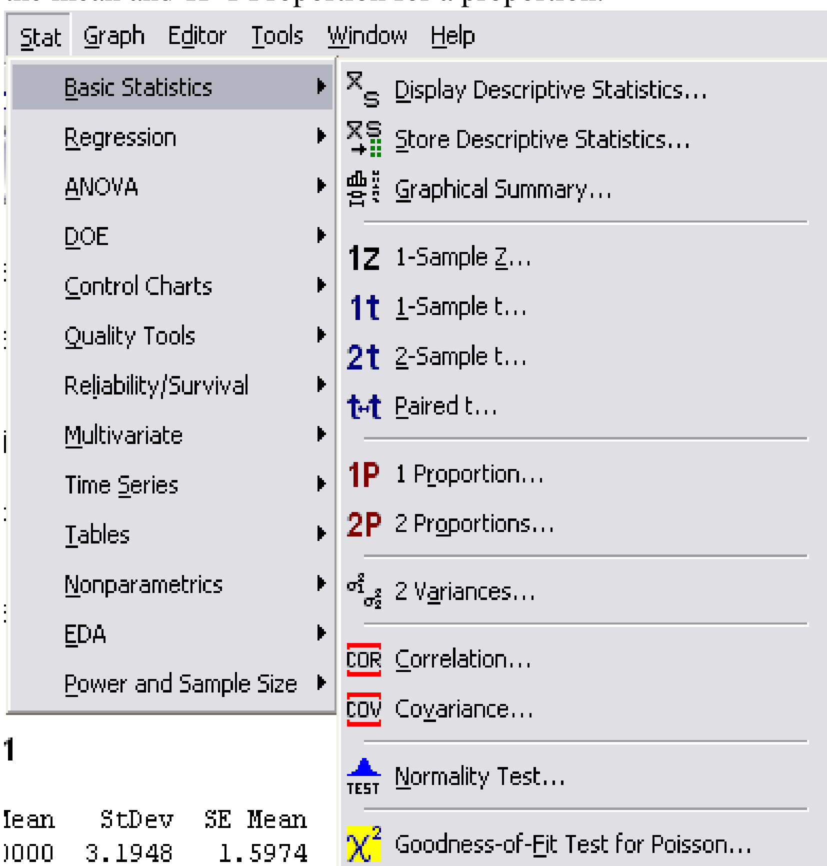
Stat 538 - Biostatistics I - Fall 2005

Lab 6

Hypothesis Testing in the one-sample situation of a population mean and population proportion

Minitab can perform the calculations for a hypothesis test in the same way we constructed Confidence intervals by choosing the Stat/Basic Statistics/1t 1-Sample t for the mean and 1P 1 Proportion for a proportion.



While Minitab will do the calculations, most of a hypothesis test is done by considering a condition of the population you want to test (mean is or is not 3, proportion is less than .5), and describing the results of the test.

First, a quick review of the five steps for conducting a hypothesis test. Refer to lecture notes #7 at http://math.unm.edu/~schrader/biostat/bio1/Fall_2005/lec7b.pdf. See page 2 for a nice picture of the rejection region.

1. [**Hypothesis**] Set up the null and alternative hypotheses: $H_0 : \mu = \mu_0$ and $H_A : \mu \neq \mu_0$, where μ_0 is specified by the context of the problem. State the hypothesis in words in terms of the alternative. Ideally, this step is done before any data is collected.
2. [**Significance level**] Choose the size or significance level of the test, denoted by α. In practice, α is set to a small value, say, .01 or .05, but theoretically can be any value between 0 and 1.
3. [**Statistic**] (similar to 4 in lecture notes) Compute the test statistic $t_s = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$, which determines the *p*-value.
4. [***p*-value**] (similar to 5 in lecture notes) Compare the *p*-value to α. If *p*-value<α, we reject H_0 in favor of H_A. Otherwise, we fail to reject H_0. (Note, if t_s is in the rejection region, α will be small and we will reject H_0.)
5. [**Conclusion**] State your conclusion in terms of the original problem.

Let's perform a hypothesis test of a mean. See Example 6.6 on page 188 of SW, *soybean growth*. I state a hypothetical hypothesis to illustrate.

1. Is the average soybean stem length different from 20cm?
   $H_0 : \mu = 20$cm and $H_A : \mu \neq 20$cm. (Note that the hypothesis is a question about the mean of the population, phrased in terms of the hypothesis, that it is different, or not equal to, 20. Here, μ_0=20cm.
2. Let's test this hypothesis at the α=0.05 level. This gives us a 5% chance of being wrong due to sampling variability even when H_0 is true, 2.5% in each tail of the distribution, exactly like the region outside a confidence interval.
3. At this point we conduct the experiment, survey, take measurements. We get the data:

```
Variable      N    Mean  SE Mean  StDev
stem-length  13  21.338    0.338  1.219
```

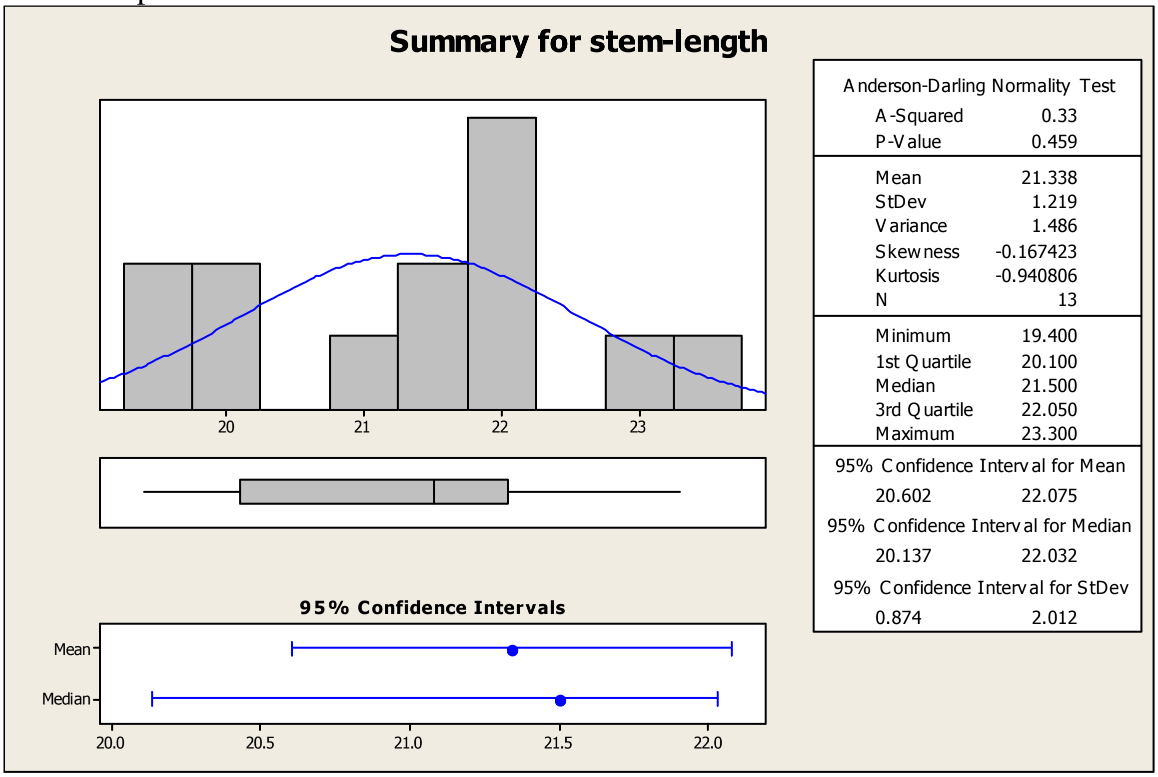
Calculate $t_s = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{21.338 - 20}{1.219/\sqrt{13}} = 3.96$.

Below, we see this corresponds to a *p*-value=0.002.

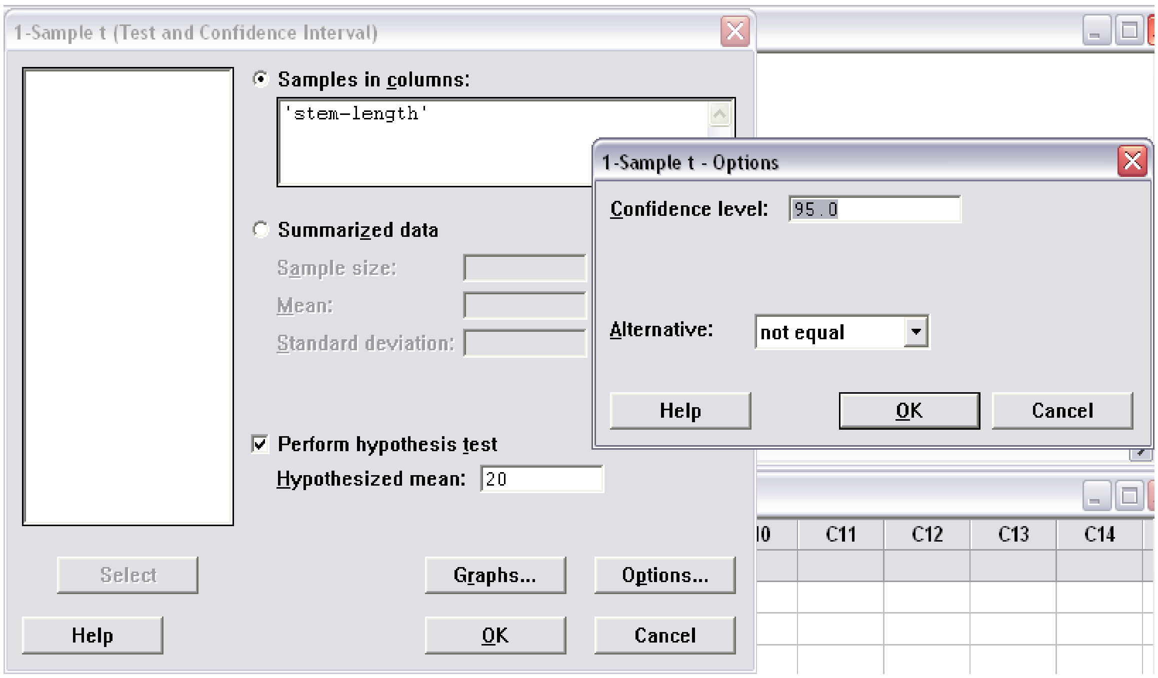
4. (and 5.) Because the *p*-value=0.002 < α=0.05 (the *p*-value is less than α) we Reject H_0 in favor of H_A, concluding that the average soybean stem length is different from 20cm. (Note that I stated the conclusion both in terms of H_0 and H_A and in terms of the language of the original problem.)

Let's see how Minitab can do this for us. Get the data from the website under chapter 6 soybean.txt.

Let's first plot the data:



Choose the menu Stat/Basic Statistics/1t 1-Sample t.
Select the column with the data in it.
Select "Perform hypothesis test" and input the mean μ_0=20cm.
Select the Options button, input your desired confidence level.
Note that this is 1-α=1-0.05=0.95=95%.
Specify that you want an alternative "not equal". (other options are "less than" and "greater than")



One-Sample T: stem-length

```
Test of mu = 20 vs not = 20

Variable      N     Mean    StDev  SE Mean        95% CI              T      P
stem-length  13  21.3385  1.2190   0.3381  (20.6018, 22.0751)  3.96  0.002
```

The results above give us the relavant sample statistics, a confidence interval, the t_s test statistic, and the associated *p*-value.

In practice, you will state your hypothesis (1), put the data in Minitab, and state your conclusion (4) based on the *p*-value.
Note that the confidence interval is the same as in the text.

What is the relationship between a confidence interval and a hypothesis test?

Our hypothesized mean 20 was outside the confidence interval and we rejected H_0. In fact, at the α=0.05 significance level, we will reject any hypothesized mean outside the 95% confidence interval (20.6018, 22.0751). At those endpoints, the *p*-value will equal α=0.05. Let's try it. Redo the hypothesis test inputing 20.6 and 22.075 as the mean.

Note the *p*-value is 0.05 because there is 0.05 probability outside the 95% confidence interval.

One-Sample T: stem-length

```
Test of mu = 20.6 vs not = 20.6


Variable      N     Mean   StDev  SE Mean         95% CI            T      P
stem-length  13  21.3385  1.2190   0.3381  (20.6018, 22.0751)  2.18  0.050
```

One-Sample T: stem-length

```
Test of mu = 22.075 vs not = 22.075

Variable      N     Mean   StDev  SE Mean         95% CI             T      P
stem-length  13  21.3385  1.2190   0.3381  (20.6018, 22.0751)  -2.18  0.050
```

A one-sided hypothesis.

See Example 2.19 on page 29 of SW, *cricket singing times*.

1. Is the average cricket singing time longer than 3 minutes?
   $H_0 : \mu = 3\text{min}$ and $H_A : \mu > 3\text{min}$.
2. Let's test this hypothesis at the α=0.01 level. This gives us a 1% chance of being wrong due to sampling variability even when H_0 is true, 1% in the right tail.
3. At this point we use Minitab. This time input 3 as the "Hypothesized mean", and under options select a 99 "confidence level", and the "Alternative" "greater than".

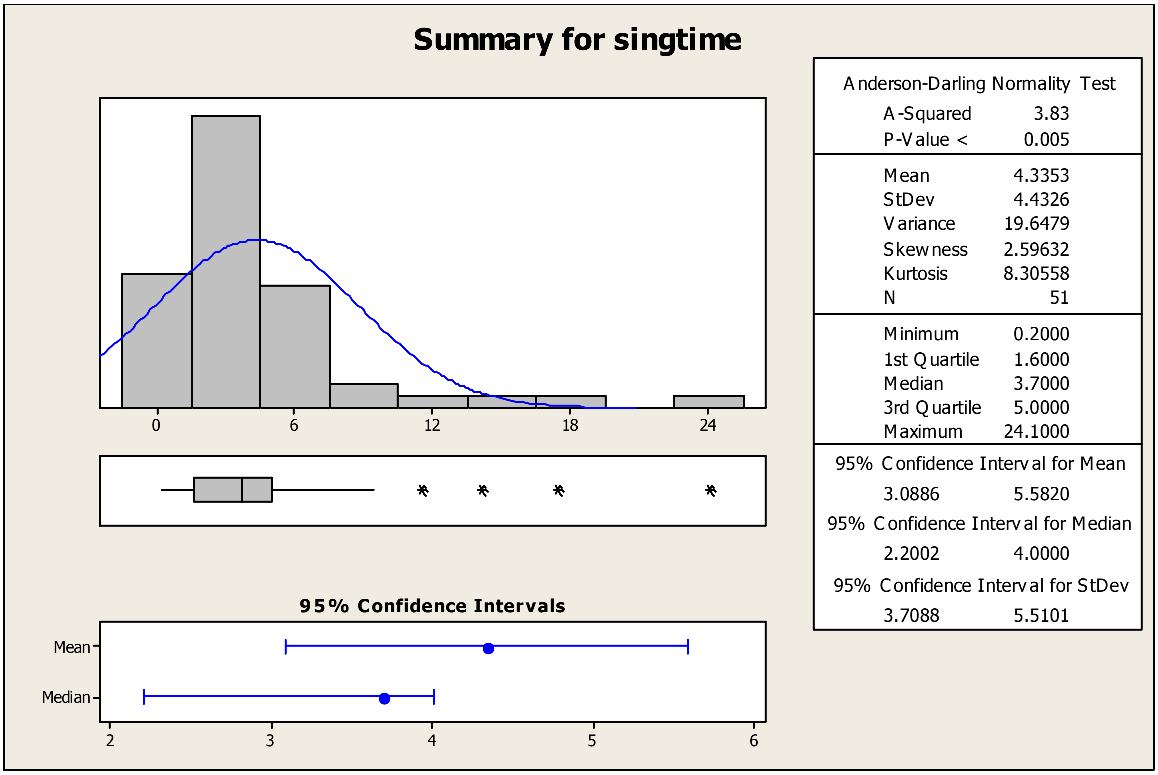
One-Sample T: singtime

```
Test of mu = 3 vs > 3
                                                 99%
                                               Lower
Variable   N     Mean    StDev  SE Mean    Bound     T      P
singtime  51  4.33529  4.43260  0.62069  2.84361  2.15  0.018
```

We have a sample mean of 4.33, which appears larger than 3. However, we have a *p*-value=0.018 > α=0.01.

4. (and 5.) Because the *p*-value=0.018 > α=0.01 (the *p*-value is not less than α) we Fail to Reject H_0, concluding that the average cricket singing time is not different from 3 minutes.

Below is a plot of the cricket singtime data.



Tests for population proportions can be done in exactly the same way.

Type I and Type II errors

We can make two types of errors when conducting a hypothesis test:

Type I error = Reject H_0 when H_0 is true
Type II error = Fail to Reject H_0 when H_0 is false

	State of Nature	
Decision	*H_0 is true*	*H_0 is false*
Fail to Reject H_0	Correct Decision	Type II Error
Reject H_0	Type I Error	Correct Decision

Typically fix probability of making a Type I error = α,
the probability of rejecting H_0 when H_0 is true.
The probability of making a Type II error = β,
the probability of Failing to Reject H_0 when H_0 is false.
The Power of the Test = 1-β,
the probability of Rejecting H_0 when H_0 is false.
The power is computed for specific values of the parameter in the alternative hypothesis.

How to view Type I and Type II errors graphically

The probability of making a Type I error = α
the probability of rejecting H_0 when H_0 is true.

Let's consider a $\alpha = 0.05$ significance test, or equivalently, a 95% confidence interval.

Then $\alpha=0.05$, the probability of rejecting H_0 when H_0 is true, the small tail probability in the direction of the alternative hypothesis.

If $H_A : \mu \neq \mu_0$, then there is $\alpha/2=0.025$ in each of the two tails (the area outside a confidence interval).
If $H_A : \mu > \mu_0$, then there is $\alpha=0.05$ in the right tail.
If $H_A : \mu < \mu_0$, then there is $\alpha=0.05$ in the left tail.

The probability of making a Type II error = β,
the probability of Failing to Reject H_0 when H_0 is false.

Construct a 100(1-α)% confidence interval (95% in our example) under the null hypothesis (what is automatically printed by Minitab for a hypothesis test).
Next, consider a specific value of the parameter in the alternative hypothesis and compute the area between the confidence bounds under the alternative distribution.

As an example, consider the soybean growth example at the beginning of the lab. This was the test results:

```
Test of mu = 20 vs not = 20
Variable      N     Mean   StDev  SE Mean         95% CI            T      P
stem-length  13  21.3385  1.2190   0.3381  (20.6018, 22.0751)  3.96  0.002
```

Say the true population mean was 22cm. In this case H_0 is false since $\mu \neq 20$cm, but instead $\mu = 22$cm. What is the probability a Type II error?
Construct a 95% confidence interval centered at $\mu = 20$cm, using the same sample size *n*=13 and our estimate of the standard deviation *s*=1.2190 (in practice you would guess this standard deviation or use previous studies or expert opinion to estimate it before collecting data). That gives (19.3374, 20.6626), calculate the area in this region under the Normal distribution with standard deviation σ=1.2190. (We use the normal distribution here because we assumed the distribution was normal, and based on the plot this assumption appears reasonable.)

Cumulative Distribution Function

```
Normal with mean = 22 and standard deviation = 1.219

      x  P( X <= x )
19.3374     0.014472
20.6626     0.136293
```

Pr(Type II error) = β = 0.136293 - 0.014472 = 0.121821.

The Power of the Test = 1-β = 1-0.121821=0.878179.
This says that if $\mu = 22$cm, we would be 87.8% likely to detect this difference from $\mu = 20$cm. The power increases as the sample size increases, and as the null value of the parameter and this value are farther apart. It also increases as the standard deviation decreases, but we typically have no control over that.