STAT 1220
Common Final Exam

FALL 2019
December 6, 2019

PLEASE PRINT THE FOLLOWING INFORMATION:

Name: _____   Instructor: _____

Student ID #: _____   Section/Time: _____

THIS EXAM HAS TWO PARTS.

PART I.

Part I consists of 30 multiple choice questions. Each correct answer is scored 2 points; each incorrect (or blank) answer is scored 0, so there is no penalty for guessing. You may do calculations on the test paper, but your answers must be marked on the OPSCAN sheet with a soft lead pencil (HB or No. 2 lead). Any question with more than one choice marked will be counted as incorrect. If more than one choice seems correct, chose the one that is most complete or most accurate. Make sure that your name and ID number are written and correctly bubbled on the OPSCAN sheet.

PART II.

Part II consists of 3 free response questions, with values as indicated. You must show all work in the space provided or elsewhere on the exam paper in a place that you clearly indicate. Work on loose sheets will not be graded.

FOR DEPARTMENT USE ONLY:

Part II.

Question	1	2	3
Score			

Part I	Part II	TOTAL

Part I

Problems 1 and 2 pertain to the following information: The following is a sample of maximum temperature readings (in Degrees Celsius) in May 2018 in Chicago:

$$5, \quad -2, \quad 0, \quad -3, \quad 2, \quad 6, \quad 1, \quad 4, \quad -1$$

1. The sample mean of this data set is about

  (a) 1.5   (b) 1.33   (c) 9.0   (d) 12.0   (e) 2.92

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

  (a) 2.98   (b) 80.00   (c) 3.16   (d) 10.00   (e) 1.33

Problems 3–5 pertain to following information: The ages of 77 richest people as listed in the *Forbes Magazine* are given by the following frequency distribution table:

x	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76
f	1	2	2	3	4	5	4	6	8	12	9	6	6	4	3	2

3. The sample median is about:

  (a) 71   (b) 70   (c) 71.5   (d) 72.5   (e) impossible to tell from the table

4. The sample range is about:

  (a) 15   (b) 77   (c) 78   (d) 7.5   (e) 11

5. The percentile rank of the measurement 73 is about:

  (a) 88   (b) 81   (c) 75   (d) 86   (e) 78

6. The distribution of the ages of accountants at Three River Corp. is roughly bell-shaped with a mean of 27 years and a standard deviation of 2.5 years. The percentage of accountants that have an age at most 32 is about:

  (a) 50%   (b) 68%   (c) 95%   (d) 97.5%   (e) 47.5%

7. In order to determine the average price per gallon of regular gasoline, the U.S. department of energy conducted a survey of 1000 gas stations in the country and found the average price to be $2.27 per gallon in the Thanksgiving weekend last year. The *population* of interest is

  (a) All the 1000 gas stations from which price data were collected.
  (b) All gas stations in the country.
  (c) All gas stations excluding the 1000 gas stations from which price data were collected.
  (d) All gas stations where the price per gallon is close to $ 2.27 per gallon.
  (e) None of the above.

8. The *mode* of a numerical data set measures the _____ of the data set.

   (a) range

   (b) percentile

   (c) most frequent observation

   (d) variability

   (e) size

Problems 9 and 10 pertain to the following information: A consumer agency randomly selected 1700 flights of two major airlines, A and B. The following table gives the two-way classification of the flights based on airlines and arrival time. Note that "less than 30 minutes late" includes flights that arrive early or on time. One flight is selected at random.

	Less Than 30 Minutes Late	30 Minutes to 1 Hour Late	More than 1 Hour Late
Airline A	429	390	92
Airline B	393	316	80

9. The probability that a randomly selected flight is more than 1 hour late and from airline A is about:

   (a) 0.535   (b) 0.054   (c) 0.946   (d) 0.101   (e) 0.087

10. The probability that a randomly selected flight is less than 30 minutes late, given that it is from airline B, is about:

   (a) 0.498   (b) 0.231   (c) 0.393   (d) 0.326   (e) 0.252

Problems 11 and 12 pertain to the following information: In a group of 2500 persons, 1400 are females, 600 are vegetarian, and 400 are female and vegetarian.

11. The probability that a randomly selected person from this group is a male or a vegetarian is about:

   (a) 0.40   (b) 0.60   (c) 0.88   (d) 0.26   (e) impossible to tell (insufficient information)

12. The events F: *a randomly selected adult is female* and V: *a randomly selected adult is vegetarian* are:

   (a) Independent because $P(F) \cdot P(V) = P(F \cap V)$

   (b) Independent because $P(F) \cdot P(V) \neq P(F \cap V)$

   (c) Dependent because $P(F) \cdot P(V) = P(F \cap V)$

   (d) Dependent because $P(F) \cdot P(V) \neq P(F \cap V)$

   (e) impossible to tell if independent or not (insufficient information)

Problems 13–15 pertain to the following information: Choose an American household at random and let the random variable X be the number of cars (including SUV's and light trucks) they own. Here is the probability distribution if we ignore the households that own more than 4 cars.

x	0	1	2	3	4
$P(x)$	p	0.41	0.27	0.08	0.03

13. The missing entry p is about:

   (a) 0.79   (b) 0.07   (c) 0.18   (d) 0.21   (e) 0.00

14. The probability that a randomly selected household will exceed two cars is about:

   (a) 0.62   (b) 0.38   (c) 0.89   (d) 0.11   (e) 0.27

15. The average number of cars is about:

   (a) 1.31   (b) 2.00   (c) 0.21   (d) 2.62   (e) 2.50

16. According to Harris Interactive poll, 52% of American college graduates have Facebook accounts. Suppose this result is true for the current population of American graduates. The probability that exactly 9 American college graduates in a sample of 15 have Facebook account is about:

   (a) 0.0520   (b) 0.8298   (c) 0.1702   (d) 0.3404   (e) 0.6000

17. According to Harris interactive survey conducted for world vision and released in February 2009, 56% of teens volunteered time for charitable causes. Assume that this result is true for the current population of U.S. teens. A sample of 60 teens is selected. The average number of teens who volunteer time for charitable causes is about:

   (a) 28.2   (b) 56.1   (c) 30.2   (d) 33.6   (e) 60.3

18. Hupper corporation produced many types of soft drinks, including orange cola. The filling machines are adjusted to pour 12 ounces of soda into each 12-ounce can of Orange Cola. However, the actual amount of soda poured into each can is not exactly 12 ounces; it varies from can to can. It has been observed that the net amount of soda in such a can has normal distribution with a mean of 12 ounces and standard deviation of 0.015 ounce. The probability that a randomly selected can of Orange Cola contains 11.97 to 11.99 ounces of soda is about:

   (a) 0.0228   (b) 0.2514   (c) 0.2286   (d) 0.9772   (e) 0.7486

19. Assume that the monthly electric power consumptions during winter by all households in a city have a normal distribution with a mean of 1650 kilowatt-hours and a standard deviation of 320 kilowatt-hours. The company sent a notice to Bill Johnson from that city informing him about 90% of the households use less electricity per month than he does. Bill Johnson's monthly electricity consumption (in kwh) is about:

(a) 2060   (b) 1650   (c) 320   (d) 1970   (e) 2290

20. According to Sallie Mae surveys and Credit Bureau data, college students carried an average of $3173 credit card debt in 2018. Suppose the probability distribution of the current credit card debts of all college students in the United States is unknown but its mean is $3173 and the standard deviation is $750. A random sample of 400 college students is selected. The probability that the mean of the current credit card debts for this sample is within $70 of the population mean is about:

(a) 0.9693   (b) 0.0614   (c) 0.0307   (d) .0718   (e) 0.9386

21. Brooklyn Corporation manufactures CDs. The machine that is used to make these CDs is known to produce 6% defective CDs. The quality control inspector selects a sample of 100 CDs every week and inspects them for being good or defective. If 8% or more CDs in the sample are defective, the process is stopped and the machine is readjusted. The probability that based on a sample of 100 CDs, the process will be stopped to readjust the machine is about:

(a) 0.7995   (b) 0.8400   (c) .2005   (d) 0.0800   (e) 0.0600

22. A consumer agency that proposes that lawyers' rates are too high, wanted to estimate the mean hourly rate for all lawyers in New York city. A sample of 70 lawyers taken from New York city showed that the mean hourly rate charged by them is $420 and standard deviation is $110. The 99% confidence interval for the mean hourly charges for all lawyers in New York city is about:

(a) [406.86, 433.14]

(b) [310.11, 530.22]

(c) [200.22, 640.33]

(d) [386.08, 453.92]

(e) [350.08, 490.16]

23. A jumbo mortgage is a mortgage with a loan amount above the industry-standard definition of conventional conforming loan limits. As of January 2019, approximately 2.57% of people who took out a jumbo mortgage during the previous 12 months were at least 60 days late on their payments. Suppose that this percentage is based on a random sample of 1430 people who took out a jumbo mortgage during the previous 12 months. The 95% confidence interval for the proportion of all people who took out a jumbo mortgage during the previous 12 months and were at least 60 days late on their payments is about:

    (a) 0.9743 ± 0.0028

    (b) 0.0257 ± 0.0082

    (c) 0.0257 ± 0.0042

    (d) 0.9743 ± 0.05

    (e) 0.0257 ± 0.9918

24. Jack's Auto Insurance Company customers sometimes have to wait a long time to speak to a customer service representative when they call regarding disputed claims. Assume that the waiting times for the population follow a normal distribution. A random sample of 25 such calls yielded a mean waiting time of 22 minutes with a standard deviation of 6 minutes. The 99% confidence interval for the population mean of such waiting times is about:

    (a) 22 ± 3.36 (b) 22 ± 3.09 (c) 22 ± 2.48 (d) 6 ± 3.09 (e) 22 ± 6

25. A consumer agency wants to estimate, with 99% confidence and to within three percentage points, the proportion of all drivers who wear seat belts while driving. A preliminary study has shown that 76% of all drivers wear seat belts while driving. The minimum sample size required to meet this objective is about:

    (a) 1097   (b) 779   (c) 760   (d) 1849   (e) 1350

26. In a test of hypotheses of the form $H_0 : \mu = 12.5$ versus $H_a : \mu \neq 12.5$ using $\alpha = 0.01$, when the sample size is 18 and the population is normally distributed but of unknown standard deviation, the rejection region will be the interval

    (a) $(-\infty, -2.121] \cup [2.121, \infty)$

    (b) $(-\infty, -2.576] \cup [2.576, \infty)$

    (c) $(-\infty, -2.326] \cup [2.326, \infty)$

    (d) $(-\infty, -2.898] \cup [2.898, \infty)$

    (e) $(-\infty, -1.740] \cup [1.740, \infty)$

27. In a test of hypotheses $H_0 : \mu = 50$ versus $H_a : \mu < 50$ in a normally distributed population, the rejection region is the interval $(-\infty, -1.895]$, the value of the sample mean computed from a sample of size 8 is $\bar{x} = 44.98$, and the value of the test statistic is $t = -2.097$. The correct decision and justification are:

   (a) Reject H_0 because 44.98 is less than 50.

   (b) Reject H_0 because 40.98 lies in the rejection region.

   (c) Reject H_0 because $-2.097 < -1.895$.

   (d) Do not reject H_0 because the sample is small.

   (e) Do not reject H_0 because $-1.895 > -2.097$.

28. In a test of hypotheses $H_0 : \mu = 90$ vs. $H_a : \mu < 90$ at 10% level of significance from a normally distributed population with standard deviation 7, a sample of size 20 resulted in a mean of 85 which produced the test statistic $z = -3.19$. The p-value (the observed significance) of the test is about:

   (a) 0.1000   (b) 0.9993   (c) 0.0007   (d) 0.0014   (e) .9000

29. A company wanted to know if attending a course on "how to be a successful salesperson" can increase the average sales of its employees. The company sent six of its sales persons to attend this course. The average sales of these employees after attending this course (x_2) are compared to their average sales before attending this course (x_1). The difference is defined as $d = x_1 - x_2$. The setup of the null and alternative hypotheses for this test must be:

   (a) $H_0 : \bar{x}_1 - \bar{x}_2 = 0$ vs. $H_a : \bar{x}_1 - \bar{x}_2 > 0$

   (b) $H_0 : \bar{d} = 0$ vs. $H_a : \bar{d} < 0$

   (c) $H_0 : \mu_d = 0$ vs. $H_a : \mu_d > 0$

   (d) $H_0 : \mu_d = 0$ vs. $H_a : \mu_d \neq 0$

   (e) $H_0 : \mu_d = 0$ vs. $H_a : \mu_d < 0$

30. A researcher took a sample of 25 electronics companies and found the relationship $\hat{y} = 11.75x + 3.6$ between x and y where x represents the amount of money (in millions of dollars) spent on advertising and y represents the total gross sales (in millions of dollars) of that company in 2018. For each additional million spent on advertising, the total gross sales (in millions of dollars) :

   (a) increases by about 3.6 millions, on average

   (b) increases by about 11.75 millions, on average

   (c) increases by about 15.35 millions, on average

   (d) increases by about 25 millions, on average

   (e) changes by an amount that cannot be determined from the information given

Part II

1. A mail-order company claims that more than 60% mails are mailed within 48 hours. From time to time, the quality control department at the company checks if this promise is fulfilled. Recently, the quality control at this department took a sample of 400 orders and found that 208 of them were mailed within 48 hours of the placement of orders. In order to test the company's claim at 1% level of significance:

(a) State the null and alternative hypotheses for the test. [2 points]

(b) State the formula for the test statistic and compute its value. Justify your answer. [4 points]

(c) Construct the rejection region and make a decision. [4 points]

(d) State a conclusion about the proportion of all mails mailed within 48 hours, based on the test you performed. [2 points]

(e) Compute the p-value (the observed significance) of the test and state what it means in the context of this problem. [2 points]

2. The U.S. Department of Agriculture (USDA) uses sample surveys to produce important economic estimates. One pilot study estimated wheat prices (in dollars/bushel) in July and in January using independent samples of wheat producers in the two months. A randomly selected 15 wheat producers in July had a mean of 6.93 and standard deviation of 0.27. A randomly selected 10 wheat producers in January had a mean of 6.66 and standard deviation of 0.24. Assume that the wheat prices of both the months are normally distributed with equal but unknown standard deviations. In order to test the claim at 5% level of significance that the national average wheat prices in the two months are different:

(a) State the null and alternative hypotheses for the test. [2 points]

(b) State the formula for the test statistic and compute its value. Justify your answer. [4 points]

(c) Construct the rejection region and make a decision. [4 points]

(d) State a conclusion in the context of the problem, based on the test you performed. [2 points]

3. The following summary information is about ticket price x (in dollars) and average attendance y (rounded to the nearest thousand) for the last year for six football games:

$$26.50 \leqslant x \leqslant 59.50 \qquad 42 \leqslant y \leqslant 71 \qquad \bar{x} = 40 \qquad \bar{y} = 59.67$$

$$SS_{xx} = 647 \qquad SS_{xy} = -103.5 \qquad SS_{yy} = 591.33 \qquad SSE = 574.77$$

(a) Find the proportion of the variability in average attendance that is accounted for by ticket price. [4 points]

(b) Find the regression line for predicting y from x. [4 points]

(c) Find the average attendance predicted for by the regression equation found in part (b) when the ticket price is \$35. [1 point]

(d) Construct a 95% confidence or prediction interval, whichever is appropriate, for average attendance. [5 points]