CONCORDIA UNIVERSITY
Department of Economics

ECON 222/4 SECTIONS A, B and BB
STATISTICAL METHODS II
WINTER 2016 – ASSIGNMENT 1 (SOLUTIONS)
Due: Monday, February 15, before 3:00 pm

1. (**4 marks**) Simplify the following expressions.

   a. Find $\int_0^1 2x^3\,dx$

   $$\int_0^1 2x^3\,dx = \left.\frac{x^4}{2}\right|_0^1 = \frac{1}{2}$$

   b. Find $\frac{dy}{dx}$, where $y = Ae^{(a+bX)}$

   $$\frac{dy}{dx} = Abe^{(a+bX)}$$

2. (**4 marks**) The population mean and variance of the random variable X are μ and σ^2, respectively. Prove that, for a sufficiently large sample size n, the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

   a. $E(\bar{x}) = \mu$

   $$E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}(n\mu) = \mu$$

   b. $\text{var}(\bar{x}) = \frac{1}{n}\sigma^2$

   $$\text{var}(\bar{x}) = \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\text{var}\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{var}(x_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

3. (**6 marks**) Let X be a random variable with a probability density function (PDF) given by

$$f(x) = \begin{cases} cx^2, \text{ if } |x| \le 1 \\ 0, \text{ otherwise} \end{cases}$$

a. Solve for c.

$$\int_{-1}^{1} f(x)dx = 1 \Rightarrow \int_{-1}^{1} cx^2 dx = 1 \Rightarrow \left.\frac{cx^3}{3}\right|_{-1}^{1} = 1 \Rightarrow \frac{c}{3} - \left(-\frac{c}{3}\right) = 1 \Rightarrow \frac{2c}{3} = 1 \Rightarrow c = \frac{3}{2}$$

b. Calculate $E(X)$.

$$E(X) = \int_{-1}^{1} xf(x)dx = \int_{-1}^{1} \frac{3}{2}x^3 dx = \left.\frac{3x^4}{8}\right|_{-1}^{1} = \frac{3}{8} - \frac{3}{8} = 0$$

c. Calculate $\operatorname{var}(X)$.

$$\operatorname{var}(X) = E(X^2) - E(X)^2 = \int_{-1}^{1} x^2 f(x)dx - 0^2 = \int_{-1}^{1} \frac{3}{2}x^4 dx = \left.\frac{3x^5}{10}\right|_{-1}^{1} = \frac{3}{10} - \left(-\frac{3}{10}\right) = \frac{3}{5}$$

4. (**2 marks**) A random sample of n voters is selected to estimate the proportion of voters who plan to vote for Candidate A in an election, θ. How large does n need to be so that we can obtain a 90 percent confidence interval with a margin of error of ± 0.03.

$$\Pr\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < \theta < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1-\alpha \Rightarrow ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = z_{\alpha/2}\sqrt{\frac{0.5(1-0.5)}{n}}$$

$$\Rightarrow n = \left(\frac{0.5 z_{\alpha/2}}{ME}\right)^2 = \left(\frac{0.5 \cdot 1.645}{0.03}\right)^2 = 751.67 \Rightarrow n^* = 752$$

5. (**10 marks**) Let X and Y be two continuous variables with a joint PDF given by

$$f(x,y) = \begin{cases} 6xy, 0 \le x \le 1;\ 0 \le y \le \sqrt{x} \\ 0, \text{otherwise} \end{cases}$$

a. Calculate the marginal PDF of X.

$$f_X(x) = \int_0^{\sqrt{x}} 6xy\,dy = 3xy^2\Big|_0^{\sqrt{x}} = 3x^2$$

b. Calculate the marginal PDF of Y.

$$f_Y(y) = \int_0^1 6xy\,dx = 3x^2y\Big|_0^1 = 3y$$

c. Briefly explain if X and Y are independent.

X and Y are independent $\Leftrightarrow f(x,y) = f_x(x) f_y(y)$.

But, $6xy = f(x,y) \ne f_X(x) f_Y(y) = 3x^2 \cdot 3y = 9x^2y$. *Therefore, X and Y are not independent.*

d. Calculate $E(X|Y)$.

$$E(X|Y) = \int_0^1 xf(x|y)\,dx = \int_0^1 x\frac{f(x,y)}{f_Y(y)}\,dx = \int_0^1 x\frac{6xy}{3y}\,dx = \int_0^1 2x^2\,dx = \frac{2x^3}{3}\Bigg|_0^1 = \frac{2}{3}$$

e. Calculate $\text{var}(X|Y)$.

$$E(X^2|Y) = \int_0^1 x^2 f(x|y)\,dx = \int_0^1 x^2\frac{f(x,y)}{f_Y(y)}\,dx = \int_0^1 x^2\frac{6xy}{3y}\,dx = \int_0^1 2x^3\,dx = \frac{x^4}{2}\Bigg|_0^1 = \frac{1}{2}$$

$$\text{var}(X|Y) = E(X^2|Y) - \left[E(X|Y)\right]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

6. (**10 marks**) The data file *assignment.xlsx* contains the grades for 33 students on assignment 1 (X_1) and assignment 2 (X_2). Let $d_i = X_{1i} - X_{2i}$ be normally and independently distributed with a mean and variance of μ and σ^2, respectively.

   a. Calculate $E(d)$.

   $E(d) = E(X_1) - E(X_2) = \mu - \mu = 0$

   b. Calculate $\text{var}(d)$.

   X_1 and X_2 are independent $\Leftrightarrow \text{cov}(X_1, X_2) = 0$

   $\text{var}(d) = \text{var}(\bar{X}_1) + \text{var}(\bar{X}_2) + \text{cov}(\bar{X}_1, \bar{X}_2) = \text{var}(\bar{X}_1) + \text{var}(\bar{X}_2) = \dfrac{\sigma^2}{n} + \dfrac{\sigma^2}{n} = \dfrac{2\sigma^2}{n}$

   c. State the appropriate null and alternative hypotheses to test whether the performance on the assignments does not differ.

   $H_0 : \bar{d} = 0$ and $H_1 : \bar{d} \neq 0$

   d. Briefly explain whether a *t*- or *Z*-test is more appropriate.

   The t-test is more appropriate because the population variance is unknown.

   e. Perform the appropriate test at the 5-percent level of significance and briefly explain your conclusion.

$$\bar{d} = \frac{1}{33}\sum_{i=1}^{33} d_i = \frac{1}{33}\left(\sum_{i=1}^{33} X_{1i} - \sum_{i=1}^{33} X_{2i}\right) = \frac{1}{33}\left(2680 - 2418.\overline{18}\right) = 7.9339$$

$$s_d^2 = \frac{\sum_{i=1}^{33}\left(d_i - \bar{d}\right)^2}{n-1} = \frac{\sum_{i=1}^{33} d_i^2 - n\bar{d}^2}{n-1} = \frac{15080.3471 - 33 \cdot 7.9339^2}{33-1} = 406.3472$$

   Since the t-statistic ($t_{STAT} = \dfrac{\bar{d}}{\sqrt{s_d^2/n}} = \dfrac{81.2121 - 73.2782}{\sqrt{406.3472/33}} = 2.261$ *) exceeds the critical value (* $t_{0.025,32} = 2.037$ *), reject the null hypothesis in favour of the alternative. That means that there is enough evidence to suggest that the performance on the assignments differs.*

7. (**10 marks**) The data file *fultonfish.dat* shows the daily sales of fish (in pounds) for a period of time.

a. Test $H_0 : \mu \geq 10{,}000$ against $H_0 : \mu < 10{,}000$ at the 5-percent level of significance. Briefly explain your result using a diagram showing the estimated value of the test statistic and the critical value.

Since the z-statistic ($z_{STAT} = \dfrac{\bar{X} - \mu}{\sqrt{s^2/n}} = \dfrac{6334.\dot{6} - 10000}{\sqrt{16322567/111}} = -9.5583$ *) is substantially lower than the critical value* $z_{0.025} = 1.96$ *), reject the null hypothesis in favour of the alternative. This means that there is enough evidence to suggest that the daily sales of fish is less than 10,000 pounds. (Note that we use a z-statistic even though the population variance is unknown because of the large sample size.)*

b. Calculate the *p*-value of the test statistic and briefly explain how it can be used to perform the hypothesis test. Show the *p*-value in the diagram.

The p-value is the area under a standard normal curve from $-\infty$ *to the calculated z-statistic. This value equals zero and confirms that there is zero implies that there zero probability that the null hypothesis could be true while observing this z-statistic or one more extreme.*

Let total weekly sales be given by $W = \sum_{i=1}^{5} X_i$, where $X_i \sim iid\ N(\mu, \sigma^2)$ represents sales on weekday *i*.

c. Calculate $E(W)$.

$$E(W) = E\left(\sum_{i=1}^{5} \bar{X}_i\right) = \sum_{i=1}^{5} E(\bar{X}_i) = \sum_{i=1}^{5} \mu = 5\mu$$

d. Calculate $\text{var}(W)$.

$$\text{var}(W) = \text{var}\left(\sum_{i=1}^{5} \bar{X}_i\right) = \sum_{i=1}^{5} \text{var}(\bar{X}_i) = \sum_{i=1}^{5} \frac{\sigma^2}{n} = \frac{5\sigma^2}{n}$$

e. Derive the probability distribution of $\hat{\mu} = \bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \bar{X}_4 + \bar{X}_5$ and calculate a 95-percent confidence interval estimate for μ.

$\hat{\mu} \sim N\left(5\mu, \dfrac{5\sigma^2}{n}\right)$. *The 95-percent confidence interval estimate for* μ *is* $\hat{\mu} \pm 1.96\left(\dfrac{5s^2}{n}\right)$.

8. (**6 marks**) A police chief claims that the standard deviation in the length of response times is less than 3.7 minutes. A random sample of 9 response times from a normal population has a standard deviation of 3.0 minutes.

a. State the appropriate null and alternative hypotheses.

$H_0 : \sigma^2 \geq 3.7 \text{ and } H_1 : \sigma^2 < 3.7$

b. Briefly explain whether a *t*- or χ^2 -statistic is more appropriate.

The χ^2 *-statistic is more appropriate because we are testing variance.*

c. Perform the appropriate test for $\alpha = 0.05$ and briefly explain your conclusion.

Since the test statistic, $\chi^2_{STAT} = \frac{(n-1)s^2}{\sigma^2} = \frac{(9-1)\cdot 3.0^2}{3.7^2} = 5.259$ *, exceeds the critical value,* $\chi^2_{n-1,1-\alpha} = 2.733$*, we cannot reject the null hypothesis. Therefore, the police chief's claim is rejected.*