Lab4

1. In this example, X is a binomial random variable because:
2. We are using identical trials in 5 samples.
3. X has 2 possible outcomes, either it does use tap water it will be counted as x or it doesn’t in this case it will not be counted as x
4. The probability of x remains the same from trial to trial
5. One trial is independent of the other, since one outcome does not change the next
6. X is the number of Successes in the trails.
7. The probability distribution for x is as follows  
   p(x) = (n! / x! \* ((n-x)!) \* (p ^ x) \* (1 - p) ^ (n – x))
8. The probability for this trial where x = 2 is 0.263671875
9. The probability for this trial where x <= 1 is 0.6328125

In python I coded the following function:

import math

fact = math.factorial

def bimonrand(n, x, p):

y = (fact(n) / (fact(x) \* (fact(n - x)))) \* (p \*\* x) \* ((1 - p) \*\* (n - x))

return y

To return the following results:

>>> stats.bimonrand(5, 2, .25)

0.263671875

>>> a = stats.bimonrand(5, 1, .25)

>>> b = stats.bimonrand(5, 0, .25)

>>> a + b

0.6328125

LAB 5

The Test of Knowledge about Epilepsy (KAE), which is designed to measure attitudes towards persons with epilepsy, uses 20 multiple-choice items, all of which are incorrect. For each person, two scores (ranging from 0 to 20) are obtained: an attitude score (KAE-A) and a general-knowledge score (KAE-GK). On the basis of a large-scale study of college students, the distribution of KAE-A scores has a mean of   = 11.92 and a standard deviation of   = 2.95 while the distribution of KAE-GK scores has a mean of   = 6.35 and a standard deviation of   = 2.12 (Rehabilitative Psychology, Spring 1995). Consider a random sample of 100 college students and suppose you observe a sample mean KAE score of 6.5. Is this result more likely to be the mean of the attitude scores (KAE-A) or the general-knowledge scores (KAE-GK)? Explain.

19, 16, 13, 10, 7

10, 8, 6, 4, 2

Based on the large-scale study, a mean of 6.5 would be more likely to be the mean score of KAE-GK. The large scale KAE-GK Mean score is 6.35, which is less than 1 stand deviation from its mean. In comparison, the large-scale study found that the KAE-A mean score was 11.92. A Mean score of 6.35 would be 2 standard deviations away from the large scale study Mean.

LAB 6

*Aquatic Biology* (Vol. 9, 2010) reported on a study of green sea turtles inhabiting the Grand Cayman South Sound lagoon. The data on curved carapace (shell) length (in centimeters) for 76 captured turtles are saved in the TURTLES file. The time-depth recorders allowed the environmentalists to track the movement of sea turtles in the lagoon. These 6 turtles had a mean shell length of 52.9 cm with a standard deviation of 6.8 cm.

a) Use the information on the 6 tracked turtles to estimate, with 99% confidence, the true mean shell length of all green sea turtles in the lagoon. Interpret the result.

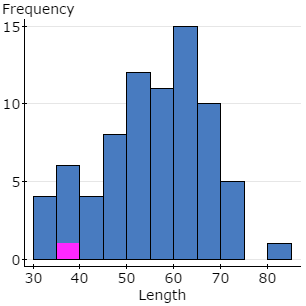
b) What assumption about the distribution of shell lengths must be true in order for the confidence interval, part a, to be valid? Is this assumption reasonably satisfied? Use the data in the TURTLES file to help you answer this question.

**Summary Statistics for a sample of 6 turtles:**

|  |  |
| --- | --- |
| **Mean** | 52.9 |
| **Std. dev.** | 6.8 |
| **L.Limit** | 41.7 |
| **U.Limit** | 64.1 |

In a sample of 6 turtles from a population of 76 with a Standard Deviation of 6.8 and mean of 52.9 we can be 99% confident that the mean shell length of all green turtles in the lagoon will be between 41.7 cm and 64.1 cm.

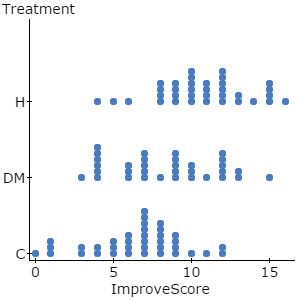
In order for this confidence interval to be valid the random sample must be selected from the target population. In this case the 6 turtles come from the population of 67 turtles. In addition, the population has to have a relative frequency distribution that is approximately normal. The below histogram shows the population of 67 turtles are approximately normal.



Lab 7

In this example we have 33 out of 105 children are using over the counter medicine labeled as DM , 35 out of 105 children are using honey labeled as H and 37 out of 105 are the control grouped labeled as CC in the Data sheet.

LAB 7



In this example we have 33 out of 105 children are using over the counter medicine labeled as DM , 35 out of 105 children are using honey labeled as H and 37 out of 105 are the control grouped labeled as C in the Data sheet. The large sample validation works with these current numbers as n^p !>= 15 and n^q !>= 15.

**Summary Statistics for Over-the-counter Medicine:**

|  |  |
| --- | --- |
| **Mean** | 8.333 |
| **Std. dev.** | 3.256 |
| **N** | 33 |
| **T-Interval** | 7.3729, 9.2931 |

**Summary Statistics for Honey:**

|  |  |
| --- | --- |
| **Mean** | 10.714 |
| **Std. dev.** | 2.855 |
| **N** | 35 |
| **T-Interval** | 9.924, 11.556 |

In both cases, the means are within the intervals. Honey as a narrower range between high and low intervals and therefore has a less of a probability to effect on children’s recovery.

LAB 8

|  |  |
| --- | --- |
| Sample Number | 100 |
| Standard Deviation | 10 |
| Mean | 50 |
| Null Mean Cases | 50, 47 |
| Z-score range for .05 rejection level | z < -1.96 or z > 1.96 |

With a risk level of .05 we can conclude that the test hypotheses with a null mean of 50, a standard deviation of 10 and a sample of 100, results in status quo with a description “the sample evidence is insufficient to reject” to avoid a Type II error. Status quo is always accepted until it is proven false.

With a risk level of .05 we can conclude that the test hypotheses with a null mean of of 10 and a sample of 100 is rejected. The true mean falls outside of two tailed alternative hypothesis z scores of z < -1.96 or z > 1.96. In this scenerio there is a .05 probability of a Type I error, where a null hypothesis is rejected when in fact it is true.

LAB 9

Pond's has discontinued the production of Age-Defying Complex, a cream with alpha-hydroxy acid, with Age-Defying Towlettes. Pond's advertised that the product could reduce wrinkles and improve the skin. In a study published in Archives of Dermatology (June 1996), 33 middle-aged women used a product with alpha-hydroxy acid for 22 weeks. At the end of the study period, a dermatologist judged whether each woman exhibited any improvement in the condition of her skin. The results for the 33 women (where I = improved skin and N - no improvement) are listed in the SKINCREAM file.

a) Do the data provide sufficient evidence to conclude that the cream will improve the skin of more than 60% of middle-aged women? Test, using    = .05.

b) Find and interpret the p-value of the test.

In this sample of 33 women, 24 had improved skin after using the Age-Defying Complex cream and 9 showed no improvement. We are to determine if data provides sufficient evidence to conclude that the cream will improve the skin of more than 60% of the middle-aged women.

First we have to determine if the sample size is large enough to use the normal approximation to the sampling distribution of p (npo and npq must be => 15). In this case, neither equation is >= 15:

.6 \* 24 = 14.4

.6 \* 9 = 5.4

In this scenario we use the sign test. If we use α = .05 where null theory is no improvement at 9 women, then P = P(x>= 9) = 1 – (p(x<= 8) = 1- .994 = **.006**. We get the .994 from Table II binomial table where f.n = 10 selecting 9 at .60.

Since p = .006 is less than α = .05 we reject the null hypothesis with a α = .05 level of significance.

Lab 10

In the state of Florida, elementary school performance is based on the average score obtained by students on a standardized exam, called the Florida Comprehensive Assessment Test (FCAT). An analysis of the link between FCAT scores and sociodemographic factors was published in the Journal of Educational and Behavioral Statistics(Spring 2004). Data on average math and reading FCAT scores of third graders, as well as the percentage of students below the poverty level, for a sample of 22 Florida elementary schools are listed in the FCAT file.

a) Propose a straight-line model relating math score (y) to percentage (x) of students below the poverty level. (Utilize "math editor" in your discussion forum for needed math symbols).

b) Use the method of least squares to fit the model to the data in the FCAT file.

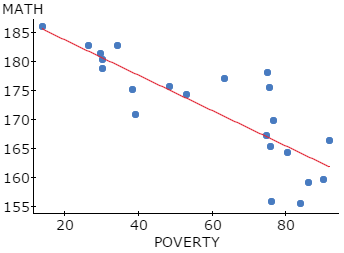
Reminder: remember to respond to at least 2 class members in order to satisfy discussion requirements.

c) Graph the least squares line on a scatterplot of the data. (You do not need to submit your graph.) Is there visual evidence of a relationship between the two variables? Is the relationship positive or negative?

d) Interpret the estimates of the y-intercept and slope in the words of the problem.

Y = β0 + β1x + ε where β0 is the Y intercept, β1x is the slope of the line, Y is the dependent or response variable and x is the independent or predictor variable. With the FCAT data we get:

y = 189.81582 – 0.30544451x



Visual inspection shows a negative relationship where math scores decreases by the amount of β1.