Professor Joyce Jacobsen

Economics 385
Spring Semester 2013-14
March 24, 2014

Suggested Answers for Midterm

1. random, fixed, random, random, fixed, random, fixed, random, fixed, random, random, random

2. a) sample 1 (as n increases, lower variance on the β estimates)

   b) sample 2 (higher leverage, so lower variance on the β estimates)

   c) sample 1 (lower variance of the error, so lower variance on the β estimates)

   d) sample 1 (same reason as for **a**)

3. a) decreases

   b) no, there is no necessary relationship between coefficient signs and data means

   c) negative

4. a) ten percent

   b) note you can actually solve for these answers formally using Bayesian analysis (as discussed in lecture 1); in comparing the two hypotheses in play (H_1: 10% gold vs. H_2: no gold):

   posterior odds ratio, $\frac{\Pr(H_1 \mid \text{data})}{\Pr(H_2 \mid \text{data})} = \frac{\Pr(\text{data} \mid H_1)}{\Pr(\text{data} \mid H_2)} * \frac{\Pr(H_1)}{\Pr(H_2)}$, likelihood ratio times prior odds ratio

   then if we take ratios of the posterior odds ratios as the sample size increases, the prior odds ratio drops out (so long as it is not zero):

   if n = 1, the likelihood ratio is .9/1 = .9
   if n = 10, the likelihood ratio is $(.9)^{10}$ = .35, so 2.6 times less sure that the seller is telling the truth
   if n = 1000, the likelihood ratio is $(.9)^{1000}$ = 0, so 100% sure the seller is lying

   c) It is not necessary to choose between these extreme positions. As shown in **b**, Bayesian analysis allows us to combine prior beliefs with likelihood functions to create posterior probabilities, thereby combining the two ways of forming opinions

5. a) $E(\hat{\beta}_1) = E\left(\frac{1}{n-2}\sum_{i=3}^{n}\frac{y_i - y_2}{x_i - x_2}\right) = \frac{1}{n-2}\sum_{i=3}^{n}\frac{\beta_0 + \beta_1 x_i + E(u_i) - \beta_0 - \beta_1 x_2 - E(u_2)}{x_i - x_2} = \frac{\beta_1}{n-2}\sum_{i=3}^{n}\frac{x_i - x_2}{x_i - x_2} = \frac{\beta_1(n-2)}{(n-2)} = \beta_1$

   b) no, because this is an inefficient method relative to OLS (can't be more efficient, as OLS is BLUE)

   c) no, because it doesn't use all available information in the sample and thus would have higher variance (be less efficient) than an estimate of MSE that used more observations.

   d) greater

6. a) either run $Y = \beta_0 + \beta_2(2X_1 + X_2) + \beta_3(X_3 - X_1)$, or run $Y = \beta_0 + \delta X_1 + \beta_2(2X_1 + X_2) + \beta_3(X_3 - X_1)$

   b) either create an F statistic with the restricted and unrestricted SSRs (or R-squareds) to test the null hypothesis that the constraint is true (equivalent to testing if your friend is telling the exact truth), or use the t statistic for δ to test the null hypothesis that $\delta = 0$

7. a) e.g., years of work experience

   b) e.g., gender (women get more education, but have lower earnings)

   c) e.g., whether or not you have a hidden tattoo; the day of the week you were born

8. a) False. Multiplied by itself, it doesn't return itself (or can answer Uncertain as $X = I$ is a special case where it is idempotent).

   b) True. It shows what proportion of S_{YY} is explained; you can't explain more than 100%

   c) False. It can be used to choose between models where S_{YY} is the same.