

1 (7 Points)

Which of the following sampling methods is the least appropriate method to estimate the percentage of college students who have used a student tutor during the previous year?

- [1] stratified random sampling to randomly select students from various college classes
- [2] simple random sampling
- [3] the use of self-selected volunteers
- [4] taking a census
- [0] no answer or skip this item

[Submit Answer](#)

2 (7 Points)

A professor is interested in the proportion of students who consider the recreational facilities to be adequate within the university. The survey design attempted to capture variation between majors by stratifying the sample into four strata: first year, second year, third year, and fourth year (or more) students.

The population to which the results of this study can be applied is most likely:

- [1] all students
- [2] all students who used the recreational facilities within the university
- [3] all students who used any recreational facility
- [4] only the students who participated in the study
- [0] no answer or skip this item

[Submit Answer](#)

3 (7 Points)

A study is done to compare the cholesterol levels in people who follow a particular diet plan, to the cholesterol levels in people who do not follow this plan. The researcher is able to study 200 individuals of each type.

Other factors that may affect cholesterol levels are an individual's weight, exercise habits and stress levels. The weight characteristics of the two groups of people are similar, their exercise habits are different and their stress levels are unknown.

In this study, the explanatory variable is:

- [1] exercise
- [2] cholesterol level
- [3] weight
- [4] diet plan
- [0] no answer or skip this item

Submit Answer

4 (7 Points)

A population contains 1000 students and a random sample of 125 students is obtained. Data collected includes sex, height, and weight.

Which of the following statements is true?

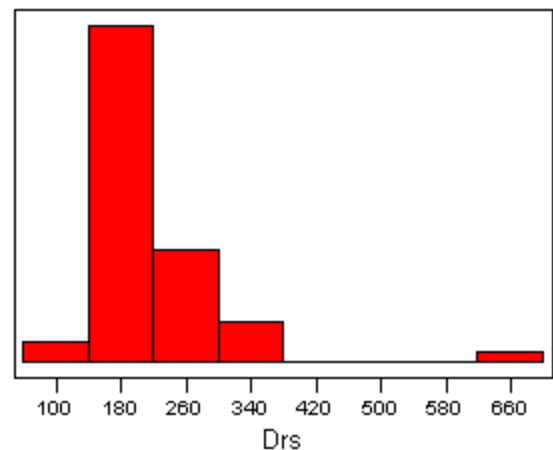
- [1] The population of measurement and the sampling frame would change if a new random sample were drawn.
- [2] The population of units and the population of measurements would change if a new random sample were drawn.
- [3] The sample statistics would change if a new random sample were drawn.
- [4] The population parameters would change if a new random sample were drawn.
- [0] no answer or skip this item

Submit Answer

5 (7 Points)

The histogram below is based on the number of doctors per 100,000 population in the 50 states plus DC.

Based on this histogram, which is larger, the average or the median?



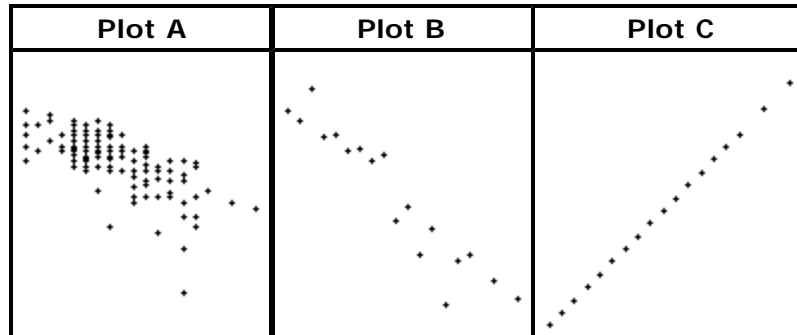
- [1] Average is larger
- [2] Median is larger
- [3] Impossible to tell from the information given
- [4] They are equal.
- [0] no answer or skip this item

Submit Answer

6 (7 Points)

Next are three plots based on data from a car race. There is one plot each (in some order) with

- X = Speed in miles per hour and Y = Time in hours;
- X = Speed in miles per hour and Y = Average Time in hours for everyone of traveling at speed X ;
- X = Speed in miles per hour and Y = Speed in kilometers per hour for the 126 participants.



Which plot has X = Speed in miles per hour and Y = Speed in kilometers per hour for the 126 participants?

- [1] Plot A
- [2] Plot B
- [3] Plot C
- [4] None of these plots.
- [0] no answer or skip this item

Submit Answer

7 (7 Points)

The color of a seed might be green or yellow. Suppose green (G) is dominant and yellow (y) is recessive. Suppose both parents have one gene of each color. Here is a Punnett square to illustrate this situation:

Seed Color		Seed Color	
		Father	
		G	y
Mother	G	(G, G)	(G, y)
	y	(G, y)	(y, y)

Suppose we cross these seeds. What are the probabilities for seed color in this cross?

- [1] Green = 1/4, Yellow = 1/4
- [2] Green = 4/4, Yellow = 1/4
- [3] Green = 2/4, Yellow = 2/4
- [4] Green = 3/4, Yellow = 1/4
- [0] no answer or skip this item

8 (7 Points)

Lie detector tests have often been criticized as unreliable. It is suspected that false positives, results indicating the subjects were lying when they were not, occur far too frequently. Suppose 300 people submitted to a lie detector test, with some being told to always tell the truth and others to lie to certain questions. The individuals administering the lie detector test were not told who was truthful and who was not truthful. Positive results indicate the subject is lying.

Lie Detector Testing	Lied	Truthful	Total
Positive Test	59	18	77
Negative Test	7	216	223
Total	66	234	300

Among people with a negative test result, what is the probability that the person is lying?

- [1] 0.89
 [2] 0.0314
 [3] 0.0769
 [4] 0.197
 [0] no answer or skip this item

9 (7 Points)

Suppose the amount of detergent delivered through a high pressure sprayer is a uniform distribution. If X = gallons of detergent per minute and X has values between 5.5 and 6.5 gallons which are equally probable. The probability density function is $f(x) = 1$ for $5.5 < x < 6.5$, $f(x) = 0$ otherwise.

What is the probability that less than 5.8 gallons of detergent per minute are dispensed?

- [1] 0.1
 [2] 0.2
 [3] 0.3
 [4] 0.7
 [0] no answer or skip this item

10 (7 Points)

The following table shows the probability distribution for the number of heads that can result when a fair coin is flipped three times.

Heads	0	1	2	3
Prob	0.125	0.375	0.375	0.125

The expected value of this distribution is 1.5 heads.

Calculate the standard deviation of the number of heads in three flips of the coin.

- [1] 1.00
- [2] 0.866
- [3] 0.75
- [4] 0.5
- [0] no answer or skip this item

Submit Answer

11 (7 Points)

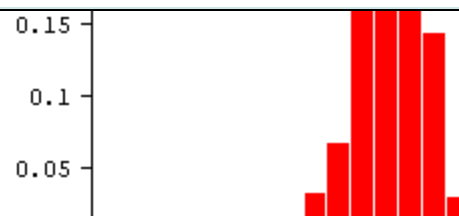
Suppose that the probability of a prior conviction for driving under the influence of alcohol is .01 and that the police stop 5 motorists who are driving under the influence of alcohol.

What is the probability that at least one of these 5 motorists has a prior conviction?

- [1] about 0.00001
- [2] about 0.049
- [3] about 0.0585
- [4] about 0.999
- [0] no answer or skip this item

Submit Answer

regarding the sampling distribution for the sample mean?



- [1] The shape of the sampling distribution is approximately a normal curve although not perfectly so.
- [2] The center of the sampling distribution is about 32 inches.
- [3] The sample means are spread from about 26 inches to about 32 inches.
- [4] All of the above are false.
- [0] no answer or skip this item

Submit Answer

13 (7 Points)

A bowl contains a huge number of coins. One third of the coins are pennies, one third are nickels and one third are dimes. The population mean of these coins is \$0.0533 (5.33 c), and the standard deviation is \$0.028 (2.8 c). Imagine choosing a sample of 100 coins from the bowl, where each coin has an equal chance of being picked.

The standard deviation of the sum of the coins is closest to:

- [1] \$0.028
- [2] \$0.0533
- [3] \$0.28
- [4] \$2.80
- [5] \$5.33
- [0] no answer or skip this item

Submit Answer

14 (7 Points)

Assume that 38% of all faculty employed at universities are female. The females have value '1', and males have value '0.' Thus the population mean is 0.38. The population standard deviation is 0.485. Assume we are choosing a sample of size 100 employees.

To answer this question, you should use a normal curve calculator or a normal table of your choice.

What is the probability that the proportion is in the interval from 31% to 45%?

- [1] about 0.9256
- [2] about 0.8414
- [3] about 0.0796
- [4] almost 1
- [0] no answer or skip this item

Submit Answer

15 (7 Points)

In wanting to estimate the average travel time between customers, a manager measured the travel time for a random sample of 84 customers. Using MINITAB, she calculated a 90% confidence interval for the average travel time:

The assumed sigma = 3

Variable	N	Mean	StDev	SE Mean	90.0 % CI
TravelTime	84	22.27	3.02	.327	(21.73, 22.81)

Which statement correctly summarizes the manager's findings?

- [1] The manager can be 90% confident that 90% of the travel times between customers is between 21.73 and 22.81 minutes.
- [2] The manager can be 90% confident that the true average travel time between customers is between 21.73 and 22.27 minutes.
- [3] The manager can be 90% confident that the true average travel time between customers is between 21.73 and 22.81 minutes.
- [4] The manager can be 95% confident that the true average travel time between customers is between 21.73 and 22.81 minutes.
- [0] no answer or skip this item

Submit Answer

16 (7 Points)

A coin is tossed 10,000 times to see if it is fair (i.e., 'heads' and 'tails' are equally likely). In particular, the investigator thought that a head came up more than it should. Let p be the probability of a head. If the coin is fair, then $p = 1/2$.

What is the alternative hypothesis?

- [1] $H_a: p = 1/2$
- [2] $H_a: p < 1/2$
- [3] $H_a: p > 1/2$
- [4] $H_a: p \neq 1/2$
- [0] no answer or skip this item

Submit Answer

17 (7 Points)

Researchers are interested in testing whether there are an excessive number of rat hairs in jars of peanut butter produced at a particular factory. They examine a random sample of 144 jars, and find an average of 6.3 rat hairs in each jar. The sample standard deviation is 2. They would like to do a one-sided z-test of whether the population average is equal to five (the maximum permitted by law) versus the alternative that it is greater than five.

What is the standard error of the sample mean?

- [1] about 1.3
- [2] about 0.1667
- [3] about 0.014
- [4] about 0.01
- [0] no answer or skip this item

Submit Answer

18 (7 Points)

The recommended dietary allowance of folic acid for adult females is 400 mcg. Folic acid is found naturally in leafy dark green vegetables, legumes (dried beans and peas), citrus fruits and juices, and most berries. A vitamin supplement is supposed to contain $\mu = 400$ mcg of folic acid.

A random sample of 100 such vitamin tablets was obtained and the amount of folic acid contained in each tablet was determined. The sample mean was 399.92 mcg, and the sample standard deviation was $s = .5$ mcg. Any deviation from the null is of interest.

What are the null and the alternative hypotheses?

- [1] H_0 : Population mean = 400; H_a : Population mean = 399.92
- [2] H_0 : Population mean \neq 400; H_a : Population mean = 400
- [3] H_0 : Population mean = 400; H_a : Population mean > 400
- [4] H_0 : Population mean = 400; H_a : Population mean < 400
- [5] H_0 : Population mean = 400; H_a : Population mean \neq 400
- [0] no answer or skip this item

Submit Answer

19 (7 Points)

Which statement best describes a t-test?

- [1] A t-test is a tool used to answer questions about the population mean when the population standard deviation is known.
- [2] A t-test is a significance testing procedure used to evaluate hypothesis statements about means.
- [3] A t-test is a useful model for describing the behavior of sample means when the population standard deviation is known.
- [4] A t-test is the theoretical model that describes the possible values of t and is also called the t curve.
- [0] no answer or skip this item

Submit Answer

20 (7 Points)

What is the average age at the time a person is married for the first time? This question was posed to a random selection of 36 people. Use this Minitab output to answer the question.

Test of mu = 23 vs mu < 23						
N	Mean	StDev	SE Mean	T	P	
36	20.84	3.85	0.642	-3.37	0.0003	

How can we interpret the p-value of this test?

- [1] If the population mean is really 20.84, the probability is 0.0003 that the mean of the 36 ages could be different from 23.
- [2] If the population mean is really 20.94, the probability is 0.0003 that the mean of the 36 ages is 23.
- [3] If the population mean is really 23, the probability is 0.0003 that the mean of the 36 ages could be less than 23.
- [4] If the population mean is really 23, the probability is 0.0003 that the mean of the 36 ages could be different from 23.
- [0] no answer or skip this item

Submit Answer

21 (7 Points)

The table below shows the results of four random samples conducted by four different instructors who want to determine whether students study enough for the final:

Sample	Study Enough	Don't Study Enough
No. 1	4	7
No. 2	72	94
No. 3	86	73
No. 4	350	250

For which of the above samples would it be inappropriate to use the z-interval to estimate the true proportion of students who study enough?

- [1] Sample No. 1
- [2] Sample No. 2
- [3] Sample No. 3
- [4] Sample No. 4
- [0] no answer or skip this item

Submit Answer

22 (7 Points)

A significance test is done to compare two population means. The null hypothesis is that the means are the same. The alternative is that they are different. The p-value of the test is found to be 0.23. Each sample mean is based on $n = 7$ observations.

What is the most appropriate conclusion?

- [1] The population means are different.
- [2] The sample means are the same.
- [3] The population means are the same.
- [4] There's not enough evidence to say the means are different.
- [0] no answer or skip this item

Submit Answer

23 (7 Points)

The variables gender and job stability were found to have a significant relationship.

What might be possible confounding variables that could be responsible for the observed relationship?

- [1] Two possibilities are home ownership and parental level of education.
- [2] Two possibilities are occupation and seniority.
- [3] Two possibilities are political party and number of sick days used.
- [4] Two possibilities are place of residence and driving habits.
- [0] no answer or skip this item

Submit Answer

24 (7 Points)

The height and weight of a group of basketball players are related by the linear model

$$Y = -208 + 5 \times X + e$$

where X is height (in inches) and Y is weight (in pounds).

What is the expected weight difference between a person 70 inches tall and a person 72 inches tall?

- [1] 152 pounds
- [2] 147 pounds
- [3] 13 pounds
- [4] 10 pounds
- [0] no answer or skip this item

Submit Answer

25 (7 Points)

In data on 60 locations, the least squares line for $X = \text{Rain}$ versus $Y = \text{Mortality}$ is

$$\text{Mortality} = 850 + 2.37 \times (\text{Rain}).$$

What is the least squares estimate of the intercept a ?

- [1] 2.37
- [2] 850
- [3] 2014.5
- [4] 358.6
- [0] no answer or skip this item

Submit Answer

26 (11 Points)

In an observational study, the extent of heart disease in a sample of smokers is compared to the extent of heart disease in a sample of nonsmokers. The researcher finds that a higher percentage of smokers have heart disease.

What else do we have to know to be able to conclude that there is a causal link between smoking and the risk of heart disease?

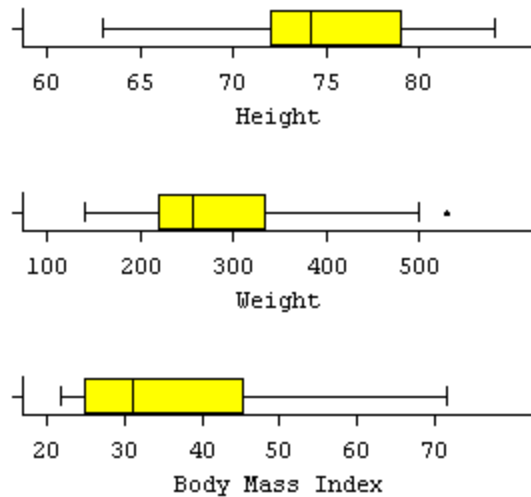
Submit Answer

27 (11 Points)

Below are **Heights** in inches and **Weights** in pounds of 132 professional male athletes, in two sports. Also included are their **body mass index** numbers, which are defined by

$$\text{BMI} = \text{Body Mass Index} = (\text{Weight in Pounds}) * 703 / ((\text{Height in inches})^2)$$

BMI is supposed to measure how overweight or underweight a person is. A value in the range 20-25 is fine; more is deemed overweight; under 20 is deemed underweight. It is fairly easy to be overweight under this measure. Here are boxplots for the three variables:



- (a) What is the median height, approximately?
- (b) How tall is the tallest person?
- (c) About what percentage of these athletes is taller than 72 inches?

Submit Answer

28 (11 Points)

The next interactivity is based on 11 observations with 6 variables. Use it to construct some plots. Choose among the following choices for the best description of the plot:

- 1. Linear
- 2. Linear, but with an outlier
- 3. Curved
- 4. None of the above

Use the **Interactive Tool** below.

Best description for $X = X1$ and $Y = Y4$?

CLICK HERE FOR
**Interactive
Tool!**

Submit Answer

29 (11 Points)

In 1989, scientists identified the genetic cause of cystic fibrosis (CF): an error on chromosome 7. This has helped identify carriers of the gene within high-risk families. The genetic test for CF is 85% accurate in identifying carriers.

Suppose a husband and wife are both carriers. What is the chance that they will both test positive and be identified as carriers?

Submit Answer

30 (11 Points)

The heights of women have a distribution that is approximated by a normal curve with a mean of 65 inches and a standard deviation equal to 2.7 inches.

To answer this question, you should use a normal curve calculator or a normal table of your choice.

About what proportion of women are between 62 and 70 inches tall?

Submit Answer

31 (11 Points)

The heights of women have a distribution that is approximated by a normal curve with a mean of 65 inches and a standard deviation equal to 2.7 inches.

To answer this question, you should use a normal curve calculator or a normal table of your choice.

For women, approximately what is the 60th percentile of heights?

Submit Answer

32 (11 Points)

CNN recently conducted a survey on their web site, asking visitors: "Is there life out there?" Of the 613 people (or aliens?!) who voted, 80% said "yes". The margin of error for the poll was plus or minus four percentage points.

Do you think the survey results are believable and could be used to estimate the true percentage of Americans who believe in life "out there"? Explain.

Submit Answer

33 (11 Points)

In US Census data on the 50 states, two of the variables are X = Employment percentage and Y = Poverty percentage. It makes sense that the variables would have negative correlation, because one would expect that the less employment, the more poverty.

State the null and alternative hypotheses in terms of the population correlation coefficient for testing whether there is a negative correlation.

Submit Answer

34 (11 Points)

A null hypothesis is that the mean systolic blood pressure of women is 115. The alternative hypothesis is that the mean is higher than 115. Based on a sample of 100 women, a t-test is performed and the p-value is 0.35.

What is the appropriate conclusion?

Submit Answer

35 (11 Points)

Suppose you were given a 95% confidence interval for the difference in two population means.

What could you conclude about the two population means if the confidence interval contained only negative numbers?

Submit Answer

36 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Body-temps-62_adults.dat**". This is one out of 8 questions that will work with this data set.

Calculate (and report) the mean, median, and standard deviation for the variable temperature.

Submit Answer

37 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Body-temps-62_adults.dat**". This is one out of 8 questions that will work with this data set.

Note that the temperature data is reported in degrees Fahrenheit. Calculate (and report) the mean, median, and standard deviation for the variable temperature in degrees Celsius. Recall that $\text{degrees C} = (\text{degrees Fahrenheit} - 32) / 1.8$.

Submit Answer

38 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Body-temps-62_adults.dat**". This is one out of 8 questions that will work with this data set.

Construct a histogram for the variable temperature, using a bin width of 0.5. Describe the overall shape of the histogram (i.e., is it fairly symmetric, skewed towards the higher values, skewed towards the lower values). Are there any outliers? What is the modal bar?

Submit Answer

39 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Body-temps-62_adults.dat**". This is one out of 8 questions that will work with this data set.

Construct a boxplot of the variable temperature, using fences to identify outliers. Are there any outliers - and does the boxplot suggest that the data is fairly symmetric or whether there is skewness?

Submit Answer

40 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Body-temps-62_adults.dat**". This is one out of 8 questions that will work with this data set.

Conduct a test for the mean temperature (assuming the data originates from a random sample from all patients in a hospital - the population of interest), where H_0 : population mean = 99 versus H_a : population mean \neq 99. Can we use a z-test here (if yes, do so) or do we have to use t-test (in this case, indicate the degrees of freedom). Explain. Report the p-value for the test you choose and provide a verbal conclusion.

Submit Answer

41 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Body-temps-62_adults.dat**". This is one out of 8 questions that will work with this data set.

Construct (and report) 90% confidence intervals for the mean temperature of all patients in the hospital (assuming the data originates from a random sample from all patients in a hospital - the population of interest). Do this based on a z-interval and t-interval. How do these two intervals differ?

Submit Answer

42 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Body-temps-**

62_adults.dat". This is one out of 8 questions that will work with this data set.

Assume you manually have to construct a 90% confidence interval for the mean temperature of all patients in the hospital (assuming the data originates from a random sample from all patients in a hospital - the population of interest), based on a z-interval. The general formula is "sample mean +/- multiplier * standard error". Indicate the value for the multiplier and show how you would calculate the standard error.

Submit Answer

43 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Body-temps-62_adults.dat**". This is one out of 8 questions that will work with this data set.

Conduct a test for the variance of the temperatures (assuming the data originates from a random sample from all patients in a hospital - the population of interest), where H_0 : population variance = 0.4 versus H_a : population variance > 0.4. Explain what type of a test we are using and how many degrees of freedom we have. Report the p-value and provide a verbal conclusion.

Submit Answer

44 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Physiologic_data-130_adults.dat**". This is one out of 7 questions that will work with this data set.

Construct a scatterplot of Temperature (x) versus HeartRate (y). Describe the overall appearance of this plot. Is there an apparent trend? Are there any obvious outliers? Do you notice two obvious clusters for the two genders in this plot?

Submit Answer

45 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Physiologic_data-130_adults.dat**". This is one out of 7 questions that will work with this data set.

Calculate (and indicate) the correlation between Temperature and HeartRate. What does this value indicate?

Submit Answer

46 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Physiologic_data-130_adults.dat**". This is one out of 7 questions that will work with this data set.

Construct (and report) the regression line, predicting HeartRate from Temperature. Do we have a significant slope? And can we use the regression equation to predict HeartRate from Temperature?

Submit Answer

47 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Physiologic_data-130_adults.dat**". This is one out of 7 questions that will work with this data set.

Construct histograms of HeartRate for the two genders separately (group by Gender and choose a separate graph for each gender - start bins at 50 with a bin width of 5). Explain how these two histograms look. Do they look (somewhat) similar and what are the main differences?

Submit Answer

48 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Physiologic_data-130_adults.dat**". This is one out of 7 questions that will work with this data set.

Construct (and report) regression lines for both Genders separately, predicting HeartRate from Temperature.

Submit Answer

49 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Physiologic_data-130_adults.dat**". This is one out of 7 questions that will work with this data set.

Based on your regression calculations above, for which Gender(s) do we have a significant slope, i.e., for which Gender can we use Temperature to predict HeartRate?

Submit Answer

50 (11 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "**Physiologic_data-130_adults.dat**". This is one out of 7 questions that will work with this data set.

Based on your regression calculations above, what would be the predicted HeartRates for someone from Gender 2 with Temperatures of 97, 99, and 104 (degrees Fahrenheit)? Which of these 3 predicted values is least reliable?

Submit Answer

When you are done answering all questions above to your satisfaction, press the button below to complete your test.

Mark Test Completed