Statistics Exam

University of Houston Math Contest 2024

1. Given a set of 15 data points we get the following stem-and-leaf plot. What is the correct mean and median of these data points?

The decimal point is 1 digit(s) to the right of the |

```
0 | 4
1 | 00
2 | 466
3 | 2446
4 | 0
5 | 2
6 | 48
7 |
8 | 5
```

a. Mean = 3.63, Median = 3.4
b. Mean = 36.33, Median = 34
c. Mean = 44.5, Median = 44.5
d. Mean = 3.4, Median = 3.63
e. Mean = 4, Median = 4
f. Mean = 15, Median = 3.4

2. Which of the following list of data points has the largest standard deviation?

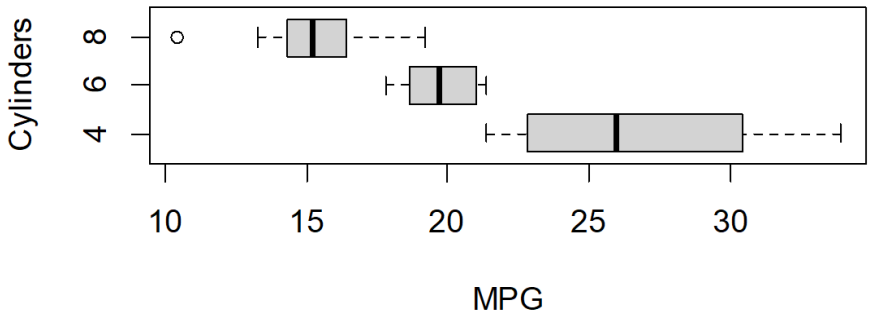
a. 0, 0, 5, 10, 10
b. 0, 2.5, 5, 7.5, 10
c. 0, 1, 5, 9, 10
d. 0, 4, 5, 6, 10
e. 0, 3, 5, 8, 10
f. All of them have the same standard deviation.

3. Given the table below, what proportion of customers will order hot fudge for a sauce given that they ordered vanilla ice cream.

		Ice Cream Flavor		
		Vanilla	Chocolate	Strawberry
Sauce	Hot Fudge	6	3	1
	Carmel	4	1	0
	Strawberry	3	2	5

a. 0.60
b. 0.24
c. 0.40
d. 0.46
e. 0.0
f. None of these above.

4. The boxplot below is the MPG "miles per gallon" of several automobiles by the number of cylinders of the automobile. Which statement cannot be determined by this plot?



a. An 8-cylinder automobile has lower mpg than a 4-cylinder automobile.
b. There is a higher variability of the mpg in the 4-cylinder automobiles than in the 6-cylinder automobiles.
c. The mean mpg of the 8-cylinder automobiles is 15 mpg.
d. The minimum mpg of the 4-cylinder automobiles is the same as the maximum value of the 6-cylinder automobiles.
e. The 8-cylinder automobiles have at least one outlier.
f. All of the statements above can be determined.

5. From 50 data points, the five number summary is:

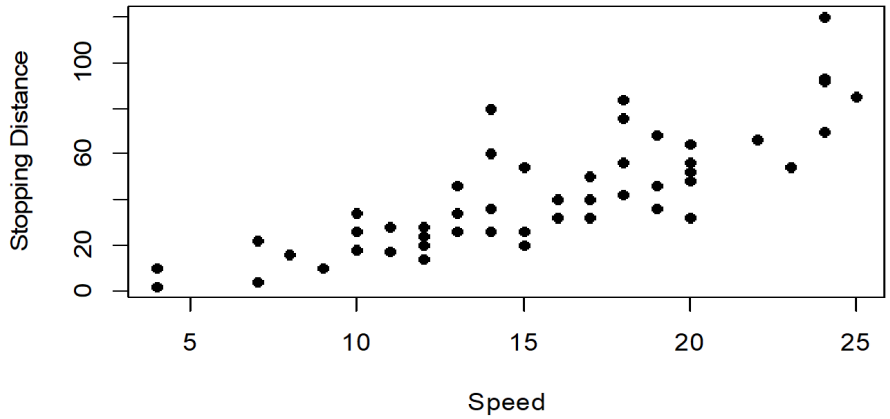
2, 26, 36, 56, 120

The minimum 4 values are: 2, 4, 10, 10
The maximum 4 values are: 85, 92, 93, 120
What value(s) are considered as outliers?

a. 2
b. 2, 4
c. 2, 4, 10
d. 92, 93, 120
e. 120
f. None of these above

6. A list of 5 data points resulted in a mean of 15. Here are 4 of the data points: 14, 15, 24, 14. Determine the value of the missing data point.
a. 15
b. 20
c. 8
d. 10
e. 13
f. None of these above.

7. Given the scatterplot below determine the best correlation between stopping distance and speed.



a. 0.81
b. 0.35
c. 0.99
d. -0.53
e. -0.91
f. 0.02

8. Suppose we have a random sample of 3 data points X_1, X_2, and X_3. These are from a population with mean μ and standard deviation, σ. Suppose we want to estimate the population mean from these three data points, which of the following estimators are unbiased estimates of the population mean?

i. $\hat{\mu}_1 = X_1$
ii. $\hat{\mu}_2 = 0.5\,X_1 + 0.2X_2 + 0.3X_3$
iii. $\hat{\mu}_3 = \frac{X_1+X_2+X_3}{3} = \bar{X}$

a. $\hat{\mu}_1$
b. $\hat{\mu}_2$
c. $\hat{\mu}_3$
d. $\hat{\mu}_1$ and $\hat{\mu}_3$
e. All of them are unbiased estimator
f. None of these are unbiased estimators

9. An experiment compares two brands of automobile tires. Each of a number of cars is equipped with one tire of each brand on a rear wheel (the order is randomized from car to car) and tread wear is measured periodically. This experiment is called

a. Simple random sample.
b. Stratified random sample.
c. Completely randomized design.
d. Matched pairs design.
e. Double-blind design.
f. None of these above.

10. Which of the following statements about a randomized block design with two treatments is **not** true?

    a. Every subject has a 50/50 chance of being given the first treatment.
    b. Block A is chosen randomly from among the available subjects.
    c. In every block, some subjects are assigned the first treatment and some the second treatment.
    d. Treatments are assigned randomly within each block.
    e. All of the statements above are true.
    f. All of the statements above are false.

11. Mr. Stat's library of literary classics includes issues of certain scholarly periodicals. The total number of issues of 4 such periodicals, together with the percentages of Mr. Stat's entire library that they comprise, are as follows:

Periodical	Spiderman	Superman	X-men	Batman
Number	156	52	26	78
% of Library	18.6%	6.2%	3.1%	3.0%

    Which of the percentages are wrong?

    a. 18.6% (Spiderman)
    b. 6.2% (Superman)
    c. 3.1% (X-men)
    d. 3.0% (Batman)
    e. All of these percentages are wrong.
    f. All of these percentages are correct.

12. For a certain experiment the subjects are 8 rats, of which 4 are female (call them F1, F2, F3, F4) and 4 are male (call them M1, M2, M3, M4). There are to be 4 treatment groups A, B, C, and D. If a randomized block design is used, with the rats blocked by gender, which of the following assignments of treatments to rats is **impossible**?

    a. $A \rightarrow (F1, M1); B \rightarrow (F2, M2); C \rightarrow (F3, M3); D \rightarrow (F4, M4)$
    b. $A \rightarrow (F1, M2); B \rightarrow (F2, M3); C \rightarrow (F3, M4); D \rightarrow (F4, M1)$
    c. $A \rightarrow (F2, M1); B \rightarrow (F3, M2); C \rightarrow (F4, M3); D \rightarrow (F1, M4)$
    d. $A \rightarrow (F1, F2); B \rightarrow (F3, M1); C \rightarrow (F4, M2); D \rightarrow (M3, M4)$
    e. $A \rightarrow (F3, M1); B \rightarrow (F4, M2); C \rightarrow (F1, M3); D \rightarrow (F2, M4)$
    f. All of these are possible.

13. If we increase the size of a probability sample, which statement is true?

    i. The margin of error for a sample statistic decreases.
    ii. The variability of a sample statistic increases.
    iii. The variability of a sample statistic decreases.

    a. Only statement i.
    b. Only statement ii.
    c. Only statement iii.
    d. Both statements i and ii.
    e. Both statements i and iii.
    f. None of the above.

14. A dishonest butcher has a scale on which he weighs the meat his customers buy. In order to increase his profits, he has doctored the scale so that it always reads very close to 10 percent more than the actual weight. The measurements from this scale are

    a. Biased and unreliable
    b. Biased and reliable
    c. Unbiased and unreliable
    d. Unbiased and reliable
    e. All of these above.
    f. None of these above.

15. SAT scores are normally distributed with mean 500 and standard deviation 100. Julie's standard score on the test is 1.5. What is her raw SAT score?

    a. 150
    b. 200
    c. 350
    d. 500
    e. 650
    f. None of these above.

16. Suppose that the least squares regression line for predicting *y* is $y = 4x - 17.5$. Given:

$$\bar{x} = 25\ s_x = 2, \bar{y} = 82.5, and, s_y = 10$$

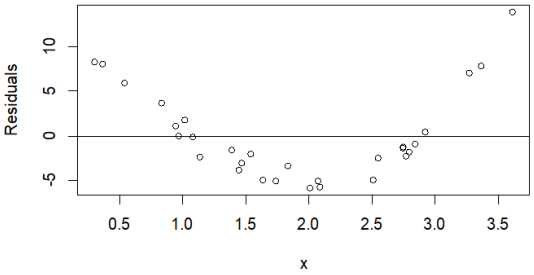
    What is the correct value for the correlation coefficient, *r*?

    a. -0.8
    b. -0.25
    c. 4
    d. 0.8
    e. 0.25
    f. None of these above.

17. The following is an output and residual plot for a least-squares regression equation to predict y. What statement best describes this regression?

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -11.854      2.116  -5.603 5.35e-06 ***
x             12.415      1.010  12.296 8.36e-13 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.082 on 28 degrees of freedom
Multiple R-squared:  0.8438,    Adjusted R-squared:  0.8382
F-statistic: 151.2 on 1 and 28 DF,  p-value: 8.359e-13
```



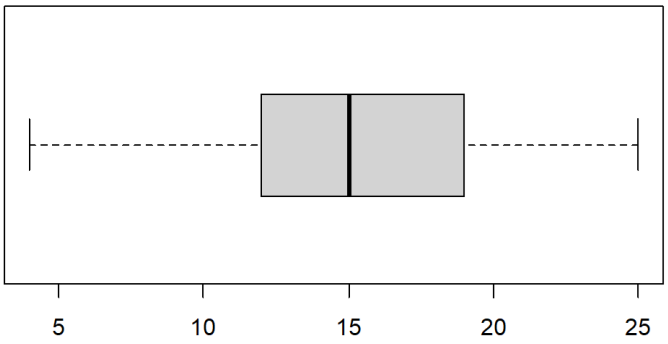
a. The linear equation is the best fit for predicting y.
b. A non-linear equation is the best fit for predicting y.
c. The correlation coefficient between x and y is zero (0).
d. The variable *x* is not significant in predicting *y*.
e. About 5% of the variation in *y* can be accounted for by this equation.
f. None of the above are true.

18. Given the following statistics determine a least square estimate to predict y based on x.

$$\bar{x} = 20\ s_x = 2, \bar{y} = 50, s_y = 5\ and\ r = 0.8$$

a. $y = 50 + 0.8x$
b. $y = 50 + 20x$
c. $y = 50 - 0.8x$
d. $y = 10 - 2x$
e. $y = 10 + 2x$
f. None of these above.

19. Based on the boxplot below, approximately what percent of data points are between 12 and 15?



a. 12.5%
b. 25%
c. 50%
d. 75%
e. 100%
f. None of these are true.

20. For categorical variables (like a person's occupation or sex) it only makes sense to speak of _____. It makes no sense to speak of the _____ or the _____ sex. The sequence of words to correctly complete this statement is

    a. Mean, standardize, median
    b. Median, mean, standard deviation
    c. Standard deviation, mean, median
    d. Counts or percents, mean, median
    e. Mean, percents, standard deviation
    f. None of these above.

21. An urn contains 7 red balls and 13 blue balls. Suppose I pick with replacement 5 balls. What is the probability (to nearest .001) that at least one of the 5 balls is red?

    a. 0.312
    b. 0.2
    c. 0.884
    d. 0.917
    e. 0.926
    f. None of these above.

22. Suppose that 60% of people have heart disease. A primary care physician sees 30 patients a day. What is the expected number of patients per day that have heart disease?

    a. 18
    b. 30
    c. 15
    d. 13
    e. 8
    f. None of these above.

23. Suppose that 5% of American households have 1 gig for the speed of the in-home internet service. A survey was done in a neighborhood and randomly selected 20 households. What is the probability (to nearest .001) that at least 4 of the homes have 1 gig of internet speed?

    a. 0.997
    b. 0.200
    c. 0.016
    d. 0.013
    e. 0.003
    f. None of these above.

24. Suppose that 10% of the US population has diabetes. A doctor sees several patients in a day. What is the probability (to nearest .001) that the 10^{th} patient is the first patient of that day that has diabetes?

    a. 0.132
    b. 0.100
    c. 0.264
    d. 0.039
    e. 0.961
    f. None of these above.

25. A man's weight is normally distributed with mean of 165 pounds with a standard deviation of 15 pounds. A women's weight is normally distributed with a mean of 136 pounds and a standard deviation of 25 pounds. A man and women are randomly selected. What is the probability (to nearest .001) that the man is at least 25 pounds heavier than the woman?

    a. 0.445
    b. 0.555
    c. 0.952
    d. 0.159
    e. 0.579
    f. None of these above.

26. A regular hamburger has an average of 569.8 calories. At the most 80% of the hamburgers have 574 calories. What is the variance (to nearest .1) of the number of calories in a regular hamburger, assuming that the number of calories is normally distributed?

    a. 4.2
    b. 17.6
    c. 5.9
    d. 24.9
    e. 2.1
    f. None of these above.

27. Suppose that the monthly mobile cost of a population of households are normally distributed with a mean of $150 and standard deviation of $20. A random sample of 25 households is taken from this population. What is the probability (to nearest .001) that the mean monthly mobile cost of this sample is at least $160?

    a. 0.309
    b. 0.345
    c. 0.655
    d. 0.006
    e. 0.994
    f. None of these above.

28. A manufacturer of TVs found out that 2% of their TVs are defective. A shipment of 500 TVs was sent to a retail store. What is the expected proportion of TVs that are defective from this shipment of 500 and what is the standard error (to nearest .001) for that proportion?
    a. $E(\hat{p}) = 0.02,\ SE(\hat{p}) = 0.006$
    b. $E(\hat{p}) = 10,\ SE(\hat{p}) = 3.130$
    c. $E(\hat{p}) = 0.2,\ SE(\hat{p}) = 0.060$
    d. $E(\hat{p}) = 0.02,\ SE(\hat{p}) = 0.001$
    e. $E(\hat{p}) = 5,\ SE(\hat{p}) = 9.800$
    f. None of these above.

29. A box contains 5 chips marked 1, 2, 3, 4, and 5. One chip is drawn at random, the number on it is noted, and the chip is replaced. The process is repeated with another chip. Let X_1, X_2 be the outcomes of the two draws. A game is played and the player receives the amount of the mean of the numbers on the two chips drawn. What is the expected value and the variance of the amount the payer receives?
    a. Expected value: \$3; Variance: \$2.50
    b. Expected value: \$1.50; Variance: \$1.25
    c. Expected value: \$2; Variance: \$1
    d. Expected value: \$3; Variance: \$1
    e. Expected value: \$3; Variance: \$0.71
    f. None of these above.

30. Consider the following function.

$$f(x) = P(X = x) = \begin{cases} c\left(\frac{1}{2}\right)^x, & for\ x = 1, 2, 3, 4 \\ 0, & otherwise \end{cases}$$

    Find the constant *c* so that *f(x)* is a probability mass function.
    a. 1
    b. ½
    c. 1/16
    d. 16/15
    e. ¾
    f. None of these above.

31. A truck can carry a maximum load of 4000 lb. A manufacturer wants to ships an order of 50 boxes. The weights of the boxes are normally distributed with mean $\mu = 78$ lb. and standard deviation $\sigma = 12$ lb. What is the probability (to nearest .001) that all 50 boxes can be sent in one shipment? Assume that the weights of each box are independent.
    a. 0.119
    b. 0.434
    c. 0.566
    d. 0.881
    e. 0.999
    f. None of these above.

32. The weight of a randomly selected bag of corn chips coming off an assembly line is a random variable with mean $\mu = 24$ oz. and standard deviation $\sigma = 0.16$ oz. Suppose we pick four bags at random and determine the combined weight of these bags, *T*. Assume that weight of each of the bags are independent. What is the mean and the standard deviation of the combined weight of the bags?

a. $\mu_T = 24, \text{and } \sigma_T = 0.16$
b. $\mu_T = 96, \text{and } \sigma_T = 0.64$
c. $\mu_T = 96, \text{and } \sigma_T = 0.32$
d. $\mu_T = 6, \text{and } \sigma_T = 0.04$
e. $\mu_T = 24, \text{and } \sigma_T = 0.4$
f. None of these above.

33. One of your teacher's claims that boys do better in math classes than girls. Together you run two independent simple random samples and calculate the given summary statistics of the boys and the girls for comparable math classes. 15 boys had a mean percentage of 82.3 with standard deviation of 5.6 while 12 girls had a mean percentage of 81.2 with standard deviation of 6.7. Which of the following would be the most appropriate test for establishing whether boys do better in math classes than girls?

a. two-sample z-test for means
b. two-sample t-test for means
c. chi-square test of independence
d. two-sample z-test for proportions
e. One-sample t-test for means
f. none of these tests would be appropriate

34. A sample of 100 delivery drivers indicated the mean amount of time they spend driving during the day is 6.5 hours. Three interns for this delivery service independently calculated different two-sided confidence intervals of the true mean amount of time for all drivers at this company.

The confidence intervals of the interns are:
A) (6.38, 6.62) B) (6.19, 8.81) C) (5.19, 7.81)

a. All are calculated correctly with different levels of confidence.
b. A and C have reasonable intervals, but B does not.
c. A and B have reasonable intervals, but C does not.
d. B and C have reasonable intervals, but A does not.
e. B has a reasonable interval, but A and C does not.
f. None of these intervals are reasonable.

35. It has been estimated that 70% of people have high cholesterol. Determine an approximate 95% range for the percentage (to nearest .01) of people with high cholesterol in a sample of 150 people.

    a. (62.67%, 77.33%)
    b. (66.26%, 73.74%)
    c. (57.67%, 72.33%)
    d. (50.67%, 80.33%)
    e. (91.5%, 98.5%)
    f. None of these above.

36. A study examines a drug's effectiveness on lowering high blood pressure. The study indicates a null hypothesis that the drug does not work on lower high blood pressure. The data shows that the drug does lower high blood pressure, but in reality, the drug does not work. This indicates

    a. Type 1 error
    b. Type 2 error
    c. Type 3 error
    d. Bad study
    e. The drug works
    f. None of these above.

37. A food manufacturer wants to estimate the mean weight of bags of chips. It is estimated that the standard deviation of weight is 0.14 oz. How many bags should be sampled in order to find a 97% confidence interval having a margin of error of no more than 0.05 ounces.

    a. 37
    b. 36
    c. 30
    d. 6
    e. 7
    f. None of these above.

38. The SAT scores from a random sample of freshman class of 50 at a liberal arts university had a mean score of 1250 with a sample standard deviation of 100. Assume that this class represents a random sample of all future students. Determine a 95% confidence interval (to nearest .01) for the mean SAT score of all future students.

    a. (1222.28, 1277.72)
    b. (1221.58, 1278.42)
    c. (1235.86, 1264.14)
    d. (1150, 1350)
    e. (1213.57, 1286.43)
    f. None of these above.

39. A random sample of 100 voter were surveyed to determine if they will vote yes for proposal 1 on the ballot. From this sample, 85% said they would vote yes for the proposal with a margin of error of ±3%, with 95% confidence. Which statement is true about this survey?

    a. 85% of all voters will vote yes for proposal 1.
    b. 95% of all voters will vote yes for proposal 1.
    c. We are 95% confident that between 82% and 88% of the 100 voters in the sample will vote yes for proposal 1.
    d. We are 95% confident that between 82% and 88% of all voters will vote yes for proposal 1.
    e. We are 3% confident that 85% of the voters will vote yes for proposal 1.
    f. None of these statements are true.

40. A study was determined if exercise affected the performance on an exam. A random sample of 35 students took an exam without exercising (group A) and another 30 students took the same exam after exercising (group B). A significance test for the difference of the mean scores, $H_0;\ \mu_A = \mu_B$ against the alternative hypothesis $H_0;\ \mu_A < \mu_B$ produced a p-value of 0.68. What is the conclusion of this significance test?

    a. Reject the null hypothesis, group B had statistically significant higher mean exam scores than group A.
    b. Fail to reject the null hypothesis, there is no evidence that group B had statistically significant higher mean exam scores than group A.
    c. Accept the null hypothesis, the mean scores for group A and group B are the same.
    d. Reject the null hypothesis, group B had statistically significant lower mean exam scores than group A.
    e. Fail to reject the null hypothesis, group B had statistically significant higher mean exam scores than group A.
    f. None of these are correct.

41. Two brands of water filters are to be compared in the mean reduction in impurities measured in parts per million (ppm). Twenty-five water samples were tested with each filter and reduction in the impurity level was measured, resulting in the following data:

$$\text{Filter 1: } n_1 = 25, \bar{x}_1 = 8.064, s_1^2 = 5.21$$
$$\text{Filter 2: } n_2 = 25, \bar{x}_2 = 6.76, s_2^2 = 3.03$$

    We want to determine a statistical difference of the mean between the two filters, $H_0: \mu_1 - \mu_2 = 0$ against $H_0: \mu_1 - \mu_2 \neq 0$, assuming that the variances are equal. Give the p-value (to nearest .001) and the decision of the test, use α = 0.05.

    a. *P*-value = 0.028, reject the null hypothesis.
    b. *P*-value = 0.028, fail to reject the null hypothesis.
    c. *P*-value = 0.0125, reject the null hypothesis.
    d. *P*-value = 0.0125, fail to reject the null hypothesis.
    e. *P*-value = 0.806, reject the null hypothesis.
    f. None of these above.

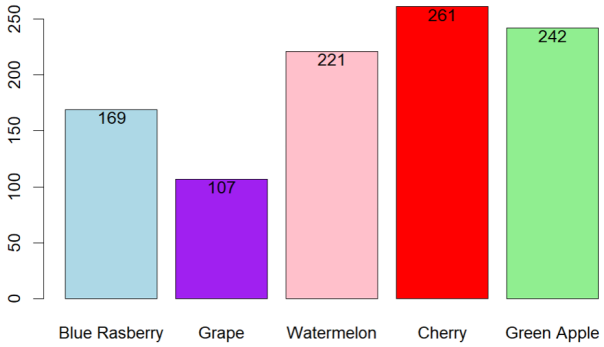
42. The following data set is from a student that gives the incidence of cold among 279 French skiers who were randomized to the Vitamin C and Placebo groups.

Group	Cold		Total
	Yes	No	
Vitamin C	17	122	139
Placebo	31	109	140

Is there a significant difference in the incident rates for cold between the Vitamin C and Placebo groups at α = 0.05? Give the test statistic and the decision.

a. $\chi^2 = 122.65$, reject the null hypothesis
b. $\chi^2 = 4.1407$, reject the null hypothesis
c. $Z = -2.2145$, reject the null hypothesis
d. $\chi^2 = 4.1407$, accept the null hypothesis
e. $Z = -2.2145$, accept the null hypothesis
f. None of these above.

43. A bag of Jolly Rancher hard candy has five different types of candy, blue raspberry, grape, watermelon, cherry, and green apple. A random sample of ten bags, each has 100 pieces of candy. The chart shows the number obtained for each type of candy.



The null hypothesis is the proportions of the five flavors in this bag are the same. Give the test statistic of this hypothesis test.

a. χ^2 = 77.68
b. χ^2 = 18.621
c. $Z = 12.73$
d. Z = 5.72
e. T = -3.69
f. None of these above.