DIRECTIONS:

- This exam contains two parts:
  - Part 1. 20 Multiple Choice `[40 points]`
  - Part 2. 4 Short Answer/Computation `[60 points]`
- On Part 1, circle the correct response for each question. Make sure that your answer is clearly marked. You will not receive partial credit for any work done in Part 1. On Part 2, show all of your work to receive full (or partial) credit.
- This is a closed-book examination. However, you may use one or two 8.5×11 sheet(s) of notes if you wish. You may also use a calculator.
- The standard normal and t distributions are attached at the end of this examination.
- Any discussion or otherwise inappropriate communication between examinees, as well as the appearance of any unnecessary material, will be dealt with severely.
- This exam is worth a total of 100 points. **Print** your name **at the top of this page in the upper right hand corner**. *Good Luck!!*

PART 1: MULTIPLE CHOICE. Circle the correct response for each question. Make sure that your answer is clearly marked.

1. The random variable X counts the number of errors on a particular accounting invoice. The probability distribution for X is given below.

x_i	0	1	2
$P(X = x_i)$	0.5	0.3	0.2

What is the **mean** number of errors that is recorded?
(a) 0.33
(b) 0.50
(c) 0.70
(d) 1.00

2. A sprinkler system contains two devices (D1 and D2). If device 1 (D1) works with probability 0.9, device 2 (D2) works with probability 0.8, and the two devices are **independent**, what is the probability that **both** devices will work?
(a) 0.28
(b) 0.72
(c) 0.85
(d) 1.70

3. The **length** of a confidence interval for μ does **not** depend on
(a) the sample mean, $\overline{x}$
(b) the sample size, n
(c) the population standard deviation, σ
(d) the confidence level, $1 - \alpha$

4. What is the 95th percentile of a t distribution with 8 degrees of freedom?
(a) 1.860
(b) 1.895
(c) 2.306
(d) 2.365

5. Which of the following statements is **false**?
(a) The t distributions are centered around zero.
(b) The t distributions are symmetric.
(c) The t distributions are less variable than a standard normal distribution.
(d) The area under a t density curve is one.

6. In conducting a hypothesis test, we compute the probability value to be 0.03. Which of the following statements is **true**?
(a) H_0 is rejected at both the $\alpha = 0.01$ and $\alpha = 0.05$ significance levels.
(b) H_0 is rejected at the $\alpha = 0.05$ level, but not at the $\alpha = 0.10$ level.
(c) H_0 is rejected at the $\alpha = 0.01$ level, but not at the $\alpha = 0.05$ level.
(d) H_0 is rejected at the $\alpha = 0.10$ level, but not at the $\alpha = 0.01$ level.

7. True or False. In a hypothesis test, if we reject H_0, then we know that H_1 is true.
(a) True
(b) False

8. True or False. In an SRS, as the sample size increases, the sample standard deviation s more precisely estimates the population standard deviation σ.
(a) True
(b) False

9. Suppose that I repeatedly take simple random samples from the same population. With each sample (each of the same size), I perform an $\alpha = 0.01$ level hypothesis test for H_0 versus H_1. What percent of the time will I reject H_0 when, in fact, H_0 is **true**?
(a) 0.005
(b) 0.01
(c) 0.05
(d) 0.10

10. In the late 1800's, William Gosset obtained a post as a chemist and brewer in the Guinness brewery in Dublin and did important work in statistics. What was his most famous contribution to the statistical community?
(a) He used regression to predict the Chicago White Sox would win the World Series in 2005!
(b) He invented the use of statistical process control charts.
(c) He pioneered early statistical methodology work in drug-testing.
(d) He discovered the t distribution.

11. Loosely speaking, what does the **Central Limit Theorem** say?
(a) For any probability distribution, the probabilities are guaranteed to sum to 1.
(b) The larger the sample size, the "better" the estimate.
(c) Averaging reduces variation.
(d) The sampling distribution of $\overline{x}$ is approximately normal.

12. We have computed the probability of an event to be 0.003. Which of the following statements would **not** be correct?
(a) The event is unlikely to occur.
(b) We would expect the event to occur about 0.3 percent of the time.
(c) If the event did occur, we would consider it to be unusual.
(d) The event will never occur.

13. With the one and two-sample t tests, the normal assumption may be violated slightly, but the tests will still produce meaningful results. Which term describes this property?
(a) random
(b) robust
(c) significant
(d) rejection point

14. Which of the following statements concerning probability values (P-values) is **false**?
(a) The smaller the P-value, the more evidence against the null hypothesis.
(b) The P-value is computed under the assumption that H_0 is true.
(c) The P-value for a hypothesis can never be larger than 1.
(d) Large P-values suggest that the alternative hypothesis is true.

15. What was the message taken from the comic strip that we looked at in class?
(a) The consequences of committing a Type I Error could be disastrous.
(b) Sample sizes should be large for the Central Limit Theorem to work.
(c) Statistical software should be used whenever possible.
(d) Outliers always have an effect on the final conclusions.

16. True or False. In a hypothesis testing situation, the **significance level** is another name for P-value.
(a) True
(b) False

17. True or False. Other things being equal with a simple random sample, as the sample size increases, the standard deviation for the sampling distribution of the sample mean also increases.
(a) True
(b) False

18. My colleague made the following statement. What do you think?

"We decided to test $H_0 : \overline{x} = 34.5$ versus $H_1 : \overline{x} \neq 34.5$."

(a) My colleague should have used a one-sided test.
(b) My colleague should have used a confidence interval.
(c) My colleague doesn't know the difference between the sample mean and the population mean.
(d) My colleague committed a Type II Error.

19. I have a statistic whose sampling distribution has mean equal to θ. What do I know?
(a) My statistic was computed from a random sample.
(b) My statistic is an unbiased estimator for θ.
(c) My statistic will always equal θ.
(d) My statistic's sampling distribution is approximately symmetric.

20. I have computed a 95 percent confidence interval for a population mean μ. The interval is $(45.54, 56.31)$. Which of the following statements must be **true**?
(a) $H_0 : \mu = 55$ would not be rejected at the $\alpha = 0.10$ level.
(b) $H_0 : \mu = 55$ would not be rejected at the $\alpha = 0.01$ level.
(c) $H_0 : \mu = 60$ would not be rejected at the $\alpha = 0.05$ level.
(d) $H_0 : \mu = 60$ would be rejected at all α levels less than 0.05.

PART 2: SHORT ANSWER/COMPUTATION. Show all of your work, and explain your reasoning.

1. [15] Sheila's doctor is concerned that she may suffer from gestational diabetes. There is variation both in the actual glucose level and in the blood test that measures the level. A patient is classified as having gestational diabetes if the glucose level is above 140 milligrams per deciliter (mg/dl) one hour after a sugary drink is ingested. Sheila's measured glucose level one hour after ingesting the sugary drink varies according to a **normal** distribution with $\mu = 125$ mg/dl and $\sigma = 10$ mg/dl.

(a) If a single glucose measurement is made, what is the probability that Sheila is diagnosed as having gestational diabetes?

(b) If measurements are made on 4 separate days and the mean result is compared with the criterion 140 mg/dl, what is the probability that Sheila is diagnosed as having gestational diabetes? (Assume that the measurements are independent).

2. [15] You work for a manufacturing company that specializes in trenching equipment. A popular piece of equipment manufactured by the company is a Model CC145 Concrete Cutter. The length of a critical part for this model, measured in millimeters (mm), has mean μ and varies according to a **normal** distribution. Based on a random sample of $n = 16$ parts taken from an assembly line, you have computed $\overline{x} = 95.2$ mm. From past knowlege, a reliable guess of the population standard deviation is $\sigma = 1.5$ mm.

(a) Find a 90 percent **confidence interval** for μ, the mean length of this critical part.

(b) A colleague examines the confidence interval that you computed in (a) and says, "It must be that 90 percent of the **parts** have lengths between your lower and upper endpoints" (you should have numerical values for these endpoints above). Do you agree with your colleague's assertion? Explain why you do or why you do not.

3. [15] In 2004, a study was conducted to examine the cost of education at schools in the Atlantic Coast Conference (ACC). From a random sample of $n = 26$ ACC students, the average cost per academic year (this includes tuition, room, board, books, etc.) was $\overline{x} = 18,700$ dollars. The sample standard deviation for the 26 students was equal to 2,500 dollars.

According to the College Board, the 2003-2004 average costs were 17,660 dollars for students attending major conference schools (like the ACC) in the United States. Based on the data from this study, is the ACC's population mean yearly cost significantly **greater** than the national mean of 17,660 dollars? Set up appropriate null and alternative hypotheses and carry out a statistical test to answer this question. Use an $\alpha = 0.10$ level, and state your conclusion using "plain English." Also, make sure to draw a pertinent picture with everything properly labeled.

4. [15] A production supervisor at a major chemical company must determine which of two catalysts (substances that cause a chemical reaction), catalyst XA-100 or catalyst ZB-200, maximizes the hourly yield of a chemical process. In order to compare the mean hourly yields obtained by using the two catalysts (μ_1 and μ_2, respectively), the supervisor runs the process using each catalyst for five one-hour periods. The resulting yields (in pounds per hour) for each catalyst, along with some summary statistics are given below.

XA-100 (Catalyst 1)	ZB-200 (Catalyst 2)
801	752
814	718
784	776
836	742
820	763
$n_1 = 5$	$n_2 = 5$
$\overline{x}_1 = 811.0$	$\overline{x}_2 = 750.2$
$s_1^2 = 386.0$	$s_2^2 = 484.2$

Here we will assume that all other factors affecting yields of the process have been held constant; thus, the five observations for each catalyst construe a random sample from each process. We will also assume that the process variances σ_1^2 and σ_2^2 are equal.

(a) Compute a pooled estimate of the common variance. Show all of your work.

(b) Suppose that the supervisor (your supervisor, say) desires to test

$$H_0 : \mu_1 - \mu_2 = 0$$
versus
$$H_1 : \mu_1 - \mu_2 \neq 0.$$

Using software, he (correctly) computes the test statistic

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{811.0 - 750.2}{\sqrt{435.1\left(\frac{1}{5} + \frac{1}{5}\right)}} = 4.61,$$

but he has no idea how to interpret it (he went to Clemson). First, making sure to label everything, draw a pertinent picture to illustrate how he can make his decision. Second, explain to your supervisor, in "plain English" terms, what information is conveyed by this test statistic. Use the next page for all of your work.

This is an extra page for Question 4.