MATH-1230 (DUPRÉ) TEST 2 SPRING 2018 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 MATH COURSE AND SECTION NUMBERS UNDERNEATH YOUR LAB INFORMATION IN THE UPPER RIGHT CORNER.

THE REMAINING PROBLEMS ALL USE THE FOLLOWING. READ THIS INFORMATION CAREFULLY.

Suppose that SAM has a pond containing EXACTLY 1000 fish, and SAM KNOWS that the fish in his pond have an average weight of 100 grams with a standard deviation of 12 grams. Of the 1000 fish in Sam's pond, SAM KNOWS 200 are redfish and 800 are bluefish. SAM decides to RANDOMLY catch 100 fish from his pond WITH REPLACEMENT, forming an INDEPENDENT RANDOM SAMPLE in the FUTURE which is henceforth called SAM'S FUTURE SAMPLE.

JOE does NOT KNOW anything in advance about the weights of the fish in Sam's pond. JOE weighs THIRTYSIX fish from Sam's pond, one after another, WITH REPLACEMENT (or, catch and release, in fisherman terminology) forming an INDEPENDENT RANDOM SAMPLE but Sam does not see which fish Joe weighs or the sample results. JOE found that the sample MEAN for his THIRTYSIX fish is 110 GRAMS, and the STANDARD DEVIATION for his SAMPLE is 18 GRAMS. Moreover, JOE observed that 9 of the fish in his sample were redfish and the rest were bluefish.

Of JOE'S sample, SAM only knows that JOE'S sample is an INDEPENDENT RANDOM SAMPLE of size 36.

ALL ANSWERS BELOW will use X for the weight of a fish from LAKE WOBEGONE, will use σ_e for standard error, and $SD(Y) = \sigma_Y$ for the population standard deviation of Y. We use $\bar{X}_n$ to denote the sample mean random variable for samples of X of size n, and $\bar{x}$ to denote an observed value of it.

5. What should SAM EXPECT is the AVERAGE weight of the THIRTYSIX fish Joe SAMPLED?

ANSWER: In general, for any random sampling method of any unknown Y, we know

$$E(\bar{Y}_n) = E(Y) = \mu_Y.$$

Therefore, SAM should expect that JOE's sample mean is simply

$$\mu_X = 100.$$

6. What should SAM EXPECT is the SQUARED ERROR in his expected AVERAGE weight of the THIRTYSIX fish Joe SAMPLED?

ANSWER: As the standard error of the mean is

$$\sigma_e = SD(\bar{X}_n) = \frac{\sigma_X}{\sqrt{n}},$$

the squared error SAM should expect here is σ_e^2, the result of squaring the previous expression, so the result is

$$\sigma_e^2 = \frac{\sigma_X^2}{n} = \frac{12^2}{36} = (12/6)^2 = 2^2 = 4.$$

7. If SAM tells JOE that his favorite fish weighs between 90 and 120 grams, but nothing else, so JOE has no idea what fish SAM refers to, then what should JOE think is the distribution in weight of THAT FISH, that is what is the NAME of ITS DISTRIBUTION?

ANSWER: any time a person only knows the minimum and maximum possible values of an unknown, for that person the distribution of that unknown is UNIFORM.

8. If SAM tells JOE that his favorite fish weighs between 90 and 120 grams, but nothing else, what does JOE EXPECT for that weight?

ANSWER: For any unknown Y with uniform distribution having minimum m and maximum M, the expected value is simply the average of m and M, so

$$\mu_Y = \frac{m+M}{2} = m + \frac{M-m}{2}.$$

This means that for JOE the expected weight of that fish is

$$\mu = \frac{90+120}{2} = 90 + \frac{30}{2} = 105.$$

9. If Sam tells Joe that his favorite fish weighs between 90 and 120 grams, but nothing else, what should Joe think is the probability Sam's favorite fish weighs between 90 and 110 grams?

ANSWER: For any unknown Y with uniform distribution having minimum m and maximum M, and for any numbers a, b with $m \leq a \leq b \leq M$,

$$P(a \leq Y \leq b) = P(a < Y < b) = \frac{b-a}{M-m},$$

so here,

$$P(90 < X < 110) = \frac{110-90}{120-90} = \frac{20}{30} = \frac{2}{3}.$$

10. HOW MANY of the fish in SAM'S FUTURE SAMPLE does SAM EXPECT will turn out to be REDFISH?

ANSWER: Sam knows his pond has only 200 redfish out of the 1000 fish in the pond, so each time a fish is caught it has a probability of 1/5 of being a redfish. Thus, the success rate $p = .2$ **for catching redfish. As SAM's sample is an independent random sample and therefore a random sample of size** $n = 100$, **the expected number of redfish is**

$$\mu = np = (100)(1/5) = 20.$$

11. For SAM'S FUTURE SAMPLE what is the standard deviation FOR SAM in the NUMBER of redfish he will get?

ANSWER: SAM's sample is not just a random sample, it is an independent random sample, so his count has the binomial distribution which has standard deviation

$$\sigma = \sqrt{np(1-p)},$$

which here is

$$\sigma = \sqrt{(100)(.2)(.8)} = \sqrt{16} = 4.$$

Keep in mind, the redfish count is simply the sample total T_n **for the indicator** I_R **of a redfish on a single catch, and the sample is an independent random sample of size** $n = 100$ **of the random variable** $Y = I_R$. **But for any unknown** Y, **we know the sample total** T_n **for random samples of size** n **in case of independent random sampling has standard deviation given by**

$$SD(T_n) = \sqrt{n} \cdot SD(Y),$$

and the variance of Y **in case** $Y = I_R$ **as then** $Y^2 = Y$, **is simply**

$$Var(Y) = E(Y^2) - E(Y)^2 = E(Y) - E(Y)^2 = p - p^2 = p(1-p).$$

The result is then

$$Var(T_n) = np(1-p).$$

12. What is the PROBABILITY (for SAM) that exactly 25 of the fish in SAM'S FUTURE SAMPLE turn out to be REDFISH?

ANSWER: Since SAM's sample is going to be an independent random sample, the redfish count has the binomial distribution, so the probability of 25 of the 100 fish turning out to be redfish is

$$P(T_{100} = 25) = C(100, 25)(.2)^{25}(.8)^{75},$$

where $C(100, 25)$ denotes "100 choose 25",

$$C(100, 25) = \frac{100!}{25!75!}.$$

13. If JOE uses his SAMPLE DATA, then what should Joe think is the STANDARD ERROR he must use to compute the margin of error for his 93 percent confidence interval for the true mean weight of all the 1000 fish in the pond, if JOE ASSUMES that the DISTRIBUTION of fish weight in the pond is NORMAL?

ANSWER: Since JOE does not know the population standard deviation, he must use his sample standard deviation in its place in calculating the standard error so in reality his standard error is only an estimate based on his sample data. He will use

$$\sigma_e(\textbf{estimate}) = \frac{s}{\sqrt{n}} = \frac{18}{\sqrt{36}} = \frac{18}{6} = 3$$

for the standard error used in calculating the margin of error, no matter what confidence level is used.

14. If JOE found that the sample mean for the FIRST SIX fish of his THIRTYSIX fish sample is 90 grams, and if the sample standard deviation for the FIRST SIX fish in the sample is 24 grams, then what should Joe think is the TRUE NUMBER OF DEGREES OF FREEDOM in the student t-distribution that governs the statistic used to compute the margin of error for his 99 percent confidence interval for the true mean weight of all the 1000 fish in the pond BASED ONLY ON the FIRST SIX FISH, if JOE ASSUMES that the DISTRIBUTION of fish weight in the pond is NORMAL?

ANSWER: If JOE only uses the first 6 fish, then his sample size is only $n = 6$, so the number of degrees of freedom he must use is $v = n - 1 = 5$. Keep in mind that when JOE does not know the population standard deviation, in confidence intervals and hypothesis tests, he is replacing σ_e with an estimate he gets using s in place of σ_X which means his standardization only results in a student $t-$distribution, not the standard normal, and this distribution has $n - 1$ degrees of freedom. If n is larger than 30, it is acceptable to ignore this fact and go ahead and use the standard normal, but for such a small n as $n = 6$, one must use the student $t-$distribution.

15. If JOE uses his SAMPLE DATA, then what should Joe think is the STANDARD ERROR he must use to compute the margin of error for his 93 percent confidence interval for the true mean weight of all the 1000 fish in the pond, if JOE ASSUMES that the DISTRIBUTION of fish weight in the pond is NORMAL AND that the POPULATION standard deviation is 12 grams because SAM TOLD HIM so?

ANSWER: Now JOE knows the population standard deviation so he calculates the precise standard error, exactly the same as SAM would for a sample of size $n = 36$, **namely**

$$\sigma_e = \frac{12}{\sqrt{36}} = \frac{12}{6} = 2.$$

16. Since 9 of the THIRTYSIX fish Joe caught are redfish, what should JOE guess is the TRUE PROPORTION of redfish in Sam's pond, IGNORING any margin of error?

ANSWER: Since JOE does not know the true proportion of redfish in the pond, he can only use the proportion in his sample for his guess for the true proportion, that is he guesses the value $\hat{p}$ **from his sample data,**

$$\textbf{Guess } (p) = \hat{p} = \frac{9}{36} = \frac{1}{4} = 0.25.$$

17. Since 9 of the THIRTYSIX fish JOE caught are REDFISH, what is the STANDARD ERROR he should use to compute the 99 percent confidence interval for the true proportion of redfish in Sam's pond based on JOE'S sample DATA ONLY?

ANSWER: There are 2 count them, TWO possible answers here that are acceptable. The true standard error here is the standard deviation of $\hat{p}$ but we know that using $Y = I_R$ for the indicator of a redfish catch, $Var(Y) = p(1-p)$, so

$$Var(\hat{p}) = \frac{p(1-p)}{n}.$$

This means the true standard error is

$$SD(\hat{p}) = \sqrt{p(1-p)/n},$$

whose evaluation would require we know what we are sampling to try and determine. The logically correct way to proceed is to be conservative and use the fact that

$$SD(I_R) \leq .5,$$

for sure, a fact that is true of any indicator. In general for any situation of trying to estimate a standard error when you do not know σ, if you happen to know for sure that $\sigma \leq b$, for some number b, then for sure you can be safe to use b in place of σ as it only increases the margin of error resulting by using b in place of σ in calculating σ_e. Thus a conservative statistician would use $\sigma_e = (.5)/\sqrt{n}$, which here gives

$$\sigma_e = \frac{.5}{\sqrt{36}} = \frac{.5}{6} = \frac{1}{12} \textbf{ and } \sigma_e^2 = \frac{1}{144}.$$

For LARGE samples used in practice for proportion sampling say in political polling, the sample sizes are typically hundreds or better, thousands, so the value of $\hat{p}$ from the sample data is usually close to p and the standard error formula is not very sensitive to small variations of the value of $\hat{p}$ as can be seen by noticing that the variance as a function of p is a simple quadratic whose graph is a parabola, downward with axis of symmetry at .5 and which passes through zero at $p = 0$ and $p = 1$, so as long as p is not far from .5, the tangent slope is not too steep and small changes do not make much difference. The upshot is that most textbooks simply say use $\hat{p}$ in place of p in the standard error calculation which would give

$$\sigma_e^2(\textbf{estimate}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{(.25)(.75)}{36} = \frac{3}{4^2 6^2} = \frac{1}{(12)(16)} = \frac{1}{192} < \frac{1}{144},$$

and therfore

$$\sigma_e(\textbf{estimate}) = \frac{1}{\sqrt{192}} < \frac{1}{12}.$$

Again, we will accept both answers here, so unless told otherwise, in this situation, you can take your pick. To my way of thinking, the use of the upper bound of $SD(\hat{p}) \leq .5$ vastly simplifies the calculation and leads to a logically clearer result.

18. What should JOE think is the value of the TEST STATISTIC he will compute in seeking to PROVE THE HYPOTHESIS that the true mean weight of all the 1000 fish in the pond, is over 90 grams, if JOE ASSUMES that the DISTRIBUTION of fish weight in the pond is NORMAL?

ANSWER: Using TS for test statistic, JOE computes standardization of $\bar{x}$ by using the assumed null hypothesis value of the true mean, $\mu_0 = 90$, and uses his sample standard deviation $s = 18$ in place of $\sigma_X = 12$ in calculating his estimate of the standard error σ_e, so for JOE, his test statistic for testing the null hypothesis $\mu_X = 90$ against any alternative, is simply

$$TS = \frac{\bar{x} - \mu_0}{\sigma_e} = \frac{110 - 90}{18/\sqrt{36}} = \frac{6(20)}{18} = \frac{20}{3}.$$

19. If JOE uses his SAMPLE DATA, then what should Joe think is the value of the TEST STATISTIC he will compute in seeking to PROVE the true mean weight of all the 1000 fish in the pond, is over 90 grams, based on his sample data, if JOE ASSUMES that the DISTRIBUTION of fish weight in the pond is NORMAL AND the POPULATION standard deviation is 12 GRAMS because SAM TOLD HIM so?

ANSWER: As soon as JOE knows the population's standard deviation, he no longer needs to use his sample standard deviation so instead of estimating the standard error, he calculates the true standard error, simply using 12 in place of the 18 in the previous problem, so the test statistic is now

$$TS = \frac{\bar{x} - \mu_0}{\sigma_e} = \frac{110 - 90}{12/\sqrt{36}} = \frac{(20)}{2} = 10.$$

20. Suppose that ANOTHER SAMPLE of 6 redfish in the pond had a mean weight of 110 grams with sample standard deviation 12 and ANOTHER SAMPLE of 10 bluefish had a sample mean weight of 95 grams with a standard deviation of 20, then what would be the STANDARD ERROR ESTIMATE used to calculate the margin of error in a 95 percent confidence interval for the amount that true mean redfish weight exceeds the true mean bluefish weight for the fish in Sam's pond?

ANSWER: Since the samples are independent, the estimate of $\mu_R - \mu_B$ is accomplished with the statistic $\bar{X}_R - \bar{X}_B$, the difference in sample means, whose variance is the sum of the variances of each of the terms,

$$Var(\bar{X}_R - \bar{X}_B) = Var(\bar{X}_R) + Var(\bar{X}_B) = \frac{Var(X_R)}{n_R} + \frac{Var(X_B)}{n_B}.$$

But we do not know the population standard deviations for the redfish or the bluefish. This means the sample variances must be used in estimating the standard error, and this results that the standardization of the difference in sample means will be a $t-$distribution, not the standard normal. Since they are both part of the population of the fish in the pond, one might get away with assuming both standard deviations are equal and use the pooled variance in place of each true variance which allows us to add the degrees of freedom. The pooling of the variances is done according to degrees of freedom, so if you make the assumption of equal variances, then you must use the pooled variance

$$S_{pool}^2 = \frac{(n_R - 1)s_R^2 + (n_B - 1)s_B^2}{n_R - 1 + n_B - 1} = \frac{(5)(144) + (9)(400)}{14},$$

and then

$$\sigma_e^2 = S_{pool}^2((1/6) + (1/10)).$$

This means

$$\sigma_e = \sqrt{\frac{(5)(144) + (9)(400)}{14}}\sqrt{(1/6) + (1/10)}$$

But, if you do not make the assumption of equal variances then you would use the Satterthwaite formula to calculate the degrees of freedom and simply use the sample variances in place of population variances giving,

$$\sigma_e^2 = \frac{12^2}{6} + \frac{20^2}{10} = 24 + 40 = 64 = 8^2$$

giving

$$\sigma_e(\textbf{estimate}) = \sqrt{\frac{12^2}{6} + \frac{20^2}{10}} = 8.$$

Since we do not need to calculate the degrees of freedom to answer this problem, dispensing with the assumption of equal variances makes for a simpler calculation. To proceed here with a confidence interval using the standard error estimate of 8, one can use the minimum of the two degrees of freedom which is 5 and be safe, and avoid the use of the Satterthwaite formula for calculating degrees of freedom. One should keep in mind that the variance estimate here does not strictly speaking obey a chi-square distribution so it is only approximately chi-square any way, but for the final exam, you have to know how to compute it. As the standard error estimate does not strictly obey a chi-square distribution in case the variances are not equal, that means the standardization of the difference in means used is not strictly speaking a $t-$distribution UNLESS we make the assumption of equal population variances. Remember, if the problem does not explicitly state which assumption is being made, you can choose which to make.