

PL 9239: Assignment 2 (Resit)

Please write up all your answers in a document. For submission, the Learning Central portal only accepts .doc or .pdf documents. If you use R to calculate or visualise anything, please also include the R code in your document. Please make sure that you structure your document and R-code well so that it is easy to read.

p-Value

What does the p-value mean? Explain it with the help of an exemplary p-value of $p = 0.05$. (1 point, 2-3 sentences)

Formulating Hypotheses

In one of the labsheets, we have been working with the Welsh WIMD data and took a look at some interesting socioeconomic variables, for example the mental health rate, the chronic condition rate, the long term illness rate, income deprivation, employment deprivation, the proximity to green space, the time to travel to a post office, or the time to travel to a food shop

You have the impression that the propensity to vote for labour depends on a number of those factors and that it might be higher for some parts of the population. Given how you understand electoral dynamics, what might affect a citizen to cast their vote in favour of the labour party in Wales? Please formulate 4 hypothesis that you could investigate with a statistical test on the basis of survey data.

Please note that the points will only be awarded for the correct formulation of the hypothesis. (It does not matter whether the content actually makes sense). (1 point)

Central Limit Theorem

Imagine you toss a fair coin 10 times and record the number of tails that you observe. You repeat this experiment 50 times, each time recording the number of tails. How will these overall 50 results be distributed? And why is that the case? (1 point, 3-4 sentences)

Calculating Probabilities

Given the results from a (hypothetical) survey, you estimate that it takes the average citizen in Wales 7 minutes to get to the next post office. You also estimate that the standard deviation in the population is 2. You assume that the distribution follows a Normal distribution function and that you now know the true population sample.

- What is the probability to observe someone living at a distance larger than 11 minutes?
- What is the probability to observe someone living at a distance between 5 minutes and 7 minutes from the post office?

(Overall 1 point)

z-Scores

Take a look at this data:

Name	Distance to Post Office
Peter	10
Eve	12
John	18
Mary	21

- Calculate the z-scores for Eve and John using the data from the table. Please show how you are calculating the values (1 point)
- How can you interpret the results? What does the z-score mean? (1 point, 3-5 sentences)

Hypothesis Testing in Practice

Please load the data `assignment_2_data_resit.csv` into R. It is a (hypothetical) sample that contains the time it takes to travel to a post office in minutes. Test whether the true average travel time in the population is actually 6.

- Interpret the results from your test using a p-value. What does the p-value mean here? What do you conclude regarding your 0-hypothesis? (1 point, 3-4 sentences)
- Interpret the results from your test using a confidence interval. What does the confidence interval mean here? What do you conclude regarding your 0-hypothesis? (1 points, 3-4 sentences)
- Calculate a confidence interval that uses three standard errors instead the 1.96 standard errors that are the default in R. How can you interpret the result? What do you conclude? (2 points, 5-6 sentences)