

MATH 1065 — Quantitative Methods in Health

Statistical Analysis Report (SP5 2021)

DUE by 11:00pm, Friday 29 October

- This assignment is worth 20% of your final grade. It is due **no later than 11 pm on Friday the 29th of October in Week 12**.
- You will need to submit your assignment via learnonline. Marked assignments will be returned to you electronically.
- The file you submit needs to be in a **pdf** format and prepared **using the template provided**.
- Relevant MINITAB output should be copied and pasted into Word as a picture (Windows) or pdf (Mac). You may use Print Screen but check that the quality is not compromised, i.e. image does not become fuzzy.
- For full marks, ensure that appropriate axis labels, meaningful titles and legends are included with all graphical displays.
- **Failure to follow the template, poor communication or a messy layout will attract a penalty of up to 10 marks (10% of maximum marks available).**
- **Any late submission will attract a penalty of 10 marks (10% of maximum marks available) per working day, or part thereof, the assignment is late. The cut-off time is 11pm each day. Weekend counts as one day.**
- An example of a 'good' and 'bad' report is available from the course website.
- A suggested schedule to work through this assignment:

Week	Suggested tasks to be completed
Week 9	<ul style="list-style-type: none">• Download the assignment instructions and the submission template; read through them to get an idea of what is required.• Download the data file and make sure that it can be opened from the computer you will use to work on this assignment.• Match questions to weekly topics. They are not necessarily in the same order!• Work on Question 1.
10	<ul style="list-style-type: none">• Work on Questions 2, 3 and 4.
11	<ul style="list-style-type: none">• Finalise your answers to Questions 1 to 4.• Write and check the report – it should be consistent with results from Questions 1 to 4.
12	<ul style="list-style-type: none">• Check the entire submission for completeness and adherence to the template, and submit 😊

- Major components of the assignment are as follows:

Assignment component	Mark
Statistical Analysis (Questions 1 to 4)	70
Statistical Analysis Report	30
TOTAL	100



Let them eat fruit!

The effect of fruit and vegetable consumption on well-being

The health benefits of fruit and vegetables are well established. Studies have shown that people who consume more fruit and vegetables have better cardiovascular health, reduced risk of some cancers and greater longevity. There is also growing evidence that people who eat more fruit and vegetables have better mental health, including lower incidence of depression and anxiety, higher life satisfaction and greater socio-emotional well-being. This suggests that the importance of what food we eat goes beyond the already established physical health benefits.

A recent study¹ tested the psychological benefits of a 14-day clinical intervention designed to increase fruit and vegetable (FV) consumption among low FV consuming adults, identified as anyone aged 18 to 25 who is eating fewer than 3 combined serves of fruit and vegetables per day. Study participants were 171 students at the University of Otago, New Zealand. There were 56 men and 115 women, with majority of participants (64%) identifying themselves as of European descent. Recruited volunteers were included in the study if they were young adults (18-25 in age), had an internet enabled smartphone, were low FV consuming (no more than 3 combined serves of FV per day), without any known FV allergies and not on any anti-depressant medication.

Once formally enrolled in the study, students provided written consent to participate and were randomly assigned to three intervention conditions. The *control* condition was diet as usual. Those assigned to the *ecological momentary intervention* (EMI) condition were given a challenge to increase their FV consumption to at least 5 serves a day. They received a voucher for a local green grocer and were explicitly told to use this voucher to assist in the challenge. These participants also received two text-messages a day at targeted times for maximum impact (e.g., before typical meal times to influence food choices). Finally, participants assigned to the *fruit and vegetable intervention* (FVI) condition were supplied with a bag with enough fruit and vegetables for one serve of each for every day of the intervention. It was a mixture of large carrots, apples, oranges and kiwi fruit. Fruit and vegetables were selected based on freshness, seasonal availability, and their vitamin C (oranges, kiwi fruit) and carotenoids content (carrots). These participants were asked to consume one additional serve of fruit and one additional serve of vegetables every day, in addition to their normal consumption.

Answer the questions and prepare a report where you describe your findings and draw appropriate conclusions. Follow instructions provided in this document. The data file for this assignment is called *FruitVeg.xlsx*. It can be downloaded from the Data tab on the course website.

You are not required to consult additional sources (e.g. internet articles or scientific papers) but if you do, ensure you include a reference list and cite them in text appropriately.

¹Conner TS, Brookie KL, Carr AC, Mainvil LA, Vissers MCM (2017). Let them eat fruit! The effect of fruit and vegetable consumption on psychological well-being in young adults: A randomized controlled trial. PLoS ONE 12(2): e0171206.doi: 10.1371/journal.pone.0171206

Question 1 (20 marks)

Are there any significant differences in the composition of study groups? Random assignment to treatments can sometimes lead to an unbalanced split of subjects. We want to check this using *Gender*, *Ancestry* and *BMI*. Answer the questions that follow. In each case, you will need to decide on the most appropriate hypothesis test to perform. Variable descriptions are given in Appendix A, Table A.1. It will help you to first decide which variables are categorical and which are numerical. The decision tree from Week 7 workshop called 'Which statistical test?' will also help.

- (a) **(6 marks)** Is there is a statistically significant relationship between the variables *Gender* and *Intervention*? Formulate and perform an appropriate hypothesis test at a 5% significance level. Use Minitab to obtain the test statistic and the *P*-value. For full marks, include appropriate Minitab output and check the requirements. Use the STATE-FORMULATE-SOLVE-CONCLUDE procedure and perform follow-up analysis if appropriate.
- (b) **(8 marks)** Now test to see if there is a statistically significant relationship between *Ancestry* and *Intervention*. Formulate and perform an appropriate hypothesis test at a 5% significance level. Use Minitab to obtain the test statistic and the *P*-value. For full marks, include appropriate Minitab output and check the requirements. Use the STATE-FORMULATE-SOLVE-CONCLUDE procedure and perform follow-up analysis if appropriate.
- (c) **(6 marks)** Are there statistically significant differences in BMI among the three groups? Formulate and perform an appropriate hypothesis test at a 5% significance level. Use Minitab to obtain the test statistic and the *P*-value. For full marks, include appropriate Minitab output and check the requirements. Use the STATE-FORMULATE-SOLVE-CONCLUDE procedure and perform follow-up analysis if appropriate.

Question 2 (16 marks)

Have interventions been effective? Descriptive statistics for participants' baseline consumption of fruit and vegetables are given in Table 1. Information about fruit and vegetable consumption during the intervention period is stored in variable *FV_daily*. See variable descriptions in Table A.1 for more details.

	Control (<i>n</i> = 59)	EMI (<i>n</i> = 55)	FVI (<i>n</i> = 57)
Baseline	2.5 (1.2)	2.6 (1.1)	2.5 (0.9)

Table 1. Fruit and vegetable consumption at baseline. Values are reported as Mean (Std Dev).

- (a) **(6 marks)** Use Minitab .to obtain boxplots that show daily fruit and vegetable consumption by type of intervention. Comment on the effectiveness of the interventions based on Table 1 and what you conclude from your boxplots. Have any study participants reached the recommended five serves or more of fruit and vegetables per day? How do you know? Explain briefly.
- (b) **(10 marks)** Vitamin C and carotenoids are the two most responsive micronutrient biomarkers of fruit and vegetable consumption. In this study, participants' plasma vitamin C and total plasma carotenoids were assessed using fasting blood samples collected at baseline and at conclusion of the intervention period. Do blood test results confirm increases in fruit and vegetable consumption reported by participants? Formulate and perform an appropriate hypothesis test for each intervention group separately using variables *VitC_change* and *Carot_change* (refer to Table A.1 for variable descriptions). Use Minitab to obtain test statistics and *P*-values (one for each group). For full marks, include appropriate Minitab output and check the requirements. Do not use the full STATE-FORMULATE-SOLVE-CONCLUDE procedure. Instead, state the hypotheses, report on requirement checks, then give your decision and conclusion at a 5% significance level.

Hint: You will need to unstack the data. Once the data is unstacked, you will be able to get multiple test results at once, e.g. vitamin C results for all three interventions in one go.

Question 3 (12 marks)

Which intervention had the most impact on feelings of vitality? The daily survey the participants completed using their smartphones included questions relating to how energetic or fatigued they felt. Participants were asked: 'Did you feel full of life today? Did you have a lot of energy? Did you feel worn out? Did you feel tired?' and answered on a 6-point scale that ranged from 'none of the time' to 'all of the time'. Responses were then combined into an overall 'vitality' score. Answer the questions that follow.

- (a) **(4 marks)** Use Minitab and variable *Vitality_change* to produce an error diagram of the change in vitality score by type of intervention. Comment briefly on how the mean changes in vitality scores compare across groups and whether you expect to find any statistically significant differences. Refer to variable descriptions in Table A.1. Instructions on how to produce an error diagram are given in the Minitab Booklet.
- (b) **(8 marks)** Is there a significant difference in mean change in vitality scores based on type of intervention? Formulate and perform an appropriate hypothesis test at a 5% significance level. Use Minitab to obtain the test statistic and the *P*-value. For full marks, include appropriate Minitab output and check the requirements. Use the STATE-FORMULATE-SOLVE-CONCLUDE procedure.

Question 4 (22 marks)

What is the relationship between vitality and depressive symptoms? Study participants also answered questions designed to capture depressive symptoms, on a daily basis. Participants' overall, post-intervention period scores are represented by variables *CESD* and *Vitality*. See variable descriptions are in Table A.1 for more information. Answer the questions that follow.

- (a) **(5 marks)** Now use Minitab to obtain a scatterplot with *Vitality* as the independent variable (*x*) and *CESD* as the dependent variable (*y*). Also obtain the Pearson correlation coefficient and the corresponding *P*-value. Does there appear to be a linear relationship? Is it positive or negative? How strong is it? Does it make sense to fit a simple linear regression model in this case? Justify your answers briefly.
- (b) **(2 marks)** Use Minitab to fit a simple linear regression model that represents the relationship between *CESD* and *Vitality*. Show (but do not interpret) the full regression model output from the Minitab session window (tables) and the corresponding residual plots.
- (c) **(6 marks)** Are conditions for linear regression satisfied? Answer in terms of *Linearity*, *Independence*, *Normality* and *Population standard deviations*. Use residual plots from part (b).
- (d) **(3 marks)** Comment on the strength of the relationship between *CESD* and *Vitality* using the coefficient of determination. What is its value? What precisely does it measure in this scenario? Use the regression model output from part (b).
- (e) **(2 marks)** What is the value of the slope? What does it measure in this scenario? Use the regression model output from part (b).
- (f) **(4 marks)** Is the relationship between *CESD* and *Vitality* statistically significant? In other words, is the slope estimate statistically significant at a 5% significance level? How do you know? Explain briefly. Use the regression model output from part (b).

Statistical Analysis Report (30 marks)

Your report should consist of sections described below.

Introduction (4 marks)

Provide the context and rationale for the study. Use your own words! *There is no word limit, just ensure you have explained what the report will contain. As a guideline, one paragraph will be sufficient.*

Methods (6 marks)

Discuss the methods used to collect and analyse data from this study:

- What type of study was conducted? Name the study design and briefly describe, in your own words, the interventions that were part of this study.
- Describe the sample (including the sample size and any demographic information, e.g. who the study participants were, their age, gender split etc).
- Briefly describe variables that you have analysed.
- Provide a list of statistical displays and procedures that you have used.

There is no word limit. As a guideline, one paragraph for this section will be sufficient.

Results & Discussion (16 marks)

First, summarise the main results of your analyses from Questions 1 to 4. You may use subsections, tables etc. as you see fit. Present and discuss results in a clear and simple way:

- Present findings of statistical analyses in a logical sequence. Descriptive statistics about variables of interest are usually presented first, followed by the results of further statistical analyses.
- Include copies of key diagrams from Questions 1 to 4 as relevant to your presentation of results. Useful diagrams to include in a report are histograms, boxplots, error diagrams and scatterplots.
- State each result and the corresponding statistical procedure, and report *P*-values to three decimal places. However, do not include numerical calculations or full details of statistical procedures and condition checking (e.g. full Minitab output).

Next, interpret your statistical findings by discussing their practical significance. Use plain language; there should be no technical details or statistical terminology. Are any of the results surprising in any way? Ensure you address the following within one short paragraph:

- Did participants end up eating more fruit and vegetables?
- Which intervention proved to be effective and in what way?
- Were there any significant improvements to wellbeing of participants and in what sense?

Finally, in another short paragraph indicate shortcomings, if any, of the study design and analyses that were performed. Are there any issues with *internal* and *external validity* of this study?

There is no word limit. As a guideline, one and a half pages (two pages at most) will be sufficient for this section, including any tables and graphs. Remember, marks are awarded for quality not quantity!

Conclusion (4 marks)

What can you conclude overall about the effects of increased fruit and vegetable consumption on psychological well-being of young adults, and the effectiveness of proposed interventions?

There is no word limit. As a guideline, one paragraph will be sufficient. Do not introduce any new information in this section, and do not simply repeat statements made elsewhere in your report!

Appendix A. Data file and variable descriptions

Some of the data from the study is stored in the file called *FruitVeg.xlsx* that can be downloaded from the **Data** tab within the course website. Below are descriptions of variables in that data file:

Name	Description
<i>ID</i>	Study participant ID
<i>Gender</i>	Two categories: <i>Male</i> or <i>Female</i>
<i>Intervention</i>	Type of intervention: <i>Control</i> = diet as usual and a voucher for a local green grocer received at the end of the study; <i>EMI</i> = Ecological momentary intervention with twice-daily text messages plus a voucher for a local green grocer; <i>FVI</i> = Fruit and vegetable intervention with a bag containing two weeks' worth of extra fruit and vegetables.
<i>Ancestry</i>	Two categories: <i>European</i> or <i>Other</i>
<i>BMI</i>	Body Mass Index (kg/m ²)
<i>FV_daily</i>	Mean number of fruits and vegetables (combined) per day, based on the information recorded in participants' diaries each day during the two-week intervention period. Participants were required to use their smartphones to record how many fruits and vegetables they consumed each day.
<i>VitC_Change</i>	Change in plasma Vitamin C level from baseline to post intervention period (μM), calculated as post intervention minus baseline.
<i>Carot_Change</i>	Change in total plasma carotenoids (beta-carotene equivalents) from baseline to post intervention period, calculated as post intervention minus baseline. Measurements are mmol/L of β -carotene equivalents.
<i>Vitality</i>	Total vitality score based on the energy/fatigue scale taken from the <i>Rand 36-Item Short Form Survey</i> (SF-36). Post-intervention scores with range from 0 to 100. The higher the score, the greater the feeling of vitality.
<i>Vitality_change</i>	Difference in participants' vitality scores, calculated as the score post intervention minus the baseline score.
<i>CESD</i>	Total depressive symptom score based on responses to the <i>Centre for Epidemiological Studies Depression Scale</i> (CES-D) questionnaire. Post-intervention scores with range from 0 to 100. The higher the score, the worse the depressive symptoms.

Table A.1. Descriptions of variables in *FruitVeg.xlsx*