Your Name:

E370: Statistical Analysis For Business and Economics
Sectional Final Exam Version A

Instructor: Naoya Nagasaka (Section 8106)

December 17, 2024 (5:20 - 7:20 PM)

Instructions

1. Second part of your final exam is the set of 22 multiple choice questions administered to students in E370 instructed by Naoya Nagasaka. Please, proceed to bubble in answers for this exam on the same bubble sheet where you provided the answers to your Departmental exam.
2. Questions on the Departmental and sectional exams are equally weighted and worth 7 points.
3. The maximum total points on the final exam is 315 (including both Departmental and sectional exams).

24. The years in which the students in this class were born are examples of

    (a) interval data

    (b) ratio data

    (c) nominal data

    (d) ordinal data

Exhibit ?

Indiana Police Department is tracing the number of traffic accidents occurred per day in Bloomington. The following table relates the number of accidents and the associated frequency for the randomly selected 200 days in 2023.

Number of Accidents	Number of Days
0	10
1	20
2	10
3	50
4	40
5	50
6	20

The average number of accidents based on this sample is calculated to be 3.6.

25. **Refer to Exhibit ?.** The Police Department would like to learn the property of average number of accidents for the whole days in 2023 using the data on those 200 days. In this context, the average number of accidents based on this sample (3.6) is an example of

    (a) inferential statistics

    (b) population parameters

    (c) descriptive statistics

    (d) point estimator

26. **Refer to Exhibit ?.** The cumulative relative frequency associated with 3 accidents per day is

    (a) 0.45

    (b) 0.2

    (c) 0.25

    (d) 0.65

27. **Refer to Exhibit ?.** Which of the following is true about the data summarized in the table?

    (a) This is a bimodal distribution.

    (b) The median number of accidents per day is 3.

    (c) The proportion of days on which we observe more than five accidents is larger than 0.25.

    (d) When presenting this data visually, a bar chart is the appropriate type of graph to use.

28. We collect the data on daily prices of apples and oranges sold at Kroger. The means and standard deviations of the daily price are summarized in the table below.

	Apple	Orange
Mean ($)	1	2
Standard Deviation ($)	0.15	0.25

    What can we say about the relative variability of apples' and oranges' prices?

    (a) Price of apples is more variable.

    (b) Price of oranges is more variable.

    (c) Price of apples is as variable as that of oranges.

    (d) We are not given enough information to compare the variability.

29. A crucial difference between the covariance and the correlation coefficient is that

    (a) the correlation coefficient is used to measure both the strength and the direction of the linear relationship between variables while the covariance can only describe the strength.

    (b) the correlation coefficient is used to measure both the strength and the direction of the linear relationship between variables while the covariance can only describe the direction.

    (c) the correlation coefficient is used to measure the strength of the linear relationship between variables while the covariance describes both the direction and the strength.

    (d) the correlation coefficient is used to only measure the strength of the linear relationship between variables while the covariance describes only the direction.

30. The sampling distribution of the sample mean $\bar{x}$ is

    (a) the point estimate of the population mean μ.

    (b) the mean of all possible values of the sample mean $\bar{x}$.

    (c) the probability distribution of the population mean μ.

    (d) the probability distribution of all possible values of the sample mean $\bar{x}$.

31. It is believed that the average of cumulative GPA of current IU students is 2.91. A randomly selected sample of 569 IU students was taken, and the students in the sample were asked their cumulative GPA. Thinking about the average of GPA in the sample, which of the following is correct?

    (a) It is the same number as for all college students (2.91) because a random sample is taken.

    (b) It is smaller than 2.91 because the sample of 569 college students is smaller than the entire population of college students.

    (c) It is greater than or equal to 2.91 because the data in the sample vary more than in the population.

    (d) It can be bigger, equal to, or smaller than 2.91.

32. Historically, a large university had 15% of students who failed to pass at least one course each semester. The university is interested in the outcome of a course standardization process. From a random sample of 1,000 students in the current semester, they learned that 13% of students in the sample failed to pass at least one course. What is the mean and the standard error of the sampling distribution of the proportion of students who fail to pass at least one course in this university?

    (a) mean = 0.15, standard error = 0.0113

    (b) mean = 150, standard error = 11.292

    (c) mean = 0.13, standard error = 0.0106

    (d) mean = 130, standard error = 10.635

33. Let the weight of babies born this year be μ on average. Based on the randomly selected sample of newborns, you calculate the 99% confidence interval of weight to be $[6.4\text{lb}, 6.8\text{lb}]$. Three classmates in your statistics class interpret the result in the following way.

    - Classmate A: If you prepare many samples with the same sample size as yours and calculate the confidence interval for each of them, 99% of those confidence intervals includes μ.
    - Classmate B: μ is between 6.4lb and 6.8lb with probability of 99%.
    - Classmate C: 99% of newborns in your sample weighs between 6.4lb and 6.8lb.

    Who interpret your confidence interval correctly?

    (a) A and B

    (b) B and C

    (c) Only A

    (d) Only B

34. The following five numbers are drawn from the normal distribution with unknown mean μ and known standard deviation 1.

$$5, \quad 3, \quad 2, \quad 3, \quad 4$$

Which of the following intervals is closest to the 95% confidence interval of based on these five numbers? If needed, you can use

$$z_{0.025} = 1.960, \quad z_{0.05} = 1.645, \quad t_{0.025} = 2.776 \ (df = 4), \quad t_{0.05} = 2.132 \ (df = 4)$$

Recall that the subscript refers to the area on the right of the value in the corresponding distribution.

(a) [2.420, 4.380]

(b) [2.578, 4.223]

(c) [2.012, 4.788]

(d) [2.334, 4.466]

35. Susan, a music critic, feels that recent hit songs are shorter than the previous ones, possibly due to the appearance of music streaming services. To see whether her conjecture is supported by data, she randomly picks up 12 songs from Spotify Top Hits, and calculates the average and standard deviation of the duration of those songs to be 220 seconds and 38 seconds, respectively. It is known that the (population) average duration of hit songs was 240 seconds 10 years ago. Which of the following are appropriate hypotheses to test her conjecture?

(a) $H_0 : \mu \geq 240, \quad H_1 : \mu < 240$

(b) $H_0 : \mu \leq 240, \quad H_1 : \mu > 240$

(c) $H_0 : \mu \geq 220, \quad H_1 : \mu < 220$

(d) $H_0 : \mu \leq 220, \quad H_1 : \mu > 220$

36. We would like to test $H_0 : \mu \leq 5$ against $H_1 : \mu > 5$. We obtain a sample with 9 observations from which the sample mean and standard deviation are computed to be 5.8 and 1, respectively. Assume that the population is normally distributed. Which of the following is the correct pair of test statistics and conclusion from the hypothesis testing with the significance level of 1%? If needed, you can use

$$z_{0.005} = 2.576, \quad z_{0.01} = 2.326, \quad t_{0.005} = 3.356 \ (df = 8), \quad t_{0.01} = 2.897 \ (df = 8)$$

(a) The test statistic is 2.40, and we reject the null hypothesis.

(b) The test statistic is 2.40, and we do not reject the null hypothesis.

(c) The test statistic is 7.20, and we reject the null hypothesis.

(d) We cannot conclude due to the insufficient sample size.

37. If the p-value for a two-tail hypothesis test is 0.11 and the significance level (α) is 10%, then the correct conclusion is to

    (a) reject the null hypothesis since $\frac{p\text{-value}}{2} < \alpha$.

    (b) accept the null hypothesis since p-value $> \alpha$.

    (c) not reject the null hypothesis since p-value $> \alpha$.

    (d) reject the null hypothesis since p-value $> \frac{\alpha}{2}$.

Exhibit ?

In exams with multiple choice questions, we expect that each choice (A, B, C, and D) appears equally likely (i.e., with probability of 25%). The table below shows the frequency of correct answers in our Midterm 2 Version A.

Choice	Appearance	Proportion
A	8	0.276
B	8	0.276
C	5	0.172
D	8	0.276
Total	29	1

38. **Refer to Exhibit ?.** We would like to test whether B appears with probability of 25%. To calculate the critical value or p-value for this test, we use

    (a) Standard normal distribution

    (b) t distribution with degrees of freedom 29

    (c) t distribution with degrees of freedom 28

    (d) None of them is appropriate.

39. **Refer to Exhibit ?.** We would like to test whether C appears with probability of 25%. The p-value associated with the test statistic is found to be 0.335. How do we interpret this p-value?

    (a) The null hypothesis is correct with probability of 33.5%.

    (b) The alternative hypothesis is correct with probability of 33.5%.

    (c) The probability of having the proportion of C as a correct answer in randomly selected 29 questions to be more extreme than or as extreme as 17.2% is equal to 33.5% if C actually appears as a correct answer with probability of 25%.

    (d) The probability of having the proportion of C as a correct answer in randomly selected 29 questions to be more extreme than or as extreme as 25% is equal to 33.5% if C actually appears as a correct answer with probability of 17.2%.

40. Based on the sample on heights of fathers and sons, we get the following estimated regression equation.

$$\hat{y} = 1 + 0.8x$$

where x is height of father in feet and y is height of son in feet. Which of the following statements is **NOT** correct about the equation above?

(a) This equation summarizes the linear relationship between fathers' and sons' heights at the population level.

(b) If a father is 6 feet tall, his son is predicted to be 5.8 feet tall on average.

(c) The underlying assumption is that "variation in fathers' height explains variation in sons' height, but variation in son's height should not explain variation in fathers' height".

(d) When a father's height decreases by 0.5 feet, his son's height is predicted to decrease by 0.4 feet on average.

Exhibit ?

Policymakers are interested in whether government spending effectively stimulates gross domestic product (GDP). A useful measure to answer this question is the "government spending multiplier", which is defined as how much GDP increases when the government spending increases by \$1. To quantify the multiplier, staffs at the Congressional Budget Office (CBO) consider the following regression model.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where y is GDP and x is government spending, both of which are measured with dollar. The slope coefficient associated with x is the government spending multiplier. To estimate this model, they use the data on GDP and government spending for 51 US states in 2022. The screenshot below is the output from the Excel data analysis tool.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.32673
R Square	0.10675
Adjusted R Square	0.08852
Standard Error	860.72144
Observations	51

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4338409.013	4338409.013		0.019
Residual	49	36301228.823	740841.405		
Total	50	40639637.836			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	10156.570	1164.614		0.000
X Variable 1	1.428	0.590		0.019

41. **Refer to Exhibit ?.** The estimated government spending multiplier (the slope coefficient) is

    (a) 1.428

    (b) 10,156.570

    (c) 5.856

    (d) 2.420

42. **Refer to Exhibit ?.** Which of the following statements about the goodness of fit is correct?

    (a) 10.7% of variation in state-level GDP is explained by variation in state-level government spending.

    (b) 10.7% of variation in state-level government spending is explained by variation in state-level GDP.

    (c) 86.1% of variation in state-level GDP is explained by variation in state-level government spending.

    (d) 86.1% of variation in state-level government spending is explained by variation in state-level GDP.

43. **Refer to Exhibit ?.** When we test the significance of the coefficient of determination for this regression, the null and alternative hypotheses take the form

    (a) $H_0 : \rho^2 \leq 0$, $H_1 : \rho^2 > 0$

    (b) $H_0 : \rho^2 \geq 0$, $H_1 : \rho^2 < 0$

    (c) $H_0 : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$

    (d) $H_0 : \beta_1 \leq 0$, $H_1 : \beta_1 > 0$

44. **Refer to Exhibit?.** We would like to test whether the government spending multiplier (the slope coefficient) is different from zero. The test statistic we use is

    (a) 2.420

    (b) 8.721

    (c) 5.856

    (d) 0.019

45. **Refer to Exhibit ?.** We again consider the test on whether the government spending multiplier is different from zero. Which of the following statements is the most appropriate about this hypothesis testing?

    (a) Under the significance level of 5%, we conclude that the government spending multiplier is different from zero.

    (b) Under the significance level of 1%, we conclude that the government spending multiplier is different from zero.

    (c) In this hypothesis testing, the null hypothesis is set as $\beta_1 \leq 0$, and the alternative hypothesis is set as $\beta_1 > 0$.

    (d) We use the standard normal distribution to compute the critical value or p-value.

This is the end of exam. Before turning in your exam, **please make sure that you have bubbled in your IU username and version number on the bubble sheet, and written down your first and last name both on the bubble sheet and the front page of question sheets!**