i - \overline{y})^2$; explained variation as $\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$. Hence,

$$\begin{aligned} \sum_{i=1}^{n} (y_i - \overline{y})^2 &= \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ \text{total variation} &= \text{explained variation} + \text{unexplained variation}. \end{aligned}$$

The *coefficient of determination* is $r^2$:

$$r^2 = \frac{\text{explained variation}}{\text{total variation}}.$$

The coefficient of determination increases with the proportion of explained variation by the linear relation above. In the extreme cases where $r^2 = 1$, all of the variation is explained by the linear regression. The other extreme, $r^2 = 0$, is where the empirical covariance of $X$ and $y$ is zero. However, the coefficient of determination is influenced by the number of regressors $p$. For a given sample size $n$, the $r^2$ value will increase by adding more regressors into the linear model. The value of $r^2$ may therefore be high even if possibly irrelevant regressors are included. A corrected coefficient of determination for $p$ regressors and a constant intercept ($p$+1 parameters) is

$$r^2_{\text{adj}} = r^2 - \frac{p(1 - r^2)}{n - (p+1)}.$$

Now we consider a data set consisting of 10 measurements of 4 variables. The story: A textile shop manager is studying the sales of "classic blue" pullovers over 10 periods. He uses three different marketing methods and hopes to understand his sales as a fit of these variables using statistics. The variables measured are

$X_1$ : Numbers of sold pullovers,
$X_2$ : Price (in EUR),
$X_3$ : Advertisement costs in local newspapers (in EUR),
$X_4$ : Presence of a sales assistant (in hours per period).

| | Sales | Price | Advert. | Ass. Hours |
|---|---|---|---|---|
| 1 | 230 | 125 | 200 | 109 |
| 2 | 181 | 99 | 55 | 107 |
| 3 | 165 | 97 | 105 | 98 |
| 4 | 150 | 115 | 85 | 71 |
| 5 | 97 | 120 | 0 | 82 |
| 6 | 192 | 100 | 150 | 103 |
| 7 | 181 | 80 | 85 | 111 |
| 8 | 189 | 90 | 120 | 93 |
| 9 | 172 | 95 | 110 | 86 |
| 10 | 170 | 125 | 130 | 78 |

Now please answer the following questions

(a) Find the least square estimate $\hat{\beta}$.

(b) Find the coefficient of determination.

(c) Find the corrected coefficient of determination.