620-270 Applied Statistics

Problem Set 2

2.1 A random sample of size 40 is selected from a population to test \( H_0 : \mu = 60 \) vs \( H_1 : \mu \neq 60 \). It can be assumed that \( \sigma = 5 \).

(a) If \( \bar{x} = 57.6 \), find the \( P \)-value.

(b) Is \( H_0 \) accepted or rejected at the 0.05 significance level? At the 0.01 significance level?
At the 0.001 significance level?

(c) If the alternative hypothesis is one-sided, i.e. \( H_1 : \mu < 60 \), what is the \( P \)-value?

2.2 The following 14 observations were made on a random variable which is normally distributed about \( \theta \):

36, 37, 64, 53, 40, 42, 56, 50, 30, 45, 33, 49, 57, 44

(a) Find a 95\% confidence interval for \( \theta \).

(b) Test the hypothesis that \( \theta = 40 \), at the 5\% significance level, against
i. a two-sided alternative
ii. the one-sided alternative \( \theta > 40 \).

2.3 A researcher has a theory that a particular disease is due to a high concentration of a substance \( S \) in the blood. To verify this theory the researcher conducts an experiment in which the concentration of \( S \) in the blood is measured for twenty individuals who have the disease and twenty individuals who do not have the disease. What should the null hypothesis be? What should the alternative hypothesis be? Explain your choices.

2.4 A random sample of 8 cigarettes of a certain brand has an average nicotine content of 4.2 mg and a s.d. of 1.4 mg. We want to know if this is in line with the manufacturer’s claim that the average nicotine content does not exceed 3.5 mg? To be accommodating to the manufacturer, we will use a 0.01 significance level. Think first about whether this is a one-sided or two-sided test.

(a) Use both the \( P \)-value approach and the rejection region approach to test the null hypothesis.

(b) What are your answers to (a) if it is known that the population s.d. is 1.6?

2.5 (a) Find the power of the test in Q2.4, if the true mean nicotine content is actually 4.0 mg, assuming that the population s.d. \( \sigma = 1.6 \).

(b) Is the magnitude of the power satisfactory, from a consumer’s point of view? Is it satisfactory from the manufacturer’s point of view? Explain.

(c) Explain how the power can be increased.

(d) Find the power if the significance level is changed to 0.05.

2.6 For the problem in Q2.4, suppose we want the test to have power 0.8 when the mean is 4.0, and significance level 0.01. How many cigarettes should be analysed? (Assume \( \sigma = 1.6 \).)

How many cigarettes should be analysed if the hypothesis test is two-sided?

2.7 A study is planned to estimate the mean concentration of a pollutant in stream sediments in a defined area at a given point in time. The sediments have been sampled extensively in the past and the variability in the measurements is known to be \( \sigma = 10 \) ppm. Study objectives require that enough measurements be taken so that a 95\% CI for the mean concentration has width no more than 4 ppm. How many measurements should be taken?
2.8 (for Assignment 1)
Consider the data in Q1.4 in Problem Set 1. Here again are the carbon monoxide (CO) levels for the 12 cars.

\[5.02 \ 5.79 \ 2.03 \ 4.62 \ 6.78 \ 8.60 \ 8.32 \ 6.07 \ 3.99 \ 5.22 \ 6.26 \ 11.13\]

Suppose we are concerned about whether the true mean vehicle emission level \( \mu \) is greater than 5 g/km. Assume that the CO levels are normally distributed with \( \sigma = 2.5 \).

(a) Formulate this as a hypothesis-testing problem, i.e. state the null and alternative hypotheses.

(b) Find the \( P \)-value. What is the conclusion, if the significance level is 0.05?

(c) Find a 95% confidence interval for \( \mu \) and confirm your conclusion.

(d) Perform the hypothesis test using a rejection region approach.

(e) How many vehicles should be tested if the width of the CI is to be 2 g/kg?

(f) What is the power of the above test if the actual mean is 7 g/km?

(g) How many vehicles should be tested if we want the test to have significance level 0.05 and power 0.90 for \( \mu = 7 \)? What is the rejection region for such a test (in terms of CO levels)?

2.9 (for Assignment 1)
As part of a study investigating the effect of smoking on infant birthweight, a physician examined the records of births to 40 non-smoking mothers, 40 light-smoking mothers, and 40 heavy-smoking mothers. The mean birthweights (in kg) for the three groups were respectively 3.43, 3.29 and 3.21.

(a) What is the response variable? What is the explanatory variable? What are the experimental units?

(b) For both the response variable and explanatory variable, state whether it is numerical or categorical. If numerical, state whether it is continuous or discrete, and if categorical, whether it is ordinal or nominal.

(c) Is this a designed experiment or an observational study? Explain your choice.

(d) Suggest two potential confounding variables in this study. Explain how you would eliminate the effect of one of these confounding variables.

(e) Assuming that birthweights in Australia are normally distributed with mean 3.4 kg and standard deviation 0.7 kg, test the hypothesis (at the 0.05 significance level) that light-smoking mothers give birth to lighter babies than average.

(f) Test the same hypothesis for heavy-smoking mothers.

(g) What is the power of this test when the mean is 3.1 kg?