Chapter 8 Estimating with Confidence

One of the main powers of stats is the ability to know what a whole population is like based on a sample.

This is used to influence every part of your lives.

Vending machines
Clothing
Music airtime
Next year's cars
Medicine
Social Services
What classes are offered by your school
How the government “suggests” you live your life.

Anger Study

Part 1 – anger survey.

Fill out (confidentially) the anger survey.

Clinical Anger Scale (CAS) Survey Scoring Guide

Score A = 0, B = 1, C = 2, and D = 3
Score your responses and add up the scores.

Score	CAS Interpretation
0 – 13	minimal clinical anger
14 – 19	mild clinical anger
20 – 28	moderate clinical anger
29 – 63	severe clinical anger

Anger leads to Heart Disease

A Newspaper reports: Compared to "Low Anger" people, "High Anger" people are 2.2 times more likely to get heart disease (HD) and 2.7 times more likely to have an acute heart attack.

The article was based on a study published in the medical journal "Circulation", 101(2000) pp 2034-2039

The Study:

8474 people were given a Trait Anger Scale Test.

633 were High Anger on anger scale
4731 were Medium Anger
3110 were Low Anger

The people tested were:

45-64 years old
randomly selected
diverse race and socio-economic
normal blood pressure

The results:

Six years later the people were evaluated for heart disease (HD) and adjusted for "lurking" variables.

High Anger people were 2.2 times more likely to get HD than Low Anger and were 2.7 times more likely to have an acute heart attack.

The results:

Does anger sound like a serious problem?

53 out of the 3110 Low Anger people got HD

27 out of the 633 High Anger people got HD

Still sound serious?

Was the newspaper misleading its readers?

As intelligent and educated readers you will be asking yourself two questions.

What does this study say about the sample?

What does this study say about the population?

What does this study say about the sample?

"Compared to the Low Anger people the High Anger people are 2.2 times as likely to get HD and 2.7 times as likely to have an acute heart attack."

A single number is called a point estimate

The mean is the most commonly used point estimate.

What does this study say about the population?

"Compared to the Low Anger people, the High Anger people are between 1.35 and 3.55 times as likely to get HD, with 95% certainty, and are between 1.48 and 4.90 times as likely to have an acute heart attack, with 95% certainty."

A range is called an interval estimate

The confidence interval is the most commonly used interval estimate.

The formula for finding the confidence interval is:

$$\bar{X} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Where $z_{\frac{\alpha}{2}}$ depends on the level of confidence you want.

For 99% confidence $z_{\frac{\alpha}{2}} = 2.58$

For 95% confidence $z_{\frac{\alpha}{2}} = 1.96$

For 90% confidence $z_{\frac{\alpha}{2}} = 1.65$

I'm sure you recognized these from the z-table!

e.g.1

The president of a large university wishes to estimate the average age of the students enrolled. From past studies, the standard deviation is known to be 2 years. A sample of 50 students is selected, and the mean found to be 23.2 years. Find the 95% confidence interval of the population mean.

Solution to example 1

95% means $z_{\frac{\alpha}{2}}$ = 1.96, so, using the CI formula, we get:

$$23.2 - 1.96\left(\frac{2}{\sqrt{50}}\right) < \mu < 23.2 + 1.96\left(\frac{2}{\sqrt{50}}\right)$$

$23.2 - 0.6 < \mu < 23.2 + 0.6$

$22.6 < \mu < 23.8$

With 95% confidence the mean age of students lies between 22.6 and 23.8 years.

e.g. 2

A certain medication is known to increase the pulse rate of its users. The s.d. is 5 beats per minute (bpm). A sample of 30 users had an average pulse rate of 104 bpm. Find the 99% CI of the population mean.

Solution to e.g. 2

99% means $z_{\frac{\alpha}{2}}$ = 2.58, so using the CI formula,

$$104 - 2.58\left(\frac{5}{\sqrt{30}}\right) < \mu < 104 + 2.58\left(\frac{5}{\sqrt{30}}\right)$$

$104 - 2.4 < \mu < 104 + 2.4$

$101.6 < \mu < 106.4$

With 99% confidence the mean pulse rate of users of this medicine lies between 101.6 and 106.4 bpm.

e.g.3

Find the true mean daily revenue from this sample of 30 parking meters, use a 95% CI.

$2.60	1.05	2.45	2.90	1.30
3.10	2.35	2.00	2.40	2.35
2.40	1.95	2.80	2.50	2.10
1.75	1.00	2.75	1.80	1.95
1.95	2.05	2.15	2.45	3.00
2.85	2.65	1.95	2.50	2.00

Of course you calculated mean and s.d.!

ü = 2.235, and Sx = 0.5279

Solution to e.g. 3

95% means $z_{\frac{\alpha}{2}}$ = 1.96, so using the CI formula,

$$2.235 - 1.96\left(\frac{0.5279}{\sqrt{30}}\right) < \mu < 2.235 + 1.96\left(\frac{0.5279}{\sqrt{30}}\right)$$

$$2.235 - 0.1889 < \mu < 2.235 + 0.1889$$

$$2.0461 < \mu < 2.4239$$

With 95% confidence the mean daily revenue of parking meters lies between $2.05 and $2.42.

Note that we used sample s.d. in the last example. We didn't know population s.d. and n ≥ 30.

The formula used was:

$$\bar{X} - z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)$$

<u>This formula can only be used when n ≥ 30</u>

Scores on the Stanford-Binet IQ test are normally distributed with $\sigma = 16$

Here are the IQs from 20 randomly selected people. Find the 85% confidence interval of the population mean.

104.32	98.19	107.16	102.28
100.38	93.97	95.89	104.43
108.73	104.11	100.67	96.87
99.84	100.25	101.32	94.24
102.23	94.32	97.66	101.44

Here are the prices in $US of 20 randomly selected Corvettes listed for sale at cars.com.

Find the 92% confidence interval of the population mean.

33,900	24,900	27,995	26,490
37,900	39,990	33,995	21,500
22,900	28,900	47,995	53,995
49,900	10,995	34,995	19,995
12,000	47,995	44,900	37,488

Souce: http://www.cars.com 2007