Testing for Differences - Continuous Data

(aka Hypothesis Testing)

$\leq \neq \cong \geq$

There are many forms of Testing:

Variable Data
(discussed in this tab)

1. F -test

   (Compares Variances)

   - Levene's Test
   - Bartlett's Test

2. t-test

   (Compares Means)

   - 1 Sample t-test
   - 2 Sample t-test
   - Paired t -test

Discrete Data

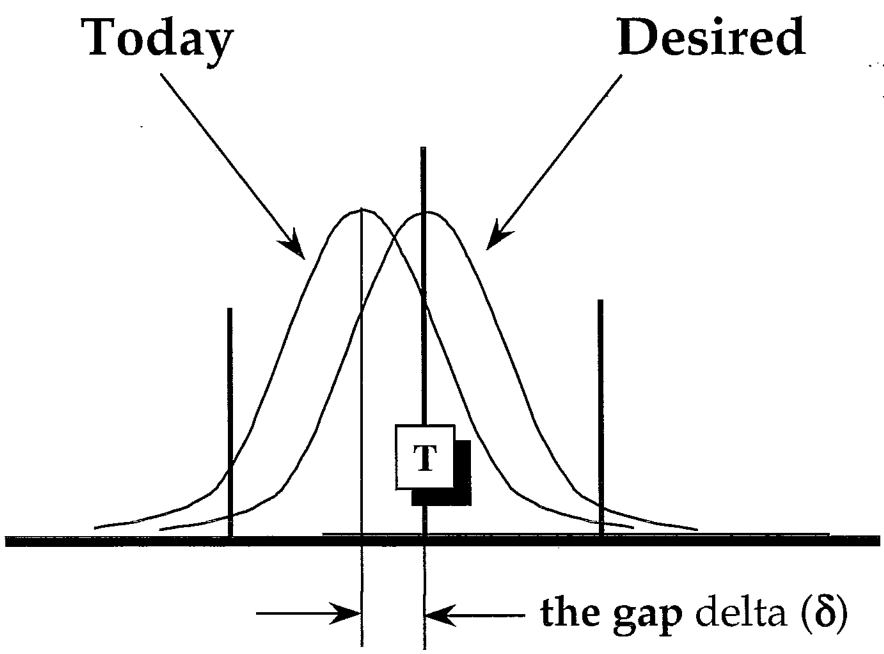
Chi Square Test

(Compares Counts or Frequencies)

- Goodness of Fit
- Contingency Table

Choose the Test based on what you want to compare

Comparing the Observed to the Expected



Determining a Statistical Difference

Hypothesis Tests as an alternative method for determining difference

Confidence intervals give a **range of plausible values** for a population value (parameter).

Hypothesis tests determine if an apparent difference is **real** or could be due to **chance**. We can quantify our level of confidence that the difference is real.

Why Do We Use Hypothesis Tests and Confidence Intervals?

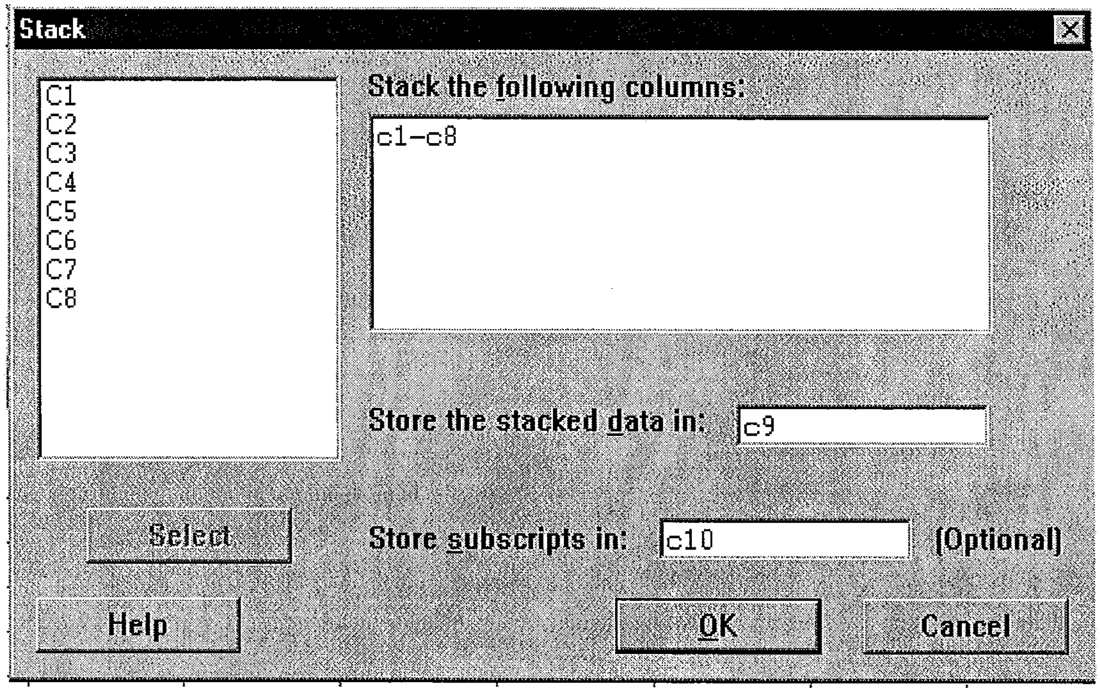
2. Stack in one column.

>Manip >Stack/Unstack > Stack

Stack c1-c8.

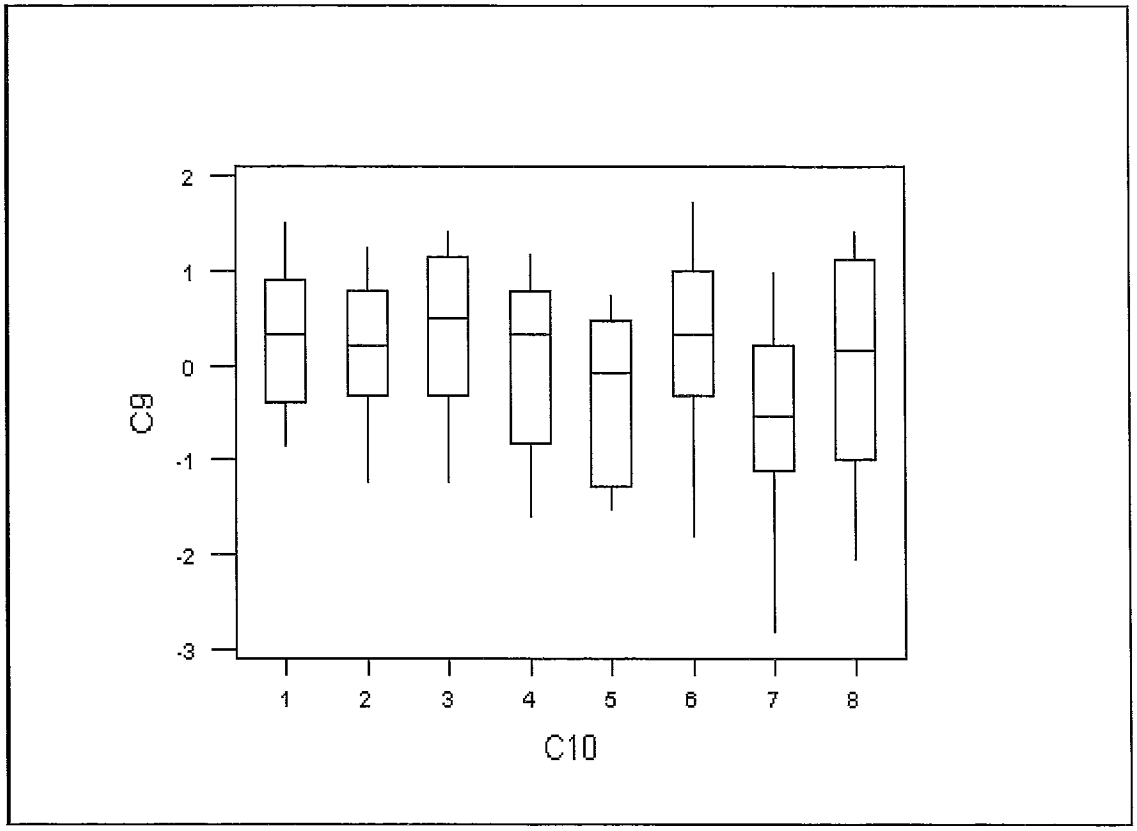
Store the stacked data in c9.

Store subscripts in c10.



Why Do We Use Hypothesis Tests and Confidence Intervals?

4. Note that there are differences in the sample averages and variation, even though all 8 sets of data came from the same population.



Interpreting Test Results:

Hypotheses can be accepted or rejected by 3 different methods:

Method 1

If the **Calculated Value is < or = to Table (Critical) Value, no conclusions can be drawn** (fail to reject H_o).

If the **Calculated Value is > Table (Critical) Value, then a difference exists** (reject H_o, accept H_a).

OR

Method 2

If the **p value is > or = to α, no conclusions can be drawn** (fail to reject H_o).

If the **p value is < α, then a difference exists** (reject H_o, accept H_a).

OR

Method 3

If **"0" falls within the Confidence Interval** of the difference of two means, **then no conclusions can be drawn** (fail to reject H_o - the difference could be zero, i.e. no difference).

If **"0" falls outside the Confidence Interval** of the difference of two means, **then a difference exists** (reject H_o, accept H_a).

Recommend Methods 2 & 3 because Minitab does the work!

Compare the Variances: F - Test

Why use it?

- The F - test is for testing that the variances of two or more distributions are equal. (Null Hypothesis is equal variances)

- To compare "Xs" that were recorded during baselining

- If you make a change to a process and want to determine whether or not the variance in the output was changed, you can compare samples before and after the change using the F - test.

When should it be used?

- If you make a change to a process and want to determine if the variance in the output was changed, you can compare variances before and after the change
- **Before the t-test** (need to understand if the variances are equal first)
- On normal or non-normal data

How is an F-test performed?*

- Hand calculations for 2 distributions
- Minitab for 2 or more distributions

F-Test is calculated using "Homogeneity of Variance" in Minitab

*See Appendix

F - Test calculated by Hand

Example: Compare the variances of fixture 1 and fixture 2

Data (height) : Fixture 1 -- 5.39, 5.389, 5.39, 5.389, 5.388,
5.391, 5.391, 5.391, 5.391, 5.389
Fixture 2 -- 5.387, 5.387, 5.387, 5.387, 5.388,
5.388, 5.389, 5.389, 5.388, 5.387

σ_1 = .00110 standard deviation of fixture 1
σ_2 = .000823 standard deviation of fixture 2
Sample size is 10 for each -- 9 degrees of freedom for each.

Calculated F = $.00110^2 / .000823^2$ = **1.79**

The critical value for a F distribution with 9 degrees of freedom in the numerator and 9 degrees of freedom in the denominator is **3.18**, from the F table (on the next page).

The calculated F is lower than the tabled value of F, so we accept the null hypothesis that the variances are equal.

Decision criteria: If the Calculated value > Table value, reject H_o

Conclusion: There is not sufficient evidence to claim with 95% confidence that the variation has changed.

F - Test by Minitab

Example :

The machined lower transmission housing on the washing machine has 10 CTQs, two of which primarily affect stack-up height of brake and clutch assemblies, which influences brake performance.

At this point, data has been collected with a rational subgroup plan, taking into consideration shift, cavity and fixture.

- The data was checked for normality; found to be non-normal
- Histogram showed a bi-modal distribution
- Normality Test was performed on each fixture to see if the fixtures were causing the non-normal condition (some fixtures were normal, some were non-normal)

There are 8 different fixtures. Could fixture be a possible "X"?

- Graph the data (look for obvious differences)
- Choose a Hypothesis Test
- Develop the Hypotheses
- Perform the test(s)
- Draw a conclusion

Open file (worksheet file) "lth.mtw" in Minitab

C:\Program Files\Mtbwin\Data\Ph1_data\lth.mtw

Perform an F-test to compare the variances of the 8 fixtures

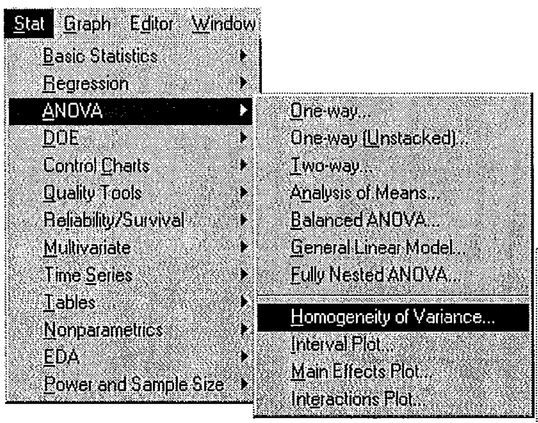
What are the Hypotheses?

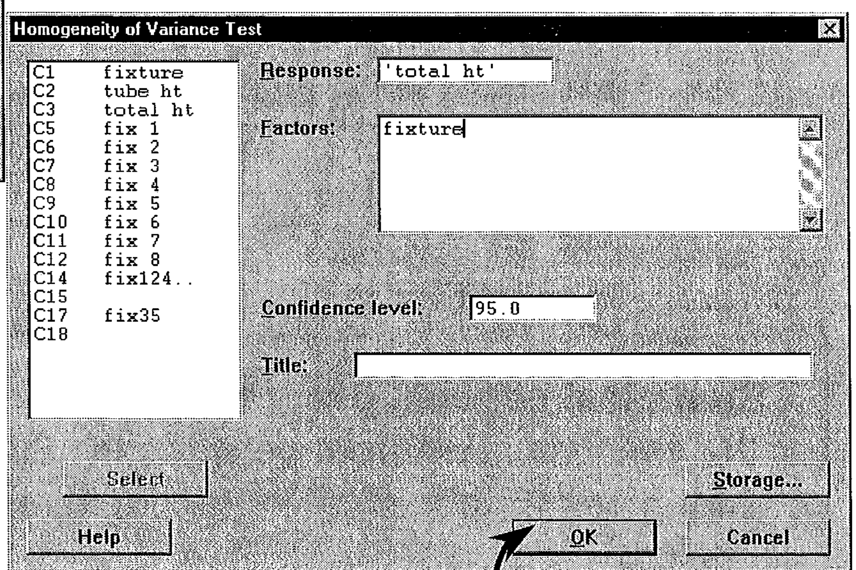
H_o:

H_a:

What is the rejection criteria?

Stat>ANOVA>Homogeneity of Variance





Click OK to run command

Minitab Session Window Report for Homogenity of Variance

```
Response     total ht
Factors      fixture
ConfLvl      95.0000

Bonferroni confidence intervals for standard deviations

  Lower      Sigma      Upper    N  Factor Levels

6.62E-04  1.10E-03  2.66E-03  10   1
4.95E-04  8.23E-04  1.99E-03  10   2
6.98E-04  1.16E-03  2.81E-03  10   3
8.24E-04  1.37E-03  3.32E-03  10   4
8.51E-04  1.41E-03  3.42E-03  10   5
8.53E-04  1.42E-03  3.43E-03  10   6
8.99E-04  1.49E-03  3.62E-03  10   7
6.62E-04  1.10E-03  2.66E-03  10   8

Bartlett's Test (normal distribution)

Test Statistic: 4.298
P-Value       : 0.745

Levene's Test (any continuous distribution)

Test Statistic: 0.818
P-Value       : 0.576
```

Displays the same information as the Graph

Compare the Means: t - Test

Why use it?

- A t - test is for testing that the means of two distributions are equal.

- To compare baseline, "X" information

- If you make change to a process, and want to determine whether or not the output was changed, you can compare samples before and after the change using the t -test.

When should it be used?

- **<u>After the F-test</u>**
- On Normal or close to Normal Data
- When there are 30 samples or fewer
  (if more than 30, do a Z transformation)
- When comparing a <u>mean to a target value</u>
- When comparing <u>means of 2 or more distributions</u>
- When comparing <u>paired data</u>
  (Do multiple steps change the "Y" response?)

How to perform a t-test:

- Calculated by hand for 2 distributions
- Minitab for 2 or more distributions

How to Perform the t-Test...

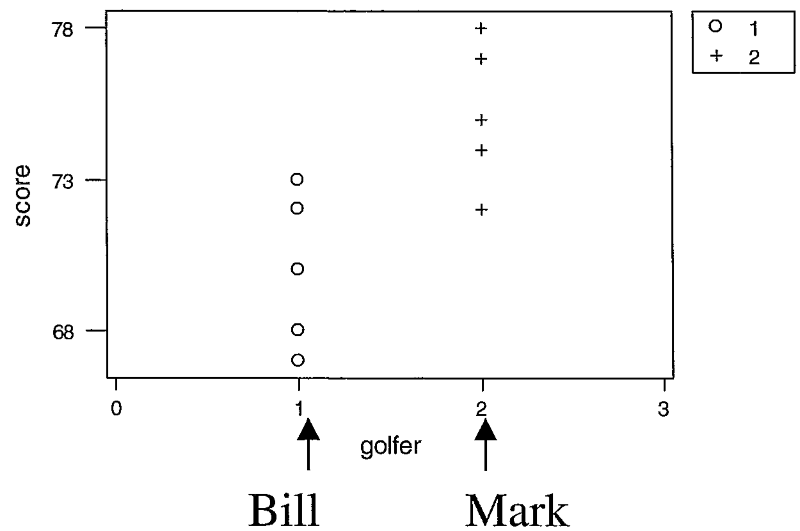
Minitab Method (Cont'd)

- **2-sample t** Compares 2 distribution means
  Compares a paired distribution (Paired t - test)

Hypotheses

$H_o = \mu_1 - \mu_2 = 0$
$H_a = \mu_1 - \mu_2 \neq 0$
$H_a = \mu_1 - \mu_2 < 0$
$H_a = \mu_1 - \mu_2 \geq 0$

Comparing means of 2 or more distributions
& Paired t -test



Fail to reject H_o when $p \not> .05$; Accept H_a when $p < .05$

Confidence Interval

Lower Confidence Limit = $x - t_{(\alpha/2,\ df)} \frac{s}{\sqrt{n}}$

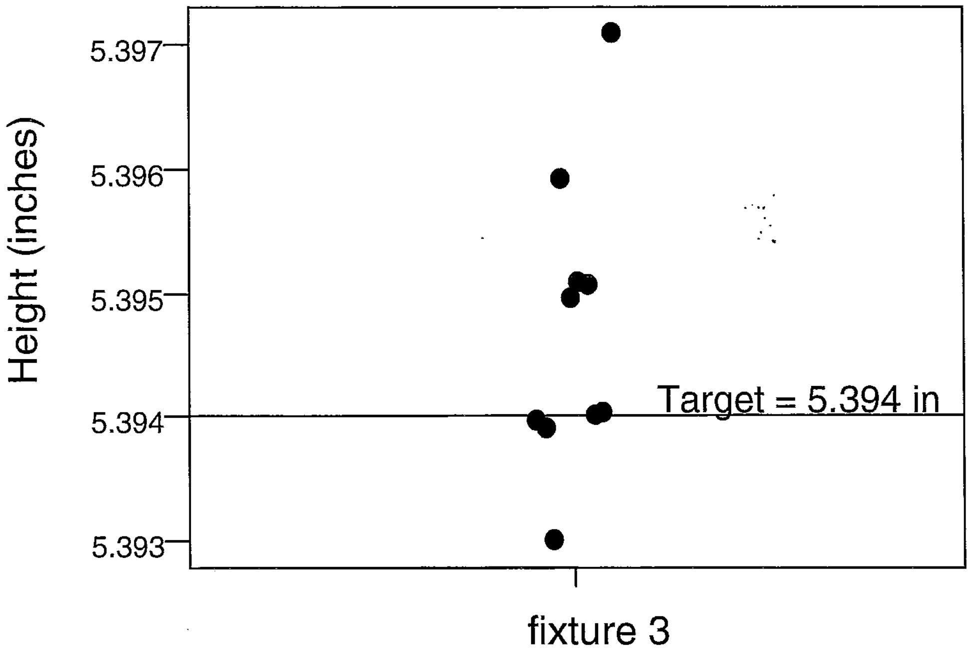
Upper Confidence Limit = $x + t_{(\alpha/2,\ df)} \frac{s}{\sqrt{n}}$

Where:

x = sample mean
t = t statistic from t table
α = alpha risk
df = degrees of freedom = n -1
s = sample standard deviation
n = number of data points in the sample

Graph the Data

Height of 10 parts from fixture 3
5.394
5.394
5.393
5.394
5.394
5.395
5.396
5.397
5.395
5.395



$n = 10$

$\bar{x} = 5.3947$

$\hat{\sigma} = 0.00116$

What is a range of plausible values for the population average? Could the difference between $\bar{x}$ (5.3947) and the target (5.394) be due to chance?

t table

$\alpha = .05$
$\alpha/2 = .025$

9 degrees of freedom

df	1-α .600	.700	.800	.900	.950	.975	.990	.995
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

Example - Continued

Confidence Interval Interpretation

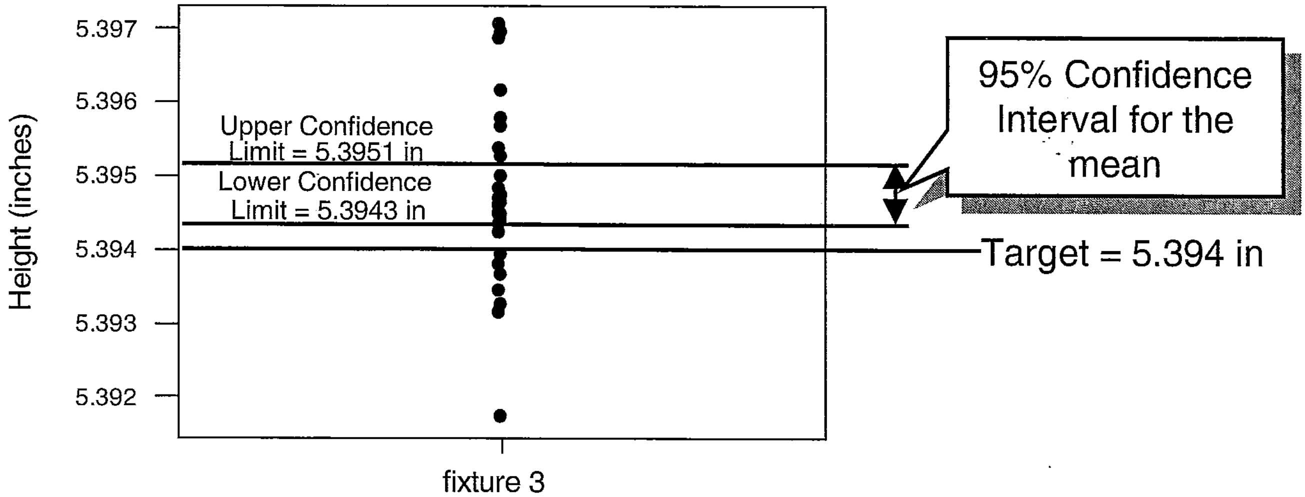
95% of intervals constructed in this manner will be correct (contain the true population mean), and 5% of intervals constructed in this manner will be incorrect.

- The target, 5.394, is contained in the interval.
  Statistical evaluation: There is no evidence to prove that the average height of the parts made in fixture 3 are not on target.
  Practical evaluation: The target is just barely in the confidence interval. The calculation was based on only 10 data points and alpha = 0.05.

- Things you can do to investigate fixture 3 further using confidence intervals...
  ... Obtain additional samples (if practical) and calculate the confidence interval
  ... Calculate the confidence interval using different values of alpha

The Confidence Interval quantifies the uncertainty in the data.

Effect of Sample Size on the Confidence Interval - continued



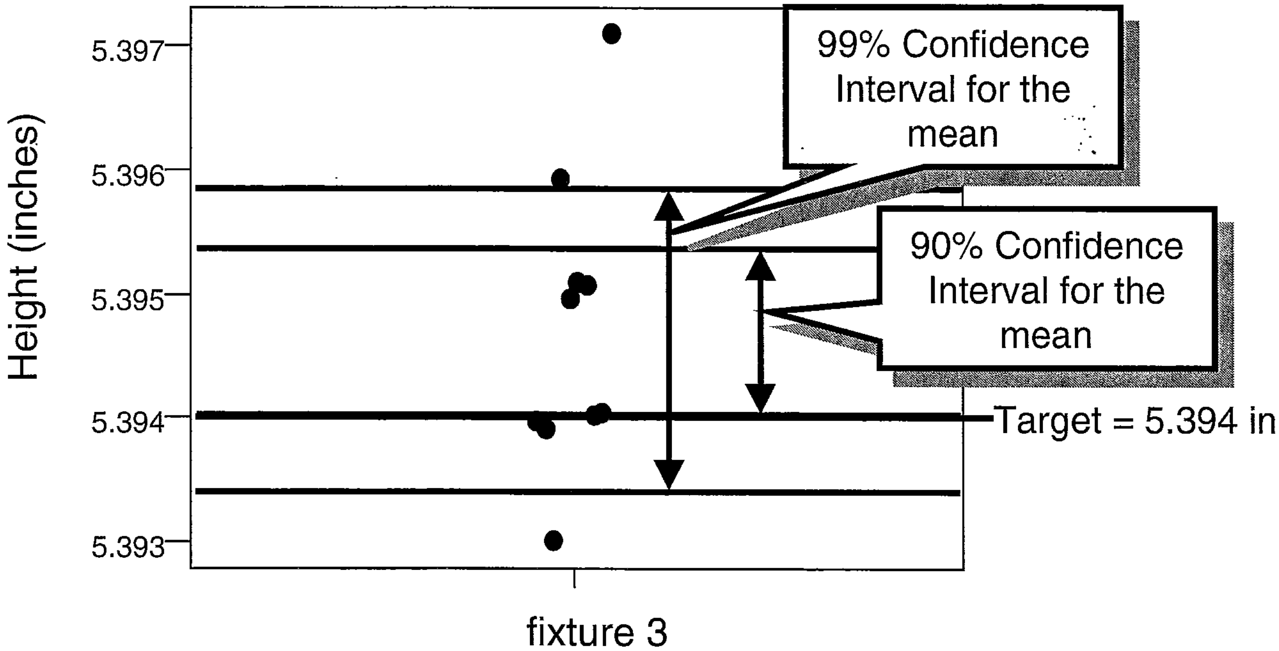
The 95% confidence interval with n = 10 was 5.3939 - 5.3955.
The 95% confidence interval with n = 30 was 5.3943 - 5.3951.

The only thing that changed was **n**.

With the additional samples, there **<u>is evidence to prove</u>** that the average height of the parts made in fixture 3 are not on target.

The Confidence Interval gets smaller as the sample size increases.

Effect of α on the Confidence Interval



The only thing that changed was α.

We **can** say with **90% confidence** that the parts made on fixture 3 are not on target.

We **can not** say with **99% confidence** that the parts made on fixture 3 are not on target.

The Confidence Interval gets larger as α decreases.

1-sample t - Test by Minitab
-- Perform the t-test to compare the means

What are the Hypotheses?

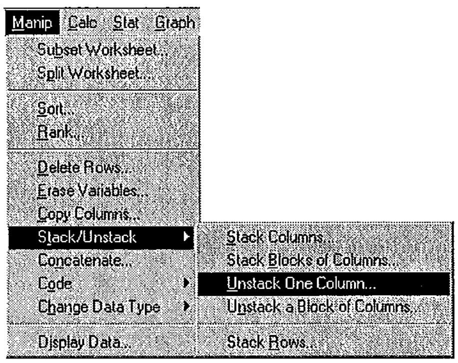
H_o:

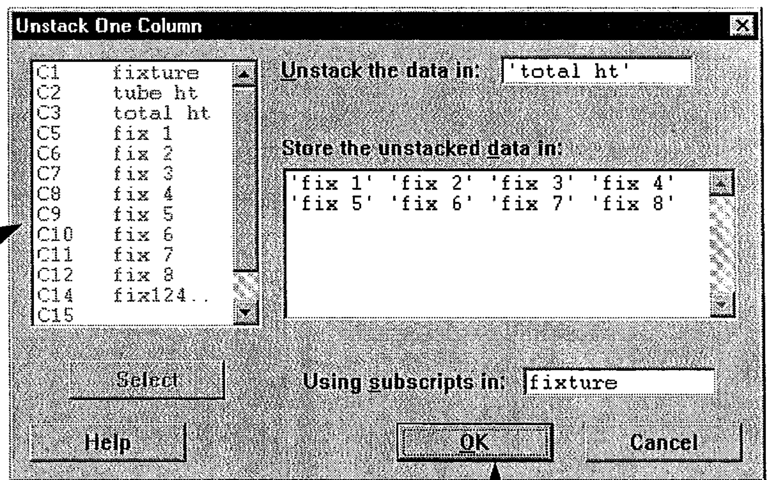
H_a:

What is the rejection criteria?

First: Unstack the data so each fixture can be compared separately

Manip>Stack/Unstack>Unstack One Column...





Enter the column headings (C5 - C12) in the spreadsheet before running the transaction

Click on “OK” to run

Running a 1-Sample t-Test

What does a 1 sample t-test compare?

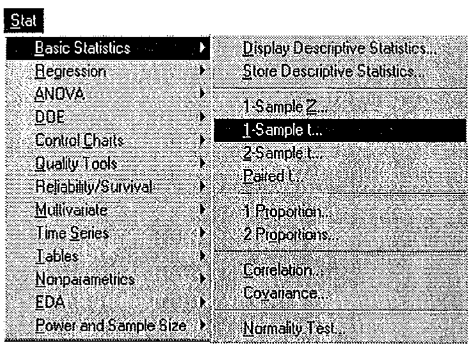
What are the Hypotheses?

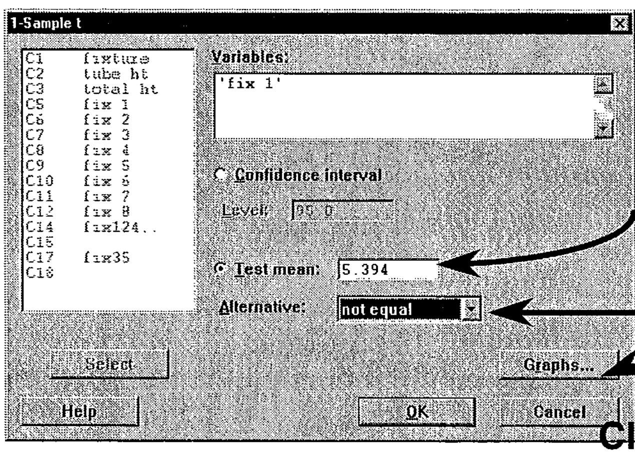
H_o:

H_a:

What is the rejection criteria?

Stat>Basic Statistics>1-Sample t



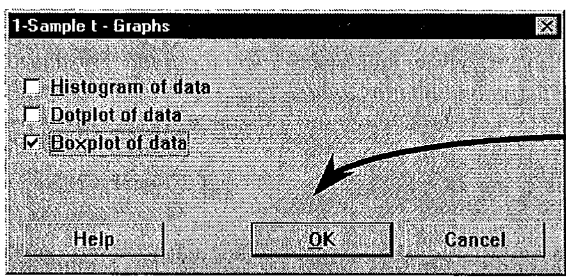


Type in the target mean

Must choose H_a

Click on "Graph"

Click "Boxplot of data"



Click "OK" twice to run

Let's try another example:
1-Sample t-Test for Fixture 3 to the Target Mean

Stat>Basic Statistics>1-sample t

T-Test of the Mean

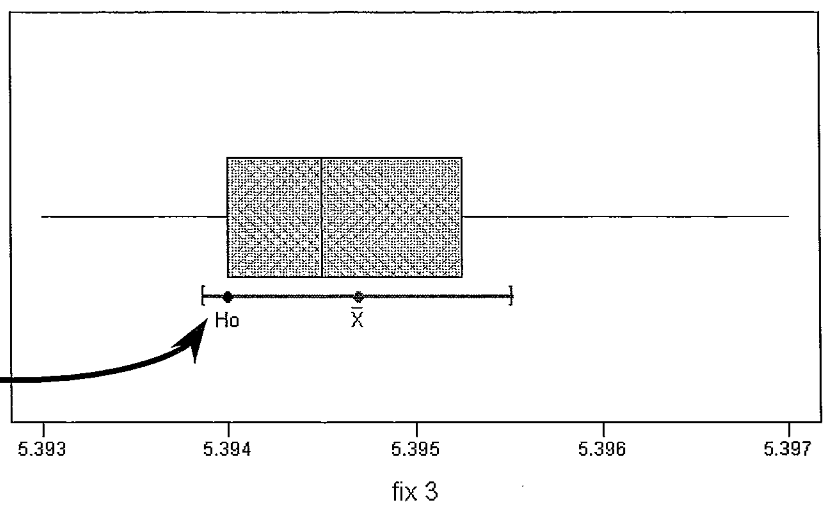
The Hypotheses for H_o and H_a

```
Test of mu = 5.39400 vs mu not = 5.39400

Variable      N      Mean     StDev   SE Mean        T        P
fix 3        10   5.39470   0.00116   0.00037     1.91    0.089
```

P value is > .05; accept H_o

Boxplot of fix 3

(with Ho and 95% t-confidence interval for the mean)



Ho indicates where the target of Fixture 3 falls in the Confidence Interval for the mean.

No statistical conclusion can be drawn at this time.
May need to take more data to show a statistical difference

2-sample t - Test by Minitab -- Compare Fixture-to-Fixture, or Groups of Fixtures to other Groups

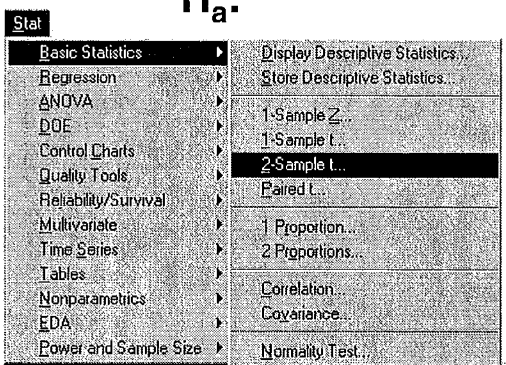
Compare Fixture 1 to Fixture 3:

What are the Hypotheses?

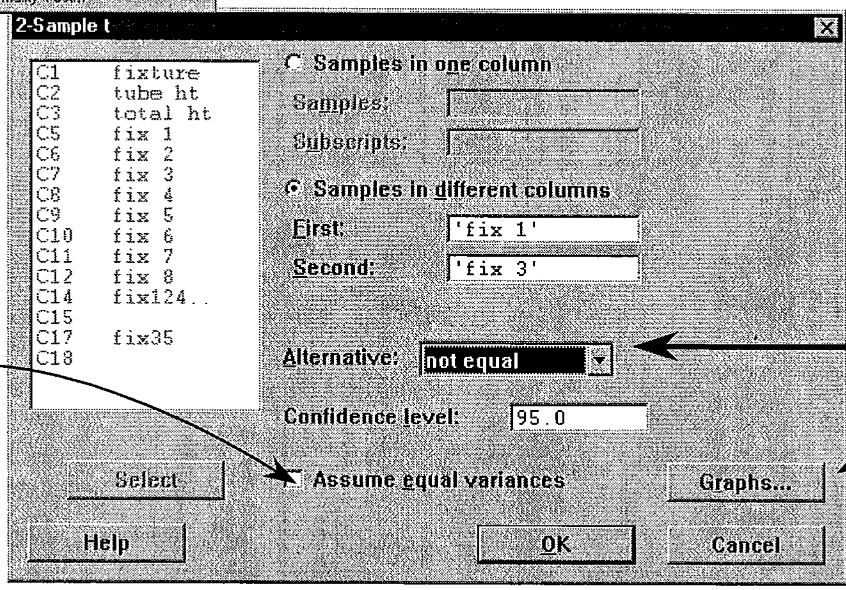
H_o:

H_a:

What is the rejection criteria?



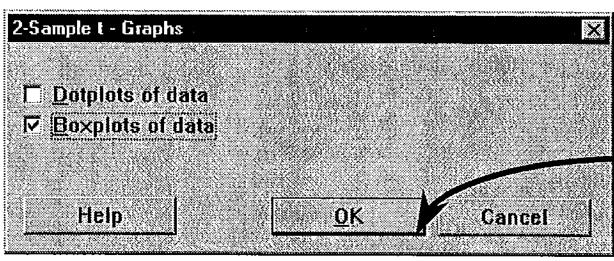
Stat>Basic Statistic>2-Sample t



Choose H_a

Click on "Assume Equal Variance" if the F-test accepted H_o

Click on "Graphs"



Click "Boxplots of data"

Click "OK" twice to run

** Please find 2-sample t-test by hand in the Appendix

Interpretation of the Confidence Interval

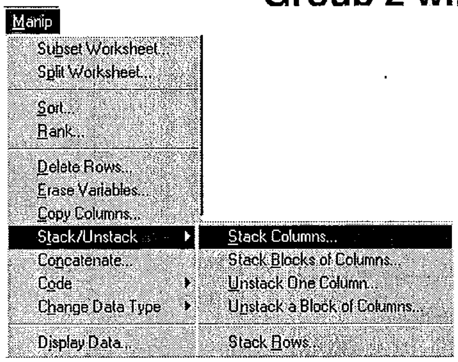
- The ***most likely estimate*** of the difference in the averages is: 5.3899 - 5.3947 = -0.0048.
- The ***true difference*** (if we had all the data in the population) may be higher or lower than this.
- We are ***95% confident*** that the true value is between -0.00586 and -0.00374.
- This is a ***range of plausible values*** (values consistent with the data) for the population difference.
- 95% of intervals constructed in this manner will contain the true population value. (You will be wrong 5% of the time.)
- 0.0 is not in the interval, so we have strong evidence that the difference between the fixtures is real, and not due to chance alone. This means we reject the null hypothesis that the 2 means are equal (H_o: $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$).

Perform a 2-Sample t-Test to Compare Fixtures 1,2,4,6,7 & 8 to Fixtures 3 & 5

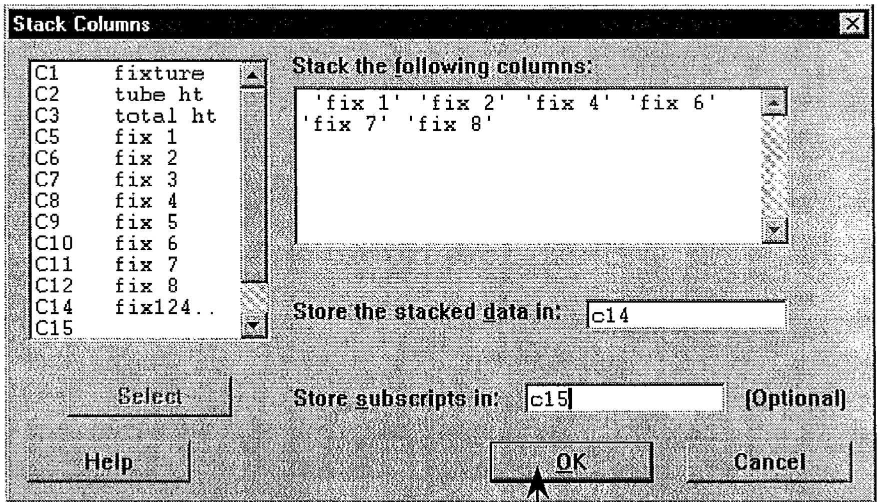
First, name the columns and then stack the data into groups

Group 1 will contain fixtures 1,2,4,6,7,8

Group 2 will contain fixtures 3 & 5

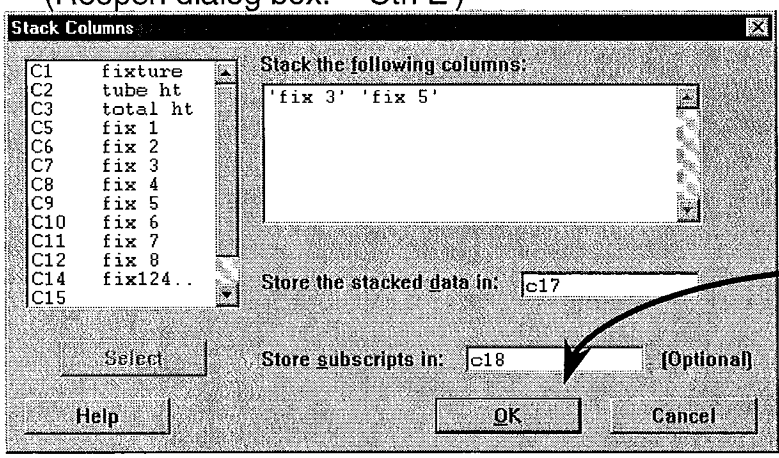


Manip>Stack/Unstack>Stack



Click "OK" to run

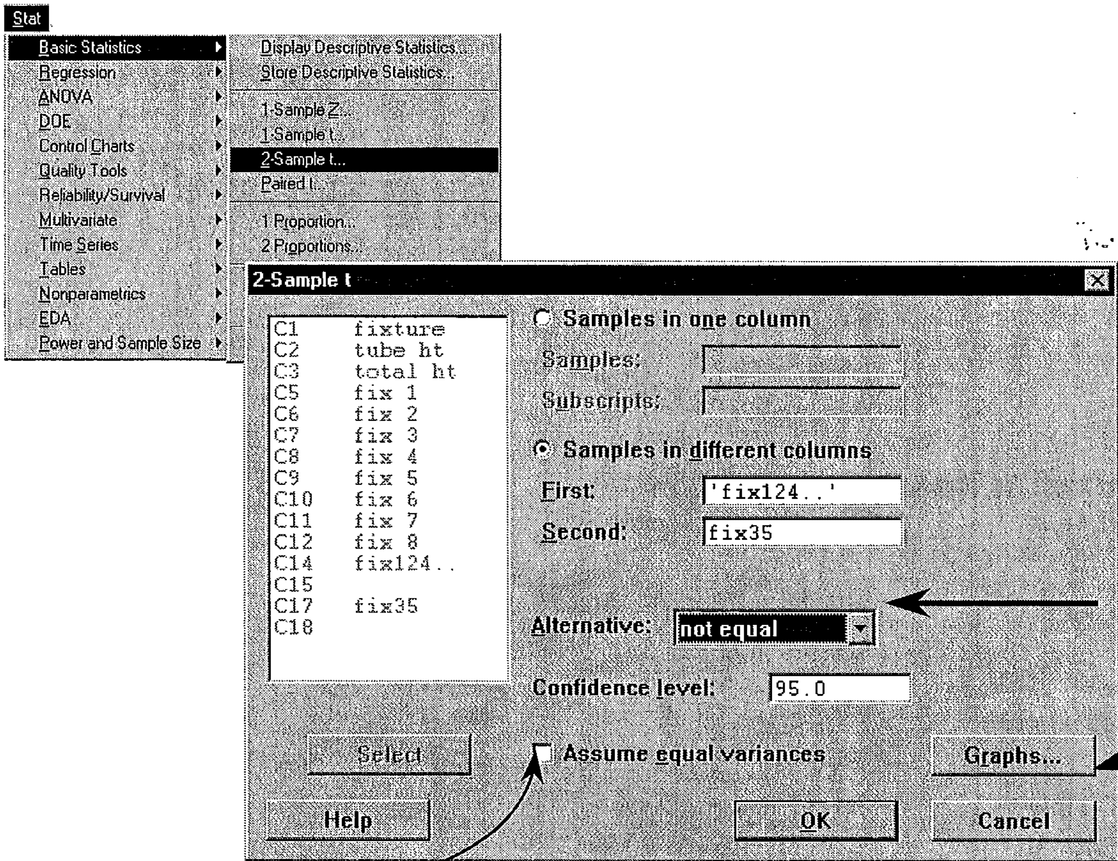
Repeat for columns C7 & C9
(Reopen dialog box: "Ctrl E')



Click "OK" to run

Running a 2-Sample t-Test

Stat>Basic Statistics>2-Sample t

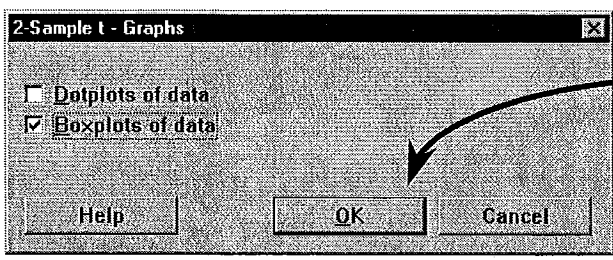


Choose H_a

Click on "Assume Equal Variance" if the F-test accepted H_o

Click on "graph"

Click "Boxplots of data"



Click "OK" twice to run

Class Exercise:

Perform a 2-Sample t-Test to Compare Fixture to Fixture, or Fixture(s) to Fixture(s)

1. Fixture 1 to Fixture 5

What are the Hypotheses?

H_o:

H_a:

What is the rejection criteria?

What conclusion can be drawn?

2. Fixture 3 to Fixture 5

What are the Hypotheses?

H_o:

H_a:

What is the rejection criteria?

What conclusion can be drawn?

3. Fixtures 1,2,4,6,7, & 8 to Fixture 3

What are the Hypotheses?

H_o:

H_a:

What is the rejection criteria?

What conclusion can be drawn?

4. Fixtures 1,2,4,6,7, & 8 to Fixture 5

What are the Hypotheses?

H_o:

H_a:

What is the rejection criteria?

What conclusion can be drawn?

Let's work an example...

Problem:

One operator has two calipers that he uses to perform his quality checks. He has been asked why he has two pairs of calipers, and whether having two pairs of calipers affects his accuracy of inspection.

The inspector collected data to answer the question: "Do the two calipers provide the same measurements?"

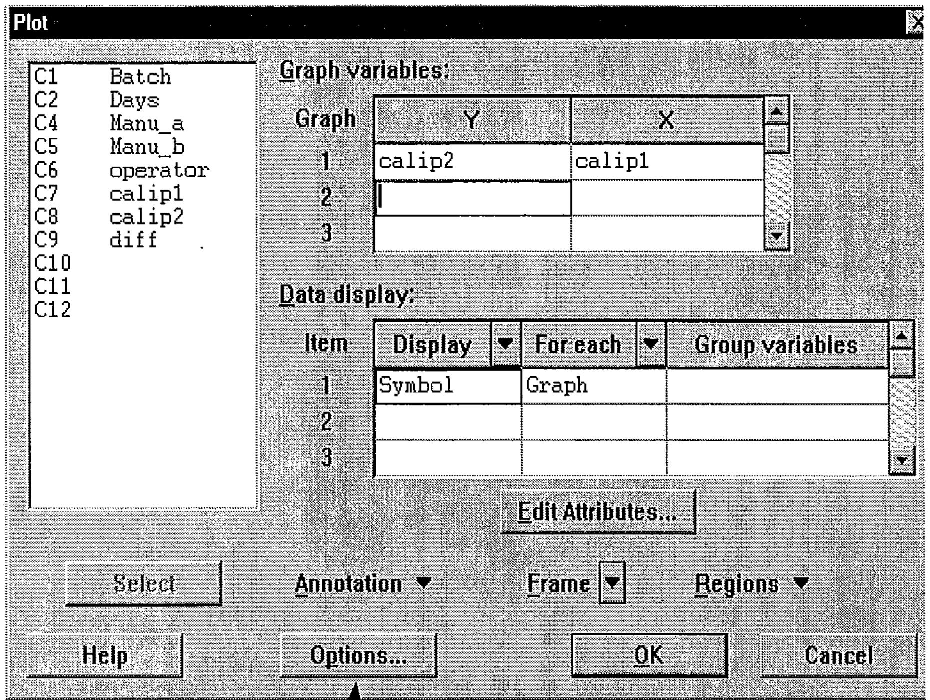
Open worksheet file "ttests" in Minitab
C:\Program Files\Mtbwin\Data\Ph1_data\ttest.mtw

- Decide which test(s) to perform
- Write the correct Hypotheses
- Perform the test(s)

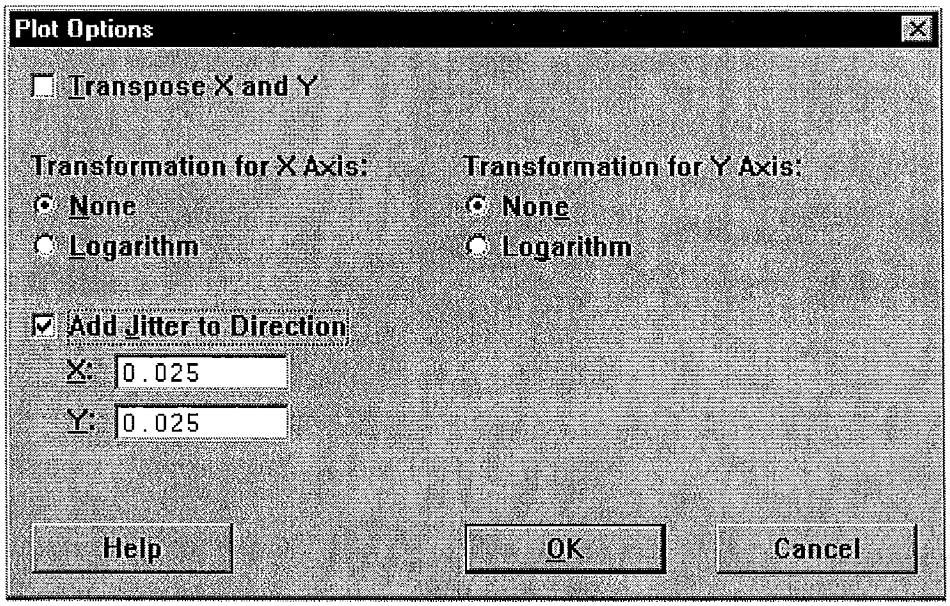
Note: Focus on columns C6 - C12. However, you may need to modify the worksheet depending on which test(s) are run. (F-test, 1 Sample t-test, 2 Sample t-test)

Graph the data first

Scatter plot of calip2 versus calip1

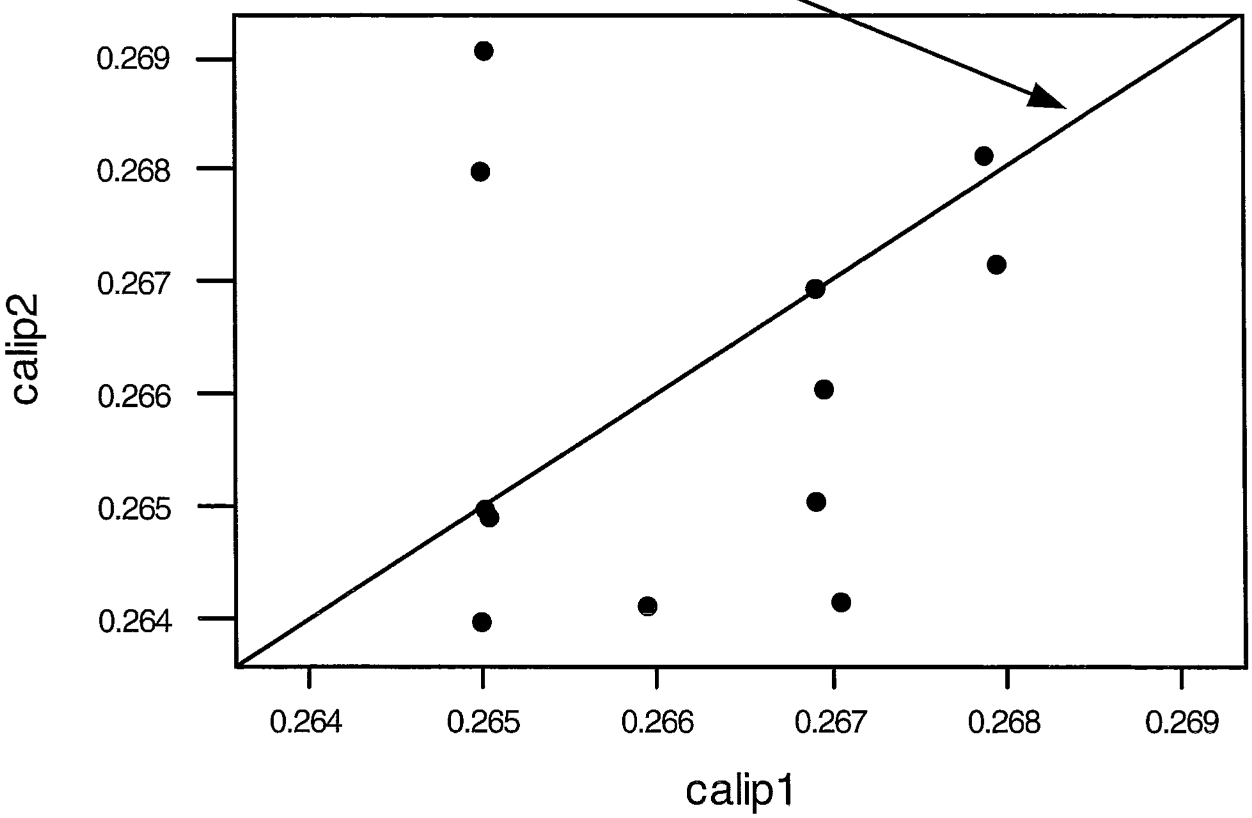


Click Options to add jitter



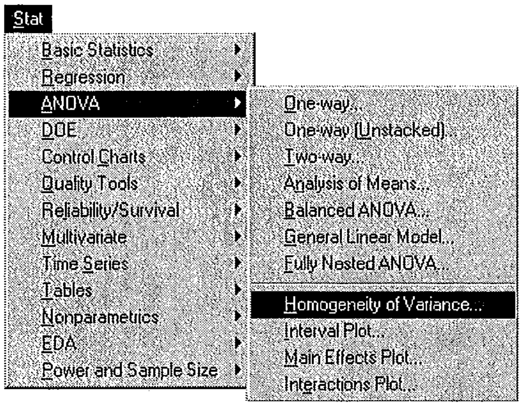
Graph the data first

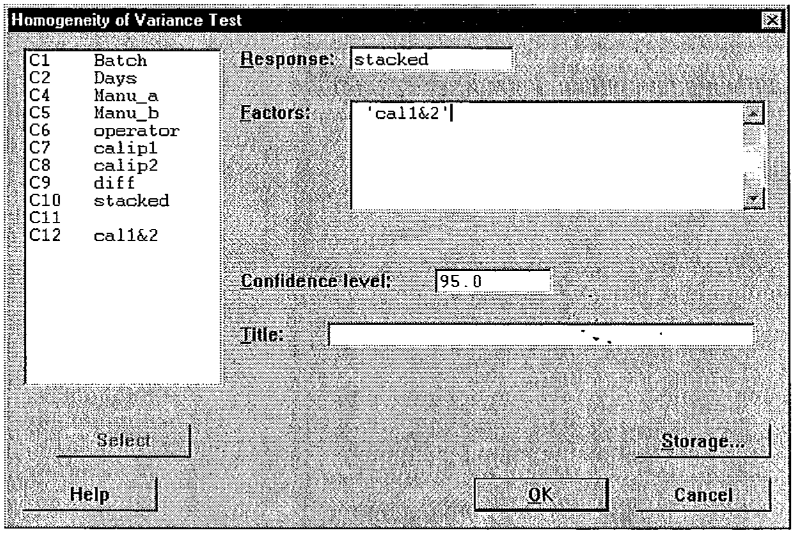
Scatter plot of calip2 versus calip1
- With jitter
- Equal axis scales
- Line of equality calip1 = calip2



Some points fall above and below the line of equality. Sometimes calip1 is higher; sometimes calip2 is higher.

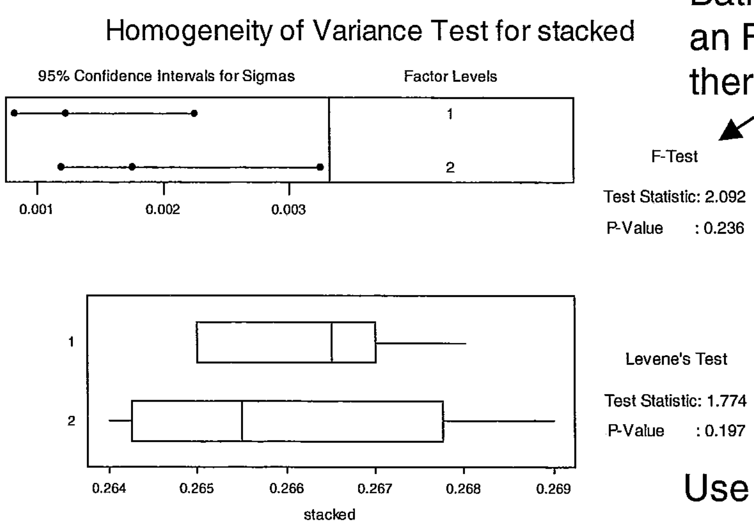
Next, Compare the Variances of the Stacked Samples





H_o: Variances of the calipers are the same

H_a: Variances of the calipers are not the same



Minitab replaces Batlett's test with an F-test when there's only 2 factors

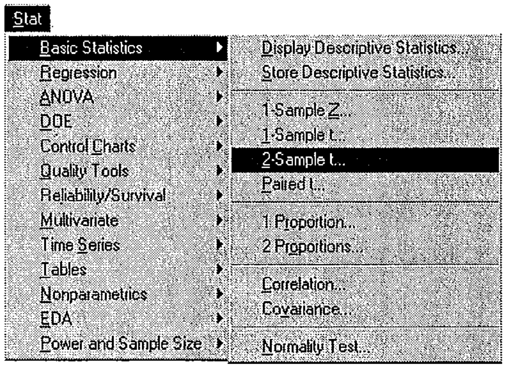
p value > .05, accept H_o

Use Levene's test for non-Normal data

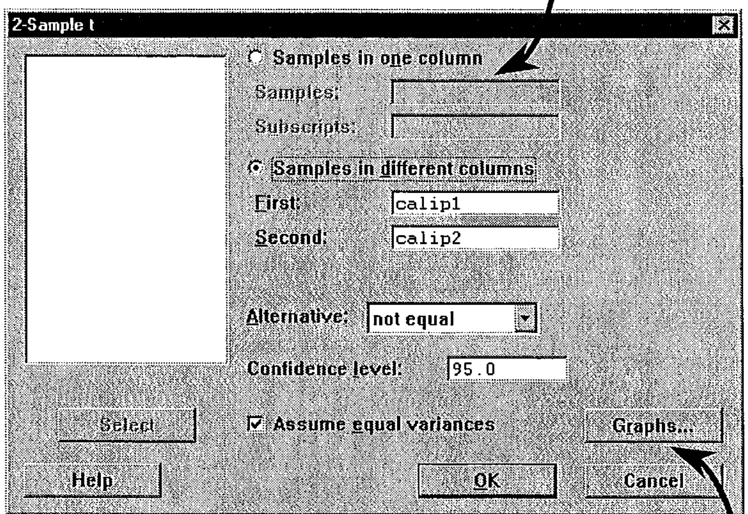
Variation Appears <u>Not</u> to be Statistically Different

Compare the Means by using 2-Sample t-Test

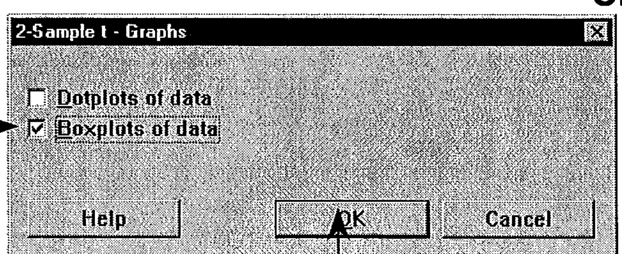
Stat>Basic Statistics>2-Sample t



Could use samples in one column by using the stacked data (C10 & C12)



Click "Graphs"



Click on "Boxplots of data"

Click "OK" twice to run

What do you want to know about Averages?

Question	Method
What is the range of plausible values for the average?	One sample confidence interval
Is the average different than the hypothesized value?	One sample hypothesis test
What is the range of plausible values for the difference between 2 averages?	Confidence interval for the difference between 2 averages
Do 2 populations have the same average?	Two sample hypothesis Test
With ***paired data***, do 2 populations have the same average?	Paired t-test, and confidence interval for the difference.

Key Questions to Ask:

- How are the data distributed?
  - Normal
  - Non-normal
- How large is the sample size?
- What parameters must be estimated?
  - Mean
  - Variance (Standard Deviation)
- Which type of comparison is required?
  - Compare variances
  - Compare single mean to a target
  - Compare two or more independent means
  - Comparison of paired group means
  - Comparisons of distributions

Class Exercise: Catapult Manipulation

Your team has baselined the Z-value of the missile-launcher. This characterization was done without tweaking the process. Based upon the results of the baselining, you know the capability gap between where you are and where you need to be (in terms of launch distance).

In the Analyze Phase, we are searching for potential Vital Few "X"s, and we can begin to manipulate the process to focus the search.

Process:

1. Decide as a team what *ONE* X-variable you would like to change on the catapult to attempt to improve its Z-value.

2. Determine the size of the sample needed to assess whether a change has occurred (see sample size table last page in Appendix of this tab).

3. Make the *ONE* change to the catapult, and launch the appropriate number of missiles

4. Record the "Y" results in Minitab. Graph the data, run Basic Statistics.

5. In Minitab, run F-Test ('Homogeneity of Variance') and then t-test to determine whether a statistically significant change has occurred.

6. Is the change practically significant?

Appendix

Steps in Testing for Differences.....

7. Establish the effect size (Delta)

8. Establish the sample size

9. Develop the sampling plan

10. Select samples

11. Collect data and conduct test

12. Calculate the test statistic (t, F or χ^2) from the data

13. Determine the probability of the calculated test statistic occurring by chance

14. If the probability is $< \alpha$, reject H_o and accept H_a
    If the probability is $\geq \alpha$, fail to reject H_o

15. Replicate results and translate statistical conclusion to practical solution.

In every Test, we are trying to prove H_a

Hypothesis Testing Terms

1. **Null Hypothesis (Ho)** - statement of no change or difference. This statement is assumed true until sufficient evidence is presented to reject it.

2. **Alternative Hypothesis (Ha)** - statement of change or difference. This statement is considered true if Ho is rejected.

3. **Type I Error** - The error in rejecting Ho when it is true, or in saying there is a difference when, in fact, there is no difference.

4. **Alpha Risk** - The maximum risk or probability of making a Type I Error. This probability is always greater than zero, and is usually established at 5%. The researcher makes the decisions to the greatest level of risk that is acceptable for a rejection of Ho.

5. **Significance Level** - Same as Alpha Risk.

6. **Type II Error** - The error in failing to reject Ho when it is false, or in saying there is no difference when, in fact, there really is a difference.

7. **Beta Risk** - The risk or probability of making a Type II Error, or overlooking an effective treatment or solution to the problem.

8. **Significant Difference** - The term used to describe the results of a statistical hypothesis test where a difference is too large to be reasonably attributed to chance.

9. **Power** - The ability of a statistical test to detect a real difference when there really is one, or the probability of being correct in rejecting Ho. Commonly used to determine if sample sizes are sufficient to detect a difference in treatments if one exists.

10. **Test Statistic** - a standardized value (z, t, F, etc.) which represents the feasibility of Ho, and is distributed in a known manner such that a probability for this observed value can be determined. Usually, the more feasible Ho is, the smaller the absolute value of the test statistic, and the greater the probability of observing this value within its distribution.

What is $F_{critical}$?

- If you want to know the critical value, you can look it up in a table. The value depends upon the alpha level, and the df for both the numerator and the denominator (use α of 0.05).

- When the calculated F value exceeds the critical $F_{\alpha=.05}$, the p-value will be less than 0.05. A high calculated F and a low p value indicate that at least one of the factor levels is different from the rest.

Nature of the t Distribution

$$Z = \frac{X - \mu}{\sigma}$$

Distribution of
Individual Measurements

df = infinity

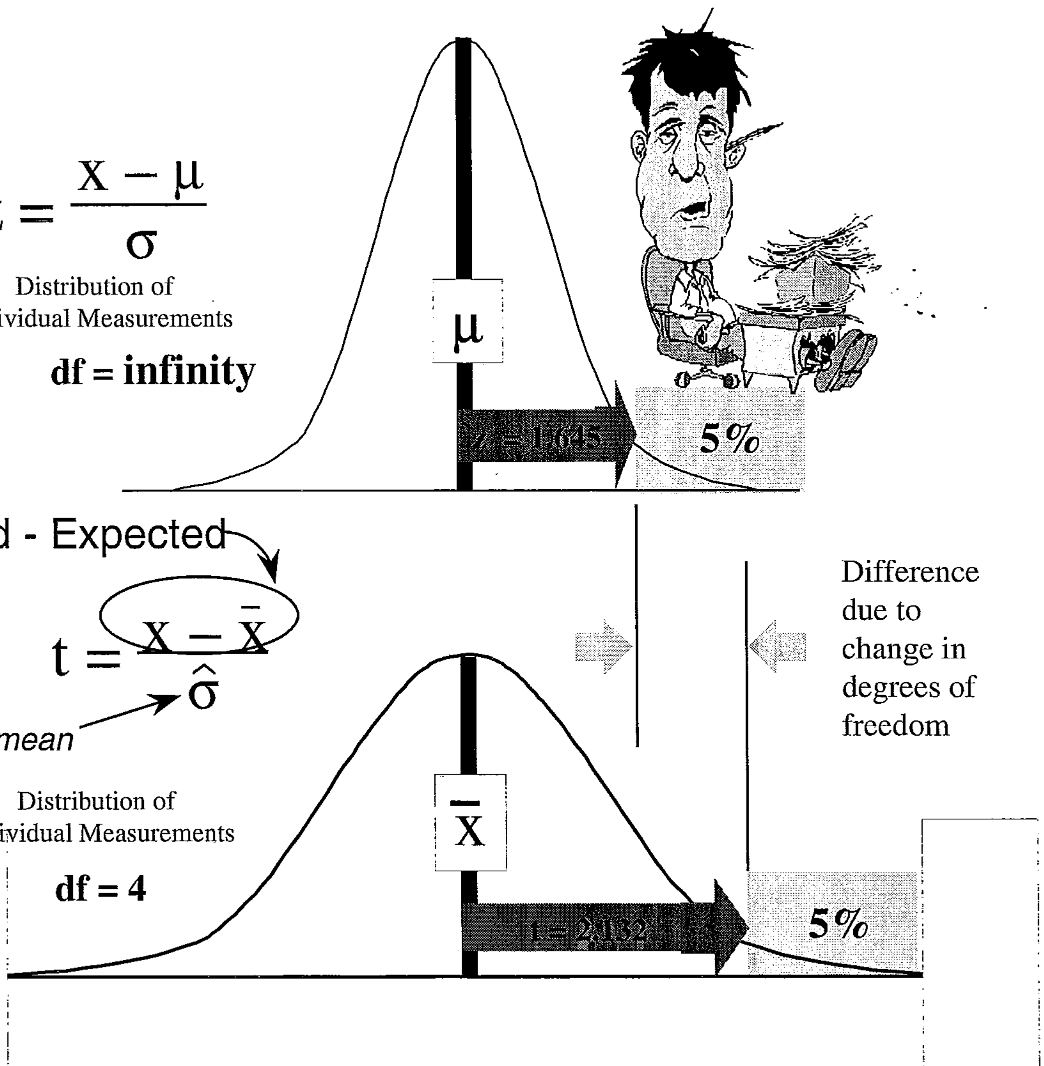
Observed - Expected

$$t = \frac{X - \bar{X}}{\hat{\sigma}}$$

Std error of the mean

Distribution of
Individual Measurements

df = 4



If the universe distribution is unknown, we may estimate it with a random sample. When the sample size is infinite, there is no uncertainty of estimation; hence, we apply the normal (z) distribution to discover a given probability of chance occurrence. However, as the sample size declines, our uncertainty increases; consequently, we must expand the range of prediction for the same probability. In other words, we must correct z for the loss in degrees of freedom.

(Reference pg. 26)

T-Distribution

Confidence Level (1 - α)

df	60.0%	70.0%	80.0%	90.0%	95.0%	97.5%	99.0%	99.5%
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
inf	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

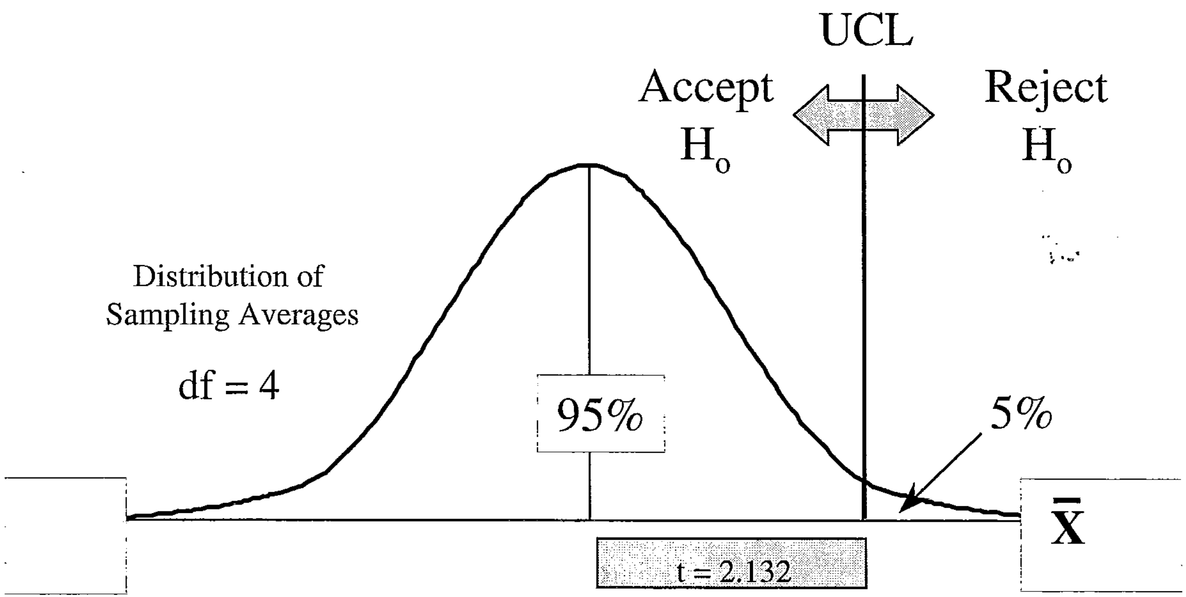
Same as Z table? Why?

t Test: Testing a Single Mean

Utilizing t Tests to test for a difference between what is observed and what is expected:

- Compares a sample to a target.
- Used for paired data.
- Test is performed on a single column which contains the difference between the pairs of data.

One-Sided Use of the t Distribution



95% Confidence Interval | Risk

$$UCL = \bar{X} + t_{\alpha} \frac{\hat{\sigma}}{\sqrt{n}}$$

$$UCL = \bar{X} + 2.132 \frac{\hat{\sigma}}{\sqrt{5}}$$

There is 95% certainty that the true universe mean will be less than the UCL. If we observe a sampling average greater than UCL, we may conclude that such an event could only occur 5 out of 100 times by random chance (sampling variations).

Two sample t, by hand. . .

Confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{(n_1+n_2-2)} * s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$\bar{x}_1$ is the average of the first sample
$\bar{x}_2$ is the average of the first sample
n_1 is the sample size
n_2 is the sample size
$t_{(n1 + n2 -2, \alpha/2)}$ is from a t-table, with $(n_1 + n_2 -2)$ degrees of freedom, and $\alpha/2$ in each tail
s_p is the pooled standard deviation

$$s_p = \sqrt{\frac{((n_1 - 1) * s_1^2) + ((n_2 - 1) * s_2^2)}{n_1 + n_2 - 2}}$$

s_p is the weighted average of the variances.
The weights are the degrees of freedom, n_i-1.

Hypothesis test

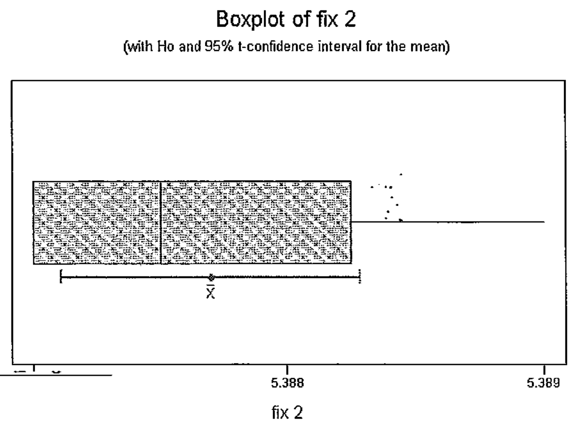
$$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

To determine statistical significance, compare t_{calc} to a tabled value of t with $(n_1 + n_2 -2, \alpha/2)$ degrees of freedom.

Class Exercise Answers: Page 46

Perform a 1-Sample t-Test to Compare Fixtures 2 & 5 to the Target Mean

1. Fixture 2 to the target mean of 5.394"



What are the Hypotheses?

H_o:Fixture 2 = target mean
H_a:Fixture 2 ≠ target mean

What is the rejection criteria?

p ≥ .05, Fail to reject H_o
p< .05, reject H_o

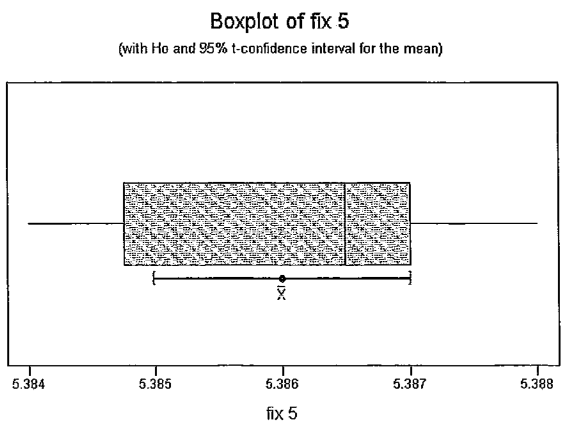
What conclusion can be drawn?

T-Test of the Mean

```
Test of mu = 5.39400 vs mu not = 5.39400

Variable      N      Mean     StDev    SE Mean        T          P
fix 2        10   5.38770   0.00082    0.00026   -24.20     0.0000
```

The mean of Fixture 2 is Statistically Different from the target mean

2. Fixture 5 to target mean of 5.394"



What are the Hypotheses?

H_o:Fixture 5 = target mean
H_a:Fixture 5 ≠ target mean

What is the rejection criteria?

p > .05, Fail to reject H_o
p ≤ .05, reject H_o

What conclusion can be drawn?

T-Test of the Mean

```
Test of mu = 5.39400 vs mu not = 5.39400

Variable      N      Mean     StDev    SE Mean        T          P
fix 5        10   5.38600   0.00141    0.00045   -17.89     0.0000
```

The mean of Fixture 5 is Statistically Different from the target mean

1. Fixture 1 to Fixture 5

What are the Hypotheses?

H_o:Fixture 1 = Fixture 5

H_a:Fixture 1 $\neq$ Fixture 5

What is the rejection criteria?

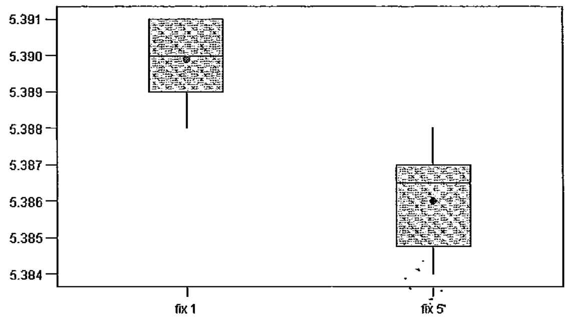
$p \geq .05$, Fail to reject H_o

$p < .05$, reject H_o

What conclusion can be drawn?

Boxplots of fix 1 and fix 5

(means are indicated by solid circles)



Two Sample T-Test and Confidence Interval

```
Two sample T for fix 1 vs fix 5
         N      Mean     StDev   SE Mean
fix 1   10   5.38990   0.00110   0.00035
fix 5   10   5.38600   0.00141   0.00045

95% CI for mu fix 1 - mu fix 5: ( 0.00271,  0.00509)
T-Test mu fix 1 = mu fix 5 (vs not =): T= 6.88  P=0.0000  DF=  18
Both use Pooled StDev = 0.00127
```

The means of the Fixtures are Statistically Different

2. Fixture 3 to Fixture 5

What are the Hypotheses?

H_o:Fixture 3 = Fixture 5

H_a:Fixture 3 $\neq$ Fixture 5

What is the rejection criteria?

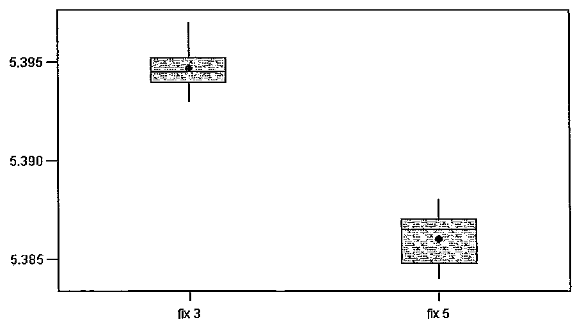
$p \geq .05$, Fail to reject H_o

$p < .05$, reject H_o

What conclusion can be drawn?

Boxplots of fix 3 and fix 5

(means are indicated by solid circles)



Two Sample T-Test and Confidence Interval

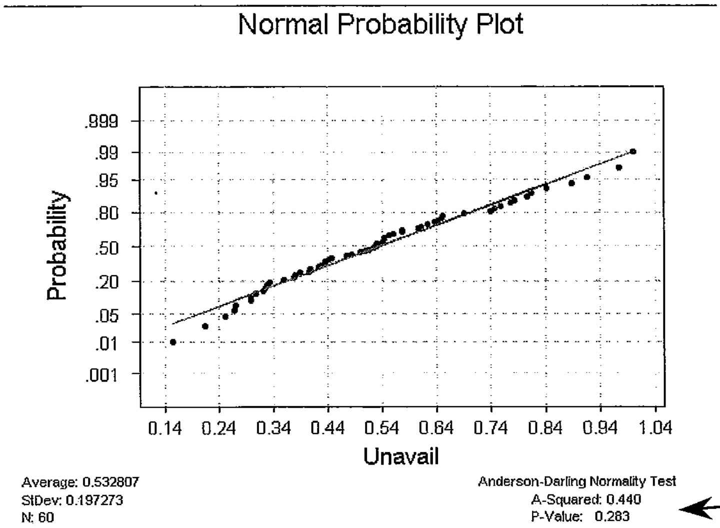
```
Two sample T for fix 3 vs fix 5
         N      Mean     StDev   SE Mean
fix 3   10   5.39470   0.00116   0.00037
fix 5   10   5.38600   0.00141   0.00045

95% CI for mu fix 3 - mu fix 5: ( 0.00748,  0.00991)
T-Test mu fix 3 = mu fix 5 (vs not =): T= 15.04  P=0.0000  DF=  18
Both use Pooled StDev = 0.00129
```

The means of the Fixtures are Statistically Different

Class Exercise Answers: Page 68

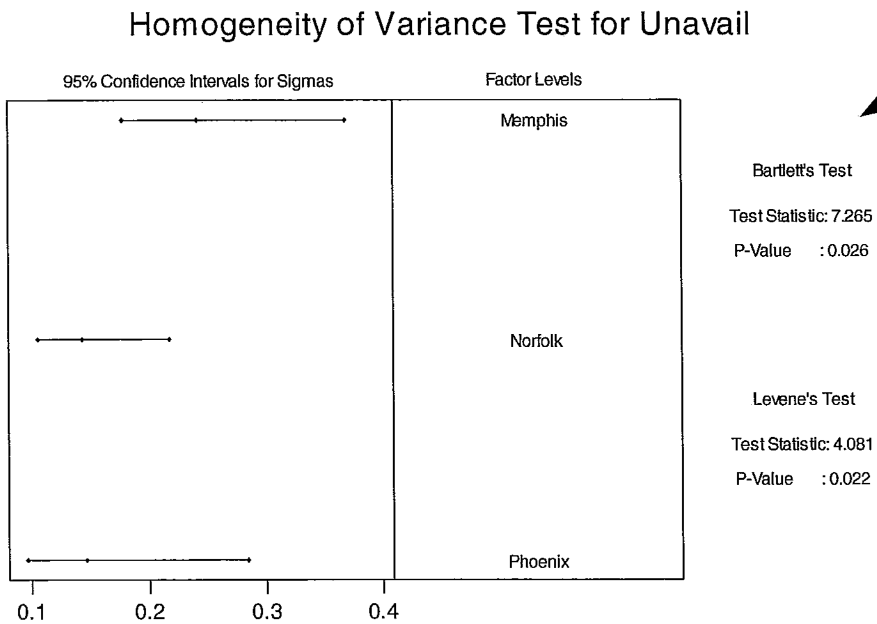
Check for Normality and Different Variances

Stat>Basic Statistics>Normality Test



Data appears to be normal

Stat>ANOVA>Homogenity of Variances



There <u>is</u> a statistical difference between the variances

2-Sample t-Test to Compare Region 1 to Region 2

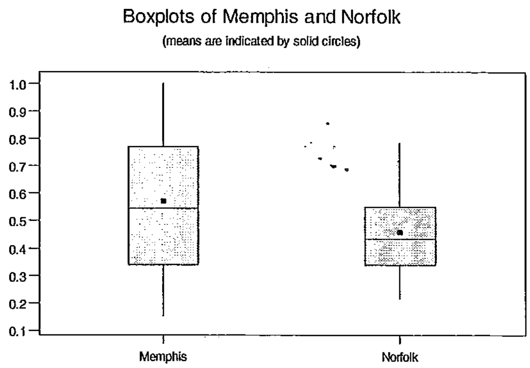
Stat>Basic Statistics>2-Sample t

Two Sample T-Test and Confidence Interval

Two sample T for Memphis vs Norfolk

	N	Mean	StDev	SE Mean
Memphis	24	0.571	0.242	0.049
Norfolk	24	0.457	0.143	0.029

95% CI for mu Memphis - mu Norfolk: (-0.002, 0.230)
T-Test mu Memphis = mu Norfolk (vs not =): T= 1.99 P=0.054 DF= 37



Fail to reject H_0, NO statistical difference can be proven

2-Sample t-Test to Compare Region 1 to Region 3

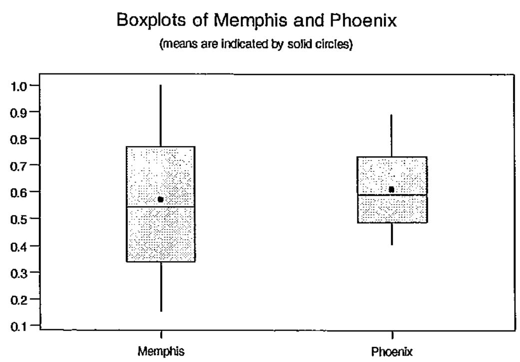
Two Sample T-Test and Confidence Interval

Two sample T for Memphis vs Phoenix

	N	Mean	StDev	SE Mean
Memphis	24	0.571	0.242	0.049
Phoenix	12	0.610	0.147	0.042

95% CI for mu Memphis - mu Phoenix: (-0.172, 0.093)
T-Test mu Memphis = mu Phoenix (vs not =): T= -0.60 P=0.55 DF= 32

Fail to reject H_0, NO statistical difference can be proven



A Basic Sample Size Table

Applies to continuous data only

α →	α = 20%				α = 10%				α = 5%				α = 1%			
δ/σ	20%	10%	5%	1%	20%	10%	5%	1%	20%	10%	5%	1%	20%	10%	5%	1% ← β
0.2	225	328	428	651	309	428	541	789	392	525	650	919	584	744	891	1202
0.3	100	146	190	289	137	190	241	350	174	234	289	408	260	331	396	534
0.4	56	82	107	163	77	107	135	197	98	131	162	230	146	186	223	300
0.5	36	53	69	104	49	69	87	126	63	84	104	147	93	119	143	192
0.6	25	36	48	72	34	48	60	88	44	58	72	102	65	83	99	134
0.7	18	27	35	53	25	35	44	64	32	43	53	75	48	61	73	98
0.8	14	21	27	41	19	27	34	49	25	33	41	57	36	46	56	75
0.9	11	16	21	32	15	21	27	39	19	26	32	45	29	37	44	59
1.0	9	13	17	26	12	17	22	32	16	21	26	37	23	30	36	48
1.1	7	11	14	22	10	14	18	26	13	17	21	30	19	25	29	40
1.2	6	9	12	18	9	12	15	22	11	15	18	26	16	21	[illegible]	[illegible]
1.3	5	8	10	15	7	10	13	19	9	12	15	22	14	18	[illegible]	[illegible]
1.4	5	7	9	13	6	9	11	16	8	11	13	19	12	15	18	25
1.5	4	6	8	12	5	8	10	14	7	9	12	16	10	13	16	21
1.6	4	5	7	10	5	7	8	12	6	8	10	14	9	12	14	19
1.7	3	5	6	9	4	6	7	11	5	7	9	13	8	10	12	17
1.8	3	4	5	8	4	5	7	10	5	6	8	11	7	9	11	15
1.9	2	4	5	7	3	5	6	9	4	6	7	10	6	8	10	13
2.0	2	3	4	7	3	4	5	8	4	5	6	9	6	7	9	12
2.1	2	3	4	6	3	4	5	7	4	5	6	8	5	7	8	11
2.2	2	3	4	5	3	4	4	7	3	4	5	8	5	6	7	10
2.3	2	2	3	5	2	3	4	6	3	4	5	7	4	6	7	9
2.4	2	2	3	5	2	3	4	5	3	4	5	6	4	5	6	8
2.5	1	2	3	4	2	3	3	5	3	3	4	6	4	5	6	8
2.6	1	2	3	4	2	3	3	5	2	3	4	5	3	4	5	7
2.7	1	2	2	4	2	2	3	4	2	3	4	5	3	4	5	7
2.8	1	2	2	3	2	2	3	4	2	3	3	5	3	4	5	6
2.9	1	2	2	3	1	2	3	4	2	2	3	4	3	4	4	6
3.0	1	1	2	3	1	2	2	4	2	2	3	4	3	3	4	5
3.1	1	1	2	3	1	2	2	3	2	2	3	4	2	3	4	5
3.2	1	1	2	3	1	2	2	3	2	2	3	4	2	3	3	5
3.3	1	1	2	2	1	2	2	3	1	2	2	3	2	3	3	4
3.4	1	1	1	2	1	1	2	3	1	2	2	3	2	3	3	4
3.5	1	1	1	2	1	1	2	3	1	2	2	3	2	2	3	4
3.6	1	1	1	2	1	1	2	2	1	2	2	3	2	2	3	4
3.7	1	1	1	2	1	1	2	2	1	2	2	3	2	2	3	4
3.8	1	1	1	2	1	1	1	2	1	1	2	3	2	2	2	3
3.9	1	1	1	2	1	1	1	2	1	1	2	2	2	2	2	3
4.0	1	1	1	2	1	1	1	2	1	1	2	2	1	2	2	3

Typical Usage