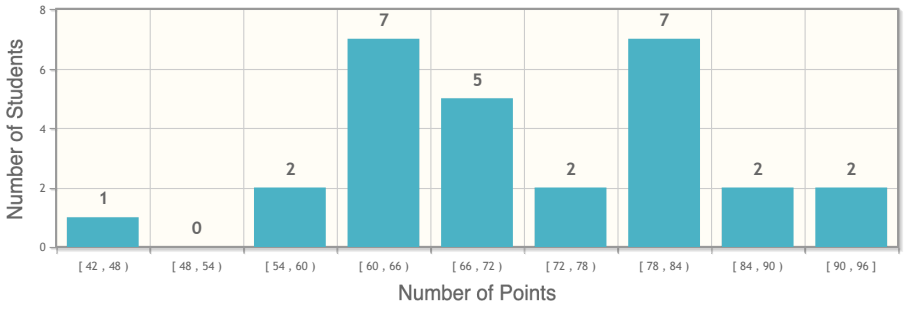
Statistics: Mid-term exam

View Highest Submission

Grade Statistics	
Submissions	28
Total Score Possible	100
Mean	72
Median	69.5
Mode	80
Range	42 - 96
Quartile 1	64
Quartile 3	80
Standard Deviation	12.47

Final Score Distribution



Questions

Part : All Parts

Part 1, Question 1 (True/False Question)

A scientist wanted to learn the effect of a new drug. She randomly selected 50 patients from a hospital where this new drug is tried out, and also randomly selected another 50 patients from a different hospital where a traditional treatment was adopted. She followed up with all patients after 6 months to observe their recovery, and compared the outcomes between patients in the two hospitals. Her analysis WILL reveal a causal relationship between the new drug and patient recovery.

Answer Options	Correctness	Number of Responses
True	Not Correct	14
False	Correct	14

28 Responses, **50%** Answered Correctly

Part 1, Question 2 (True/False Question)

For data with a right-skewed distribution, the sample mean is definitely larger than the median.

Answer Options	Correctness	Number of Responses
True	Correct	25
False	Not Correct	2

27 Responses, **92%** Answered Correctly

Part 1, Question 3 (True/False Question)

Box plots can help us identify extreme values and potential outliers, but histograms can not.

Answer Options	Correctness	Number of Responses
True	Not Correct	4
False	Correct	24

28 Responses, **85%** Answered Correctly

Part 1, Question 4 (True/False Question)

The average percentage of Stephen Curry making a 3-point shot during the 20-21 season was 0.429. (That is, on average, 42.9% of his 3-point shot attempts were successful.) Thus, the probability of him making three 3-point shots in a row should be $0.429^3 \approx 0.079$.

Answer Options	Correctness	Number of Responses
True	Not Correct	27
False	Correct	1

28 Responses, **3%** Answered Correctly

Part 1, Question 5 (True/False Question)

If a given value (for example, the actual population mean of household income) is within a 90% confidence interval, it will definitely also be within a 95% confidence interval calculated based on the same sample.

Answer Options	Correctness	Number of Responses
True	Correct	25
False	Not Correct	3

28 Responses, **89%** Answered Correctly

Part 1, Question 6 (True/False Question)

The zip code of Duke campus is 27708. A "zip code" can be considered as a numerical variable since it consists of digits only.

Answer Options	Correctness	Number of Responses
True	Not Correct	0
False	Correct	28

28 Responses, **100%** Answered Correctly

Part 1, Question 7 (True/False Question)

Suppose we want to investigate if there is a difference in the gender proportions between Math majors and Sociology majors at Duke. It is reasonable to set up our null and alternative hypotheses in the following manner:

H_0: the proportion of female students in the Math department is equal to that in the Sociology department.

H_A: the proportion of female students in the Math department is NOT equal to that in the Sociology department.

Answer Options	Correctness	Number of Responses
True	Correct	27
False	Not Correct	1

28 Responses, **96%** Answered Correctly

Part 1, Question 8 (True/False Question)

Suppose we want to test a hypothesis using a dataset. Prior to analyzing the data, we set the significance level at $\alpha = 0.05$, and after analyzing the data we get a p-value of 0.04. Then the null hypothesis must be false.

Answer Options	Correctness	Number of Responses
True	Not Correct	10
False	Correct	18

28 Responses, **64%** Answered Correctly

Part 2, Question 1 (Numeric Response)

The height of US women aged 20 or above has a mean of 63.7 inches, with a standard deviation of 2.5 inches. Assume that the height of a US woman (aged 20 or above) follows a normal distribution. Then the probability that a woman is taller than 68.7 inches is ___.
(Note: round to 3 digits after the decimal point.)

Answer Options	Number of Correct Responses
0.021|0.024	18

28 Responses, **64%** Answered Correctly

Part 2, Question 2 (Numeric Response)

To estimate the proportion of Duke students satisfied with their living conditions, Alice took a random sample of 100 students. Upon calculation, she found out that the standard error of her estimate is 0.1. To reduce the standard error down to 0.05, she has to take a larger sample, with sample size increased to ___ at least.

Answer Options	Number of Correct Responses
400	23

28 Responses, **82%** Answered Correctly

Part 2, Question 3 (Numeric Response)

Load the Boston housing dataset included in the "MASS" package in R using the following two commands: (you can do it in your RStudio container or your local R environment)

```
library(MASS)
data("Boston")
```

Check out the dataset and fill in the blanks: there are ___ observations and ___ variables in total.

Answer Options	Number of Correct Responses
506	28
14	28

28 Responses, **100%** Answered Correctly

Part 2, Question 4 (Numeric Response)

A rapid test for HIV has a 99% accuracy on HIV-positive cases and a 95% accuracy on HIV-negative cases. (That is, it returns a positive result 99% of the time for someone with HIV, and a negative result 95% the time for someone without it.)

The HIV prevalence of a certain country is 6%. A randomly selected person from this country gets this rapid test and receives a positive result. The probability that he actually has HIV is ___%.

(Note: round to 1 digit after the decimal point.)

Answer Options	Number of Correct Responses
54.5|57.0	13

28 Responses, **46%** Answered Correctly

Part 2, Question 5 (Numeric Response)

Tom buys lunch from a sandwich shop every day. He typically has a sandwich and a side. The price of sandwiches has a mean of $5 and a standard deviation of $1.6, while the price of sides has a mean of $3 and a standard deviation of $1.2. The prices of a sandwich and a side are independent. Then Tom's average cost of a lunch is $___ with a standard deviation of $___.
(NOTE: round to 1 digit after the decimal point.)

Answer Options	Number of Correct Responses
8.0	27
1.8|2.2	22

28 Responses, **78%** Answered Correctly

Part 2, Question 6 (Numeric Response)

On average, 3 major earthquakes (i.e., earthquakes with a magnitude of 5 or higher) occur in Japan during a week. Let's assume that the number of earthquakes in Japan follow a Poisson distribution, then the probability of 4 major earthquakes taking place across Japan during a week is ___.

(Note: round to 3 digits after the decimal point.)

Answer Options	Number of Correct Responses
0.155|0.180	25

28 Responses, **89%** Answered Correctly

Part 2, Question 7 (Numeric Response)

A political scientist randomly surveyed 100 Durham voters on whether or not they voted for Joe Biden in the 2020 presidential election. The results showed that 80 of them did. She used her survey outcomes to construct a 95% confidence interval for the proportion among all Durham voters who voted for Biden. The 95% confidence interval would be (____,____).

(Note: round to 2 digits after the decimal point.)

Answer Options	Number of Correct Responses
0.70|0.74	21
0.85|0.89	21

28 Responses, **75%** Answered Correctly

Part 2, Question 8 (Numeric Response)

Jack is a knife throwing performer and he can successfully hit the target 80% of the time. Assume that each of his knife-throw attempts has the same success rate (80%) and all his knife-throws are independent. The probability that 3 out of 5 knives that he throws hit the target is ___.

(Note: round to 3 digits after the decimal point.)

Answer Options	Number of Correct Responses
0.180|0.220	19

28 Responses, **67%** Answered Correctly

Part 2, Question 9 (Numeric Response)

Among all customers in a bar, 80% would order alcoholic drinks, 69% would order snacks, and 55% would order both drinks and snacks. Then ___% of all customers in this bar would order either drinks or snacks.

Answer Options	Number of Correct Responses
94	23

28 Responses, **82%** Answered Correctly

Part 3, Question 1 (Multiple Correct, Multiple Selection)

The attached plot (see attachment) is a scatterplot that visualizes the relationship between the frontal lobe size and rear width of 200 Leptograpsus crabs in west Australia. Which of the following statements are correct? (Select ALL that are correct.)

Answer Options	Correctness	Number of Responses
Frontal lobe sizes and rear widths are strongly associated.	Correct	25
Frontal lobe sizes and rear widths are independent variables.	Not Correct	2
The relationship between frontal lobe sizes and rear widths appears linear.	Correct	27
The relationship between frontal lobe sizes and rear widths is nonlinear.	Not Correct	1

28 Responses, **82%** Answered Correctly

Part 3, Question 2 (Single Correct)

A political scientist is interested in the effect of economic development on social equality. She wants to use a sample of 50 countries evenly represented among the Americas, Europe, Asia, and Africa to conduct her analysis. What type of study and strategy should she use to ensure that countries are selected from each region of the world?

Answer Options	Correctness	Number of Responses
Observational study, with simple random sampling.	Not Correct	8
Observational study, with cluster sampling.	Not Correct	5
Observational study, with stratified sampling.	Correct	14
Experiment, with random assignment.	Not Correct	0
Experiment, with blocking.	Not Correct	1

28 Responses, **50%** Answered Correctly

Part 3, Question 3 (Single Correct)

The "iris" data contain 50 observations of 3 species of iris flowers, including the sepal length and width and petal length and width for each flower.
You may use the following command to load the data in R:
`data("iris")`
Check out the iris dataset. Which type of plot would be the most useful in visualizing the relationship between the species and petal length of these flowers?

Answer Options	Correctness	Number of Responses
histogram	Not Correct	0
side-by-side box plot	Correct	13
side-by-side bar plot	Not Correct	15
dot plot	Not Correct	0

28 Responses, **46%** Answered Correctly

Part 3, Question 4 (Single Correct)

A comprehensive survey conducted on Duke students show that the true proportion of all Duke students who have taken at least one Statistics course is 0.4. You survey 60 students in your dorm and record that the proportion of students who have taken Statistics courses is 0.25. The proportion of all students at this college who have taken Statistics courses is a _____ and the proportion of students who have taken Statistics courses in your dorm is a _____.

Answer Options	Correctness	Number of Responses
parameter; statistic.	Correct	12
statistic; parameter.	Not Correct	0
population; sample.	Not Correct	15
measure of central tendency; measure of variability.	Not Correct	0
None of the other options is correct.	Not Correct	1

28 Responses, **42%** Answered Correctly

Part 3, Question 5 (Multiple Correct, Multiple Selection)

Which of the following statements about z-scores is/are true?

Answer Options	Correctness	Number of Responses
Larger z-scores are always better.	Not Correct	0
The z-score for an observation that is equal to the mean is 0.	Correct	27
If a z-score is 2 that means that the observation is two times the value of the mean.	Not Correct	0
If a z-score is negative that means that the observation is less than mean.	Correct	28

28 Responses, **96%** Answered Correctly

Part 3, Question 6 (Single Correct)

About conditions for applying the Central Limit Theorem when estimating single proportions, which of the following statements is true?

Answer Options	Correctness	Number of Responses
The observations in the data can be dependent.	Not Correct	1
If sample size n is larger than 1000, then we can definitely assume that the sample proportion approximately follows a normal distribution.	Not Correct	4
We can use the sample proportion as an approximation of the true proportion to check the success-failure condition.	Correct	14
The size sample is considered as sufficiently large if one of np and n(1-p) is larger than 10.	Not Correct	9

28 Responses, **50%** Answered Correctly

Part 3, Question 7 (Matching)

You are given 4 datasets with their distributions visualized in the 4 histograms (a), (b), (d), and (d) (see the attached image). Please match them with the correct descriptions of these data distributions

Answer Options	Number of Correct Responses
Bimodal	17
Multi-modal	17
Right-skewed	17
Uniform	19

28 Responses, **50%** Answered Correctly

Part 3, Question 8 (Single Correct)

Kim wants to test if a coin is fair. She conducts hypothesis testing where the null hypothesis (H_0) is "the coin is fair". After tossing the coin many times, she decides not to reject the hypothesis that the coin is indeed fair. What decision error could she have made?

Answer Options	Correctness	Number of Responses
Type 1 error.	Not Correct	0
Type 2 error.	Correct	26
Both type 1 and type 2 errors.	Not Correct	2
Neither type 1 nor type 2 errors.	Not Correct	0

28 Responses, **92%** Answered Correctly