Practice Problems SOLUTIONS: I recommend you try all problems yourself before reading the solutions. Email me if you see a mistake.

1. A family wishes to have 3 children. Let X be the number of girls.
   a. Find the probability distribution of X.

      X can take on the values 0,1,2,3 so I need to find the probabilities of each. There are two approaches.

      Approach 1: Since the probability of having a boy or girl is equally likely, I can write out the sample space and count the number of outcomes in each case for X and divide by the total number of outcomes in the sample space. S = {ggg, ggb, gbg, bgg, bbg, bgb, gbb, bbb}. So the probability distribution function is:

      p(0) = 1/8; p(1) = 3/8; p(2) = 3/8; p(3) = 1/8 .(Note it is good to check that this adds up to 1).

      Approach 2: I can think about X as a binomial random variable. A "success" is having a girl. There are n=3 trials, p=0.5 each time and the trials are independent. Hence we can use the formula for the probability distribution function for binomial:

      $p(x) = {}_nC_x(0.5)^x(0.5)^{n-x}$

   b. What is the expected number of girls?

      Approach 1: The formula for E[X] = 0(1/8) + 1(3/8) +2(3/8) +3(1/8) = 12/8 = 1 ½ girls. Note that we understand E[X] to be an average so even though 1 ½ girls doesn't make sense we don't round it, we leave it as 1 ½.

      Approach 2: Since X is binomial the expected value is just np = 3*(0.5) = 1 ½ girls.

2. Suppose that X is the random variable representing the number of days a patient needs to be in the hospital (where 4 is the maximum possible). Suppose X has the probability distribution function:

$$p(x) = \frac{5-x}{10} \ for\ x = 1,2,3,4.$$

   a. What is the probability that a patient will be in the hospital for exactly three days?

      I want the probability of x=3 days so I just plug 3 in for x in the formula for the probability distribution function: p(3) = 2/10.

b. What is the expected number of days the patient will be in the hospital?

I use the formula: E[X] = 1(4/10) + 2(3/10) + 3(2/10) + 4(1/10) = 20/10 = 2 days.

c. If the patient is to receive from an insurance company $200 for each of the first two days in the hospital and $100 for each day after the first two days, what is the expected payment for a hospitalization?

I need to figure out the possible outcomes for the payment and the probability of each outcome:
1 day in the hospital: $200 p(1) = 4/10
2 days in the hospital: $400 ($200 per day) p(2) = 3/10
3 days in the hospital: $500 ($200 for first two and $100 for third) p(3) = 2/10
4 days in the hospital: $600 ($200 for first two and $100 for second two) p(4)=1/10

Now I use the formula:
Expected payment = 200(4/10) + 400(3/10) + 500 (2/10) + 600 (1/10) = 3600/10 = $360.

3. It is believed that 20% of Americans do not have any health insurance. Suppose that this is true and let X equal the number with no health insurance in a random sample of n=15 Americans.

a. Explain why X has a binomial distribution. What are n and p?

This is binomial since we have a <u>fixed number of trials</u> (15 people), since we are using a random sample we can assume <u>the trials are independent</u> (this might not be true if our trial consisted of people who were in the same family, for example), and we can assume that the probability of any given person having no health insurance is 0.20 (<u>same prob. of success for each trial</u>) since even though the problem states that 20% of Americans have no health insurance and we are sampling without replacement, our sample is smaller than 5% of our population. n=15, p = 0.20

b. Find the probability that the number with no health insurance is <u>at most 3</u>.
(Note: AT MOST X means from 0 to X: 0,1,2,3…,X)

$P(X\leq3) = P(0)+P(1)+P(2) +P(3) = 0.035+0.132+0.231+0.250=0.648$
(I used the binomial table in the back of the book with n=15, p=0.2 and x=0,1,2,3.)

c. Find the probability that the number with no health insurance is <u>at least 6</u>.
(Note: AT LEAST X means X or more: X,X+1,…,N)

$P(X \geq 6) = P(6)+P(7)+\ldots+P(15) = 0.043+0.014+0.003+0.001=0.061$
(I used the binomial table in the back of the book with n=15, p=0.2 and x=6,7,8,…15)

d. What is the expected number of people in the sample who will have no health insurance? (e.g. the expected value of X)

This is binomial so I can use the formula $E[X] = np = 15(0.2) = 3$ people.

e. What is $P(2 \leq X \leq 5)$?

$P(2 \leq X \leq 5) = P(2) + P(3) + P(4)+P(5) = 0.231+0.250+0.188+0.103=0.772$
(I used the binomial table in the back of the book with n=15, p=0.2, x = 2,3,4,5)

f. What is the standard deviation of X?

Since this is binomial I can use the formula: $\sigma = \sqrt{15(0.2)(0.8)} = 1.55$ people

g. Would it be unusual for 6 of the people to have no health insurance? <u>Explain.</u>
(Note: Explain means explain. Use some mathematics to support your explanations.)

The z-score for 6 people is $z = \frac{6-3}{1.55} = 1.94$. This is <u>not unusual</u> because the z-score is not less than -2 or greater than 2 (i.e. 6 is fewer than 2 standard deviations below the mean).

4. If Z has a normal distribution with $\mu=0$ and $\sigma=1$, find:
   a. $P(Z \leq 1.25) = 0.8944$
   b. $P(Z \geq -0.36) = 1 - 0.3594 = 0.6406$
   c. $P(0.45 \leq Z \leq 2.79) = P(Z \leq 2.79) - P(Z \leq 0.45) = 0.9974-0.6736=0.3238$.
   d. Find the 45th percentile. $P_{45} \approx -0.13$
   e. Between which two values does the middle 80% of the data lie?
   The middle 80% of the data lies between -1.28 and 1.28.

   For all of the #4 problems I used the normal table.

5. Is X has a normal distribution with $\mu=15$ and $\sigma=8$, find:
   a. $P(X \geq 19) = P(Z \geq (19-15)/8) = P(Z \geq 0.5) = 1-0.6915 = 0.3085$. (normal table)
   b. $P(10 \leq X \leq 20) = P(Z \leq (20-15)/8) - P(Z \leq (10-15)/8) = P(Z \leq 0.63) - P(Z \leq -.63) = 0.7357-0.2643 = 0.4714$. (normal table)

c. Find the 80th percentile.

First I will look up the 80th percentile on the standard normal Z-curve using the table and then convert it to the X curve. $P_{80} = 0.84$. Then I use the formula $x = \mu + z\sigma = 15 + 0.84(8) = 21.72$. The 80th percentile is 21.72.

d. Between which two values does the middle 90% lie?

Again I will look this up on the z-table and then convert it. On the z-table the middle 90% lies between -1.645 and 1.645.
$15 + (-1.645)(8) = 1.84$
$15 + (1.645)(8) = 28.16$
So the middle 90% lies between 1.84 and 28.16

6. A teacher gave the same test to two sections of an Algebra class. Both sets of scores were normally distributed. The scores for Class 1 had a mean of $\mu=80$ and standard deviation of $\sigma=5$. The scores for Class 2 had a mean of $\mu=75$ and a standard deviation of $\sigma=10$.

   a. Carefully sketch both normal curves on the <u>same axis (label your axis well)</u>. (Pay attention to the height of each curve – which should be taller?)

   I'm not going to try to do it here, but the Class 1 curve would be taller and skinnier because it has a smaller standard deviation. It's peak would be shifted 5 to the right of class 2 since the mean is 5 points higher.

   b. A paper was misplaced. The teacher doesn't remember anything except that the score was more than 89. Which class is it more likely the test came from? Explain using probabilities.

To solve this I will compute the probability that the score was more than 89 for each class:
Class 1: $P(X \geq 89) = P(Z \geq (89-80)/5) = P(Z \geq 1.80) = 1 - 0.9641 = 0.0359$. (normal table)

Class 2: $P(X \geq 89) = P(Z \geq (89-75)/10) = P(Z \geq 1.40) = 1 - 0.9192 = 0.0808$. (normal table)

It is more likely that the test came from class 2 because the probability is higher.

7. Assume that the number of red M&M's in a regular size package follows a normal distribution with mean $\mu=5$ and standard deviation $\sigma=1.2$.

   a. What is the probability that my package of M&M's has 7 or more red M&M's?
   $P(X \geq 7) = P(Z \geq (7-5)/1.2) = P(Z \geq 1.67) = 1 - 0.9525 = 0.0475$. (normal table)

   b. Would it be unusual if my package only had 2 red M&M's?

I'll check the z-score: z = (2-5)/1.2 = -2.5. Yes, this would be unusual because it is more than 2 standard deviations below the mean.

c. If I bought 20 bags of M&M's what is the probability that the average number of red M&M's per bag would be 4.7 or less?

Note: I see the word "average" in this problem so I need to use the Central Limit Theorem to figure out the correct standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$.

$$P(\bar{X} \leq 4.7) = P\left(Z \leq \frac{(4.7-5)}{\left(\frac{1.2}{\sqrt{20}}\right)} \right) = P(Z \leq -1.12) = 0.1314$$

(normal table)

8. Assume that the weights of full-term newborn babies in the U.S. follows a normal distribution with mean μ=7.5 lbs and standard deviation σ= 0.94 lbs.
   a. Mercy Hospital in St. Louis delivered 50 babies in the month of November. What is the probability that the average weight of the babies was 7.65 lbs or less?

   (I want a probability about an average so use CLT)

$$P(\bar{X} \leq 7.65) = P\left(Z \leq \frac{(7.65-7.5)}{\left(\frac{0.94}{\sqrt{50}}\right)} \right) = P(Z \leq 1.13) = 0.8708$$

   (normal table)

   b. Mercy Hospital in St. Louis delivered 120 babies in the month of November. What is the probability that the average weight of the babies was 7.65 lbs or less?

   (Same problem, but the sample size has changed – this changes the standard deviation

$$P(\bar{X} \leq 7.65) = P\left(Z \leq \frac{(7.65-7.5)}{\left(\frac{0.94}{\sqrt{120}}\right)} \right) = P(Z \leq 1.75) = 0.9599$$

(normal table)

9. Students took n=35 samples of water from a lake and measured the amount of sodium in parts per million. Assume that σ = 4.94. From their data they found $\bar{x} = 24.11$.

a. Find a 90% confidence interval for μ, the mean amount of sodium in parts per million in the lake.

$$z_{\frac{\alpha}{2}} = 1.645; E = 1.645\frac{4.94}{\sqrt{35}} = 1.37;\ The\ confidence\ interval\ is\ (\bar{x} - E, \bar{x} + E)$$

= (24.11 – 1.37, 24.11 + 1.37) = (22.74,25.48). I am 90% confident

(24.11 – 1.37, 24.11 + 1.37) = (22.74,25.48). I am 90% confident that the amount of sodium in parts per million is in the interval (22.74,25.48).

b. What is the margin of error in your interval? 1.37 parts per million

c. The "Citizens for the Lake" group claimed that the mean amount of sodium in the lake is 20 parts per million. Is this claim supported by the confidence interval you found? Explain.

No, the claim is not supported. 20 parts per million is not in the confidence interval, the interval only contains higher numbers, therefore I am 90% confident that the mean amount of sodium in the lake is greater than 20 parts per million.

10. Researchers are trying to determine the average credit card debt, μ, of a student graduating from college. Suppose we know this follows a normal distribution. In a random sample of 30 graduating college students, the average credit card debt was found to be $\bar{x} = \$3{,}500$ and the sample standard deviation to be s= \$300. Find a 95% confidence interval for μ, the average credit card debt.

Since the population standard deviation is unknown, I need to use the t-table with 29 d.f. at the 95% confidence level (the d.f. is one less than the sample size):

$$t_{\alpha/2} = 2.045 \text{ then } E = 2.045\left(\frac{300}{\sqrt{30}}\right) = 112 \text{ so}$$

the confidence interval is $\left(\bar{x}\text{-E},\bar{x}\text{-E}\right) = (\$3388, \$3612)$

11. The Merchant's Association claims that the average outlet mall shopper will spend more than \$250 on the Friday after Thanksgiving (aka Black Friday). Let X be the amount spent by such a shopper and suppose it is known that the standard deviation of X is σ=200. In a random sample of 35 outlet mall shoppers on Black Friday, the average spending was found to be $\bar{x} = 270$.

a. Do a hypothesis test at the significance level of α=0.05 to determine if the Merchant's Association claim is true. Include a concluding statement in your answer

$H_0 : \mu = 250$

$H_1 : \mu > 250$ (the alternate hypothesis supports the claim)

The standard deviation is known so this is a z-test

$$z = \frac{270-250}{200/\sqrt{35}} = 0.59$$

It is a right tailed test with test significance level α=0.05 so the critical value is $z_\alpha = 1.645$

Since z=0.59 is less than 1.645 it is NOT in the critical region so

we fail to reject the null hypothesis.

Concluding statement: There is not enough data to support the claim that shoppers spend an average of more than $250 on Black Friday.

.

b. Find a 90% confidence interval for μ, the average spent by shoppers.

$z_{\alpha/2} = 1.645$

$$E = 1.645\left(\frac{200}{\sqrt{35}}\right) = 55.61$$

The confidence interval is $\overline{x} \pm E$:

$(214.39, 325.61)$

c. Do your answers to a. and b. agree? Explain why or why not.

Yes, the answers agree. In part a) we saw there was not enough evidence to reject the idea that the average spent was $250. In part b) we see that $250 is in our confidence interval which shows that there is a possibility that $250 is in fact the average.

12. It has been said that a majority of people living in Roswell, New Mexico, believe that we have been visited by extraterrestrial life. In a random sample of n=120 residents of Roswell, 81 said they believe we have been visited by extraterrestrial life. Test this claim to a significance level of α=0.10. Include a concluding statement in your answer.

If the <u>majority</u> of people believe then the claim is that p > 0.5.

$H_0 : p = 0.5$

$H_1 : p > 0.5$

So this is a right tailed test. The significance level is α=0.10 so the critical value is $z_{\alpha} = 1.28$. The point estimate for p is

$\hat{p} = \frac{81}{120} = 0.675$ so the test statistic is

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.675 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{120}}} = 3.83$$

$3.83 > 1.28$ so the test statistic is in the critical region so we reject the null hypothesis $H_{0.}$

Concluding statement: The data supports the claim that the majority of people living in Roswell, New Mexico believe that we have been visited by extraterrestrial life.