**Module 8: Statistical Hypothesis Testing**

# Setting up a Hypothesis Test

Recall the intent to purchase example from Module 7.

The status quo belief of management is that 25% of a population of consumers intends to purchase a computer. A new sample seems to contradict that claim; only 14% in this sample intend to purchase a computer. Management needs to decide whether to reject the claim based on the new evidence.

The general problem of statistical hypothesis testing concerns using data to make a decision. In this setting, the decision is whether a hypothesis of interest should be rejected.

Jargon:

The hypothesis to be tested is called the *null hypothesis* and is denoted H0.

An *alternative hypothesis*, denoted Ha, is considered as an alternative to H0.[[1]](#footnote-2)

In the intent to purchase example, these two hypotheses can be expressed as

H0: μ = .25 versus Ha: μ ≠ .25

where μdenotes the unknown true population proportion.

Don’t take such hypotheses too literally. If μ =.2500001, it makes sense to retain H0 because it remains a rather good description of the population. Instead, think of the null here as saying “It’s reasonable to treat the mean of the population as .25.”

# The One-Sample t Test

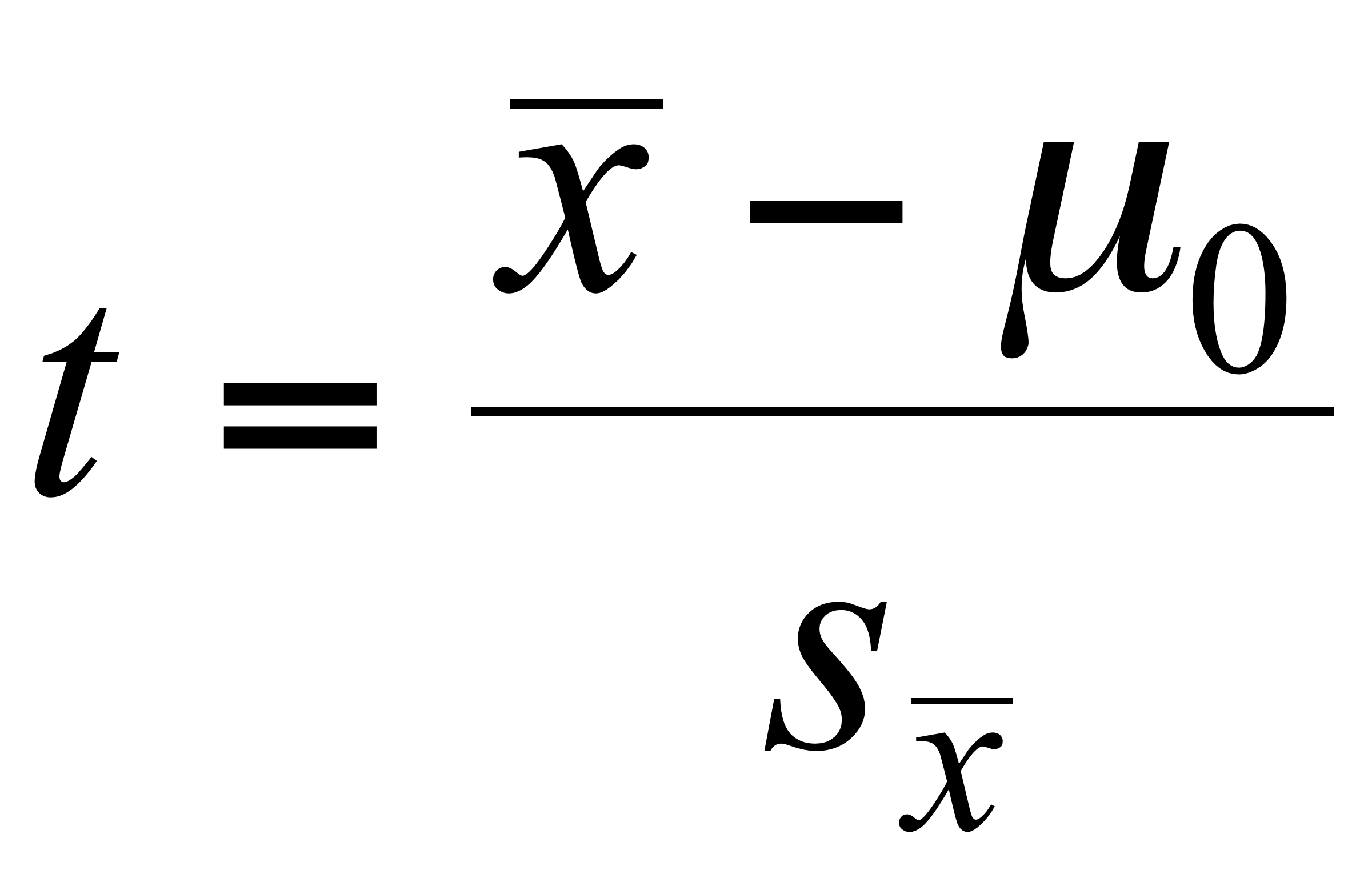
Suppose we have a random sample from a population with unknown mean μ.

A common set of hypotheses often considered for this setup is[[2]](#footnote-3)

H0: μ = μ0 versus Ha: μ ≠ μ0

Note that the intent to purchase hypotheses is the special case with μ0 = .25.

The key quantity for testing H0 here is

,

which is called a *t statistic* or a *t ratio*. The *t* statistic counts the number of standard errors between the observed statistic () and the hypothesized population parameter (μ0).

Intuition: a large *t* statistic implies that the data are implausible if the null hypothesis were true. We interpret a large *t* statistic as evidence against H0.

Example: Testing[[3]](#footnote-4) the intent to purchase hypothesis H0: μ = .25 vs. Ha: μ ≠ .25.



The estimate of the unknown proportion μ is the sample proportion  = .14.

In units of standard errors, how far is .14 from the hypothesized value μ0 = .25?

# The p-value: How Extreme is Enough to Reject H0?

Key issue: How large should *t* be in magnitude (positive or negative) in order to convince us to reject H0?

To answer this question in the previous example, JMP provides the quantity:

Prob > | t | = .0021

which is the probability of observing a *t* statistic more extreme than -3.15 (positive or negative) if in fact H0: μ = .25 is true.[[4]](#footnote-5)

Thus, if μ were .25, a |t-statistic| larger than 3.15 would occur only 0.21% of the time!

The quantity Prob > | t | = .0021 is called a *p-value* and it measures the rarity of the data when H0 is true.

“Small” p-values indicate one of two things: either H0 is false or else something very unusual has happened.

Faced with these two choices, statistical practice is to reject H0 when the p-value is “small” enough.

# The “Official” Rules for hypothesis Testing

Procedure to test a pair of hypotheses is:

1) Pick a value ** called the *significance level* (traditionally ** = .05 or .01).

2) If the p-value ≤ **, reject H0 and declare the result to be *statistically significant*(at the ** level of significance).

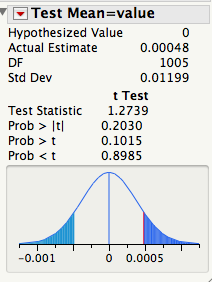
3) Otherwise, if the p-value > **, the result is said to be *not statistically significant.*

Should the intent to purchase hypothesis H0: μ = .25 be rejected at the .05 level of significance? At the .01 level of significance?

Example: Microsoft returns. Based on the four years of data in *Microsoft\_Subset.JMP*, let’s test

H0: μ = 0 versus Ha: μ ≠ 0

where μ is the unknown mean of the population of GM daily returns. JMP yields[[5]](#footnote-6)



Can H0 be rejected at the .05 level of significance?

If you don’t reject *H*0, does that mean that it’s true?

# The "2 Standard Error" Rule of Thumb

For practical purposes, a good *approximate* test of

H0: μ = μ0 versus Ha: μ ≠ μ0

is to

reject H0: μ = μ0 at the =.05 level when |t-statistic| ≥ 2

(i.e., reject H0 when  is more than 2 standard errors away from the hypothesized value μ0).

# Using Confidence Intervals to Test Hypotheses

The traditional choice of the significance level of a test is **= 5%. Also, the traditional choice of the level of confidence of a confidence interval is 95% (= 100% - 5%).

This correspondence[[6]](#footnote-7) is not accidental!

An alternative way to test H0: μ = μ0 versus Ha: μ ≠ μ0 at the **=.05 level of significance uses a confidence interval:

Reject H0: μ = μ0 when μ0 lies outside the 95% CI for μ

For example, for the intent to purchase example, the 95% CI for the population proportion is (.07, .21).

Because the conjectured value 0.25 does not lie inside the confidence interval, we can reject H0: μ = .25 at the **=.05 level since μ0 = .25 is not contained in (.07, .21).

This seems reasonable since a confidence interval is the set of plausible values for μ given the data.

# Summary of Testing H0: μ = μ0 versus H1: μ ≠ μ0

The following conditions are equivalent.[[7]](#footnote-8)

(a) The *p*-value is less than 0.05.

(b) The absolute value of the *t*-statistic is larger than 2 (i.e., |t| > 2).

(c) The 95% confidence interval does not contain *μ*0.

In each case, we reject H0: μ = μ0.

# The Two-Sample t Test

Suppose we have *two* *independent* random samples[[8]](#footnote-9):

*x*1,…,*xn* from a population with unknown mean μx

*y*1,…,*yn* from a population with unknown mean μy

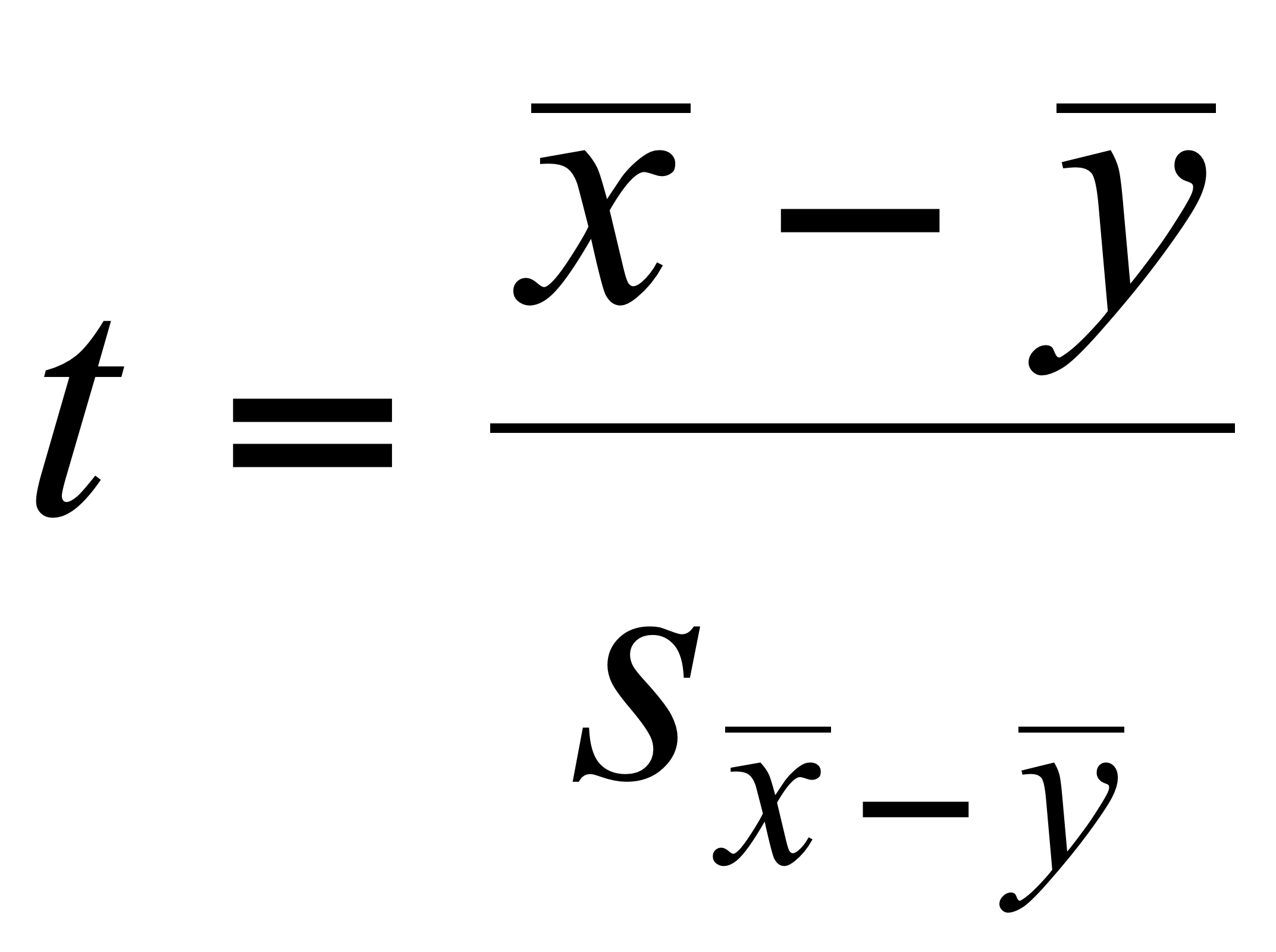
and we want to test the null hypothesis that the means of the populations are the same [[9]](#footnote-10)

H0: μx = μy versus Ha: μx ≠ μy

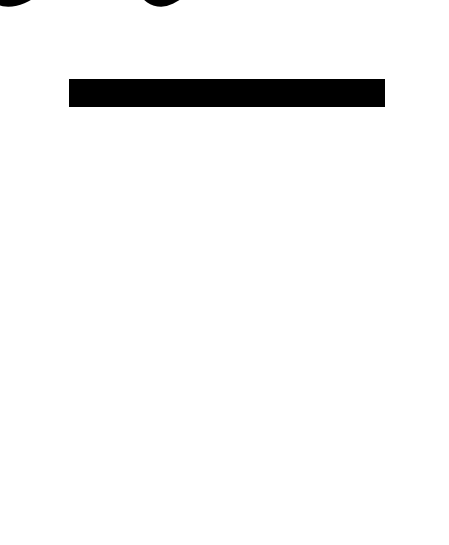
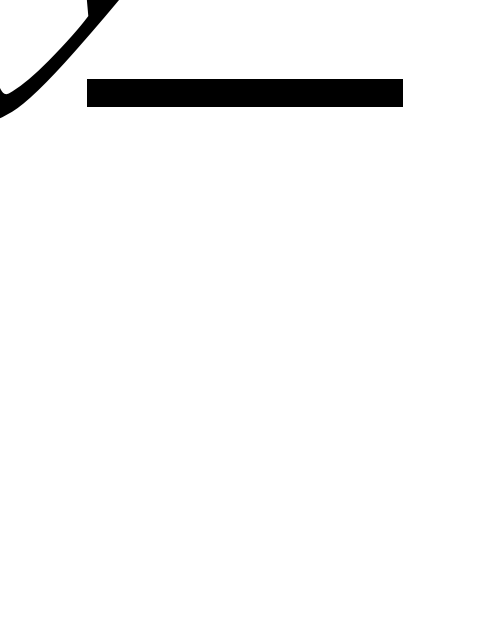
These hypotheses can also be written as

H0: μx – μy = 0 versus Ha: μx – μy ≠ 0

To test this hypothesis, the test statistic of interest is the two-sample *t* statistic



[NEED TO INSERT SYMBOLS HERE, X-BAR, YBAR]

counts the number of standard errors between the observed statistic (-) and 0. [[10]](#footnote-11)

Associated with this *t* statistic is a p-value. The p-value is the probability (under H0) of observing a larger *t* statistic (positive or negative) than the observed *t*.

The test proceeds just as before:

If p-value ≤ **, reject H0 and declare the difference to be statistically significant (at the ** level of significance).

Alternatively, we can perform the test with a confidence interval:

Reject *H*0 if 0 lies outside the 95% confidence interval for the difference in the two means.

# Example

A car manufacturer uses the price of used cars to determine the initial cost charged to customers who lease its automobiles.

The file *UsedCars.JMP* contains the prices of 155 used BMW automobiles divisions. Some are the xi model (4-wheel drive) and the others are the standard i model.

Comparison boxplots of the prices of the two types of cars show considerable overlap.[[11]](#footnote-12)



To judge if the population means differ, it is useful to compare the 95% confidence intervals for μi and μxi (BBS, p.164).[[12]](#footnote-13)



Because these intervals *do not* overlap, we are assured that the difference between the means is significantly different from zero.[[13]](#footnote-14)

We can confirm these results (and get a more detailed comparison) with the two-sample test. A two-sample *t* test of (BBS, p.164)

H0: μi = μxi versus Ha: μi ≠ μxi

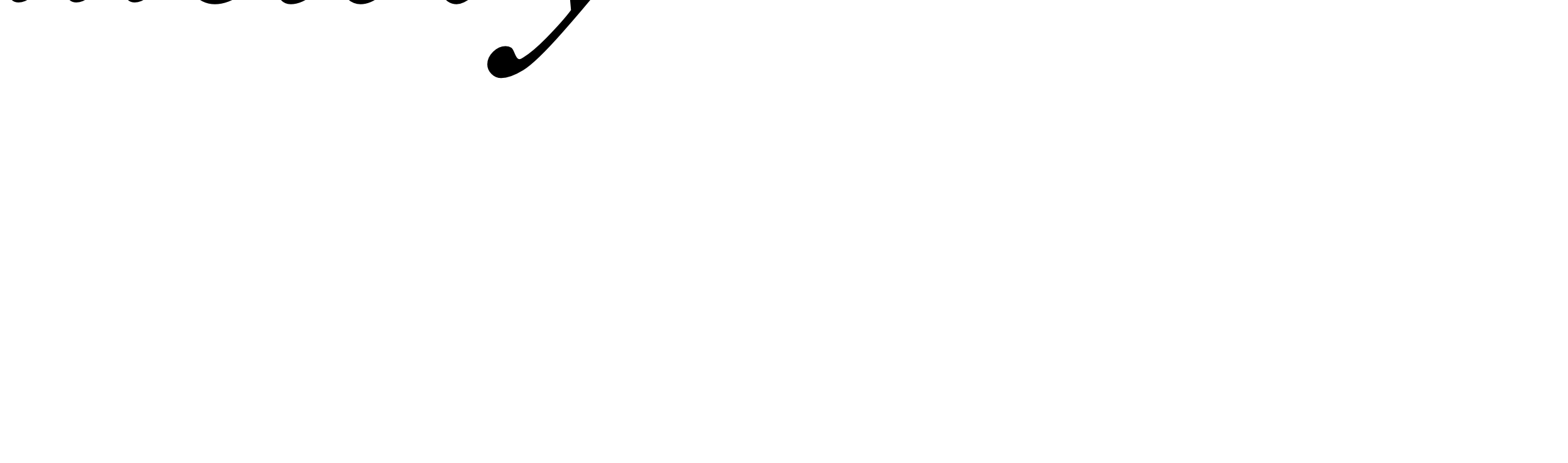
yields[[14]](#footnote-15)



This result is statistically significant because the p-value = 0.0023 is less than 0.05. Thus, we can reject H0 at the .05 level.

Note that we can also reach this conclusion by noting that |*t*| = 3.11is larger than 2 or that the 95% CI for (μxi - μi), here (619, 2780), *does not* contain 0.

# The Paired t Test

When comparing the population means of two samples, say *x*1,…,*xn* and *y*1,…,*yn*, it is sometimes useful to consider the paired differences , and treat*d*1,…,*dn* as a random sample from a population with unknown mean μd = (μx - μy). When might such pairings be natural?

In this case, testing

H0: μx = μy versus Ha: μx ≠ μy

is equivalent to testing

H0: μd = 0 versus Ha: μd ≠ 0

H0: μd = 0 can then be tested using the one-sample *t* test.

# Example (BBS p 163)

Management of a newly merged pharmaceutical company needs to reduce its sales force. Does the sales force from the “GL” division differ from the sales force from the “BW” division? The sales data in *Pharmasal-split.JMP* suggest a natural pairing within each of the 20 sales districts which leads to the variable: Differences = BW – GL. The hypothesis H: μBW = μGL is then equivalent to H: μd = 0. What should we conclude?[[15]](#footnote-16)



# Testing Other Statistical Hypotheses

Hypothesis testing is not restricted to statements about mean values alone.

Many other hypotheses are of interest in statistical analysis. For example, useful null hypotheses make claims about other features of populations.

H0: population is normal

H0: population correlation is 0

H0: two populations are identical

H0: time series is iid

In all of these settings, conventional statistical practice proceeds as follows:

1) Calculate a p-value for the disparity between the null hypothesis H0 and the data.

2) If p-value ≤ **, reject H0 and declare the result to be statistically significant (at the ** level of significance).

3) If p-value > **, the result is not statistically significant.

We will describe some of these tests in more detail in Stat 621.

# What You Will Need to Know For 621

Graphical tools - Histogram, boxplot, comparison boxplots, and scatterplot.

Mean, variance and correlation - Mean is the average value. Variance is the average squared deviation from mean. Correlation measures the strength of linear association.

Normal distribution - 95% of the distribution lies within  ± 2*σ*. Normal quantile plot as a diagnostic.

Sampling Distributions - Random Sampling; iid samples. Sample-to-sample variation of a statistic. Standard Error: SE() = s/√n .

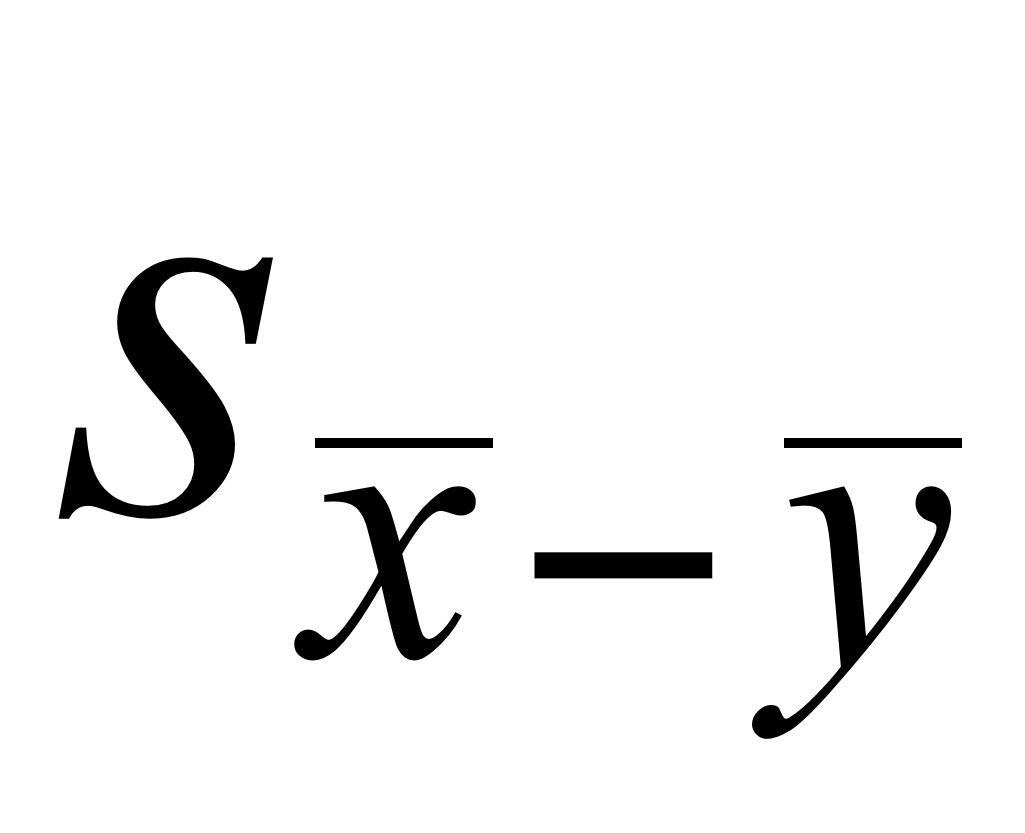
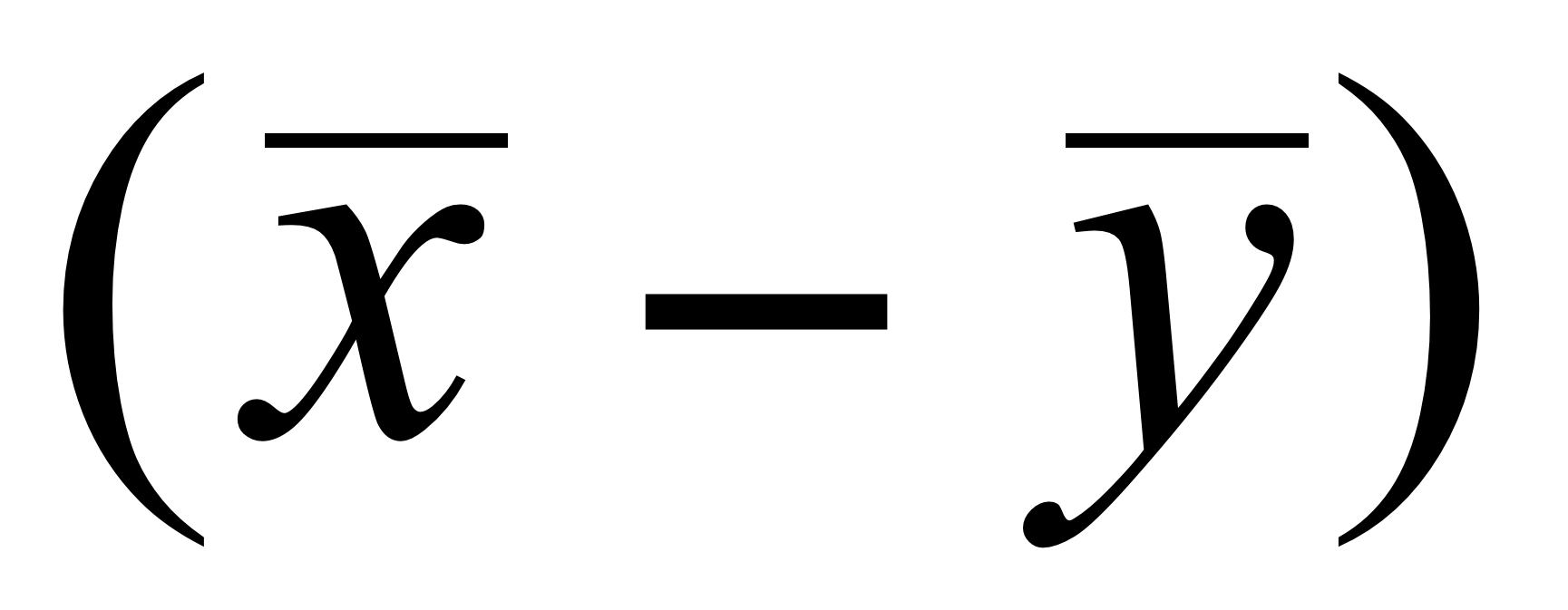
Confidence interval - 95% CI - Estimate ± 2 SE(Estimate). Interpretation.

Hypothesis test - t-statistic/t-ratio counts the SE’s from conjectured value.  
p-value measures “plausibility” of H0. Reject H0 at the .05 significance level <=>

p-value < 0.05 <=> | t-statistic| > 2 <=> hypothesized value lies outside 95% CI.

**JMP Software**

See you in September!

1. Some textbooks denote the alternative hypothesis as H1 rather than Ha. Conventions for the notation of hypothesis tests vary. [↑](#footnote-ref-2)
2. Hypotheses of this form are sometimes called “two-sided” hypotheses to distinguish them from hypotheses of the form H0: μ ≥ μ0 vs. Ha: μ < μ0 or H0: μ ≤ μ0 vs Ha: μ > μ0, which are called “one-sided.” Such refinements of the hypotheses are used in certain situations. We’ll stick to two-sided methods here. [↑](#footnote-ref-3)
3. For the data in *CompPur.JMP*, apply Analyze > Distribution to the column *Intend*, select Test Mean after clicking on the title bar, and enter .25 for the Hypothesized Mean to obtain this output. [↑](#footnote-ref-4)
4. For testing the hypotheses H0: μ ≤ .25 versus Ha: μ > .25, JMP reports the p-value of  = .14 as Prob > t = .9989. For testing the hypotheses H0: μ ≥.25 versus Ha: μ < .25, JMP reports the p-value of  = .14 as Prob < t = .0011. [↑](#footnote-ref-5)
5. Again, use the Test Mean command with Analyze > Distribution. [↑](#footnote-ref-6)
6. The correspondence described here works for “two-sided” hypotheses. See the Basic Business Statistics (BBS) casebook (BBS), pp. 102-103. [↑](#footnote-ref-7)
7. This equivalence assumes that the sample size is large enough so that we can use the form ±2 SE to form a 95% interval. For small samples with t-statistics close to ±2, use the *p*-value rather than the empirical rule. [↑](#footnote-ref-8)
8. For inferential purposes, the ideal method for getting two independent samples is from a *randomized experiment* as done in the pharmaceutical industry to show the value of a new drug. Otherwise, we must sort out the possibility of *confounding* (BBS, Class 9). [↑](#footnote-ref-9)
9. It would be pretty rare to find two populations with *exactly* the same mean, measured to infinite precision. As before, interpret H0 as saying that the means are close and offer a description of the populations that is consistent with the data. [↑](#footnote-ref-10)
10.  stands for the standard error of  which has a formula that you can safely ignore. If you are interested, it is found in precisely the same way that we found the variance of portfolios, see BBS, p 212. It is based on the fact the variance of the difference of the two means is the sum of there variances when the two means are independent, a consequence of Fact 2 in Module 5. [↑](#footnote-ref-11)
11. To horizontally perturb the points so you can see them all, as in this plot, click on the red triangle and select Display Options > Points Jittered. [↑](#footnote-ref-12)
12. To get this output, use Fit Y by X. But rather than select Box Plots as you would for comparison boxplots, select Mean Diamonds. To get the output underneath, select Means and Standard Deviation from the red triangle menu. [↑](#footnote-ref-13)
13. The algebra for this claim appears in BBS, p.145. If the intervals don’t overlap, the difference is significant. If they overlap, you need to be more careful. [↑](#footnote-ref-14)
14. Use the JMP Fit Y by X command, then right-click on the title bar and select t Test. If you have good reason to justify assuming the two populations have equal variances, you can get a bit more precision by selecting the Means/Anova/Pooled t. [↑](#footnote-ref-15)
15. A two-sample t test without pairing fails to find a significant difference. Use Analyze >Distribution and select Test Mean to obtain these results. JMP also offers a specific method for paired comparisons (see BBS pp.165-166). That procedure yields the same result. [↑](#footnote-ref-16)