

✓ Congratulations! You passed!

TO PASS 70% or higher

Keep Learning

GRADE
100%

Practice Statistics Assessment II

TOTAL POINTS 31

1. Which of the following definitions is the correct definition of Nonparticipation Bias?

1 / 1 point

- When participants respond differently from how they actually feel.
- When a representative sample is chosen but a subset cannot be contacted or does not respond.
- The method used to select participants does not produce a representative sample of the population.

✓ Correct

This considered nonparticipation bias.

2. What is the 95% (non-conservative) confidence interval for the population proportion of all teens who get 6 to 8 hours of sleep per night on average?

1 / 1 point

Hours of Sleep	Number of Responses
Less than 4	119
4 to 6	404
6 to 8	460
8 to 10	298
More than 10	87
Total	1368

- 26.83% to 32.23%
- 28.30% to 30.76%
- 31.07% to 36.19%
- 30.93% to 36.33%

✓ Correct

Note that this is the non-conservative confidence interval. The formula sheet has equations to help calculate the value.

Topics to review: confidence intervals, population proportion, conservative confidence intervals

3. Construct the least squares regression line based in the following output:

1 / 1 point

Call:

```
lm(formula = egg_production ~ age, data = chickendata)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.24521	.981048	0.342	0.831
age	-0.34231	0.03453	2.874	0.0020 **

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 ' '

Correlation Matrix:

	egg_production	age
egg_production	1.0000000	0.6213532
age	0.6213532	1.0000000

Please, structure your response as $y = b_0 + b_1x$ so, for example, if the value of b_0 was 1.301 and the value of b_1 was 0.422, then your answer would be: $y = 1.301 + 0.422x$. Please round to the third decimal place.

y = 3.245 - 0.342x

✓ Correct

topics to review for this question: linear regression, least squares regression line

4. True or False - the following scenario depicts an independent relationship between variables (tree growth and air quality): 20% of trees growing in a particular region are not growing to their expected height. In a particular neighborhood in that region, the Air Quality Index is labeled as "Unhealthy for Sensitive Groups" or worse 30% of the time. 10% of the trees in the region grow in this neighborhood. If you randomly measured the growth of a tree in that neighborhood, then the probability that that tree is not growing to its expected height is 33.33%.

True
 False

Correct

Topic to review: Conditional probability for independent events

5. Say we have a random sample of $n = 15$ people who visit a library's website to browse their catalogue. With $p = 0.1$ of the population proportion who reserve a book while browsing, what is the probability of selecting exactly 14 users who actually make a reservation in the random sample? Please round to the fourth decimal place.

0.0000

Correct

Topic to review: Binomial Probability

Note that in this case, because the number is so small we typically round to zero.

6. For the next three questions, use the following information to determine your answers: A tree farming company is testing how many items customers purchase during their visits. Based on many results, the (partial) probability distribution below was determined for the discrete random variable X = number of pieces of information remembered (during a fixed time period).

What is the missing probability $P(X=6)$? Note that the missing probability should be reported to the second decimal place.

$X = \# \text{ items}$	1	2	3	4	5	6
Probability	0.58	0.18	0.10	0.07	0.05	---

0.02

Correct

Topic to review: PDF (Probability Distribution Function)

7. Complete the table below to provide the cumulative distribution function of X .

Please format your response in the following way:

(val_one, val_two, val_three, val_four, val_five, val_six)

Note that for each value, you should report the number to the second decimal place.

$X = \# \text{ items}$	1	2	3	4	5	6
Probability						

(0.58, 0.76, 0.86, 0.93, 0.98, 1.00)

Correct

Topic to review: CDF (Cumulative Distribution Function)

Also, remember to follow the formatting, as this is not hand graded.

8. Given that the person purchases at least 2 items from the tree farm, what is the probability that they purchase a total of 6 items? Please round to the second decimal place.

0.05

Correct

Topic to review: Conditional Probability

9. For the next three questions, use the following information to determine your answers: A survey was sent out to re-evaluate the proportion of people who play games on pc computers, as the last study on the topic had been gathered four years prior. The current study, with 861 participants, found that 53% of people who responded play on a pc computer.

Calculate the 90% confidence interval. Enter your answer in the following format: (lower_value, upper_value). Please round your values to the fourth decimal point.

(0.5020, 0.5580)

 **Correct**

Topic to review: Confidence Interval, Population Proportion

1 / 1 point

10. This survey was done to test the possibility that fewer people are playing games on pc computers. The previous study found that 81% of people were playing games on pc computers. Which of the following represents the hypotheses that we will be testing, assuming that p represents the most recent findings and that p_0 represents the older findings in the previous study.

- H₀: $p \leq 0.81$ versus H_a: $p > 0.81$
- H₀: $p_0 = 0.81$ versus H_a: $p_0 \neq 0.81$
- H₀: $p \geq 0.81$ versus H_a: $p < 0.81$
- H₀: $p_0 \geq 0.81$ versus H_a: $p_0 < 0.81$

 **Correct**

Consider what is being tested. Do we assume that the data is less than before, greater than before, or different, but not in what way it is different?

11. Calculate the p-value and determine if we should **accept** or **reject** H₀ under alpha = 0.05.

1 / 1 point

 **Correct**

Topics to review: testing a hypothesis, calculating the p-value, interpreting the p-value, Type I and Type II errors.

12. True or False: An article published recently about the results of a study, which found an association between sentiment in social media posts and an online communities overall satisfaction with the group. The author concludes their article stating that "the tone and word choices by members in an online community may affect how satisfied members are with the group, but more research is still needed to determine what factors are the most important in determining community satisfaction." Based on the information provided, did they accurately report the findings throughout the entire article?

- True
- False

 **Correct**

Correlation does not mean causation

13. Which of the following correlation coefficient values would match the following description:

1 / 1 point

little or no negative relationship

- 0.15

 **Correct**

topic to review: correlation coefficient meaning

- 0.87
- 0.67

14. Which of the following correlation coefficient values would match the following description:

1 / 1 point

strong to very strong negative relationship

- 0.87

 **Correct**

topic to review: correlation coefficient meaning

- 0.67
- 0.15
- 0.25

15. Which of the following correlation coefficient values would match the following description:

1 / 1 point

moderate to strong positive relationship

- 0.15

-0.15

0.2553

0.67

 Correct

topic to review: correlation coefficient meaning

-0.86

16. For the next two questions use the following information to determine your answers. A makeup company wants to know if all the shades of their foundation are sold at equal rates. Below is the gathered data

1 / 1 point

Assuming that each shade of foundation is sold at an equal rate, how many foundations would you expect to sell in shade # 4? Please round to the second decimal point.

Shade #1	Shade #2	Shade #3	Shade #4	Shade #5	Shade #6	Shade #7
218	191	239	189	178	168	149
Pearson's Chi-square test						
X-squared = 28.88	df = 6	p_value = ?				

190.29

 Correct

Topic to review: Chi-square tests

17. Based on the information above, determine if you should **accept** or **reject** the null hypothesis that there is no relationship between shade number and number of units sold when alpha = 0.05?

1 / 1 point

reject

 Correct

Topic to review: Chi-square tests, hypothesis testing, Type I and Type II errors, calculating the p-value, interpreting the p-value

18. Based on the following data, determine the median. This should be a whole number.

1 / 1 point

34, 44, 56, 58, 58, 60, 62, 63, 63, 63, 69, 72, 72

62

 Correct

topic to review: calculating a mean, median, and mode

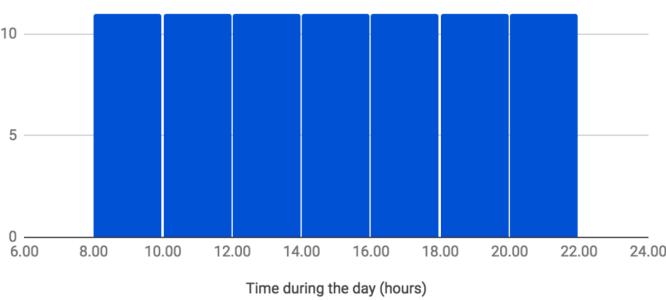
19. Which of the following graphs match the following distribution description?

1 / 1 point

uniform, symmetrical, no apparent outliers

15

Customers in the Store

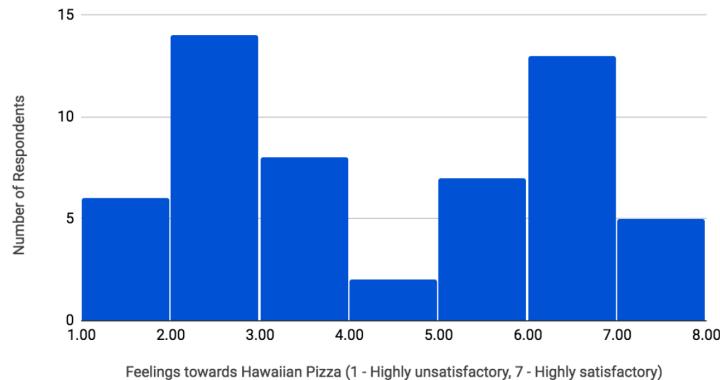


Time during the day (hours)

Image Description: A histogram describes the number of customers in store in different time periods of a day. The horizontal x-axis shows time slots of day ranging from 6 o'clock to 24 o'clock, each bin representing a 2-hour time period. The vertical y-axis shows the number of customers in store ranging from 0 to 15, in increments of 5. Data in the histogram can be summarized accurately as:

- 0-0 hours, 0 customers
- 8-10 hours, 11 customers
- 10-12 hours, 11 customers
- 12-14 hours, 11 customers
- 14-16 hours, 11 customers
- 16-18 hours, 11 customers
- 18-20 hours, 11 customers
- 20-22 hours, 11 customers
- 22-24 hours, 0 customers

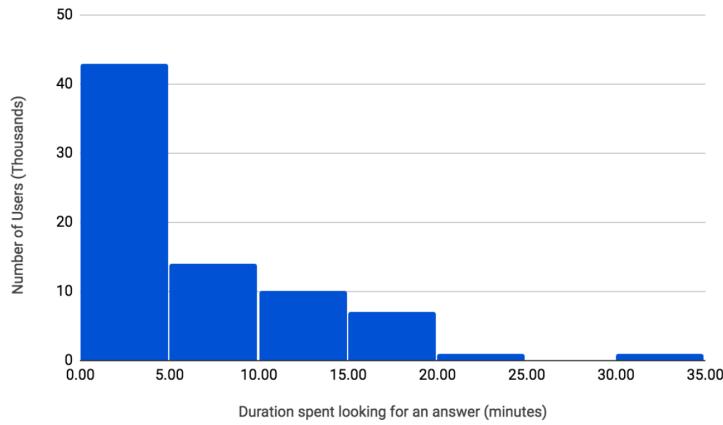
○



A histogram describes the distribution of respondents' satisfaction level towards Hawaiian pizza. The horizontal x-axis shows respondents' feelings, ranging from 1 to 7, 1 means highly unsatisfactory, 7 means highly satisfactory, each bin representing a 1 satisfactory level. The vertical y-axis shows the number of respondents, ranging from 0 to 15, in increments of 5. Data in the histogram can be summarized accurately as:

- 6 respondents voted for level 1
- 14 respondents voted for level 2
- 8 respondents voted for level 3
- 2 respondents voted for level 4
- 7 respondents voted for level 5
- 13 respondents voted for level 6
- 5 respondents voted for level 7

○

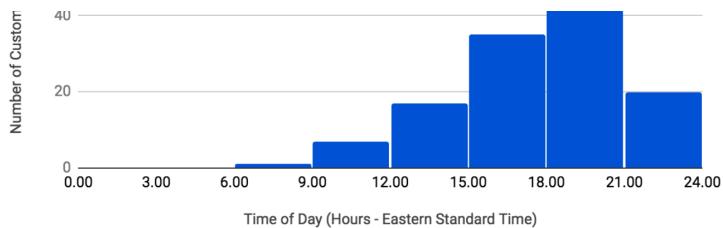


A histogram describes the time users spent looking for answers. The horizontal x-axis shows duration spent looking for an answer in minutes, ranging from 0.00 to 35.00, each bin representing a 5-minute time period. The vertical y-axis shows the number of users (in thousands) ranging from 0 to 50, in increments of 10. Data in the histogram can be summarized accurately as:

- 0-5 minutes, 43 people
- 5-10 minutes, 14 people
- 10-15 minutes, 10 people
- 15-20 minutes, 7 people
- 20-25 minutes, 1 people
- 25-30 minutes, 0 people
- 30-35 minutes, 1 people

○





A histogram describes the number of customers on website in different time periods of a day. The horizontal x-axis shows time of day in hours in eastern standard time, ranging from 0 o'clock to 24 o'clock, each bin representing a 3-hour time period. The vertical y-axis shows number of customers on website in different time periods, ranging from 0 to 80, in increments of 20. Data in the histogram can be summarized accurately as:

- 0-3 o'clock, 0 customer
- 3-6 o'clock, 0 customer
- 6-9 o'clock, 1 customer
- 9-12 o'clock, 7 customers
- 12-15 o'clock, 17 customers
- 15-18 o'clock, 35 customers
- 18-21 o'clock, 68 customers
- 21-24 o'clock, 20 customers

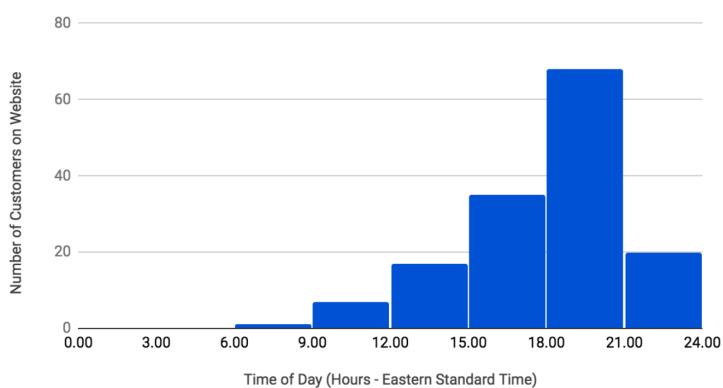
Correct

Topic to review: distributions, histograms, interpreting graphs and diagrams

20. Which of the following graphs match the following distribution description?

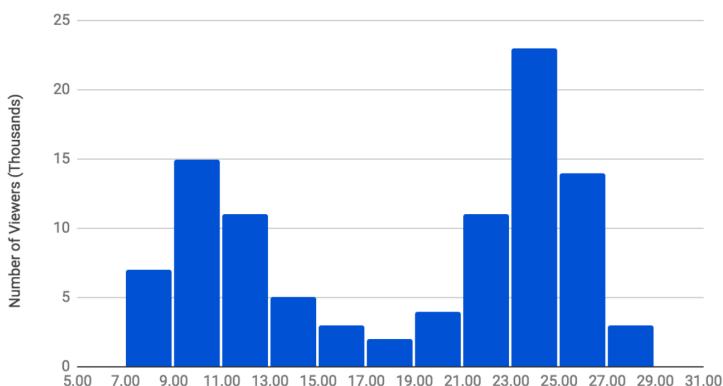
1 / 1 point

unimodal, skewed to the left, no apparent outliers



A histogram describes the number of customers on a website in different time periods of a day. The horizontal x-axis shows time of day in hours in eastern standard time, ranging from 0 o'clock to 24 o'clock, each bin representing a 3-hour time period. The vertical y-axis shows number of customers on website in different time periods, ranging from 0 to 80, in increments of 20. Data in the histogram can be summarized accurately as:

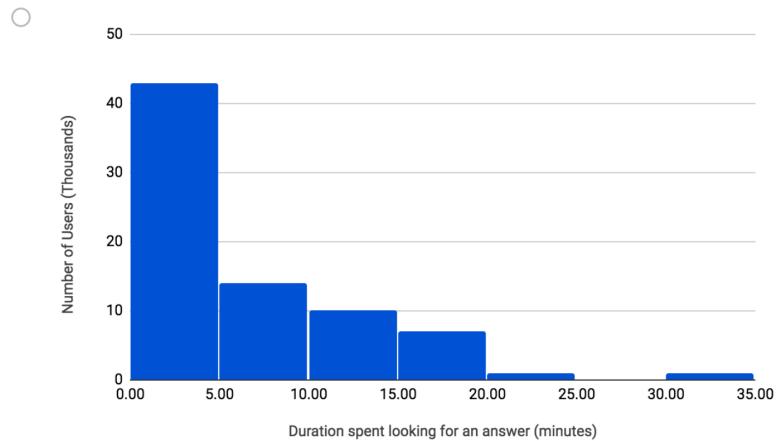
- 0-3 o'clock, 0 customer
- 3-6 o'clock, 0 customer
- 6-9 o'clock, 1 customer
- 9-12 o'clock, 7 customers
- 12-15 o'clock, 17 customers
- 15-18 o'clock, 35 customers
- 18-21 o'clock, 68 customers
- 21-24 o'clock, 20 customers



Age of Viewership (Years old)

Image Description: A histogram describes the number and age of viewership. The horizontal x-axis shows the age of viewership, ranging from 5 years old to 30 years old, each bin representing a 2-year old period. The vertical y-axis shows the number of viewers in thousands, ranging from 0 to 25, in increments of 5. Data in the histogram can be summarized accurately as:

- 5-7 years old, 0
- 7-9 years old, 7
- 9-11 years old, 15
- 11-13 years old, 11
- 13-15 years old, 5
- 15-17 years old, 3
- 17-19 years old, 2
- 19-21 years old, 4
- 21-23 years old, 11
- 23-25 years old, 23
- 25-27 years old, 14
- 27-29 years old, 3
- 29-31 years old, 0



A histogram describes the time users spent looking for answers. The horizontal x-axis shows duration spent looking for an answer in minutes, ranging from 0.00 to 35.00, each bin representing a 5-minute time period. The vertical y-axis shows the number of users (in thousands) ranging from 0 to 50, in increments of 10. Data in the histogram can be summarized accurately as:

- 0-5 minutes, 43 people
- 5-10 minutes, 14 people
- 10-15 minutes, 10 people
- 15-20 minutes, 7 people
- 20-25 minutes, 1 people
- 25-30 minutes, 0 people
- 30-35 minutes, 1 people

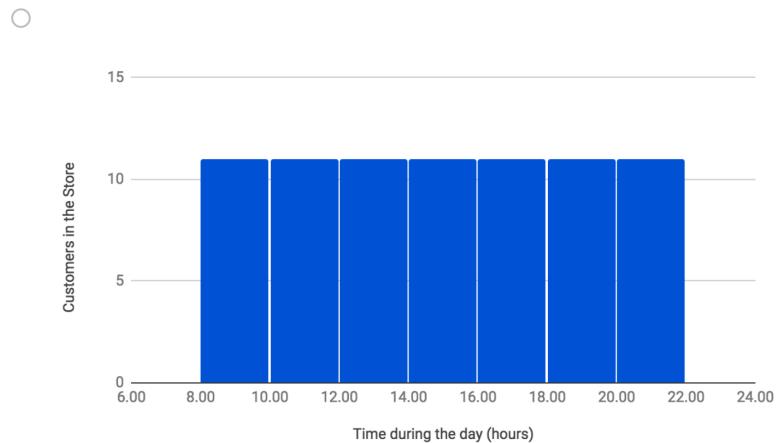


Image Description: A histogram describes the number of customers in store in different time periods of a day. The horizontal x-axis shows time slots of day ranging from 6 o'clock to 24 o'clock, each bin representing a 2-hour time period. The vertical y-axis shows the number of customers in store ranging from 0 to 15, in increments of 5. Data in the histogram can be summarized accurately as:

- 6-8 hours, 0 customers
- 8-10 hours, 11 customers
- 10-12 hours, 11 customers
- 12-14 hours, 11 customers
- 14-16 hours, 11 customers
- 16-18 hours, 11 customers

- 18-21 hours, 11 customers
- 20-22 hours, 11 customers
- 22-24 hours, 0 customers

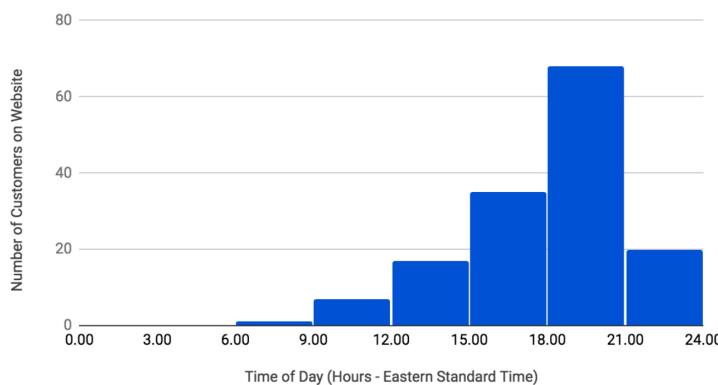
 Correct

Topic to review: distributions, histograms, interpreting graphs and diagrams

21. Which of the following graphs match the following distribution description?

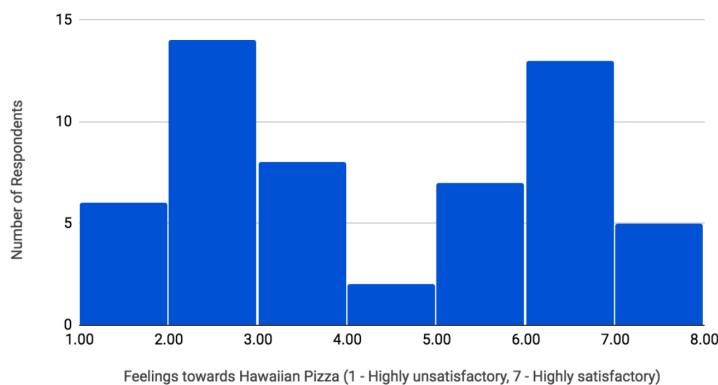
1 / 1 point

bimodal, symmetrical, no apparent outliers



A histogram describes the number of customers on website in different time periods of a day. The horizontal x-axis shows time of day in hours in eastern standard time, ranging from 0 o'clock to 24 o'clock, each bin representing a 3-hour time period. The vertical y-axis shows number of customers on website in different time periods, ranging from 0 to 80, in increments of 20. Data in the histogram can be summarized accurately as:

- 0-3 o'clock, 0 customer
- 3-6 o'clock, 0 customer
- 6-9 o'clock, 1 customer
- 9-12 o'clock, 7 customers
- 12-15 o'clock, 17 customers
- 15-18 o'clock, 35 customers
- 18-21 o'clock, 68 customers
- 21-24 o'clock, 20 customers



Feelings towards Hawaiian Pizza (1 - Highly unsatisfactory, 7 - Highly satisfactory)

A histogram describes the distribution of respondents' satisfaction level towards Hawaiian pizza. The horizontal x-axis shows respondents' feelings, ranging from 1 to 7, 1 means highly unsatisfactory, 7 means highly satisfactory, each bin representing a 1 satisfaction level. The vertical y-axis shows the number of respondents, ranging from 0 to 15, in increments of 5. Data in the histogram can be summarized accurately as:

- 6 respondents voted for level 1
- 14 respondents voted for level 2
- 8 respondents voted for level 3
- 2 respondents voted for level 4
- 7 respondents voted for level 5
- 13 respondents voted for level 6
- 5 respondents voted for level 7



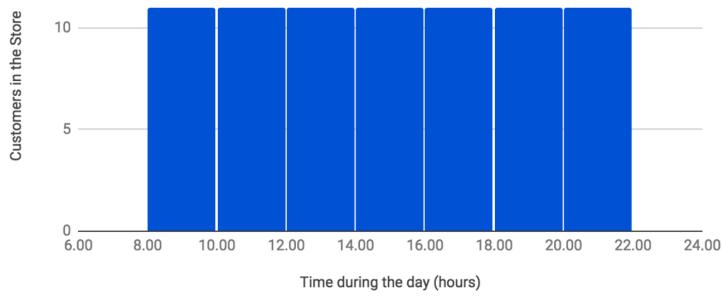
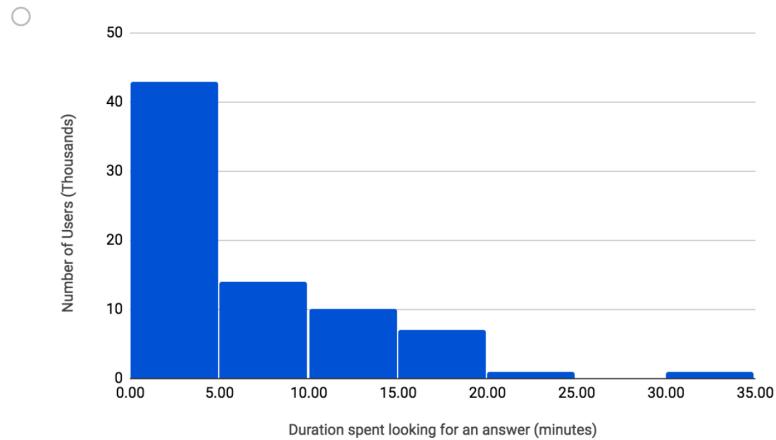


Image Description: A histogram describes the number of customers in store in different time periods of a day. The horizontal x-axis shows time slots of day ranging from 6 o'clock to 24 o'clock, each bin representing a 2-hour time period. The vertical y-axis shows the number of customers in store ranging from 0 to 15, in increments of 5. Data in the histogram can be summarized accurately as:

- 6-8 hours, 0 customers
- 8-10 hours, 11 customers
- 10-12 hours, 11 customers
- 12-14 hours, 11 customers
- 14-16 hours, 11 customers
- 16-18 hours, 11 customers
- 18-20 hours, 11 customers
- 20-22 hours, 11 customers
- 22-24 hours, 0 customers



A histogram describes the time users spent looking for answers. The horizontal x-axis shows duration spent looking for an answer in minutes, ranging from 0.00 to 35.00, each bin representing a 5-minute time period. The vertical y-axis shows the number of users (in thousands) ranging from 0 to 50, in increments of 10. Data in the histogram can be summarized accurately as:

- 0-5 minutes, 43 people
- 5-10 minutes, 14 people
- 10-15 minutes, 10 people
- 15-20 minutes, 7 people
- 20-25 minutes, 1 people
- 25-30 minutes, 0 people
- 30-35 minutes, 1 people

Correct

Topic to review: distributions, histograms, interpreting graphs and diagrams

22. When testing whether the means of more than two populations are equal, which test would we use?

1 / 1 point

- ANOVA
- CHI-Square
- T-test
- F-test

Correct

Topics to review: Situations for different types of tests - ANOVA, CHI-Square, Population Proportion, Population Mean, Two Population Proportions, Two Population Means, Regression

23. For the next three questions, use the following information to determine your answers: The length of a movie falls on a normal distribution. About 95% of movies fall between 75 minutes and 163 minutes.

1 / 1 point

What is the mean value for average movie length in minutes? Your answer should be a whole number.

119

✓ Correct

Topics to review: distribution, calculating the mean

24. What is the value of the standard deviation for average movie length in minutes? Please round to the second decimal place.

1 / 1 point

22.45

✓ Correct

Key topics to review for this question: distribution, standard deviation

25. Suppose you watched a movie that was 182 minutes long. How many standard deviations is that movie from the mean? Please round to the second decimal point.

1 / 1 point

2.81

✓ Correct

Key topics to review for this question: distribution, standard deviation

26. For the next three questions use the following information to determine your answers. A research group is curious about features that can be attributed to music genres. A music streaming service provides a few different attributes for songs such as speechiness, danceability, and valence. They suspect that there is a difference between the average valence (positive or negative emotion) of metal songs compared to blues songs. However, they must conduct a study to determine if that is true. From a sample of 87 metal songs, the sample mean for valence is 0.451 and the sample standard deviation is 0.139. From a sample of 94 blues songs, the sample mean for valence is 0.581 and the sample standard deviation is 0.167.

1 / 1 point

Assume that sample1 comes from the sample metal songs and that sample2 comes from the sample blues songs

Compute the 90% confidence interval. Please round the values to the fourth decimal point and format your response as follows: (lower_value, upper_value)

(-0.1679, -0.0921)

✓ Correct

Topics to review: confidence intervals, population means

27. Which of the following represents the hypotheses that we will be testing, assuming that μ_1 represents the population mean of valence for all metal songs and that μ_2 represents the population mean of valence for all blues songs.

1 / 1 point

- H₀: $\mu_1 = \mu_2$ versus H_a: $\mu_1 \neq \mu_2$
 H₀: $\mu_1 = \mu_2$ versus H_a: $\mu_1 > \mu_2$
 H₀: $\mu_1 = \mu_2$ versus H_a: $\mu_1 < \mu_2$

✓ Correct

Consider what you're testing: is it that μ_1 is greater than, less than, or just different from μ_2 but not in a particular direction?

28. Calculate the p-value and determine if we should accept or reject H₀ under alpha = 0.10.

1 / 1 point

reject

✓ Correct

Topic to review: population mean, hypothesis testing, Type I and Type II errors, calculating the p-value, interpreting the p-value

29. Select the correct p-values that match the statements. Some conclusions may be applicable to more than one p-value.

1 / 1 point

There is not enough evidence to reject the null hypothesis at the 5% level.

0.0960

✓ Correct

This number is above 0.05 so when alpha is set at 5%, then this number is not significant.

0.0740

 **Correct**

This number is above 0.05 so when alpha is set at 5%, then this number is not significant.

0.0956

 **Correct**

This number is above 0.05 so when alpha is set at 5%, then this number is not significant.

0.0354

30. Select the correct p-values that match the statements. Some conclusions may be applicable to more than one p-value.

1 / 1 point

The evidence is significant at the 5% level, but not at the 1% level.

0.0192

 **Correct**

This number is below 0.05, but is above 0.01, so when alpha is set at 5%, then this number is significant. When alpha is set at 1% then this number is not significant.

0.0556

0.0290

 **Correct**

This number is below 0.05, but is above 0.01, so when alpha is set at 5%, then this number is significant. When alpha is set at 1% then this number is not significant.

0.0097

31. Select the correct p-values that match the statements. Some conclusions may be applicable to more than one p-value.

1 / 1 point

Evidence against the null hypothesis is significant at the 1% level.

0.0037

 **Correct**

This is below 0.01, so if alpha is set to 1%, then we would consider the evidence to be significant.

0.0621

0.0949

0.0209