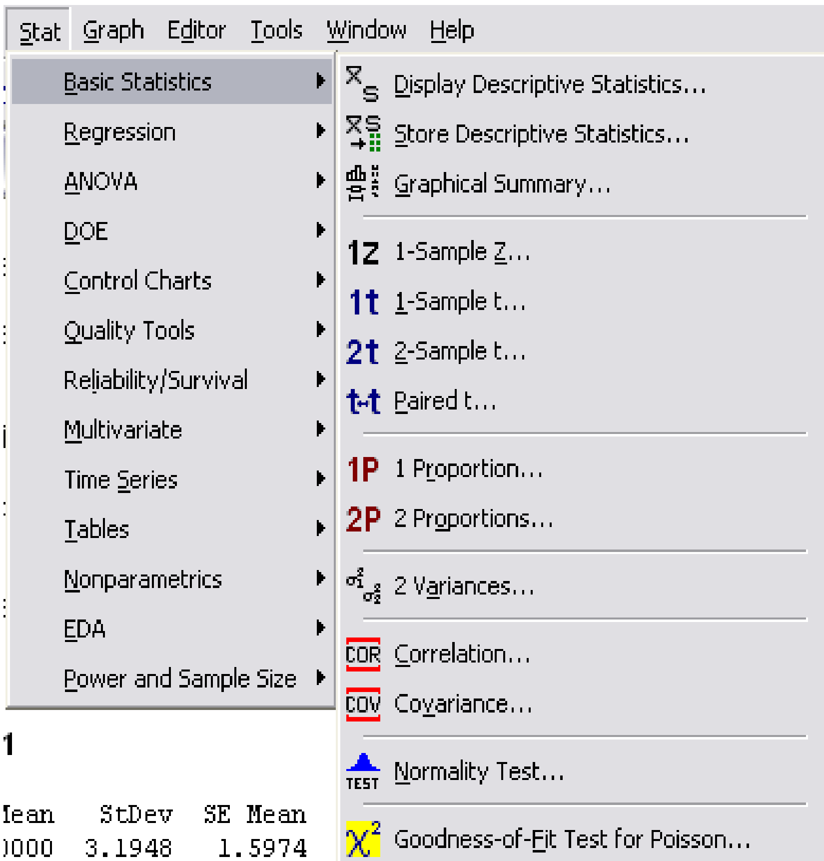
Stat 538 - Biostatistics I - Fall 2005

Lab 7

Hypothesis Testing in the two-sample situation of a population mean

Minitab can perform the calculations for a hypothesis test and Confidence intervals for two samples using the 2t option below. These are very similar to the one-sample situation that we've used in Labs 5 and 6..



The lecture notes for the two-sample inference for means are very good.
http://math.unm.edu/~schrader/biostat/bio1/Fall_2005/lec8b.pdf

The example we will use today comes from SW p. 247 exercise 7.38, available on the Labs website in the Chpt 7 directory called fertiliz.

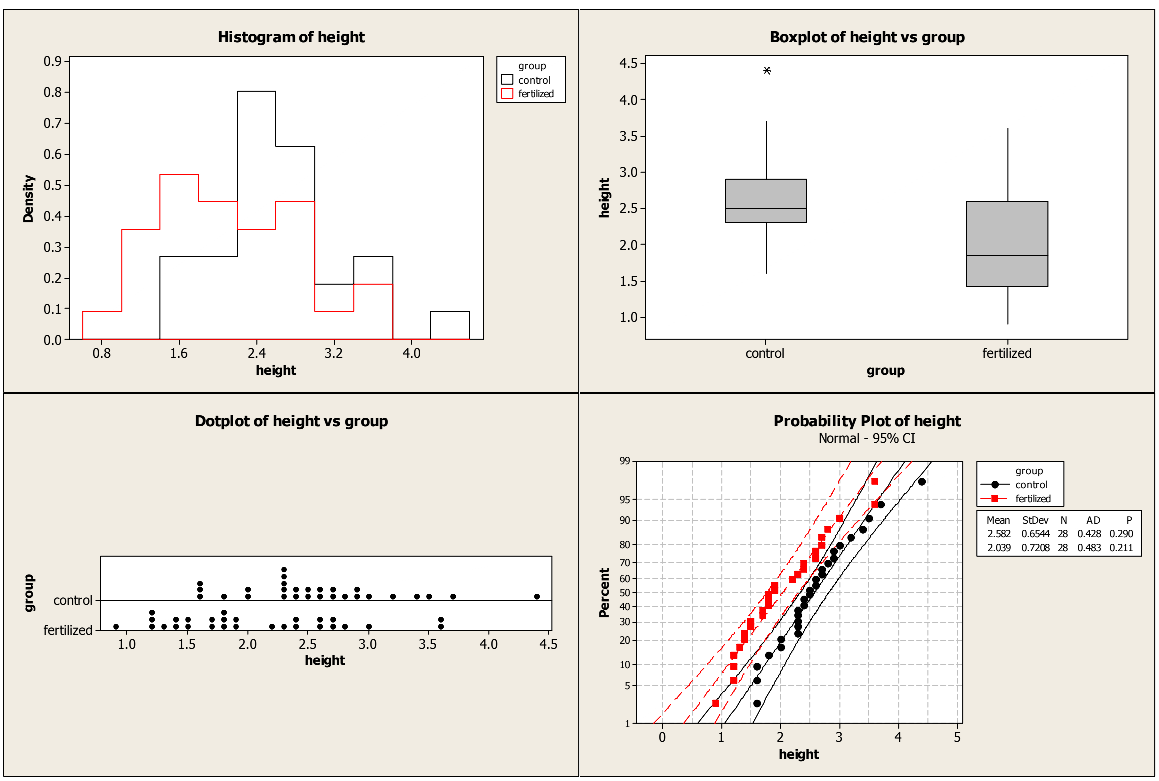
These are two independent samples, so we will perform a hypothesis test and construct a confidence interval for the population means being different. First, let's look at the data.

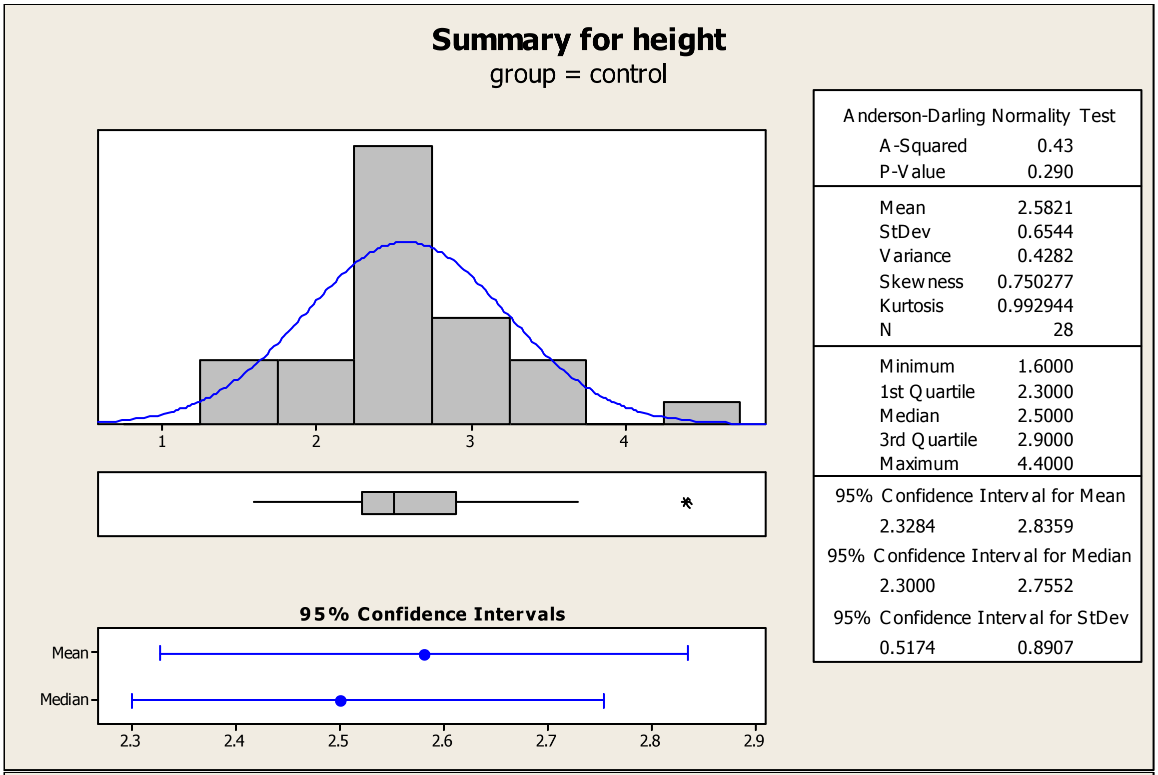
Below are summary statistics, histograms, boxplots, dotplots, normal probability plots, and full graphical summaries of the variables selecting the group as the "by variables" are below.

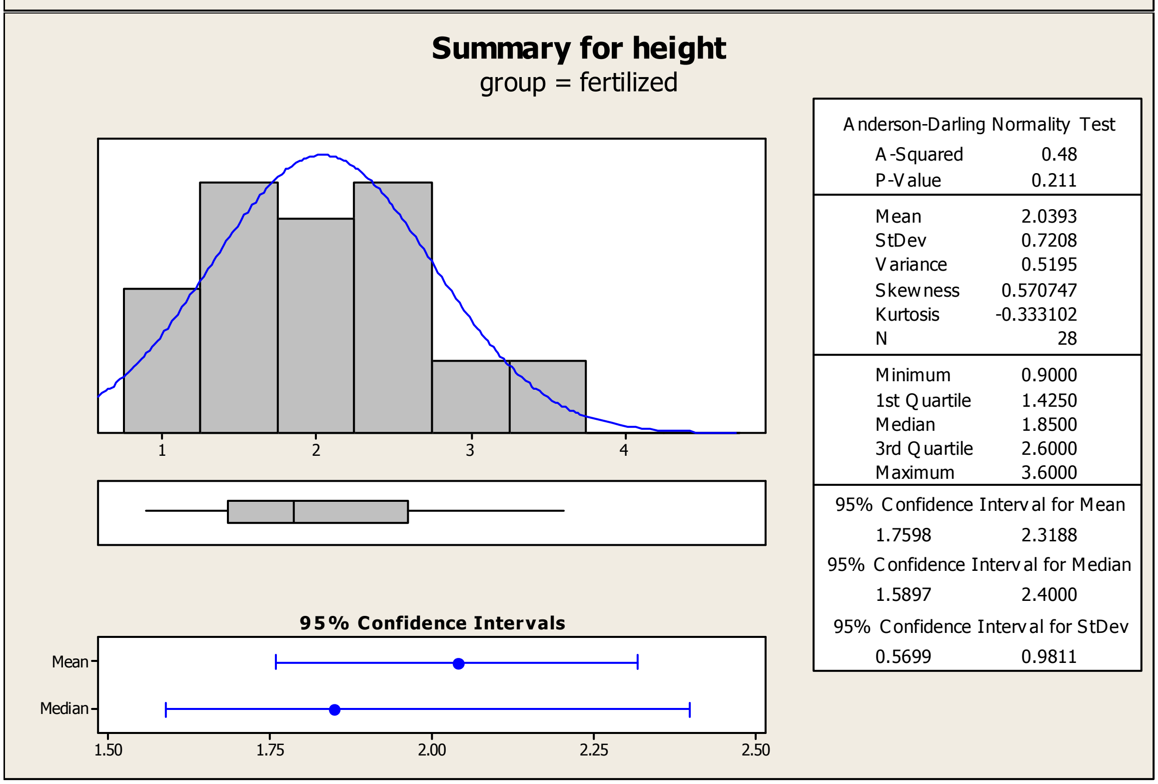
Both groups, control and fertilized, appear normal. This is supported by the shapes in the histograms and boxplots, the points following close to a line in the normal probability plots, and the Anderson-Darling normality test p-values of 0.290 and 0.211, both large. There is one potential outlier in the control group.

Descriptive Statistics: height

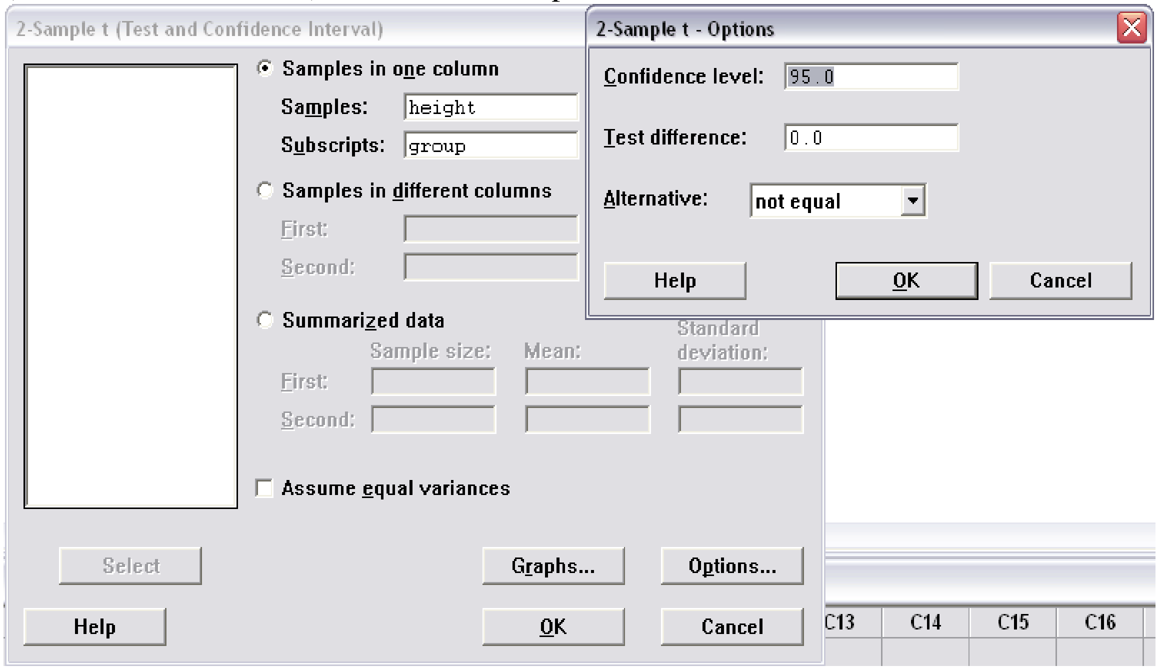
Variable	group	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
height	control	2.582	0.124	0.654	1.600	2.300	2.500	2.900	4.400
	fertilized	2.039	0.136	0.721	0.900	1.425	1.850	2.600	3.600

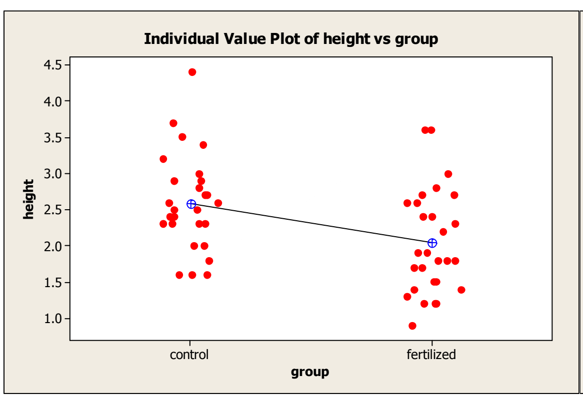


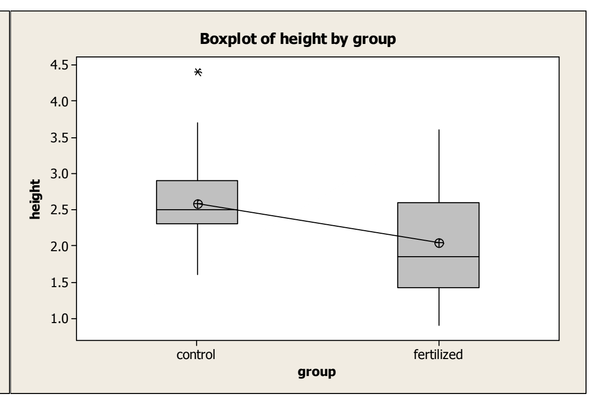

Summary for height
group = control
Anderson-Darling Normality Test
A-Squared 0.43
P-Value 0.290
Mean 2.5821
StDev 0.6544
Variance 0.4282
Skewness 0.750277
Kurtosis 0.992944
N 28
Minimum 1.6000
1st Quartile 2.3000
Median 2.5000
3rd Quartile 2.9000
Maximum 4.4000
95% Confidence Interval for Mean
2.3284 2.8359
95% Confidence Interval for Median
2.3000 2.7552
95% Confidence Interval for StDev
0.5174 0.8907
1 2 3 4
95% Confidence Intervals
Mean
Median
2.3 2.4 2.5 2.6 2.7 2.8 2.9


Summary for height
group = fertilized
Anderson-Darling Normality Test
A-Squared 0.48
P-Value 0.211
Mean 2.0393
StDev 0.7208
Variance 0.5195
Skewness 0.570747
Kurtosis -0.333102
N 28
Minimum 0.9000
1st Quartile 1.4250
Median 1.8500
3rd Quartile 2.6000
Maximum 3.6000
95% Confidence Interval for Mean
1.7598 2.3188
95% Confidence Interval for Median
1.5897 2.4000
95% Confidence Interval for StDev
0.5699 0.9811
1 2 3 4
95% Confidence Intervals
Mean
Median
1.50 1.75 2.00 2.25 2.50

The screen shot below shows us doing a 2-sample t test. Because our data are STACKED in one column, we use the first option "Samples in one column", indicating the group in the subscripts field. I have checked both plots in the graphs button, and will select 95% confidence level, testing whether the means are different.
The check-box "Assume equal variances" is unchecked for unequal variances (Satterthwaite's method), or checked for a pooled variance.







"Assume equal variances" box checked for pooled variance. We do this because the two standard deviations are roughly the same from the summary statistics above. This is the correct one to choose in this case.

Two-Sample T-Test and CI: height, group

```
Two-sample T for height

group        N   Mean  StDev  SE Mean
control     28  2.582  0.654     0.12
fertilized  28  2.039  0.721     0.14

Difference = mu (control) - mu (fertilized)
Estimate for difference:  0.542857
95% CI for difference:  (0.174012, 0.911702)
T-Test of difference = 0 (vs not =): T-Value = 2.95  P-Value = 0.005  DF = 54
Both use Pooled StDev = 0.6884
```

"Assume equal variances" box unchecked for Satterthwaite's method which does not require similar standard deviations. Because the two standard deviations ARE close, we see that the results are virtually indistinguishable from the pooled-variance test and confidence interval. This was just done for comparison.

Two-Sample T-Test and CI: height, group

```
Two-sample T for height

group        N   Mean  StDev  SE Mean
control     28  2.582  0.654     0.12
fertilized  28  2.039  0.721     0.14

Difference = mu (control) - mu (fertilized)
Estimate for difference:  0.542857
95% CI for difference:  (0.173852, 0.911862)
T-Test of difference = 0 (vs not =): T-Value = 2.95  P-Value = 0.005  DF = 53
```

The full hypothesis test goes like this:

1. [**Hypothesis**] Is there a difference in population means between the height of control and fertilized radish sprout growth?
   a. $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_A : \mu_1 - \mu_2 \neq 0$.
2. [**Significance level**] Let α=0.05.
3. [**Statistic**] Minitab reports `T-Value = 2.95`.
4. [***p*-value**] Minitab reports `P-Value = 0.005`.
5. [**Conclusion**] Because p-value = 0.005 < α = 0.05, we reject H_0 in favor of H_A, concluding that there is a difference in population means between the height of control and fertilized radish sprout growth.

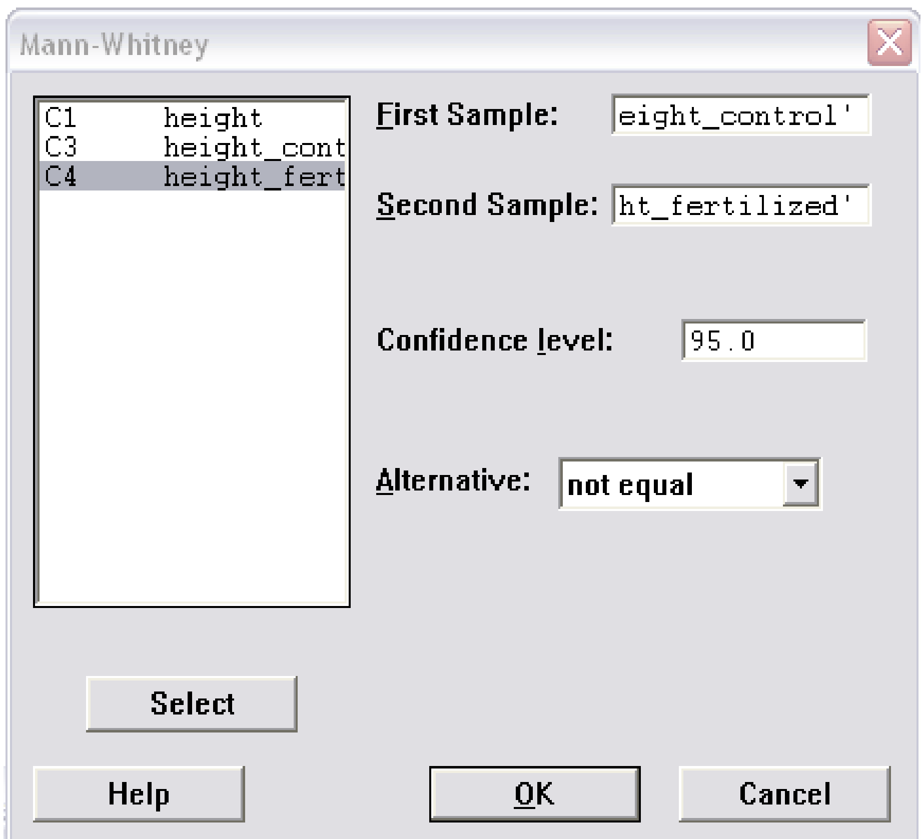
An interpretation of the confidence interval goes like this:

We are 95% confident that the true difference in population means is in the interval (0.174012, 0.911702). (This interpretation is more common.)
Or, 95% of confidence intervals constructed in this way will contain the true difference in population means.

What if the two samples suggested that our populations are NOT normal?
We can still compare the two population distributions using the **Mann-Whitney** test. The test is described well on p. 295, as well as the conditions for use. This test is a **nonparametric test**, since it does not depend on a distribution (which is specified by parameters such as mean and standard deviation). In short, it doesn't use the actual data values, but the ranks of the values of the two populations combined. If the two samples intersperce well, then the distributions are similar. If most of the values of one sample are lower in rank than those of the other sample, then the distributions are different.

In order to use the test, first we need to UNSTACK our data. Do this with Data/Unstack columns.

Next, choose Stat/Nonparametrics/Mann-Whitney. There aren't many options. We needed to unstack the data since the box takes the samples as separate columns.



The test statistic is *W* (instead of *t* or *z*, called so after Whitney). The hypothesis is conducted in the same way. In this case, when we get to step 4, we find a *p*-value = 0.0063, which is less than $\alpha = 0.05$, so we again reject H_0 in favor of H_A. I prefer the *p*-value adjusted for ties (ties are when the data includes multiples of a value), but unless there are many ties, these will be very close.

Mann-Whitney Test and CI: height_control, height_fertilized

```
                     N  Median
height_control      28  2.5000
height_fertilized   28  1.8500

Point estimate for ETA1-ETA2 is 0.6000
95.2 Percent CI for ETA1-ETA2 is (0.2001,0.9001)
W = 965.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0064
The test is significant at 0.0063 (adjusted for ties)
```