2-Sample Hypothesis Tests for the Difference in Means

Requirements for complete responses to free response questions that require 1-sample hypothesis tests for means:

1. Identify the population parameter of interest, the type of inferential procedure to be used, the null and the alternative hypothesis, and the significance level of the test.
   - The both population parameters of interest must be defined in context of the problem.
   - The null hypothesis is <u>always</u> an "equals" statement.
   - The alternative hypothesis is <u>always</u> an inequality statement (use $<$, or $>$, or $\neq$)
   - A t-distribution is used for a hypothesis test for means, since we are estimating the population standard deviation using the sample standard deviation, s. Don't forget to include the degrees of freedom! (Note: use the smaller of $n_1 - 1$ and $n_2 - 1$ or your calculator to determine the degrees of freedom.)
   - Remember that the significance level of the test α is the probability of Type I error.
2. State and verify whether the conditions (assumptions) needed for this procedure are met.
   - The conditions for 2-sample hypothesis test for the difference of means are as follows:
     - Both samples must be independent random samples from the populations of interest or, in an experimental setting, random assignment to treatments must be established. You must state this condition and verify whether or not this condition is met.
     - The samples must be large. To verify this condition, check that $n_1 \geq 30$ and $n_2 \geq 30$. You must show your work to receive credit for verifying this condition. If the sample sizes are small, you must show a graph of both sample data sets and state how each graph shows whether it is reasonable to assume each population is normally distributed or not.
     - The observations within each sample are independent. When sampling without replacement, each of the two populations must be at least 10 times greater than the sample size.
3. Show the results of the test including the formula for the test statistic, the value of the test statistic, the degrees of freedom, the p-value, and a picture of the sampling distribution showing the hypothesized mean, the test statistic, and the associated p-value shaded.
   - Every test statistic computation follows the same general form (on the formula sheet):

$$\frac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$$

4. State the conclusion in context of the situation.
   - Include a statement linking the p-value to the significance level and the decision about the null hypothesis (fail to reject or reject) and a second statement about whether or not there is evidence to support the alternative hypothesis. The alternative should always be stated in context. Remember you <u>never</u> accept the null hypothesis!

Hypothesis Tests for Matched Pairs:

When the observations are collected in pairs or the observations in one group are related to the observations in the other, a hypothesis test for the mean difference can be used. The conditions for this type of hypothesis test are the same as a one-sample t-test, using the differences that must be independent of each other.

The formula is: $t = \dfrac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$, where $\bar{d}$ is the mean difference, μ_d is the pairwise difference, and s_d is the standard deviation of the differences.

Other things to remember about Hypothesis testing in general:

- The p-value of a test gives the probability that the observed outcome, or one even more different from the claim of the null hypothesis, would occur just by chance if the null hypothesis were true.
- Small p-values indicate that there is little chance that the observed outcome, or one even more different from the claim of the null hypothesis, would occur just by chance if the null hypothesis were true. Small p-values lead us to reject the null hypothesis.
- Large p-values indicate that the observed outcome, or one even more different from the claim of the null hypothesis, is not rare and would be likely to occur just by chance if the null hypothesis were true. So large p-values lead us to conclude that the null hypothesis should not be rejected.

		Reality	
		H_0 is true	H_0 is false
Decision	Fail to reject H_0	Correct decision	Type II error
	Reject H_0	Type I error	Power (correct decision)

- A *Type I error* occurs when you reject a null hypothesis that is true.
- A *Type II error* occurs when you fail to reject a null hypothesis that is not true.
- The *power* of a test is the probability that a test will reject a false null hypothesis. It makes sense to think of the *power* of the test as how *sensitive* the test is to detect a false null hypothesis.
- $\text{Power} = 1 - P(\text{Type II error})$
- Power increases when sample size increases (the standard error decreases); power increases when the α-level of a test increases (the test is more likely to reject the null hypothesis at the 10% level than at the 5% level);

Writing templates for Hypothesis Testing:

Conclusion for a hypothesis test:

Since the p-value is (less/greater) than alpha, (show inequality statement). I (reject/ fail to reject) the null hypothesis. There (is/ is not) statistically significant evidence that (state the alternative hypothesis in context.)

Interpretation of p-value:

The probability of observing a sample (statistic of interest) as extreme or more extreme than (state the observed value), assuming (state the null hypothesis in context) is true, is (state the p-value).

Type I error:

A type I error would be rejecting the null hypothesis that (state the null in context) when in fact it is true.

Type II error:

A type II error would be failing to reject the null hypothesis that (state the null in context) when in fact (state the alternative in context).

Multiple Choice Questions:

1.

A major department store wants to compare the mean number of customers an employee can process per hour using two new scanning systems. One hundred employees are randomly selected and then randomly assigned to one of the new systems. After each group of 50 employees have been trained on one scanning system and 50 employees on the other scanning system, the number of customers processed per hour is recorded. Which of the following would be the most appropriate inferential statistical test to use in this situation?

A. One-sample t-test
B. Two-sample t-test
C. Paired t-test
D. One-sample z-test
E. Two sample z-test

2.

Dan, a trainer at the Popular Gym, was interested in comparing levels of physical fitness of students attending a nearby community college and those attending a 4-year college in town. He selected 320 students from the community college. The mean and standard deviation of their fitness scores were 95 and 10, respectively. Dan also selected a random sample of 320 students from the 4-year college. The mean and standard deviation of their fitness scores were 92 and 13, respectively. He then conducted a two-sided t-test that resulted in a t- value of 3.27. Which of the following is an appropriate conclusion from this study?

A. Because the sample means only differed by 3, the population means are not significantly different.
B. Because the second group had a larger standard deviation, their mean fitness score is significantly higher.
C. Because the second group had a larger standard deviation, the mean fitness score of the first group is significantly higher.
D. Because the p-value is less than $\alpha = 0.05$, the mean fitness scores for the two groups of students are significantly different.
E. Because the p-value is greater than $\alpha = 0.05$, the mean fitness scores for the two groups of students are significantly different.

3.

Two students, who were registering for college classes, had a choice of either Dr. Brown or Dr. Edwards. One of the students heard that Dr. Brown's class was easier since the average grade in his class was higher than Dr. Edwards. A random sample of grades from 35 students from each professor was obtained with the results recorded below:

	$\bar{x}$	s	n
Dr. Brown	81	3	35
Dr. Edwards	78	5	35

Since both students took AP Statistics in high school, they decided to conduct a test of significance using the following hypotheses.

$$H_0 : \mu_B = \mu_E$$
$$H_a : \mu_B > \mu_E$$

The p-value was 0.0018. Which conclusion(s) is supported by the results of the significance test?

I. If Dr. Brown's grades do in fact have a higher average than Dr. Edwards' grades, we would expect to get results as extreme as those from our sample 0.18% of the time.
II. The test results are valid, since the conditions necessary to perform the test were met.
III. We do have sufficient evidence to suggest that Dr. Brown's average grade is higher than Dr. Edwards' average grade at the $\alpha = 0.05$ level.

A. I only
B. II only
C. III only
D. I and II only
E. II and III only

4.

According to research, black bear cubs have a better of chance of surviving outside of the den if they weigh at least five pounds when they first leave the den with their mother. At birth, the cubs can weigh 8 to 12 ounces which means that they must gain several pounds in a very short period of time. A study was conducted on a random sample of 10 pairs of cubs to determine if the first born cub weighed more that the second born cub. If μ_1 represents the true mean birth weight of all first born cubs, μ_2 represents the true mean birth weight of all second born cubs, and u_D represents the true mean difference in birth weight for all pairs of black bear cubs, $(\mu_1 - \mu_2)$.

Which of the following hypothesis statements should be used for this study?

A. $H_0 : \mu_1 - \mu_2 = \mu_D$ and $H_a : \mu_1 - \mu_2 > \mu_D$
B. $H_0 : \mu_1 - \mu_2 = \mu_D$ and $H_a : \mu_1 - \mu_2 \neq \mu_D$
C. $H_0 : \mu_1 = \mu_2$ and $H_a : \mu_1 < \mu_2$
D. $H_0 : \mu_1 = \mu_2$ and $H_a : \mu_1 \neq \mu_2$
E. $H_0 : \mu_D = 0$ and $H_a : \mu_D > 0$

5.

A history teacher wants to compare the mean times needed to access a particular website using two different internet browsers. He believes that Browser A is faster than Browser B (Browser A's time is less than Browser B). Twenty-five randomly selected students accessed the website using Browser A with a mean of 39 seconds and a standard deviation of 5 seconds. Another twenty-five randomly selected students accessed the website using Browser B and had a mean time 42 seconds and a standard deviation of 7 seconds. Assume the conditions for inference are met. Which of the following statements is true?

A. If $\alpha = 0.01$, Browser B is significantly faster than Browser A.
B. If $\alpha = 0.01$, Browser A is significantly faster than Browser B.
C. If $\alpha = 0.05$, Browser B is significantly faster than Browser A
D. If $\alpha = 0.05$, Browser A is significantly faster than Browser B.
E. If $\alpha = 0.10$, there is no significant difference in the speeds of the two browsers.

6.

A football team owner believes that attendance at the home stadium increases when the team has a winning season. He randomly selects 10 seasons when the team won more games than they lost and determines that the mean attendance is 88,532 people and the standard deviation is 1341 people. Ten more seasons are randomly selected from years when the team lost more games than they won with a mean of 86,910 people and a standard deviation of 1521 people. Assume all conditions for conducting a significance test have been met and that the p-value is 0.011. Which of the following conclusions should the owner make?

A. The mean attendance at the home stadium is greater when the team is having a winning season approximately 1.1% of the time.

B. The p-value indicates that the mean attendance at the home stadium is greater when the team is having a winning season at the 1% level of significance. There is a significant difference in mean attendance when the team is winning than when the team is losing.

C. At the 5% significance level, there is evidence that the mean attendance at the home stadium is greater when the team is having a winning season than when the team is having a losing season. We would expect to get a test statistic at least as extreme as the one observed 1.1% of the time if the null hypothesis is true.

D. At the 5% significance level, there is evidence that the mean attendance at the home stadium is greater when the team is having a winning season than when the team is having a losing season. We would expect to get a test statistic at least as extreme as that observed 98.9% of the time if the null hypothesis is true.

E. Because the p-value is so small, the results show that there is not a significant difference between the mean attendance during a winning season and the mean attendance during a losing season.

Free Response Questions:

7. Investigators at the U. S. Department of Agriculture wished to compare methods of determining the level of *Salmonella* bacteria contamination in lettuce. Two different models (A and B) of determining the level of contamination were used on each of twelve randomly selected specimens of a certain type of lettuce. The data obtained, in millimicrobes/liter of chopped lettuce, for each of the methods are shown in the table below.

		Specimens											
		1	2	3	4	5	6	7	8	9	10	11	12
Method	A	45.7	43.8	43.6	46.1	47.4	48.5	52.4	55.2	60.4	67.4	68.3	69.1
	B	45.0	43.1	42.7	46.5	46.6	48.1	51.2	55.0	60.5	67.3	67.9	69.4

Is there a significant difference in the mean amount of *Salmonella* bacteria detected by the two methods for this type of lettuce? Provide a statistical justification to support your answer.

8. A blood sugar level above 160 milligrams per deciliter (mg/dl) of blood 2 hours after eating is a risk factor for gestational diabetes in pregnant women. At a medical center in Chicago, a study to test the effectiveness of a new drug to control blood sugar levels was conducted. One hundred women diagnosed with gestational diabetes were available for this study. Fifty women were assigned at random to each of two treatment groups. One group received the standard medication and the other group received the new drug. After taking the drug for three weeks, the 50 subjects who received the standard treatment had a mean decrease in blood sugar level of 12 mg/dl with a standard deviation of 8 mg/dl, and the 50 subjects who received the new drug had a mean decrease of 15 mg/dl with a standard deviation of 9 mg/dl.

   Does the new drug appear to be more effective than the standard treatment in lowering mean blood sugar level? Give appropriate statistical evidence to support your conclusion.

9. A researcher wants to conduct a study to test whether listening to soothing music while preparing for diagnostic tests helps short term memory loss, for patients with minor to moderate dementia, compared to sitting in a noise-free environment while preparing for diagnostic tests. One hundred patients at a large neurological institute are available to participate in this study.

    (a) Propose a design for this study to compare these two treatments.

(b) The null hypothesis for this study is that there is no difference in the mean short term memory retention for the two treatments and the alternative hypothesis is that the mean short term memory retention is greater for the music treatment. If the null hypothesis is rejected, the institute will offer this music therapy as a free service to their patients with short term memory loss. Describe Type I and Type II errors and the consequences of each in the context of this study, and discuss which one you think is more serious.