Answers:

	A	B	C	D
1.	A: 2	B: 7	C: 20	D: 17
2.	A: $\frac{4}{5}$ *or* 0.8	B: 80	C: 92	D: 1
3.	A: 0	B: $\frac{\sqrt{10}}{2}$	C: 2	D: $\frac{7}{2}$ *or* 3.5
4.	A: 0	B: 2	C: 2	D: 14
5.	A: 6	B: 9	C: 50	D: 36
6.	A: $\frac{7}{20}$	B: $\frac{3}{4}$	C: $\frac{1}{10}$	D: $\frac{3}{13}$

NOTE: All answers must be in the form of a fraction – do not accept decimals!

	A	B	C	D
7.	A: $\frac{3}{2}$	B: $\frac{2\sqrt{5}}{5}$	C: $\frac{128}{625}$	D: $\frac{36}{625}$
8.	A: 6	B: 7	C: $-\frac{1}{2}$ *or* -0.5	D: $\frac{3}{110}$
9.	A: 12	B: $2\sqrt{10}$	C: -8	D: $12\sqrt{2}$
10.	A: 0.475	B: 0.815	C: 0.185	D: 0.354
11.	A: -18	B: 10	C: 0.84	D: 0.975
12.	A: 0	B: 1	C: $\frac{1}{40}$ *or* 0.025	D: 0.135
13.	A: 8	B: $\frac{1}{7}$	C: $20 + 10\sqrt{2}$	D: 1

Solutions:

1. A: 2 B: 7 C: 20 D: 17

A: A stacked (or segmented) bar graph is used to display the conditional distribution of a categorical response variable for each value (or level) of a categorical explanatory variable. Thus, A = 2.

B: Boxplots (or modified boxplots) are graphical displays of the distribution of a quantitative dataset that are useful for identifying outliers (according to the 1.5IQR rule); however, they are only minimally informative with respect to assessing the symmetry versus the skewness of the distribution of a quantitative dataset, nor do they provide any information as to the location of the peak (or peaks) of the distribution and they only show gaps in the data between outliers and non-outliers but not elsewhere within the dataset. Thus, B = 7.

C: When a dotplot or a stemplot are created with an appropriate scale and sufficient detail (such as a key for a stemplot), then we can extract the raw data values from them. Thus, C = 4(5) = 20.

D: The only two graphical displays of the relationship between a quantitative explanatory variable and a quantitative response variable in a bivariate dataset listed are a scatterplot and a residual plot.
Thus, D = 8 + 9 = 17.

2. A: $\frac{4}{5}$ *or* 0.8 B: 80 C: 92 D: 1

A: The scaling coefficient of a linear transformation from one quantitative dataset into another is equivalent to the slope of the linear transformation which is equivalent to the ratio of the standard deviation of the original dataset to that of the transformed dataset. Thus, $A = b = \frac{s_{T-Score}}{s_{Raw\ Score}} = \frac{4}{5}$ *or* 0.8.

B: Since $\overline{T - Score} = a + b(\overline{Raw\ Score})$ by the properties of linear transformations when applied to measures of location within a quantitative dataset, we have that $\overline{Raw\ Score} = \frac{\overline{T - Score} - a}{b} = \frac{84-20}{4/5} = 80$.

C: Jeff's $T - Score$ is $T - Score = 20 + \frac{4}{5}(90) = 92$.

D: Since the set of $T - Scores$ is a direct linear transformation of the set of raw scores, the exact value of the correlation coefficient between the two sets of data when treated as a bivariate dataset is 1.

3. A: 0 B: $\frac{\sqrt{10}}{2}$ C: 2 D: $\frac{7}{2}$ *or* 3.5

A: $\bar{X} = \frac{-2-1+0+1+2}{5} = \frac{0}{5} = 0$

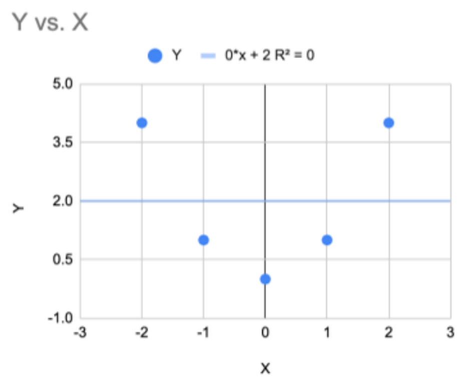
B: $s_x = \sqrt{\frac{\sum(x-\bar{X})^2}{n-1}} = \sqrt{\frac{(-2-0)^2 + (-1-0)^2 + (0-0)^2 + (1-0)^2 + (2-0)^2}{5-1}} = \sqrt{\frac{10}{4}} = \frac{\sqrt{10}}{2}$.

C: $\bar{Y} = \frac{4+1+0+1+4}{5} = \frac{10}{5} = 2$

D: $s_y^2 = \frac{\sum(y-\bar{Y})^2}{n-1} = \frac{(4-2)^2+(1-2)^2+(0-2)^2+(1-2)^2+(4-2)^2}{5-1} = \frac{14}{4} = \frac{7}{2}$ *or* 3.5.

4. A: 0 B: 2 C: 2 D: 14

A: Thinking conceptually and noting that each y-coordinate is just the square of each corresponding x-coordinate and plotting the points will lead us to the realization that the coefficient of linear correlation between X and Y within this bivariate dataset is $r = 0$. Here are the scatterplot and the details of the calculations for verification:



Using the sample means and the sample standard deviations calculated in the previous question, we can set up a table to help us verify the calculation of the coefficient of linear correlation as follows:

X	Y	Z_X	Z_Y	$(Z_X)(Z_Y)$
-2	4	$\frac{-2-0}{\sqrt{2.5}}$	$\frac{4-2}{\sqrt{3.5}}$	$\frac{-4}{\sqrt{8.75}}$
-1	1	$\frac{-1-0}{\sqrt{2.5}}$	$\frac{1-2}{\sqrt{3.5}}$	$\frac{1}{\sqrt{8.75}}$
0	0	$\frac{0-0}{\sqrt{2.5}}$	$\frac{0-2}{\sqrt{3.5}}$	$\frac{0}{\sqrt{8.75}}$
1	1	$\frac{1-0}{\sqrt{2.5}}$	$\frac{1-2}{\sqrt{3.5}}$	$\frac{-1}{\sqrt{8.75}}$
2	4	$\frac{2-0}{\sqrt{2.5}}$	$\frac{4-2}{\sqrt{3.5}}$	$\frac{4}{\sqrt{8.75}}$
$\bar{X} = 0$ $s_x = \sqrt{2.5}$	$\bar{Y} = 2$ $s_y = \sqrt{3.5}$			$r = \frac{\sum(Z_X)(Z_Y)}{n-1} = \frac{0}{4} = 0$

B: Since the coefficient of linear correlation between X and Y within this bivariate dataset is $r = 0$, then the least-squares linear regression model equation fit to this bivariate dataset is $\hat{y} = a + bx = 2 + 0x = 2$ which has a slope of 0 and a y-intercept of 2; and so, their sum is 0 + 2 = 2.

C: Once again, since the coefficient of linear correlation between X and Y within this bivariate dataset is 0, then the least-squares linear regression model equation fit to this bivariate dataset is $\hat{y} = \bar{Y} = 2$. Thus, the estimated or predicted mean value of Y when X is equal to 2 is $\hat{y} = \bar{Y} = 2$.

D: Since the explanatory variable, X, is of no use in helping us linearly predict the response variable, Y, then the sum of the squared residuals is equal to $\sum(y - \bar{Y})^2 = (n-1)s_y^2 = 4\left(\frac{14}{4}\right) = 14$.

5. A: 6 B: 9 C: 50 D: 36

A: Since there are two blocking factors, gender and age group, with 2 levels and 3 levels each, respectively; there is a total of $2 \times 3 = 6$ blocks.

B: Since each experimental factor, avocado oil extract supplement and soybean oil extract supplement, has 3 levels; there are $3 \times 3 = 9$ total unique treatments within each of the 6 unique blocks.

C: Taking the 2700 available volunteer subjects and dividing by the 6 unique blocks gives us $\frac{2700}{6} = 450$ subjects per block. Then, dividing 450 by the 9 unique treatments gives us an expected value of $\frac{450}{9} = 50$ subjects randomly assigned to each unique treatment within each unique block.

D: Since there are 9 unique treatments within each of the 6 unique blocks, there are $\binom{9}{2} = 36$ total possible pair-wise comparisons between two distinct treatments within each unique block.

6. A: $\frac{7}{20}$ B: $\frac{3}{4}$ C: $\frac{1}{10}$ D: $\frac{3}{13}$

The following two-way table summarizes the given information in terms of proportions 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

A: $P(2\ yr) = 0.35 = \frac{7}{20}$

B: $P(E \cup 4\ yr) = 0.65 + 0.30 - .020 = 0.75 = \frac{3}{4}$

C: $P(B \cap 2\ yr) = 0.1 = \frac{1}{10}$

D: $P[(H \cup P) \mid 4\ yr] = \frac{0.10+0.05}{0.65} = \frac{0.15}{0.65} = \frac{3}{13}$

7. A: $\frac{3}{2}$ B: $\frac{2\sqrt{5}}{5}$ C: $\frac{128}{625}$ D: $\frac{36}{625}$

A: Since random variable X is a binomial random variable with $n = 5$ trials and a probability of success of $p = 0.3$, the expected value of random variable X is $E(X) = np = 5(0.3) = 1.5 = \frac{3}{2}$.

B: Since random variable Y is a binomial random variable with $n = 5$ trials and a probability of success of $p = 0.2 = \frac{1}{5}$, the standard deviation of random variable Y is $SD(Y) = \sqrt{np(1-p)} = \sqrt{5\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)} = \frac{2\sqrt{5}}{5}$.

C: This is a binomial setting with $n = 5$ trials and a probability of success $p = 0.2 = \frac{1}{5}$ and we are looking for $x = 2$ successes in the five trials because the expected value of the number of Health Sciences Majors in the SRS of 5 students is $2(5)(0.2) = 2$. Thus, we have: $P(X = 2) = \binom{5}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^3 = \frac{10\times4^3}{5^5} = \frac{128}{625}$.

D: This is a multinomial setting and so the requested probability is: $\frac{5!}{2!\times1!\times1!\times1!}\left(\frac{2}{5}\right)^2\left(\frac{3}{10}\right)\left(\frac{1}{5}\right)\left(\frac{1}{10}\right) = \frac{36}{625}$.

8. A: 6 B: 7 C: $-\frac{1}{2}$ *or* -0.5 D: $\frac{3}{110}$

A: $E(X^2) - [E(X)]^2 = (1)^2(0.4) + (6)^2(0.6) - [1(0.4) + 6(0.6)]^2 = 22 - 16 = 6.$

B: $E(Y^2) - [E(Y)]^2 = (2)^2(0.5) + (4)^2(0.3) + (9)^2(0.2) - [2(0.5) + 4(0.3) + 9(0.2)]^2 = 23 - 16 = 7.$

C: Given that $Var(X+Y) = Var(X) + Var(Y) + 2 \times Cov(X,Y) = 12$ and noting that the answers to parts A and B are just $Var(X) = E(X^2) - [E(X)]^2 = 6$ and $Var(Y) = E(Y^2) - [E(Y)]^2 = 7$, we can easily solve for the covariance between X and Y as follows: $Cov(X,Y) = \frac{12-6-7}{2} = -\frac{1}{2}$ $or -0.5$.

D: Since the covariance between X and Y was computed in part C as $Cov(X,Y) = -0.5$, random variables X and Y are <u>not independent</u>. Therefore, $D = \frac{9!}{(2!)(2!)(3!)} \div \frac{11!}{(3!)(2!)(3!)} = \frac{9!}{(2!)(2!)(3!)} \times \frac{(3!)(2!)(3!)}{11!} = \frac{3}{11\times 10} = \frac{3}{110}$.

9. A: 12 B: $2\sqrt{10}$ C: -8 D: $12\sqrt{2}$

A: $\mu_{\frac{1}{2}X+2Y} = \frac{1}{2}(4) + 2(5) = 12$

B: $\sigma_{\frac{1}{2}X+2Y} = \sqrt{\left(\frac{1}{2}\right)^2(4)^2 + (2)^2(3)^2} = \sqrt{4+36} = \sqrt{40} = 2\sqrt{10}$

C: $\mu_{3X-4Y} = 3(4) - 4(5) = -8$

D: $\sigma_{3X-4Y} = \sqrt{(3)^2(4)^2 + (-4)^2(3)^2} = \sqrt{144+144} = \sqrt{288} = 12\sqrt{2}$

10. A: 0.475 B: 0.815 C: 0.185 D: 0.354

A: $P(\mu_X \le X \le 12) = \frac{0.95}{2} = 0.475$

B: $P(-1 \le Y \le 8) = \frac{0.95}{2} + \frac{0.68}{2} = 0.815$

C: $P(X \le 0 \cup X \ge 12) = \frac{1-0.68}{2} + \frac{1-0.95}{2} = 0.16 + 0.025 = 0.185$

D: $P[(X \le 0 \cup X \ge 8) \cup (Y \le -1 \cup Y \ge 11)] = 0.32 + 0.05 - (0.32)(0.05) = 0.354$

11. A: -18 B: 10 C: 0.84 D: 0.975

A: $E(8\bar{X} - 10\bar{Y}) = 8(4) - 10(5) = 32 - 50 = -18.$

B: $SD(8\bar{X} - 10\bar{Y}) = \sqrt{(8)^2\frac{4^2}{16} + (10)^2\frac{3^2}{25}} = \sqrt{64+36} = \sqrt{100} = 10.$

C: Due to independence: $P[(8\bar{X} \le 40) \mid (10\bar{Y} \ge 20)] = P(8\bar{X} \le 40) = P(\bar{X} \le 5) = P\left(Z \le \frac{5-4}{4/\sqrt{16}} = 1\right) =$ 0.84 according to the Empirical Rule.

D: $P\left(\bar{X} \leq \frac{5\bar{Y}+1}{4}\right) = P(4\bar{X} \leq 5\bar{Y}+1) = P(4\bar{X} - 5\bar{Y} \leq 1) = P(8\bar{X} - 10\bar{Y} \leq 2)$. Now, using the mean and standard deviation of random variable $8\bar{X} - 10\bar{Y}$ calculated in parts A and B as $\mu = -18$ and $\sigma = 10$, we have that $P(8\bar{X} - 10\bar{Y} \leq 2) = P\left(Z \leq \frac{2-(-18)}{10} = 2\right) = 0.975$ according to the Empirical Rule.

12. A: 0 B: 1 C: $\frac{1}{40}$ *or* 0.025 D: 0.135

A: The sum of the residuals in any least-squares liner regression analysis is always equal to 0.

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 $0 + 1 = 1$.

C: 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 is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(0.5)}{400}} = \frac{1}{40} = 0.025$.

D: Let random variable X represent the number of heads we observe in the 400 flips of the fair coin. Thus, random variable X is a binomial distribution with a mean of $\mu = np = 100(0.5) = 200$ and a standard deviation of $\sigma = \sqrt{np(1-p)} = \sqrt{400(0.5)(0.5)} = 10$. Now, since we are told to use the Empirical Rule and the normal approximation to the binomial distribution with the aforementioned mean and standard deviation, we have $P(210 \leq X \leq 220) = \frac{0.95 - 0.68}{2} = 0.135$.

13. A: 8 B: $\frac{1}{7}$ C: $20 + 10\sqrt{2}$ D: 1

A: $\sigma = 2\sqrt{14} \rightarrow 2\sqrt{14} = \frac{\sqrt{1-p}}{p} \rightarrow 56 = \frac{1-p}{p^2} \rightarrow 56p^2 + p - 1 = 0 \rightarrow (8p-1)(7p+1) = 0 \rightarrow p = \frac{1}{8}$ *or* $\frac{-1}{7}$; however, since we must have $0 < p < 1, p = \frac{1}{8}$ is correct. Thus, $\mu = \frac{1}{1/8} = 8$.

B: $\sqrt{24p(1-p)} = \frac{1}{2}(24p) \rightarrow 24p(1-p) = 144p^2 \rightarrow 1 - p = 6p \rightarrow 7p = 1 \rightarrow p = \frac{1}{7}$.

C: $\mu = \frac{16+24}{2} = 20$ and $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{24-16}{4} = 2 \rightarrow \frac{\sigma_X}{\sqrt{50}} = 2 \rightarrow \sigma_X = 10\sqrt{2}$ and their sum is $20 + 10\sqrt{2}$.
NOTE: This calculation is made possible due to the Central Limit Theorem and the fact the sample size of n = 50 is sufficiently large for the sampling distribution of the sample mean to be approximately normal in spite of the fact that the population distribution is not specified in the given scenario.

D: Since the sum of two odd integers is always even, then their mean (which is the result of dividing their sum by 2) is always an integer.