CONCORDIA UNIVERSITY
Department of Economics

ECON 222 SECTIONS A, B
STATISTICAL METHODS II
Fall 2018 – ASSIGNMENT 1
Due: Tuesday, October 9th, by 4:00 pm (in the instructor's mailbox)

1. **(4 marks)** Calculate the first derivative for the following functions.

a. **(2 marks)** $f(x) = \ln(x^2 + 3x)$

$f'(x) = \frac{2x+3}{(x^2+3x)}$

b. **(2 marks)** $f(x) = \frac{1}{\sqrt{x}}$ $\Rightarrow f(x) = (x)^{-\frac{1}{2}}$

$f'(x) = -\frac{1}{2}(x)^{-\frac{3}{2}} \Rightarrow f'(x) = \frac{-1}{x^{3/2}}$

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

a. **(2 marks)** $E(\bar{x}) = \mu$

$E(\bar{X}) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum E(X_i)$

$= \frac{1}{n}\sum \mu = \frac{n\mu}{n} = \mu$

$\Rightarrow E(\bar{X}) = \mu$

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

$Var(\bar{X}) = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum Var(X_i)$ $\because$ X_i's are independent

$\Rightarrow Var(\bar{X}) = \frac{1}{n^2}\sum \sigma^2 = \frac{n\sigma^2}{n^2}$

$\Rightarrow Var(\bar{X}) = \frac{\sigma^2}{n}$

EXERCISE B.13

(a) The volume under the joint *pdf* is

$$\int_0^2 \int_0^y \left(\frac{1}{2}\right) dx\, dy = \int_0^2 \left[\frac{x}{2}\Big|_0^y \right] dy = \int_0^2 \left(\frac{y}{2}\right) dy = \frac{y^2}{4}\Big|_0^2 = 1$$

(b) The marginal *pdf* for X is

$$f(x) = \int_x^2 \left(\frac{1}{2}\right) dy = \frac{y}{2}\Big|_x^2 = 1 - \frac{x}{2}$$

The marginal *pdf* for Y is

$$f(y) = \int_0^y \left(\frac{1}{2}\right) dx = \frac{x}{2}\Big|_0^y = \frac{y}{2}$$

(c)
$$P\left(X < \tfrac{1}{2}\right) = \int_0^{1/2} \left(1 - \tfrac{x}{2}\right) dx = \left(x - \frac{x^2}{4}\right)\Big|_0^{1/2} = \frac{1}{2} - \frac{1}{16} = \frac{7}{16}$$

(d) The *cdf* for Y is

$$F(y) = \int_0^y \left(\frac{t}{2}\right) dt = \frac{t^2}{4}\Big|_0^y = \frac{y^2}{4}$$

(e) The conditional *pdf* $f(x \mid y)$ is given by

$$f(x \mid y) = \frac{f(x,y)}{f(y)} = \frac{1/2}{y/2} = \frac{1}{y} \qquad \text{implying} \qquad f\left(x \middle| Y = \tfrac{3}{2}\right) = \frac{2}{3}$$

The required probability is

$$P\left(X < \tfrac{1}{2} \middle| Y = \tfrac{3}{2}\right) = \int_0^{1/2} \left(\tfrac{2}{3}\right) dx = \left(\frac{2x}{3}\right)\Big|_0^{1/2} = \frac{1}{3}$$

X and Y are not independent because $P\left(X < \tfrac{1}{2} \middle| Y = \tfrac{3}{2}\right) \neq P\left(X < \tfrac{1}{2}\right)$.

(f) The mean of Y is

$$E(Y) = \int_0^2 y\,f(y)\,dy = \int_0^2 \left(\frac{y^2}{2}\right) dy = \frac{y^3}{6}\bigg|_0^2 = \frac{4}{3}$$

The second moment of Y is

$$E(Y^2) = \int_0^2 y^2 f(y)\,dy = \int_0^2 \left(\frac{y^3}{2}\right) dy = \frac{y^4}{8}\bigg|_0^2 = 2$$

The variance of Y is

$$\text{var}(Y) = E(Y^2) - [E(Y)]^2 = 2 - \left(\tfrac{4}{3}\right)^2 = \frac{2}{9}$$

(g) From part (e),

$$E(X \mid Y) = \int_0^y x f(x \mid y)dx = \int_0^y \left(\frac{x}{y}\right) dx = \frac{x^2}{2y}\bigg|_0^y = \frac{y}{2}$$

$$E(X) = E_Y\left[E(X \mid Y)\right] = \int_0^2 \left(\frac{y}{2}\right) f(y)dy = \int_0^2 \left(\frac{y^2}{4}\right) dy = \frac{y^3}{12}\bigg|_0^2 = \frac{2}{3}$$

We can check this result by using the marginal *pdf* for X to find $E(X)$:

$$E(X) = \int_0^2 x f(x)dx = \int_0^2 \left(x - \frac{x^2}{2}\right) dx = \left(\frac{x^2}{2} - \frac{x^3}{6}\right)\bigg|_0^2 = 2 - \frac{4}{3} = \frac{2}{3}$$

e. **(2 marks)** Calculate the conditional mean and variance of X given $Y = 0.5$.

See the previous page

f. **(2 marks)** ***Briefly*** explain whether X and Y are independent.

see the previous page.

4. **(2 marks)** Let X have a normal distribution with mean μ and variance σ^2. If $Y = aX + b$, where (a and b are constants), calculate the probability density function of Y.

$E(Y) = E(aX+b) = a\,E(X) + b = a\mu + b$

$Var(Y) = Var(aX+b) = a^2 Var(X) = a^2\sigma^2$

$\because X \sim$ Normal distribution & Y is a linear function of $X \Longrightarrow Y \sim$ Normal

$\Longrightarrow Y \sim N(a\mu+b, a^2\sigma^2)$

Q5 (4-Marks)

EXERCISE C.5

2 (a) We set up the hypotheses $H_0: \mu \le 170$ versus $H_1: \mu > 170$. The alternative is $H_1: \mu > 170$ because we want to establish whether the mean monthly account balance is more than 170.

The test statistic, given H_0 is true, is:

$$t = \frac{\bar{X} - 170}{\hat{\sigma}/\sqrt{N}} \sim t_{(399)}$$

The rejection region is $t \ge 1.649$. The value of the test statistic is

$$t = \frac{178 - 170}{65/\sqrt{400}} = 2.462$$

Since $t = 2.462 > 1.649$, we reject H_0 and conclude that the new accounting system is cost effective.

2 (b) $$p = P\left[t_{(399)} \ge 2.462\right] = 1 - P\left[t_{(399)} < 2.462\right] = 0.007$$

Q6 (8-Marks)

EXERCISE C.1

2 (a) A linear estimator is one that can be written in the form $\sum a_i Y_i$ where a_i is a constant. Rearranging Y^* yields,

$$Y^* = \frac{Y_1 + Y_2}{2} = \frac{1}{2}Y_1 + \frac{1}{2}Y_2 = \sum_{i=1}^{2} \frac{1}{2} Y_i$$

Thus, Y^* is a linear estimator where $a_i = 1/2$ for $i = 1,2$ and $a_i = 0$ for $i = 3,4,\ldots,N$.

2 (b) The expected value of an unbiased estimator is equal to the true population mean.

$$E(Y^*) = E\left(\frac{Y_1 + Y_2}{2}\right) = \frac{1}{2}E(Y_1) + \frac{1}{2}E(Y_2) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

2 (c) The variance of Y^* is given by

$$\begin{aligned}
\text{var}(Y^*) &= \text{var}\left(\frac{Y_1 + Y_2}{2}\right) = \text{var}\left(\frac{1}{2}Y_1 + \frac{1}{2}Y_2\right) \\
&= \left(\frac{1}{2}\right)^2 \text{var}(Y_1) + \left(\frac{1}{2}\right)^2 \text{var}(Y_2) + 2\frac{1}{2}\frac{1}{2}\text{cov}(Y_1, Y_2) \\
&= \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{\sigma^2}{2} \qquad \text{since } \text{cov}(Y_1, Y_2) = 0
\end{aligned}$$

2 (d) The sample mean is a better estimator because it uses more information. The variance of the sample mean is σ^2/N which is smaller than $\sigma^2/2$ when $N > 2$, thus making it a better estimator than Y^*. In general, increasing sample information reduces sampling variation.

Q:7 (6 Marks)

EXERCISE C.20

2

(a) The test statistic for testing $\sigma^2_{ADV} = 2500$ against the alternative $\sigma^2_{ADV} > 2500$ is

$$V = \frac{(N_{ADV}-1)\hat{\sigma}^2_{ADV}}{2500} \sim \chi^2_{(N_{ADV}-1)}$$

The rejection region is $\chi^2_{(256)} \geq \chi^2_{(0.95,256)} = 294.32$. The value of the test statistic is

$$V = \frac{256 \times 44.78422^2}{2500} = 205.38$$

The *p*-value is $P\left(\chi^2_{(256)} > 205.38\right) = 0.991$. We do not reject $H_0: \sigma^2_{ADV} = 2500$. There is insufficient evidence to conclude that the variance of incomes in advanced-degree households is greater than 2500.

2

(b) The test statistic for testing $\sigma^2_{ADV} = 2500$ against the alternative $\sigma^2_{ADV} < 2500$ is

$$V = \frac{(N_{ADV}-1)\hat{\sigma}^2_{ADV}}{2500} \sim \chi^2_{(N_{ADV}-1)}$$

The rejection region is $\chi^2_{(256)} \leq \chi^2_{(0.05,256)} = 219.95$. The value of the test statistic is

$$V = \frac{256 \times 44.78422^2}{2500} = 205.38$$

The *p*-value is $P\left(\chi^2_{(256)} < 219.95\right) = 0.009$. We reject $\sigma^2_{ADV} = 2500$ in favor of the alternative $H_1: \sigma^2_{ADV} < 2500$. There is evidence to conclude that the variance of incomes in advanced-degree households is less than 2500.

2

(c) The test statistic for testing $\sigma^2_{ADV} = 2500$ against the alternative $\sigma^2_{ADV} \neq 2500$ is

$$V = \frac{(N_{ADV}-1)\hat{\sigma}^2_{ADV}}{2500} \sim \chi^2_{(N_{ADV}-1)}$$

The rejection region is $\chi^2_{(256)} \leq \chi^2_{(0.025,256)} = 213.57$ or $\chi^2_{(256)} \geq \chi^2_{(0.975,256)} = 302.21$. The value of the test statistic is

$$V = \frac{256 \times 44.78422^2}{2500} = 205.38$$

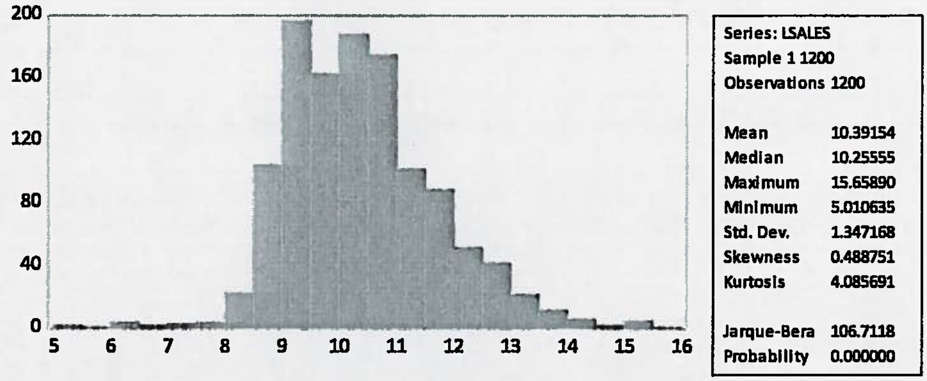
We reject $\sigma^2_{ADV} = 2500$ in favor of the alternative $H_1: \sigma^2_{ADV} \neq 2500$. There is evidence to conclude that the variance of incomes in advanced-degree households is not equal to 2500.

Q8 (8-Marks)

EXERCISE C.22

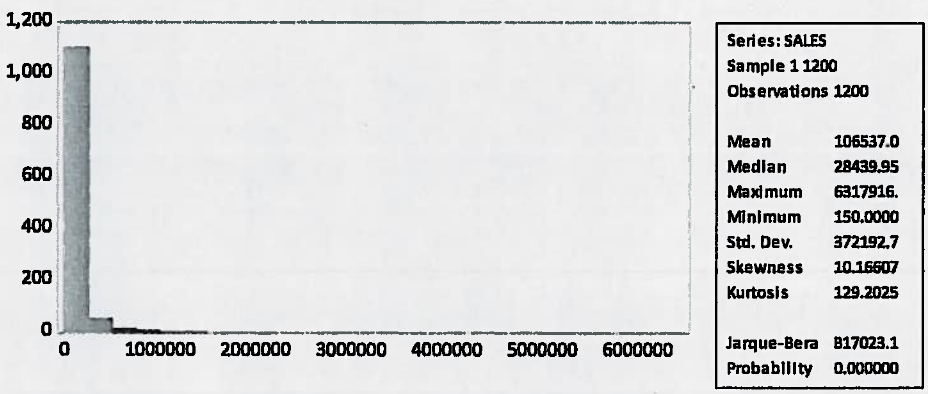
2

(a) A histogram for *LSALES*, the value of the Jarque-Bera (*JB*) test statistic, and its *p*-value, are displayed below. The 10% critical value is $\chi^2_{(0.9,\,2)} = 4.605$. Given that $JB = 106.7$ exceeds the critical value (or the *p*-value = 0.0000 < 0.1), we reject a null hypothesis that *LSALES* is normally distributed.



3

(b) A histogram for *SALES*, the value of the Jarque-Bera (*JB*) test statistic, and its *p*-value, are displayed below. Because there are a small number of very large firms, the observations are concentrated at the left and there is a long tail in the distribution. The 10% critical value is $\chi^2_{(0.9,\,2)} = 4.605$. Given that $JB = 817{,}023$ exceeds the critical value (or the *p*-value = 0.0000 < 0.1), we reject a null hypothesis that *SALES* is normally distributed. Neither *SALES* nor *LSALES* can be described as normally distributed, but *LSALES* more closely resembles a normal distribution.



2

(c) An estimate of $\mu_1 - \mu_0$ is

$$\bar{y}_1 - \bar{y}_0 = 11.52219 - 10.18282 = 1.33937$$

We estimate that sales from firms who export are 134% higher than sales from firms who do not export.

2

(d) To test $H_0:\mu_1=\mu_0$ versus the alternative $H_1:\mu_1\neq\mu_0$, it is useful to do some preliminary calculations

$$\widehat{\operatorname{var}}(\bar{y}_1)=\frac{\hat{\sigma}_1^2}{N_1}=\frac{1.48038^2}{187}=0.01171938 \qquad \widehat{\operatorname{var}}(\bar{y}_0)=\frac{\hat{\sigma}_0^2}{N_0}=\frac{1.211444^2}{1013}=0.00144876$$

Then, the value of the test statistic is

$$t=\frac{\bar{y}_1-\bar{y}_0}{\sqrt{\dfrac{\hat{\sigma}_1^2}{N_1}+\dfrac{\hat{\sigma}_0^2}{N_0}}}=\frac{1.33937}{\sqrt{0.01171938+0.00144976}}=11.67$$

The degrees of freedom are

$$df=\frac{\left(\hat{\sigma}_1^2/N_1+\hat{\sigma}_0^2/N_0\right)^2}{\dfrac{\left(\hat{\sigma}_1^2/N_1\right)^2}{N_1-1}+\dfrac{\left(\hat{\sigma}_0^2/N_0\right)^2}{N_0-1}}=\frac{(0.01171938+0.00144876)^2}{\dfrac{(0.01171938)^2}{186}+\dfrac{(0.00144876)^2}{1012}}=234$$

Because $t=11.67>t_{(0.995,234)}=2.597$, we reject $H_0:\mu_1=\mu_0$ at the 1% significance level. We can conclude that the means of log sales for firms who export and firms who do not export are not equal.

Test for Equality of Means of LSALES
Categorized by values of EXPORT
Sample: 1 1200
Included observations: 1200

Method	df	Value	Probability
Satterthwaite-Welch t-test*	234.1718	-11.67182	0.0000