CHAPTER FIVE   FINDING CONFIDENCE INTERVALS FOR PROPORTIONS and MEANS

What proportion of undergraduate students at American University smoke cigarettes?

POPULATION   SAMPLE

π ? (population proportion)

69/300 .23 (sample proportion) **π(hat)**

n = 300 (sample size)

We use **π(hat)** to create a confidence interval that **π** falls in.

We do not know exactly what **π** is, but we would like to say that **π** is between two numbers with a certain amount of confidence.

That is, LB < **π** < UB or **π** is inside of a confidence interval (LB , UB) given an acceptable confidence level.

π(hat) is also called a point estimate.

We find the confidence interval by using the point estimate along with the standard error (the standard deviation of the sampling distribution) The formula is given below:

π(hat) ± zsqrt((π(hat)(1 – π(hat))/n))

zsqrt((π(hat)(1 – π(hat))/n)) is called the margin of error

(π(hat)(1 – π(hat))/n) is the standard error and is also referred to as **se**

How do we get the **z** in the formula? It is a function of the confidence level indicated

Confidence Level	**z**
.99	2.58
.95	1.96
.90	1.64

Let's find a 95% confidence interval for our example:

.23 ± (1.96) sqrt((.23(1-.23))/300)

= .23 ± (1.96) sqrt(.23(.77)/300)

= .23 ± (1.96) sqrt(.1771/300)

= .23 ± (1.96) sqrt(.00059) = .23 ± .0476 ; confidence interval (.1824 , .2776)

R code for finding the confidence interval

prop.test(x = 69, n = 300, conf.level = .95)

R confidence output

95 percent confidence interval:

0.1844451   0.2826189

Lets compare this interval to the interval that we obtained by using the formula

(.1824 , .2776)

Apart from rounding differences, the intervals are practically the same.

In either case our population proportion approximately falls between **.18** and **.28**

Interpretation: We are 95% confident that **π** falls between .18 and .28

Also, if this process for finding a confidence interval is repeated, 95% of your confidence intervals will capture **π.**

For example, if you sampled 100 times from the population you would get 100 confidence intervals. 95 of these intervals would capture **π**, 5 would not.

If you sampled from the population 140 times and you wanted a 95% confidence interval, 133 of the intervals would work, 7 would not capture the desired population parameter **π**

There is no guarantee that **π falls in** your interval, but chances are it is one of the 95 intervals that work.

Text book problem,

The General Survey asks whether you agree or disagree with the following statement; "It is much better for everyone involved if the girls in the family were not educated as much as the boys." The sample proportion agreement was .78 in 1977 and .17 in 2016 (n = 1845)

a) Construct the 95% confidence interval for 2016 and interpret it.

x/1845 = .17 , x = .17(1845) = 313

R code

prop.test(x = 313, n = 1845, conf.level = .95)

prop.test(x = 313, n = 1845, conf.level = .95)

95 percent confidence interval:

0.1529526   0.1877373

Interpretation:

We are 95% confident that **π falls between** 0.1529526 and 0.1877373

b) Show that the estimated standard error is **.0087**

Solution:

by definition

se =   sqrt((π(hat)(1- π(hat))/n))

se =   sqrt((.17(.83)/1845)) = .0087

Another Text book problem

Of 1824 voters sampled in an exit poll, 60.5% said that they voted for a certain candidate. Is there enough evidence to predict the winner of the election. Base your decision on a 95% confidence interval.

x/1824 = .605 , x = 1824(.605) = 1103.52 ≈ 1103 , n = 1824

prop.test(x = 1103 , n = 1824 , conf.level = .95)

95 percent confidence interval:

0.5818040   0.6271803

Yes, there is enough evidence to predict a winner for the population parameter lies between the interval 0.5818040   0.6271803 and the lower bound is .5818940; well above 50%.

CHAPTER FIVE FINDING CONFIDENCE INTERVALS FOR PROPORTIONS and MEANS

POPULATION SAMPLE

mean of the population **μ** ?

mean from the sample **y(bar)**

standard deviation from the sample **s**

sample size **n**

intervals will be based on a t distribution instead of a z distribution.

The formula for building a confidence interval for a population mean is given below

y(bar) ± t_k (**se**)

where **se** = s / sqrt(n) and df = n - 1

We will use R to produce the confidence intervals for means when possible. For some problems we will use the t distribution table

Example 1 (Summarized Data)

Find a 95% confidence interval for a population mean if a sample is taken of sample size 10, sample mean of 380.67, and a sample standard deviation of 42.47.

Solution:

Step 1) compute **se** **42.47 / sqrt(10) = 13.43**

Step 2) Find the t distribution factor t_k (use the t distribution table) **2.262**

Step 3) Find y(bar) y(bar) = **380.67**

Step 4) **Build the confidence interval**

380.67 ± 2.262(13.43)

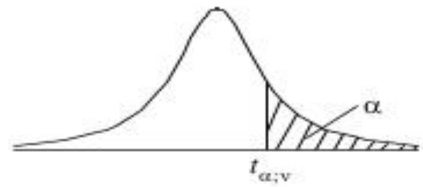
380.67 ± 30.3787

(350.29 , 411.05)

Interpretation: We are 95% confident that the population mean falls between 350.29 and 411.05.

Table of the Student's *t*-distribution

The table gives the values of $t_{\alpha;\nu}$ where
$\Pr(T_\nu > t_{\alpha;\nu}) = \alpha$, with ν degrees of freedom



ν \ α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

CHAPTER FIVE FINDING CONFIDENCE INTERVALS FOR PROPORTIONS and MEANS

Example 2 (Un summarized / Raw Data)

The following 20 Physics test scores were randomly sampled from a population of Physics test scores. Find a 99% confidence interval for the mean test score of the population.

S = {67, 72, 77, 68, 80, 85, 77, 60, 72, 72, 91, 58, 74, 73, 62, 73, 88, 81, 67, 71}

R code

```
S = c(67, 72, 77, 68, 80, 85, 77, 60, 72, 72, 91, 58, 74, 73, 62, 73, 88, 81, 67, 71)

S

t.test(S, conf.level = .99)
```

R Output

```
One Sample t-test

data:  S

t = 37.513, df = 19, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

99 percent confidence interval:

 67.80219  78.99781

sample estimates:

mean of x

    73.4
```

Interpretation:

We are 99% confident that the mean test score of the population lies between 67.80219 and 78.99781