Introduction, 1 sample t-test and t-interval

87 individuals were given a flu vaccination. After 28 days, blood samples were taken to assess the concentration of antibody (X) in their serum.

Some summary statistics for the sample are as follows.

$n = 87$, $\bar{X} = 1.689$, $s = 1.549$.

Assume that the sample is from a normal population with mean μ and variance σ^2.

Confidence interval for the mean of a normal population

The form of the $100(1-\alpha)$ % confidence interval for μ is

$$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

where $t_{\alpha/2,n-1}$ is the upper $\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom.

- To find a 99% CI, $\alpha = .01$, so $\alpha/2 = .005$
- $n = 87$, so there are 86 degrees of freedom
- $t_{.005,86} \approx t_{.005,75} = 2.643$

  (using 75 degrees of freedom, which is the df nearest 86 from t table in Deveaux et al). If you are using Devore, the closest df is 60, which gives $t_{.005,60} = 2.660$.
- lower confidence limit $= 1.689 - 2.643(1.549)/\sqrt{(}87) = 1.25$.
- upper confidence limit $= 1.689 + 2.643(1.549)/\sqrt{(}87) = 2.13$.
- The 99% CI for μ is (1.25,2.13).

The **confidence coefficient** is the number $1-\alpha$ (or the percentage $100(1-\alpha)$).

The statement that "the probability that μ lies in (1.25,2.13)" is .99 is INCORRECT

The only **interpretation of the confidence coefficient** is that among the collection of all such intervals, $100(1-\alpha)$ % of them will contain the true but unknown mean μ.

Hypothesis test for the mean of a normal population

In general, we test the null hypothesis $H_0 : \mu = \mu_0$, where μ_0 is some specified value.

The test statistic used is a standardized form of $\bar{X}$,

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

This has a t-distribution with $n-1$ degrees of freedom if the population sampled is normal with mean μ_0 and unknown variance.

Let t_{obs} be the observed value of t. The p-value calculations for the test of $H_0 : \mu = \mu_0$ against the three possible alternative hypotheses H_A are given in the following table.

H_A	p-value
$\mu > \mu_0$	$P(t_{n-1} > t_{obs})$
$\mu < \mu_0$	$P(t_{n-1} < t_{obs})$
$\mu \neq \mu_0$	$2P(t_{n-1} > \lvert t_{obs} \rvert)$

where t_{n-1} denotes a t random variable with $n-1$ degrees of freedom. The probabilities are approximated using the t-table.

eg. test $H_0 : \mu_1 = 2$ against the alternative $H_A : \mu_1 < 2$. Find the p-value and report your conclusion when testing at level of significance $\alpha = .05$.

(Recall that the **level of significance** is the probability that the test will lead to a type 1 error - that the null hypothesis is incorrectly rejected).

$n = 87$, $\bar{X} = 1.689$, $s = 1.549$, so

$$t_{obs} = (1.689 - 2)/(1.549/\sqrt{87}) = -1.87$$

The p-value is the probability that a t variable with 86 degrees of freedom is less than -1.87.

- Go to the t-table. Approximate using 75 degrees of freedom, which is the df closest to 86.
- The t-distribution is symmetric about 0. Therefore the probability that $t < -1.87$ is the same as the probability that $t > 1.87$.
- From the table $P(t_{75} > 1.665) = .05$ and $P(t_{75} > 1.992) = .025$. Therefore the p-value is between .025 and .05.
- If p-value $< \alpha$, reject the null hypothesis in favour of H_A at level of significance α, otherwise do not reject H_0. Conclusion: reject H_0 at level .05.

What is the p-value if we test $H_0 : \mu_1 = 2$ against the two sided alternative $H_A : \mu_1 \neq 2$?

$t_{obs} = -1.87$ as before.

The p-value for the two sided alternative is $2P(t_{86} > |-1.87|) \approx 2P(t_{90} > 1.87)$.

We just saw that $.025 < P(t_{90} > 1.87) < .05$, so $.05 < p - value < .1$. Do not reject H_0 in favour of the two sided alternative at level .05.

Recall that in general, **the p-value is the probability of obtaining a sample presenting greater evidence against the null hypothesis than does the observed data, when in fact, the null hypothesis is true.**

Thus a small p-value is taken as evidence against the null hypothesis.

Here's a minitab computer output. The 87 observations are in column C9.

```
MTB > onet c9;
SUBC> confidence .99;
SUBC> test 2.

One-Sample T: C9

Test of mu = 2 vs not = 2

Variable   N   Mean  StDev  SE Mean       99% CI           T      P
C9        87  1.689  1.549    0.166  (1.252, 2.126)  -1.87  0.065
```

Relationship of p-values to significance tests: The null hypothesis is rejected in favour of the alternative at level α if and only if the p-value is less than α.

Another confidence interval example: Suppose we evaluate vitamin C levels (mg/100 gm) in 8 batches of corn soy blend (CSB) from a production run and get:

```
26 31 23 22 11 22 14 31
```

Find a 95% confidence interval for the mean vitamin C content of CSB produced during this run.

$$\bar{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

Summary statistics are:

$n = 8$
$\bar{x} = \frac{\sum x_i}{n} = 22.50$
$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = 51.714$, so $s = 7.19$

- $(1-\alpha) = .95$, $\alpha/2 = .025$, $\nu = n - 1 = 7$ and $t_{\alpha/2,\nu} = t_{.025,7} = 2.365$ from the table.

- $U = 22.50 + 2.365(\frac{7.19}{\sqrt{8}}) = 22.50 + 6.012 = 28.5$

- $L = 22.50 - 2.365(\frac{7.19}{\sqrt{8}}) = 22.50 - 6.012 = 16.5$

- 95% Confidence Interval for μ: $(L, U) = (16.5, 28.5)$