CHAPTER 9: ESTIMATING THE VALUE OF A PARAMETER

[This chapter is based on Chapter 9 of the textbook]

In Chapter 8 we learned that statistics, such as the sample mean $\overline{x}$ and the sample proportion $\hat{p}$, are random variables. We studied their mean, standard deviation and distribution.

In statistics, our goal is to say something about population parameters using sample statistics. In this chapter, we learn how to estimate the value of a parameter using a statistic. We start with a definition.

Definition 9.1. *A point estimate is the value of a statistic that estimates the value of a parameter.*

In other words, when we compute a statistic for a specific sample, we are computing a point estimate of the population parameter. That is, if we want to estimate the population parameter, we may use the sample statistic. This inference is called point estimate. In the rest of this chapter we will learn about two point estimates: the sample proportion and the sample mean.

We already know how to compute the point estimates for the population proportion and the population mean. The point estimate for the population proportion p is the sample proportion

$$\hat{p} = \frac{x}{n}$$

and for the population mean is the sample mean

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

9.1 Estimating a population proportion

Let's start with a quick example.

Example 9.1. *The Gallup Organization conducted a poll in which a simple random sample of 1016 Americans 18 and older were asked, "Do you consider the amount of federal income tax you have to pay is fair?" Of the 1016 individuals surveyed, 558 said yes. Obtain a point estimate for the proportion of Americans 18 and older who believe the amount of federal income tax they pay is fair.*

Solution. We compute the sample proportion as follows:

$$\hat{p} = \frac{x}{n} = \frac{558}{1016} = 0.549$$

Then, we estimate that 54.9% of Americans 18 and older believe that the amount of federal income tax they have to pay is fair. □

As we learned in Chapter 8, the sample proportion is a random variable and depends on the sample. If we take a different sample, we may obtain a different proportion. For example, if we ask a different group, we may obtain that 49% of adult Americans believe that the amount of federal income tax they pay is fair.

A common practice in statistics to deal with this issue is reporting an interval and a level of confidence instead of a point estimate. For example, we may say that we are 90% confident that the proportion of adult Americans who believe that the amount of federal income tax they pay is fair is between 49.9% and 59.9% or, equivalently, 54.9% ± 5%.

If we want to increase the level of confidence, it means that we are more sure about our statement. Hence, the width of the interval we can give increases. For example, we may say that with 95% of confidence, the proportion of adult Americans who believe that the amount of federal income tax they pay is fair is 54.9% ± 7%.

We just made up those numbers, but we will learn a systematic way to compute them. Let's start with a definition.

Definition 9.2.

(i) *A confidence interval for an unknown parameter consists of an interval of numbers based on a point estimate.*

(ii) The level of confidence represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained. The level of confidence is denoted $(1-\alpha)\cdot 100\%$.

For example, a 90% level of confidence represents $\alpha = 0.1$ and implies that if a 100 confidence intervals were constructed, we expect (or, on average) 90 of them will include the parameter and 10 will not.

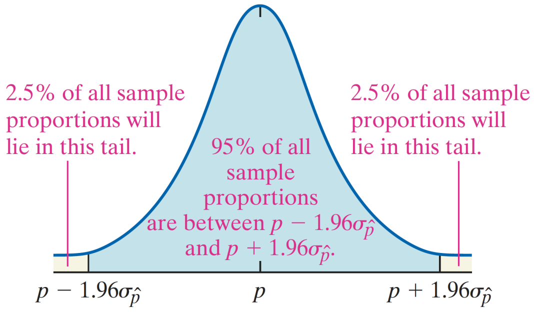
Confidence intervals are typically represented as follows:

$$\text{point estimate} \pm \text{margin of error}$$

To compute the confidence interval we use what we know about the sample proportion $\hat{p}$. Recall:

- The mean of the sample proportion is the population proportion: $\mu_{\hat{p}} = p$
- The standard deviation of the sample proportion is $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$
- If the sample is small with respect to the population (less than 5%) and the sample size is large enough ($np(1-p) \geq 10$), then the sample proportion is normally distributed.

In a standard-normal distribution, 95% of the values are between $z = -1.96$ and $z = 1.96$. We can compute these numbers using the standard-normal table. Therefore, 95% of the values of any normal distribution are within 1.96 standard deviations from the mean, as shown in the following picture:



Hence, 95% of the sample proportions $\hat{p}$ will satisfy:

$$p - 1.96\sigma_{\hat{p}} < \hat{p} < p + 1.96\sigma_{\hat{p}}$$

However, we want to use $\hat{p}$ to compute the extremes of the interval and have p in the middle. With some algebraic manipulations, we obtain

$$\hat{p} - 1.96\sigma_{\hat{p}} < p < \hat{p} + 1.96\sigma_{\hat{p}}$$

that is, with 95% confidence, the parameter p lies in the interval

$$\hat{p} \pm 1.96\sigma_{\hat{p}}$$

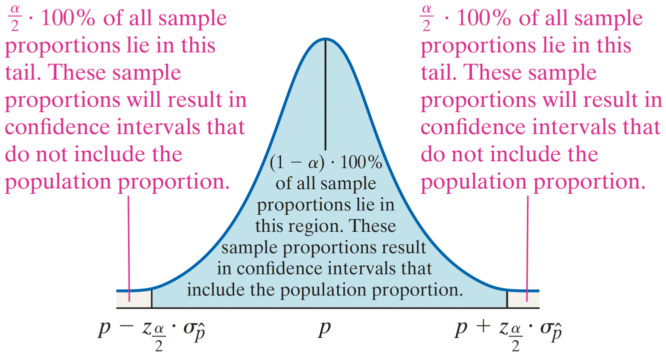
Definition 9.3. *The margin of error of a confidence interval is half of its width.*

For example, a 95% confidence interval for the population proportion has margin of error $1.96\sigma_{\hat{p}}$. Before moving on, let's introduce some notation.

> From now on, we will use z_α to denote the value of a standard-normal random variable such that the area to the right of the value is α. Since the normal density function is symmetric, we have
>
> $$P(Z \geq z_\alpha) = P(Z \leq -z_\alpha) = \alpha$$

For example, based on the analysis above, $z_{0.025} = 1.96$.

We just constructed a confidence interval for $\alpha = 5\%$, that is, with a level of confidence 95%. In practice, we can construct conficence intervals for any confidence level α. To do it, we must remember that $(1-\alpha)\cdot 100\%$ is the area under the normal curve that we want to cover, and the left and right tail need to weigh the same. In the following figure we illustrate this idea:



Then, we obtain the following general confidence interval:

$$\hat{p} - z_{\frac{\alpha}{2}}\sigma_{\hat{p}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sigma_{\hat{p}}$$

We always can obtain the value $z_{\frac{\alpha}{2}}$ from the standard normal table, but here are some of the most frequently used:

Level of confidence $(1-\alpha)\cdot 100\%$	Area in each tail $\frac{\alpha}{2}$	Critical value $z_{\frac{\alpha}{2}}$
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.575

An important observation is that the confidence level is in the method; not in the interval. In other words, 90% confidence means that the method works for 90% of all the samples. The level of confidence is **not** the probability that the parameter lies in the interval.

Constructing confidence intervals

The following result shows us how to compute a $(1-\alpha)\cdot 100\%$ confidence interval for a population proportion.

Theorem 9.1. *Suppose that a simple random sample of size n is taken from a population. A $(1-\alpha)\cdot 100\%$ confidence interval for p is given by*

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where n is the sample size and $\hat{p}$ is the sample proportion.

The margin of error of the confidence interval above is

$$E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Note that we must have $n\hat{p}(1-\hat{p}) \geq 10$ and n smaller than 5% of the population size to construct this interval.

Notice that we use the sample proportion $\hat{p}$ instead of the population proportion p to estimate the standard deviation of $\hat{p}$.

Let's do an example.

Example 9.2. *In the Parent-Teen Cell Phone Survey conducted by Princeton Survey Research Associates International, 800 randomly sampled 16- to 17-year-olds living in the United States were asked whether they have ever used their cell phone to text while driving. Of the 800 teenagers surveyed, 272 indicated that they text while driving. Obtain a 95% confidence interval for the proportion of 16- to 17-year-olds who text while driving.*

Solution. We use the theorem above. The first step is to compute the point estimate $\hat{p}$. We obtain:

$$\hat{p} = \frac{x}{n} = \frac{272}{800} = 0.34$$

Next, we check that the conditions are satisfied. Indeed, 800 is below 5% of teenagers in the United States and

$$n\hat{p}(1-\hat{p}) = 800 \cdot 0.34 \cdot 0.66 = 179.52 \geq 10$$

Hence, we can compute the confidence interval as indicated in the theorem.

We want a 95% confidence interval, so $\alpha = 1 - 0.95 = 0.05$ and $\frac{\alpha}{2} = 0.025$. Using the standard normal table (or the summary table above), we obtain

$$z_{\frac{\alpha}{2}} = 1.96$$

Hence, the margin of error is

$$\begin{aligned} E &= z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &= 1.96\sqrt{\frac{0.34 \cdot 0.66}{800}} \\ &= 0.033 \end{aligned}$$

and the confidence interval is

$$0.34 \pm 0.033$$

Therefore, the lower bound is

$$0.34 - 0.033 = 0.307$$

and the upper bound is

$$0.34 + 0.033 = 0.373$$

Hence, we are 95% confident that the proportion of teenagers who text while driving is between 0.307 and 0.373

□

Example 9.3. *For the above example, construct a 99% confidence interval and compare your answers.*

Solution. We already know that the assumptions are satisfied. All we need to do is compute the interval again. In this case, $\alpha = 1 - 0.99 = 0.01$ and $\dfrac{\alpha}{2} = 0.005$. Using the standard-normal table, we obtain

$$z_{\frac{\alpha}{2}} = 2.575$$

Then, the new margin of error is

$$\begin{aligned} E &= z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &= 2.575\sqrt{\frac{0.34 \cdot 0.66}{800}} \\ &= 0.043 \end{aligned}$$

Then, the confidence interval is

$$0.34 \pm 0.043$$

that is, the lower bound is

$$0.34 - 0.043 = 0.297$$

and the upper bound is

$$0.34 + 0.043 = 0.383$$

When we increased the confidence level from 95% to 99%, the margin of error increased from 0.033 to 0.043. Hence, the confidence interval became wider. $\square$

If we look at the margin of error (width of the confidence interval) observe that it depends on the confidence level α, the point estimate $\hat{p}$ and the sample size n. We already learned how the confidence level affects the interval. The next question we address is how big should the sample size be to obtain a certain margin error.

Theorem 9.2. *If we have an estimate of $\hat{p}$, the sample size required to obtain a $(1-\alpha)\cdot 100\%$ confidence interval for p with a margin error E is given by*

$$n = \hat{p}(1-\hat{p})\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2$$

rounded up to the next integer.

If an estimate of $\hat{p}$ is not available, we use $\hat{p} = 0.5$ and obtain

$$n = 0.25\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2$$

rounded up to the next integer.

Let's see an example.

Example 9.4. *An economist wants to know if the proportion of the US population who commutes to work via carpooling is on the rise. What sample size should be obtained if the economist wants an estimate with 90% confidence and margin of error 0.02 if:*

- *(a) the economist uses the 2009 estimate of 10% obtained from the American Community Survey?*
- *(b) the economist does not use any prior estimates?*

Solution. In both cases, we want an estimate with 90% confidence, so $\alpha = 1 - 0.9 = 0.1$, which implies $\frac{\alpha}{2} = 0.05$. Hence,

$$z_{\frac{\alpha}{2}} = 1.645$$

(a) In this case we use the first formula with $\hat{p} = 0.1$. We obtain:

$$\begin{aligned} n &= \hat{p}(1-\hat{p})\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2 \\ &= 0.1\cdot 0.9\left(\frac{1.645}{0.02}\right)^2 && = 608.9 \end{aligned}$$

Hence, the economist must survey 609 people.

(b) In this case we don't use the estimate from 2009, so we obtain:

$$\begin{aligned} n &= 0.25\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2 \\ &= 0.25\left(\frac{1.645}{0.02}\right)^2 \\ &= 1691.3 \end{aligned}$$

Hence, the economist must survey 1692 people.

The second case is considerably larger than the first one, which represents the cost of not having an estimate. $\square$

9.2 Estimating a Population Mean

The goal of this section is to construct confidence intervals for the sample mean $\overline{x}$. Inspired from our construction of the confidence intervals for the sample proportion $\hat{p}$, we want an interval of the form:

$$\text{point estimate} \ \pm \ \text{margin of error}$$

The point estimate in this case is $\overline{x}$. From Chapter 8, we know that $\overline{x}$ is normally distributed with mean $\mu_{\overline{x}} = \mu$ and standard deviation $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$, where μ and σ are the population mean and standard deviation, respectively. Hence, using the same logic as in Section 9.1, we would obtain the following interval:

$$\overline{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

However, computing the standard deviation of the sample proportion and the sample mean are very different processes. In the case of the sample proportion, we had

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

and we were able to easily estimate $\sigma_{\hat{p}}$ using the sample proportion $\hat{p}$. In the case of the sample mean, instead, the standard deviation is

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

and we could estimate it as

$$\sigma_{\overline{x}} \approx \frac{s}{\sqrt{n}},$$

where s is the sample standard deviation. However, s is a statistic and, as such, it is a random variable too. Therefore, the value

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$

is **not** standard normal. In other words, changing σ by s in the computation of the z-score of $\overline{x}$ changes the distribution. Then, we need a new model and we describe it below.

Theorem 9.3. *Suppose that a simple random sample of size n is taken from a population. If the population from which the sample is drawn follows a normal distribution, the distribution of*

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$

is the Student's t distribution with $n-1$ degrees of freedom.

The Student's t distribution is the probability density function of continuous random variables, and has the following properties:

(1) The t-distribution is different for different degrees of freedom

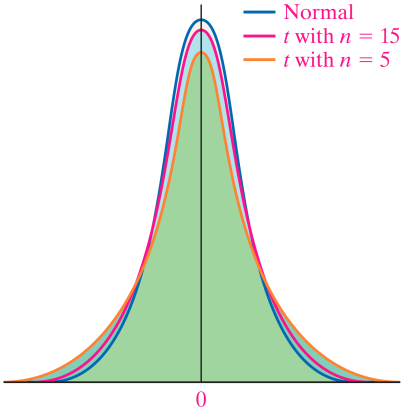
(2) The t-distribution is centered at 0 and is symmetric about 0. Hence, the area under the curve to the right of 0 is 0.5, and the area to the left of 0 is also 0.5.

(3) As the value of t increases or decreases, the graph approaches 0 but it never reaches 0

(4) The area in the tails of the t-distribution is a big greater than the area in the tails of a standard normal distribution. This increased area in the tails is due to the use of the sample standard deviation s (instead of the population standard deviation σ), which introduces more variability.

(5) As the sample size n increases, the density curve of the t-distribution approaches a standard normal density function.

In conclusion, the t-distribution is very similar to the standard-normal distribution. The only difference is that its variability is larger than the normal distribution and, hence, the tails of the bell are a bit thicker. In the following picture we show a standard-normal distribution, and the t-distribution for 2 values of the sample size n.



Similarly to the standard-normal distribution, we compute probabilities associated to the t-distribution using a table.

In this case, the table gives us the value of the t random variable such that the area to the right of t is α (columns) for different values of the degrees of freedom (rows). Extending the notation introduced for the normal distribution, the table gives us the value of t_α. We present the table below[1].

Of course, this table does not show all the possible values of the degrees of freedom. If we have a sample size that is not listed, we can approximate the degrees of freedom by the closest number in the table. Additionally, for $n > 1000$ we can use the standard-normal distribution as a very close approximation. Hence, in the last row of the table the values of z_α are presented.

---

[1] Table VII in Appendix A of the textbook

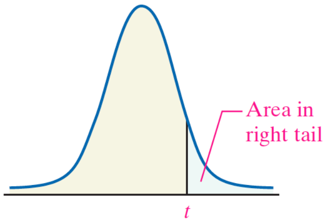


Table VII

***t*-Distribution**
Area in Right Tail

Degrees of Freedom	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
31	0.682	0.853	1.054	1.309	1.696	2.040	2.144	2.453	2.744	3.022	3.375	3.633
32	0.682	0.853	1.054	1.309	1.694	2.037	2.141	2.449	2.738	3.015	3.365	3.622
33	0.682	0.853	1.053	1.308	1.692	2.035	2.138	2.445	2.733	3.008	3.356	3.611
34	0.682	0.852	1.052	1.307	1.691	2.032	2.136	2.441	2.728	3.002	3.348	3.601
35	0.682	0.852	1.052	1.306	1.690	2.030	2.133	2.438	2.724	2.996	3.340	3.591
36	0.681	0.852	1.052	1.306	1.688	2.028	2.131	2.434	2.719	2.990	3.333	3.582
37	0.681	0.851	1.051	1.305	1.687	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.681	0.851	1.051	1.304	1.686	2.024	2.127	2.429	2.712	2.980	3.319	3.566
39	0.681	0.851	1.050	1.304	1.685	2.023	2.125	2.426	2.708	2.976	3.313	3.558
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
70	0.678	0.847	1.044	1.294	1.667	1.994	2.093	2.381	2.648	2.899	3.211	3.435
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
90	0.677	0.846	1.042	1.291	1.662	1.987	2.084	2.368	2.632	2.878	3.183	3.402
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291

Before constructing the confidence intervals of $\overline{x}$, let's do an example of the use of the table of the t-distribution.

Example 9.5. *Find the value of a t-distributed random variable such that the area to the right of the value is 0.10. Assume 15 degrees of freedom.*

Solution. In other words, we are asked to compute $t_{0.10}$ with 15 degrees of freedom. We go to the row of 15 degrees of freedom and the column corresponding to 0.10. We obtain

$$t_{0.01} = 1.341$$

In the following picture we show how to find this number in the t table.

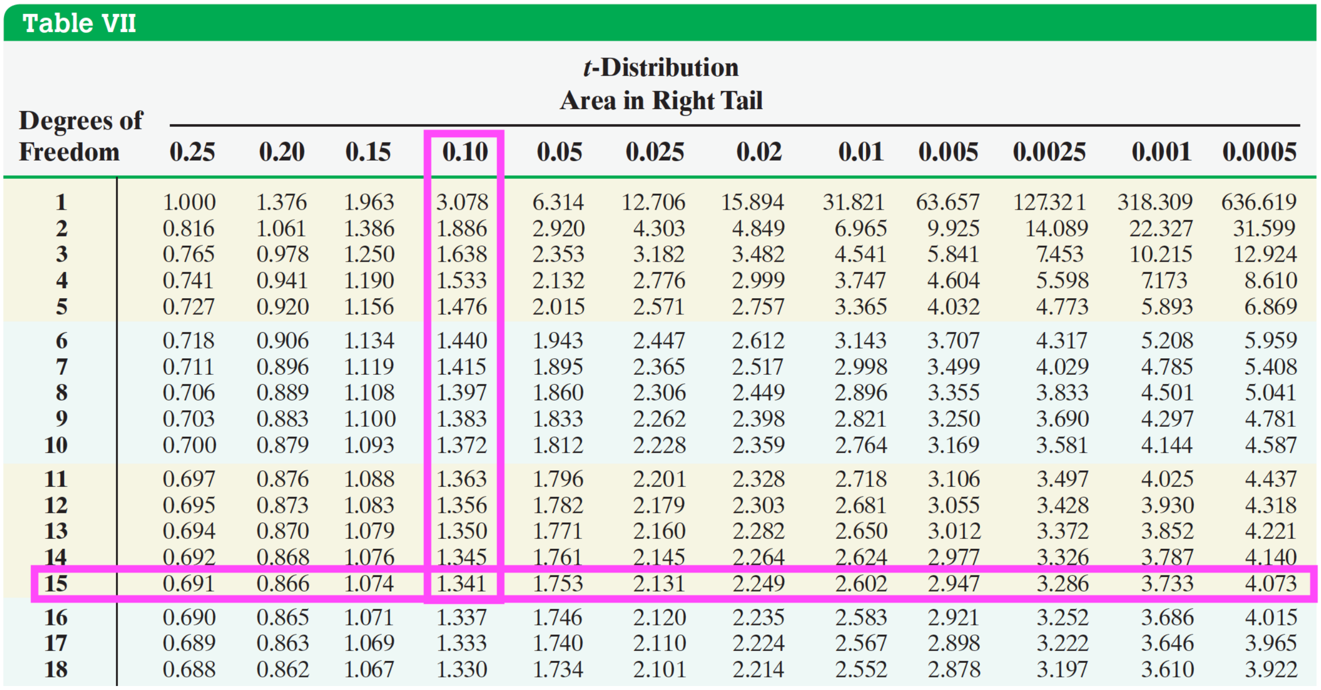


Table VII

***t*-Distribution**
Area in Right Tail

Degrees of Freedom	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922

□

Constructing the confidence interval

The process is very similar to the construction of the confidence interval of the sample proportion $\hat{p}$. The main difference is that now we use the t-distribution, as indicated in the following theorem.

Theorem 9.4. *If*

(i) the sample size is small relative to the population (no more than 5%) and

(ii) the data comes from a normally-distributed population <u>or</u> the sample size is large ($n \geq 30$)

Then, a $(1 - \alpha) \cdot 100\%$ confidence interval for the population mean μ is given by

$$\overline{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

where $t_{\frac{\alpha}{2}}$ is the critical value with $n - 1$ degrees of freedom.

The margin of error of such confidence interval is

$$E = t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Let's do an example.

Example 9.6. *The website fueleconomy.gov allows drivers to report the miles per gallon of their vehicle. The data in the following table*[2] *show the reported miles per gallon on 2011 Ford Focus automobiles for 16 different owners. You may assume that the miles per gallon of 2011 Ford Focus automobiles is normally distributed.*

35.7	*37.2*	*34.1*	*38.9*
32.0	*41.3*	*32.5*	*37.1*
37.3	*38.8*	*38.2*	*39.6*
32.2	*40.9*	*37.0*	*36.0*

Construct a 95% confidence interval for the mean miles per gallon of a 2011 Ford Focus. Interpret the interval.

Solution. We first verify the conditions of the theorem.

(i) There are probably thousands of Ford Focus automobiles in the streets, so our sample size of $n = 16$ cars is well below the 5% of the population.

(ii) The sample size is not large because $n = 16 < 30$. However, we are told that we can assume that the data come from a normally-distributed population.

Hence, we can compute the confidence interval as

$$\overline{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

We may compute $\overline{x}$ and s using a graphic calculator or Excel, and we obtain:

$$\overline{x} = 36.8 \qquad \text{and} \qquad s = 2.92$$

To compute $t_{\frac{\alpha}{2}}$ we use the t-distribution table. We want a 95% confidence interval. Then,

$$\alpha = 1 - 0.95 = 0.05$$

and we obtain $\frac{\alpha}{2} = 0.025$. Our sample size is $n = 16$, so we search in the table for $n - 1 = 15$ degrees of freedom and obtain $t_{0.025} = 2.131$. The picture below shows how we find this number in the t-distribution table:

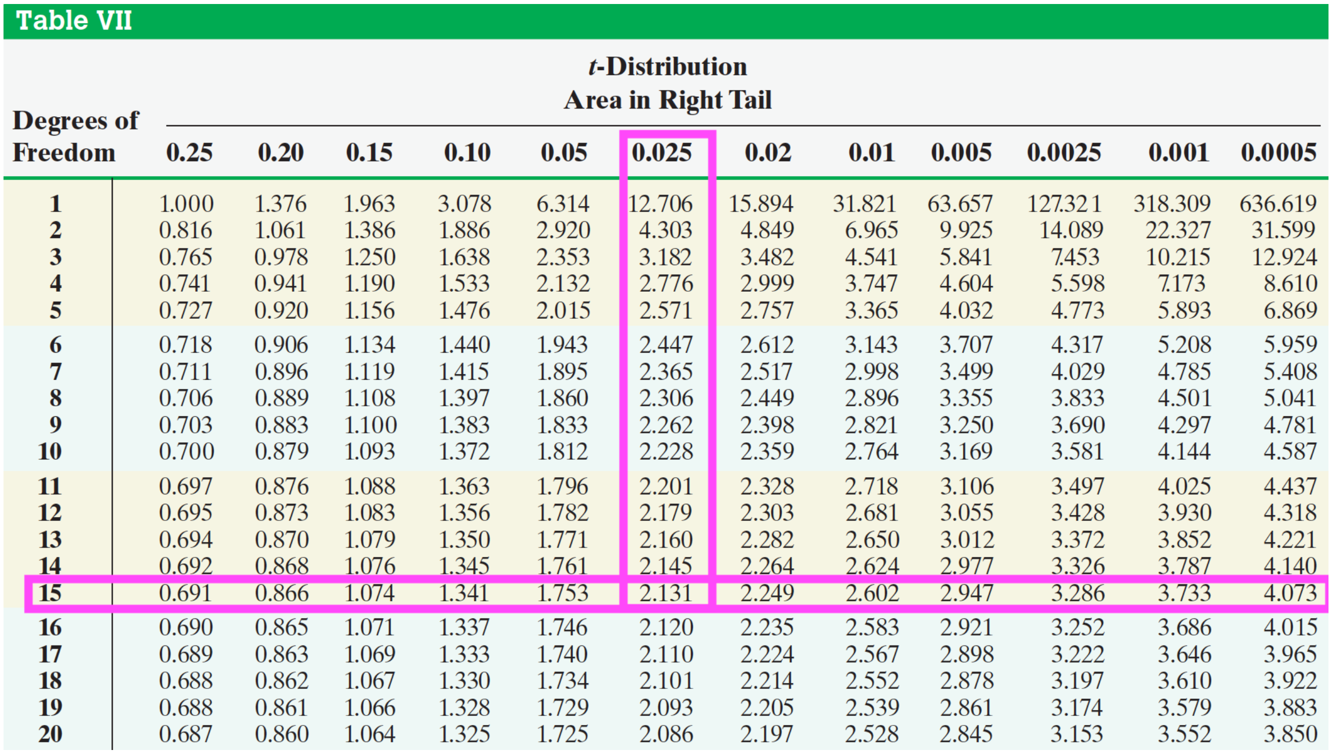


Table VII

***t*-Distribution**
Area in Right Tail

Degrees of Freedom	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850

---

[2] Source: www.fueleconomy.gov

Then, the confidence interval is

$$36.8 \pm 2.131 \cdot \frac{2.92}{\sqrt{16}} = 36.8 \pm 1.56$$

Hence, the lower bound of the interval is 35.24 and the upper bound is 38.36, that is, we are 95% confident that the mean miles per gallon of all 2011 Ford Focus automobiles is between 35.24 and 38.36.

□

The last question we answer is the sample size needed to obtain a certain margin of error. When we computed the sample size needed for the sample proportion, we simply solve the algebraic equation to obtain n. In this case, we cannot simply do that because the margin of error depends on $t_{\frac{\alpha}{2}}$, and $t_{\frac{\alpha}{2}}$ depends on the sample size n. In other words, we would obtain the following formula:

$$n = \left(\frac{t_{\frac{\alpha}{2}} \cdot s}{E}\right)^2$$

However, in order to compute the right-hand side we need $t_{\frac{\alpha}{2}}$, and in order to obtain $t_{\frac{\alpha}{2}}$ from the table, we need n. Hence, the formula is useless.

Instead, we use an approximation to $t_{\frac{\alpha}{2}}$ inspired in the motivation to use the t-distribution. We present the result below.

Theorem 9.5. *The sample size n required to estimate the population mean μ with a level of confidence $(1-\alpha)\cdot 100\%$ within a specified margin of error E is*

$$n = \left(\frac{z_{\frac{\alpha}{2}} s}{E}\right)^2,$$

rounding up to the nearest whole number.

Let's see an example.

Example 9.7. *Consider the situation described in Example 9.6. How large a sample is required to estimate the mean of miles per gallon within 0.5 of error with 95% confidence?*

Solution. We compute n as indicated in the theorem above and obtain

$$\begin{aligned} n &= \left(\frac{z_{\frac{\alpha}{2}} s}{E}\right)^2 \\ &= \left(\frac{1.96 \cdot 2.92}{0.5}\right)^2 \\ &= 131.02 \end{aligned}$$

Rounding up, we obtain that if we want a margin of error of 0.5 with 95% confidence, we need 132 cars. □