Question # 0 Statistics Bowl MAΘ National Convention 2022

Compute each of the requested values for the following sample data set:

{336, 337, 338}

A: What is the sample mean of the data set as an exact value?

B: What is the sample standard deviation of the data set as an exact value?

C: What is the median of the data set?

D: How may outliers are there in this data set according to the 1.5(IQR) rule?

Summary: Compute 3(AB + C + D).

Question # 0 Statistics Bowl MAΘ National Convention 2022

Compute each of the requested values for the following sample data set:

{336, 337, 338}

A: What is the sample mean of the data set as an exact value?

B: What is the sample standard deviation of the data set as an exact value?

C: What is the median of the data set?

D: How may outliers are there in this data set according to the 1.5(IQR) rule?

Summary: Compute 3(AB + C + D).

Question # 1 Statistics Bowl   MAΘ National Convention 2022

Compute each of the requested values for the following sample data set of 20 values:

{1, 2, 3, 4, 5, 5, 7, 8, 10, 10, 10, 10, 10, 12, 15, 15, 16, 17, 20, 20}

A: Determine the mean, median, and mode of the data set. What is the product of these three values?

B: What is the sample standard deviation of the data set rounded to the thousandths place?

C: Convert each and every value in the data set into a Z-score using the sample mean of the data set and the exact sample standard deviation of the data set. What is the sum of the mean and the standard deviation of this resulting set of Z-scores?

D: How many outliers are there in this data set according to the 1.5(IQR) rule?

Summary: Compute A(B + C + D) using your final rounded values from above.

Question # 1 Statistics Bowl   MAΘ National Convention 2022

Compute each of the requested values for the following sample data set of 20 values:

{1, 2, 3, 4, 5, 5, 7, 8, 10, 10, 10, 10, 10, 12, 15, 15, 16, 17, 20, 20}

A: Determine the mean, median, and mode of the data set. What is the product of these three values?

B: What is the sample standard deviation of the data set rounded to the thousandths place?

C: Convert each and every value in the data set into a Z-score using the sample mean of the data set and the exact sample standard deviation of the data set. What is the sum of the mean and the standard deviation of this resulting set of Z-scores?

D: How many outliers are there in this data set according to the 1.5(IQR) rule?

Summary: Compute A(B + C + D) using your final rounded values from above.

Question # 2 Statistics Bowl — MAΘ National Convention 2022

Fill in the blank for each numbered description in Column X with a valid response number from the list in Column Y to form a set of ordered pairs (x, y). NOTE: More than one choice in column Y is valid for some descriptions in column X and it is even possible that none are valid. Therefore, the correspondence between Column X and Column Y is neither one-to-one nor onto and so you must include all possible valid ordered pairs. For example, suppose the blank in Column X Statement 1 maps to both #1 and #2 in Column Y – this would give you two valid ordered pairs: (1, 1) and (1, 2). It is even possible that none of the values in Column Y correctly correspond to a value in Column X, and so there is no valid ordered pair contributing to the summary answer from that statement. HINT: You should have a total of 8 valid ordered pairs when you are done!

Column X

1. A _____ is a valid way to display the distribution of data from a single categorical variable.
2. A _____ is a valid way to display the distribution of data from a single quantitative variable.
3. A _____ is a valid way to display the relationship between two categorical variables.
4. A _____ is a valid way to display the relationship between two quantitative variables.

Column Y

1. Bar Graph
2. Boxplot
3. Dotplot
4. Histogram
5. Pie Chart
6. Residual Plot
7. Scatterplot
8. Stemplot

Summary: What is the sample Pearson product-moment coefficient of linear correlation between column X and Column Y using the set of 8 ordered pairs you just formed rounded to the hundredths place?

Question # 2 Statistics Bowl — MAΘ National Convention 2022

Fill in the blank for each numbered description in Column X with a valid response number from the list in Column Y to form a set of ordered pairs (x, y). NOTE: More than one choice in column Y is valid for some descriptions in column X and it is even possible that none are valid. Therefore, the correspondence between Column X and Column Y is neither one-to-one nor onto and so you must include all possible valid ordered pairs. For example, suppose the blank in Column X Statement 1 maps to both #1 and #2 in Column Y – this would give you two valid ordered pairs: (1, 1) and (1, 2). It is even possible that none of the values in Column Y correctly correspond to a value in Column X, and so there is no valid ordered pair contributing to the summary answer from that statement. HINT: You should have a total of 8 valid ordered pairs when you are done!

Column X

1. A _____ is a valid way to display the distribution of data from a single categorical variable.
2. A _____ is a valid way to display the distribution of data from a single quantitative variable.
3. A _____ is a valid way to display the relationship between two categorical variables.
4. A _____ is a valid way to display the relationship between two quantitative variables.

Column Y

1. Bar Graph
2. Boxplot
3. Dotplot
4. Histogram
5. Pie Chart
6. Residual Plot
7. Scatterplot
8. Stemplot

Summary: What is the sample Pearson product-moment coefficient of linear correlation between column X and Column Y using the set of 8 ordered pairs you just formed rounded to the hundredths place?

Question # 3 Statistics Bowl MAΘ National Convention 2022

An experiment is conducted where a potent antibacterial mixture is added to a bacterial culture that initially contains 1000 live bacteria. The table below displays the number of living bacteria (Y) after each minute of exposure (X) to the antibacterial mixture. Fit a least squares linear regression model to the data and answer each of the questions that follow. IMPORTANT NOTE: Perform each of the requested calculations in spite of the fact that a linear regression analysis of the data is potentially not valid. Also, you will need this data again for the next question, so keep it in your calculator!

X = Exposure Time (minutes)	0	1	2	3	4	5	6	7	8	9
Y = Number of Living Bacteria	1000	740	490	270	120	50	15	11	7	2

A: What is the sum of the slope and the y-intercept of the least squares linear regression model fit to this data rounded to the nearest thousandth?

B: What is the sum of the coefficient of linear correlation and the linear coefficient of determination of the least squares linear regression model fit to this data rounded to the nearest thousandth?

C: What is the residual of the point (4, 120) rounded to the nearest whole bacterium?

D: Use the least squares linear regression model fit to this data to predict how many bacteria are left after 8.5 minutes rounded to the nearest whole bacterium. If this prediction is reasonably accurate, let D equal the sum of 150 and your rounded result. If this prediction is not reasonably accurate, then let D just equal your rounded result.

Summary: Using your final rounded answers from above, compute A + B + C + D to the thousandths place.

Question # 3 Statistics Bowl MAΘ National Convention 2022

An experiment is conducted where a potent antibacterial mixture is added to a bacterial culture that initially contains 1000 live bacteria. The table below displays the number of living bacteria (Y) after each minute of exposure (X) to the antibacterial mixture. Fit a least squares linear regression model to the data and answer each of the questions that follow. IMPORTANT NOTE: Perform each of the requested calculations in spite of the fact that a linear regression analysis of the data is potentially not valid. Also, you will need this data again for the next question, so keep it in your calculator!

X = Exposure Time (minutes)	0	1	2	3	4	5	6	7	8	9
Y = Number of Living Bacteria	1000	740	490	270	120	50	15	11	7	2

A: What is the sum of the slope and the y-intercept of the least squares linear regression model fit to this data rounded to the nearest thousandth?

B: What is the sum of the coefficient of linear correlation and the linear coefficient of determination of the least squares linear regression model fit to this data rounded to the nearest thousandth?

C: What is the residual of the point (4, 120) rounded to the nearest whole bacterium?

D: Use the least squares linear regression model fit to this data to predict how many bacteria are left after 8.5 minutes rounded to the nearest whole bacterium. If this prediction is reasonably accurate, let D equal the sum of 150 and your rounded result. If this prediction is not reasonably accurate, then let D just equal your rounded result.

Summary: Using your final rounded answers from above, compute A + B + C + D to the thousandths place.

Question # 4 Statistics Bowl — MAΘ National Convention 2022

Every good statistician knows that you should always graph your data in order to reveal any meaningful patterns. So, hopefully you created a scatterplot of the data from the previous question where an experiment was conducted with a potent antibacterial mixture added to a bacterial culture that initially contained 1000 live bacteria. Again, the table below displays the number of living bacteria (Y) after each minute of exposure (X) to the antibacterial mixture. However, a scatterplot of the original bivariate data reveals that the pattern is clearly not linear! Thus, transform the dependent variable Y using the natural logarithm into ln(Y). Create a new least squares linear regression model with ln(Y) as the new dependent variable (thus, keeping X as the original explanatory variable) and answer the questions that follow.

X = Exposure Time (minutes)	0	1	2	3	4	5	6	7	8	9
Y = Number of Living Bacteria	1000	740	490	270	120	50	15	11	7	2

A: What is the absolute value of the increase in the proportion of explained variance between the two linear regression models (the original Y vs. X and the transformed ln(Y) vs. X) expressed as a decimal rounded to the thousandths place?

B: Round each regression coefficient for the linear model of ln(Y) vs. X to four decimal places and use them to predict the number of bacteria remaining after 7 minutes rounding the final result to the nearest whole bacterium. What is the now the residual of the point (7, 11) using this result? NOTE: Your final answer should be an integer.

C: Using the model for ln(Y) vs. X, after how many minutes (to the nearest whole minute) will the predicted / expected number of living bacteria remaining round to 0? HINT: You must extrapolate beyond the observed range of x-values given above by guessing and checking whole minutes starting at 10 minutes – don't worry, it won't take long!

D: Is the new model [ln(Y) vs. X] you created a POWER model or an EXPONENTIAL model? Let D = the number of letters in the correct word choice. Thus, either D = 5 or D = 11 accordingly – but which is it?

Summary: Using your final rounded values from above, calculate $A + B + C + D$ to the nearest thousandth.

Question # 4 Statistics Bowl — MAΘ National Convention 2022

Every good statistician knows that you should always graph your data in order to reveal any meaningful patterns. So, hopefully you created a scatterplot of the data from the previous question where an experiment was conducted with a potent antibacterial mixture added to a bacterial culture that initially contained 1000 live bacteria. Again, the table below displays the number of living bacteria (Y) after each minute of exposure (X) to the antibacterial mixture. However, a scatterplot of the original bivariate data reveals that the pattern is clearly not linear! Thus, transform the dependent variable Y using the natural logarithm into ln(Y). Create a new least squares linear regression model with ln(Y) as the new dependent variable (thus, keeping X as the original explanatory variable) and answer the questions that follow.

X = Exposure Time (minutes)	0	1	2	3	4	5	6	7	8	9
Y = Number of Living Bacteria	1000	740	490	270	120	50	15	11	7	2

A: What is the absolute value of the increase in the proportion of explained variance between the two linear regression models (the original Y vs. X and the transformed ln(Y) vs. X) expressed as a decimal rounded to the thousandths place?

B: Round each regression coefficient for the linear model of ln(Y) vs. X to four decimal places and use them to predict the number of bacteria remaining after 7 minutes rounding the final result to the nearest whole bacterium. What is the now the residual of the point (7, 11) using this result? NOTE: Your final answer should be an integer.

C: Using the model for ln(Y) vs. X, after how many minutes (to the nearest whole minute) will the predicted / expected number of living bacteria remaining round to 0? HINT: You must extrapolate beyond the observed range of x-values given above by guessing and checking whole minutes starting at 10 minutes – don't worry, it won't take long!

D: Is the new model [ln(Y) vs. X] you created a POWER model or an EXPONENTIAL model? Let D = the number of letters in the correct word choice. Thus, either D = 5 or D = 11 accordingly – but which is it?

Summary: Using your final rounded values from above, calculate $A + B + C + D$ to the nearest thousandth.

Question # 5 Statistics Bowl MAΘ National Convention 2022

Match the definition in Column X with the correct term in Column Y to fill in the blank and to form a set of ordered pairs (x, y). This time, there is a one-to-one and onto correspondence between Column X and Column Y (thus, a bijection).

Column X

1. Using a regression model to make predictions far outside the range of the original x data values used to create it is called _____ .
2. A sampling method in which each individual from the population that is chosen for the sample is chosen via some form of chance process is called _____ .
3. A sampling method where each sample of a specified size has the same chance of being the sample actually chosen is a _____ .
4. A sampling method where a random sample is chosen from each one of two or more non-overlapping subcategories within the population is a _____ .
5. A sampling method where the population is subdivided into smaller non-overlapping subsets and all individuals in a random sample of these subsets are used together as the whole sample is a _____ .
6. The design of a statistical study shows ____ if it is very likely to underestimate or very likely to overestimate the value you want to know.
7. When the tail of the distribution of a quantitative data set or variable extends out in one direction or the other, we call the distribution _____ .
8. If a data point is removed from a regression analysis and it significantly changes the results of many regression statistics, we call that data point _____ .

Column Y

1. Bias
2. Influential
3. Cluster Random Sample
4. Extrapolation
5. Random Sampling
6. Simple Random Sample
7. Skewed
8. Stratified Random Sample

Summary: What is the correlation coefficient between Column X and Column Y using the correct set of ordered pairs you just created rounded to four decimal places?

Question # 5 Statistics Bowl MAΘ National Convention 2022

Match the definition in Column X with the correct term in Column Y to fill in the blank and to form a set of ordered pairs (x, y). This time, there is a one-to-one and onto correspondence between Column X and Column Y (thus, a bijection).

Column X

1. Using a regression model to make predictions far outside the range of the original x data values used to create it is called _____ .
2. A sampling method in which each individual from the population that is chosen for the sample is chosen via some form of chance process is called _____ .
3. A sampling method where each sample of a specified size has the same chance of being the sample actually chosen is a _____ .
4. A sampling method where a random sample is chosen from each one of two or more non-overlapping subcategories within the population is a _____ .
5. A sampling method where the population is subdivided into smaller non-overlapping subsets and all individuals in a random sample of these subsets are used together as the whole sample is a _____ .
6. The design of a statistical study shows ____ if it is very likely to underestimate or very likely to overestimate the value you want to know.
7. When the tail of the distribution of a quantitative data set or variable extends out in one direction or the other, we call the distribution _____ .
8. If a data point is removed from a regression analysis and it significantly changes the results of many regression statistics, we call that data point _____ .

Column Y

1. Bias
2. Influential
3. Cluster Random Sample
4. Extrapolation
5. Random Sampling
6. Simple Random Sample
7. Skewed
8. Stratified Random Sample

Summary: What is the correlation coefficient between Column X and Column Y using the correct set of ordered pairs you just created rounded to four decimal places?

Question # 6 Statistics Bowl MAΘ National Convention 2022

The observed correct letter choice distribution across a random sample of 8000 past Mu Alpha Theta individual questions for which there was one and only one correct answer is summarized in the table below:

Letter	A	B	C	D	E	Total
Observed Proportion	20.9%	23.8%	24.1%	22.2%	9.0%	100%

Current test writing guidelines recommend that less than 10% of the questions have E as the correct answer with the remaining proportion divided up equally amongst the other four letter choices. Does it seem that this sample of questions provides evidence that individual test writers follow this recommended distribution? Let us first investigate by testing the hypotheses $H_0: P_E = 0.10 \; vs \; H_A: P_E < 0.10$ where P_E represents the population proportion of questions for which the correct answer is E. Perform the indicated left-tail one-proportion Z-test and provide the requested values below. You may assume all inference assumptions and conditions are satisfied.

A: What is the expected number of questions for which the correct answer is E in a sample of this size according to H_0?

B: What is the observed number of questions for which the correct answer is E in this sample?

C: What is the sum of the test statistic and the p-value of the test rounded to four decimal places?

D: If H_0 is rejected at the 1% level of significance, let D = 20. If H_0 is not rejected at the 1% level of significance, let D = 10.

Summary: Using your final rounded results from above, compute $C(A + B - D)$ to the nearest integer.

Question # 6 Statistics Bowl MAΘ National Convention 2022

The observed correct letter choice distribution across a random sample of 8000 past Mu Alpha Theta individual questions for which there was one and only one correct answer is summarized in the table below:

Letter	A	B	C	D	E	Total
Observed Proportion	20.9%	23.8%	24.1%	22.2%	9.0%	100%

Current test writing guidelines recommend that less than 10% of the questions have E as the correct answer with the remaining proportion divided up equally amongst the other four letter choices. Does it seem that this sample of questions provides evidence that individual test writers follow this recommended distribution? Let us first investigate by testing the hypotheses $H_0: P_E = 0.10 \; vs \; H_A: P_E < 0.10$ where P_E represents the population proportion of questions for which the correct answer is E. Perform the indicated left-tail one-proportion Z-test and provide the requested values below. You may assume all inference assumptions and conditions are satisfied.

A: What is the expected number of questions for which the correct answer is E in a sample of this size according to H_0?

B: What is the observed number of questions for which the correct answer is E in this sample?

C: What is the sum of the test statistic and the p-value of the test rounded to four decimal places?

D: If H_0 is rejected at the 1% level of significance, let D = 20. If H_0 is not rejected at the 1% level of significance, let D = 10.

Summary: Using your final rounded results from above, compute $C(A + B - D)$ to the nearest integer.

Here again is the observed correct letter choice distribution across a simple random sample of 8000 past Mu Alpha Theta individual questions for which there was one and only one correct answer:

Letter	A	B	C	D	E	Total
Observed Proportion	20.9%	23.8%	24.1%	22.2%	9.0%	100%

This time, suppose that current test writing guidelines recommend that 10% of the questions have E as the correct answer with the remaining proportion divided up equally amongst the other four letter choices. Does it seem that this sample of questions provides evidence that individual test writers do <u>not</u> follow this recommended overall answer distribution? Compare the observed distribution of letter choices to the expected one using the appropriate statistical test and answer the following questions as you investigate. Again, you may assume all inference assumptions and conditions are satisfied.

A: What is the mean of the expected counts of the five letter choices (A, B, C, D, and E) for the correct answer to a Mu Alpha Theta individual test question in this sample of 8000 questions according to the recommended distribution?

B: What is the sum of the test statistic and degrees of freedom of the appropriate statistical test to determine if these data provide evidence that test writers do <u>not</u> follow the recommended distribution? Round the final answer to the <u>nearest tenth</u>.

C: If this sample of questions provides evidence that individual test writers do <u>not</u> follow the recommended overall answer distribution at the 1% level of significance, then let C = 1. If not, let C = 2.

D: One of the components (a.k.a. contributions) to the test statistic is an integer. What is that integer component?

Summary: Compute $\frac{ABC}{D}$ to the <u>nearest integer</u>.

Question # 7 Statistics Bowl MAΘ National Convention 2022

Here again is the observed correct letter choice distribution across a simple random sample of 8000 past Mu Alpha Theta individual questions for which there was one and only one correct answer:

Letter	A	B	C	D	E	Total
Observed Proportion	20.9%	23.8%	24.1%	22.2%	9.0%	100%

This time, suppose that current test writing guidelines recommend that 10% of the questions have E as the correct answer with the remaining proportion divided up equally amongst the other four letter choices. Does it seem that this sample of questions provides evidence that individual test writers do <u>not</u> follow this recommended overall answer distribution? Compare the observed distribution of letter choices to the expected one using the appropriate statistical test and answer the following questions as you investigate. Again, you may assume all inference assumptions and conditions are satisfied.

A: What is the mean of the expected counts of the five letter choices (A, B, C, D, and E) for the correct answer to a Mu Alpha Theta individual test question in this sample of 8000 questions according to the recommended distribution?

B: What is the sum of the test statistic and degrees of freedom of the appropriate statistical test to determine if these data provide evidence that test writers do <u>not</u> follow the recommended distribution? Round the final answer to the <u>nearest tenth</u>.

C: If this sample of questions provides evidence that individual test writers do <u>not</u> follow the recommended overall answer distribution at the 1% level of significance, then let C = 1. If not, let C = 2.

D: One of the components (a.k.a. contributions) to the test statistic is an integer. What is that integer component?

Summary: Compute $\frac{ABC}{D}$ to the <u>nearest integer</u>.

Question # 8 Statistics Bowl MAΘ National Convention 2022

M² chocolate candies come in a variety of package sizes. One small bag is labeled as 1.69 oz which is equivalent to 47.91 g. A simple random sample of n bags is chosen and weighed and the sample mean weight is found to be $\bar{X} = 48.208\ g$ and the sample standard deviation is $s_x = 0.8825\ g$. Let μ represent the mean weight of all M² candy bags labeled as 47.91 g. After verifying all inference assumptions and conditions are met, a test of the hypotheses $H_0: \mu = 47.91\ g\ vs. H_A: \mu > 47.91\ g$ is conducted and the test statistic is calculated as $t = 1.3078$ when rounded. Answer each of the following:

A: How many degrees of freedom does the test statistic's distribution have? (Rounded to the nearest integer – obviously!)

B: What is the p-value of the test rounded to the thousandths place?

C: What is the lower bound of a 95% confidence interval for μ rounded to the thousandths place?

D: Based on the results of the hypothesis test and confidence interval, are you convinced that the mean weight of all M² candy bags labeled as 47.91 g is actually greater than 47.91 g at the 5% level of significance? If so, let D = 10; if not, let D = 0.

Summary: Using your final rounded values from above, compute A + B + C + D.

Question # 8 Statistics Bowl MAΘ National Convention 2022

M² chocolate candies come in a variety of package sizes. One small bag is labeled as 1.69 oz which is equivalent to 47.91 g. A simple random sample of n bags is chosen and weighed and the sample mean weight is found to be $\bar{X} = 48.208\ g$ and the sample standard deviation is $s_x = 0.8825\ g$. Let μ represent the mean weight of all M² candy bags labeled as 47.91 g. After verifying all inference assumptions and conditions are met, a test of the hypotheses $H_0: \mu = 47.91\ g\ vs. H_A: \mu > 47.91\ g$ is conducted and the test statistic is calculated as $t = 1.3078$ when rounded. Answer each of the following:

A: How many degrees of freedom does the test statistic's distribution have? (Rounded to the nearest integer – obviously!)

B: What is the p-value of the test rounded to the thousandths place?

C: What is the lower bound of a 95% confidence interval for μ rounded to the thousandths place?

D: Based on the results of the hypothesis test and confidence interval, are you convinced that the mean weight of all M² candy bags labeled as 47.91 g is actually greater than 47.91 g at the 5% level of significance? If so, let D = 10; if not, let D = 0.

Summary: Using your final rounded values from above, compute A + B + C + D.

Question # 9 Statistics Bowl MAΘ National Convention 2022

The color distribution of the aforementioned M^2 candies varies by factory with green allegedly having the largest difference in the proportions that appear in large quantities of candies from two particular factories: one in New Jersey and one in Tennessee. A random sample of 500 candies is chosen from each factory and 12.6% of the candies in the sample from the New Jersey factory are green while 19.6% of the candies in the sample from the Tennessee factory are green. Perform a two-tail two-proportion Z-test to determine if these data provide statistically significant evidence that the proportion of green M^2 candies produced by the two factories differs at the 1% level of significance. You may assume all inference assumptions and conditions are satisfied.

A: What is the absolute value of the test statistic rounded to four decimal places?

B: What is the two-tail p-value of the test rounded to four decimal places?

C: Is the null hypothesis of the test rejected at the 1% level of significance? If yes, let C = 1000; and if not, let C = 2000.

D: If green M^2 candies are your favorite, then would you prefer to get your M^2 candies from the NEW JERSEY factory or the TENNESSEE factory? Let D equal the number of unique permutations of the letters in the word(s) of the correct preferred choice and do NOT treat the space in New Jersey as a letter and make no distinction between lower case and capital letters! If it does not matter which factory you get your M^2 candies from, then let D = 9! (9 factorial).

Summary: Using your final rounded answers from above, calculate $C(A + B) + D$ to the nearest integer.

Question # 9 Statistics Bowl MAΘ National Convention 2022

The color distribution of the aforementioned M^2 candies varies by factory with green allegedly having the largest difference in the proportions that appear in large quantities of candies from two particular factories: one in New Jersey and one in Tennessee. A random sample of 500 candies is chosen from each factory and 12.6% of the candies in the sample from the New Jersey factory are green while 19.6% of the candies in the sample from the Tennessee factory are green. Perform a two-tail two-proportion Z-test to determine if these data provide statistically significant evidence that the proportion of green M^2 candies produced by the two factories differs at the 1% level of significance. You may assume all inference assumptions and conditions are satisfied.

A: What is the absolute value of the test statistic rounded to four decimal places?

B: What is the two-tail p-value of the test rounded to four decimal places?

C: Is the null hypothesis of the test rejected at the 1% level of significance? If yes, let C = 1000; and if not, let C = 2000.

D: If green M^2 candies are your favorite, then would you prefer to get your M^2 candies from the NEW JERSEY factory or the TENNESSEE factory? Let D equal the number of unique permutations of the letters in the word(s) of the correct preferred choice and do NOT treat the space in New Jersey as a letter and make no distinction between lower case and capital letters! If it does not matter which factory you get your M^2 candies from, then let D = 9! (9 factorial).

Summary: Using your final rounded answers from above, calculate $C(A + B) + D$ to the nearest integer.

Question # 10 Statistics Bowl — MAΘ National Convention 2022

The company who manufactures and distributes the M² candies claims that the value of 47.91 g on the packaging label is intended to be an approximate minimum for the weight of candies in a bag labeled as such and that the true distribution is approximately normal with a mean of $\mu = 49\ g$ and a standard deviation of $\sigma = 0.75\ g$. Additionally, they claim that less than 10% of all bags labeled as 47.91 g actually contain less than 47.91 g of candy. Well, the SRS of 15 bags labeled as 47.91 g used in the previous questions had exactly two out of the 15 that were less than 47.91 g. Let's investigate these two claims.

A: If the first claim is true that the 47.91 g bag distribution is approximately normal with a mean of $\mu = 49\ g$ and a standard deviation of $\sigma = 0.75\ g$**,** then 10% of bags have less than how many grams total weight (to the nearest whole gram)?

B: If exactly 10% of all bags labeled as 47.91 g actually contain less than that amount, then we would expect how many in a random sample of 15 independent bags to have less than 47.91 g of candy in them?

C: If exactly 10% of all bags labeled as 47.91 g actually contain less than that amount, then what is the probability that at most 2 bags in a random sample of 15 independent bags contain less than 47.91 g of candy. Rounded your final answer to the thousandths place?

D: If the first claim is true in that the true 47.91 g bag weight distribution is approximately normal with a mean of $\mu = 49\ g$ and a standard deviation of $\sigma = 0.75\ g$, then what proportion of all such bags will actually contain less than 47.91 g as a decimal rounded to the nearest thousandth? (For example 0.037 and not 3.700%.)

Summary: Using your final rounded values, compute $A + B + C + D$.

Question # 10 Statistics Bowl — MAΘ National Convention 2022

The company who manufactures and distributes the M² candies claims that the value of 47.91 g on the packaging label is intended to be an approximate minimum for the weight of candies in a bag labeled as such and that the true distribution is approximately normal with a mean of $\mu = 49\ g$ and a standard deviation of $\sigma = 0.75\ g$. Additionally, they claim that less than 10% of all bags labeled as 47.91 g actually contain less than 47.91 g of candy. Well, the SRS of 15 bags labeled as 47.91 g used in the previous questions had exactly two out of the 15 that were less than 47.91 g. Let's investigate these two claims.

A: If the first claim is true that the 47.91 g bag distribution is approximately normal with a mean of $\mu = 49\ g$ and a standard deviation of $\sigma = 0.75\ g$**,** then 10% of bags have less than how many grams total weight (to the nearest whole gram)?

B: If exactly 10% of all bags labeled as 47.91 g actually contain less than that amount, then we would expect how many in a random sample of 15 independent bags to have less than 47.91 g of candy in them?

C: If exactly 10% of all bags labeled as 47.91 g actually contain less than that amount, then what is the probability that at most 2 bags in a random sample of 15 independent bags contain less than 47.91 g of candy. Rounded your final answer to the thousandths place?

D: If the first claim is true in that the true 47.91 g bag weight distribution is approximately normal with a mean of $\mu = 49\ g$ and a standard deviation of $\sigma = 0.75\ g$, then what proportion of all such bags will actually contain less than 47.91 g as a decimal rounded to the nearest thousandth? (For example 0.037 and not 3.700%.)

Summary: Using your final rounded values, compute $A + B + C + D$.

Question # 11 Statistics Bowl — MAΘ National Convention 2022

Recall the two claims made by the company who manufactures and distributes the M^2 candies from the previous question:

Claim 1: The 47.91 g on the label is intended as an approximate minimum for the weight of candies in a bag labeled as such and that the true distribution is approximately normal with a mean of $\mu = 49\ g$ and a standard deviation of $\sigma = 0.75\ g$.

Claim 2: Less than 10% of all bags labeled as 47.91 g actually contain less than 47.91 g of candy.

Also, recall that the sample of 15 bags labeled as 47.91 g used in the previous questions had exactly two out of the 15 that were less than 47.91 g and that the sample mean of the 15 bags was $\bar{X} = 48.208\ g$. Let's investigate these claims further.

A: In Part D of the previous question, you found that approximately 7.3% of all bags labeled as 47.91 g should actually contain less than 47.91 g, if we assume Claim 1 is true. Supposing this exact 7.3% is actually true, then what is the probability that a random sample of 15 independent bags labeled as 47.91 g has more than 3 bags that actually contain less than 47.91 g rounded to four decimal places?

B: Assuming Claim 1 is true, what is the approximate probability that the sample mean weight of 15 randomly selected independent bags labeled as 47.91 g is between 48.5 g and 49.5 g rounded to four decimal places?

C: Assuming Claim 1 is true, does the sample mean of $\bar{X} = 48.208$ provide statistically significant evidence at the 1% level of significance that the population mean is actually less than what is claimed? If so, let C = the absolute value of the test statistic rounded to four decimal places; and if not, let C = 0. You may assume all inference assumptions and conditions are satisfied.

D: According to your results in the previous part, is Claim 1 plausible? If so, let D = 1; and if not, let D = 0.

Summary: Using your final rounded answers from above, compute A + B + C + D.

Question # 11 Statistics Bowl — MAΘ National Convention 2022

Recall the two claims made by the company who manufactures and distributes the M^2 candies from the previous question:

Claim 1: The 47.91 g on the label is intended as an approximate minimum for the weight of candies in a bag labeled as such and that the true distribution is approximately normal with a mean of $\mu = 49\ g$ and a standard deviation of $\sigma = 0.75\ g$.

Claim 2: Less than 10% of all bags labeled as 47.91 g actually contain less than 47.91 g of candy.

Also, recall that the sample of 15 bags labeled as 47.91 g used in the previous questions had exactly two out of the 15 that were less than 47.91 g and that the sample mean of the 15 bags was $\bar{X} = 48.208\ g$. Let's investigate these claims further.

A: In Part D of the previous question, you found that approximately 7.3% of all bags labeled as 47.91 g should actually contain less than 47.91 g, if we assume Claim 1 is true. Supposing this exact 7.3% is actually true, then what is the probability that a random sample of 15 independent bags labeled as 47.91 g has more than 3 bags that actually contain less than 47.91 g rounded to four decimal places?

B: Assuming Claim 1 is true, what is the approximate probability that the sample mean weight of 15 randomly selected independent bags labeled as 47.91 g is between 48.5 g and 49.5 g rounded to four decimal places?

C: Assuming Claim 1 is true, does the sample mean of $\bar{X} = 48.208$ provide statistically significant evidence at the 1% level of significance that the population mean is actually less than what is claimed? If so, let C = the absolute value of the test statistic rounded to four decimal places; and if not, let C = 0. You may assume all inference assumptions and conditions are satisfied.

D: According to your results in the previous part, is Claim 1 plausible? If so, let D = 1; and if not, let D = 0.

Summary: Using your final rounded answers from above, compute A + B + C + D.

Once again, here is Claim 2 made by the company who manufactures and distributes the M^2 candies:

Claim 2: Less than 10% of all bags labeled as 47.91 g actually contain less than 47.91 g of candy.

Also, recall that the sample of 15 bags labeled as 47.91 g used in the previous questions had exactly two out of the 15 that were less than 47.91 g and we calculated the probability that at most 2 bags in a random sample of 15 independent bags labeled 47.91 g actually containing less than 47.91 g under the assumption Claim 2 is true is 0.816. Answer the following:

A: If we assume Claim 2 is true and we end up with at least ______ of these bags (to the nearest whole bag) containing less than 47.91 g in a random sample of 15 independent bags, it is considered rather unusual because it is more than 2 standard deviations above the expected value for such a sample. What is the minimum possible integer answer to fill in the blank?

B: If we assume Claim 2 is true, then we expect (on average) that we need to randomly select at least how many bags (to the nearest whole bag) labeled 47.91 g until we obtain the first one that actually contains less than 47.91 g?

C: Suppose we want to estimate the true proportion of M^2 candy bags labeled as 47.91 g that actually contain less than 47.91 grams with a 95% confidence interval and we use 10% as our estimate for this true proportion and we want a margin of error of 2%. What is the minimum sample size required for this confidence interval? Remember to always round this up!

D: Suppose we select an SRS of 900 independent bags labeled as 47.91 g and we find that 117 in our sample actually contain less than 47.91 g. Use this to construct a 95% confidence interval for the true proportion of all such bags that actually contain less than 47.91 g. Based on the interval, can you reject Claim 2? If so, let D = the margin of error of the confidence interval rounded to the thousandths place; and if not, let D = the width of the confidence interval rounded to the thousandths place.

Summary: Using your final rounded values from above, compute A + B + C + D.

Question # 12 Statistics Bowl MAΘ National Convention 2022

Once again, here is Claim 2 made by the company who manufactures and distributes the M^2 candies:

Claim 2: Less than 10% of all bags labeled as 47.91 g actually contain less than 47.91 g of candy.

Also, recall that the sample of 15 bags labeled as 47.91 g used in the previous questions had exactly two out of the 15 that were less than 47.91 g and we calculated the probability that at most 2 bags in a random sample of 15 independent bags labeled 47.91 g actually containing less than 47.91 g under the assumption Claim 2 is true is 0.816. Answer the following:

A: If we assume Claim 2 is true and we end up with at least ______ of these bags (to the nearest whole bag) containing less than 47.91 g in a random sample of 15 independent bags, it is considered rather unusual because it is more than 2 standard deviations above the expected value for such a sample. What is the minimum possible integer answer to fill in the blank?

B: If we assume Claim 2 is true, then we expect (on average) that we need to randomly select at least how many bags (to the nearest whole bag) labeled 47.91 g until we obtain the first one that actually contains less than 47.91 g?

C: Suppose we want to estimate the true proportion of M^2 candy bags labeled as 47.91 g that actually contain less than 47.91 grams with a 95% confidence interval and we use 10% as our estimate for this true proportion and we want a margin of error of 2%. What is the minimum sample size required for this confidence interval? Remember to always round this up!

D: Suppose we select an SRS of 900 independent bags labeled as 47.91 g and we find that 117 in our sample actually contain less than 47.91 g. Use this to construct a 95% confidence interval for the true proportion of all such bags that actually contain less than 47.91 g. Based on the interval, can you reject Claim 2? If so, let D = the margin of error of the confidence interval rounded to the thousandths place; and if not, let D = the width of the confidence interval rounded to the thousandths place.

Summary: Using your final rounded values from above, compute A + B + C + D.

Question # 13 Statistics Bowl MAΘ National Convention 2022

Continuous random variable *X* is defined on all positive real numbers within the closed interval $[a, b]$. It has a probability density function (or pdf) of $f(x) = 2x^{-2}$ and cumulative distribution function (or cdf) of $F(x) = 2 - 2x^{-1}$. Find the values of a and b that make $f(x)$ a valid pdf then answer each of the following providing all answers as exact values:

A: What is a + b?

B: What is $P\left(\frac{6}{5} \leq X < \frac{9}{5}\right)$?

C: What is the median of random variable *X*?

D: Is the mean of random variable *X* closer to a or closer to b? Let D equal the numerical value of either a or b accordingly.

Summary: Compute the product $ABCD$ as an exact value.

Question # 13 Statistics Bowl MAΘ National Convention 2022

Continuous random variable *X* is defined on all positive real numbers within the closed interval $[a, b]$. It has a probability density function (or pdf) of $f(x) = 2x^{-2}$ and cumulative distribution function (or cdf) of $F(x) = 2 - 2x^{-1}$. Find the values of a and b that make $f(x)$ a valid pdf then answer each of the following providing all answers as exact values:

A: What is a + b?

B: What is $P\left(\frac{6}{5} \leq X < \frac{9}{5}\right)$?

C: What is the median of random variable *X*?

D: Is the mean of random variable *X* closer to a or closer to b? Let D equal the numerical value of either a or b accordingly.

Summary: Compute the product $ABCD$ as an exact value.

Three events X, Y, and Z are all independent of each other such that P(X) = 2P(Y), P(Y) = 2P(Z), and P(X) + P(Y) + P(Z) = $\frac{7}{8}$. Compute each of the following as a simplified fraction.

A: P(X and Y and Z)

B: P(X^C and Y^C and Z^C)

C: P(X or Y or Z)

D: P(X^C or Y^C or Z^C)

Summary: Compute A + B + C + D as an exact value.

Question # 14 Statistics Bowl MAΘ National Convention 2022

Three events X, Y, and Z are all independent of each other such that P(X) = 2P(Y), P(Y) = 2P(Z), and P(X) + P(Y) + P(Z) = $\frac{7}{8}$. Compute each of the following as a simplified fraction.

A: P(X and Y and Z)

B: P(X^C and Y^C and Z^C)

C: P(X or Y or Z)

D: P(X^C or Y^C or Z^C)

Summary: Compute A + B + C + D as an exact value.