Understand that valid statistical conclusions are based on randomly selected, representative samples of the population. That is to say, if the sample is not representative of the population, then all bets are off.

Understand the caveats that statistical conclusions are based on the sample data and the level of significance. That is to say, if we were to use a different sample and/or a different level of significance, then we might arrive at a different conclusion regarding the hypotheses.

Be able to apply hypothesis testing for ***One Sample*** and both ***Dependent*** **and** ***Independent*** **Two-Sample Testing**.

One Sample (Large Sample n >= 30, Small Sample n < 30)

Two Sample Dependent

Paired *t*-test (same subjects measured twice {Example: *Pre & Post Testing* and *Different Instruments*}).

Two Sample Independent

Population Means (H_0: $\mu_1 = \mu_2$) for both cases where population variances are known and unknown.

Use a ***One-Way*** **ANOVA** table to determine whether or not there is a statistically significant difference among means for a multi-level single-factor experiment. In the case where we rejected the null hypothesis, the proper statement for the conclusion would be "At least one sample (or equivalently, one level of treatment) is significantly different".

Refer to General Rules for Hypothesis Testing (M&R, Section 9-1.6, 4ed, page 301, 5ed page 296).
For any problems dealing with hypothesis testing: you must be able to correctly

- state both the null hypothesis and the alternate hypothesis;
- determine the appropriate critical value(s);
- calculate the pertinent test statistic;
- decide whether to "Reject the Null" or to "Fail to Reject the Null";
- draw a proper conclusion about the problem specifics based on your decision regarding the null hypothesis.

Remember "Reject the Null" is a strong conclusion.
In general, "Fail to Reject the Null" is at its best, a suggestion that there is
insufficient evidence to warrant rejecting the H_0, or equivalently, *insufficient evidence to warrant accepting* H_1.

One- Sample Hypothesis Testing examples of conclusions in cases where we failed to reject the null hypothesis:

if H_1: $\mu \neq \mu_0$, then an appropriate statement would be "insufficient evidence to say $\mu \neq \mu_0$".

if H_1: $\mu < \mu_0$, then an appropriate statement would be "insufficient evidence to say $\mu < \mu_0$".

if H_1: $\mu > \mu_0$, then an appropriate statement would be "insufficient evidence to say $\mu > \mu_0$".

Two-Sample Hypothesis Testing examples of conclusions in cases where we failed to reject the null hypothesis:

if H_1: $\mu_1 \neq \mu_2$, then an appropriate statement would be "insufficient evidence to say $\mu_1 \neq \mu_2$".

if H_1: $\mu_1 < \mu_2$, then an appropriate statement would be "insufficient evidence to say $\mu_1 < \mu_2$".

if H_1: $\mu_1 > \mu_2$, then an appropriate statement would be "insufficient evidence to say $\mu_1 > \mu_2$".

More than Two-Sample Hypothesis Testing (ANOVA) where we failed to reject the null hypothesis:

"Insufficient evidence to say that at least one sample is significantly different."

Be mindful that hypotheses are assumptions about population parameters and that conclusions are not facts but only suppositions with some associated degree of confidence. We can never know whether the null hypothesis is true or false and therefore we risk encountering either a Type I or a Type II Error regardless of how carefully we conducted the experiment and how diligently we collected the data

Define *Type I Error*, *Type II Error*, *Power of the Test*.
Describe the effect of changing the *Level of Significance* has on both the *Probability of Type I* & *Type II Errors*.
Explain the difference between *Statistical Significance* and *Practical Significance*.
Know the definition and application of *p-values* (it will not be necessary to calculate p-values).

For the test, you will be given the sample size, mean, and standard deviation as needed;
you will not be required to calculate these sample statistics, nor will you be required to determine any p-values.
The test will be open notes, you may use your own generated notes and all course handouts, appropriate tables from the textbook as needed. You must have your own calculators, notes, and tables, no sharing!
Cell phones use is NOT permitted during the test.

Practice Review Problems: (Use $\alpha = 5\%$ for all problems. It is not necessary to calculate any p-values.)

Homework #8 One Sample Hypothesis Testing
Homework #9 Two Sample Hypothesis Testing
Homework #10 One-Way ANOVA

Forming the Null Hypothesis Quiz

Hypothesis Testing Quizzes for One Sample, Two, Sample, and ANOVA

State the Claim
State the Null Hypothesis
State the Alternate Hypothesis
Type of Test, Number of Tails, Degrees of Freedom
Determine Critical Value
Calculate Test Value
Conclusion Regarding the Null Hypothesis
Conclusion Regarding the Alternate Hypothesis
Conclusion Regarding the Claim

Example Problems from Previous Exams (see pages 3 & 4)

State the Claim
State the Null Hypothesis
State the Alternate Hypothesis
Type of Test, Number of Tails, Degrees of Freedom
Determine Critical Value
Calculate Test Value
Conclusion Regarding the Null Hypothesis
Conclusion Regarding the Alternate Hypothesis
Conclusion Regarding the Claim

One Sample Hypothesis Testing

The specified amount of the active ingredient for an over-the-counter analgesic is 485 milligrams. An assay of 36 randomly sampled tablets averaged 487.4 milligrams, with a sample standard deviation of 6.7 milligrams. Based on the sample data, does the over-the-counter drug meet specifications?

The package label for 3/4 inch nylon rope states that the average breaking strength exceeds 5000 pounds. A safety expert used a sample of 100 different pieces of rope and calculated the average breaking strength to be 5045 pounds with a sample standard deviation of 245 pounds. What can you conclude about the manufacturer's statement regarding the average breaking strength, does the sample data provide strong statistical evidence that the average breaking strength exceeds 5000 pounds?

The brightness of an LCD screen is in some aspects determined by illumination current. The design for a certain level of brightness specifies an average current of 300 milliamps. Twenty sample circuits were measured and the average current was 318 milliamps with a sample standard deviation of 42 milliamps. Note: If the average illumination current is either too low or too high, the LCD brightness will be adversely affected. Based on the sample average illumination current, does the circuit meet specifications for average current of 300 milliamps?

The average tensile strength for a safety break-away link is specified to be less than 2400 pounds. A safety expert used a sample of 25 different links and calculated the average break-away strength to be 2385 pounds with a sample standard deviation of 40 pounds. What can you conclude about the manufacturer's statement regarding the average break-away strength, does the sample data provide strong statistical evidence that the average break-away strength is less than 2400?

Two Sample Hypothesis Testing

Two different analytical procedures were used to determine the impurity levels in well water. Fifteen specimens were tested using each procedure. The data is shown in the table below. The average difference between the two procedures is 26.87, with a standard deviation of the differences of 19.04. Based on the sample data, is there sufficient evidence to conclude that the procedures produced significantly different results? Note: Use a two-tailed test, $\alpha = 5\%$.

Specimen	Procedure 1	Procedure 2	Difference
1	265	229	36
2	240	231	9
3	258	227	31
4	295	240	55
5	251	238	13
6	245	241	4
7	287	234	53
8	314	256	58
9	260	247	13
10	279	239	40
11	283	246	37
12	240	218	22
13	238	219	19
14	225	226	-1
15	247	233	14

Test the data from two independent samples to determine whether or not their means are significantly different. Assume equal variances; use Level of Significance $\alpha = 0.05$

$\overline{X}_1 = 91.0$ $\overline{X}_2 = 92.2$ $Sp = 3.42$

$s_1 = 8.74$ $s_2 = 12.2$

$n_1 = 10$ $n_2 = 10$

ANOVA

Suppose as a newly hired engineer (albeit whether you might be a biomedical, human factors, or industrial systems, or electrical, or mechanical, or even a computer science/engineering) your supervisor at a major pharmaceutical facility has asked that you investigate whether or not a series of formulating machines are generating equals outputs of product per hour. Data from five different machines are taken at random to be processed by a statistical software program. As sometimes happens, the computer crashes before completing the calculations. Wanting to impress your boss, you continue the analysis, using the partial computer output and determine whether or not the machines produce equal output of product per hour (α = 5%).

Machine	**Output of Product in Pounds per Hour**				
1	242	246	254	242	228
2	246	252	243	244	216
3	246	234	237	216	241
4	224	234	241	233	229
5	242	226	242	222	241

Partial Computer ANOVA Results

Source	**Sum Squares**	**df**	**Mean Sum Squares**	F_{test}
Machines	365	________	___________	________
Error	________	________	___________	
Total	2565			

Additional Questions Test 4

Indicate whether or not the following statements are True or False (circle the correct answer).

1. Type I and Type II Errors:

T / F Increasing the Level of Significance increases the probability of a Type I Error.

T / F Decreasing the Level of Significance increases the probability of a Type II Error.

T / F *Rejecting the Null Hypothesis* is logically equivalent to *Accepting the Alternate Hypothesis*.

T / F *Failing to Reject the Null Hypothesis* may lead to Type Two Errors.

T / F Figures never lie, but liars sometimes figure.

2. With regard to Dependent Two-Sample Hypothesis Testing, when is it appropriate to use a *paired t-test*?

T / F When the population variances are known, but are statistically significantly different.

T / F When the population variances are known, but there is no statistically significant difference.

T / F When the population variances are unknown, but are statistically significantly different.

T / F When the population variances are unknown, but there is no statistically significant difference.

T / F When the samples are taken from two distinctively separate populations.

T / F When the same sample is tested twice using different instruments.

T / F When the same sample is measured both before and after an experimental procedure.

3. Practical Significance versus Statistical Significance

T / F For large sample sizes, small departures from the hypothesized value μ_0, will probably be detected, even though the difference is of little or no practical difference.

T / F For large sample sizes, small departures from the hypothesized value μ_0, will probably not be detected, unless we use a level of significance $\alpha <= 1\%$.

T / F Practical differences can only be detected for sample size $n < 30$, if the p-value is less than 1%.

T / F Practical differences can only be detected for sample size $n < 30$, if the p-value is greater than 5%.

T / F To insure discriminating between practically significant and statistically significant differences, we should use sample size $n > 30$.

T / F To insure discriminating between practically significant and statistically significant differences, we should use a level of significance $\alpha <= 1\%$.

4. For the most part, textbook problems, our quizzes, and even this test provided instructions with regards to two sample independent hypothesis testing as to whether or not to assume the population variances to be equal or unequal. For real world practical situations, describe a procedure for making this determination: include

the null hypothesis,
the formula for calculating the test value,
the degrees of freedom,
the formula for calculating the two sample test value when the null hypothesis is rejected,
the formula for calculating the two sample test value when we fail to reject the null hypothesis.

5. Define

Type I Error
Type II Error