

## BUEC 333: Midterm Fall 2010

**RULES:** No notes or other source materials are permitted. All such items must remain on the floor. You may use only: pens, pencils, rulers, erasers, calculators and watches. Be sure to carefully write down all assumptions that you use to answer the question, including any calculations that you may make. Write your NAME and SECTION NUMBER (or Tutorial Leader) on the front page of your exam booklet. There are 25 points: I recommend 4 minutes per point.

**No looking, copying or cheating. All such behaviour will result in a zero for the exam.**

1. [3 points] Consider the sum of three **4-sided dice**. A four-sided die is shaped like a pyramid made of equilateral triangles, and has the numbers 1,2,3,4 on its sides.
  - a. Construct the pdf and cdf for the sum of three 4-sided dice.
  - b. Using the cdf, show the probability of getting a sum in the range 6 to 8.
2. [3 points] Consider the sum of 3 coin flips, where heads is assigned the value 1 and tails is assigned the value 0.
  - a. What is the mean of the sum of 3 coin flips? What proportion of the time would you expect to see the sum of 3 coin flips take on its mean value?
  - b. What is the variance of the sum of 3 coin flips?
3. [1 point] Which of the following is a linear regression model:
  - a.  $Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$
  - b.  $\log(Y_i) = \beta_0 + \beta_1 \log(X_i) + \varepsilon_i$
  - c.  $Y_i = \beta_0 + \beta_1 e^{X_i} + \varepsilon_i$
  - d. all of the above
  - e. none of the above

4. [2 points] Why does  $s^2$  have  $n-1$  in its denominator rather than  $n$ ?

5. [6 points] Let 2 random variables have the following joint pdf, with  $X$  in rows, and  $Y$  in columns:

	$Y=-1$	$Y=0$	$Y=1$
$X=-1$	0.125	0.15	0.125
$X=0$	0	0.2	0
$X=1$	0.125	0.15	0.125

- a. What is the conditional mean of  $Y$  for each value of  $X$ ?
  - b. What is the covariance of  $X$  and  $Y$ ? Are  $X$  and  $Y$  independent?
  - c. What is the probability that  $X=1$ ?
  - d. Assume the even simpler regression model  $Y_i = \beta X_i + \varepsilon_i$ . Given a large number of observations from this joint pdf, what would you expect the estimated coefficient to be?
6. [2 points] Prove that the sample mean is an unbiased estimator of the population mean. Be sure to write down all your assumptions.
7. [2 points] Suppose that  $Z$  is the average of  $n$  iid observations of a random variable  $X$ . Suppose that  $V(X)=1$ . Suppose that  $n$  is very large. Construct a standard normally distributed test statistic for the hypothesis that  $X$  is drawn from a distribution whose mean is 2.
8. [2 points] Consider the even simpler regression model  $Y_i = \beta X_i + \varepsilon_i$ , where  $Y$  earnings in thousands of dollars and  $X$  is age in years. Say that your estimated value of the coefficient is 4, and that the variance of this estimated coefficient is 4.
  - a. What is estimated difference in earnings between a person who is 25 and a person who is 40?
  - b. How would you use the tables in the back of your textbook to test the hypothesis that the slope of earnings with respect to age is zero?
9. [2 points] Suppose that  $X \sim N(0,1)$ ,  $U \sim N(0,1)$  and  $Y=2X+U$ .
  - a. What is the conditional expectation of  $Y$  given that  $X=2$ ?
  - b. What is the distribution of  $Z=X^2+U^2$ ?
10. [2 points] Suppose that  $X \sim N(2,2)$  and  $Y=2X+5$ .
  - a. What is the mean of  $Y$ ? What is the variance of  $Y$ ?
  - b. What is the distribution of  $Y$ ?

**USEFUL FORMULAE**

$$E(X) = \sum_{i=1}^k p_i x_i$$

$$Var(X) = E[(X - \mu_x)^2] = \sum_{i=1}^k (x_i - \mu_x)^2 p_i$$

$$Pr(X = x) = \sum_{i=1}^k Pr(X = x, Y = y_i)$$

$$Pr(Y = y | X = x) = \frac{Pr(X = x, Y = y)}{Pr(X = x)}$$

$$E(Y | X = x) = \sum_{i=1}^k y_i Pr(Y = y_i | X = x)$$

$$E(Y) = \sum_{i=1}^m E(Y | X = x_i) Pr(X = x_i)$$

$$Var(Y | X = x) = \sum_{i=1}^k [y_i - E(Y | X = x)]^2 Pr(Y = y_i | X = x)$$

$$E(a + bX + cY) = a + bE(X) + cE(Y)$$

$$Cov(X, Y) = \sum_{i=1}^k \sum_{j=1}^m (x_j - \mu_x)(y_i - \mu_y) Pr(X = x_j, Y = y_i)$$

$$Var(a + bY) = b^2 Var(Y)$$

$$Corr(X, Y) = \rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X, Y)$$

$$E(Y^2) = Var(Y) + E(Y)^2$$

$$Cov(a + bX + cY, Y) = bCov(X, Y) + cCov(Y, Y)$$

$$E(XY) = Cov(X, Y) + E(X)E(Y)$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$$Z = \frac{X - \mu}{\sigma}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$r_{XY} = s_{XY} / s_X s_Y$$

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}$$

<p>For the simple linear regression model <math>Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i</math> we have: <math>\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}</math>     <math>\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}</math></p>
--

<p>For the even simpler regression model <math>Y_i = \beta X_i + \varepsilon_i</math> we have: <math>\hat{\beta} = \frac{\sum_{i=1}^n (X_i Y_i)}{\sum_{i=1}^n (X_i)^2}</math></p>
---