Module 3 solutions

Module 3: Solutions to Learning Activities

Activity 3.1

An investigator wishes to study people living with agoraphobia (fear of open spaces). The investigator places an advertisement in a newspaper asking for volunteer participants. A total of 100 replies are received of which the investigator randomly selects 30. However, only 15 volunteers turn up for their interview.

1. Which of the following statements is true?

a) The final 15 participants are likely to be a representative sample of the population available to the investigator
b) The final 15 participants are likely to be a representative sample of the population of people with agoraphobia
c) The randomly selected 30 participants are likely to be a representative sample of people with agoraphobia who replied to the newspaper advertisement
d) None of the above

2. The basic problem confronted by the investigator is that:

a) The accessible population might be different from the target population
b) The sample has been chosen using an unethical method
c) The sample size was too small
d) It is difficult to obtain a sample of people with agoraphobia in a scientific way

Answers

Part 1

The correct answer is C. The only point at which random selection occurs is when the researcher selects 30 participants from the 100 replies. These 30 participants represent a random sample of the 100 replies.

A is not correct. The 15 who turned up for interview were most likely not similar to those who did not turn up for interview. Perhaps those who turned up had less severe disease, as they were more likely to leave their house and meet the research team.

B is not correct. As with A, the final 15 participants probably have less severe disease than the population living with agoraphobia.

Part 2

The correct answer is A.

B is not correct. Recruiting research participants through mass media is not unethical.

C is not necessarily correct. While 15 participants is a very small sample size, the main issue here is about recruiting a representative sample.

D is not necessarily correct. There may be valid ways of recruiting people living with agoraphobia. The researchers could work with anxiety support groups to codesign a method for enrolling participants that would increase the chance of recruiting a representative sample.

Activity 3.2

A dental epidemiologist wishes to estimate the mean weekly consumption of sweets among children of a given age in her area. After devising a method which enables her to determine the weekly consumption of sweets by a child, she conducted a pilot survey and found that the standard deviation of sweet consumption by the children per week is 85 gm (assuming this is the population standard deviation, σ). She considers taking a random sample for the main survey of:

- 25 children, or
- 100 children, or
- 625 children or
- 3,000 children.

a) Estimate the standard error and maximum likely (95% confidence) error of the sample mean for each of these four sample sizes.
b) What happens to the standard error as the sample size increases? What can you say about the precision of the sample mean as the sample size increases?

Answers

Part a

Table 1: Estimated standard errors of the mean and maximum likely errors for different sample sizes

Sample size	Standard error of the mean	Maximum likely error
25	17 gm	33 gm
100	9 gm	17 gm
625	3 gm	7 gm
3000	2 gm	3 gm

[Note: these results have been presented with no decimal places, the same precision as for the standard deviation.]

Part b

When the sample size increases, the standard error of the mean (and hence the maximum likely error) decreases. Thus, sample means from larger samples are more precise than from smaller samples.

Process

The standard error of the mean for a sample of 25 = 85/√25 = 17 gm, and the maximum likely error = 1.96 × 17 = 33.32 gm.

The standard error of the mean for a sample of 100 = 85/√100 = 8.5 gm, and the maximum likely error = 1.96 × 8.5 = 16.66 gm.

The standard error of the mean for a sample of 625 = 85/√625 = 3.4 gm, and the maximum likely error = 1.96 × 3.4 = 6.66 gm.

The standard error of the mean for a sample of 3,000 = 85/√3000 = 1.55 gm, and the maximum likely error = 1.96 × 1.55 = 3.04 gm.

Activity 3.3

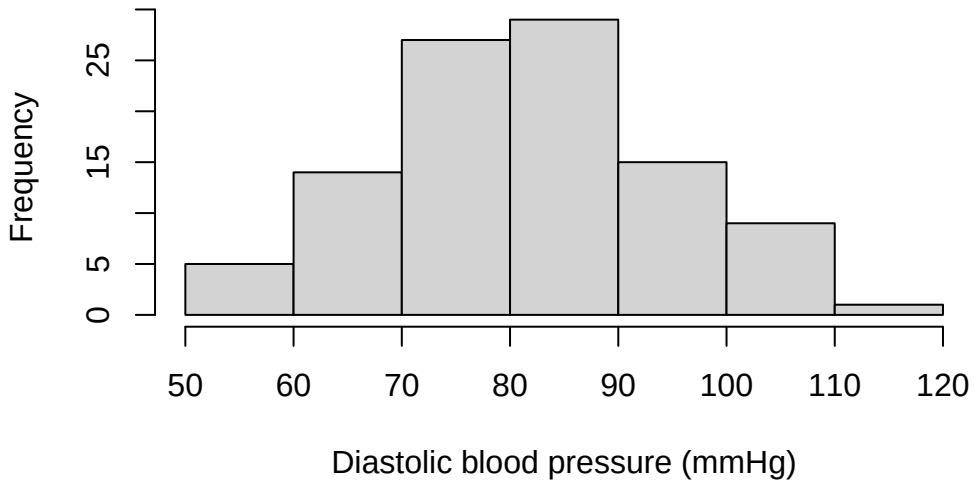
The dataset for this activity is the same as the one used in Activity 1.4 in Module 1. The file is `Activity1.4.dta` or `Activity1.4.rds` on Moodle.

a) Plot a histogram of diastolic BP and describe the distribution.
b) Use Stata or R to obtain an estimate of the mean, standard error of the mean and the 95% confidence interval for the mean diastolic blood pressure.
c) Interpret the 95% confidence interval for the mean diastolic blood pressure.

Answers

a) See Figure 1. The distribution is roughly symmetrical, centred about the mean.

Figure 1: Distribution of diastolic blood pressure from a sample of 100 participants



b) The sample mean is estimated as 82.2 mmHg, and the standard error (SE) of the mean is 1.30 mmHg. The 95% confidence interval is from 79.6 to 84.8 mmHg.

c) We are 95% confident that the true mean diastolic blood pressure of the population from which we sampled lies between 79.6 mmHg and 84.8 mmHg.

Process

a) The histogram can be constructed using the following code. Note the use of `/n` to indicate a new line when defining a figure title.

```
bp <- readRDS("data/activities/Activity_1.4.rds")

hist(bp$diabp,
     xlab="Diastolic blood pressure (mmHg)",
     main="Figure 1: Distribution of diastolic blood pressure /n
           from a sample of 100 participants")
```

b) The mean and its standard error can be obtained using the `descriptives` package, using the `se=TRUE` and `ci=TRUE` options:

```
library(jmv)

descriptives(data=bp, vars=diabp, se=TRUE, ci=TRUE)
```

```
DESCRIPTIVES

Descriptives

                              diabp

  N                               100
  Missing                           0
  Mean                       82.23000
  Std. error mean            1.301522
  95% CI mean lower bound    79.64750
  95% CI mean upper bound    84.81250
  Median                     83.00000
  Standard deviation         13.01522
  Minimum                    56.00000
  Maximum                    118.0000

  Note. The CI of the mean assumes
  sample means follow a
  t-distribution with N - 1 degrees
  of freedom
```

Activity 3.4

Suppose that a random sample of 81 newborn babies delivered in a hospital located in a poor neighbourhood during the last year had a mean birth weight of 2.7 kg and a standard deviation of 0.9 kg. Calculate the 95% confidence interval for the unknown population mean. Interpret the 95% confidence interval.

Answer

We are 95% confident that the true mean birthweight of babies born in the hospital located in a poor neighbourhood lies between 2.5 kg and 2.9 kg.

Process

As discussed in Section 3.9 of the course notes, R does not have a built-in function for calculating a confidence interval of a mean from summarised data. We can use the function as supplied in this section to perform the calculation.

First, we define the function by copy-and-pasting the code from the notes:

```
ci_mean <- function(n, mean, sd, width=0.95, digits=3){
  lcl <- mean - qt(p=(1 - (1-width)/2), df=n-1) * sd/sqrt(n)
  ucl <- mean + qt(p=(1 - (1-width)/2), df=n-1) * sd/sqrt(n)

  print(paste0(width*100, "%", " CI: ",
               format(round(lcl, digits=digits), nsmall = digits),
               " to ", format(round(ucl, digits=digits), nsmall = digits) ))

}
```

We then use the function by supplying the sample size (`n`), the estimated mean (`mean`) and the estimated standard deviation (`sd`). By default, this function will calcualte a 95% confidence interval, but this can be changed by supplying a different value for `width`. For example, to calculate a 90% confidence interval, we would define `width = 0.9`.

Putting all this together gives:

```
ci_mean(n=81, mean=2.7, sd=0.9)
```

```
[1] "95% CI: 2.501 to 2.899"
```

Rounding our confidence interval to the same precision as the mean given a confidence interval from 2.5kg to 2.9kg.