Problem Set 02
Linear Models -- Fall 2015
Due date: 2015-09-30

Problems sets are due at the start of class on the due date. Please hand write or type up and print the solutions; we will not accept e-mail solution sets except in exceptional circumstances. You may discuss problem sets with others, but must write up your own solutions. This means that you should have no need to look at other's final written solutions. Many of these problems come from a variety of textbooks, which are referenced in the problems. These are for citation purposes and not because you will need to consult the text itself (though you may feel free to do so).

1. Consider the case of a simple linear regression (no intercept) with a random design; specifically, assume:

$$x_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1) \tag{1}$$
$$\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma^2) \tag{2}$$
$$\mathbb{E}(\epsilon_i|x) = 0 \tag{3}$$
$$y_i = x_i\beta + \epsilon \tag{4}$$

Define $\widetilde{\beta} = \frac{1}{n}\sum_i x_i y_i$.

(a) Give both an argument and counterargument for using $\widetilde{\beta}$ in lieu of $\widehat{\beta}$. Hint: What is $\frac{1}{n}\sum_i x_i^2$ an estimate of?

(b) Calculate $\mathbb{E}(\widetilde{\beta}|X)$. Is this estimator unbiased when conditioned on X? Is it unbiased when calculating the unconditional expectation?

(c) Compute the unconditional variance of $\widetilde{\beta}$ and compare to the unconditional variance of $\widehat{\beta}$. Which estimator would you rather use?

2. [Fumio Hayashi, pg 32, #6] Prove that, under assumptions I-IV, $Cov(\widehat{\beta}, r|X) = 0$. The covariance can be written as:

$$Cov(\widehat{\beta}, r|X) = \mathbb{E}\left\{ (\widehat{\beta} - \beta)(r - \mathbb{E}(r|X))^t \middle| X \right\} \tag{5}$$

Hint: Use the relations $M\epsilon = r$ and $\widehat{\beta} - \beta = A\epsilon$.

3. [Fumio Hayashi, pg 46, #7] Prove that, under assumptions I-V, the estimator s^2 has variance:

$$\mathbb{V}(s^2|X) = \frac{2\sigma^4}{n-p} \tag{6}$$

4. [Fumio Hayashi, pg 74, #5] In the restricted least squares, the sum of squared residuals is minimized subject to the constraint implied by the null hypothesis $D\beta = d$. Form the Lagrangian

as:

$$\mathcal{L}(b,\lambda) = \frac{1}{2}(y - Xb)^t(Y - Xb) + \lambda^t(Db - d) \tag{7}$$

Where λ is a k-dimensional vector. Let $\widetilde{\beta}$ be the solution to the restricted regression.

(a) Let $\widehat{\beta}$ be the standard (unrestricted) OLS estimator. Show:

$$\widetilde{\beta} = \widehat{\beta} - (X^tX)^{-1}D^t\left[D(X^tX)^{-1}D^t\right]^{-1}(D\widehat{\beta} - d) \tag{8}$$

$$\lambda = \left[D(X^tX)^{-1}D^t\right]^{-1}(D\widehat{\beta} - d) \tag{9}$$

(b) Let $\widetilde{e} = y - X\widetilde{\beta}$, the residuals from the restricted regression. Show that:

$$SSR_R - SSR_U = (\widehat{\beta} - \widetilde{\beta})^t(X^tX)(\widehat{\beta} - \widetilde{\beta}) \tag{10}$$

$$= (D\widehat{\beta} - d)^t\left[D(X^tX)^{-1}D^t\right]^{-1}(D\widehat{\beta} - d) \tag{11}$$

$$= \lambda^tD(X^tX)^{-1}D^t\lambda \tag{12}$$

$$= \widetilde{e}^tP\widetilde{e} \tag{13}$$

In other words, justify each step.

(c) Show how this verifies that the two F-statistics are the same:

$$\frac{(SSR_R - SSR_U)/k}{SSR_U/(n-p)} = \frac{(D\widehat{\beta} - d)^t[D(X^tX)^{-1}D^t]^{-1}(D\widehat{\beta} - d)^t/k}{s^2} \tag{14}$$

5. [Sheather 2009, pg 147] Chateau Latour is widely acknowledged as one of the world's greatest wine estates with a rich history dating back to at least 1638. The Grand Vin de Chateau Latour is a wine of incredible power and longevity. At a tasting in New York in April 2000, the 1863 and 1899 vintages of Latour were rated alongside the 1945 and the 1961 vintages as the best in a line-up of 39 vintages ranging from 1863 to 1999 (*Wine Spectator*, August 31, 2000). Quality of a particular vintage of Chateau Latour has a huge impact on price. For example, in March 2007, the 1997 vintage of Chateau Latour could be purchased for as little as \$159 per bottle while the 2000 vintage of Chateau Latour costs as least \$700 per bottle (`http://www.wine-searcher.com`).

While many studies have identified that the timing of the harvest of the grapes has an important effect on the quality of the vintage, with quality improving the earlier the harvest. A less explored issue of interest is the effect of unwanted rain at vintage time on the quality of icon wine like Chateau Latour. This question addresses this issue.

The Chateau Latour web site (`http://www.chateau-latour.com`) provides a rich source of data. In particular, data on the quality of each vintage, harvest dates and weather at harvest time were obtained from the site for the vintages from 1961 to 2004. An example of the information on weather at harvest time is given below for the 1994 vintage:

> Harvest began on the 13th September and lasted on the 29th, frequently interrupted by storm showers. But quite amazingly the dilution effect in the grapes was very limited ...(`http://www.chateau-latour.com/commentaires/1994uk.html`

Each vintage was classified as having had "unwanted rain at harvest" (e.g., the 1994 vintage) or not (e.g., the 1996 vintage) on the basis of information like that reproduced above. Thus, the data consist of:

> Vintage = year the grapes were harvested
> Quality – on a scale from 1 (worst) to 5 (best) with some half points
> End of harvest – measured as the number days since August 31
> Rain – a dummy variable for unwanted rain at harvest = 1 if yes.

The data can be found at `euler.stat.yale.edu/~tba3/psets/pset02/data/Latour.txt`. The first model considered was:

$$\text{Quality} = \beta_0 + \beta_1 \cdot \text{End of Harvest} + \beta_2 \text{Rain} + \beta_3 \text{End of Harvest} \times \text{Rain} + \epsilon \tag{15}$$

(a) Show that the coefficient of the interaction term in model **??** is statistically significant at the 0.05 level. In other words, show that the rate of change in quality rating depends on whether there has been any unwanted rain at vintage.

(b) Estimate the number of days of delay to the end of harvest it takes to decrease the quality rating by 1 point when there is:

(i) No unwanted rain at harvest

(ii) Some unwanted rain at harvest

(c) Construct a confidence interval for the (total) end of harvest slope variable when there is unwanted rain.