MATH-1140 (DUPRÉ) SPRING 2016 LECTURE QUIZ 5 ANSWERS

1. PRINT YOUR LAST NAME IN THE UPPER RIGHT CORNER IN LARGE CAPITAL LETTERS.

2. PRINT YOUR FIRST NAME UNDERNEATH YOUR LAST NAME IN THE UPPER RIGHT CORNER IN CAPITAL LETTERS.

3. PRINT YOUR LAB DAY AND LAB START TIME UNDERNEATH YOUR FIRST NAME IN THE UPPER RIGHT CORNER.

4. WRITE YOUR SPRING 2016 MATH-1110 COURSE SECTION NUMBER UNDERNEATH YOUR LAB DAY IN THE UPPER RIGHT CORNER.

For the following problems suppose that X is the weight in GRAMS of a fish in my pond, and that X is normally distributed. The word fish here always means a fish from my pond.

5. If we assume the true mean of X is $\mu_X = 100,$ and $\sigma_X = 10,$ then what is the fish weight in GRAMS that divides the lower 95 percent from the upper 5 percent in fish weight?

ANSWER: The score x we are looking for when standardized must give the critical value z which cuts off a right tail of area .05 on the standard distribution. This is quickly found in the $t-$table of critical values using infinity for the degrees of freedom and $z = 1.645.$ To find x we just de-standardize z:

$$x = \mu_X + z\sigma_X = 100 + (1.645)(10 = 100 + 16.45 = 116.45.$$

6. What is the chance a fish's STANDARD score on the weight scale is between 1.23 and 1.84?

ANSWER: With Z denoting the standard normal random variable, the problem simply asks for the probability that $1.23 \leq Z \leq 1.84.$ The cdf for Z, denoted F_Z is tabulated in your textbook via the right body table which allows F_Z to be easily computed from symmetry of the bell curve, so if RB denotes the right body area, then $F_Z(z) = .5 + RB(z),$ if $z \geq 0,$ and

$$P(1.23 \leq Z \leq 1.84) = RB(1.84) - RB(1.23) = .4671 - .3907 = .0764$$
$$= F_Z(1.84) - F_Z(1.23) = .9671 - .8907 = .0764.$$

7. If we know $\sigma_X = 3,$ then what is the MARGIN OF ERROR in the 99 percent CONFIDENCE INTERVAL for estimating the true mean weight μ_X using the sample mean of a sample of 9 fish?

ANSWER: If we denote the margin of error by $M,$ then as we know the population standard deviation here, we have $M = z\sigma_X/\sqrt{n},$ where n is the sample size and z is the critical value of the standard normal distribution which cuts off an upper tail of area .005 in order to have 99 percent confidence. We look up this last value in the $t-$table of critical values using infinity for the degrees of freedom and find $z = 2.576.$ Thus,

$$M = z\frac{\sigma_X}{\sqrt{n}} = (2.576)\frac{3}{\sqrt{9}} = 2.576.$$

8. What is the MARGIN OF ERROR in the 99 percent CONFIDENCE INTERVAL for estimating the true mean weight μ_X using the sample mean of a sample of 9 fish if we only know the sample standard deviation is $s = 3$?

ANSWER: Again, denoting the margin of error by M, we now have an unknown population standard deviation and only know the sample standard deviation. Keep in mind that any time you get the sample data, you will always have the sample standard deviation, but it is only when you do not know the population standard deviation that you would be forced to use the sample standard deviation in place of that for the whole population. For a sample of size 30 or more we often simply ignore the difference between the two, and thus if our sample sizes in this and the previous problem had been 90 instead of 9, we could say both problems have the same answer. But for a sample of only size 9, the difference between knowing the population standard deviation and only knowing that of the sample is substantial. For this problem you must use t in place of z and s in place of σ_X and here the $t-$distribution has $n - 1 = 8$ degrees of freedom. Looking up the critical value of t for 8 degrees of freedom which cuts off the right tail of area .005, we find $t = 3.355$. We now have

$$M = t\frac{s}{\sqrt{n}} = (3.355)\frac{3}{\sqrt{9}} = 3.355.$$

Notice there is a substantial increase in the margin of error due to simply not knowing the population standard deviation, even though numerically, the standard deviations in both problems is simply 3. In general, less information causes a larger margin of error.

9. What is the critical value, z, of the standard normal Z with the property that

$$P(Z \leq z) = .01?$$

ANSWER: By symmetry of the standard normal distribution, we have

$$P(Z \leq z) = P(Z \geq -z).$$

From the $t-$table of critical values for infinity degrees of freedom, we see that to have a right tail of area .01, the critical value is 2.326, so we have $-z = 2.326$ and therefore $z = -2.326$.

10. If we hope to establish that the true mean weight of fish exceeds 100 GRAMS with a sample of 9 fish, then what is the critical value the test statistic, that is, what is the critical value the $t-$score of our data, must exceed, for us to succeed in this, at level of significance .05?

ANSWER: The test statistic for this hypothesis test is simply

$$t_{data} = \frac{\bar{x} - 100}{s/\sqrt{9}},$$

where here we are dealing with the $t-$distribution for $n - 1 = 8$ degrees of freedom. To establish that $\mu > 100$ at the $\alpha = .05$ level of significance means that the significance of our data must be numerically smaller than .05. The significance or P-value of our data here is

$$P - Value = P(t \geq t_{data}),$$

which is the area of the right tail cut off by t_{data}, the $t-$score of our data. In order for this tail to have area no more than .05, its cutoff point must be to the right of the critical value in the table of critical values which cuts off a tail of area .05. For 8 degrees of freedom, the table gives this critical value as 1.860. Thus, the final answer is simply 1.860 from the table. Notice, that as soon as you know what you are trying to establish and the level of significance, the critical value you have to beat with your data is given in the table and you know it before even getting the sample data. If we had been trying to establish that $\mu < 100$, the result we need would have to be in the left tail of the distribution, so by symmetry, the critical value would be -1.860. If you were simply trying to prove that $\mu \neq 100$, then you have to allow for the possibility that the data can now go on either side of 100. For instance, in this case, if two samples happen to have the same standard deviation but one has mean 90 and the other has mean 110, then both have the same power to contradict the null hypothesis that $\mu = 100$. Thus, the significance of the data has to include both tail areas and thus the significance of the data is now

$$P - Value = 2P(t \geq |t_{data}|),$$

so for this to be no more than .05 would mean that $P(t \geq |t_{data}|) \leq .025$, so in this case, the critical value would be the critical value which cuts off a tail of only area .025 instead of the .05 from before. In this case, we say we have a "two tail test", so we have to allow each tail to have half the area allowed in the level of significance. The condition in the two tail test for the data to prove $\mu \neq 100$ is simply that

$$|t_{data}| \geq t_{\alpha/2}, \text{ for 8 degrees of freedom}.$$

In our problem, the critical value from the table cutting off a tail of area .025 is 2.306, so if the absolute value of the test statistic is 2.306 or more, then we have proven that the population mean is not 100 at level of significance .05. Notice this means that if we have two researchers using the same data and the $t-$score of the data is between 1.860 and 2.306, and if one of the researchers is trying to prove the mean is greater than 100, then he has proof, whereas if the other was only trying to prove the mean is not 100, then the other researcher fails to have proof. This is an inherent problem with the logic of hypothesis testing. We have two people using the same data so if one proves $\mu > 100$ you would think that he has certainly proven that $\mu \neq 100$, in particular, so the other researcher's failure seems logically inconsistent. If we ask the second researcher to try and guess in advance of looking at the data which one tail test he should have tried, then he has 50 percent chance of guessing the correct tail which will work. Remember, the significance of the data is the probability that the data can fool you into thinking you have proven your alternative hypothesis given the null hypothesis is actually true. When working at level of significance α, this means that you are allowing the probability α that you will think you have proven something when in fact you have not. So, if the null hypothesis that $\mu = 100$ is true, there is a chance t_{data} happens to be in the range between 1.860 and 2.306, then the researcher who is trying to prove $\mu > 100$ has just made a type I error whereas the researcher trying to prove $\mu \neq 100$ just thinks the data to be inconclusive. So the problem cuts both ways.