True/False

For each of the following questions indicate true or false, then **fully explain** your answer. You may use examples to illustrate your answer.

(I) To create a confidence interval for $\mu_1 - \mu_2$ (a two sample problem), we must assume that the sample sizes n_1, n_2 are above 30.

FALSE. We need a random sample, and that either the populations are normally distributed, or the sample sizes are above 30 (but not both).
Note: An answer that states it should be greater than or equal to 30 is still incorrect, as it is missing the main point of the question. Recall the assumptions are:

1. A random sample was taken from both populations
2. The groups are independent
3. $\bar{Y}_1$ and $\bar{Y}_2$ are normally distributed, **which requires one of the two following assumptions**
   - The populations are normal, **or**
   - The sample sizes are each ≥ 30

   Another reasonable answer is to say false and list all the assumptions. But in cases where the population is normally distributed, it is not necessary to have a sample size ≥ 30.

(II) The calculation of the p-value depends on your value of α.

FALSE. The p-value calculation has to do with the test-statistic, but not the value of α. We compare the p-value to α, however. For example, to calculate a p-value for an ANOVA problem, the formula is $Pr\{F > F_s\}$, which is not effected by the value of α.

(III) The larger the value of t_s in a two-sided test for $\mu_1 - \mu_2$, the less our sample data agrees with H_0.

TRUE. This is because the test-statistic is the number of estimated standard deivations our data is from the hypothesized value under the null hypothesis, so the larger the absolute value of t_s the further our sample means are from the hypothesized value.

(IV) In general, the larger the value of MSW, the more variance there was within each group.

TRUE. You can see this by looking at the equation: $SSW = \sum_i (n_i - 1)s_i^2$ - the larger the values of s_i, the larger the value of SSW, and therefore MSW.

Initial: __________

Full Detail

Work out the following problems. **Show your work.**

1. Different methods of treating cloth to repeal mosquitoes were being tested, with treatments $T1$ = Chemical I, $T2$ = Chemical II, and C = Combination of chemicals I and II. The number of bites suffered by the subjects wearing the cloth in one day was measured, and summary statistics follow:

Number of Bites	C	$T1$	$T2$	Overall
Sample mean	5.37	8.03	8.13	7.18
Sample Std. Dev	3.07	3.01	3.46	3.40
Sample Size	30	30	30	90

Further, SSB = 147.033.

(a) Calculate the value of MSW.

$SSW = \sum_{i=1}^{I}(n_i - 1)s_i^2 = 29(9.4249) + 29(9.0601) + 29(11.9716) = 273.3221 + 262.7429 + 347.1764 = 883.2414$
$MSW = \frac{SSW}{d.f.\{W\}} = \frac{883.2414}{87} = 10.1522$

(b) The test-statistic is $F_S = 7.2432$, with p-value 0.001231. State your decision and conclusion about the appropriate null hypothesis in terms of the problem, assuming $\alpha = 0.05$.

Since p-value $< \alpha$, we reject the null and conclude there is statistical evidence to suggest that the true average number of bites for the three different treatments is not equal (or that at least one true average is different).

(c) Which 95% Bonferroni confidence interval for $\mu_a - \mu_b$ would be certain to have lower and upper bounds that do not cover zero? Explain your answer.

The interval for C and $T2$. This is because it has the largest difference in means, and the larger the difference in means the less likely it is for the confidence interval to contain zero.

(d) Someone states that the probability the null hypothesis is true is 0.001231. Do you agree with this statement? Explain your answer.

No. The p-value is the probability **of observing our sample data or more extreme**, if in reality the null hypothesis was true. The null hypothesis is assumed to be true when we calculate the p-value. In reality, the null is either true, or is not, so the probability is either 1, or 0.

Initial: ___________

2. Out of 150 plants infected with a parasite, 88 of them survived after they were treated with a product. You may assume a random sample was taken.

   (a) Find the 90% Wilson-Adjusted confidence interval for the true proportion, p.

   The confidence interval is $\tilde{p} \pm Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$, or $0.5844 \pm (1.645)\sqrt{\frac{0.5844(1-0.5844))}{150+4}}$ or $(0.5191, 0.6497)$

   (b) Interpret your interval in terms of the problem.

   We are 90% confident that the true proportion of infected plants that will survive after being treated is between 0.5191 and 0.6497.

   (c) A farmer has 1000 infected plants, and asks you to estimate how many of them would survive if they are treated with this product. What is your estimate? Explain your answer.

   Based on our confidence interval, we are 90% confident that between 1000*0.5191 and 1000*0.6497 should survive, in other words between 519.1 and 649.7 (520 and 650 would also be appropriate).

   (d) List all the assumptions needed for this confidence interval.

   A random sample was taken, and that there are at least 5 subjects with the trait (who survived) and 5 that did not (who did not survive).

Initial: ___________

3. The number of hours of sleep per night was measured for subjects who smoked and did not smoke, with the following summary statistics:

	Smokers	Non-Smokers
Sample Mean	7.04	5.9
Sample Standard Deviation	1.25	1.31
Sample Size	60	39

You may assume a random sample was taken from each group, and that the appropriate degrees of freedom are: 78. Assume the question of interest is if the average hours of sleep differs between the two groups.

(a) Find the 99% confidence interval for $\mu_1 - \mu_2$.

The 99% confidence interval is: $(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, or $(7.04 - 5.9) \pm (2.648)\sqrt{\frac{1.25^2}{60} + \frac{1.31^2}{39}}$ or $(1.14) \pm (2.648)(0.2647)$ or $(0.4391, 1.8409)$.

(b) Interpret your interval in terms of the problem, being as specific as you can.

We are 99% confident that the true average hours of sleep for a smoker is more than that of a non-smoker by between 0.4391 and 1.8409 hours.

(c) What does your confidence interval suggest about the range of the p-value for the hypothesis test $H_0 : \mu_1 = \mu_2$? Explain your answer, and **do not calculate the test-statistic directly**.

Since the range does not cover zero, it suggests we would reject the null hypothesis of equal means, so that our p-value would be < 0.01 (which is the α we used in the problem).

(d) Interpret a Type I Error for the appropriate hypothesis in terms of the problem.

If we concluded that the true average hours of sleep are not equal for smokers and non-smokers, when in reality they are.

Scratch Work