Answers:

0. **2022**

1. **6767**

2. **0.55**

3. **285.285**

4. **23.186**

5. **– 0.2857**

6. **– 4470**

7. **7300**

8. **61.825**

9. **6794**

10. **50.389**

11. **5.1003**

12. **879.022**

13. $\frac{40}{9}$

14. **2**

SOLUTIONS:

0. Answers: $A = 337, B = 1, C = 337, D = 0,$ Summary $= 2022$
Entering the three data values into a list in the calculator and running the *1-Variable Statistics* program gives a sample mean of $A = \bar{X} = 337$, a sample standard deviation of $B = s_x = 1$, and a median of C = 337 as well. There are clearly no outliers in the data set, so D = 0. Summary = $3(337 \times 1 + 337 + 0) = 2022$.

1. Answers: $A = 1000, B = 5.767, C = 1, D = 0,$ Summary $= 6767$
Solutions: Carefully entering the 20 data values into a list in the calculator and running the *1-Variable Statistics* program gives us the following summary statistics: $\bar{X} = 10, s_x = 5.767422119, Q_1 = 5, Q_2 = 10, Q_3 = 15$ and we can clearly see from the already sorted list that the mode is 10. This makes $A = (10)(10)(10) = 1000$ and $B = 5.767$. C = 0 + 1 = 1 because the mean and standard deviation of any data set converted to Z-scores become 0 and 1, respectively. Using the 1.5(IQR) rule, the upper fence is 15 + 1.5(15 – 5) = 30 and the lower fence is 5 – 1.5(15 – 5) = –10, so there are no outliers making D = 0. Summary = 1000(5.767 + 1 + 0) = 6767.

2. Answer: Summary = 0.55
Solutions: The correct set of ordered pairs is {(1, 1), (1, 5), (2, 2), (2, 3), (2, 4), (2, 8), (4, 6), (4, 7)}. Entering the x-coordinates and y-coordinates into a pair of lists and running any one of the *Linear Regression* programs gives us a sample Pearson correlation coefficient of r = 0.55 when rounded to the hundredths place.

3. Answers: A = 638.382, B = –0.097, C = –203, D = – 150, Summary = 285.285
Solutions: Enter the X data into L_1 and the Y data into L_2 and run one of the *Linear Regression* programs to obtain the slope of b = –105.109, the y-intercept of a = 743.491, a correlation coefficient of r = –0.892, and a coefficient of determination of $r^2 = 0.795$ when all are rounded to the thousandths place. Thus, A = –105.109 + 743.491 = 638.382 and B = –0.892 + 0.795 = –0.097. The residual for the point (4, 120) is C = 120 – [–105.109(4) + 743.491] = –203 when rounded to the nearest whole bacterium. The predicted value for the number of bacteria remaining after 8.5 minutes is [–105.109(8.5) + 743.491] = –150 when rounded to the nearest whole bacterium. This is clearly not accurate because it is negative! This makes D = –150. Summary = 638.382 – 0.097 – 203 – 150 = 285.285.

4. Answers: A = 0.186, B = 0, C = 12, D = 11, Summary = 23.186
Solutions: The most efficient thing to do is to enter the X data into L_1, the Y data into L_2 and use $L_3 = \ln(L_2)$ for the transformed data. Then, one person should run a *Linear Regression* program on L_1 and L_2 while another runs the same program on L_1 and L_3. This gives us the two coefficients of determination as $r^2 = 0.795$ and $r^2 = 0.981$, respectively. This makes $A = |0.795 - 0.981| = 0.186$ and the coefficients for part B are $a = 7.3684$ and $b = -0.7096$. Thus, our model is $\hat{y} = e^{7.3684-0.7096x}$ and $B = y - \hat{y} = 11 - e^{7.3684-0.7096(7)} = 11 - 11 = 0$ when rounded correctly. Guessing and checking whole minutes in the model $\hat{y} = e^{7.3684-0.7096x}$ will soon lead us to C = 12 minutes as the answer since $\hat{y} = e^{7.3684-0.7096(12)} \approx 0.32$ which rounds to 0 whole bacterium. The model ln(Y) vs. X is an EXPONENTIAL model, so D = 11. Summary = 0.186 + 0 + 12 + 11 = 23.186.

5. Answer: Summary = –0.2857
Solutions: The correct set of ordered pairs is (1, 4), (2, 5), (3, 6), (4, 8), (5, 3), (6, 1), (7, 7), and (8, 2). Entering the x-coordinates into L_1 and the corresponding y-coordinates into L_2 and running a *Linear Regression* program gives us a correlation coefficient of r = –0.2857 when rounded to four decimal places. For the sake of thoroughness and to any disputes over the definition of any of the terms listed in Column Y of the question, here are screenshots (complete with typos) of the exact word-for-word definitions from the glossary of the updated 6th edition of *The Practice of Statistics*:

bias The design of a statistical study shows bias if it is very likely to underestimate or very likely to overestimate the value you want to know.

extrapolation Use of a regression model for prediction outside the interval of *x* values used to obtain the model. The further we extrapolate, the less reliable the predictions. (p. 178)

influentialpoint Any point that, if removed, substantially changes the slope, *y* intercept, correlation, coefficient of determination, or standard deviation of the residuals. (p. 200)

random sampling Using a chance process to determine which members of a population are chosen for the sample. (p. 253)

skewed A distribution of quantitative data is *skewed to the right* if the right side of the graph (containing the half of the observations with larger values) is much longer than the left side. It is *skewed to the left* if the left side of the graph is much longer than the right side. (p. 32)

simple random sample (SRS) Sample chosen in such a way that every group of *n* individuals in the population has an equal chance to be selected as the sample. (p. 254)

cluster sampling Method of sampling that divides the population into non-overlapping groups (*clusters*) of individuals that are located near each other, randomly chooses clusters, and includes each member of the selected clusters in the sample. (p. 258)

stratified random sampling Method of sampling that divides the population into non-overlapping groups *(strata)* of individuals who share characteristics thought to be associated with the variables being measured in a study, selects an SRS from each stratum, and combines the SRSs into one overall sample. (p. 257)

6. Answers: A = 800, B = 720, C = – 2.9800, D = 20, Summary = – 4470
Solutions: $A = 8000(0.10) = 800$ and $B = 8000(0.09) = 720$. Running the *1-Proportion Z-Test* program with the inputs $P_0 = 0.10$, x = 720, and n = 8000 gives us a test statistic of Z = –2.9814 and a left-tail p-value of p = 0.0014 when rounded. So, $C = -2.9814 + 0.0014 = -2.9800$. Since the p-value of 0.0014 is less than 0.01, the null hypothesis is rejected and so D = 20. The summary answer is $-2.9800(720 + 800 - 20) = -4470$.

7. Answers: A = 1600, B = 36.5, C = 1, D = 8, Summary = 7300
Solutions: Since the sample size is n = 8000, the mean of the expected counts of the five letter choices is $A = \frac{8000}{5} = 1600$. According to the recommended distribution of letter choices, we expect $8000(0.10) = 800$ E's, and $\frac{7200}{4} = 1800$ of each other letter in a sample of n = 8000 questions. Also, multiplying each observed percentage gives us the following list of observed counts for A, B, C, D, and E, respectively: 1672, 1904, 1928, 1776, 720. This makes our chi-square goodness of fit test statistic $\chi^2 = \sum \frac{(Observed-Expected)^2}{Expected} = \frac{(1672-1800)^2}{1800} + \frac{(1904-1800)^2}{1800} \dots \frac{(720-800)^2}{800} = 32.53333 \dots$ and the p-value is $p = \chi^2 cdf(32.5333, 999999, 4) \approx 0.0000015$. This makes B = 32.5 + 4 = 36.5 and C = 1 because the p-value is less than 0.01. Examining the components one at a time or by looking at the list of components provided by the $\chi^2 GOF - Test$ program we see that $D = \frac{(720-800)^2}{800} = 8$. Summary = $\frac{1600(36.5)(1)}{8} = 7300$.

8. Answers: A = 14, B = 0.106, C = 47.719, D = 0, Summary = 61.825
Solutions: The one-sample t-test statistic with n – 1 degrees of freedom is $t = \frac{\bar{X} - \mu}{s_x/\sqrt{n}}$. Since we have everything but the sample size, we can solve $1.3078 = \frac{48.208 - 47.91}{0.8825/\sqrt{n}}$ for n. Doing so we get $n = \left(\frac{1.3078 \times 0.8825}{48.208 - 47.91}\right)^2 \approx 15$ when rounded to the nearest integer. This makes the degrees of freedom A = 15 – 1 = 14. This also makes the p-value of the test is $B = p = tcdf(1.3078, 999999, 14) \approx 0.106$ when rounded. The critical t-score for a 95% confidence interval is given by $t^* = invT(0.975, 14) \approx 2.1448$ and the lower bound of the confidence interval is $C = 48.208 - 2.1448 \times \frac{0.8825}{\sqrt{15}} \approx 47.719$ when rounded. Since the p-value of the test is not less than 0.05 and the lower limit of the 95% confidence interval captures 47.91g, we are not sufficiently convinced that the mean minimum weight of all M^2 candy bags labeled as 47.91g is actually greater than 47.91g at the 5% significance level. Thus, D = 0. Summary = 14 + 0.106 + 47.719 + 0 = 61.825.

9. Answers: A = 3.0114, B = 0.0026, C = 1000, D = 3780, Summary = 6794
Solutions: This is a test of $H_0: P_N = P_T$ $vs.$ $H_A: P_N \neq P_T$. Running the *2-Proportion Z-Test* program with $x_1 = 500(0.126) = 63$, $x_2 = 500(0.196) = 98$ and $n_1 = n_2 = 500$ with the two-tail alternative selected gives us a test statistic of Z = –3.0114 and a p-value of p = 0.0026 when rounded to four decimal places. Thus, A = 3.0114, B = 0.0026, and C = 1000 since the p-value is well below 0.01. The TENNESSEE factory clearly seems to pump out a significantly higher proportion of green M^2 candies and the data supports this since the one-tail p-value is half the two-tail one. Thus, $D = \frac{9!}{4! \times 2! \times 2!} = 3780$. Summary = $1000(3.0114 + 0.0026) + 3780 = 6794$.

10. Answers: A = 48, B = 1.5, C = 0.816, D = 0.073, Summary = 50.389
Solutions: $A = invNorm(0.10, 49, 0.75) \approx 48.0388$ which rounds to 48 when rounded to the nearest whole gram. B is asking for the mean or expected value of a binomial setting with n = 15 and p = 0.10: B = μ = E(X) = np = 15(0.10) = 1.5. $C = binomcdf(15, 0.1, 2) \approx 0.816$ and $D = normalcdf(-999999, 47.91, 49, 0.75) \approx 0.073$ when each are rounded to the nearest thousandth. Summary = 48 + 1.5 + 0.816 + 0.073 = 50.389.

11. Answers: A = 0.0202, B = 0.9902, C = 4.0899, D = 0, Summary = 5.1003
Solutions: $A = 1 - binomcdf(15, 0.073, 3) \approx 0.0202$ when rounded. $B = normalcdf\left(48.5, 49.5, 49, \frac{0.75}{\sqrt{15}}\right) \approx 0.9902$ when rounded. Part C is actually a one-sample Z-test and not a t-test because we have $\sigma = 0.75\ g$ when we assume Claim 1 is true. Fortunately, however, we obtain the same result for the test statistic as long as we use $\sigma = 0.75\ g$ in the right place in the formula: $|Z| = |t| = \left|\frac{48.208-49}{0.75/\sqrt{15}}\right| \approx 4.0899$ when rounded. Treating this test statistic as either a Z-score or a t-score will give us a p-value well below 0.01, so the null $H_0: \mu = 49$ is rejected and the alternative $H_A: \mu < 49$ is supported. This makes C = 4.0899 and D = 0 since Claim 1 (which is represented by the null hypothesis) is rejected by the test in part C; thus, making it implausible. Summary = 0.0202 + 0.9902 + 4.0899 + 0 = 5.1003.

12. Answers: A = 4, B = 10, C = 865, D = 0.022, Summary = 879.022
Solutions: In part A we need the mean and standard deviation of a binomial setting with n = 15 and p = 0.10. They are $\mu = E(X) = np = 15(0.10) = 1.5\ and\ \sigma = \sqrt{np(1-p)} = \sqrt{15(0.1)(0.9)} \approx 1.16$. Two standard deviations above the mean is approximately 1.5 + 2(1.16) = 3.82 which rounds up to A = 4 bags. Part B is essentially a geometric setting modeled by a geometric distribution with p = 0.10 and we are looking for B = μ = E(X) = 1 / 0.10 = 10. NOTE: Even if p < 0.10, E(X) still rounds to 10 when p is sufficiently close to 0.10; for example, 1 / 0.099999 = 10.0001 which rounds to 10. Also, using p = 0.10 will minimize E(X), which is what we want since we are asked to find "<u>at least</u> how many bags…" $C = n = 0.10(0.90)\left(\frac{1.96}{0.02}\right)^2 = 864.36$ which rounds up to C = 865. Running the *1-Proportion Z-Interval* program with x = 117, n = 900, and 95% confidence we obtain the interval (0.10803, 0.15197) which does not quite capture the claimed value of p = 0.10 made by the company, so we can reject that claim. This makes $D = \frac{0.15197-0.10803}{2} = 0.022$ when rounded to the thousandths place. Summary = 4 + 10 + 865 + 0.022 = 879.022.

13. Answers: A = 3, B = $\frac{5}{9}$, C = $\frac{4}{3}$, D = 2, Summary = $\frac{40}{9}$
Solutions: Since we are given the cumulative distribution function $F(x) = 2 - 2x^{-1}$, we know that $F(a) = 0$ and $F(b) = 1$. So, we must solve the equation $2 - 2a^{-1} = 0$ for *a* and the equation $2 - 2b^{-1} = 1$ for *b*. Doing so, we obtain a = 1 and b = 2 which makes A = 1 + 2 = 3. $B = P\left(\frac{6}{5} \le X < \frac{9}{5}\right) = F\left(\frac{9}{5}\right) - F\left(\frac{6}{5}\right) = 2 - \frac{2}{9/5} - \left(2 - \frac{2}{6/5}\right) = \frac{5}{9}$. Let *M* represent the median of random variable *X*. To find *M*, we must solve $F(M) = \frac{1}{2}$. Solving $2 - \frac{2}{M} = \frac{1}{2}$ for *M* gives us the median of $C = M = \frac{4}{3}$. Graphing the pdf $f(x) = 2x^{-2}$ in the calculator shows that $f(x)$ is monotonically decreasing from left to right, thus making the density curve skewed to the right which will make the mean of the distribution closer to D = *b* = 2. Summary = $3 \times \frac{5}{9} \times \frac{4}{3} \times 2 = \frac{40}{9}$.

14. Answers: A = $\frac{1}{64}$, B = $\frac{21}{64}$, C = $\frac{43}{64}$, D = $\frac{63}{64}$, Summary = 2
Solutions: First, solve for P(X), P(Y), and P(Z) based on the given information. Since $P(X) = 2P(Y),\ P(Y) = 2P(Z), and\ P(X) + P(Y) + P(Z) = \frac{7}{8}, we\ have\ 4P(Z) + 2P(Z) + P(Z) = \frac{7}{8}$. Solving for P(Z) and back substituting for P(Y) and P(X) we get: $P(X) = \frac{1}{2}, P(Y) = \frac{1}{4}, P(Z) = \frac{1}{8}$.
A = $\frac{1}{64}$ since $P(X\ and\ Y\ and\ Z) = \frac{1}{2} \times \frac{1}{4} \times \frac{1}{8} = \frac{1}{64}$.
B = $\frac{21}{64}$ since $P(X^C\ and\ Y^C\ and\ Z^C) = \frac{1}{2} \times \frac{3}{4} \times \frac{7}{8} = \frac{21}{64}$.
C = $\frac{43}{64}$ since $P(X\ or\ Y\ or\ Z) = 1 - \frac{1}{2} \times \frac{3}{4} \times \frac{7}{8} = \frac{43}{64}$.
D = $\frac{63}{64}$ since $P(X^C\ or\ Y^C\ or\ Z^C) = 1 - \frac{1}{2} \times \frac{1}{4} \times \frac{1}{8} = \frac{63}{64}$.
Summary: $A + B + C + D = \frac{1}{64} + \frac{21}{64} + \frac{43}{64} + \frac{63}{64} = \frac{128}{64} = 2$.