STAT 1222 Common Final Exam

December 2019
December 6, 2019

Please print the following information:

Name: ______________________ Instructor: ______________________

Student ID #: ______________________ Section/Time: ______________________

THIS EXAM HAS TWO PARTS

PART I. Consists of 30 multiple choice questions worth a total of 60 points. Read all questions carefully. You may do calculations on the test paper. Mark the number of the opscan sheet corresponding to the test question number with a Number 2 pencil or a mechanical pencil with HB lead. Mark only one answer; otherwise the answer will be counted as incorrect. In case there is more than one answer, mark the best answer. Please make sure that your name appears on the opscan sheet in the spaces provided.

PART II. This part consists of 3 questions (40 points in total). You MUST show all work for each question in the space provided to receive full credit for that question. If you write your explanations in another part of the test, please indicate accordingly.

At the end of the examination, you MUST hand in this test booklet, your answer sheet and all scratch paper.

FOR DEPARTMENTAL USE ONLY:
PART II:

Questions	1	2	3
Maximum	12	12	16
Score			

Part I Part II Total

Use the following information for questions 1 to 3.
Given the following sample data:

6, 7, 7, 9, 5, 8, 6, 12

1. The median of this data set is

   (a) 5.00 (b) 6.00 (c) 6.50 (d) 7.00 (e) 7.50

2. The sample standard deviation of this data set is about

   (a) 2.06 (b) 2.20 (c) 4.25 (d) 4.86 (e) 7.50

3. The sample mean of this data set is about

   (a) 5.00 (b) 6.00 (c) 6.50 (d) 7.00 (e) 7.50

4. If Z denotes the standard normal random variable, then $P(-0.57 \leq Z \leq 0.22)$ is about

   (a) 0.7900 (b) 0.5871 (c) 0.2843 (d) 0.3500 (e) 0.3028

5. A survey was conducted to estimate the proportion of all eligible voters who actually voted in the previous presidential election. The survey investigated 2000 eligible voters and found that 1360 or 68% of them actually voted. Which of the following is true?

   (a) The sample is the 1360 eligible voters who actually voted in the previous presidential election.
   (b) The population in this question consists of 2000 eligible voters.
   (c) The sample size is 2000.
   (d) The sample size is 1360.
   (e) The true proportion of all eligible voters who actually voted in the previous presidential election is 0.68.

   Use the following information to answer the questions 6 to 7.
   During a test, the class average is 74 with a standard deviation of 6. Assume that the test scores are symmetrically distributed.

6. What is the approximate proportion of all test scores which are above 86?

   (a) 2.5% (b) 5% (c) 10% (d) 13.5% (e) 34%

7. Emmy's score in this exam was 65. What was her corresponding z-score in this test?

   (a) 1.50 (b) 1.20 (c) -1.20 (d) -1.50 (e) There is no way to tell.

8. You are tossing a fair coin. Let $X = 1$ if you observe a head and $X = 0$ if you observe a tail. Which of the following tables represents the resulting probability distribution for the random variable X?

(a)

X	$\frac{1}{2}$	$\frac{1}{2}$
$P(X)$	0	1

(b)

X	$\frac{1}{2}$	$\frac{1}{2}$
$P(X)$	$\frac{1}{3}$	$\frac{2}{3}$

(c)

X	0	1
$P(X)$	$\frac{1}{2}$	$\frac{2}{3}$

(d)

X	0	1
$P(X)$	$\frac{1}{2}$	$\frac{1}{2}$

Use the following information to answer questions 9 and 10.
At Hopewell Electronics, all 140 employees were asked about their political affiliations: Democrat, Republican or Independent. The employees were grouped by type of work, as executives or production workers. The results with row and column totals are shown in the following table. Suppose an employee is selected at random from the 140 Hopewell employees.

	Democrat	Republican	Independent	Total
Executive	5	34	9	48
production worker	63	21	8	92
Total	68	55	17	140

9. The probability that this employee is a Republican is about

(a) 0.393 (b) 0.486 (c) 0.121 (d) 0.243 (e) 0.657

10. The probability that this employee is a production worker **or** is a Republican is about

(a) 0.657 (b) 0.90 (c) 0.15 (d) 0.393 (e) 0.507

11. In a survey of 1000 eligible voters, 680 said that they voted in the last presidential election. let p denote the proportion of all eligible voters who actually voted. Which of the following represents 95% confidence interval for p?

(a) $680 \pm (1.96)(\frac{680}{\sqrt{1000}})$

(b) $680 \pm (1.96)\sqrt{\frac{0.68\times0.32}{1000}}$

(c) $0.68 \pm (1.645)\sqrt{\frac{0.68\times0.32}{1000}}$

(d) $0.68 \pm (1.645)\sqrt{\frac{0.68\times0.32}{680}}$

(e) $0.68 \pm (1.96)\sqrt{\frac{0.68\times0.32}{1000}}$

Use the following information for questions 12 to 13.

The following is a probability distribution for a discrete random variable X.

X	-1	0	1	10
$P(X)$	0.2	0.5	0.2	0.1

12. The mean (expected value) of X is

    (a) 2.5 (b) 1.0 (c) -1.0 (d) -2.5 (e) None of the above

13. The **standard deviation** of X is closest to

    (a) 9.40 (b) 3.07 (c) 25.67 (d) 5.07 (e) 1.0

Use the following information to answer questions 14 and 15.
Scores on the common final exam in **Elementary Statistic I** course are normally distributed with mean 72 and standard deviation 8.

14. Find the probability that a random selected student had score **above** 78.

    (a) 0.8413 (b) 0.7734 (c) 0.2266 (d) 0.1596 (e) 0.5468

15. The department decides to give A to all students whose scores are in top 10% on this exam. What is the minimum score for a student to receive A?

    (a) 80 (b) 82.24 (c) 85.16 (d) 87.68 (e) 90

Use the following information to answer questions 16 and 17: The mean annual salary for flight attendants is about \$65,700 with a standard deviation of \$14,400. A random sample of 64 flight attendants is selected from this population. Let $\bar{x}$ represent the mean annual salary of the sample.

16. Find the mean and standard deviation of the sampling distribution, i.e., $(\mu_{\bar{x}}, \sigma_{\bar{x}})$.

    (a) $\mu_{\bar{x}} = 14,400, \quad \sigma_{\bar{x}} = 64$
    (b) $\mu_{\bar{x}} = 14,400, \quad \sigma_{\bar{x}} = 65,700$
    (c) $\mu_{\bar{x}} = 65,700, \quad \sigma_{\bar{x}} = 14,400$
    (d) $\mu_{\bar{x}} = 65,700, \quad \sigma_{\bar{x}} = 64$
    (e) $\mu_{\bar{x}} = 65,700, \quad \sigma_{\bar{x}} = 1,800$

17. What is the approximate probability that the sample mean salary $\bar{x}$ is less than \$63,400?

    (a) 0.1003 (b) 0.4364 (c) 0.5636 (d) 0.8997 (e) 0.6179

18. You are running a political campaign and wish to construct a 95% confidence interval to estimate the true **proportion** of registered voters who will vote for your candidate. You will conduct a survey and will assume the margin of error $E = .03$. The minimum sample size required in this suvey is

    (a) 215 (b) 752 (c) 1068 (d) 1096 (e) 1842

19. In a random sample of 16 microwave ovens, the sample mean repair cost was \$80 with the sample standard deviation \$13. Assume that the repair costs are normally distributed. Construct a 90% confidence interval for μ, where μ represents the average repair cost for all microwave ovens.

   (a) $80 \pm (1.645)(\frac{13}{\sqrt{16}})$
   (b) $80 \pm (1.96)(\frac{13}{\sqrt{16}})$
   (c) $80 \pm (1.341)(\frac{13}{\sqrt{16}})$
   (d) $80 \pm (1.753)(\frac{13}{\sqrt{16}})$
   (e) $16 \pm (1.753)(\frac{13}{\sqrt{80}})$

Use the following information for questions 20 to 21.
Do larger universities tend to have more property crimes? University crime statistics are affected by a variety of factors such as surrounding community, accessibility given to outside visitors, etc. Let x represent student enrollment (in thousands) and let y represent the number of burglaries in a year on the university campus. A random sample of $n = 8$ universities in California yielded the following data regarding the enrollments and annual burglary incidents.

x	12.5	30.0	24.5	14.3	7.5	27.7	16.2	20.1
y	26	73	39	23	15	30	15	25

The equation of the regression line relating y to x as well as the coefficient of correlation are computed to be

$$\hat{y} = -4.13 + 1.83x, \qquad r = 0.76$$

20. The predicted number of annual burglary incidents for a California university with 17 (thousands) students is about

   (a) 15 (b) 20 (c) 27 (d) 30 (e) 39

21. Which of the following conclusions may be made?

   (a) The number of burglaries and student enroll number are very poorly correlated.
   (b) The number of burglaries and student enroll number are very strongly correlated, and burglaries tends to increase as the enroll number decreases.
   (c) The number of burglaries and student enroll number are moderately correlated, and burglaries tends to increase as the enroll number decreases.
   (d) The number of burglaries and student enroll number are moderately correlated, and burglaries tends to increase as the enroll number increases.

22. Determine the minimum sample size required when you want to be 95% confident that the sample mean is within one unit of the population mean (i.e., $E = 1$) and $\sigma = 4.8$. Assume the population is normally distributed.

   (a) 30 (b) 63 (c) 89 (d) 153 (e) 538

23. In the test of hypothesis $H_0 : \mu = 100$ vs $H_a : \mu \neq 100$, a sample of size is 250 yields the standardized test statistic $z = 1.28$. The p-value for the test is closest to

(a) 1.80 (b) 0.90 (c) 0.10 (d) 0.43 (e) 0.20

24. A pet association claims that the mean annual costs of food for dogs and cats are the same. The results for samples of the two types of pets are shown in the following table. On wished to test the claim. Assume both populations follow normal distributions

Dogs	Cats
$n_1 = 16$	$n_2 = 18$
$\bar{x}_1 = 239$	$\bar{x}_2 = 203$
$s_1 = 32$	$s_2 = 28$

with equal variances. Which of the following formula should be used to calculate the standardized test statistic?

(a) $Z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

(b) $Z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2 + s_2^2}{n_1 + n_2}}}$

(c) $T = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

(d) $T = \frac{\bar{d} - D_0}{s_D/\sqrt{n}}$

(e) $Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

The following is used for questions 25 to 27.
In a random survey of 1000 people in the United States, 790 said that they prepare and file their income taxes before April 15th. Let p be the true proportion of people in the United States prepare and file their income taxes before April 15th. One wants to test the following hypotheses

$$H_0 : p = 0.75 \qquad H_a : p > 0.75.$$

25. The standardized test statistics is about

(a) -3.11 (b) -2.92 (c) 2.92 (d) 3.11 (e) 4.11

26. Suppose one already calculated the P-value for the above test is 0.0018, which of the following will be the conclusion for $\alpha = .01$?

(a) Reject H_0
(b) Fail to reject H_0.
(c) No way to tell.

27. Assume the significant level for this test is 1%. Find the rejection region for the test.

(a) $(-\infty, 2.33)$
(b) $(-\infty, -2.33)$
(c) $(1.96, \infty)$
(d) $(2.33, \infty)$
(e) $(-\infty, 1.96)$

The following is used for questions 28 to 30.
A teacher proposes a course designed to increase reading speed (measured by words per minute). To evaluate the effectiveness of this course, the teacher tests students before and after the course. The result is following:

Student	1	2	3	4	5
Speed before class	100	160	120	167	200
Speed after class	136	170	135	169	210

The difference in the reading speeds (before − after) for this sample results in $\bar{d} = -14.6$ and $s_d = 12.84$. Assume that the reading speeds are approximately normally distributed. Let μ_d be the mean difference between two populations (i.e., $\mu_d = \mu_{before} - \mu_{after}$).

28. Is the course effective to **improve** the reading speed? Choose the appropriate hypotheses to test the claim.

(a) $H_0 : \mu_d = 0$ versus $H_a : \mu_d \neq 0$
(b) $H_0 : \bar{d} = 0$ versus $H_a : \bar{d} > 0$
(c) $H_0 : \mu_d = 0$ versus $H_a : \mu_d > 0$
(d) $H_0 : \mu_d = 0$ versus $H_a : \mu_d < 0$.
(e) $H_0 : \bar{d} = 0$ versus $H_a : \bar{d} < 0$.

29. The standardized test statistics is about

(a) 2.54 (b) -0.51 (c) 0.51 (d) -1.14 (e) -2.54

30. Find the rejection region at $\alpha = .05$.

(a) $(1.645, \infty)$
(b) $(-\infty, -1.645)$
(c) $(-\infty, -2.132)$
(d) $(2.132, \infty)$
(e) $(-\infty, -2.776)$

End of Multiple Choice Section

1. In June 2002, chemical analyses of $n_1 = 110$ water samples from various parts of Catawba River were made. In June 2004, the experiment was repeated using $n_2 = 85$ samples. We have the following result:

Chlorine Content (2002)	Chlorine Content (2004)
$\bar{x}_1 = 17.8$	$\bar{x}_2 = 18.3$
$s_1 = 1.8$	$s_2 = 1.2$

One wants to test whether the mean Chlorine content in the river has **increased** from 2002 to 2004?

(a) Set up the correct hypothesis to test the claim. [3 pts.]

H_0: H_a :

(b) Find the standardized test statistic. [3 pts.]

(c) Find the rejection region with $\alpha = 0.05$. [3 pts.]

(d) Find P-value for the above test. [3 pts.]

2. In the past, it is generally agreed that a certain standard treatment yields a mean survival period of 4.2 years for certain cancer patients. Recently, a new treatment is administered to 25 patients and their duration of survival is recorded. The sample mean and standard deviation of the duration is 4.5 years and 0.8 years, respectively. Assume that all survival times follow a normal distribution.

   (a) Set up the null and alternative hypotheses to test whether the new treatment **increases** the mean survival period. [3 pts.]

   (b) In the context of the problem, explain Type I and Type II errors. [3 pts.]

   (c) Find the value of the standardized test statistic. [3 pts.]

   (d) Find the rejection region at $\alpha = .05$ and state your conclusion in the context of the problem. [3 pts.]

3. In one of Boston's public parks, mugging in summer months has been a serious issue. A police cadet took a random sample of 10 days and compiled the data. For each day, x represents the number of police officers on duty in the park and y represents the number of reported muggings on that day.

x	10	15	16	1	4	6	18	12	14	7
y	5	2	1	9	7	8	1	5	3	6

$n = 10, \ \sum x = 103, \ \sum y = 47, \ SS_{xx} = 286.1, \ SS_{yy} = 74.1, \ SS_{xy} = -141.1.$

(a) Construct a scatter plot for the data. [2 points]

(b) From the above scatter plot what information can we learn? [2 points]

(c) Find the linear correlation coefficient between x and y and interpret its meaning in the context of the problem. [3 points]

(d) Find the equation of the regression line between y and x. [3 points]

(e) Can we predict the number of reported mugging if there were 30 police officer on duty in a day? Why or why not? [3 points]

(f) At $\alpha = .05$, one wishes to test $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$, where β_1 is the slope of the population regression line. Please propose a test statistic. **(No actual calculation needed.)** [3 points]