STATISTICS TEST – OPEN
SOLUTIONS

1. A Use the fact that the probabilities must add up to 1, and $E(X) = 1.4 = \sum_{i=1}^{4} X_i \cdot P_i(X)$.
   $0.4 + a + b + a + b = 1$ and $a + 2b + 3(a + b) = 1.4 \Rightarrow a = 0.1, b = 0.2 \Rightarrow a^2 + b^2 = .05$

2. A $P(A \cup B) = 0.6 = 2P(A) - P(A \cap B)$ and $0.8P(A) = P(A \cap B) \Rightarrow P(A) = 0.5, P(A \cap B) = 0.4$
   $P[(A \cap \overline{B}) \cup (\overline{A} \cap B)] = P(A \cup B) - P(A \cap B) = 1/5$

3. B $Q_3 + 1.5(Q_3 - Q_1) = 3.6$ and $Q_1 - 1.5(Q_3 - Q_1) = 0.5 \Rightarrow Q_1 = 1.6625$, $Q_3 = 2.4375$
   The median is equidistant from the two quartiles so the median equals 2.05.

4. A $r / r^2 = 1.25 \Rightarrow r = 0.8$ and $m = r \cdot \frac{s_y}{s_x} \Rightarrow m = 0.4$
   The point $(\overline{x}, \overline{y})$ lies on the regression line. Thus $4 = 0.4(5) + b \Rightarrow b = 2$, so $m - b = -1.6$.

5. A The sample mean is obviously 65. Thus the margin of error is 2.
   $2 = z^* \frac{\sigma}{\sqrt{n}} = z^* \frac{6}{\sqrt{20}} \Rightarrow z^* \approx 1.4907$
   If we look at the standard normal table this value corresponds closely to the 86% confidence level.

6. D Let n equal the number of lifetime at-bats. The number of lifetime home runs, triples, doubles, and singles are $(0.1)(0.3)n$, $(0.2)(0.3)n$, $(0.3)(0.3)n$, and $(0.4)(0.3)n$ respectively.
   Slugging Percentage $= \frac{Total\ Bases}{At-bats} = \frac{4(0.1)(0.3)n + 3(0.2)(0.3)n + 2(0.3)(0.3)n + (0.4)(0.3)n}{n} = 0.6$

7. *D P*(Sum of two die is prime) = 15/36
   For this binomial distribution $\mu = 300\left(\frac{15}{36}\right)$ and $\sigma = \sqrt{300\left(\frac{15}{36}\right)\left(1 - \frac{15}{36}\right)}$.
   Normal Approximation w/continuity correction $\Rightarrow P\left(z < \frac{140.5 - \mu}{\sigma}\right) - P\left(z < \frac{99.5 - \mu}{\sigma}\right) = 0.96$

8. C $r = {}_{13}C_2\left(\frac{{}_4C_2 \cdot {}_4C_2 \cdot {}_{44}C_1}{{}_{52}C_5}\right)$ Run a two-sided one-proportion z-test. $z = \frac{4/200 - r}{\sqrt{\frac{r(1-r)}{200}}} \approx -1.830$
   This corresponds with a p-value of $2 \times 0.0336 = 0.0672 \approx 0.067$.

9. C The survival rate for Dr. Brown's good patients and poor patients must be higher than Dr. Benton's.
   $\frac{a}{68} > \frac{120}{137}$ and $\frac{190 - a}{182} > \frac{80}{113} \Rightarrow a > 59\ \&\ a < 62 \Rightarrow b = 59\ \&\ c = 62 \Rightarrow b - c = 3$

STATISTICS TEST – OPEN
SOLUTIONS

10. A

# of tails out of 6 coins	Observed	Expected
0	5	$320\left({}_6C_0(1/2)^6\right)=5$
1	40	$320\left({}_6C_1(1/2)^6\right)=30$
2	86	$320\left({}_6C_2(1/2)^6\right)=75$
3	89	$320\left({}_6C_3(1/2)^6\right)=100$
4	67	$320\left({}_6C_4(1/2)^6\right)=75$
5	29	$320\left({}_6C_5(1/2)^6\right)=30$
6	4	$320\left({}_6C_6(1/2)^6\right)=5$

$$\chi^2 \text{ Statistic} = \frac{(5-5)^2}{5}+\frac{(40-30)^2}{30}+\frac{(86-75)^2}{75}+\frac{(89-100)^2}{100}+\frac{(67-75)^2}{75}+\frac{(29-30)^2}{30}+\frac{(4-5)^2}{5}\approx 7.24$$

11. B With 6 degrees of freedom and a χ^2 statistic of 7.24, the p-value is greater then 0.25.

12. A Since the two variables, cholesterol before and cholesterol after, are not independent and since we can't know the population standard deviation we must do a matched pairs t-test. In this experiment, each person is his or her own control. First, find the set of differences.

Subject	A	B	C	D	E	F	G
Cholesterol After - Before	11	13	-8	-3	7	-2	9

The null hypothesis for this test is that the mean difference is 0 and the alternative hypothesis is that the mean difference is greater than 0. Because that would mean the drug has worked; that is what we are testing.

$$\bar{x}_d \approx 3.85714286 \ \& \ s_d \approx 8.09173594 \Rightarrow t=\frac{\bar{x}_d-\mu_d}{s_d}\approx 1.26 \Rightarrow p-value \approx 0.13$$

13. E When a one sample, two-sided z-test is performed, the critical value for $|z|$ at a 5% α-level is 1.96. Thus the data will be significant at this level if z is less than -1.96 or greater than 1.96.

$$z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \Rightarrow z=1.96<\frac{n-130}{6/\sqrt{10}} \text{ or } z=-1.96>\frac{n-130}{6/\sqrt{10}} \Rightarrow n>133.7 \text{ or } n<126.2$$

Since n must be a natural number, $c=126$ and $d=134$. Thus, $d-c=8$.

14. C The primes up to 20 are 2, 3, 5, 7, 11, 13, 17, & 19. The probability that a z is within 0.5 units of a prime number is

$$\frac{1}{20}\begin{bmatrix}(2.5-1.5)+(3.5-2.5)+(5.5-4.5)+(7.5-6.5)+(11.5-10.5)+\\(13.5-12.5)+(17.5-16.5)+(19.5-18.5)\end{bmatrix}=\frac{2}{5}\Rightarrow$$

$$\frac{M}{N}=1-\frac{2}{5}=\frac{3}{5}\Rightarrow M+N=8$$

STATISTICS TEST – OPEN
SOLUTIONS

15. C Let p be the probability of getting a full house. $p = {}_{13}C_2\left(\dfrac{{}_4C_3 \cdot {}_4C_2}{{}_{52}C_5}\right)$

Let X represent the payoff. X takes two values –1 and m – 1 with probabilities of 1 – p and p respectively. To break even in the long run, the expected value of the payoff must be 0.

$$E(X) = \sum_{i=1}^{2} X_i \cdot P_i(X) = -1\cdot(1-p) + (m-1)\cdot p = 0 \Rightarrow m \approx 1388.33 \Rightarrow \log m \approx 3.1424$$

16. B There are 30 marbles total, thus n + m = 12. And remember m must be greater or equal to than n.

$$P(Same\ Color\ Picked) = \frac{107}{435} = \frac{10\cdot 9 + 8\cdot 7 + m(m-1) + n(n-1)}{30\cdot 29} =$$

$$\frac{90 + 56 + m(m-1) + (12-m)(11-m)}{30\cdot 29}$$

Thus $m = 8,\ n = 4 \Rightarrow m^2 + nm + m = 100$.

17. D The probability that a certain team is not in its correct spot is $\left(1-\dfrac{1}{25}\right)$. The probability that no teams are in their correct spots is $\left(1-\dfrac{1}{25}\right)^{25} \approx 0.360$.

18. A The value of $\dfrac{n}{100}\cdot\dfrac{m}{100}$ represents the percentage of applicants who actually decided to attend the school. $\dfrac{n}{100}\cdot\dfrac{m}{100} = \dfrac{1100}{9000} \Rightarrow nm \approx 1222$

19. D The possible combinations of Cubs winning in five games are WWWLW, WWLWW, WLWWW, & LWWWW. The probabilities for these are $(0.9)(0.6)(0.6)(0.4)(0.9)$, $(0.9)(0.6)(0.4)(0.6)(0.9)$, $(0.9)(0.4)(0.6)(0.6)(0.9)$, and $(0.1)(0.6)(0.6)(0.6)(0.9)$ respectively. The sum of these probabilities is approximately 0.369.

20. C

$$\sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2}\ \&\ \mu_{x+y} = \mu_x + \mu_y \Rightarrow 10 = 4 + \tau\ \&\ \delta^2 = \delta + \tau \Rightarrow \tau = 6,\ \delta = 3 \Rightarrow \ln(\tau + \delta) \approx 2.197$$

21. E For variables that follow the Poisson distribution, the standard deviation is equal to the square root of the mean.

22. C Total area under any probability density curve is 1, and $c = \dfrac{1}{b}$, $d = \dfrac{1}{b} + 3$

$$\frac{1}{2}\left[b\left(\frac{1}{b}+1\right)-1\right] + b + \frac{1}{2}\left[-b\left(\frac{1}{b}-1\right)+1\right] = 1 \Rightarrow b = \frac{1}{2} \Rightarrow c = 2,\ d = 5 \Rightarrow \ln(0.5+2+5) \approx 2.0149$$

STATISTICS TEST – OPEN
SOLUTIONS

23. C In order for the class mean to go from 75 to 65 and the standard deviation to go from 8 to 5, each person's grade must be multiplied by a number b and then a number c must then be added to each.
$$\sigma_{New} = b \cdot \sigma_{Old} \ \& \ \mu_{New} = b \cdot \mu_{Old} + c \Rightarrow 5 = b \cdot 8 \ \& \ 65 = b \cdot 75 + c \Rightarrow b = 0.625, c = 18.125$$
$$Andrea's\ New\ Grade = b \cdot (Andrea's\ Old\ Grade) + c \Rightarrow Andrea's\ New\ Grade = 74.375$$

24. C Probability of a Type I error is equal to the significance level, $\alpha = 0.05$. Probability of a Type II error is equal to 1 minus the power of the test, $\beta = 1 - 0.86 = 0.14$. So $\alpha + \beta = 0.19$.

25. A
$$\mu_4 = 4 \cdot 190 \ \& \ \sigma_4 = \sqrt{10^2 + 10^2 + 10^2 + 10^2} \Rightarrow$$
$$P(Weight > 800) = P\left(z > \frac{800 - \mu_4}{\sigma_4}\right) = P(z > 2) = 0.023$$

26. A The number of distinguishable combinations of MISSISSIPPI is equal to $\frac{11!}{4! \cdot 2! \cdot 4!} = 34650$. The number of distinguishable combinations of MISSISSIPPI where the P's are in the first and last spots is equal to the number of distinguishable combinations of MISSISSII, which is $\frac{9!}{4! \cdot 4!} = 630$. And $\frac{630}{34650} \approx 0.018$.

27. C An increase in the population variance causes the overlap between the hypothesized distribution and the true distribution to get larger. Thus the probability of committing a Type II error increases.

28. C
$$n(Analysis) = n(LinearA\lg ebra) \Rightarrow n(Analysis \bigcup LinearA\lg ebra) = n(LinearA\lg ebra) +$$
$$n(LinearA\lg ebra) - n(Analysis \bigcap LinearA\lg ebra) \Rightarrow 80 = 2 \cdot n(LinearA\lg ebra) - 20$$
$$\Rightarrow n(LinearA\lg ebra) = 50$$

29. D Systematic sampling occurs when you sample every nth person.

30. E The point $(15,26)$ is on the regression line. And since the data point $(3,6)$ has a residual of 4 the point $(3,2)$ is on the regression line because $Residual = y - \hat{y}$. Now that you have two points on the line, you can find the equation of the line.
$$a = 2 \ \& \ b = -4 \Rightarrow \sqrt[b]{a} \approx 0.840$$