Answer Key:

1. E
2. D
3. C
4. D
5. A

6. C
7. D
8. C
9. C
10. B

11. D
12. A
13. C
14. E
15. B

16. B
17. D
18. B
19. B
20. C

21. C
22. D
23. B
24. B
25. A

26. B
27. C
28. D
29. C
30. A

Solutions:

1. E: Describing the shape of a distribution does not apply to categorical variables, such as the ordinal categorical variable age group. Thus, none of the listed choices are correct.

2. D: A geometric distribution has a mean of $\mu_X = \frac{1}{p}$. So, given that $\mu_X = \frac{1}{p} = 15$ results in $p = \frac{1}{15}$. The variance of this geometric random variable is $\sigma_X^2 = \frac{1-p}{p^2} = \frac{1-\frac{1}{15}}{\left(\frac{1}{15}\right)^2} = 210 = 2(3)(5)(7)$ and 210 has (2)(2)(2)(2) = 16 total positive integral factors.

3. C: Using $invNorm(0.997, 0, 1) \approx 2.74778$ for the 99.7th percentile, we can set up the z-score formula of $2.74778 = \frac{89.8-\mu}{\sigma}$ for Daniel and (in a similar fashion) Rick's z-score corresponds to the 68th percentile which gives us $0.467699 = \frac{80.3-\mu}{\sigma}$. Solving this system of equations by first solving for σ by multiplying each side of each equation by σ then subtracting the two resulting equations gives us $\sigma = \frac{89.8-80.3}{2.27448-0.467699} \rightarrow \sigma \approx 4.1665$ and then substituting this value into either original equation yields $\mu \approx 78.3513$ for a sum of 4.1665 + 78.3513 = 82.5178 which rounds to 83.

4. D: Effie is first splitting the country into the 50 states (stratifying by state); then, she divides each state into regions by county and randomly selects some counties (or clusters), and then takes a random sample of high schools within each county (more clusters). Then finally takes a random sample of students from each of these schools (or clusters). Thus, since this sampling method uses both stratified random sampling, cluster random sampling, and simple random sampling, it is clearly a multistage random sample.

5. A: Let random variable $D = N - H$, in which $\mu_D = \mu_{N-H} = \mu_N - \mu_H = 88.4 - 72.5 = 15.9$ and $\sigma_D = \sigma_{N-H} = \sqrt{\sigma_N^2 + \sigma_H^2} = \sqrt{2.3^2 + 5.6^2} = \sqrt{36.65}$. Thus, D is $N\left(15.9, \sqrt{36.65}\right)$. To find the approximate probability that Harry eats more than Normie: $P(D < 0) = normalcdf(-\infty,\ 0, 15.9,\ \sqrt{36.65}) \approx 0.0043$ when rounded.

6. C: Since N is approximately Normally distributed with a mean of 88.4 and a standard deviation of 2.3, then $T = \sum_{k=1}^{50} N_k$ is Normally distributed with a mean of $\mu_T = \sum_{k=1}^{50} 88.4 = 50 \cdot 88.4 = 4420$ and a standard deviation of $\sigma_T = \sqrt{\sum_{k=1}^{50} 2.3^2} = \sqrt{50 \cdot 2.3^2} = \sqrt{264.5}$. Now, we are looking for $P(T > 4400) = normalcdf(4400, \infty, 4420, \sqrt{264.5}) \approx 0.8906$ when rounded.

7. D: If $P(A \cup B^C) = 0.9$, then P(B) = 0.4. Thus, $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.3}{0.4} = \frac{3}{4}$. The following two-way table summarizes the given information:

	A	A^C	Total
B	0.3	0.1	0.4
B^C	?	?	?
Total	?	?	1

8. C: This scenario is modeled by a binomial distribution with n = 25 independent trials and a probability of success of $p = \frac{1}{3}$. The requested probability is now $P(X \geq 14) = 1 - binomcdf(25, \frac{1}{3}, 13) \approx 0.0164$ when rounded.

9. C: $P(choosing\ bag\ 1) = P(choosing\ bag\ 2)\ = \frac{1}{2}$.
Let event R = drawing a red Lego piece; thus: $P(R \mid bag\ 1) = \frac{2}{10}$ and $P(R \mid bag\ 2) = \frac{6}{8}$.
Now, using Bayes' Theorem:

$$P(bag\ 1 \mid R) = \frac{P(bag\ 1) \times P(red \mid bag\ 1)}{P(bag\ 1) \times P(red \mid bag\ 1) + P(bag\ 2) \times P(red \mid bag\ 2)} = \frac{\left(\frac{1}{2}\right)\left(\frac{2}{10}\right)}{\left(\frac{1}{2}\right)\left(\frac{2}{10}\right) + \left(\frac{1}{2}\right)\left(\frac{6}{8}\right)} = \frac{4}{19}$$

10. B: The equation of the LSRL based on the Minitab output is $\widehat{PPG} = 41.92 - 0.518(rank)$. If the basketball player was ranked 7th, their predicted points per game is $\widehat{PPG} = 41.92 - 0.518(7) = 38.294$. To calculate the residual (the actual value of PPG minus the predicted value of PPG), we would obtain $residual = 36.13 - 38.294 = -2.164$.

11. D: The coefficient of determination, r^2, tells us the percent of variation in the points per game that is explained by the least-squares regression line. So, the percent of variation that is unexplained is $1 - r^2 = 1 - .5519 = .4481$, or 44.81%.

12. A: The standard error of the regression, $s = 2.83451$, and the correlation coefficient, $r \approx -\sqrt{.5519}$, as the slope of the LSRL is negative resulting in a negative correlation coefficient. Thus, their sum is $2.83451 + (-\sqrt{.5519}) \approx 2.092$ when rounded.

13. C: Grade level (Freshman, Sophomore, Junior, Senior) form a set of strata. Thus, it is a stratified random sample that Tim is using.

14. E: Since X and Y are independent random variables and $Z = 2X - Y - 3$, $Var(Z) = Var(2X - Y - 3) = Var(2x) + Var(Y) = 2^2(Var(X)) + Var(Y) = 4(2) + 4 = 12$.

15. B: Using the given information, we have $P(toasted \mid turkey\ sub) = x, P(toasted \mid Italian\ sub) = 2x$ and $P(toasted \mid chicken\ tender\ sub) = 4x$. Solving $x + 2x + 4x = 1$ for $x\ \rightarrow x = \frac{1}{7}$. This will now help us set up the following two-way table:

	Turkey Sub	Italian Sub	Chicken Tender Sub	Total
Toasted	$0.5\left(\frac{1}{7}\right) = \frac{1}{14}$	$0.3\left(\frac{2}{7}\right) = \frac{3}{35}$	$0.2\left(\frac{4}{7}\right) = \frac{4}{35}$	$\frac{19}{70}$
Not Toasted	$0.5\left(\frac{6}{7}\right) = \frac{3}{7}$	$0.3\left(\frac{5}{7}\right) = \frac{3}{14}$	$0.2\left(\frac{3}{7}\right) = \frac{3}{35}$	$\frac{51}{70}$
Total	0.5	0.3	.02	1

Thus, $P(chicken\ tender\ sub \mid toasted) = \frac{\frac{4}{35}}{\frac{19}{70}} = \frac{8}{19}$.

16. B: There is one explanatory variable, which is the over-the-counter experimental drug, and it has four levels (the four different dosages).

17. D: A histogram is used to display quantitative data. Three of the given variables are considered quantitative: age, number of years lived at the current address, and annual salary.

18. B: The given formula $P(X = n) = \binom{16}{n}(0.2)^n(0.8)^{16-n}$ is for a binomial distribution with 16 independent trials with a probability of success of 0.2. The mean of this binomial random variable is $\mu_x = np = 16(0.2) = 3.2$ and the standard deviation is $\sigma_x = \sqrt{np(1-p)} = \sqrt{16(0.2)(0.8)} = 1.6$. The resulting product is $3.2(1.6) = 5.12$.

19. B: Plugging the data sets into the *1-Var Stats* program, we can determine that the median of each data set is 25. The sample standard deviation of the data set on the left of the stemplot is approximately 14.3324 and that of the data set on the right of the stemplot is approximately 10.1675. Using the 1.5(IQR) rule, the data set on the left has no outliers because the lower fence is $14.5 - 1.5(30.5 - 14.5) = -9.5$ and the upper fence is $30.5 + 1.5(30.5 - 14.5) = 54.5$. The dataset on the right has no outliers because the lower fence is $18.5 - 1.5(31.5 - 18.5) = -1$ and the upper fence is $31.5 + 1.5(31.5 - 18.5) = 51$. Now, since none of the data values in either dataset are strictly <u>outside</u> of their respective fences, neither dataset has any outliers. Thus, only statements I and III are true.

20. C: Let F = the event that the Florida Flamingoes win any game and let A = the event that the NY Apples win any game. Thus, $P(F) = \frac{2}{3}$, $P(F^C) = \frac{1}{3}$, $P(A) = \frac{1}{3}$, and $P(A^C) = \frac{2}{3}$. If there is no winner of the series until after the third game is played because the series is tied at one game apiece after the first two games are played, then either the Florida Flamingoes won the first game and then the NY Apples won the second game; or the NY Apples won the first game and then the Florida Flamingoes won the second game. Thus, we have $P(F)P(A) + P(A)P(F) = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{4}{9}$.

21. C: This is the conditional probability that the New York Apples won the series in the third game given that they won the series in general:

$$P(\textit{won the series in the third game} \mid \textit{won the series}) = \frac{\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)+\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)}{\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)+\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)+\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)} = \frac{4}{7}.$$

22. D: The mean of a binomial random variable is $\mu_x = np = 18.24$ and the variance is $\sigma^2 = npq = 4.3776$. Dividing the variance by the mean results in $\frac{\sigma^2}{\mu} = \frac{npq}{np} = q$. Thus, $q = \frac{4.3776}{18.24} = 0.24$. This means that $p = 1 - q = 1 - 0.24 = 0.76$ and $n = \frac{18.24}{0.76} = 24$. Now, $P(X > 14) = 1 - binomcdf(24, 0.76, 14) \approx 0.958$ when rounded.

23. B: In a Standard Normal distribution, the mean, the median, and the mode are all equivalent.

24. B: Plugging the midpoints of the time intervals of 20, 30, 40, 50, 60, 70, and 80 into L_1 and the corresponding probabilities into L_2 and running the *1-Vars Stats* program gives us a mean and standard deviation of $E(X) = \mu = 55$ minutes and $\sigma \approx 21.7945$ minutes, respectively. The waiting times that are within one standard deviation of the mean are the waiting times with the midpoints of 40, 50, 60, and 70 minutes. The total probabilities of these waiting times are $0.05 + 0.20 + 0.10 + 0.10 = 0.45$ or 45%. NOTE: A potential unique interpretation exists if a student assumes uniformity of the probability distribution withing each subinterval for the waiting times. If so, then this would add approximately 0.036 to 0.45 which yields approximately 48.6% which rounds up to 50%. However, only grant a student this particular unique interpretation if they specifically dispute along these lines.

25. A: The graph described is a *mosaic plot*: $\frac{10!}{2!} = 1{,}814{,}400$.

26. B: $P(X\ is\ even) = P(X=2) + P(X=4) + P(X=6) + \ldots = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \frac{1}{2^8} + \ldots =$
$\frac{1}{4}\left(1 + \frac{1}{4} + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \ldots\right) = \frac{1}{4} \cdot \frac{1}{1-\frac{1}{4}} = \frac{1}{3}$.

27. C: Since X is a binomial random variable with a probability of success of p on each independent trial, n, we know that $P(X=x) = \binom{n}{x}p^x(1-p)^{n-x}$, and so:

$$P(X=n) = \binom{n}{n}p^n(1-p)^{n-n} = \binom{n}{n}p^n(1-p)^0 = p^n = 0.00032$$

$$P(X=n-1) = \binom{n}{n-1}p^{n-1}(1-p)^1 = np^{n-1}(1-p) = n(p^{n-1} - p^n) = 0.00128n$$

Thus, $p^{n-1} - p^n = 0.00128$. Adding these two equations together results in $p^{n-1} = 0.0016$ and therefore, $p = \frac{p^n}{p^{n-1}} = \frac{0.00032}{0.0016} = 0.2$.

28. D: Since X is a continuous uniform distribution over the real interval $(4, 8)$, it follows that $E(X) = \frac{4+8}{2} = 6$ and $Var(X) = \frac{(8-4)^2}{12} = \frac{4}{3}$. The volume of the box is $V = 9X^2$, so the expected value of the volume of the box is $E(V) = E(9X^2) = 9E(X^2) = 9(Var(X) + (E(X))^2) = 9\left(\frac{4}{3} + 36\right) = 336\ in^3$.

29. C: Since X is a geometric random variable with a probability of success of p on each independent trial until the first success, x, we know that $P(X=x) = (1-p)^{x-1}p$. Hence: $P(X=x) = \left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right)$ and $P(X \le x) = \frac{1}{6}\left(\frac{1-\left(\frac{5}{6}\right)^x}{\frac{1}{6}}\right) = 1 - \left(\frac{5}{6}\right)^x \rightarrow 1 - \left(\frac{5}{6}\right)^x \ge \frac{1}{2} \rightarrow \left(\frac{5}{6}\right)^x \le \frac{1}{2}$ which means that $x \le \frac{ln\left(\frac{1}{2}\right)}{ln\left(\frac{5}{6}\right)}$ or $x \le$ approximately 3.8. This means that the smallest value of x in order to have $P(X \le x) \ge \frac{1}{2}$ is $x = 4$.

30. A: Recalling random variables N, H, and X from the appropriate prior questions, we have that $N \sim N(88.4, 2.3)$, $H \sim N(72.5, 5.6)$, and $X \sim Bi(n = 24, p = 0.76) \rightarrow \mu_X = 18.24$ and $\sigma_X^2 = 4.3776$ as was given to us. Thus, random variable $Y = (2N - 3) - (4H + 10) + (5X + 2)$ has a mean of $\mu_Y = (2 \times 88.4 - 3) - (4 \times 72.5 + 10) + (5 \times 18.24 + 2) = -33$ and a standard deviation of $\sigma_Y = \sqrt{2^2(2.3^2) + 4^2(5.6^2) + 5^2(4.3776)} = \sqrt{632.36}$. Now, thanks to the Central Limit Theorem and the fact that $n = 100$, random variable $\bar{Y}$ is approximately Normally distributed with a mean of $\mu_{\bar{Y}} = -33$ and a standard deviation of $\sigma_{\bar{Y}} = \sqrt{\frac{632.36}{100}} = \sqrt{6.3236}$. Hence, the requested probability is:
$P(-40 < \bar{Y} < -30) = normalcdf(-40, -30, -33, \sqrt{6.3236}) \approx 0.88$ when rounded.