CENTRAL UNIVERSITY

FAITH • INTEGRITY • EXCELLENCE

END OF SECOND SEMESTER EXAMINATION
2017 / 2018 ACADEMIC YEAR

FACULTY OF ARTS AND SOCIAL SCIENCES

DEPARTMENT OF ECONOMICS

ECON 106: (3 CREDITS)

INTRODUCTION TO STATISTICS II

LEVEL 100

MAY 2018

DURATION: 3 HOURS

STUDENT ID No..

INSTRUCTIONS:

ANSWER ALL QUESTIONS IN SECTION A
AND
ANSWER QUESTION '1' AND ANY OTHER TWO QUESTIONS IN SECTION B.
ALL QUESTIONS ARE TO BE ANSWERED IN THE ANSWER BOOKLET.
FORMULA SHEET HAS BEEN PROVIDED

DO NOT TURN OVER THIS PAGE UNTIL YOU HAVE BEEN TOLD TO DO SO BY THE INVIGILATOR.

Lecturer: Mr Ernest Somua–Wiafe

SECTION A

(Answer all questions in this section)

1. The following are examples of continuous random variables except ....

a) the number of cars sold at a dealership during a given month

b) the time taken to commute from home to work

c) the weight of a manufactured A4 paper

d) the price of a house

2. Which of the following variables is a binomial random variable?

a) The time it takes a randomly selected student to complete a multiple choice exam

b) The number of textbooks a randomly selected student bought this semester

c) The number of women taller than 63 inches in a random sample of 10 women

d) The number of CDs a randomly selected person owns

3. Given that the probability of success is p, the binomial probability distribution is skewed to the right if

a) $p > 0.5$

b) $p \geq 0.5$

c) $p \leq 0.5$

d) $p < 0.5$

4. Which of the following is not a characteristic of the normal distribution?

a) The total area under the curve is 1.0.

b) The curve is symmetric about the standard deviation.

c) The two tails of the curve extend indefinitely.

d) The curve is asymptotic to the x-axis

5. For the standard normal distribution,

a) $\mu = 0 \; and \; \sigma = 0$

b) $\mu = 1 \; and \; \sigma = 0$

c) $\mu = 0 \; and \; \sigma = 1$

d) $\mu = 1 \; and \; \sigma = 1$

6. Which of the following distributions is suitable to model the length of time that elapses before the first employee passes through the security door of a company?

a) Exponential

b) Poisson

c) Binomial

d) Uniform

7. Which of the following is not a correct statement?

a) the exponential distribution is a family of curves, which are completely described by the mean

b) the mean and variance of the Poisson distribution are equal.

c) the Poisson is a probability distribution for a continuous random variable

d) the area under the curve for an exponential distribution equals 1

8. A larger standard deviation for a normal distribution with an unchanged mean indicates that the distribution becomes:

a) narrower and more peaked

b) flatter and wider

c) more skewed to the right

d) a change in the standard deviation does not change the shape of the distribution

9. Which of the following statements regarding the probability function, $P(x)$, of the uniform distribution is correct?

a) the height of the density function differs for different values of X

b) the density function increases as the values of X increase

c) the density function is roughly "bell-shaped"

d) the density function is constant for all values that X can assume

10. In a popular shopping centre, the waiting time for an ATM machine is found to be uniformly distributed between 1 and 5 minutes. What is the probability of waiting between 2 and 3 minutes to use the ATM?

a) 0.25

b) 0.50

c) 0.75

d) 0.20

For Questions 11 and 13, let Z be the standard normal random variable.

11. Find $P\ (0 \leq Z \leq 0.61)$

a) 0.1500

b) 0.2820

c) 0.2420

d) 0.2291

12. Find $P\ (Z < -1.282)$

a) 0.05

b) 0.10

c) 0.40

d) 0.45

13. Find $P\ (Z\ > 1.645)$

a) 0.05

b) 0.10

c) 0.40

d) 0.50

14. For a set of data, if the mean and the standard deviation are 2.70 and 1.80 respectively, the z-score for the point $x\ =\ 3.5$ is (approximately)

a) 0.00

b) 0.32

c) 0.44

d) 0.65

15. The listed occupations of stockholders of a national computer company included 9% who were housewives. If six of these stockholders are randomly selected, what is the probability that none are housewives?

a) 0.568

b) 0.011

c) 0.083

d) 0.282

16. Which statement correctly compares t-distribution to the normal distribution?

I - As the degrees of freedom increase, the t-models look more and more like the standard normal.

II - They are unimodal, symmetric and bell shaped.

III - They have fatter tails than the standard normal models.

a) I only

b) II only

c) II and III only

d) I, II and III

17. A point estimator is defined as:

a) the average of the sample values

b) the average of the population values

c) a single value that is the best estimate of an unknown population parameter

d) a single value that is the best estimate of an unknown sample statistic

18. The confidence interval is an interval that is used to estimate a:

a) population parameter based on the information from a population

b) population parameter based on the information from a sample

c) sample statistic based on the information from a population

d) sample statistic based on the information from a sample

19. The width of a confidence interval depends on the size of the:

a) population mean

b) population size

c) coefficient of variation

d) margin of error

20. Which of the following is not part of the procedure for estimating the value of a population parameter?

a) Selecting a sample size

b) Collecting the required information from the members of the sample

c) Calculating the value of the sample statistic

d) Calculating the exact value of the corresponding population parameter

21. Which statement is *not* true about confidence intervals?

a) A confidence interval is an interval of values computed from sample data that is likely to include the true population parameter value.

b) An approximate formula for a 95% confidence interval is *sample estimate ± margin of error.*

c) A confidence interval of 10% to 90% implies the parameter definitely lies between 80% of the data

d) A 99% confidence interval procedure has a higher probability of producing intervals that will include the population parameter than a 95% confidence interval procedure.

22. Which statement is correct about a p-value interpretation?

a) The smaller the p-value the stronger the evidence against the alternative hypothesis

b) The smaller the p-value the stronger the evidence against the null hypothesis

c) The larger the p-value the stronger the evidence against the alternative hypothesis

d) The larger the p-value the stronger the evidence against the null hypothesis

23. After constructing a confidence interval estimate for a population mean, you believe that the interval is useless because it is too wide. In order to correct this problem, you need to:

a) Increase the population standard deviation

b) Increase the sample size

c) Increase the level of confidence

d) Increase the sample mean

24. A Type II error occurs when we:

a) reject a false null hypothesis

b) reject a true null hypothesis

c) do not reject a false null hypothesis

d) do not reject a true null hypothesis

25. What is the behavior of the power of the test, $(1-\beta)$ as the hypothesized mean approaches the population mean?

a) The power of the test decreases

b) The power of the test increases

c) The power of the test approaches infinity

d) The power of the test remains constant

26. To test a hypothesis involving proportions, both np and $n(1-p)$ should

a) be greater 30

b) be less than 5

c) Lie in the range from 0 to 1

d) be more than 5

27. Two samples of sizes 27 and 35 are independently drawn from two normal populations, where the unknown variances are assumed to be equal. The number of degrees of freedom for the equal-variances t-test statistic is:

a) 58

b) 60

c) 62

d) 59

28. Which of the following statements is correct?

a) The pooled-variances t-test is used whenever the population standard deviations can be assumed to be equal, regardless of the sample size

b) The unequal-variances t-test is used whenever the population standard deviations are unknown and cannot assumed to be equal

c) the z-test can be used as a close approximation to the unequal-variances t-test when the population standard deviation are not assumed to be equal but sample sizes are typically greater than 30

d) all of the above statements are true

Use the following information for questions 29 and 30.

Suppose that we are interested in comparing the academic success of college students who belong to fraternal organizations with the academic success of those who do not belong to fraternal organizations. Cumulative GPA is used to measure academic success. Random samples of size 40 are taken from each population. If at 0.05 significance level, we are to test the claim that the GPAs of the two populations are different,

29. Choose the correct hypotheses to test the claim.

a) $H_0 : \mu_1 - \mu_2 = 0 \; against \; H_1 : \mu_1 - \mu_2 = 0$

b) $H_0 : \mu_1 - \mu_2 = 0 \; against \; H_1 : \mu_1 - \mu_2 \neq 0$

c) $H_0 : \mu_1 - \mu_2 = 0 \; against \; H_1 : \mu_1 - \mu_2 > 0$

d) $H_0 : \mu_1 - \mu_2 = 0 \; against \; H_1 : \mu_1 - \mu_2 < 0$

30. At the significance level of 0.05, which of the following is the correct conclusion?

a) Accept H_0 since the P-value is less than 0.05

b) Fail to accept H_0 since the P-value is greater than 0.05

c) Accept H_0 since the P-value is greater than 0.05

d) Fail to accept H_0 since the P-value is less than 0.05

SECTION B

(Answer question one and any other two questions)

1. In Ghana, research reveals that roughly 60% of Ghanaian households own one or more televisions. In a survey, 6 Ghanaian households are selected at random.

a) Using statistical tables, find the probability distribution of the random variable X, the number of Ghanaian households in a random sample of six that own one or more televisions.

b) Hence or otherwise, find the probability that, of the six households sampled, the number that own one or more televisions is
i) exactly one;
ii) at least two;
iii) at most three.

c) Compute for the mean and variance of the random variable X, the number of Ghanaian households in a random sample of six that own one or more televisions.

d) Without referring to the probability distribution obtained in part (a) or constructing a probability histogram, decide whether the probability distribution is right skewed, symmetric, or left skewed. Explain your answer.

e) The selection of the six households was done without replacement. Strictly speaking, then, why is the probability distribution that you obtained in (a) only approximately correct? With further explanations, what should be the exact distribution?

f) Do you think the probability distribution that you obtained is a reasonable approximation to the actual one? Explain your answer.

2. The weekly income of Uber drivers followed the normal distribution with a mean of \$1,000 and a variance of \$10,000.

a) What is the likelihood of selecting a driver whose weekly income is between \$1,000 and \$1,100?
b) What is the probability of selecting a driver whose income is
   i) between \$790 and \$1,000?
   ii) more than \$790?
c) What is the area under this normal curve between \$840 and \$1,200?
d) What is the area under the normal curve between \$1,150 and \$1,250?

3. PrintTech, Inc. is introducing a new line of ink-jet printers and would like to promote the number of pages a user can expect from a print cartridge. A sample of 10 cartridges revealed the following number of pages printed.

2,698 2,028 2,474 2,395 2,372 2,475 1,927 3,006 2,334 2,379

a) What is the point estimate of the population mean?
b) Develop a 99% confidence interval for the population mean. Interpret your result.
c) Suppose you decided to switch from a 99% to a 95% confidence interval. Without performing any calculations, will the interval increase, decrease, or stay the same? Briefly explain why.

4. The mean life of a battery used in a digital clock is 305 days. The lives of the batteries follow the normal distribution. The battery was recently modified to last longer. A sample of 20 of the modified batteries had a mean life of 311 days with a variance of 144 days. Did the modification increase the mean life of the battery?

a) State the null hypothesis and the alternate hypothesis.
b) Show the decision rule graphically. Use the 0.05 significance level.
c) Compute the value of test statistic. What is your decision regarding the null hypothesis? Briefly summarize your results.

5. The production manager at Bellevue Steel, a manufacturer of wheelchairs, wants to compare the number of defective wheelchairs produced on the day shift with the number on the afternoon shift. A sample of the production from 6 day shifts and 8 afternoon shifts revealed the following number of defects. The sample variance of day shift is 2 whilst the sample variance for afternoon shift is approximately 5.143.

Day	5	8	7	6	9	7		
Afternoon	8	10	7	11	9	12	14	9

a) Perform a two sample hypothesis test at the 0.05 significance level, if there a difference in the mean number of defects per shift? Interpret the result.
b) What are the assumptions necessary for this test?

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, $\sigma^2 = \sum\{(x-\mu)^2 P(x)\}$ 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

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

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

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

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

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 $P(X = x) = \frac{1}{b-a}, \; if \; a \leq x \leq b, \quad and \; 0 \; elsewhere$

$b > a \in \mathcal{R} \quad and \quad b - a \neq 0$

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 $P(X = x) = \lambda e^{-\lambda x}$ $\quad for \; x > 0, \; and \; \lambda > 0$

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 $\quad P(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]}$

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

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 $\quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Confidence Interval (CI) Estimation

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

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

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

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

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

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	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ with $n - 1$ degrees of freedom
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)	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ with degrees of freedom, $df = \frac{\left[(s_1^2/n_1) + (s_2^2/n_2)\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$
Paired t test (dependent samples)	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$ with $n - 1$ degrees of freedom 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 $p_c = \frac{x_1 + x_2}{n_1 + n_2}$