PROBLEM SET 1

For the first three answer true or false and explain your answer. A picture is often helpful.

1. Suppose the significance level of a hypothesis test is $\alpha=0.05$. If the p-value of the test statistic is p-value = 0.07, then the null hypothesis (H_0) should be rejected.

   False. We reject the null hypothesis when the p-value is less than α. But $0.07 > 0.05$ so we fail to reject H_0.

2. If we fail to reject the null hypothesis (H_0) that means that the test statistics was not in the rejection region.

   True. We only reject the null hypothesis if the test statistic is in the rejection region (also called critical region).

3. TRUE OR FALSE: If we fail to reject the null hypothesis (H_0) at a significance level of $\alpha=0.05$, then we also must fail to reject it at a significance level of $\alpha=0.10$.

   False. Consider the p-values. If we fail to reject H_0, then the p-value must be greater than α; so if $\alpha=0.05$, then the p-value > 0.05. However, this does not necessarily imply that the p-value must also be greater than 0.10. For example if the p-value = 0.08, then we would fail to reject H_0 at the significance level of $\alpha=0.05$ since $0.08 > 0.05$, but we would reject H_0 at the significance level of $\alpha = 0.10$ since $0.08 < 0.10$.

4. Suppose that we do reject the null hypothesis at a significance level of $\alpha=0.05$, but we **do not** reject the null hypothesis at a significance level of $\alpha=0.01$. What can you say about the p-value of the test statistic?

   This means that the p-value is less than 0.05 but greater than 0.01 (0.01 < p-value < 0.05)

5. Suppose we claim that the Candidate A will win the election against Candidate B. We take a sample and construct a confidence interval for p, the proportion of voters who will vote for Candidate A. The confidence interval is (0.489, 0.532). Does this confidence interval support the claim that Candidate A will win? Explain.

   No. In order to win Candidate A must get more than 50% of the vote (p must be greater than 0.5). However our confidence interval also contains values less than 0.5, so it is possible Candidate A will lose.

PROBLEM SET 2

1. A multiple choice test has 5 answers per question (a,b,c,d,e). Suppose you randomly guess on each question. If you get the question correct, you get 5 points, if you get the question wrong, you lose 2 points. If there are 3 questions, what is your expected point total for the test?

   There are 2 ways you could think about this problem.

   STRATEGY 1: The probability of randomly getting a problem correct is p=0.2. This is a binomial problem, X~ b(3,0.2). First figure out the possible point values you might get and the probability of each outcome:

Number Correct (+5 each)	0	1	2	3
Number wrong (-2 each)	3	2	1	0
Total Points (x)	-6	1	8	15
Probability of outcome (used binomial table (n=3,p=0.2)) (p(x))	0.512	0.384	0.096	0.008

   The expected value is the sum of the xp(x) which is:
   -6(0.512)+1(0/384)+8(0.096)+15(0.008) = -1.8 <u>You would expect to have -1.8 points if you randomly guessed each answer.</u>

   STRATEGY 2: The expected number correct is μ=np = 3 (0.2) = 0.6. So you expect to get 0.6 problems correct and 3-0.6 = 2.4 problems wrong.

If you get the problem correct you get 5 points. If you get it wrong you get -2 points so your expected number of points is

5(0.6) -2(2.4) = -1.8 points

2. Two cards are dealt at random from a standard 52 card deck without replacement.
   a. What is the probability that the first is a Jack and the second is a 3?

$$\frac{4}{52}\bullet\frac{4}{51}=\frac{16}{2652}=0.006$$

   b. What is the probability that the first is a face card and the second card is a club?

There are two possibilities to consider

1. The face card is a club and the second card is a club OR
2. The face card is not a club and the second card is a club

We compute both probabilities and add them up:
(Note: There are 13 clubs. There are 12 face cards 3 of which are clubs)

$$\frac{3}{52}\bullet\frac{12}{51}+\frac{9}{52}\bullet\frac{13}{51}=\frac{153}{2652}=0.058$$

3. I was not paying attention to where I was walking and I stepped on a very large bag of tortilla chips. The probability that any given chip in the bag was broken is p=0.25. Suppose I reach in and grab 10 chips. Let X be the number of cracked chips.
   a. Explain why X is binomial. What is n? p?

Even though we are sampling without replacement, we can consider the trials independent since the number of chips I am choosing (10) is likely less than 5% of the whole population of chips since I said it was a very large bag of chips. The outcome of each trial is only one of two choices: a success (cracked chip) or failure (not cracked). The probability of success for each trial is 0.25.

n=10, p = 0.25

   b. What is the probability that at most 3 of my chips are broken?
   (Note you can't use the tables unless the exact values for p are listed – don't average, use the formula if you can't use the table)

$P(X \leq 3) = P(0)+P(1)+P(2)+P(3) =$
${}_{10}C_0(0.25)^0(0.75)^{10}+{}_{10}C_1(0.25)^1(0.75)^9+{}_{10}C_2(0.25)^2(0.75)^8+{}_{10}C_3(0.25)^3(0.75)^7=$
$0.0563+0.1877+0.2816+0.2503=0.7759$

   c. What is the probability that at least 4 of my chips are broken?
   $P(X\geq4) = 1 - P(X\leq3) = 1 - 0.7759 = 0.2241$
   (I used the result for $P(X\leq3)$ from the last problem. I also could have added up $P(4)+P(5)+\ldots+P(10)$

PROBLEM SET 3

A disgruntled customer complained to a candy bar company that although the label weight of the candy bar was 58.7 grams, it actually weighed less. The company claims that the weights of the candy bars follow a normal distribution with mean $\mu = 58.7$ grams and standard deviation $\sigma=0.5$ grams.

1. Assuming the company's claim is true, what is the probability that a given candy bar will weigh less than 58 grams?

$$P(X \leq 58) = P\left(Z \leq \frac{58 - 58.7}{0.5}\right) = P(Z \leq -1.4) = 0.0808$$

2. The customer bought 20 candy bars. Find the probability that the average weight, $\bar{x}$, was less than 58.5 grams.

(Note this is AVERAGE weight so we use the central limit theorem to find the standard deviation) $P(\bar{X} \leq 58.5) = P\left(Z \leq \frac{58.5-58.7}{\frac{0.5}{\sqrt{20}}}\right) = P(Z \leq -1.79) = 0.0367$

3. Suppose that the customer's 20 candy bars had an average weight of $\bar{x} = 58.5$ grams. Do a hypothesis test to determine if his claim that the average is less than 58.7 grams is valid. State the null and alternate hypothesis. Use a significance level of $\alpha=0.05$ and clearly state your conclusion.

   H_0: $\mu=58.7$ grams H_1: $\mu<58.7$ grams

   This is a left-tailed test with significance level $\alpha=0.05$. The critical value is $z_\alpha = -1.645$ and the critical region is $z \leq -1.645$. The test statistic is

$$z = \frac{58.5 - 58.7}{\frac{0.5}{\sqrt{20}}} = -1.79$$

   Since $z < -1.645$ we reject H_0. Our conclusion is that the data supports the claim that the mean weight of the candy bars is less than 58.7 grams.

4. Again suppose that the customer's 20 candy bars had an average weight of $\bar{x} = 58.5$. Find a 90% percent confidence interval for the true mean, μ, of the candy bars. Assume $\sigma=0.5$ grams.

   $z_\alpha = 1.645$ so $E = 1.645\frac{0.5}{\sqrt{20}} = 0.1839$. The confidence interval for μ is (58.32,58.68).

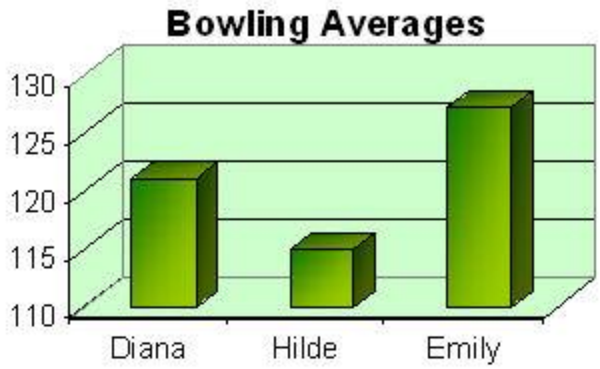
This says that I am 90% confident that the mean weight of the candy bars is between 58.32 grams and 58.69 grams.

5. Explain clearly why your confidence interval in 4 supports the result you got in 3.

My hypothesis result in (3) supported the claim that the true mean was less than 58.7 grams. This is supported by my confidence interval since all values in my interval are less than 58.7 grams.

PROBLEM SET 4

1. Why is this graph misleading?



The bars are not proportional. For example it looks like Emily has a bowling average over twice as much as Hilde, but really it is only about 12 points higher. The true differences would be more accurately represented if the y-axis started at 0.

2. Some students are comparing their scores on an exam when their friend walks in with a lower score than the min, in this case the RANGE:

   a. Decreases b. **<u>Increases</u>** c. Stays the Same d. None of the Above

3. The test score distribution for a class of about 50 students is approximately normal. The teacher finds four papers that she misplaced that <u>skew the data more to the left</u>. This means the four scores, compared to the rest of the class :

   a. Are much higher than the mean b. **<u>Are much lower than the mean</u>** c. Are about the same as the mean d. None of the Above

4. The test score distribution for a class of about 50 students is approximately normal. The teacher finds four papers that she misplaced <u>decrease the mean test score</u>. This means the average of the four scores, compared to the rest of the class:

   a. Are higher than the mean b. **<u>Are lower than the mean</u>** c. Are about the same as the mean

5. The test score distribution for a class of about 50 students is approximately normal. The teacher finds four papers that she misplaced <u>increase the median test score</u>. This means the four scores, compared to the rest of the class:

   a. **<u>Are higher than the median</u>** b. Are lower than the median c. Are about the same as the median

6. Suppose you go to Fry's Electronics to buy a CD. If the CD is marked down 20% to $11.20, what was the original price of the CD?

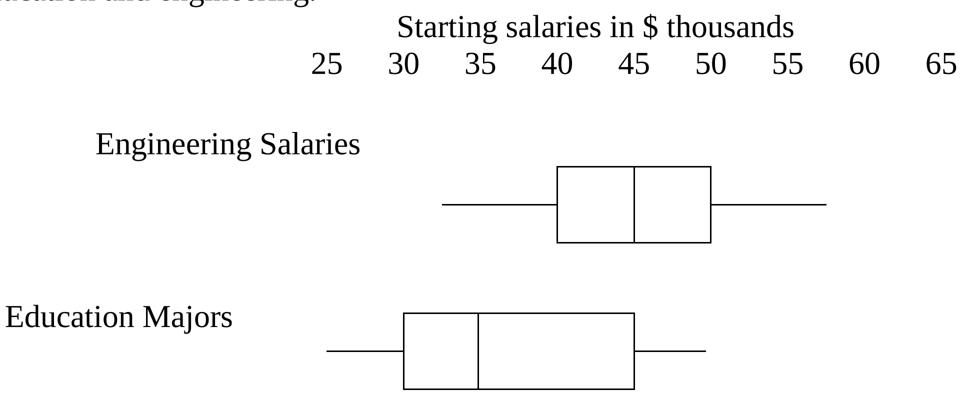
   a. $8.96 b. $11.40 c. **$14.00** d. $12.44

7. If the CD you bought at Fry's (from Question 6) is then marked back up to its original price from its discounted price of $11.20, what was the percentage increase from the discounted price to the original price. (Remember the CD was originally marked down 20%.)

   a. 20% b. **25%** c. 28% d. 30% e. 33 1/3 %

This information is for the next three problems

Use the box-and-whisker plots to compare the starting salaries for first year graduates in education and engineering.



8. If there were 504 engineering majors; how many of them had a starting salary $50,000 or more?

   a. 378 b. 252 **c. 126** d. Can't be determined

9. If there were 1200 education majors; what percent of them made $30,000 or less?

   a. 25 b. 300 c. 75 d. Can't be determined

10. Why is one part of the box in the plot for education salaries longer than the other part?

    a. There is more data between the upper part than the lower part.

    b. The data in the lower part is spread out more.

    c. There is less data between the upper part than the lower part.

    d. **The data in the upper part is more spread out.**