Initial: ___________

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) The Central Limit Theorem states that if a random sample from a population and your sample size is at least 30, then the distribution of your population is normally distributed.

FALSE. The Central limit theorem states that if you take a random sample from a popuulation and your sample size is at least 30, the distribution of the **sample mean** is normally distributed.
The size of the sample you take does not change the distribution of the population - that remains constant.

(II) The smaller the value of a test-statistic for ANOVA, the smaller the p-value associated with that test.

FALSE. The smaller the test-statistic, the closer the test-statistic is to zero, so the larger the area in the upper tail (the p-value). Or, a small F test statistic would correspond to data that strongly agrees with the null, and thus we would fail to reject, and the p-value must be large. A picture of the F distribution would illustrate this as well.

(III) For a confidence interval for $\mu_1 - \mu_2$, when we increase the sample size (assuming all other values remain constant), the confidence interval width decreases.

TRUE. This is because the estimated standard deviation $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ will decrease (since the sample size is in the denominator), which will decrease the uncertainty in the confidence interval, and narrow the width.

(IV) The spread of the random variable Y is always larger than the spread of the random variable $\bar{Y}$ when $n \geq 2$.

TRUE. When we average values, we decrease the range of possible values, and the spread decreases. Or, $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$, which is always less than σ_Y when $n \geq 2$.
Note: Trying to use the Central Limit Theorem is not appropriate here, because the spread of $\bar{Y}$ is always $\frac{\sigma_Y}{\sqrt{n}}$, whether the CLT applies or not. This can be found in lecture 08.

Initial: ___________

Full Detail

Complete the following problems. **Show your work.**

1. Suppose we take a random sample of 45 people, and ask each subject "How long does it take you to fall asleep at night?". The sample average was 45 minutes, with a standard deviation of 22.

   (a) Find a 99% confidence interval for μ, and interpret your interval in terms of the problem.

   The 99% confidence interval is calculated by: $\bar{y} \pm t_{\alpha/2}(s/\sqrt{n})$ or $45 \pm (2.704)(22/\sqrt{45})$ or (36.1321, 53.8679). We are 99% confident that the true average time until the subjects fell asleep is between 36.1321 and 53.8679 minutes.

   (b) Does your interval suggest that a sample average of 50 is unusual? Explain your answer.

   No, since 50 is in the interval this would not be unusual.

   (c) List all assumptions needed for a confidence interval for μ.

   We need a random sample from the population, and either that 30 or more subjects were measured, or that the population is normally distributed (because we need $\bar{Y}$ to be normally distributed).

   (d) If we wanted to build a confidence interval that has the same level of confidence, but whose margin of error e is 3, how many people would we have to sample at least?

   $n = \frac{t_{\alpha/2}^2 s^2}{e^2} = \frac{2.576^2(22^2)}{3^2} = 356.8572871 = 357$

2. A drug company tests three formulations of a pain relief medicine for migraine headache sufferers. The groups were independent, and the subjects randomly put into one of three groups. The subjects took the drug when they next had a migraine, and rated their pain on a scale of 1 to 10 (10 being the most pain). Sample statistics follow:

Species	A	B	C
$\bar{y}_i$	3.67	5.78	5.89
s_i	0.87	1.48	0.78
n_i	9	9	9

Further, $MSW = 1.1852$, and some possible values for the Bonferroni multiplier $t_{\alpha/(2k)}$ are : $t_{0.05/(2*1)} = 2.064$, $t_{0.05/(2*2)} = 2.391$, and $t_{0.05/(2*3)} = 2.574$.

(a) Find the 95% confidence intervals for the two pairs of means that are most likely to have a significant difference. You should have two confidence intervals calculated.

Since the sample sizes are the same, the means that are most likely to have a significant difference are those which have the largest sample difference. This would be $A - B$, and $A - C$.
Note that we are making two intervals, so $k = 2$, and the correct multipler is 2.391
The interval for A vs. B is : $\bar{y}_A - \bar{y}_B \pm t_{\alpha/(2k)}\sqrt{MSW(\frac{1}{n_A} + \frac{1}{n_B})}$ or $(3.67 - 5.78) \pm 2.391\sqrt{1.1852(\frac{1}{9} + \frac{1}{9})}$ or -2.11 ± 2.391(0.5132) or -2.11 ± 1.2271 or (-3.3371, -0.8829)
The interval for A vs. C is : $\bar{y}_A - \bar{y}_C \pm t_{\alpha/(2k)}\sqrt{MSW(\frac{1}{n_A} + \frac{1}{n_C})}$ or $(3.67 - 5.89) \pm 2.391\sqrt{1.1852(\frac{1}{9} + \frac{1}{9})}$ or -2.22 ± 2.391(0.5132) or -2.22 ± 1.2271 (-3.4471, -0.9929)
Note : If you swapped the order of the groups, your confidence intervals would be: A vs B : (0.8829, 3.3371), and A vs C: (0.9929, 3.4471)

(b) Interpret one of the confidence intervals from (a) in terms of the problem, being as specific as you can.

We are overall 95% that the true average pain score for group A is less than that of group B by between 0.8829 and 3.3371.
OR We are overall 95% that the true average pain score for group A is less than that of group C by between 0.9929 and 3.4471.

(c) What is the smallest difference between two group averages your confidence intervals from (a) suggest? Explain your answer.

This would be the smallest value in both of our confidence intervals, which would be a difference of 0.8829.

(d) If you were to conduct a hypothesis test for this data, would you reject the null or fail to reject based on the above? Explain your answer.

Since we concluded that there was a significant difference between a pair of means, we would reject the null hypothesis that all true group averages are equal.

Initial: ___________

3. A test comparing two diets for dairy cows was conducted, and their milk yield (in Liters) on each diet was measured. Some sample statistics follow:

	Diet I	Diet II
Sample Mean	45.15	42.25
Sample Standard Deviation	7.998	8.74
Sample Size	33	32

The goal of the study is to determine if there is a significant difference between milk yield. Assume the groups are independent, and a random sample of cows was taken. The d.f. for this problem are : 62. Let $\alpha = 0.05$

(a) State the null and alternative, and indicate which one corresponds to the goal of the study.

$H_0 : \mu_I = \mu_{II}$, vs $H_A : \mu_I \neq \mu_{II}$ (goal)

(b) Find the test-statistic and the range for the p-value.

$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(45.15 - 42.25) - 0}{\sqrt{\frac{7.998^2}{33} + \frac{8.74^2}{32}}} = \frac{2.9}{2.0798} = 1.3944$

Note that since $H_A : \mu_I \neq \mu_{II}$, the p-value is $2Pr\{t > 1.3944\}$.

Since on the t-table at d.f. = 60, we know that $Pr\{t > 1.294\} = 0.05$, and $Pr\{t > 1.671\} = 0.1$, our p-value has the range : $2(0.05) <$ p-value $< 2(0.1)$ or $0.1 <$ p-val < 0.2

(c) Interpret the p-value in terms of the problem.

If in reality there was no difference in true average milk yield, we would observe our data or more extreme with probability between $0.1 <$ and < 0.2

(d) What type of error could we have made in this problem? Interpret that type of error in terms of the problem.

Since we failed to reject the null, we could have made a type II error. We would have concluded that the true average milk yields were equal, when in reality they were not.

Initial: ___________

Scratch Work