

CONCORDIA UNIVERSITY
Department of Economics

ECON 222/4 SECTIONS A, B and BB
STATISTICAL METHODS II
WINTER 2016 – MIDTERM 1 (SOLUTIONS)
Sunday, February 21, 1:00pm – 3:00pm

1. (4 marks) Simplify the following expressions.

a. (2 marks) Find $\int_0^1 2x(x^2 + 1) dx$

$$\int_0^1 2x(x^2 + 1) dx = \int_0^1 (2x^3 + 2x) dx = \left[\frac{1}{2}x^4 + x^2 \right]_0^1 = \frac{1}{2} + 1 = \frac{3}{2}$$

b. (2 marks) Find $\frac{dy}{dx}$, where $y = \ln(x^2)$

$$\frac{dy}{dx} = \frac{1}{x^2} \cdot 2x = \frac{2}{x}$$

2. (4 marks) The random variable $Z = aX + bY$ is a linear combination of two other random variables, X and Y , where $E(X) = \mu_x$, $\text{var}(X) = \sigma_x^2$, $E(Y) = \mu_y$ and $\text{var}(Y) = \sigma_y^2$. Prove that...

a. (2 marks) $E(Z) = a\mu_x + b\mu_y$

$$Z = aX + bY \Rightarrow E(Z) = E(aX + bY) \Rightarrow E(aX) + E(bY) \Rightarrow aE(X) + bE(Y) \Rightarrow a\mu_x + b\mu_y$$

b. (2 marks) $\text{var}(Z) = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\text{cov}(X, Y)$

$$\begin{aligned} \text{var}(Z) &= E(Z - E(Z))^2 = E(aX + bY - a\mu_x - b\mu_y)^2 = E(a(X - \mu_x) + b(Y - \mu_y))^2 \\ &= a^2E(X - \mu_x)^2 + b^2E(Y - \mu_y)^2 + 2abE((X - \mu_x)(Y - \mu_y)) = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\text{cov}(X, Y) \end{aligned}$$

3. (10 marks) X and Y are continuous random variables with a joint probability density function (PDF) of $f(x, y) = x + y$ for $x \in [0, 1]$ and $y \in [0, 1]$. To be a valid density function, $f(x, y) \geq 0$ for all $x \in [0, 1]$ and $y \in [0, 1]$ and the total area under the PDF must equal one. The first condition is true.

- a. (2 marks) Prove the second condition (ie, prove that $\int_{y=0}^1 \int_{x=0}^1 f(x, y) dx dy = 1$).

$$\int_{y=0}^1 \int_{x=0}^1 f(x, y) dx dy = \int_{y=0}^1 \int_{x=0}^1 (x + y) dx dy = \int_{y=0}^1 \left[\left(\frac{1}{2} x^2 + xy \right) \Big|_0^1 \right] dy = \int_{y=0}^1 \left(\frac{1}{2} + y \right) dy = \left[\frac{1}{2} y + \frac{1}{2} y^2 \right]_0^1 = \frac{1}{2} + \frac{1}{2} = 1$$

- b. (2 marks) Calculate the marginal PDF of X (ie, calculate $f(x) = \int_{y=0}^1 f(x, y) dy$).

$$f(x) = \int_{y=0}^1 f(x, y) dy = \int_{y=0}^1 (x + y) dy = \left[xy + \frac{1}{2} y^2 \right]_0^1 = x + \frac{1}{2}$$

- c. (2 marks) Calculate the expected value of X .

$$E(X) = \int_{x=0}^1 xf(x) dx = \int_{x=0}^1 x \left(x + \frac{1}{2} \right) dx = \left[\frac{1}{3} x^3 + \frac{1}{4} x^2 \right]_0^1 = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

- d. (2 marks) Calculate the variance of X .

$$\begin{aligned} \text{var}(X) &= E(X^2) - E(X)^2 = \int_{x=0}^1 x^2 f(x) dx - \left(\frac{7}{12} \right)^2 = \int_{x=0}^1 x^2 \left(x + \frac{1}{2} \right) dx - \frac{49}{144} = \left[\frac{1}{4} x^4 + \frac{1}{6} x^3 \right]_0^1 - \frac{49}{144} \\ &= \frac{1}{4} + \frac{1}{6} - \frac{49}{144} = \frac{11}{144} \end{aligned}$$

- e. (2 marks) Calculate the conditional PDF of Y given X (ie, calculate $f(y|x) = \frac{f(x, y)}{f(x)}$).

$$f(y|x) = \frac{f(x, y)}{f(x)} = \frac{x + y}{x + \frac{1}{2}}$$

4. (2 marks) Consider a random variable $V = Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \chi_n^2$, where $Z_i \sim iid N(0, 1)$.

- a. (2 marks) Prove that $E(V) = n$ (ie, prove that the mean of a chi-squared variable is n). (Hint: $\text{var}(Z_i) = E(Z_i^2) - E(Z_i)^2$.)

$$E(V) = E\left(\sum_{i=1}^n Z_i^2 \right) = \sum_{i=1}^n E(Z_i^2) = \sum_{i=1}^n \left[\text{var}(Z_i^2) + E(Z_i)^2 \right] = \sum_{i=1}^n \text{var}(Z_i^2) + \sum_{i=1}^n E(Z_i)^2 = \sum_{i=1}^n 1 + \sum_{i=1}^n 0^2 = n$$

5. (8 marks) A researcher draws a random sample Y_1, Y_2 and Y_3 from a $N(\mu, \sigma^2)$ population. She is deciding between two estimators of μ : $\bar{Y} = \frac{1}{3}Y_1 + \frac{1}{3}Y_2 + \frac{1}{3}Y_3$ and $\tilde{Y} = \frac{1}{2}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3$.

a. (2 marks) Prove that both \bar{Y} and \tilde{Y} are linear unbiased estimators.

$$E(\bar{Y}) = E\left(\frac{1}{3}Y_1 + \frac{1}{3}Y_2 + \frac{1}{3}Y_3\right) = \frac{1}{3}E(Y_1) + \frac{1}{3}E(Y_2) + \frac{1}{3}E(Y_3) = \frac{1}{3}\mu + \frac{1}{3}\mu + \frac{1}{3}\mu = \mu$$

$$E(\tilde{Y}) = E\left(\frac{1}{2}Y_1 + \frac{1}{3}Y_2 + \frac{1}{6}Y_3\right) = \frac{1}{2}E(Y_1) + \frac{1}{3}E(Y_2) + \frac{1}{6}E(Y_3) = \frac{1}{2}\mu + \frac{1}{3}\mu + \frac{1}{6}\mu = \mu$$

b. (2 marks) Calculate the variance of \bar{Y} .

$$\text{var}(\bar{Y}) = \text{var}\left(\frac{1}{3}(Y_1 - \mu) + \frac{1}{3}(Y_2 - \mu) + \frac{1}{3}(Y_3 - \mu)\right) = \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 = \frac{1}{3}\sigma^2$$

c. (2 marks) Calculate the variance of \tilde{Y} .

$$\text{var}(\tilde{Y}) = \text{var}\left(\frac{1}{2}(Y_1 - \mu) + \frac{1}{3}(Y_2 - \mu) + \frac{1}{6}(Y_3 - \mu)\right) = \frac{1}{4}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{36}\sigma^2 = \frac{7}{18}\sigma^2$$

d. (2 marks) **Briefly** explain which estimator is better.

Since both estimators are unbiased, the better estimator is the one with the smaller sampling variance. This is \bar{Y} .

6. (12 marks) A typical student should study two hours for each hour spent in class. This means that, for a three-hour-per-week class, students should do six hours of work outside class. A professor randomly selects eight students from his class and finds that they are studying 1, 3, 4, 4, 6, 6, 8 and 12 hours, respectively.

a. (2 marks) State the null and alternative hypotheses appropriate with testing whether students are studying an average of six hours per week?

$$H_0 : \mu = 6 \text{ and } H_0 : \mu \neq 6$$

b. (2 marks) **Briefly** explain whether a t - or z -statistic is more appropriate.

A t -statistic is better since the population variance is unknown and because the sample is small.

c. (2 marks) Calculate the test statistic.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{8}(1+3+4+4+6+6+8+12) = 5.5$$

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) = \frac{1}{7}(1^2 + 3^2 + 4^2 + 4^2 + 6^2 + 6^2 + 8^2 + 12^2 - 8 \cdot 5.5^2) = \frac{80}{7}$$

$$t_s = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \frac{5.5 - 6}{\sqrt{10/7}} = -0.418$$

d. (2 marks) Assuming that the critical values associated with a 0.10 level of significance are ± 1.895 , **briefly** explain your conclusion.

Since the test statistic lies inside the critical value, do not reject the null hypothesis. There is insufficient evidence to reject the claim that students study an average of six hours per week.

e. (2 marks) Construct a 90-percent confidence interval.

$$\Pr \left(\bar{x} - t_c \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_c \frac{s}{\sqrt{n}} \right) = 0.9 \Rightarrow 5.5 - 1.895 \sqrt{\frac{10}{7}} < \mu < 5.5 + 1.895 \sqrt{\frac{10}{7}} \Rightarrow 3.235 < \mu < 8.765$$

f. (2 marks) **Briefly** explain whether your answer to part (e) supports or contradicts your conclusion in part (d).

The procedure used to calculate the confidence interval should contain the true population value 90 percent of the time. This estimated contains the value $\mu = 6$ in support of the hypothesis test which concluded that there is insufficient evidence to reject the claim that $\mu = 6$.