

Portfolio Assessment

Introduction

The first coursework assignment for this module is a group portfolio of short tasks worth a cumulative **50% of the overall module mark**. You are asked to complete five tasks for the portfolio. All five tasks are based on real world data and require you to produce statistical analysis based on the concepts and theories we are learning throughout the term. We have now gone through enough material in the module so far, for you to be able to start working on the portfolio straightaway. There are still some topics that we will cover in the second part of the term and that will be needed for you to complete the portfolio. However, I would suggest that you get started with the portfolio and you work through the tasks over the coming weeks. For example, you could already complete tasks 1 and 2 (and possibly 3) and the remaining three over the following weeks. Do not leave the work until the last moment. Your overall performance is likely to greatly benefit from working regularly throughout the term. The portfolio should be submitted by **8th April 2022, 4pm** (London time).

Group or Individual Work

The portfolio assessment can be completed either as a group work or you can work on your own. I have a preference for you to work in a group because sharing ideas, thoughts, views, opinions helps to make your work richer and more creative. If you set up a group, I would like the group to include no more than three/four people and my expectation is that you will work together on the five tasks by actively collaborating, sharing the work and consulting each other on a regular basis. Each group should appoint a 'group leader' who will keep in regular contact with me and the class teachers and will inform us of the group progress. Any concern about the group's ability to work together should be communicated as soon as possible to me or the class teachers. Whether you decide to work with a group or on your own, please record your group names or just your individual name and ID on this online spreadsheet by **30th March 2022** (the link is also available on Moodle):

[EC1011 Portfolio Groups & Individuals.xlsx](#)

Introduction to the Portfolio Tasks

The portfolio is made up of five tasks that ask you to engage with some real world data drawn from the UK survey called '*Opinions and Lifestyle Survey*'. This survey is carried out monthly by the Office for National Statistics (ONS) and the data used for this portfolio refers to the period April – May 2015. The dataset is available on Moodle and is named "**EC1011 Portfolio Dataset.xlsx**".

A Brief Introduction about the Lifestyle Survey

The Opinions and Lifestyle Survey is a multipurpose social survey which provides quick and reliable information about topics of immediate interest. The survey has a monthly cycle and specialises in asking pertinent research questions on topics too brief to warrant full surveys of their own. Government organisations, academic institutions and charities can commission questions on the ONS Opinions Survey and use the survey in a variety of different ways. This is mostly to provide answers to questions of immediate policy interest but also to help assess public awareness of new policies, initiatives and publicity campaigns; to develop, test and pilot new survey questions; or to obtain samples of respondents for follow-up research projects.

Main Topics

Each month's questionnaire consists of two elements: core questions, covering demographic information, are asked each month together with non-core questions that vary from month to month.

The non-core questions for the April – May 2015 months were regarding *Well-Being (Module MCZ)*: this module was asked on behalf of ONS and all questions ask respondents to rate their feelings towards different aspects of their lives, generally using a scale of 0-10, where 0 is generally low and 10 high. You can find a list of the variables and their description in the appendix at the end of this document (page 7).

Coursework Brief

The portfolio is made up of five tasks. Please complete each task in separate worksheets in the same file that contains the dataset. Label each worksheet with the number of the task i.e. Task 1, Task 2 etc. It is important that for each question there is clear evidence of how the answer has been calculated i.e. there should be formulae, functions, explanations etc. that show how the solution has been computed. The comments should be typed in 'text boxes' inside the Excel document. Formulae/functions/graphs/tables do not count towards the word limit.

Task 1

This task requires you to deal with single and joint discrete probability distributions, the calculations of probabilities, functions of random variables and types of distributions. Complete this task in a worksheet labelled 'Task 1' and clearly show your solutions and, where required, add your comments. The comments should not exceed the **400 words limit**. Task 1 accounts for **20 marks**.

Task 2

This task requires you to engage with the insights of the Central Limit Theorem. Complete this task in a worksheet labelled 'Task 2' and clearly show your solutions and, where required, add your comments. The comments should not exceed the **200 words limit**. Task 2 accounts for **20 marks**.

Task 3

This task asks you to engage with point estimation and, in particular, to compute and compare two Point Estimators. Complete this task in a worksheet labelled 'Task 3' and clearly show your solutions and, where required, add your comments that should not exceed the **200 words limit**. Task 3 accounts for **20 marks**.

Task 4

This task asks you to engage with inferential statistics for the population mean with a particular focus on hypotheses concerning the level of 'happiness' across individuals who are in different employment status. Complete this task in a worksheet labelled 'Task 4' and clearly show your solutions and, where required, add your comments that should not exceed the **300 words limit**. Task 4 accounts for **20 marks**.

Task 5

This task asks you to engage with inferential statistics for the population proportion with a particular focus on hypotheses concerning the proportion of individuals with a 'degree or equivalent' level of education. The task also asks you to investigate differences in the proportion of women and men with a 'degree or equivalent' qualification. Complete this task in a worksheet labelled 'Task 5' and clearly show your solutions and, where required, add your comments that should not exceed the **300 words limit**. Task 5 accounts for **20 marks**.

Feedback

We will be happy to provide you with some feedback on one of the portfolio tasks during the term provided that any request reaches us by 30th March at the latest.

Submission Deadline

You are expected to submit one excel file that contains six worksheets: one worksheet with the actual data and one for each of the five tasks. Each worksheet should contain details of the solutions and comments wherever needed.

The portfolio coursework has one single submission point which is **Friday 8th March 2022, 4pm** (London time).

Task 1

- 1) Use the Excel functions to compute the average, median, mode, minimum, maximum, variance and standard deviation of the variable MCZ_1. Briefly comment on your findings.
- 2) The variable MCZ_1 measures the overall satisfaction of each respondent to the survey and we could regard it as a sort of measure of 'happiness'. Produce a relative frequency distribution of 'happiness' and plot an appropriate graph displaying the distribution.
- 3) Use the relative frequency distribution to:
 - a. Compute the expected rating. Comment on and interpret your findings.
 - b. Compute the variance and the standard deviation of rating. Do you get values similar to the ones you computed in 1)? Briefly comment on your findings.
- 4) Consider the two variables/events "MCZ_1" (happiness) and "QHealthr" (quality of health). Convert the "QHealthr" variable into a numerical variable where:

| Variable QHealthr | |
|----------------------|------------------------|
| <i>Ordinal Value</i> | <i>Numerical Value</i> |
| Very bad | 1 |
| Bad | 2 |
| Fair | 3 |
| Good | 4 |
| Very good | 5 |

Use the MCZ_1 variable and the new numerical 'QHealthr' variable to answer the following questions:

- a. Produce a joint probability distribution of 'happiness' and 'quality of health';
- b. What is the probability that a randomly selected individual shows 'bad' health?
- c. What is the probability that a randomly selected individual shows happiness of 7 and fair quality of health?
- d. What is the probability that a randomly selected individual shows happiness of 8 or good quality of health?
- e. What is the probability that a randomly selected individual shows happiness of 9 given that their quality of health is very good?
- f. Are happiness (MCZ_1) and quality of health (QHealthr) independent events? Explain.
- g. Is there a relationship between happiness and health? Use the Excel functions to compute the covariance and correlation coefficients between happiness (MCZ_1) and quality of health (QHealthr). Briefly comment on your findings.
- h. Use the joint probability distribution from a) to compute the covariance and the correlation coefficient. Compare your results with those obtained in g) and briefly comment.
- i. The government is concerned both about the 'happiness' of citizens and the quality of their health according to the following welfare function:

$$Welfare = 1.5Happiness + 2Quality\ of\ Health$$

Compute the 1) expected value and 2) standard deviation of welfare. Briefly comment on your findings.

- j. If the government wanted to maximise welfare, should the government change the weights associated with Happiness (1.5) and Quality of Health (2)? Briefly explain.

- 5) Define the proportion of individuals who reported a happiness of 8 or more as the probability that an individual is 'happy'. Suppose that a researcher is about to interview 30 individuals about their level of happiness.
- Use the appropriate Excel function to produce the probability distribution of the random variable "Individual is happy".
 - Plot the probability distribution graph. Briefly comment on the graph.
 - What is the probability that more than 8 individuals will declare to be "happy"?
 - What is the probability that between 12 and 20 individuals will declare to be "happy"?
 - Use the appropriate Excel function to work out the number of individuals for which the cumulative distribution is greater than or equal to any given probability value.

Task 2

In this task we focus on the variable MCZ_8 that lists the respondents' answers to the question "Overall, how satisfied are you with your mental wellbeing?" (0 = not satisfied at all; 10 = fully satisfied).

- Compute some summary statistics for the MCZ_8 variable and briefly comment on your findings.
- Draw thirty random samples of size 10 from the MCZ_8 variable and, for each sample, compute the sample mean. Compute the mean of the sample means and produce a histogram of the distribution of the sample means. Briefly comment on your findings.
- Now, draw thirty random samples of size 40 from the MCZ_8 variable and, for each sample, compute the sample mean. Compute the mean of the sample means and produce a histogram of the distribution of the sample means. Compare the mean of the sample means and the distribution of the sample means with those you computed in 2). Explain the similarity or differences.
- Draw a random sample of 40 observations from the MCZ_8 variable and:
 - Compute the sample mean.
 - Use the information from the random sample to compute the probability that the sample mean is smaller than 7.5. Briefly comment on your findings.

Task 3

For this task you should focus on the variable MCZ_17 that contains the respondents' answers to the question "Overall, how satisfied are you with the area where you live?". Suppose that the population average satisfaction with the area individuals live in is equal to 7.9 but this average is not known to the researchers. You are hired to check which estimator might be best in estimating the population average level of satisfaction with the area in which they live. To this end you are asked to complete the tasks below.

- We are interested in comparing the properties of two estimators (\bar{X}_1 and \bar{X}_2) of the population mean. To this end, complete the following tasks:
 - Draw 30 random samples of size $n = 100$ from the variable MCZ_17.
 - For each of the 30 samples drawn, calculate:

- The sample mean $\bar{X}_1 = \frac{\sum_{i=1}^{100} X_i}{100}$ where X_i are all 100 observations of MCZ_17 in the sample ($i = 1, 2, \dots, 100$);
- An alternative estimator $\bar{X}_2 = \frac{\sum_{i=1}^{20} X_i}{20}$ where X_i are the first 20 observations of MCZ_17 in the sample ($i = 1, 2, \dots, 20$).

c. Present your calculations in a table:

- The first column of this table should contain the 30 values of \bar{X}_1 you have obtained.
 - The second column of this table should contain the 30 values of \bar{X}_2 you have obtained.
2. Compute the (approximate) bias of the estimator \bar{X}_1 and the (approximate) bias of the estimator \bar{X}_2 . Is one of the two estimators more biased than the other? Explain why or why not.
 3. Compute the (approximate) variance of the estimator \bar{X}_1 and the (approximate) variance of the estimator \bar{X}_2 . Compute the relative efficiency of the two estimators and interpret it. Is one of the two estimators more efficient than the other?
 4. Overall, what estimator would you recommend the researchers use in order to estimate the population average answer to the question "Overall, how satisfied are you with the area where you live?"

Task 4

The government is interested in understanding whether there are differences in the average level of happiness (question MCZ_1, "Overall, how satisfied are you with your life nowadays?") among respondents' 'employment status': 'In Employment' or 'Economically Inactive' or 'ILO Unemployed' (variable DVILO3ar). To this end, you are asked to complete the following tasks.

- 1) Suggest one unbiased estimator that researchers can use to estimate the average level of satisfaction for each of the three 'employment status' categories. Use the sample data and your suggested estimators to compute the estimates and comment on your findings.
- 2) Produce three confidence intervals, one for each type of 'employment status', that have a high probability of capturing the true average 'happiness' of the three groups of individuals. Plot the confidence intervals and comment on your findings.
- 3) The government believes that individuals who are in work show a greater level of happiness than individuals who are either unemployed or economically inactive. You are asked to test the government hypotheses by completing the following tasks:
 - a. The government believes that the average level of happiness among the employed is 8. The unions disagree with the government hypothesis and they claim that the average happiness is different. Test this hypothesis and comment on your findings.
 - b. The government also believes that the average level of happiness among the unemployed is 7.5. The unions, again, strongly disagree and they believe that the true average level of happiness is smaller than the one claimed by the government. Test this hypothesis and comment on your findings.
 - c. Finally, the government claims that the average level of happiness among the economically inactive individuals is 7.4. Again, the unions disagree with the government and they claim that the true average level of happiness is actually greater. Test this hypothesis and comment on your findings.

Task 5

The government would like to use the survey data in order to understand the proportion of the UK population that has got a degree or equivalent. The government is also interested in understanding whether there are differences in educational achievements between men and women. To this end, you are asked to use the information provided by the variable “HighEd4r” in order to carry out the following analysis.

- 1) Complete the following tasks to estimate the population proportion of individuals with a ‘degree or equivalent’.
 - a. Suggest an unbiased estimator that could be used to estimate the population proportion of individuals with a ‘degree or equivalent’. Use the sample data in order to provide an estimate. Briefly comment on your choice of estimator and on your estimate.
 - b. Compute and plot a confidence interval that has a high probability of including the population proportion of individuals with a ‘degree or equivalent’. Briefly comment on your findings.
 - c. The government believes that 30% of the population has got a ‘degree or equivalent’ level of education. Use the sample data in order to test this hypothesis. Briefly comment on your findings.
- 2) Complete the following tasks to estimate the proportion of female and male individuals with a ‘degree or equivalent’ level of education.
 - a. Suggest two unbiased estimators that could be used to estimate: 1) the proportion of women with a ‘degree or equivalent’ and 2) the proportion of men with a ‘degree or equivalent’. Use your estimators and the sample data in order to provide estimates of the two population proportions.
 - b. Compute two confidence intervals that have a high probability of including 1) the population proportion of women and 2) the population proportion of men with a ‘degree or equivalent’ level of education. Plot the two confidence intervals and comment on your findings.
 - c. The government believes that the proportion of women and men achieving a ‘degree or equivalent’ level of education is identical and equal to 30%. Many researchers disagree. Some researchers believe that 1) the proportion of women with a ‘degree or equivalent’ level of education is smaller than 30% and 2) the proportion of men with a ‘degree or equivalent’ level of education is greater than 30%. Test these two hypotheses and comment on your findings.

Appendix – List of variables in the dataset

| Variable Name | Description | Type |
|------------------------|---|---------|
| Casenew | New random ID number | Scalar |
| MCZ_1 | Overall, how satisfied are you with your life nowadays? | Scalar |
| MCZ_2 | Overall, to what extent feel things you do in your life are worthwhile? | Scalar |
| MCZ_3 | Overall, how happy did you feel yesterday? | Scalar |
| MCZ_4 | Overall, how anxious did you feel yesterday? | Scalar |
| MCZ_5 | Overall, how satisfied are you with your relationships with family, including spouse/partner? | Scalar |
| MCZ_7 | Overall, how satisfied are you with your physical health? | Scalar |
| MCZ_8 | Overall, how satisfied are you with your mental wellbeing? | Scalar |
| MCZ_11 | Overall, how satisfied are you with your work situation? | Scalar |
| MCZ_10 | Overall, how satisfied are you with your financial situation? | Scalar |
| MCZ_17 | Overall, how satisfied are you with the area where you live? When answering, please consider the area to be within 15 to 20 minutes distance from your home | Scalar |
| MCZ_13 | Overall, how satisfied with amount of time have to do things like doing? | Scalar |
| MCZ_9 | Overall, how satisfied are you with the wellbeing of your child/children? | Scalar |
| QHealthr* | How is your health in general? | Ordinal |
| RSEX | Sex of respondent | Nominal |
| AGEXR* | Grouped age | Ordinal |
| Marstat3r ⁹ | Marital status – 3 categories (recoded) | Nominal |
| HighEd4r* | What is the highest level of qualification? | Nominal |
| Ethnicity2r | Ethnicity White/Other (recoded) | Nominal |
| DVIL03ar* | DV for International Labour Organisation (ILO) in employment | Nominal |
| FtPtWkr* | Full-time or part-time work | Nominal |
| NSECAC3 | NS-SEC 3 categories | Nominal |
| GorA | Government Office Region | Nominal |