BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #1.

Consider the following numbered list of terms and provide the number of the term for each of parts A, B, C, and D. For example, if the description given in part A corresponds to the term "Bar Graph," then the answer to part A is 1.

1. Bar Graph
2. Stacked (or Segmented) Bar Graph
3. Pie Chart
4. Dotplot
5. Stemplot
6. Histogram
7. Boxplot (or Modified Boxplot)
8. Scatterplot
9. Residual Plot

A: This type of graph is used to display the conditional distribution of a categorical response variable for each value (or level) of a categorical explanatory variable.

B: This graphical display of the distribution of a quantitative dataset is useful for identifying outliers (according to the 1.5IQR rule); however, it is only minimally informative with respect to assessing the symmetry versus the skewness of the distribution of a quantitative dataset, nor does it provide any information as to the location of the peak (or peaks) of the distribution (if there are any) and only shows gaps in the data between outliers and non-outliers but not elsewhere within the dataset.

C: When created with an appropriate scale and sufficient detail (such as a key), two of the graphical displays of the distribution of a quantitative dataset listed above can enable us to extract the raw data values from them. What is the product of the two corresponding numbers?

D: What is the sum of the two numbers corresponding to the two graphical displays of the relationship between a quantitative explanatory variable and a quantitative response variable in a bivariate dataset?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #1.

Consider the following numbered list of terms and provide the number of the term for each of parts A, B, C, and D. For example, if the description given in part A corresponds to the term "Bar Graph," then the answer to part A is 1.

1. Bar Graph
2. Stacked (or Segmented) Bar Graph
3. Pie Chart
4. Dotplot
5. Stemplot
6. Histogram
7. Boxplot (or Modified Boxplot)
8. Scatterplot
9. Residual Plot

A: This type of graph is used to display the conditional distribution of a categorical response variable for each value (or level) of a categorical explanatory variable.

B: This graphical display of the distribution of a quantitative dataset is useful for identifying outliers (according to the 1.5IQR rule); however, it is only minimally informative with respect to assessing the symmetry versus the skewness of the distribution of a quantitative dataset, nor does it provide any information as to the location of the peak (or peaks) of the distribution (if there are any) and only shows gaps in the data between outliers and non-outliers but not elsewhere within the dataset.

C: When created with an appropriate scale and sufficient detail (such as a key), two of the graphical displays of the distribution of a quantitative dataset listed above can enable us to extract the raw data values from them. What is the product of the two corresponding numbers?

D: What is the sum of the two numbers corresponding to the two graphical displays of the relationship between a quantitative explanatory variable and a quantitative response variable in a bivariate dataset?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #2.

Mr. G. does not "curve" his AP Statistics students' test scores per se. Instead, he applies a "linear transformation" to the set of raw scores to obtain what he likes to call the "$T - Score$" (which is short for "transformed score") for each student in the class before he enters them as a grade in his gradebook. Thus, he uses a linear function of the following form:

$$T - Score = a + b(Raw\ Score)$$

Suppose that the constant term in the transformation function above is 20 and that the ratio of the standard deviation of the transformed scores to that of the raw scores is $\frac{s_{T-Score}}{s_{Raw\ Score}} = \frac{4}{5}$. Answer each of the following.

A: What is the value of the scaling coefficient, b, in the linear transformation defined above as an <u>exact value</u>?

B: If the mean of the set of transformed scores ($T - Scores$) is 84, what is the mean of the set of raw scores as an <u>exact value</u>?

C: If Jeff earned a raw score of 90, what is his transformed score ($T - Score$) as an <u>exact value</u>?

D: What is the <u>exact value</u> of the correlation coefficient between the set of raw scores and the set of transformed scores when they are treated as a set of bivariate data?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #2.

Mr. G. does not "curve" his AP Statistics students' test scores per se. Instead, he applies a "linear transformation" to the set of raw scores to obtain what he likes to call the "$T - Score$" (which is short for "transformed score") for each student in the class before he enters them as a grade in his gradebook. Thus, he uses a linear function of the following form:

$$T - Score = a + b(Raw\ Score)$$

Suppose that the constant term in the transformation function above is 20 and that the ratio of the standard deviation of the transformed scores to that of the raw scores is $\frac{s_{T-Score}}{s_{Raw\ Score}} = \frac{4}{5}$. Answer each of the following.

A: What is the value of the scaling coefficient, b, in the linear transformation defined above as an <u>exact value</u>?

B: If the mean of the set of transformed scores ($T - Scores$) is 84, what is the mean of the set of raw scores as an <u>exact value</u>?

C: If Jeff earned a raw score of 90, what is his transformed score ($T - Score$) as an <u>exact value</u>?

D: What is the <u>exact value</u> of the correlation coefficient between the set of raw scores and the set of transformed scores when they are treated as a set of bivariate data?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #3.

Consider the following set of bivariate data:

X	−2	−1	0	1	2
Y	4	1	0	1	4

Compute each of the requested values below and keep them handy because you may need them for the next question as well!

A: What is the sample mean, $\bar{X}$, of the x-coordinates of this bivariate dataset as an exact value?

B: What is the sample standard deviation, s_x, of the x-coordinates of this bivariate dataset as an exact value?

C: What is the sample mean, $\bar{Y}$, of the y-coordinates of this bivariate dataset as an exact value?

D: What is the sample variance, s_y^2, of the y-coordinates of this bivariate dataset as an exact value?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #3.

Consider the following set of bivariate data:

X	−2	−1	0	1	2
Y	4	1	0	1	4

Compute each of the requested values below and keep them handy because you may need them for the next question as well!

A: What is the sample mean, $\bar{X}$, of the x-coordinates of this bivariate dataset as an exact value?

B: What is the sample standard deviation, s_x, of the x-coordinates of this bivariate dataset as an exact value?

C: What is the sample mean, $\bar{Y}$, of the y-coordinates of this bivariate dataset as an exact value?

D: What is the sample variance, s_y^2, of the y-coordinates of this bivariate dataset as an exact value?

<u>BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #4.</u>

Again, consider the following set of bivariate data from the previous question:

X	–2	–1	0	1	2
Y	4	1	0	1	4

Compute each of the requested values below treating *X* as the explanatory variable and *Y* as the response variable; and as a hint to save a <u>ton</u> of time and trouble, plot the points and think conceptually!

A: What is the <u>exact value</u> of the coefficient of linear correlation between *X* and *Y* for this bivariate dataset?

B: According the least-squares linear regression model equation fit to this bivariate dataset, what is the sum of the slope and the *y*-intercept as an <u>exact value</u>?

C: What is the estimated or predicted mean value of *Y* when *X* is equal to 2?

D: What is the sum of the <u>squared</u> residuals in this least-squares linear regression analysis as an <u>exact value</u>?

<u>BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #4.</u>

Again, consider the following set of bivariate data from the previous question:

X	–2	–1	0	1	2
Y	4	1	0	1	4

Compute each of the requested values below treating *X* as the explanatory variable and *Y* as the response variable; and as a hint to save a <u>ton</u> of time and trouble, plot the points and think conceptually!

A: What is the <u>exact value</u> of the coefficient of linear correlation between *X* and *Y* for this bivariate dataset?

B: According the least-squares linear regression model equation fit to this bivariate dataset, what is the sum of the slope and the *y*-intercept as an <u>exact value</u>?

C: What is the estimated or predicted mean value of *Y* when *X* is equal to 2?

D: What is the sum of the <u>squared</u> residuals in this least-squares linear regression analysis as an <u>exact value</u>?

BC/AHS-PB January Invitational   Statistics Team (No Calculator)   Question #5.

The extracts of avocado oil and soybean oil have been shown to slow cell inflammation in test tubes. Thus, some medical researchers wonder if taking either avocado oil and/or soybean oil extract supplements (alone or in some combination of the two) will help relieve pain for subjects with joint stiffness due to arthritis. So, they conduct a double-blind, placebo-controlled, randomized block-design experiment upon a large set of volunteer subjects - all of whom suffer from joint stiffness due to arthritis. Since pain from joint stuffiness due to arthritis varies differently in men versus women as well as for different age groups, the researchers decide to use gender (male versus female) and age group (under 50 years old, between 50 and 70 years old, and over 70 years old), as blocking factors. Answer each of the following.

A: How many unique blocks (gender and age group combinations) are there in this experiment?

B: If both the avocado oil extract supplement and the soybean oil extract supplement are each administered in either a placebo dose, a 500 mg dose, or a 1000 mg dose; then, how many total unique treatments are there in this experiment (both inert and active) within each unique block?

C: Suppose 2700 volunteer subjects are available for this experiment and an equal number of subjects are available within each unique block; then, what is the expected value of the number of subjects who will get randomly assigned to each unique treatment within each unique block?

D: How many possible pair-wise comparisons between two distinct treatment groups are possible within each unique block in this experiment?

BC/AHS-PB January Invitational   Statistics Team (No Calculator)   Question #5.

The extracts of avocado oil and soybean oil have been shown to slow cell inflammation in test tubes. Thus, some medical researchers wonder if taking either avocado oil and/or soybean oil extract supplements (alone or in some combination of the two) will help relieve pain for subjects with joint stiffness due to arthritis. So, they conduct a double-blind, placebo-controlled, randomized block-design experiment upon a large set of volunteer subjects - all of whom suffer from joint stiffness due to arthritis. Since pain from joint stuffiness due to arthritis varies differently in men versus women as well as for different age groups, the researchers decide to use gender (male versus female) and age group (under 50 years old, between 50 and 70 years old, and over 70 years old), as blocking factors. Answer each of the following.

A: How many unique blocks (gender and age group combinations) are there in this experiment?

B: If both the avocado oil extract supplement and the soybean oil extract supplement are each administered in either a placebo dose, a 500 mg dose, or a 1000 mg dose; then, how many total unique treatments are there in this experiment (both inert and active) within each unique block?

C: Suppose 2700 volunteer subjects are available for this experiment and an equal number of subjects are available within each unique block; then, what is the expected value of the number of subjects who will get randomly assigned to each unique treatment within each unique block?

D: How many possible pair-wise comparisons between two distinct treatment groups are possible within each unique block in this experiment?

BC/AHS-PB January Invitational   Statistics Team (No Calculator)   Question #6.

At Broward College, suppose 40% of the students are Business majors, 30% are Education majors, 20% are Health Sciences majors, and 10% are Public Safety majors. Also, suppose that 75% of the Business majors are in the four-year program while the rest are in the two-year program, two-thirds of the Education majors are in the four-year program while the rest are in the two-year program, and half of the Health Science majors as well as half of the Public Safety majors are in the four-year program while the rest are in the two-year program. Answer each of the following expressing all final answers as exact values in the form of a simplified fraction.

A: What is the probability that a randomly selected Broward College student is in a two-year program?

B: What is the probability that a randomly selected Broward College student is either an Education major or in a four-year program?

C: What is the probability that a randomly selected Broward College student is a Business major and in a two-year program?

D: What is the probability that a randomly selected Broward College student is either a Health Sciences major or a Public Safety major given that he or she is in a four-year program?

BC/AHS-PB January Invitational   Statistics Team (No Calculator)   Question #6.

At Broward College, suppose 40% of the students are Business majors, 30% are Education majors, 20% are Health Sciences majors, and 10% are Public Safety majors. Also, suppose that 75% of the Business majors are in the four-year program while the rest are in the two-year program, two-thirds of the Education majors are in the four-year program while the rest are in the two-year program, and half of the Health Science majors as well as half of the Public Safety majors are in the four-year program while the rest are in the two-year program. Answer each of the following expressing all final answers as exact values in the form of a simplified fraction.

A: What is the probability that a randomly selected Broward College student is in a two-year program?

B: What is the probability that a randomly selected Broward College student is either an Education major or in a four-year program?

C: What is the probability that a randomly selected Broward College student is a Business major and in a two-year program?

D: What is the probability that a randomly selected Broward College student is either a Health Sciences major or a Public Safety major given that he or she is in a four-year program?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #7.

The following two-way contingency table summarizes the given information from the scenario in the previous question in terms of proportions (or probabilities) in decimal form:

	B = Business	E = Education	H = Health Sciences	P = Public Safety	Total
4-yr	0.30	0.20	0.10	0.05	0.65
2-yr	0.10	0.10	0.10	0.05	0.35
Total	0.40	0.30	0.20	0.10	1.00

Suppose we select a simple random sample (SRS) of 5 independent Broward College students with replacement. Compute each of the following giving all final answers as exact values in the form of a simplified fraction or a rationalized radical where applicable.

A: Let random variable X represent the number of Education majors in our SRS of 5 students. What is $E(X)$?

B: Let random variable Y represent the number of Health Sciences majors in our SRS of 5 students. What is the standard deviation of random variable Y?

C: What is the probability that we obtain exactly twice the expected value of the number of Health Science majors in our SRS of 5 students?

D: What is the probability that we obtain exactly 2 Business majors, 1 Education major, 1 Health Sciences major, and 1 Public Safety major is our SRS of 5 students?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #7.

The following two-way contingency table summarizes the given information from the scenario in the previous question in terms of proportions (or probabilities) in decimal form:

	B = Business	E = Education	H = Health Sciences	P = Public Safety	Total
4-yr	0.30	0.20	0.10	0.05	0.65
2-yr	0.10	0.10	0.10	0.05	0.35
Total	0.40	0.30	0.20	0.10	1.00

Suppose we select a simple random sample (SRS) of 5 independent Broward College students with replacement. Compute each of the following giving all final answers as exact values in the form of a simplified fraction or a rationalized radical where applicable.

A: Let random variable X represent the number of Education majors in our SRS of 5 students. What is $E(X)$?

B: Let random variable Y represent the number of Health Sciences majors in our SRS of 5 students. What is the standard deviation of random variable Y?

C: What is the probability that we obtain exactly twice the expected value of the number of Health Science majors in our SRS of 5 students?

D: What is the probability that we obtain exactly 2 Business majors, 1 Education major, 1 Health Sciences major, and 1 Public Safety major is our SRS of 5 students?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #8.

The two-way contingency table below displays the joint probability distribution of random variables X and Y along with each of their respective marginal distributions in the rightmost column and the bottom row labeled as "Total." Compute each of the requested items as exact values.

	Random Variable Y				
		$y = 2$	$y = 4$	$y = 9$	Total
Random Variable X	$x = 1$	0.2	0.1	0.1	0.4
	$x = 6$	0.3	0.2	0.1	0.6
	Total	0.5	0.3	0.2	1

A: What is $E(X^2) - [E(X)]^2$?

B: What is $E(Y^2) - [E(Y)]^2$?

C: Given that $Var(X + Y) = Var(X) + Var(Y) + 2 \times Cov(X, Y) = 12$, what is $Cov(X, Y)$?
NOTE: $Var(X)$ and $Var(X)$ represent the variance of X and Y, respectively, and $Cov(X, Y)$ represents the covariance between X and Y which is a measure of the linear association between the two random variables.

D: Are random variables X and Y independent or not? If they are independent, then let **D** equal the number of unique permutations of the letters in the word INDEPENDENT divided by the number of unique permutations of the letters in the word DEPENDENT. If they are not independent, then let **D** equal the number of unique permutations of the letters in the word DEPENDENT divided by the number of unique permutations of the letters in the word INDEPENDENT.

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #8.

The two-way contingency table below displays the joint probability distribution of random variables X and Y along with each of their respective marginal distributions in the rightmost column and the bottom row labeled as "Total." Compute each of the requested items as exact values.

	Random Variable Y				
		$y = 2$	$y = 4$	$y = 9$	Total
Random Variable X	$x = 1$	0.2	0.1	0.1	0.4
	$x = 6$	0.3	0.2	0.1	0.6
	Total	0.5	0.3	0.2	1

A: What is $E(X^2) - [E(X)]^2$?

B: What is $E(Y^2) - [E(Y)]^2$?

C: Given that $Var(X + Y) = Var(X) + Var(Y) + 2 \times Cov(X, Y) = 12$, what is $Cov(X, Y)$?
NOTE: $Var(X)$ and $Var(X)$ represent the variance of X and Y, respectively, and $Cov(X, Y)$ represents the covariance between X and Y which is a measure of the linear association between the two random variables.

D: Are random variables X and Y independent or not? If they are independent, then let **D** equal the number of unique permutations of the letters in the word INDEPENDENT divided by the number of unique permutations of the letters in the word DEPENDENT. If they are not independent, then let **D** equal the number of unique permutations of the letters in the word DEPENDENT divided by the number of unique permutations of the letters in the word INDEPENDENT.

<u>BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #9.</u>

Suppose <u>independent</u> random variables X and Y have the following parameters:

$$\mu_X = 4 \qquad \sigma_X = 4 \qquad \mu_Y = 5 \qquad \sigma_Y = 3$$

Compute each of the following as <u>exact values</u>:

A: $\mu_{\frac{1}{2}X+2Y}$

B: $\sigma_{\frac{1}{2}X+2Y}$

C: μ_{3X-4Y}

D: σ_{3X-4Y}

<u>BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #9.</u>

Suppose <u>independent</u> random variables X and Y have the following parameters:

$$\mu_X = 4 \qquad \sigma_X = 4 \qquad \mu_Y = 5 \qquad \sigma_Y = 3$$

Compute each of the following as <u>exact values</u>:

A: $\mu_{\frac{1}{2}X+2Y}$

B: $\sigma_{\frac{1}{2}X+2Y}$

C: μ_{3X-4Y}

D: σ_{3X-4Y}

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #10.

Suppose independent random variables X and Y from the previous question still have the same following parameters and they are both normally distributed:

$$\mu_X = 4 \qquad \sigma_X = 4 \qquad \mu_Y = 5 \qquad \sigma_Y = 3$$

Use the Empirical Rule to compute each of the following as exact values in decimal form. For example, use 0.95 to represent "95%."

A: $P(\mu_X \leq X \leq 12)$

B: $P(-1 \leq Y \leq 8)$

C: $P(X \leq 0 \cup X \geq 12)$

D: $P[(X \leq 0 \cup X \geq 8) \cup (Y \leq -1 \cup Y \geq 11)]$

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #10.

Suppose independent random variables X and Y from the previous question still have the same following parameters and they are both normally distributed:

$$\mu_X = 4 \qquad \sigma_X = 4 \qquad \mu_Y = 5 \qquad \sigma_Y = 3$$

Use the Empirical Rule to compute each of the following as exact values in decimal form. For example, use 0.95 to represent "95%."

A: $P(\mu_X \leq X \leq 12)$

B: $P(-1 \leq Y \leq 8)$

C: $P(X \leq 0 \cup X \geq 12)$

D: $P[(X \leq 0 \cup X \geq 8) \cup (Y \leq -1 \cup Y \geq 11)]$

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #11.

Once again, suppose independent random variables X and Y from the previous question still have the same following parameters and they are both normally distributed:

$$\mu_X = 4 \qquad \sigma_X = 4 \qquad \mu_Y = 5 \qquad \sigma_Y = 3$$

Let $\bar{X}$ represent the sample mean of an SRS of size $n = 16$ independent observations from random variable X and let $\bar{Y}$ represent the sample mean of an SRS of size $n = 25$ independent observations from random variable Y.

A: What is $E(8\bar{X} - 10\bar{Y})$ as an exact value?

B: What is $SD(8\bar{X} - 10\bar{Y})$ as an exact value? NOTE: "SD" stands for the "Standard Deviation" function or operator.

C: Using the Empirical Rule, what is $P[(8\bar{X} \leq 40) \mid (10\bar{Y} \geq 20)]$ as an exact value in decimal form? For example, 0.95.

D: Using the Empirical Rule, what is $P\left(\bar{X} \leq \frac{5\bar{Y} + 1}{4}\right)$ as an exact value in decimal form? For example, 0.95.

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #11.

Once again, suppose independent random variables X and Y from the previous question still have the same following parameters and they are both normally distributed:

$$\mu_X = 4 \qquad \sigma_X = 4 \qquad \mu_Y = 5 \qquad \sigma_Y = 3$$

Let $\bar{X}$ represent the sample mean of an SRS of size $n = 16$ independent observations from random variable X and let $\bar{Y}$ represent the sample mean of an SRS of size $n = 25$ independent observations from random variable Y.

A: What is $E(8\bar{X} - 10\bar{Y})$ as an exact value?

B: What is $SD(8\bar{X} - 10\bar{Y})$ as an exact value? NOTE: "SD" stands for the "Standard Deviation" function or operator.

C: Using the Empirical Rule, what is $P[(8\bar{X} \leq 40) \mid (10\bar{Y} \geq 20)]$ as an exact value in decimal form? For example, 0.95.

D: Using the Empirical Rule, what is $P\left(\bar{X} \leq \frac{5\bar{Y} + 1}{4}\right)$ as an exact value in decimal form? For example, 0.95.

<u>BC/AHS-PB January Invitational   Statistics Team (No Calculator)   Question #12.</u>

Answer each of the following:

A: The sum of the residuals in any least-squares liner regression analysis is always equal to what number?

B: If we use the sample mean of a dataset and the sample standard deviation of the same dataset to convert each of the data values within the dataset into its corresponding z-score, then the sum of the mean and the standard deviation of this set of z-scores is always what number?

C: What is the standard deviation of the sampling distribution of the sample proportion with respect to the proportion of heads observed in 400 flips of a fair coin? Express your answer as an <u>exact value</u>.

D: Using the <u>Empirical Rule</u> and the normal approximation to the binomial distribution, what is the approximate probability that the number of heads observed in 400 flips of a fair coin is between 210 and 220, inclusive? Express your final answer in the form of a decimal; for example, 0.95.

<u>BC/AHS-PB January Invitational   Statistics Team (No Calculator)   Question #12.</u>

Answer each of the following:

A: The sum of the residuals in any least-squares liner regression analysis is always equal to what number?

B: If we use the sample mean of a dataset and the sample standard deviation of the same dataset to convert each of the data values within the dataset into its corresponding z-score, then the sum of the mean and the standard deviation of this set of z-scores is always what number?

C: What is the standard deviation of the sampling distribution of the sample proportion with respect to the proportion of heads observed in 400 flips of a fair coin? Express your answer as an <u>exact value</u>.

D: Using the <u>Empirical Rule</u> and the normal approximation to the binomial distribution, what is the approximate probability that the number of heads observed in 400 flips of a fair coin is between 210 and 220, inclusive? Express your final answer in the form of a decimal; for example, 0.95.

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #13.

Answer each of the following:

A: Suppose the standard deviation of a geometric random variable, X, is $\sigma = 2\sqrt{14}$. What is the mean of X?

B: Suppose the standard deviation of a binomial random variable with $n = 24$ trials is half the size of its mean. What is the probability of success, p, for this binomial random variable?

C: Suppose that, according to the Empirical Rule, approximately 95% of all sample means for independent simple random samples of size $n = 50$ from some population are between 16 and 24. What is the sum of the mean and the standard deviation of the population in this scenario as an exact value?

D: What is the probability that the mean of two randomly selected odd integers is also an integer?

BC/AHS-PB January Invitational Statistics Team (No Calculator) Question #13.

Answer each of the following:

A: Suppose the standard deviation of a geometric random variable, X, is $\sigma = 2\sqrt{14}$. What is the mean of X?

B: Suppose the standard deviation of a binomial random variable with $n = 24$ trials is half the size of its mean. What is the probability of success, p, for this binomial random variable?

C: Suppose that, according to the Empirical Rule, approximately 95% of all sample means for independent simple random samples of size $n = 50$ from some population are between 16 and 24. What is the sum of the mean and the standard deviation of the population in this scenario as an exact value?

D: What is the probability that the mean of two randomly selected odd integers is also an integer?