Suggested Answers for Midterm Exam. I

Nov. 1, 2002

1a False. Unbiasedness of the estimators does not require a normally distributed population error term. It does require that the error term has an expected value of zero.

1b False. Since $\text{Var}[\hat{\beta}_2] = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$, dropping the middle range of X_i in fact will make $\sum_{i=1}^n (X_i - \bar{X})^2$ smaller.

1c False. Statistical significance and practical significance are different. They don't imply each other.

1d False. A high p-value implies a small t-statistic which means that the coefficient is not significantly different from zero.

1e False. If we force $\hat{\beta}_2$ to be zero and estimate β_1 only, then the residual sum of squares will be greater, R^2 will be smaller. In fact, in this case $\hat{\beta}_1 = \bar{Y}$, $\hat{Y}_i = \bar{Y}$ and $R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 0$.

2a

$$\begin{aligned}\int_0^2 kx(2-x)\,dx &= k\left[x^2 - \frac{x^3}{3}\right]_0^2 = k(4 - \frac{8}{3}) = 1 \\ k &= \frac{3}{4}\end{aligned}$$

2b

$$\begin{aligned}\text{E}[X] &= \int_0^2 xkx(2-x)\,dx = k\int_0^2 2x^2 - x^3 = k\left[\frac{2}{3}x^3 - \frac{1}{4}x^4\right]_0^2 \\ &= k(\frac{16}{3} - 4) = 1 \\ \text{E}[x^2] &= \int_0^2 x^2kx(2-x)\,dx = k\int_0^2 2x^3 - x^4 = k\left[\frac{2}{4}x^4 - \frac{1}{5}x^5\right]_0^2 \\ &= k(8 - \frac{32}{5}) = \frac{3}{4}\cdot\frac{8}{5} = \frac{6}{5} \\ \text{Var}[X] &= \text{E}[X^2] - \text{E}[X]^2 = \frac{6}{5} - 1^2 = \frac{1}{5}\end{aligned}$$

3a $f(x) > 0$ and

$$\int_0^1 f(X)\,dX = \int_0^1 \theta X^{\theta-1}\,dX = \left[X^\theta\right]_0^1 = 1$$

Therefore, $f(x)$ is a proper density function.

3b

$$\begin{aligned} L &= \prod_{i=1}^{n} \theta X_i^{\theta-1} \\ \ln L &= \sum_{i=1}^{n} (\ln\theta + (\theta-1)\ln X_i) \\ &= n\ln\theta + (\theta-1)\sum_{i=1}^{n}\ln X_i \end{aligned}$$

3c Using the log-likelihood function derived above, the first order condition is

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln X_i = 0$$

Therefore, the ML estimator of θ is

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln X_i}.$$

4a $\tilde{\beta}_2$ is the average slope of the $n-1$ lines that connecting the n observations.

4b

$$\begin{aligned} \tilde{\beta}_2 &= \frac{1}{n-1}\sum_{i=2}^{n}\frac{Y_i - Y_{i-1}}{X_i - X_{i-1}} \\ &= \frac{1}{n-1}\sum_{i=2}^{n}\frac{\beta_2(X_i - X_{i-1}) + u_i - u_{i-1}}{X_i - X_{i-1}} \\ &= \beta_2 + \frac{1}{n-1}\sum_{i=2}^{n}\frac{u_i - u_{i-1}}{X_i - X_{i-1}} \\ \mathrm{E}[\tilde{\beta}_2] &= \beta_2 + \frac{1}{n-1}\sum_{i=2}^{n}\frac{\mathrm{E}[u_i] - \mathrm{E}[u_{i-1}]}{X_i - X_{i-1}} \\ &= \beta_2 \end{aligned}$$

4c

$$\begin{aligned}
\mathrm{Var}[\tilde{\beta}_2] &= \mathrm{E}[(\tilde{\beta}_2-\beta_2)^2] = \mathrm{E}\left[\left(\frac{1}{n-1}\sum_{i=2}^{n}\frac{u_i-u_{i-1}}{X_i-X_{i-1}}\right)^2\right] \\
&= \frac{1}{(n-1)^2}\sum_{i=2}^{n}\frac{\mathrm{E}[(u_i-u_{i-1})^2]}{(X_i-X_{i-1})^2} \\
&= \frac{1}{(n-1)^2}\sum_{i=2}^{n}\frac{2\sigma^2}{(X_i-X_{i-1})^2} \to 0
\end{aligned}$$

Since $\mathrm{E}[\tilde{\beta}_2] = 0$ and $\lim_{n\to\infty} \mathrm{Var}[\tilde{\beta}_2] = 0$, $\tilde{\beta}_2$ is consistent.

4d No. Ideally, we have to compare the variances of the two estimator. However, since $\hat{\beta}_2$ is the OLS estimator, according to the Gauss-Markov theorem, $\hat{\beta}_2$ has the smallest variance in the class of linear unbiased estimators. $\tilde{\beta}_2$ can not be more efficient than $\hat{\beta}_2$.

5a $R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{1.0188}{26.4888} = 0.0385.$

5b $F = \frac{1.0188}{.2547} = 4.$

5c The t statistic for $\hat{\beta}_2$ is $\sqrt{F} \doteq 2 > 1.96$. Therefore, it is significant different from zero at a 95% level of confidence.

5d Since $t = \frac{\hat{\beta}_2 - 0}{se(\hat{\beta}_2)}$, $\hat{\beta}_2 = 2 \cdot 0.1 = 0.2$, A 10% increase of the father's wage will increase the son's wage by $10 * 0.2 = 2\%$.