

MATH 203: PRINCIPLES OF STATISTICS 1
FALL 2024

ASSIGNMENT 4 – SOLUTIONS

1. For each of the studies presented below, define the population parameter of interest to the researchers and report a point estimate for this parameter.

(a) A 42-inch, large-screen, plasma TV is designed to consume only 350 watts and therefore produce less heat. In order to check this specification, a sample of plasma TVs was obtained and the power consumption was measured for each. The sample mean power consumption was 353.8 watts. 2 MARKS

(b) A sports physiologist is interested in the proportion of professional marathon runners who have very high haemoglobin counts. To this end, the researcher randomly selects 150 elite marathon runners, measures their haemoglobin levels and records it as either very high or normal. The researcher found that 48 subjects in the sample had very high haemoglobin counts. 2 MARKS

Solutions:

(a) The parameter of interest is the population mean power consumption, and the point estimate is 353.8 watts. 2

(b) The parameter of interest is the population proportion of high haemoglobin prevalence, and the point estimate is $48/150 = 0.32$. 2

2. In a study of student exam preparation, a sample of 411 undergraduate students preparing for a final exam in an introductory psychology course was collected. The sample mean time spent studying for the exam was 6.74 hours and the sample standard deviation of study times was 4.4 hours. It is reasonable to regard the sample as representative of the population of students taking introductory psychology at the university.

Construct a 95% confidence interval to estimate μ , the population mean time spent studying for the final exam for students taking introductory psychology at this university. Give a justification for the calculation you perform, discussing the assumptions made. 6 MARKS

Solutions: Here we have $n = 411$ large, so we can appeal to the Central Limit Theorem, and do not have to assume Normality. We are in Scenario 3, so the appropriate formula for the 95% CI is

$$\left(\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right)$$

with $\bar{x} = 6.74$, $s = 4.4$ and $z_{\alpha/2} = z_{0.025} = 1.96$. Hence the interval is

$$\left(6.74 - 1.96 \times \frac{4.4}{\sqrt{411}}, 6.74 + 1.96 \times \frac{4.4}{\sqrt{411}} \right) = (6.3146, 7.1654)$$

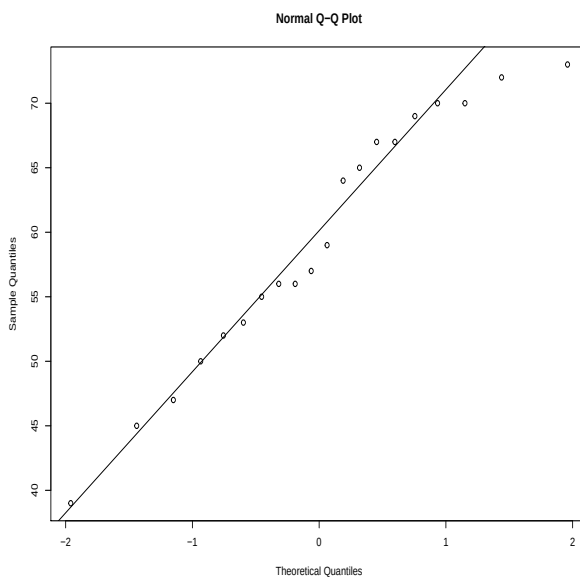
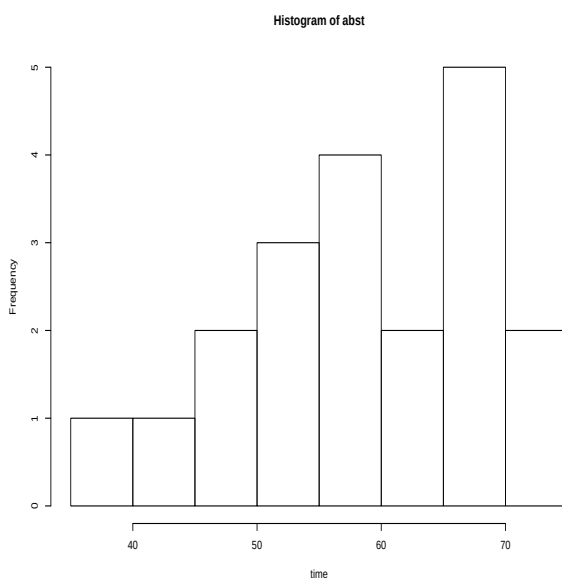
The only assumption needed is that the sample is a random sample from the target population. (4 marks for the interval, 2 marks for the assumptions) 6

3. A study investigated whether time perception, an indication of a person's ability to concentrate, is impaired during nicotine withdrawal. After a 24-hour smoking abstinence, 20 smokers were asked to estimate how much time had passed during a 45-second period. The sample mean perceived elapsed time was $\bar{x} = 49.3$ seconds and the sample standard deviation was $s = 9.14$ seconds.

Let μ denote the mean perceived elapsed time in the population of smokers who have abstained from smoking for 24 hours.

- (a) Find a 95% confidence interval for μ . Justify the calculation you carry out in terms of the assumptions made. 6 MARKS

- (b) The plots below depict the data collected in the sample. Do the plots support the assumptions made in (a)? Explain your answer with reference to each of the plots. 4 MARKS



Solutions:

- (a) Here we have $n = 20$ small, so we **must** assume Normality of the original data. We are in Scenario 1, so the appropriate formula for the 95% CI is

$$\left(\bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right)$$

with $\bar{x} = 49.3$, $s = 9.14$ and $t_{\alpha/2, n-1} = t_{0.025, 19} = 2.093$. Hence the interval is

$$\left(49.3 - 2.093 \times \frac{9.14}{\sqrt{20}}, 49.3 + 2.093 \times \frac{9.14}{\sqrt{20}} \right) = (45.0224, 53.5776)$$

The other assumption needed is that the sample is a random sample from the target population (4 marks for the interval, 2 marks for the assumptions). 6

- (b) The plots indicate that Normality is plausible, but not conclusively established;

- the histogram perhaps indicates skewness; however it is not definitive 2
- the QQ plot is more indicative of Normality, with perhaps a few upper-tail outliers 2

However, these are subjective assessments: it is also possible to argue that these plots vindicate the Normality assumption. Any reasoned argument over both plots merits marks.

4. *A cream with alpha-hydroxy acid, advertises that it can reduce wrinkles and improve the skin. In a study published in Archives of Dermatology (June 1996), 49 adults used a cream with alpha-hydroxy acid for 22 weeks. At the end of the study period, a dermatologist judged whether each participant exhibited any improvement in the condition of their skin: 31 showed improvement and 18 did not show improvement.*

- (a) Construct a 95% confidence interval for p , the true population proportion of adults with improved skin after using this cream. 6 MARKS
- (b) State the assumptions required for the confidence interval of part (a) to be valid, and report whether these assumptions are met for the study. 2 MARKS

Solutions:

- (a) Here we have $n = 49$ and $x = 31$ in a Bernoulli sampling setting. We are in Scenario 4, so the appropriate formula for the 95% CI is

$$\left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$$

with $\hat{p} = 31/49 = 0.6327$, $\sqrt{\hat{p}(1 - \hat{p})/n} = 0.0689$ and $z_{\alpha/2} = z_{0.025} = 1.96$. Hence the interval is

$$(0.6327 - 1.96 \times 0.0689, 0.6327 + 1.96 \times 0.0689) = (0.4977, 0.7676)$$

- (b) Here the key assumption is that n is large enough. We have $n\hat{p} = x = 31 > 15$ and $n(1 - \hat{p}) = n - x = 18 > 15$, so the large sample assumption is met. 6

5. A US car manufacturer advertises that its new model has a fuel consumption average of 47 miles per gallon. A consumer research group believes the average is lower than 47 miles per gallon.

(a) What is the parameter of interest, θ , for this study? 2 MARKS

(b) Write out the null hypothesis, H_0 , and alternative hypothesis, H_a , for testing the manufacturer's claim. Use mathematical notation as introduced in lectures. 2 MARKS

Solutions:

(a) The parameter of interest, θ , is the mean fuel consumption level (in miles per gallon) in the population of new models. 2

(b) Here

$$H_0 : \theta = 47$$

$$H_a : \theta < 47$$

where we may also use the typical notation with μ instead of θ . 2

6. A report states that in a certain country, 61% of high school graduates go to attend a two-year or four-year college the year after graduation. A group of researchers in certain state in this country would like to test whether the college-going rate for high school graduates in their state is different from the national figure.

(a) What is the parameter of interest, θ , for this study? 2 MARKS

(b) Write out the null hypothesis, H_0 , and alternative hypothesis, H_a , for testing the research question. Use mathematical notation as introduced in lectures. 2 MARKS

Solutions:

(a) The parameter of interest, θ , is the proportion of high school graduates who attend college in the year after graduation. 2

(b) Here

$$H_0 : \theta = 0.61$$

$$H_a : \theta \neq 0.61$$

where we may also use the typical notation with p instead of θ . 2

7. Consider a test of $H_0 : \mu = 4$, where μ is the population mean for a distribution from which a random sample is to be collected.

Give the rejection region, assuming that the test statistics follow a standard Normal distribution, for the following alternative hypotheses:

(a) $H_a : \mu > 4, \alpha = 0.1$ 3 MARKS

(b) $H_a : \mu \neq 4, \alpha = 0.05$ 3 MARKS

Give the rejection region, assuming that the test statistics follow a t -distribution and that the sample size used for the test was $n = 22$, for the following alternative hypotheses:

(c) $H_a : \mu < 4, \alpha = 0.1$ 3 MARKS

(d) $H_a : \mu \neq 4, \alpha = 0.05$ 3 MARKS

(e) Identify experimental conditions or assumptions under which the interval in (d) would be used. 2 MARKS

Solutions:

- (a) This is a one-sided, upper tailed Rejection Region derived from the Normal distribution, with $\alpha = 0.1$, so the R. R. is

$$\{T \geq z_\alpha\} = \{T \geq z_{0.1}\} = \{T \geq 1.2816\}$$

3

- (b) This is a two-sided, two tailed Rejection Region derived from the Normal distribution, with $\alpha = 0.05$, so the R. R. is

$$\{|T| \geq z_{\alpha/2}\} = \{|T| \geq z_{0.025}\} = \{|T| \geq 1.96\}$$

3

- (c) This is a one-sided, lower tailed Rejection Region derived from the t -distribution with $n - 1 = 21$ degrees of freedom, with $\alpha = 0.1$, so the R. R. is

$$\{T \leq -t_{\alpha, n-1}\} = \{T \leq -t_{0.1, 21}\} = \{T \leq -1.3232\}$$

3

- (d) This is a two-sided, two tailed Rejection Region derived from the t -distribution with $n - 1 = 21$ degrees of freedom, with $\alpha = 0.05$, so the R. R. is

$$\{|T| \geq t_{\alpha/2, n-1}\} = \{|T| \geq t_{0.025, 21}\} = \{|T| \geq 2.0796\}$$

3

- (e) Because n is small, and we are using the t -distribution, we would be using this interval when the data were a random sample from a Normal distribution, where we do not know the population mean or standard deviation (that is, Case 1 in the lecture notes).

2