

Introductory Statistics Tutorial

Chapter 8 – Confidence Intervals

Section A: Confidence intervals for a mean, proportion and difference between means

1. The 1995 exam for 528.188 (Stage I Statistics for Commerce students) had a possible total of 64 marks. A random sample of 30 was selected from all of the exam scripts. The data collected and summary statistics follow.

46 32 24 20 51 33 35 43 26 29 59 41 30 35 49
53 32 50 52 23 25 53 51 34 26 29 40 38 45 42

| sample size | sample mean | sample standard deviation |
|-------------|-------------|---------------------------|
| 30 | 38.20 | 10.85 |

You will use this sample to construct a 95% confidence interval for the mean.

- (a) State the parameter θ (using a symbol and in words).
- (b) State the estimate $\hat{\theta}$ (using a symbol, in words and as a number).
- (c) Calculate $se(\hat{\theta})$.
- (d) State the value of df .
- (e) Use the table for the Student's t -distribution to write down the value of the t -multiplier.

- (f) Calculate the 95% confidence interval for the mean.

- (g) Interpret the confidence interval.

- (h) Does the confidence interval contain the true mean? Discuss briefly.

2. Tuberculosis (TB) is known to be a highly contagious disease. In 1995 a study was carried out on a random sample of 1074 Spanish prisoners. The study investigated factors that might be associated with the tuberculosis infection. Some of the results follow.

| | Prisoners with tuberculosis | Total number of prisoners |
|--------|-----------------------------|---------------------------|
| Male | 556 | 984 |
| Female | 36 | 90 |

You will use this sample to construct a 95% confidence interval for the proportion of female prisoners who had tuberculosis.

- (a) State the parameter θ (using a symbol and in words).
- (b) State the estimate $\hat{\theta}$ (using a symbol, in words and as a number).
- (c) Calculate $se(\hat{\theta})$.
- (d) Use the table for the Student's t -distribution to write down the value of the z -multiplier.

- (e) Calculate the 95% confidence interval for the proportion of female prisoners who had tuberculosis.

- (f) Interpret the confidence interval.

3. Adapted from Example 8.4.1 (page 339) from *Chance Encounters* (the text for this course).

Banford et al. [1982] noted that thiol concentrations within human blood cells are seldom determined in clinical studies, in spite of the fact that they are believed to play a key role in many vital processes. They reported a new reliable method for measuring thiol concentration and demonstrated that, in one disease at least (rheumatoid arthritis), the change in thiol status in the lysate from packed blood cells is substantial. There were two groups of volunteers, the first group being “normal” and the second suffering from rheumatoid arthritis. We shall treat the two groups as random samples from the normal and rheumatoid populations respectively (for the area in which the study was undertaken) and will estimate $\mu_R - \mu_N$, the difference in true mean thiol levels between the rheumatoid and normal populations.

MINITAB Output

Two sample T for Rheumatoid vs Normal

| | N | Mean | StDev | SE Mean |
|----------|---|--------|--------|---------|
| Rheumato | 6 | 3.465 | 0.440 | 0.18 |
| Normal | 7 | 1.9214 | 0.0756 | 0.029 |

95% CI for mu Rheumato - mu Normal: (1.08, 2.012)

T-Test mu Rheumato = mu Normal (vs not =): T = 8.48 P = 0.0004 DF = 5

Which **one** of the following statements is **false**?

- (1) On the basis of this analysis, it is not plausible that there is no difference in the true mean thiol levels of sufferers of rheumatoid arthritis and non-sufferers.
- (2) We estimate, with 95% confidence, the true mean thiol level in non-sufferers of rheumatoid arthritis to be somewhere between 1.1 units and 2.0 units lower than the true mean thiol level of sufferers.
- (3) On average, sufferers of rheumatoid arthritis have a higher thiol level than non-sufferers.
- (4) On the basis of this analysis, there is no evidence of a difference between the true mean thiol levels for sufferers of rheumatoid arthritis and non-sufferers.
- (5) We estimate, with 95% confidence, the true mean thiol level of sufferers of rheumatoid arthritis to be somewhere between 1.1 units and 2.0 units higher than the true mean thiol level of non-sufferers.

Questions 4 and 5 refer to the following information.

1996 Semester 2 Exam

In 1990 *CNN/Time* sought information on how young American adults viewed their parents’ marriage. In a telephone poll, one of the questions they asked of six hundred and two (602) 18-29 year old Americans was “Would you like to have a marriage like the one your parents have?” Forty-four percent (44%) responded “Yes”.

4. *CNN/Time* were interested in determining what proportion of the 18-29 year old American population would answer “Yes” to this question. Which **one** of the following statements is **false**?

- (1) The value of the parameter of interest is an unknown quantity.
- (2) In this context, 0.44 is an estimate for the parameter of interest.
- (3) The parameter of interest depends on the sample and hence is a random quantity.
- (4) A confidence interval for the parameter of interest will give a range of possible values for this parameter.
- (5) The parameter of interest is the proportion of 18-29 year old Americans who would have answered “Yes” in 1990.

5. An approximate 95% confidence interval for the proportion of the 18-29 year old American population who would have answered “Yes” to this question in 1990 is [0.400, 0.480]. If two thousand four hundred (2400) 18-29 year old Americans had been sampled instead of six hundred and two (602) 18-29 year old Americans, then the new 95% confidence interval would be approximately:

- (1) twice as wide.
- (2) one-quarter as wide.
- (3) half as wide.
- (4) four times as wide.
- (5) equally as wide.

Section B: Confidence interval for a difference in proportions

1. In 1991 a random sample of New Zealand adults were surveyed about their working hours and the number of jobs they had. A similar survey was carried out in 1994.

Identify the sampling situation as:

Situation (a): *Two independent samples,*

Situation (b): *Single sample, several response categories,*

Situation (c): *Single sample, two or more Yes/No items,*

in the following cases.

- (a) We want to compare the proportion of females working 20-39 hours in 1994 with the proportion of females working 40 hours or more in 1994.
- (b) We want to compare the proportion of males working 40 hours or more in 1991 with the proportion of females working 40 hours or more in 1991.
- (c) In the same survey people were also asked if they had 2 or more jobs. We want to compare the proportion of people who had 2 or more jobs in 1994 with the proportion of people who worked 40 hours or more per week in 1994.
- (d) We want to compare the proportion of females working 40 hours or more in 1994 with the proportion of females working 40 hours or more in 1991.

Questions 2 to 4 refer to the following information.

Tuberculosis (TB) is known to be a highly contagious disease. In 1995 a study was carried out on a random sample of 1074 Spanish prisoners. The study investigated factors that might be associated with the tuberculosis infection. The results follow.

| Variable | | Prisoners with tuberculosis | Total number of prisoners |
|------------------------|--------|-----------------------------|---------------------------|
| Gender | Male | 556 | 984 |
| | Female | 36 | 90 |
| Race | White | 496 | 886 |
| | Gypsy | 74 | 152 |
| | Other | 22 | 36 |
| Intravenous Drug Users | Yes | 361 | 629 |
| | No | 231 | 445 |
| HIV Positive | Yes | 186 | 294 |
| | No | 406 | 780 |
| Re-imprisonment | Yes | 272 | 456 |
| | No | 320 | 618 |

2. Identify the sampling situation as:
- Situation (a): *Two independent samples,*
- Situation (b): *Single sample, several response categories,*
- Situation (c): *Single sample, two or more Yes/No items,*
- in the following cases:
- (a) Of those prisoners who had TB, we want to compare the proportion of white prisoners with the proportion of Gypsy prisoners.
- (b) We want to compare the proportion of prisoners who were intravenous drug users with the proportion of prisoners who had been re-imprisoned.
- (c) We want to compare the proportion of white prisoners who had TB with the proportion of Gypsy prisoners who had TB.
- (d) Of those prisoners who had TB, we want to compare the proportion who were intravenous drug users with the proportion who were HIV-positive.

3. Construct a 95% confidence interval for the difference between the proportion of White prisoners who were infected with TB and the proportion of Gypsy prisoners who were infected with TB. State what your interval tells you in plain English.

(a) State the parameter θ (using symbols and in words).

(b) State the estimate $\hat{\theta}$ (using symbols, in words and as a number).

(c) Calculate $se(\hat{\theta})$.

(d) Use the table for the Student's t -distribution to write down the value of the z -multiplier.

(e) Calculate the confidence interval.

(f) Interpret the confidence interval.

4. The standard error of the difference between the proportion of prisoners who have TB that are intravenous drug users and the proportion of prisoners who have TB that had been re-imprisoned is:

$$(1) \sqrt{\frac{0.6098(1-0.6098) + 0.4595(1-0.4595)}{592}}$$

$$(2) \sqrt{\frac{0.6098 + 0.4595 - (0.6098 - 0.4595)^2}{592}}$$

$$(3) \sqrt{\frac{0.3902 + 0.5405 - (0.6098 - 0.4595)^2}{592}}$$

$$(4) \sqrt{\frac{0.6098(1-0.6098)}{629} + \frac{0.4595(1-0.4595)}{456}}$$

$$(5) \sqrt{\frac{0.6098^2 + 0.4595^2}{592}}$$

5. 1997 Semester 2 Exam

A *Time/CNN* poll was based on a telephone survey of 800 adult Hong Kong residents conducted two weeks before the hand over of Hong Kong to China. p_c is the proportion of people in Hong Kong who think "Corruption" is the issue which worries them most, and p_f is the proportion of people in Hong Kong who think "Reduced personal freedoms" is the issue which worries them most.

A 95% confidence interval for $p_c - p_f$ is (0.012, 0.088). Which **one** of the following statements is **false**?

(1) In repeated sampling, we would expect that 95% of the 95% confidence intervals produced contain the true value of $p_c - p_f$.

(2) In light of the data, the interval (0.012, 0.088) contains the most plausible values for $p_c - p_f$.

(3) The true value of $p_c - p_f$ must be in the interval (0.012, 0.088).

(4) At this level of confidence, statements such as " p_c is bigger than p_f by somewhere between 0.012 and 0.088" are true, on average, 19 out of 20 times.

(5) With 95% confidence, the true value of $p_c - p_f$ is 0.05 with a margin of error of 0.038.