CONCORDIA UNIVERSITY
Department of Economics

ECON 221 SECTIONS DD, E, F and G

STATISTICAL METHODS I

Winter 2018 – MIDTERM EXAM 2

March 24, 9:30 am – 11:30 am

Name: Solutions	**I.D:**	**Section:**

Total: 52 marks

INSTRUCTIONS

- ☐ This is a two-hour exam.
- ☐ This paper is graded out of 52 marks.
- ☐ The examination comprises six (6) problems. You should attempt **ALL** questions.
- ☐ All answers should be written in the spaces provided. You may use the back pages for rough work.
- ☐ You may not tear pages from the examination paper package, it must be returned intact at the end of the examination.
- ☐ Statistical tables are provided.
- ☐ You are allowed to use a non-programmable calculator.
- ☐ Notes or formula crib-sheets are **NOT** allowed. Please use a pen to provide your answers.
- ☐ Write your name clearly at the top of the first page.
- ☐ **Round to 4 decimal places where necessary.**

FOR EXAMINERS' USE ONLY							
Question	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Marks							

Q1 (10 marks) A random variable X has the following probability distribution:

$X = x$	4	6	8
$P(X = x)$	1/3	1/3	1/3

a. **(2 marks)** Calculate the mean, μ, and variance, σ^2.

$$\mu = E(X) = \sum x\,p(x) = 4 \cdot \tfrac{1}{3} + 6 \cdot \tfrac{1}{3} + 8 \cdot \tfrac{1}{3} = \underline{6}$$

$$\sigma^2 = E(x^2) - [E(x)]^2 = \sum x^2 p(x) - [E(x)]^2$$

$$\sum x^2 p(x) = 4^2 \cdot \tfrac{1}{3} + 6^2 \cdot \tfrac{1}{3} + 8^2 \cdot \tfrac{1}{3} = 38.67$$

$$\Rightarrow \sigma^2 = 38.67 - 6^2 = \underline{2.67}$$

Now consider all the possible samples (sampling with replacement) of size $n = 2$, shown below:

1st observation	2nd observation 4	6	8
4	4,4	4,6	4,8
6	6,4	6,6	6,8
8	8,4	8,6	8,8

b. **(1 mark)** Complete the following table with the average of each sample, $\bar{x}$.

4	5	6
5	6	7
6	7	8

c. **(1 mark)** Complete the following table with the variance of each sample, s^2.

0	2	8
2	0	2
8	2	0

d. **(1 mark)** Use part (b) to calculate the probability distribution of the sample mean $\bar{X}$.

$\bar{X} = \bar{x}$	4	5	6	7	8
$P(\bar{X} = \bar{x})$	1/9	2/9	1/3	2/9	1/9

e. **(2 marks)** Show that $\bar{X}$ is an unbiased estimator of μ.

$E(\bar{x}) = \sum \bar{x}\, P(\bar{x})$

$= 4 \cdot 1/9 + 5 \cdot 2/9 + 6 \cdot 1/3 + 7 \cdot 2/9 + 8 \cdot 1/9$

$E(\bar{x}) = 6 = \mu$

$\therefore \bar{x}$ is unbiased

f. **(1 mark)** Use part (c) to calculate the probability distribution of the sample variance, S^2.

$S^2 = s^2$	0	2	8
$P(S^2 = s^2)$	1/3	4/9	2/9

g. **(2 marks)** Show that S^2 is an unbiased estimator of σ^2.

$E(S^2) = \Sigma s^2 p(s^2) = 0 \cdot 1/3 + 2 \cdot 4/9 + 8 \cdot 2/9$

$= 24/9 = 2.67 = \sigma^2$

$\therefore$ s^2 is unbiased

Q2 (10 marks) We wish to estimate the annual income of households in a particular community for the year 2017. In 2016, the entire community had an average annual income of C$50,000 with a standard deviation of C$5,000. A sample of 625 households is randomly selected. Let $\bar{X}$ denote the mean annual income for this sample.

a. **(2 marks)** Calculate the mean of $\bar{X}$.

$E(\bar{X}) = \mu = \$50,000$

b. **(2 marks)** Calculate the standard error of $\bar{X}$.

$S.E.(\bar{x}) = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5,000}{\sqrt{625}} = 200$

c. **(2 marks)** According to Central Limit Theorem (CLT), describe the sampling distribution of $\bar{X}$.

$\bar{X} \sim N\left(50,000, \frac{5,000^2}{625}\right)$

OR

The sampling distribution of $\bar{X}$ will be approximately normal with mean = 50,000 and standard error = 200.

d. **(2 marks)** Calculate $P(\bar{X} > C\$50{,}050)$.

$$P(\bar{X} > 50{,}050) = P\left[Z > \frac{50{,}050 - 50{,}000}{200}\right]$$

$$= P(Z > 0.25) = 1 - F(0.25)$$

$$= 1 - 0.5987 = \underline{\underline{0.4013}}$$

e. **(2 marks)** Suppose we select another random sample of 700 households ($n = 700$). Let $\bar{X}'$ denote the mean annual income of this second sample. Of the two estimators, $\bar{X}$ and $\bar{X}'$, ***briefly*** explain which is more efficient.

$$var(\bar{X}) = \frac{\sigma^2}{625} > \frac{\sigma^2}{750} = var(\bar{X}')$$

$\therefore \bar{X}'$ is the more efficient estimator

Q3 (12 marks) The US Commission on Crime wishes to estimate the fraction of crimes related to firearms in an area with one of the highest crime rates in the country. The commission randomly selects 600 files of recently committed crimes in the area and finds 300 in which a firearm was reportedly used. Let $\hat{p}$ denote the sample proportion (i.e., the fraction of crimes related to firearms).

a. **(2 marks)** Calculate the sample proportion, $\hat{p}$.

$$\hat{p} = \frac{300}{600} = 0.5$$

b. **(2 marks)** Calculate the variance of $\hat{p}$.

$$Var(\hat{p}) = \frac{\hat{p} \times (1-\hat{p})}{n} = \frac{0.5 \times 0.5}{600} = 0.0004$$

c. **(2 marks)** According to Central Limit Theorem (CLT), describe the sampling distribution of $\hat{p}$.

$$\hat{p} \sim N\left(p, \frac{p \times (1-p)}{n}\right) \Rightarrow \hat{p} \sim N(0.5, 0.0004)$$

<u>OR</u> in words

d. **(2 marks)** Calculate the margin of error (ME) for a 95 percent confidence interval for the population proportion of crimes related to firearms, *p*.

$$\alpha = 0.05 \Rightarrow ME = z_{\alpha/2} \cdot \sigma_{\hat{p}}$$

$$ME = 1.96 \times 0.02 = 0.0392$$

e. **(2 marks)** Construct the 95% confidence interval for *p*.

$$C.I. = \hat{p} \pm ME = 0.5 \pm 0.0392$$

The 95% C.I. is: $(0.4608, 0.5392)$

f. **(2 marks)** ***Briefly*** **interpret the confidence interval you produced in part (e).**

We are 95% confident that the fraction of crimes related to firearms in this area lies between 0.4608 and 0.5392.

Q4 (6 marks) Since you are particularly interested in a certain foreign sedan, you decide to estimate the average resale value of this car model with a 95% confidence level. You obtain data on 16 recently resold sedans of this model. These 16 cars were resold at an average price of $13,800 with a standard deviation of $800. Assume that the resale price for all cars of this model is normally distributed.

a. **(2 marks) Given the small-size sample and without knowing the population standard deviation (i.e., σ), describe the approximate distribution of the standardized sample mean of resale price (i.e., the standardized $\bar{X}$) with the parameter (s) indicated.**

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1} \Rightarrow \frac{\bar{X}-13,800}{800/\sqrt{16}} \sim t_{15}$$

OR in words

b. **(2 marks) Calculate the margin of error (ME) for a 95% confidence interval for the true mean resale price of all such sedans.**

$$\alpha = 0.05 \Rightarrow ME = t_{n-1,\alpha/2} \times \sigma_{\bar{x}}$$

$$= t_{15,0.025} \times \frac{s}{\sqrt{n}}$$

$$= 2.131 \times 200 = 426$$

c. **(2 marks)** Construct the 95% confidence interval for the true mean resale price of all such sedans.

The 95% C.I. is: $\bar{x} \pm ME$.

$= 13{,}800 \pm 426$

The 95% C.I. $= (13{,}374,\ 14{,}226)$

Q5 (8 marks) Canadian financial regulators have imposed stricter rules on the banking industry since the last financial crisis (2007-2009). In order to analyse the impact of new regulations on the disclosures related to credit losses, we collect annual data on credit losses (in millions of C$) from four Canadian banks, both before and after the financial crisis. Assume credit losses are normally distributed.

Bank	Before (X)	After (Y)	Difference
1	8	4	- 4
2	5	6	1
3	3	7	4
4	12	19	7
Sum			8

a. **(2 marks)** Calculate the sample average of the differences.

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{8}{4} = 2$$

b. **(2 marks)** Calculate the standard deviation of the differences.

$$S_d^2 = \frac{\Sigma(d_i - \bar{d})^2}{n-1} = \frac{6^2 + 1^2 + 2^2 + 5^2}{3} = 22$$

$$S_d = \sqrt{22} = 4.6904$$

c. **(2 marks)** Calculate the margin of error (ME) for a 90 percent confidence interval for the true mean difference in credit losses before and after the crisis.

$$\alpha = 0.1 \Rightarrow ME = t_{n-1,\alpha/2} \times S.E(\bar{d}) = t_{3,0.05}\left(\frac{S_d}{\sqrt{n}}\right)$$

$$ME = 2.353 \times 2.345$$

$$ME = 5.518$$

d. **(2 marks)** Construct the 90 percent confidence interval for the true mean difference in credit losses.

$$C.I. = \bar{d} \pm ME$$
$$= 2 \pm 5.518$$

The 90 % C.I. is $(-3.518, 7.518)$

Q6 (6 marks) A random sample of 25 temperatures over a one-week period is obtained. The sample variance is found to be $s^2 = 100$. Assume that temperature is normally distributed.

a. **(2 marks)** Describe the distribution of the sample variance and indicate the parameter (s) of the distribution.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \; ; \; df = 24$$

OR

in words

b. **(2 marks)** Find the upper-tail and lower-tail χ^2 scores for a 95% confidence interval from the distribution table.

$\alpha = 0.05 \Rightarrow \chi^2_{n-1, \alpha/2} = \chi^2_{24, 0.025} = 39.364$

$\chi^2_{n-1, 1-\alpha/2} = \chi^2_{24, 0.975} = 12.401$

c. **(2 marks)** Construct a 95% confidence interval for the true population variance in temperatures.

$UCL = \frac{(n-1)s^2}{\chi^2_{n-1, 1-\alpha/2}}$ $LCL = \frac{(n-1)s^2}{\chi^2_{n-1, \alpha/2}}$

$UCL = \frac{24 \times 100}{12.401}$ $LCL = \frac{24 \times 100}{39.364}$

$= 193.5328$ $= 60.9694$

The 95% C.I. is: $(60.9694, 193.5328)$