MATH-11X0 (DUPRÉ) 2014 SPRING TEST 2 ANSWERS

Suppose that the time X in minutes for a trolley to arrive is UNIFORMLY distributed with minimum value 10 and maximum value 60. Calculate:

1. $E(X) =$

ANSWER: For a uniformly distributed unknown, the mean is the average of the minimum and maximum values.

$$E(X) = \frac{1}{2}[Max + Min],$$

where Max denotes the maximum possible value and Min denotes the minimum possible value. This means here

$$E(X) = \frac{1}{2}[10 + 60] = 35.$$

2. $P(20 < X \leq 30) =$

ANSWER: For a uniformly distributed unknown, the probability density function is constant with height h given by

$$h = \frac{1}{Max - Min}.$$

For any values $b < c$ between the Max and Min,

$$P(b < X \leq c) = \frac{c - b}{Max - Min}.$$

This means here we have

$$h = \frac{1}{60 - 10} = \frac{1}{50}$$

and therefore

$$P((20 < X \leq 30) = \frac{30 - 20}{60 - 10} = \frac{10}{50} = \frac{1}{5} = 0.2.$$

3. $P(X > 25 \mid X \leq 40) =$

ANSWER: Using the law of conditional probability which is the multiplication rule applied to indicators,

$$P(X > 25 \mid X \leq 40) = \frac{P((X > 25) \ \&\ (X \leq 40))}{P(X \leq 40)} = \frac{P(25 < X \leq 40)}{P(X \leq 40)} = \frac{15/50}{30/50} = 0.5.$$

A box contains 4 red blocks and 6 blue blocks. We draw 5 blocks from the box at random and count the number T of red blocks. Calculate to four significant digits:

4. $E(T) =$

ANSWER: The success rate for drawing red blocks from this box is $p = 0.4$, since there are 10 blocks in the box and 4 are red. In a sample of size n, the expected total number of successes is

$$E(T) = np.$$

Here we have $n = 5$, so

$$E(T) = np = 5(.4) = 2.$$

5. Given drawing with replacement, σ_T=

ANSWER: Given drawing with replacement the success count T is binomially distributed, and we are using independent random sampling, so with $q = 1 - p$, we have

$$\sigma_T = \sqrt{npq}.$$

Here, $n = 5,\ p = .4$, and $q = .6$, so

$$\sigma_T = \sqrt{5(.4)(.6)} = \sqrt{1.2} = 1.095445115$$

or about 1.095, to three significant digits.

6. Given drawing without replacement, σ_T=

ANSWER: Given drawing without replacement, means that we are using simple random sampling instead of independent random sampling, and the success count is hypergeometrically distributed. We therefore need to multiply the SRS correction factor c_{SRS} by the result of the preceding problem:

$$c_{SRS} = \sqrt{\frac{N-n}{N-1}},$$

so here, $N = 10$ and $n = 5$, and

$$c_{SRS} = \sqrt{\frac{5}{9}} = \frac{\sqrt{5}}{3} = .7453559925.$$

This means that the standard deviation is now

$$\sigma_T = \frac{\sqrt{5}}{3}\sqrt{1.2} = \frac{\sqrt{6}}{3} = \sqrt{\frac{2}{3}} = 0.8164965809,$$

or about 0.816, to three significant digits.

7. $P(T = 2|\text{drawing with replacement}) =$

ANSWER: Given drawing with replacement, the distribution of the success count is binomial, so

$$P(T = k) = C(n, k)p^k q^{n-k}.$$

Here, as $n = 5,\ p = .4$, and $q = .6$, this means that

$$P(T = 2) = C(5, 2)(.4)^2(.6)^3 = (10)(.16)(.216) = .3456.$$

Alternately using the binomial distribution tables in the textbook gives

$$P(T \leq 2) = 0.683 \text{ and } P(T \leq 1) = .337,$$

giving

$$P(T = 2) = P(T \leq 2) - P(T \leq 1) = 0.683 - 0.337 = 0.346,$$

which is correct to three significant digits.

8. $P(T = 2|\text{drawing without replacement}) =$

ANSWER: Drawing without replacement, the success count has the hypergeometric distribution and we can simply work it out using elementary methods. The number of ways to draw five and get 2 red is simply $C(4, 2)C(6, 3)$ as the 2 red blocks must come from the 4 available and the remaining 3 blocks must then be form the 6 available blue blocks. There are $C(10, 5)$ ways to draw 5 blocks from a box containing 10 blocks and therefore

$$P(T = 2|\text{drawing without replacement}) = \frac{C(4, 2)C(6, 3)}{C(10, 5)} = \frac{(6)(20)}{252} = \frac{10}{21} = .4761904762,$$

or about 0.476 to three significant digits.

Suppose that we are studying the population of bears in Smokey Mountain National Park. We have an independent random sample of 9 bears from the population with a sample mean weight of 800 pounds and a sample standard deviation of 90 pounds. We assume that bear weight is normally distributed for bears in the population.

9. What is the MARGIN OF ERROR in the 99 percent confidence interval for the true mean weight of bears in the population if we know that the POPULATION standard deviation for the weight of bears in the population is 75 pounds?

ANSWER: Denoting the MARGIN OF ERROR by M and also using the fact that population standard deviation σ is known, we have that

$$M = z \cdot \frac{\sigma}{\sqrt{n}},$$

where n is the sample size, and z is the critical value which cuts off a right tail of area A, where

$$A = (1 - C)/2,$$

with C denoting the confidence level. Here, $C = 0.99$, so $A = .005$ which results in $z = 2.576$, from the t-table of critical values in the textbook using infinity for the degrees of freedom, as that is the normal distribution. Here $n = 9$. Therefore, here we have

$$M = z \cdot \frac{\sigma}{\sqrt{n}} = (2.576)\frac{75}{\sqrt{9}} = (2.576)(25) = 64.4.$$

10. If we know that the POPULATION standard deviation in bear weight is 75 pounds, does our sample data establish that the true mean weight of bears exceeds 725 pounds at the .005 significance level? Give the value of the standardized test statistic for the sample data and give the P-Value or significance of the data.

ANSWER: The hypothesis test is

$$H_0 : \mu = 725$$

$$vs$$

$$H_1 : \mu > 725$$

so the standardized test statistic is, using μ_0 for the null hypothesis value of the true mean,

$$\text{test statistic } = z_{data} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}},$$

and the P-Value or data significance is

$$\text{P-Value } = \text{ data significance } = P(Z \geq z_{data}),$$

where Z is the standard normal random variable.

Here $\mu_0 = 725$, $\bar{x} = 800$, and $\sigma = 75$, so

$$\text{test statistic } = z_{data} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{800 - 725}{75/\sqrt{9}} = \frac{75}{25} = 3,$$

and

$$\text{P-Value } = \text{ data significance } = P(Z \geq z_{data}) = P(Z \geq 3) = P(Z \leq -3) = 0.0013.$$

Since $0.0013 < 0.005$, we conclude, YES, our data does establish the true mean weight of bears exceeds 725 pounds at the 0.005 significance level.

11. What is the MARGIN OF ERROR in the 99 percent confidence interval for the true mean weight of bears in the population if we DO NOT know that the population standard deviation in weight of bears is 75 pounds but instead use our sample standard deviation of 90 pounds?

ANSWER: Since here we do not know the population standard deviation, we must use the sample standard deviation in its place, so the MARGIN OF ERROR, denoted M is given by

$$M = t \cdot \frac{s}{\sqrt{n}},$$

where s is the sample standard deviation and n is the sample size, and t is the critical value cutting off a right tail of area A given by

$$A = (1 - C)/2,$$

given in the t-table for $df = n - 1$ degrees of freedom. Of course, C is the confidence level.

Here, $s = 90$, $n = 9$, $C = 0.99$ so $df = 8, A = .005$, and from the t-table of critical values, $t = 3.355$.

This gives

$$M = t \cdot \frac{s}{\sqrt{n}} = (3.355)\frac{90}{\sqrt{9}} = (3.355)(30) = 100.65,$$

or 100.7, to three significant digits.

Suppose we are comparing mean weight of blue fish and red fish and we assume both populations have the same standard deviation, σ, in weight. Suppose we have a sample of 5 blue fish with variance 12 and a sample of 7 red fish with a variance of 8.

12. The pooled variance estimate (using this sample data), denoted S^2_{pool}, of the true population variance is $S^2_{pool} =$

ANSWER: To pool the variances, we simply average them according to degrees of freedom,

$$S^2_{pool} = \frac{(df_1)s_1^2 + (df_2)s_2^2}{df_1 + df_2}.$$

If we think of the bluefish as the first population and the redfish as the second population, then we have

$$df_1 = df_B = n_B - 1 = 5 - 1 = 4,$$

$$df_2 = df_R = n_R - 1 = 7 - 1 = 6,$$

$$s_1^2 = S_B^2 = 12,$$

$$s_2^2 = S_R^2 = 8.$$

Substituting into the formula then gives

$$S^2_{pool} = \frac{(df_1)s_1^2 + (df_2)s_2^2}{df_1 + df_2} = \frac{(4)(12) + (6)(8)}{4 + 6} = 9.6.$$