1-Sample Hypothesis Tests for 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 population parameter of interest must be defined in context of the problem and must be identified using μ for the population mean.
    - 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: $\text{df} = n-1$ for a one sample means test.)
    - 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 1-sample *t*-tests are as follows:
        - The sample must be a simple random sample from the population of interest. Usually this is stated in the prompt. You must state this condition and verify whether or not it's met.
        - The sample must be large OR the population must be normally distributed. To verify this condition check that $n \geq 30$. If the sample size is small, you must show a graph of the sample data and state how the graph shows whether it is reasonable to assume the population is normally distributed or not.
        - The individual observations are independent. When sampling without replacement, the population 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!

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

Interpretation of *p*-value:

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

Type I error:

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

Type II error:

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

Multiple Choice Questions:

1.

Which of the following is a condition for using a *t*-test for a population mean?

A. The mean of the population is known
B. The standard deviation of the population is known
C. The sample is not normally distributed
D. The population size is small
E. The sample size is large or the population is normally distributed

2.

In a test of the null hypothesis $\mathrm{H}_0 : \mu = 31$ *versus* the alternative hypothesis $\mathrm{H}_a : \mu < 31$, a sample from a normal population produces a mean of 28.7. The *t*-statistic for the sample is -2.01 and the *p*-value is 0.026. Based on these statistics, which of the following conclusions could be drawn?

A. 97.4 percent of the time, the mean is below 31.
B. There is reason to conclude that $\mu < 31$.
C. 2.6 percent of the time, rejecting the alternative is in error.
D. 2.6 percent of the time, the mean is above 31.
E. Due to random fluctuation, 47.4 percent of the time a sample produces a mean larger than 31.

3.

The process of producing cough suppressant tablets yields tablets with varying amounts of the active ingredient. It is claimed that the average amount of active ingredient per tablet is at least 400 milligrams. The Health Products Bureau tests a random sample of 90 tablets. The mean content of the active ingredient for this sample is 396.8 milligrams, while the standard deviation is 31 milligrams. What is the approximate *p*-value for the appropriate test?

A. 0.050
B. 0.100
C. 0.165
D. 0.330
E. 0.335

4.

In a test of the hypothesis $\text{H}_0: \mu = 65$ *versus* $\text{H}_a: \mu > 65$, the power of the test when $\mu = 66.5$ would be greatest for which of the following choices of sample size *n* and significance level α ?

A. $n = 25,\ \alpha = 0.05$
B. $n = 25,\ \alpha = 0.01$
C. $n = 30,\ \alpha = 0.05$
D. $n = 30,\ \alpha = 0.10$
E. It cannot be determined from the information given.

5.

A young man fresh out of college plans to purchase a franchise business. He needs to average more than $25,000 in sales each month to avoid bankruptcy. Before making the investment in the business, he takes a random sample of 40 of the previous month's sales and conduct a test of significance with the following hypothesis.

$H_0 : \mu = 25,000$
$H_a : \mu > 25,000$

What is a Type II error and its consequence for the young man?

A. The young man, believing that the average sales will be more than $25,000, will purchase the business, while the average sales are less than $25,000 and he ends up in bankruptcy.
B. The young man, believing that the average sales will be more than $25,000, will purchase the business and the average sales are more than $25,000 and he ends up with a successful business.
C. The young man, believing that the average sales will be less than $25,000, will not purchase the business when the average sales exceed $25,000 and he would have succeeded had he purchased the business.
D. The young man, believing that the average sales will be less than $25,000, will not purchase the business and the average sales are less than $25,000 and he would have gone bankrupt with the business.
E. The consequence of a type II error cannot be assessed unless we know the α level.

6.

A well-known nutritionist claims U. S. children drink an average of 4 classes of milk per day. Based on a random sample of 30 children, the *p*-value for the test is 0.0015. What is the correct interpretation of this *p*-value?

A. The probability that we fail to reject the null is about 0.15%.
B. The probability that we reject the null hypothesis is about 0.15%.
C. We are 99.85% confident that the alternative hypothesis is true.
D. About 0.15% of all samples would produce a test statistic at least as extreme as ours if the null hypothesis is true.
E. About 99.85% of all samples would produce a test statistic at least as extreme as ours if the null hypothesis is true.

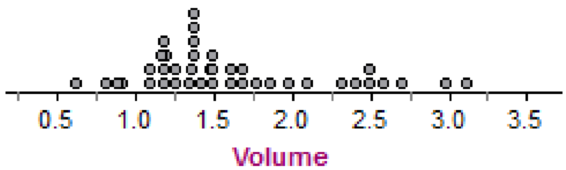
Free Response Questions:

7. The glass used for making windshields is produced using a process called "floating" where molten glass is pumped onto a bed of molten tin and cooled rapidly forming perfectly smooth sheets. The process for producing the sheets of glass is set to result in a mean thickness of 3.18 millimeters. To ensure that sheets of glass are a consistent thickness, every hour a worker randomly selects 10 sheets produced in the past hour and measures the average thickness of the sheets. If there is convincing evidence that the mean thickness is different from 3.18 millimeters or if there is convincing evidence that the standard deviation is greater than 0.04 millimeters, the entire process is shut down for recalibration. It can be assumed that the average thickness of the glass sheets is normally distributed.

   (a) During one hour, the mean thickness of 10 randomly selected sheets was 3.16 millimeters and the standard deviation was 0.055 millimeters. Perform a test of significance to determine whether the mean thickness of the sheets is different from 3.18 millimeters. Assume the conditions for inference are met.

   (b) To determine if this sample of 10 sheets provides convincing evidence that the standard deviation of the glass sheets being produced is greater than 0.04 millimeters, a simulation study was performed. In the simulation study, 500 samples, each of size 10, were randomly generated from a normal population with mean 3.18 and standard deviation 0.04. The sample standard deviation was computed for each of the 500 samples. 22 of the 500 samples had standard deviations greater than 0.055 millimeters. What would the results of this simulation study lead you to conclude?

8. Environmental law requires that the waste water from Molybdenum mining near Taos, New Mexico be pumped to tailing ponds. The pipeline running from one mine has leaks and the contaminated water runs into the Red River. An environmental group interested in local water supplies collects the volume of contaminated water leaked per day for a random sample of 40 days and finds that the pipeline is leaking an average of 1.61 million gallons of contaminated water per day with a standard deviation of 0.6 million gallons. A dotplot of their data is shown below.



Regulations on the company's mining permit states that the mining company has to keep the volume of contaminated water entering the river below 1.5 million gallons per day.

(a) Is there sufficient evidence at the $\alpha = 0.05$ level to conclude that the mining company is <u>not</u> in compliance with the regulations? Provide statistical justification for your answer.

(b) Assuming that the true mean amount of contaminated water leaked by this pipeline is 1.55 million gallons, what could the environmental group do to increase the power of the test in part (a)? Explain how your approach would increase the power of the test.

9. Since Hill Valley High School eliminated the use of bells between classes, teachers have noticed that more students seem to be arriving to class a few minutes late. One teacher decided to collect data to determine whether the students' and teachers' watches are displaying correct time. At exactly 12:00 noon, the teacher asked 9 randomly selected students and 9 randomly selected teachers to record the times on their watches to the nearest half minute. The ordered data showing minutes after 12:00 as positive values and minutes before 12:00 as negative values are shown in the table below.

Students	–4.5	–3.0	–0.5	0	0	0.5	0.5	1.5	5.0
Teachers	–2.0	–1.5	–1.5	–1.0	–1.0	–0.5	0	0	0.5

(a) Construct parallel boxplots using these data.

(b) Based on the boxplots in part (a), which of the two groups, students or teachers, tends to have watch times that are closer to the true time? Explain your choice.

(c) The teacher wants to know whether individual student's watches tend to be set correctly. She proposes to test $H_0: \mu = 0$ versus $H_a: \mu \neq 0$, where μ represents the mean amount by which all student watches differ from the correct time. Is this an appropriate pair of hypotheses to test to answer the teacher's question? Explain why or why not. <u>Do not carry out the test.</u>