Chapter 18, Fourth Edition : Inference About a Population Mean

Note : Here the 4-th and 5-th editions of the text have different chapters, but the material is the same.

The previous few Chapters introduce the idea of statistical inference for a population mean. There were a number of assumptions about the experiment and population distribution. Some of these are *critical* in the sense that inference will be invalid without these or some related property. Other assumptions were not *critical* in the sense that inferences can be made even when these assumptions are violated, although the techniques that are needed and used may be different.

Recall we made some assumptions Some Very Simple Conditions for Inferences about a Population Mean

1. A simple random sample is obtained from a population. There is no non-response or other practical difficulties with the data
2. the variable we measure has exactly a normal distribution $N(\mu,\ \mathrm{sd}\ = \sigma)$
3. we do not know μ, but we know σ.

This chapter discusses two ideas.

- How can we make inferences for a population mean and remove the not very realistic assumption that σ is known.
- In a special case of *paired* experiments how can we make inferences for the *difference in population means for two treatments.*

For the first question we now consider the conditions on the population model.

1. A simple random sample is obtained from a population. There is no non-response or other practical difficulties with the data
2. the variable we measure has exactly a normal distribution $N(\mu,\ \text{sd}\ = \sigma)$
3. we do not know μ, and we also do not know σ.

In this case we change the *test statistic* from

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

to

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \tag{1}$$

where $s = \sqrt{s^2}$ and s^2 is the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \ .$$

Except for the change in the devisor, by replacing σ by the observed value of the statistic that is the sample standard deviation s (equivalently replacing σ^2 by the observed value of the statistic s^2, which is the sample variance), the test statistic (1) looks the same as before.

There is one important difference, in that the denominator now involves a random variable or statistic. The causes the sampling distribution of t in equation (1) to be different from the sampling distribution of Z, which we know is standard normal. What is the sampling distribution of t in (1)?

Earlier we obtained the standard deviation of the random variable $\bar{x}$ as

$$\frac{\sqrt{\sigma^2}}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}$$

When we replace σ^2 by the sample variance s^2, the value

$$\frac{\sqrt{s^2}}{\sqrt{n}} = \frac{s}{\sqrt{n}}$$

is no longer the standard deviation of $\bar{x}$ as it is now a random variable, actually a statistic that we can calculate from the observed data. To help us distinguish this we use a new name and call $\frac{s}{\sqrt{n}}$ the *standard error* of the $\bar{x}$. It is also called the standard error of the statistic $\bar{x}$.

The sampling distribution of t in (1), under the population model assumption above, is the so called *Student's t* distribution. This sampling distribution looks similar to the normal distribution in that it is symmetric about 0. However it is wider or more spread out than the normal. Thus critical values will be different than those obtained from the normal distribution. Other than this we will use it in the same way as before.

For example suppose we had a two sided alternative hypothesis testing problem

$$H_0 : \mu = 21 \text{ versus } H_a : \mu \neq 21$$

Our sample is $x_1, \ldots, x_n$, which is a simple random sample from a normal population with mean μ and standard deviation σ. The test statistic is now

$$t = \frac{\bar{x} - 21}{\frac{s}{\sqrt{n}}}$$

We reject H_0 in favour of H_a if

$$|t| > t^*$$

or equivalently if

$$t < -t^* \text{ or } t > t^*$$

where is t^* is the appropriate critical value. If we make this a hypothesis test at level (or significance level) $\alpha = .05$ then we need to find the critical value t^* so that

$$P(|t| > t^* | H_0) = P_0(|t| > t^*) = \alpha = .05$$

Aside Recall in the known σ case we had to find the critical value z^* so that

$$P(|Z| > z^* | H_0) = P(|Z| > z^*) = \alpha = .05$$

and then find z^* from the standard normal distribution (Table A). This is because the sampling distribution of

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

is standard normal when *mu* is the population mean. In that case of significance level $\alpha = .05$ we used Table A to find that $z^* = 1.96$.

When we now return to thinking about our test statistic (1), it has the Student's t distribution as its sampling distribution. This (sampling) distribution is tabulated in Table C at the back part of the text. Since this sampling distribution changes depending on the sample size n, there is something called the *degrees of freedom* that we need to use in order to use the correct Student's t distribution. The degrees of freedom is $n - 1$, the same as the devisor in the sample variance. Thus we sometimes use the notation t_{n-1} (as used in this text) or $t_{(n-1)}$ to denote the distribution of (1).

What do the Student's t densities (probability curves) look like? These are shown in Figures 1 and 2 for some degrees of freedom. We can see the following useful properties

- The Student t densities are symmetric about 0
- The distributions are wider or more spread out than the standard normal density. This means that the critical values will be correspondingly larger than for the normal distribution.
- When the degrees of freedom (sample size minus 1 in our case) becomes larger, the Student t distribution becomes closer to the standard normal distribution

The second item above is useful to think about further. Since σ is not know and is replaced be a random variable s, this additional variability will make confidence intervals correspondingly larger and also increase the critical values needed to reject H_0. It becomes a little harder to detect evidence against H_0. However once n becomes even as big as about $n = 20$ this effect of not knowing σ is not so important.

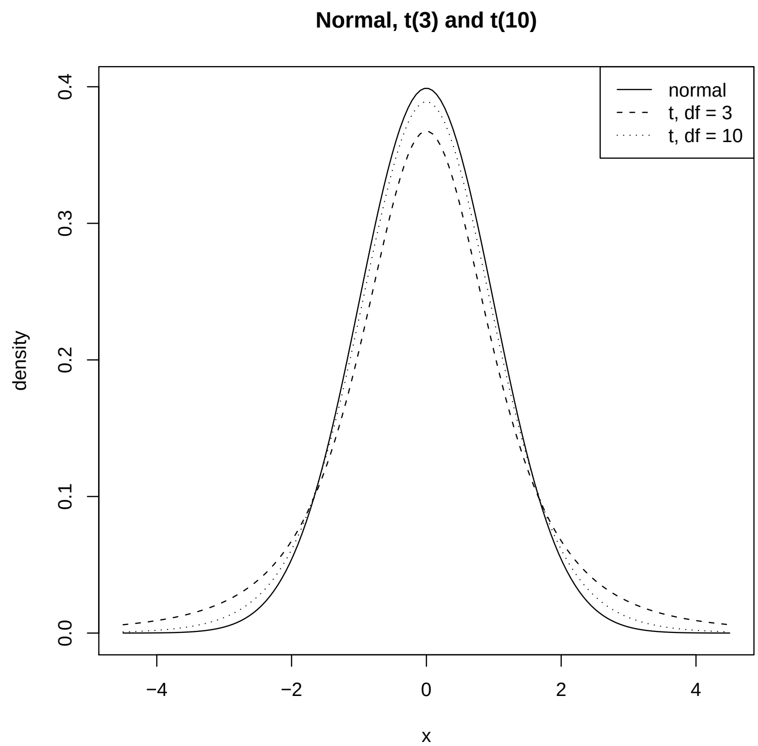

Figure 1: Normal, t density degrees of freedom 3 and 10

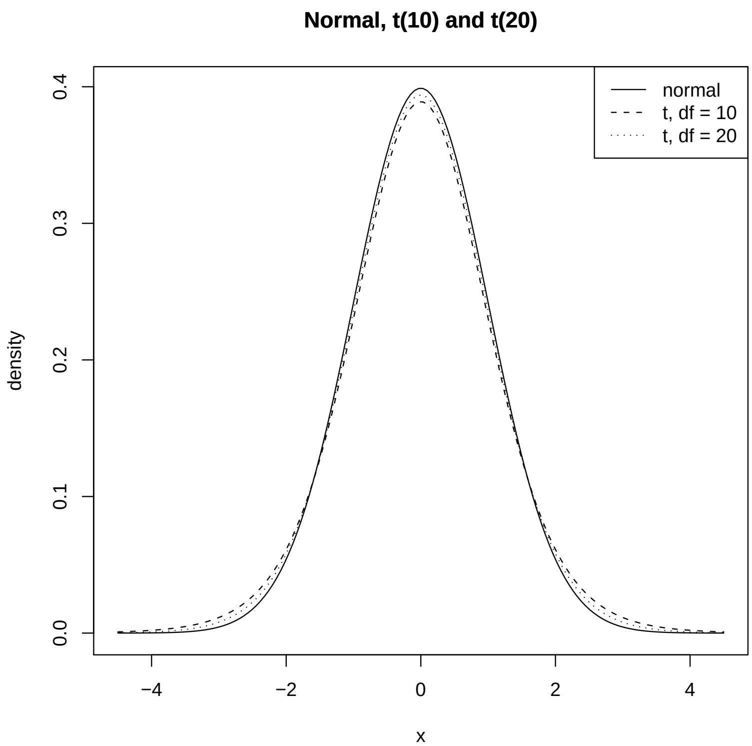

Figure 2: Normal, t density degrees of freedom 10 and 20

Reading critical values for the Student's t distribution

We use Table C from the text. Here we reproduce part of this table

	Confidence Level										
df (below)	0.5	0.6	0.7	0.8	0.9	0.95	0.996	0.98	0.99	0.998	0.999
3	0.727	0.92	1.156	1.476	2.015	2.571	5.03	3.365	4.032	5.893	6.869
10	0.700	0.879	1.093	1.372	1.812	2.228	3.716	2.764	3.169	4.144	4.587
14	0.692	0.868	1.076	1.345	1.761	2.145	3.438	2.624	2.977	3.787	4.14
20	0.687	0.86	1.064	1.325	1.725	2.086	3.251	2.528	2.845	3.552	3.85
50	0.679	0.849	1.047	1.299	1.676	2.009	3.018	2.403	2.678	3.261	3.496
z^*	0.674	0.842	1.036	1.282	1.645	1.96	2.878	2.326	2.576	3.09	3.291

Table 1: Part of Table C

If the confidence level is given as $100 * (1 - \alpha)$ then we can also determine α. For example a 95% confidence level means that the central 0.95 interval or area. The two tail areas then correspond to an area of $\frac{1-C}{2} = \frac{1-.95}{2} = .025$. This is illustrated in Figure 3.

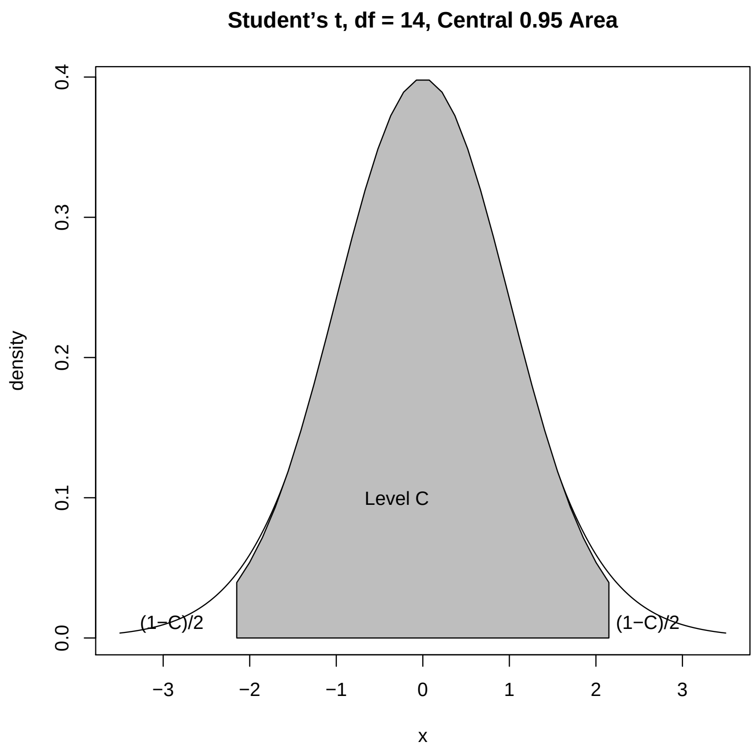

Figure 3: Student's t density and central region corresponding to confidence level C

Let us use the data from the simple random sample of size $n = 15$ from our test 1 scores. That data is

$$26, 24, 25, 24, 29, 21, 20, 15, 29, 26, 25, 27, 21, 22, 29$$

It has the summary statistics

$$\bar{x} = 24.2, s^2 = 15.171, s = \sqrt{s^2} = \sqrt{15.17} = 3.895$$

Notice that the observed value of $s = 3.895$ is not the same as the population standard deviation $\sigma = 4.04$. In fact we were to take another SRS of size $n = 15$ from this population we would obtain different values of $\bar{x}$ and s^2, as they are random variables.

We now calculate the 95% confidence interval for μ. Since $n = 15$ we need to use degrees of freedom $df = 15 - 1 = 14$. For the 95% confidence interval will will then need to used the critical value which we can obtain from Table C (or from Table 1 in these notes). We need to find the row corresponding to degrees of freedom 14, and then use critical value $t** = 2.145$.

Based on this data we then can calculate

$$\begin{aligned} \bar{x} \pm t^* \frac{s}{\sqrt{n}} &= 24.2 \pm 2.145 \frac{3.895}{\sqrt{15}} \\ &= 24.2 \pm 2.157 \end{aligned}$$

This corresponds to the interval $[22.04, 26.36]$ (rounded to 2 decimal places).

Again as before we interpret this as the values of μ which are consistent with the observed data at confidence level 95%. In a more technically specific form, these are the values of μ which are not rejected by the data at significance level 0.05.

Another useful fact about the Student's t distribution is a property called *Robustness-ness*.

The assumptions giving rise to the Student's t distribution are that the data is a simple random sample from a normal population model. It turns out that if the population distribution is not normal, but is roughly symmetric about its population mean, then the distribution of the *Studentized* sample mean centred at the population mean, that is

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

has very nearly a Student's t distribution with $n - 1$ degrees of freedom.

Robust Procedure : A statistical method is called *robust* if the probability distribution for the statistical method is approximately valid even if the conditions for the method are violated.

The Student's t statistic is an example of a robust procedure.

Another important use of the methods we have discussed here is the case of *matched pairs*. An example of this type of data is the type that we discussed with the shock absorbers example. Here we have pairs of data x_i, y_i (or for treatments 1 and 2 or treatments A and B). This type of experiment is one where we are interested in comparing the effects of two different treatments, specifically by measuring the difference in the treatment population means.

By taking the difference for each pair we obtain data $d_i = x_i - y_i$. The mean parameter of this type of difference random variable is then the parameter upon which we wish to make some statistical inference.

The data $d_1, d_2, \ldots, d_n$ is then a simple random sample from the population of possible pair differences, and one for which the *difference population mean parameter* is the one of interest. We can then analyze this as we do any other *one population* or *single population* situation.

As an example we use the body fat index data we discussed earlier in Chapter 4 and 5. Two measurement systems we discussed there, Brozek and Siri. See the scatter plots of of these data sets.

There were $n = 252$ data points. This is very large and there is no difference between the student's t distribution with 251 degrees of freedom and the normal distribution.

The first 10 data points and their differences are given in the table below

Brozek	Siri	diff.fat
12.6	12.3	0.3
6.9	6.1	0.8
24.6	25.3	-0.7
10.9	10.4	0.5
27.8	28.7	-0.9
20.6	20.9	-0.3
19.0	19.2	-0.2
12.8	12.4	0.4
5.1	4.1	1.0
12.0	11.7	0.3

For this data we have

$$\bar{x} = di\bar{f}f = -0.212, s^2 = 0.4149$$

This will yield the 95% confidence interval

$$\begin{aligned} di\bar{f}f \pm 1.96 * \frac{\sqrt{.4149}}{\sqrt{252}} \\ = -0.212 \pm 0.0796 \\ = [-0.292, -0.132] \end{aligned}$$

Thus there is evidence that these two methods do not agree over the population.

Earlier in our regression study we was there was good agreement between these two methods, but the regression line was not a line with intercept 0 and slope 1. The paired sample results here are consistent with this, in that they tell us a similar piece of information. The regression line will give information, specifically about some general cases where Brozek gives a higher value than Siri (at higher body fat levels) and where it gives a lower value (at lower body fat levels). Figure 4 shows the scatter plot of the Siri versus Brozek measurements. We see this looks nearly like a straight line. Figure 5 shows the same plot but with a line of intercept 0 and slope 1 overlaid on the scatter plot. If the Siri and Brozek measurements were the same they should fall on this straight line for each pair. This plot is another illustration that the measurements are not quite the same.

Figure 6 gives the histogram of the $n = 252$ data points. It looks fairly symmetric, so the assumptions made for the validity of the t statistic are reasonable if were to have only a small sample instead of a large sample of 252 data points. In addition the t statistic is robust, so this would also suggest the inferences are valid.

In order to illustrate a small sample data we take a random sample of size $n = 11$. This sample size is used so that we can use Table 1 given here.

The sample of size $n = 11$ obtained is

$$0.5, -0.1, 0.1, -0.1, -1.2, -0.8, -0.6, 0.5, 0.2, -0.6, -0.4$$

For this data set we have

$$\bar{x} = -0.227, s^2 = 0.297$$

Since $n = 11$, in order to calculate a 95% confidence interval we will need to use $t^* = 2.228$. This gives the confidence interval

$$\begin{aligned} &\bar{x} \pm 2.228 * \frac{\sqrt{.297}}{\sqrt{11}} \\ &= -0.227 \pm 0.366 \\ &= [-0.593, 0.138] \end{aligned}$$

Notice that we conclude that at confidence level 95% there is no difference between the two methods of measuring body fat. This is because $\mu_{\text{diff}} = 0$ is in the confidence interval and hence would not be rejected by the data. We have used the subscript diff so that we can recall from the context that we are measuring the population mean difference μ by our paired comparison method. This is different from the conclusion we obtained when we used a much bigger sample size of $n = 252$.

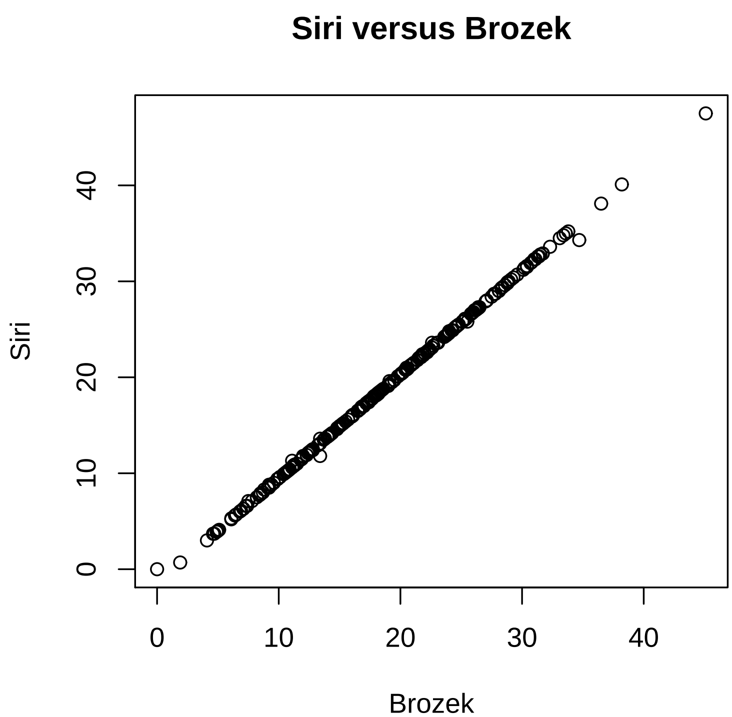

Figure 4: Scatter Plot of Siri versus Brozek Body Fat Measurements

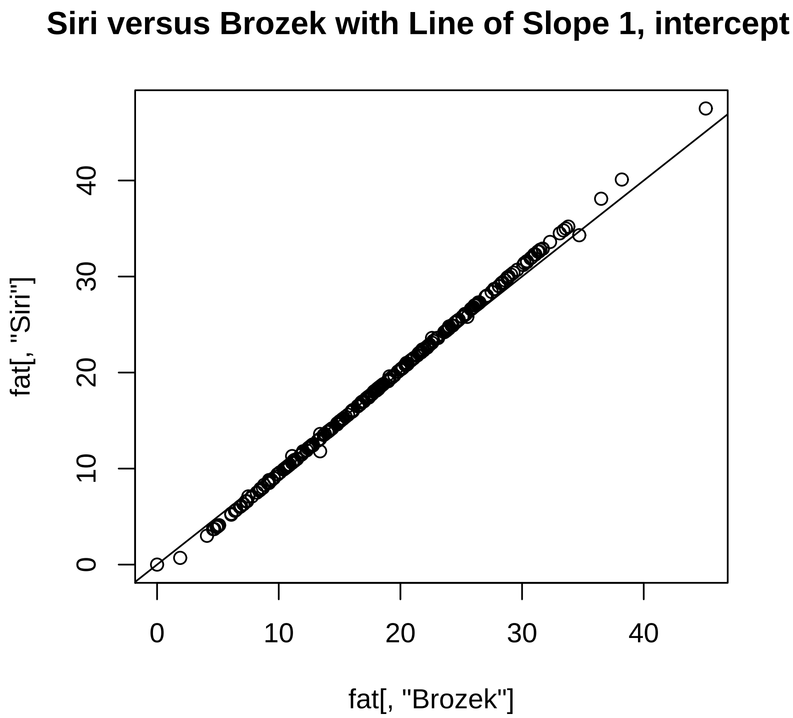

Figure 5: Scatter Plot of Siri versus Brozek Body Fat Measurements

Figure 6: Scatter Plot of Siri versus Brozek Body Fat Measurements