ECON 4135, APPLIED STATISTICS AND ECONOMETRICS

Initial test, 21 August 2006

Answers:

1.

i) Verify that the marginal distribution of *Y* is given by

y	1	2	3
$P(Y=y)$	0.6	0.2	0.2

P(Y=1)=.260.24+.10=.6 osv.

ii) E(*Y*)=.6+2*.2+3*.2=1.6.

iii) No. P(X=1)=.41. $P(X=1,Y=1)=.26 \neq P(X=1)P(Y=1)=.246$

iv) $P(X=Y)$ =.26+.10=.36

v)

P(Y=y|X=x)=P(Y=y,X=x)/P(X=x)

y	1	2	3
$P(Y=y \mid X=1)$	0.63	0.24	0.12
$P(Y=y \mid X=2)$	0.83	0	0.17
$P(Y=y \mid X=3)$	0.33	0.33	0.33

vi) Not a straight line:

x	1	2	3
$\mu(x)=E(Y \mid x)$	1.48	1.34	1.98

.

vii) $\rho = 0.17/\sqrt{0.64 \cdot 0.70} = 0.25$

viii) Let $Z = X - Y$. $E(Z) = E(X) - E(Y) = 1.89 - 1.60 = .29$,

$$\text{var}(Z) = \text{var}(X) + \text{var}(Y) - 2\,\text{cov}(X, Y) = 1.00\,.$$

2. Suppose that the heights of Norwegian males of age 20 are normally distributed with expectation, 180cm, and standard deviation, 6cm, (i.e. N(180, 6^2)).

   i) 0.047
   ii) $180 \pm 1.28 \cdot 6 = [172.3, 187.7]$

3. Let X_1, X_2, ...,X_{36}, be n=30 independent observations of the random variable *X*, which represents the height of a randomly chosen Norwegian male, regardless of age. Let X have expected value μ and standard deviation σ.

   i) The distribution has a long tail towards small values, due to the non-grown ups.

   ii)
   The mean is smaller than 180, the standard deviation is larger than 6.

   iii)
   $E\overline{X} = \mu$, $sd\overline{X} = \sigma / \sqrt{30}$. This is exact. Due to the central limit theorem, $\overline{X}$ is approximately normally distributed with these moments

4. Suppose α is an unknown parameter in an econometric model. Based on a large data set α was estimated by $\hat{\alpha} = 4$ with estimated standard deviation (standard error) 2. It is known that $\hat{\alpha}$ is unbiased, and approximately normally distributed in large samples.

   i)
   4/2 = 2 which exceeds the critical level 1.64 for a two-tailed test. $\hat{\alpha} = 4$ is thus significant at the 10% level.

   ii)
   $\hat{\alpha} \pm 1.96 \cdot 2 = (0.08, 7.92)$

   iii)
   $$p = P(Z < \frac{4 - 7.6}{2}) = 0.035$$

5. Consider the following three regression models. μ denotes the expected value of *Y* given the explanatory variables. The main interest is to study the effect of *x* on *Y*. *U* is the error term with expectation 0, independent of *x* and *z*, and with $\text{var}(U) = \sigma^2$.

   (a) $Y = \mu(x) + U = \alpha + \beta x + xU$
   (b) $Y = \mu(x,z) + U = \beta x + \gamma z + xU$
   (c) $Y = \mu(x,z) + U = \alpha + \beta_1 x + \beta_2 x^2 + \gamma_1 z + \gamma_2 z^2 + xU$

i)
The expected change in *Y* of changing *x* by one unit.
ii)
The quadratic $\beta_1 x + \beta_2 x^2$. The conditional (on *z*) marginal effect dependent on *x*: $\beta_1 + 2\beta_2 x$.

iii)
81% of the variance in *Y* is “explained”: the conditional variance of *Y* given *x* is only 19% of the marginal variance.

iv)
$\text{var}(Y^* \mid x^*) = \text{var}(U) = \sigma^2$, $E(Y^* \mid x^*) = E(Y \mid x)/x = \alpha x^* + \beta$ and the regression model is now homoscedastic and linear. Thus, OLS yields optimal results under normality, which prevails. The interpretation of β remains the same.