ACTM State Exam-Statistics

For the 25 multiple-choice questions, make your answer choice and record it on the answer sheet provided. Once you have completed that section of the test, proceed to the tie-breaker items. These will be graded in numerical order, so it's in your best interest to respond to these in that order. Be sure that your name is on each sheet of the tie-breaker items.

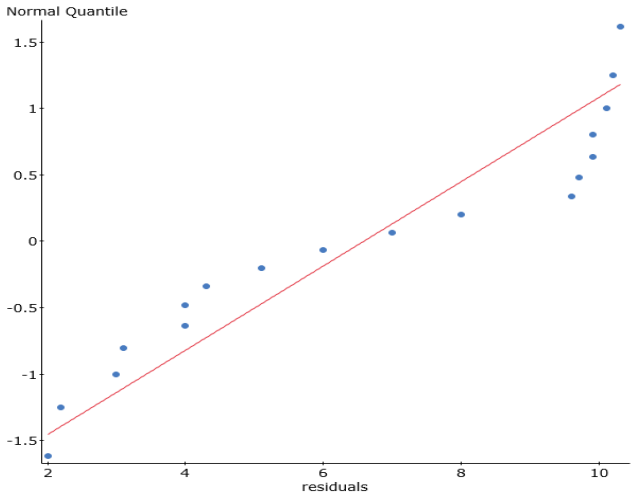
1. Generally, Greek letters are used to represent _____, and Roman letters are used to represent _____.
    a. Statistics, parameters
    b. Parameters, statistics
    c. Measures of variation, measures of spread
    d. Measures of spread, measures of variation

2. If the mode is to the left of median and the mean is to the right of the median, then the distribution is _____ skewed.
    a. Positively
    b. Negatively
    c. Not
    d. Highly

3. A set of 45 values has a mean of 23 and a set of 53 values has a mean of 47, what is the mean of the combined set of values? (Round to nearest tenths if needed.)
    a. 49.0
    b. 50.4
    c. 35.0
    d. None of these

4. Suppose we use a least squares linear regression model on a set of data points (x,y). We find the coefficient of determination is 0.81 and the regression line is given by $y = -51x + 32$. What is the value of the correlation coefficient?
    a. 0.6561
    b. 0.9
    c. -0.6561
    d. -0.9

5. Given the same scenario as in question #4 and the QQ plot of residuals below, do you believe the linear model is a good one?



   a. Yes, because the absolute value of the correlation coefficient is close to one and the residuals are close to the identity line.
   b. No, because the absolute value of the correlation coefficient is not close to one.
   c. No, because the QQ plot shows that the residuals are not normally distributed.
   d. Yes, because the QQ plot shows that the residuals are normally distributed and the absolute value of the correlation coefficient is close to one.

6. The following list gives number of hours a student studies per week. What is the largest value in the sample that is not considered an outlier?

   0, 1, 2, 2, 5, 5, 5, 6, 6, 7, 7, 7, 8, 8, 8, 8, 8, 8, 9, 10, 10, 10, 12, 20

   a. 12
   b. 20
   c. 0
   d. None of the above

7. You step onto a bus that will take between 15 and 25 minutes to arrive on campus. Assume that the distribution of these times is uniform, your statistics class begins in 20 minutes, and it will take two minutes for you to walk from the bus to the classroom. What is the probability that you arrive late to class?
   a. 70%
   b. 30%
   c. 50%
   d. None of the above

8. A drawer contains 10 different pairs of gloves. If 6 gloves are drawn randomly from the drawer, what is the probability that there will be no matching pair in the sample?
    a. $\frac{C(10,6)}{C(20,6)}$
    b. $\frac{2C(10,6)}{C(20,6)}$
    c. $\frac{2^6C(10,6)}{C(20,6)}$
    d. None of these

9. A family has two children. Given that at least one of them is a girl, was it the probability that the family has two girls?
    a. 1/2
    b. 1/3
    c. 1/4
    d. ¾

10. Suppose we know that at least 88.89% of the members of a gym have an age between 32 years 62 years. What can you conclude using Chebyshev's theorem about the mean and standard deviation of the distribution of ages for the members?
    a. You cannot draw a conclusion from the information given.
    b. The standard deviation is 3 years, but we cannot find the mean.
    c. The mean is 47 years and the standard deviation is 5 years.
    d. The mean is 47 years and the standard deviation is 1.67 years.

11. Red jelly beans are my favorite. If I randomly select a jelly bean from bowl, the probability of drawing a red one is 0.21. Suppose I grab a handful of 13 jelly beans, what is the probability of getting at least 5 that are red?
    a. 0.0797
    b. 0.1173
    c. 0.8827
    d. 0.9625

12. The random variable X is the number that shows up when a loaded 4-sided die is rolled. Its probability distribution is given in the table. Determine the mean of the distribution of X.

x	1	2	3	4
P(X=x)	0.30	0.30	0.15	0.25

    a. 2.35
    b. 2
    c. 0.5875
    d. 2.5

13. In a data set with a range of 52.2 to 114.5 and 500 observations, there are 345 observations with values less than 87.1. Find the percentile for 87.1.
    a. 69
    b. 131.458
    c. 698
    d. 32

14. Which score has a higher relative position, a score of 48 on a test for which the mean is 40 and standard deviation is 4, or a score of 288.6 on a test for which the mean is 260 and standard deviation is 26?
    a. A score of 48
    b. A score of 288.6
    c. Both scores have the same relative position

15. The brand manager for a brand of toothpaste must plan a new campaign designed to increase brand recognition. A previous survey suggested that about 76% of adults have heard of the brand. He wants to first determine the present percentage of adults who have heard of the brand. How many adults must he survey in order to be 80% confident that his estimate is within six percentage points of the true population percentage?
    a. At least 84
    b. At least 83
    c. At least 115
    d. At least 116

16. Which of the following is NOT true of the confidence level of a confidence interval?
    a. The confidence level gives us the success rate of the procedure used to construct the confidence interval.
    b. The confidence level is often expressed as an area $1-\alpha$, where α is the complement of the confidence level.
    c. The confidence level is also called the degree of confidence.
    d. There is a $1-\alpha$ chance, where α is the complement of the confidence level, that the true value of the parameter will fall in the confidence interval produced by the sample.

17. Using a 95% confidence level, find a confidence interval for estimating the true mean weight, μ, of all packages received by the parcel service given that a sample of 31 packages had a mean weight of 13.2 lb. and the standard deviation of all packages is $\sigma=2.4$ lbs.
    a. $12.2 \text{ lb} < \mu < 14.2 \text{ lb}$.
    b. $12.4 \text{ lb} < \mu < 14.0 \text{ lb}$.
    c. $12.1 \text{ lb} < \mu < 14.3 \text{ lb}$.
    d. $12.5 \text{ lb} < \mu < 13.9 \text{ lb}$.

18. The mean replacement time for a random sample of 20 washing machines is 10.8 years and the standard deviation 2.8 years. Construct a 99% confidence interval for the standard deviation, σ, of the replacement times of all washing machines of this type.
    a. 2.5 years < σ < 5.9 years
    b. 2.0 years < σ < 4.7 years
    c. 0.3 years < σ < 5.3 years
    d. 1.2 years < σ < 4.4 years

19. A farmer has decided to use a new additive to grow his crops. He divided his farm into 10 plots and kept records of the corn yield (in bushels) before and after using the additive. The results are in the table below. Using a 1% level of significance, find the p-value for testing the claim that the additive produces a higher mean yield.
    a. 0.1355
    b. 0.0004
    c. 0.0070
    d. 0.2709

Plot	Before	After
1	9	10
2	9	9
3	8	9
4	7	8
5	6	7
6	8	10
7	5	6
8	9	10
9	10	10
10	11	12

20. Which of the following is not an advantage of pooling sample variances?
    a. Confidence intervals are a little narrower.
    b. The number of degrees of freedom is a little higher.
    c. Hypothesis tests have more power.
    d. We often know that the population variances are equal.

21. A study was done using a treatment group and a placebo group. We want to test the claim that the samples are from populations with the same mean. The results are shown in the table. Assume that the two samples are independent simple random samples from normally distributed populations. Using a 0.01 significance level, state the conclusion of the test.

	Treatment	Placebo
Sample size	26	38
Sample mean	2.33	2.64
Sample standard deviation	0.51	0.99

a. Reject the null hypothesis. There is sufficient evidence to warrant rejection of the claim that the two samples are from populations with the same mean.
b. Reject the null hypothesis. There is not sufficient evidence to warrant rejection of the claim that the two samples are from populations with the same mean.
c. Fail to reject the null hypothesis. There is not sufficient evidence to warrant rejection of the claim that the two samples are from populations with the same mean.
d. Fail to reject the null hypothesis. There is sufficient evidence to warrant rejection of the claim that the two samples are from populations with the same mean.

22. The activation times for an automatic sprinkler system are a subject of study by the system's manufacturer. A sample of activation times is 27, 41, 22, 27, 23, 35, 30, 33, 24, 27, 28, 22, and 24 seconds. The design of the system calls for its activation in at most 25 seconds. Assuming a normal population, does the data contradict the validity of this design specification?
    a. Yes
    b. No
    c. Not enough data to draw a conclusion

23. Calls come into a telephone switchboard at a rate of 4 per minute. Find the probability of getting 6 calls in an interval of 2 minutes rounded to three places. Use a Poisson distribution.
    a. 0.122
    b. 0.104
    c. 0.313
    d. None of these

24. Suppose we are interested in studying the factors that affect GPA of senior high school students. Fifty schools are selected at random and we collect the GPA of one male and one female from each school. We also classify each school as public or private. Which of the following procedures is most appropriate to conduct first.
    a. An independent 2 sampled t test on the male and female GPA.
    b. A paired t test on the male and female GPA.
    c. ANOVA on the four sets of GPAs: males at private school, females at private school, males at public school, females at public school.
    d. Two-Way ANOVA using the school status as one factor and gender as the second factor.

25. The intercept in linear regression represents:
    a. The strength of the relationship between x and y.
    b. The expected x-value when y is zero.
    c. The expected y-value when x is zero.
    d. The error in the measurements.

Tie Breakers

Name: _____

1. Your teacher tells the class that the test scores on the most recent exam were normally distributed with mean of 82% and a variance of 4%. You are disappointed because your score was 76% and usually you have the top score in the class! So you randomly select 5 of your 25 classmates and find out their test scores were 56%, 78%, 84%, 67%, and 81%. Assuming the five scores you collected were a simple random sample, answer the following questions.
    a. What is the probability that a randomly selected student receives a score of 76% or lower on the exam?

    b. Assuming the teacher was truthful, what is the probability that the average of 5 randomly selected students would be higher than 83%.

    c. Based on the five scores you know, test the claim that the average for the class was actually less than 82%. Use 0.05 significance level.

    d. Suppose now that the teacher had only told you that the scores were normally distributed with a mean of 82%, but he/she did not calculate the variance for the class. Based on the five scores you know, test the claim that the average for the class was actually less than 82%. Use 0.05 significance level.

    e. Describe the interpretations of a type I and type II error in this scenario.

Name: _____

2. In a bolt factory, machines A, B and C manufacture 20%, 30% and 50% (respectively) of the total output. Of their output, 3%, 2% and 1% are defective.
    a. Suppose 500 bolts are made every hour, how many defective bolts do you expect to find over the course of a 40 hour work week?

    b. If a bolt is selected at random, what is the probability that it is defective?

    c. A bolt is selected at random and found to be defective, what is the probability that it came from machine A?

3. An Experiment was conducted to compare the effectiveness of three training programs, A, B, and C, in training assemblers of a piece of electronic equipment. Fifteen employees were randomly assigned, five each to the three programs. After completion of the courses, each person was required to assemble four pieces of the equipment, and the average length of time required to complete the assembly was recorded. Several of the employees resigned during the course of the program; the remainder were evaluated, producing the data shown in the accompanying table. Use the following output to answer the questions.

Training Program	Average Assembly Time (min)				
A	59	64	57	62	
B	52	58	54		
C	58	65	71	63	64

Source	DF	SS	MS	F	P
Program	2	170.5	85.2	5.7	.025
Error	9	134.5	14.9		
Total	11	304.9			

Level	N	Mean	StDev
A	4	60.500	3.109
B	3	54.667	3.055
C	5	64.200	4.658

    a. Do the data provide sufficient evidence to indicate a difference in mean assembly times for people trained by the three programs? Use a significance level of 0.05.

    b. Find simultaneous 99% confidence intervals for the mean assembly times between persons trained by programs A, B, and C. What can you conclude based on these confidence intervals?

Solutions

1. b
2. a
3. d
4. d
5. c
6. a
7. b
8. c
9. b
10. c
11. b
12. a
13. a
14. a
15. a
16. d
17. b
18. b
19. b
20. d
21. c
22. a
23. a
24. d
25. c

Solutions to Tie Breakers

1. Your teacher tells the class that the test scores on the most recent exam were normally distributed with mean of 82% and a variance of 4%. You are disappointed because your score was 76% and usually you have the top score in the class! So you randomly select 5 of your 25 classmates and find out their test scores were 56%, 78%, 84%, 67%, and 81%. Assuming the 6 scores you collected were a simple random sample, answer the following questions.
    a. What is the probability that a randomly selected student receives a score of 76% or lower on the exam?

P(X<76)=0.0013 by calculator. If they use empirical rule with z-score of -3, the probability is approximately 0.0015.

    b. Assuming the teacher was truthful, what is the probability that the average of 5 randomly selected students would be higher than 83%.

The sampling distribution of the mean has a mean of 82 and variance of 4/5.

P($\overline{X}$>83)=0.1318

    c. Based on the 6 scores you know, test the claim that the average for the class was actually less than 82%. Use 0.05 significance level.

H_o: μ=82 vs. H_1: μ<82 Z stat=-9,839 and p-value is very, very close to zero. Reject the null hypothesis. We have evidence that the class average was probably less than 82%. We should conclude that maybe the teacher miscalculated the average or the variance of the class.

    d. Suppose now that the teacher had only told you that the scores were normally distributed with a mean of 82%, but he/she did not calculate the variance for the class. Based on the 6 scores you know, test the claim that the average for the class was actually less than 82%. Use 0.05 significance level.

H_o: μ=82 vs. H_1: μ<82 T stat=-1.702 and p-value is 0.082. Fail to reject the null hypothesis. We do not have evidence that the class average was less than 82%.

    e. Describe the interpretations of a type I and type II error in this scenario.

A type I error would mean that we reject the null hypothesis when in fact the null hypothesis is true. Here a type I error would mean that we would conclude the teacher was wrong (or lied) about the mean of the class based on the sample collected, but in fact the average really was 82%.

A type II error would mean that we fail to reject the null hypothesis when the null hypothesis is in fact false. Here a type II error would mean that we would conclude there was not enough evidence for saying the teacher was incorrect, but actually the average of the class was less than 82%.

2. In a bolt factory, machines A, B and C manufacture 20%, 30% and 50% (respectively) of the total output. Of their output, 3%, 2% and 1% are defective.
    a. Suppose 500 bolts are made every hour, how many defective bolts do you expect to find over the course of a 40 hour work week?

A makes 100 bolts with approx. 3 defective per hour.
B makes 150 bolts with approx.. 3 defective per hour.
C makes 250 bolts with approx. 2.5 defective per hour.

Total we expect 8.5 defective bolts per hour or 340 per 40 hours. (They could also use part B)

    b. If a bolt is selected at random, what is the probability that it is defective?

P(defective)=P(defective|A)P(A)+ P(defective|B)P(B)+ P(defective|C)P(C)= 0.03*0.20+0.02*0.30+0.01*0.5=0.017

    c. A bolt is selected at random and found to be defective, what is the probability that it came from machine A?

P(A|defective)=P(defective|A)P(A) / P(defective)= (0.03*0.20) / 0.017 = 0.3529

3. a. For the hypotheses $H_0: \mu_A = \mu_B = \mu_C$; H_1:At least one mean is different.

The critical value is $F_{CV} = 4.26$, Reject H_0 since the test statistic falls inside the right tailed rejection region with p value of 0.025. There is evidence to indicate a difference in the mean assembly times for people trained by the three programs.

b. (Using Fisher's Least Significant Difference) The t-value for all confidence intervals is 3.250 and the estimate for s is square root of MSE= $\sqrt{14.9} \approx 3.86005$

The 99% Confidence intervals for the mean assembly time for programs A, B, and C are

$$54.227 < \mu_A < 66.773$$
$$47.424 < \mu_B < 61.910$$
$$58.590 < \mu_C < 69.810$$

With 99% confidence there doesn't seem to be a significant difference among the three methods. This is because there is an intersection in all confidence intervals.

This is not contradictory to our result because the test in (a) had significance level 0.05 (or confidence 95%). You can verify this by constructing the 95% simultaneous CIs (B and C do not intersect.)

It is more likely that they will compute confidence intervals for each set separately, giving even wider intervals that still overlap.

$$\mathbf{51.42} < \boldsymbol{\mu_A} < \mathbf{69.58}$$
$$\mathbf{37.161} < \boldsymbol{\mu_B} < \mathbf{72.172}$$
$$\mathbf{54.608} < \boldsymbol{\mu_C} < \mathbf{73.792}$$