Chapter 17: Significance Tests, Confidence Intervals, and Prediction Intervals

- The *t* distribution
- The *t*-test for a process mean
- Confidence intervals for a process mean
- Prediction intervals for a new observation
- The two-sample *t*-statistic
- Confidence intervals on the difference between two means

The *t*-distribution

Let $y_1, y_2, \ldots, y_n$ be a random sample from a process (or from a large population relative to the sample size) with process mean μ. Let $\bar{y}$ be the sample mean and s the sample standard deviation.

The *standard error of the mean* is

$$\frac{s}{\sqrt{n}}$$

The “standardized” mean is

$$t = \frac{\bar{y} - \mu}{se_{\bar{y}}} = \frac{\bar{y} - \mu}{\frac{s}{\sqrt{n}}} = \sqrt{n}\frac{(\bar{y} - \mu)}{s}$$

We have been using the fact that this variable has a distribution that is approximately standard normal. However, even under the best of circumstances, the distribution is not exactly standard normal. The approximation is especially poor when *n* is small.

When the sample values come from a normal distribution, the exact distribution of *t* was worked out by W. S. Gossett. He called it a ***t-distribution***.

Unfortunately, there is not one *t*-distribution. There are different *t*-distributions for each different value of *n*.

If $n = 7$ there is a certain *t*-distribution but if $n = 13$ the *t*-distribution is a little different.

We say that the variable *t* has a *t*-distribution with $n – 1$ ***degrees of freedom***.

In other situations we will have statistics with *t*-distributions with degrees of freedom determined in other ways.

The *t*-Distribution Table

(page 468)

The table gives selected percentiles for a variety of degrees of freedom.

(Infinite degrees of freedom at the bottom of the table corresponds to a standard normal distribution.)

Example: We want the 95th percentile of a *t*-distribution with 7 degrees of freedom. From the table we find 1.895.

Example: We want the value t_0 so that $-t_0$ to t_0 gives the middle 95% of a *t*-distribution with 7 degrees of freedom. That is, t_0 is the 97.5th percentile of the distribution. From the table we find $t_0 = 2.365$

[Notice that this value is *larger* than 2 (or 1.96) that we have been using from the standard normal.]

The *t*-Test for a Process Mean

Null hypothesis:
$H_0: \mu = \mu_0$ (μ_0 specified)

***t*-statistic** (or ***t*-ratio**):

$$t = \frac{\bar{y} - \mu_0}{se_{\bar{y}}} = \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}} = \sqrt{n}\frac{\bar{y} - \mu_0}{s}$$

The context determines which of three alternatives is appropriate—two-sided or one-sided—and which side.

Consider first the two-sided alternative hypothesis: $H_1: \mu \neq \mu_0$

Critical Region (**Decision Rule**):
Reject H_0 if, and only if, $|t|$ is “large enough.”

How large is large enough? Setup a required significance level or use *p*-values.

Example: “Performance of Complex Tasks Under Different Levels of Illumination”, *J. Illumination Engineering*, 1976, p.235-242

Nine subjects were required to insert a fine-tipped probe into eyeholes of 10 needles in rapid succession both for a low light level with a black background and a higher level with a white background. The times to complete the tasks under both conditions are given below for the 9 subjects.

	Lighting	
Subject	**Lower**	**Higher**
1	18.23	25.85
2	20.84	28.84
3	22.96	32.05
4	19.68	25.74
5	19.50	20.89
6	24.98	41.05
7	16.61	25.01
8	16.07	24.96
9	24.59	27.47

Are there differences in task times for the two lighting conditions?

In particular, are there differences in ***mean*** task times for the two lighting conditions?

We consider the differences from higher to lower light and ask if the mean of the differences (in the population or process) is zero or not, that is, $\mu_0 = 0$.

Let $H_0: \mu = 0$ and $H_1: \mu \neq 0$

We calculate the differences and then get the mean and standard deviation of the differences.

	Lighting		H–L
Subject	**Lower**	**Higher**	**Difference**
1	18.23	25.85	7.62
2	20.84	28.84	8.00
3	22.96	32.05	9.09
4	19.68	25.74	6.06
5	19.50	20.89	1.39
6	24.98	41.05	16.07
7	16.61	25.01	8.40
8	16.07	24.96	8.89
9	24.59	27.47	2.88

Here are the results:

mean of differences = 7.60

standard deviation of differences = 4.18
standard error of mean =

$$\frac{s}{\sqrt{n}} = \frac{4.18}{\sqrt{9}} = \frac{4.18}{3} = 1.3933333 \approx 1.39$$

The *t*-statistic is

$$t = \frac{7.60 - 0}{1.39} = 5.46$$

with 9 – 1 = 8 degrees of freedom.

Suppose we choose a significance level of 5%. Then we need to look up the 97.50th percentile from the *t*-table with 8 degrees of freedom. It is 2.306. (Notice that it is larger than 2.)
Since our observed t-statistic is larger than 2.306 we reject the null hypothesis at the 5% significance level.

We say that the two lighting conditions produce ***statistically significant*** different mean task completion times.

p-value

The *p*-value is the area outside of 5.46 in a *t*-distribution with 8 degrees of freedom.

From the *t*-table we find that the 99.9th percentile is 4.501. (This is the highest percentile given in the table.)

So the two-sided *p*-value is less than 2(1 – 0.999) = 0.002, quite a small value.

We again would say that the two lighting conditions produce statistically significant different task completion times.

From Minitab we can get a more accurate *p*-value.

Here is the printout in the session window:

```
Cumulative Distribution Function
Student's t distribution with 8 d.f.
         x     P( X <= x)
    5.4600        0.9997
```

so that the *p*-value is 2(1 – 0.9997) = 0.0006

Steps in Hypothesis Testing (page 478)

- Set up a model and/or a sampling plan
- Decide on an appropriate null hypothesis
- Clarify the alternative hypothesis
- Decide on an appropriate test statistic
- What is the sampling distribution of the test statistic when the null hypothesis is true?
- Choose a significance level (or use *p*-values)
- Set up the critical region (look at the alternative hypothesis to do so)
- Carry out the test with the appropriate distribution
- Report the results in terms of the original problem

Confidence Intervals for μ

(page 480)

A confidence interval can be related to an hypothesis test, namely:

The confidence interval for μ is the set of values for μ that cannot be rejected by the data.

If the significance level is α, then the confidence level is 1 – α.

We “undo” the inequality

$$-t_{1-\alpha/2} \le \frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}} \le t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the (1–α/2)100 percentile of the *t*-distribution with *n*–1 degrees of freedom.

We obtain

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

as a confidence interval for μ with confidence level (1–α/2)100%.

Lighting example:

The mean of the differences is 7.60 and the standard error of the mean is 1.39 with 8 degrees of freedom.

Let’s do a 99% confidence interval on the process mean difference.

We need $t_{1-\alpha/2}$ where 1–α = 0.99.

So 1–α/2 = 0.995 is the column to look in. Go down to 8 degrees of freedom and obtain $t_{1-\alpha/2}$ = 3.355.

This gives our 99% confidence interval as

$$\bar{y} \pm t_{1-\alpha/2}\frac{s}{\sqrt{n}} = 7.60 \pm 3.355(1.39)$$

$$= 7.60 \pm 4.66$$

We are 99% confident that the true process mean is in the interval: 2.94 to 12.26

Prediction Intervals

(page 483)

Let y_1, y_2,..., y_n be a random sample from a process (or from a very large population) with process mean μ. Let $\bar{y}$ be the sample mean and s the sample standard deviation.

We want to predict the value of an independent unobserved data value, y_{n+1}.

The ***prediction error*** is given by $\bar{y} - y_{n+1}$.

The mean of the prediction error is $\mu - \mu = 0$.

The Pythagorean Result

(page 444)

The standard error of the difference (or sum) of two independent estimates is the square root of the sum of the squares of the two standard errors.

In symbols, if se_1 and se_2 are the standard errors of the two estimates then the standard error of the difference (or sum) of the two estimates is:

$$se_{difference} = \sqrt{(se_1)^2 + (se_2)^2}$$

Using the pythagorean result the standard error of the prediction error is

$$se_{\bar{y} - y_{n+1}} = \sqrt{(se_{\bar{y}})^2 + (se_{y_{n+1}})^2}$$

but

$$se_{\bar{y}} = \frac{s}{\sqrt{n}} \qquad and \qquad se_{y_{n+1}} = s$$

so

$$se_{\bar{y} - y_{n+1}} = \sqrt{\left(\frac{s}{\sqrt{n}}\right)^2 + (s)^2} = s\sqrt{1 + \frac{1}{n}}$$

We start with

$$-t_{1-\alpha/2} \le \frac{\bar{y} - y_{n+1}}{s\sqrt{1 + \frac{1}{n}}} \le t_{1-\alpha/2}$$

and turn it into the ***prediction interval*** for y_{n+1}:

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

Example:

Return to the Chrysler daily stock prices. We found that the percentage changes were quite random and mound-shaped.

The mean of the 166 percentage changes was 0.082 and the standard deviation was 2.066.

Let's find a 95% prediction interval for the next percentage change.

$$\begin{aligned} &\bar{y} \pm t_{1-\alpha/2}\, s\sqrt{1 + \frac{1}{n}} \\ &= 0.082 \pm 1.96(2.066)\sqrt{1 + \frac{1}{166}} \\ &= 0.082 \pm 1.96(2.066)1.0030075 \\ &= 0.082 \pm 4.062 \end{aligned}$$

or –3.980 to 4.144 as our 95% prediction interval for the next percentage change.

Two-Sample *t*-Statistic for the Difference Between Two Means (page 485)

We have two independently selected samples from normal processes. Each sample has a mean, standard deviation, and sample size: $\bar{y}_1$ s_1, n_1 and $\bar{y}_2$, s_2, n_2.

We know from the Pythagorean result that

$$\mu_{\bar{y}_1 - \bar{y}_2} = \mu_1 - \mu_2 \qquad se_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

So we form the **two-sample *t*-statistic**

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

with (approximate) degrees of freedom ν given from

$$a = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\frac{1}{\nu} = \frac{a^2}{n_1 - 1} + \frac{(1-a)^2}{n_2 - 1}$$

Another way to obtain the (approximate)

degrees of freedom: (***not*** in the book)

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{(n_1 - 1)}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{(n_2 - 1)}\left(\frac{s_2^2}{n_2}\right)^2}$$

If $s_1 = s_2$ and $n_1 = n_2$, then the degrees of freedom reduce to $\nu = n_1 + n_2 - 2$ which is sum of the degrees of freedom for each of the samples separately.

It can be shown that the degrees of freedom is always between the smaller of $n_1 - 1$ and $n_2 - 1$ and $n_1 + n_2 - 2$. That is,

$$min(n_1 - 1, n_2 - 1) \le \nu \le n_1 + n_2 - 2$$

The most common null hypothesis is that the two processes have the same mean: $\mu_1 = \mu_2$ or H_0: $\mu_1 - \mu_2 = 0$

(This is where the term ***null*** hypothesis comes from.)
The context determines which of three alternatives is appropriate—two-sided or one-sided—and which side.

Example:

A director of training is considering changing from individual training of assembly personnel to a training method that employs team-based learning. Before implementing such a change it is important to assess the evidence that team-based learning would in fact improve productivity in the assembly operation. An experiment is designed to collect data on this issue.

A group of 29 employees has recently been hired. The training director randomly assigns 15 of them to the traditional training method and the rest (14) to the new team-based training. When the training period is completed, she evaluates the employees on the time it takes them to assemble a product.

Here are the data and some summary statistics:

Assembly Times in minutes

Individual Training	Team Training
9.2	8.5
9.6	8.4
9.1	8.8
9.1	8.6
9.0	8.8
9.1	9.7
9.3	8.7
9.7	9.0
9.2	8.2
8.4	9.5
9.5	8.8
8.9	8.9
8.7	8.5
9.3	8.7
8.8	

```
          N      Mean      StDev    SE Mean
Individ  15     9.127      0.343      0.089
Team     14     8.793      0.403      0.110
```

The two-sample *t*-statistic for testing $H_0\colon \mu_I - \mu_T = 0$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{9.127 - 8.793}{\sqrt{\frac{0.343^2}{15} + \frac{0.403^2}{14}}}$$

$$= \frac{0.334}{\sqrt{0.00784 + 0.01160}} = \frac{0.334}{0.1394274}$$

$$= 2.39$$

Now for degrees of freedom and

$$\frac{1}{\nu} = \frac{a^2}{n_1 - 1} + \frac{(1-a)^2}{n_2 - 1} = \frac{0.4^2}{14} + \frac{(1-0.4)^2}{13}$$

$$= 0.03912$$

$$a = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \frac{\frac{0.343^2}{15}}{\frac{0.343^2}{15} + \frac{0.403^2}{14}} = 0.40$$

So that

$$\nu = \frac{1}{0.03912} = 25.562372 \approx 25 \text{ rounded down}$$

Since we suspect that the team training yields ***reduced*** assembly times we use a one-sided alternative hypothesis, namely,

$H_1\colon \mu_I - \mu_T > 0$

For this one-sided alternative, we reject the null hypothesis if, and only if, our observed *t*-statistic is “too large”, that is, $t > t_{1-\alpha}$.

Choosing a significance level of 5% and using 25 degrees of freedom gives $t_{1-\alpha} = 1.708$.

Our observed *t*-statistic is 2.39 so we have statistically significant evidence to reject the null hypothesis of no difference in favor of the alternative hypothesis that says that team training produces reduced assembly times.

The one-sided *p*-value is between 0.025 and 0.01 according to the *t*-table.

According to Minitab the *p*-value is 1 – 0.9876 = 0.0124.

Confidence Interval for $\mu_1 - \mu_2$

$$\bar{y}_1 - \bar{y}_2 \pm t_{1-\alpha/2} se_{\bar{y}_1 - \bar{y}_2}$$

where, as before,

$$se_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Degrees of freedom are obtained as before.

Example: Individual versus Team Training

Let’s calculate a 95% confidence interval on the mean difference in assembly times.

With 25 degrees of freedom we find $t_{1-\alpha/2} = t_{1-0.05/2} = t_{0.975} = 2.060$.

$$\begin{aligned} &\bar{y}_1 - \bar{y}_2 \pm t_{1-\alpha/2} se_{\bar{y}_1 - \bar{y}_2} \\ &= (9.127 - 8.793) \pm 2.060(0.1394) \\ &= 0.334 \pm 0.287 \end{aligned}$$

or 0.047 to 0.621 minutes difference in the mean assembly times with 95% confidence.

The difference is statistically significant but management must decide if this size difference is of ***practical significance***.