Introductory Statistics Tutorial Answers
Chapter 8 – Confidence Intervals

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

1. **(a)** $\theta = \mu$, the population mean mark for the 1995 528.188 exam.

(b) $\hat{\theta} = \bar{x}$ = 38.20, the mean mark of the sample of 30 marks.

(c) $\text{se}(\hat{\theta}) = \text{se}(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{10.85}{\sqrt{30}} = 1.9809$

(d) $df = 30 - 1 = 29$

(e) t-multiplier = 2.045

(f) 95% c.i. is: $\bar{x} \pm t \times \text{se}(\bar{x})$ = 38.20 ± 2.045 x 1.9809 = 38.20 ± 4.0509 = (34.15, 42.25)

(g) There are many ways of interpreting a confidence interval. Three different ways follow.

(1) With 95% confidence, we estimate that the population mean mark is somewhere between 34.15 and 42.25 marks.

(2) We estimate that the population mean mark is somewhere between 34.15 and 42.25 marks. A statement such as this is correct, on average, 19 times out of every 20 times we take such a sample.

(3) We estimate the population mean mark to be 38.20 with a margin or error of 4.05. A statement such as this is correct, on average, 19 times out of every 20 times such a sample is taken.

(h) We don't know. The population mean mark is not known so we don't know whether this particular 95% confidence interval contains the population mean. However, in the long run, the population mean will be contained in 95% of the 95% confidence intervals calculated from such samples.

2. **(a)** $\theta = p$, the proportion of female Spanish prisoners in 1995 who had tuberculosis.

(b) $\hat{\theta} = \hat{p} = \frac{36}{90} = 0.4$, the proportion in the sample of female Spanish prisoners who had tuberculosis.

(c) $\text{se}(\hat{\theta}) = \text{se}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.4 \times 0.6}{90}} = 0.051640$

(d) z-multiplier = 1.96

(e) 95% c.i. is: $\hat{p} \pm t \times \text{se}(\hat{p})$ = 0.4 ± 1.96 x 0.051640 = 0.4 ± 0.1012 = (0.299, 0.501)

(f) We estimate that the proportion of female Spanish prisoners in 1995 with tuberculosis is somewhere between 29.9% and 50.1%. A statement such as this is correct, on average, 19 times out of every 20 times we take such a sample.

3. (4)

4. (3)

5. (3)

Section B: Confidence interval for a difference in proportions

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

(b) Situation (a): *Two independent samples*

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

(d) Situation (a): *Two independent samples*

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

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

(c) Situation (a): *Two independent samples*

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

3. **(a)** Let p_1 represent the proportion of white prisoners who were infected with TB and p_2 represent the proportion of Gypsy prisoners who were infected with TB.
$\theta = p_1 - p_2$, the true difference in the above proportions.

(b) $\hat{\theta} = \hat{p}_1 - \hat{p}_2 = \frac{496}{886} - \frac{74}{152} = 0.5598 - 0.4868 = 0.0730$, the estimated difference in the above proportions.

(c) $\text{se}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{0.5598(1-0.5598)}{886} + \frac{0.4868(1-0.4868)}{152}} = 0.043837$

(d) $z = 1.96$

(e) 95% c.i. is: $(\hat{p}_1 - \hat{p}_2) \pm z \times \text{se}(\hat{p}_1 - \hat{p}_2)$ = 0.0730 ± 1.96 x 0.043837 = 0.0730 ± 0.08592 = (–0.0129, 0.1589)

(f) With 95% confidence, we estimate the proportion of white prisoners who were infected with TB to be somewhere between 0.013 lower and 0.159 higher than the proportion of Gypsy prisoners who were infected with TB.

4. (3) Note: This was Situation (c): *Single sample, two or more Yes/No items* **.**

5. (3)