

Assignment 1

Marketing Research

General Note for the following tasks

Please do the following tasks and copy the solutions into a word file, which you then please convert into a PDF and upload to Learn@WU. You have to be prepared to present your assignment solutions in the next session. In case you are selected to present a subtask of this assignment, please be prepared to answer questions regarding the theoretical concepts underlying the specific task, e.g., “What is the median? What is an interquartile distance?”

Task 1

Name two examples for each of the levels of measurement of data (Nominal, Ordinal, Interval, Ratio)

Do **not** use the examples on the slides!

Be prepared to argue *why* your example is measured at a particular level. E.g. ID numbers: Nominal - there is no ordering of ID numbers in the sense that one is better than the other. Example: Student ID 01234567 is not preferable to 12345678.

Task 2

Take a look the measures of dispersion and answer the following questions:

Imagine you have the following vector t containing music listening times (in hours) of 5 WU students in the past week:

$$t = [1, 2, 3, 4, 5]$$

1. What is the
 - a) range
 - b) interquartile range (choose the range such that it includes the least amount of values that cover at least 50% of the values in t)
 - c) variance
 - d) standard deviation
2. Which of the values in 1. change if the fifth student corrects her value to 10 (i.e. now $t = [1, 2, 3, 4, 10]$)?

Be prepared to give an intuition about the magnitude of the measures of dispersion. Consider extremes such as listening times that are the same for all students asked (e.g. $t_1 = [1, 1, 1, 1]$) or a sample with very different listening times (e.g. $t_2 = [0, 5, 25, 30]$)

Task 3

You are interested in the music listening behavior of current WU students.

Therefore, you go and ask 100 students about how many hours they listen to music every week. You calculate the mean and variance of the values you have gathered. Are the calculated values population or sample parameters? Why?

Task 4

Your friend challenges you to bet with him on a series of 5 coin flips. He is strangely insistent on picking heads all the time. Sure enough he wins 5 out of the 5 coin flips. So you decide to use what you have learned in Marketing Research and try to find out if the coin is biased. You flip the coin 100 times and get **heads** 90 times.

1. Calculate an estimate for the probability of observing **tails**. Do you think the coin is biased? If yes, what amount of heads (in 100 flips) would you expect from an unbiased coin?
2. Bonus (can make up a point lost elsewhere in the assignment but will not deduct points if unanswered/answered incorrectly): Assume your estimate is correct (i.e. the true probability of observing tails). What is the probability of your friend winning all 5 coin flips?

Task 5

Central Limit Theorem

Go to learn.wu.ac.at/shiny/imsm/clt_mr/, set the number of samples to 5000 and check the “Draw Normal overlay” box. This will draw an appropriate, theoretical normal distribution over the histogram of sample means of 5000 samples.

1. Choose “Normal” as the population distribution. This could be, for example, the distribution of potential net profits of a marketing campaign. What is the minimum sample size (adjust the “sample size” slider) required to convince you that the theoretical distribution of the sample mean is normal (i.e. the histogram in the lower graphs aligns approximately with the theoretical normal distribution). Hint: It is less than 1000. So look at the fit at 1000 to get a feeling for what “approximately” means in this context.
2. Now choose “Gamma” as the population distribution. The Gamma distribution could be used to model music listening times. It has two parameters “Shape” and “Rate”. Don’t worry about them at the moment. (You can have a look at how a characteristic sample [upper graph] changes when you change those parameters.) What is the minimum sample size required to convince you that the theoretical distribution of the sample mean is normal? Hint 1: This value should be larger than the one above. Hint 2: There is not a single correct answer, but be prepared to argue for your choice.

Task 6

Confidence Intervals

Stay on the website linked to above. Set the population distribution to “Gamma”, “Shape” to 2 and “Rate” to 5 and the number of samples again to 5000. The population mean is now 0.4. Check the box to draw the normal distribution overlay. Another box appears (“Show Confidence Intervals?”). Also check that box and set/leave the Confidence Level at 95%. The red-shaded area indicates being outside of the confidence interval. Set the sample size to something around 500 (± 20 is fine).

1. Are there samples with means outside of the confidence interval around the population mean?
2. Imagine you construct the confidence interval around each of the sample means. What percentage of times is the population mean captured within that interval?