CENTRAL UNIVERSITY

FAITH • INTEGRITY • EXCELLENCE

2020/2021 END OF SECOND SEMESTER EXAMINATIONS

FACULTY OF ARTS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS

ECON 106: INTRODUCTION TO STATISTICS II
(LEVEL 100)

Time Allowed: 3 hours

Instructions:

1. Answer **ALL** questions in **section A.**
2. In **secion B**, answer question one **(1)** and any other two **(2)** questions in the answer booklet provided.

Name of Examiner: Mr. Ernest Somua-Wiafe

SECTION A (30 marks)

1. Which of the following illustrates discrete random variable?
   (A) number (0, 1, 2, or a greater integer) in favour of school uniforms
   (B) the time (in hrs) it takes for laptop batteries to fully charge
   (C) the number of one litre gallons of gasoline used by the car
   (D) weights (in kg) of boxes of cereal

2. The probability distribution for the number of defects found in a sample is shown below. Calculate the expected value and the standard deviation.

Defects, X	0	1	2	3
P(x)	0.3385	0.4481	0.1883	0.0251

   (A) $E(X) = 0.9; SD(X) = 0.8$
   (B) $E(X) = 0.9; SD(X) = 0.8$
   (C) $E(X) = 1.0; SD(X) = 0.8$
   (D) $E(X) = 1.0; SD(X) = 0.8$

For questions 3 – 5, we the Poisson probability distribution table to find the indicated probabilities for the random variable X.

3. $P(X \leq 2)$ with $\sigma = 4$
   (A) 0.6767
   (B) 0.9920
   (C) 0.2381
   (D) 0.7619

4. $P(X < 2)$ with $\sigma = 4$
   (A) 0.2707
   (B) 0.0916
   (C) 0.4060
   (D) 0.6703

5. $P(X > 2)$ with $\sigma = 4$
   (A) 0.5940
   (B) 0.3233
   (C) 0.7619
   (D) 0.9084

6. Which of the following probability distributions matches the histogram below?

(A)

Accidents, x	[illegible]	[illegible]	[illegible]	[illegible]
P(x)	0.7	0.2	0.1	0.5

(B)

Accidents, x	0	1	2	3
P(x)	0.1	0.2	0.5	0.2

(C)

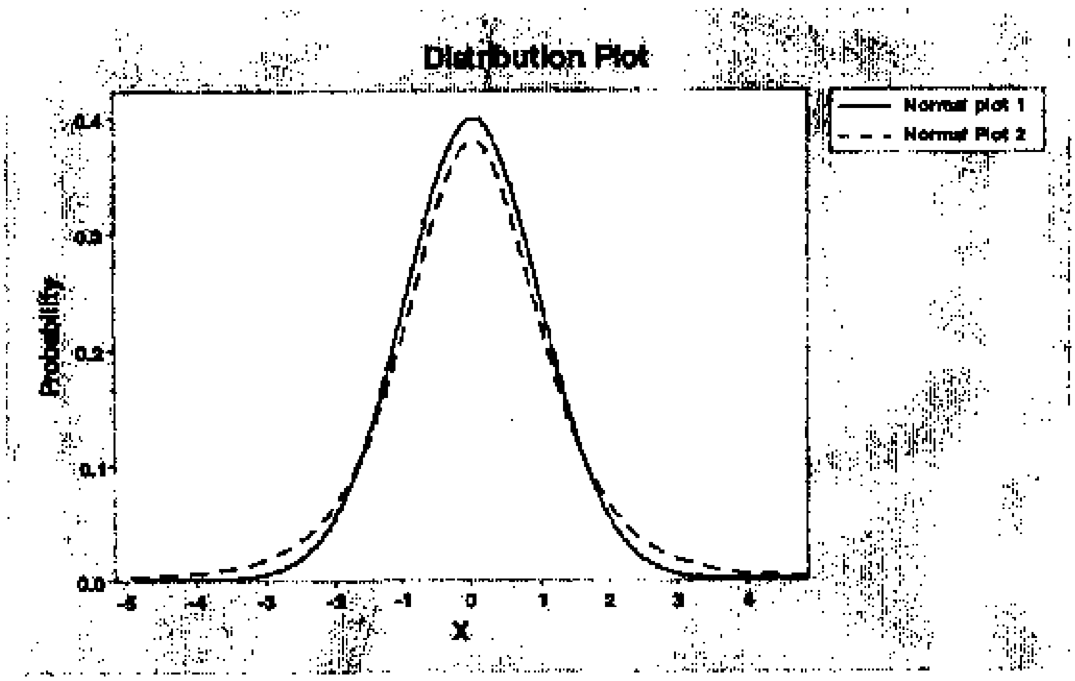
Accidents, x	[illegible]	[illegible]	2	[illegible]
P(x)	0.70	0.20	0.01	0.05

(D)

Accidents, x	0	1	2	3
P(x)	0.70	0.15	0.10	0.05

7. Which of the following is true about the student t-distribution?
   (A) The curve is not isymptotic to the x-axis.
   (B) The t-distributioı has heavy tails as compared to the standard normal distribution.
   (C) The t-distributim serves as an alternative to the standard normal distribution when the population variance is known.
   (D) The t-distribuion is defined by its mean and standard deviation.

8. Given that $X_1 \sim N(\mu_1, \sigma_1)$, and $X_2 \sim N(\mu_2, \sigma_2)$ which of the following parameters relations supports the plot below?



(A) $\mu_1 = \mu_2$; $\sigma_1 > \sigma_2$
(B) $\mu_1 = \mu_2$; $\sigma_1 \geq \sigma_2$
(C) $\mu_1 = \mu_2$; $\sigma_1 < \sigma_2$
(D) $\mu_1 = \mu_2$; $\sigma_1 \leq \sigma_2$

9. A statement made about a population for testing purposes is called?
(A) Statistical test
(B) Hypothesis
(C) Critical point
(D) Decision rule

10. Which of the following are discrete distribution(s)?

I. Binomial
II. Exponential
III. Hypergeometric
IV. Normal
V. Poisson

(A) I, III, and V
(B) II, and IV
(C) III and V
(D) I, II, and III

11. Which of these statements is true about continuous random variables and their probability distributions?

(A) The distribution is modelled by a probability distribution function that is nonnegative.
(B) The probability of an individual event E is calculated by P(X = E).
(C) The total are under the curve is 100.
(D) The curve of such a distribution is perfectly symmetric around the mean

12. Which of the statements is true about the normal probability distribution?

I. $\mu = 0$ and $\sigma = 1$ only.
II. It is symmetric about μ with tails extending to positive and negative infinity.
III. The total area under the curve cannot be determined because the tails are asymptotic to the horizontal axis.

(A) I only
(B) I and II only
(C) I and III only
(D) II only

13. In the standard normal distribution, what is the probability that $Z \geq 1.25$?

(A) 0.8925
(B) 0.8962
(C) 0.8944
(D) 0.1056

14. In the standard normal distribution, what is the probability that $-2.95 < Z \leq 0.95$?

(A) 0.0016
(B) 0.8289
(C) 0.8305
(D) 0.8273

15. If SAT scores are normally distributed with a mean of 500 and a standard deviation of 100, what minimum score is needed to ensure that you are in the top 7%?

(A) 640
(B) 500
(C) 648
(D) 650

16. A binomial random variable X counts the number of successes in n trials. What are the mean and standard deviation of X if the probability of failure is 0.29 and is repeated 145 times?

(A) 42.05; 29.86
(B) 102.95; 5.46
(C) 57.06; 7.43
(D) 126.42; 52.81

17. A binomial random variable X counts the number of successes in 40 trials. If $\mu_x = 30$, what is the probability of failure for a single trial?

(A) 0.25
(B) 0.56
(C) 0.44
(D) The probability of failure cannot be determined from the given information

18. You want to create a 95% confidence interval with a margin of error of no more than 0.05 for a population proportion. The historical data indicate that the population has remained constant at about 0.55. What is the minimum size random sample you need to construct this interval?

(A) 376
(B) 381
(C) 392
(D) 306

19. What is the critical t-value of a 95% confidence interval estimate for a sample size of 25?

(A) 2.060
(B) 2.064
(C) 2.797
(D) 2.787

20. Suppose you construct a 95% confidence interval from a random sample of size n = 20 with a sample mean 100 taken from a population with unknown μ and a known $\sigma = 10$, and the interval is fairly wide. Which of the following conditions would not lead to a narrower confidence interval?

(A) If you decreased your confidence interval
(B) If you increased your sample size
(C) If the sample size was smaller
(D) If the population standard deviation was smaller

21. A confidence interval estimate is determined from the SAT scores of an SRS of n students. All other things being equal, which of the following will result in a small margin of error?

(A) Smaller standard deviation
(B) Smaller sample size
(C) Smaller confidence interval
(D) Reworking the null hypothesis

22. In a test to check pH level, 49 samples showed its level to be 2.4 and with a standard deviation of 0.35. Find the 90% confidence interval estimate for the mean pH level.

(A) 2.4 ± 0.08
(B) 2.4 ± 0.32
(C) 2.4 ± 0.35
(D) 2.4 ± 1.51

23. In 1999, 45% of freshman at a particular college had enrolled in an introductory English course. In a sample of 800 students this year, 460 freshmen had enrolled in an introductory English course. Which type of test should be used to determine whether the percentage has increased since 1999?

(A) A matched-pairs t-test
(B) A one-sample proportion z-test
(C) A two-sample t-test
(D) A one-sample t-test

24. The mean weight of 120 male high school seniors is 156 pounds with a variance of 529 pounds. Assuming the distribution of weights is approximately normal, which type of test should be used to determine whether the mean weight of all high school seniors is more than 150 pounds?

(A) A matched-pairs t-test
(B) A one-sample proportion z-test
(C) A two-sample t-test
(D) A one-sample t-test

25. In a test of $H_0 : \mu = 12.5$ versus $H_A : \mu < 12.5$, the null hypothesis is not rejected at the 5% level. Which of the following statements must be true?

(A) In the same test, the null hypothesis will be rejected at the 10% level.
(B) The sample used had a mean that was 12.5.
(C) There is no evidence at the 5% level that the true mean is less than 12.5.
(D) The standard deviation of the sample is 12.5.

26. Which of the following statements must be true to use a one-sample proportion z-test?

I. $np \geq 5$, $n(p - 1) \geq 5$
II. The population the sample is taken from is approximately normal.
III. The hypothesized proportion is greater than 30%.

(A) I only
(B) II only
(C) I and III only
(D) I, II, and III

27. A random sample of 20 from a normal population has a mean of 15.1. In a test of $H_0 : \mu = 15$ versus $H_A : \mu \neq 15$, which of the following is the correct decision and conclusion at the 5% level?

(A) Reject H_0 concluding that there is evidence the true mean is 15.
(B) Reject H_0 concluding that there is no evidence the true mean is 15.
(C) Do not reject H_0 concluding that there is no evidence that the true mean is 15.
(D) There is not enough information to determine whether H_0 should be rejected.

28. In a test of $H_0 : p = 0.5$ versus $H_A : p < 0.5$, the point estimate of the population proportion is 0.51. Which of the following is the correct conclusion?

(A) At the 5% level, there is no evidence that the population proportion is less than 0.5.
(B) At the 5% level, there is evidence that the population proportion is less than 0.5, but not at the 10% level.
(C) At the 5% level, there is no evidence that the population proportion is less than 0.5, but there is evidence at the 10% level.
(D) A conclusion cannot be made based on this point estimate alone.

29. A sample of 15 from a normal population yields a sample mean of 43 and a sample standard deviation of 4.7. What is the *p-value* that should be used to test the claim that the population mean is less than 45?

(A) 0.0608
(B) 0.1216
(C) 0.4696
(D) 0.9392

30. A 99% confidence interval for the mean of a normal population is 153.7 ± 12.1. In a test of $H_0 : \mu = 156$ versus $H_A : \mu \neq 156$, what would be the appropriate decision and conclusion at the 1% level?

(A) Reject H_0 concluding that there is evidence the true mean is not 156.
(B) Do not reject H_0 concluding that there is evidence the true mean is not 156.
(C) Do not reject H_0 concluding that there is no evidence that the true mean is not 156.
(D) H_0 may or may not be rejected depending on the population standard deviation.

SECTION B [70 marks]
(Answer question 1 and any other two questions)

Question 1 [30 marks]

a) It's been a good day for the market, with 75% of securities gaining value. You are evaluating a portfolio of 15 securities and will assume a binomial distribution for the number of securities that lost value.

i. What assumptions are being made when you use a binomial distribution in this way?

[5 marks]

ii. How many securities in your portfolio would you expect to lose value?

[2 marks]

iii. What is the standard deviation of the number of securities in your portfolio that lose value?

[2 marks]

iv. Find the probability that all 15 securities gain value.

[4 marks]

v. Find the probability that exactly 5 securities lose value.

[4 marks]

vi. Find the probability that less than 5 securities lose value.

[5 marks]

b) Assume that the number of securities you are evaluating is increased to 150 eventually, find the probability that less than 110 securities gain value.

[8 marks]

Question 2 [20 marks]

During rush periods, accidents occur in a particular metropolitan area at the rate of two per hour. The morning rush period lasts for one hour and 30 minutes and the evening rush period lasts for two hours.

i. On a particular day, what is the probability that there will be no accidents during the morning rush period?

[3 marks]

ii. What is the probability of two accidents during the evening rush period?

[3 marks]

iii. What is the probability of four or more accidents during the morning rush period?

[6 marks]

iv. On a particular day, what is the probability that there will be an accident during both the morning and evening rush periods?

[8 marks]

Question 3 [20 marks]

Suppose that the data below were collected from a sample of gas stations in Takoradi.

Number of Stations Selling E85	Number of Sampled Stations
251	412

i. Calculate a 95% confidence interval for the percentage of all gas stations in Takoradi selling E85.

[8 marks]

ii. Suppose the confidence interval obtained in part (i) is too wide. How can the width of this interval be reduced? Discuss all possible alternatives. Which alternative is the best?

[8 marks]

iii. Using your confidence interval only, would the null hypothesis $H_0 : p = 0.61$ be rejected at the 5% level? Justify your answer.

[4 marks]

Question 4 [20 marks]

The response times for a random sample of 40 medical emergencies were tabulated. The sample mean is 13.25 minutes. The population standard deviation is believed to be 3.2 minutes. If the EMS director wants to perform a hypothesis test, with a .05 level of significance, to determine whether the service goal of 12 minutes or less is being achieved,

i. Which inferential statistical test is most appropriate in this case? Briefly explain.

[4 marks]

ii. State the null hypothesis and the alternate hypothesis.

[2 marks]

iii. State the decision rule.

[3 marks]

iv. Compute the value of test statistic.

[4 marks]

v. Obtain the p-value.

[3 marks]

vi. What is your decision regarding the null hypothesis? Explain your conclusion in words.

[4 marks]

Question 5 [20 marks]

A study has shown that 20% of all college textbooks have a price of $184.52 or higher. It is known that the standard deviation of the prices of all college textbooks is $36.35. Suppose the prices of all college textbooks have a normal distribution. What is the mean price of all college textbooks? Demonstrate your understanding with a probability distribution sketch.

[20 marks]

FORMULA SHEET

Mean of a discrete probability distribution, $\mu = \sum \{xP(x)\}$ where $P(x)$ is the probability of the discrete random variable X.

Variance of a discrete probability distribution, $\sigma^2 = \sum \left\{(x-\mu)^2 P(x)\right\}$ where $P(x)$ is the probability of the discrete random variable X

Binomial Distribution

If $X \sim Binomial(n, \pi)$ where n is the number of trials and π is the probability of success, then

$$P(X = x) = C_x^n \pi^x (1-\pi)^{n-x}, \qquad for,\ 0 \le \pi \le 1 \quad and \quad x = 0,1,2,\ldots, n$$

Mean of a binomial probability distribution, $\mu = n\pi$

Variance of a binomial probability distribution, $\sigma^2 = n\pi(1-\pi)$.

Hypergeometric Distribution

If $X \sim Hypergeometric(N, S, n)$ where N is the size of the population, S is the number of successes in the population and n is the size of the sample or the number of trials, then

$$P(X = x) = \frac{(C_x^S)(C_{n-x}^{N-S})}{C_n^N}, \qquad for,\ x = 0,1,2,\ldots$$

Mean of a hypergeometric probability distribution, $\mu = n\left(\frac{S}{N}\right)$

Variance of a hypergeometric probability distribution, $\sigma^2 = n\left(\frac{S}{N}\right)\left(1-\frac{S}{N}\right)\left(\frac{N-n}{N-1}\right)$.

The binomial distribution can be used as an approximation to the hypergeometric distribution in the case where $n < 0.05(N)$.

Poisson Distribution

If $X \sim Poisson(\mu)$, then

$$P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, \qquad for\ x = 0,1,2,\ldots \quad and \quad \mu > 0$$

μ is the mean number of occurrences (successes) in a particular interval.

x is the number of occurrences (successes).

Mean of a Poisson probability distribution, $\mu = n\pi$

Variance of a Poisson probability distribution, $\sigma^2 = \mu = n\pi$.

Uniform Distribution

If $X \sim uniform(a, b)$, then $P(X = x) = \frac{1}{b-a}$, $if\ a \le x \le b$, $and\ 0\ elsewhere$

$b > a \in \mathscr{R}$

$and \quad b - a \neq 0$

Mean of a Uniform probability distribution, $\mu = \frac{a+b}{2}$

Variance of a Uniform probability distribution, $\sigma^2 = \frac{(b-a)^2}{12}$

Exponential Distribution

If $X \sim Exp(\lambda)$, then $P(X = x) = \lambda e^{-\lambda x}$ $for\, x > 0,\ and\ \ \lambda > 0$

Also, $P(X < x) = 1 - e^{-\lambda x}$

Mean of a Uniform probability distribution, $\mu = \frac{1}{\lambda}$

Variance of a Uniform probability distribution, $\sigma^2 = \left(\frac{1}{\lambda}\right)^2$

Normal Distribution

If $X \sim N(\mu, \sigma^2)$, then $P(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]}$

Standard Normal (z-value), $z = \frac{(x - \mu)}{\sigma}$

Continuity Correction Factor

For the probability that x occurs, use the area between $(x - 0.5)$ and $(x + 0.5)$

For the probability at least x occur, use the area above $(x - 0.5)$.

For the probability that more than x occur, use the area above $(x + 0.5)$.

For the probability that x or fewer occur, use the area below $(x + 0.5)$.

For the probability that fewer than x occur, use the area below $(x - 0.5)$.

Central Limit Theorem

If all samples of a particular size are selected from any population, the sampling distribution of the sample mean is approximately a normal distribution. This approximation improves with larger samples.

Mean of the sample means $\mu_{\bar{x}} = \mu$

Standard error of the mean $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

z-value of $\bar{x}$ with population variance known $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Confidence Interval (CI) Estimation

CI for a population mean with σ known $= \left[\bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right]$

CI for a population mean with σ unknown $= \left[\bar{x} \pm t_{\frac{\alpha}{2},n-1}\left(\frac{s}{\sqrt{n}}\right)\right]$

CI for a population proportion $= \left[p \pm z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right]$

Sample size for estimating the population mean $n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$

Sample size for estimating the population proportion $n = \pi(1-\pi)\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2$

where: n is the size of the sample.
E is the maximum allowable error.
π is the population proportion

Finite-population correction factor (FPC) $= \sqrt{\frac{N-n}{N-1}}$

Hypothesis testing

One sample

	Test Statistic
Testing a mean with population variance known	$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
Testing a mean with population variance unknown	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$
Testing about a population proportion	$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$
Type II error	$z = \frac{\bar{x}_c - \mu_1}{\sigma/\sqrt{n}}$

Two sample | **Test Statistic**

Testing the means of independent samples (σ known)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Two-Sample Pooled Test of means (σ unknown)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

with $n_1 + n_2 - 2$ degrees of freedom

Test for no difference in means (unequal variance)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$df = \frac{\left[\left(s_1^2/n_1\right) + \left(s_2^2/n_2\right)\right]^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

Paired t test (dependent samples)

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

and

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$

where $\bar{d}$ is the mean of the difference between the paired or related observations.
s_d is the standard deviation of the differences between the paired or related observations.
n is the number of paired observations.

Two-sample test about proportions

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_c(1 - p_c)}{n_1} + \frac{p_c(1 - p_c)}{n_2}}}$$

where $p_c = \frac{x_1 + x_2}{n_1 + n_2}$