

ECAS7001 - Applied Research Design and Analysis

Practical Skills 2 – **PRACTICE WORK SHEET WITH CORRECT ANSWERS**

Task One: Estimating required sample size (10 marks)

Read the information below about an applied research study and answer the questions.

Healthy Eating Intervention Study:

Study Objective: to compare the effectiveness of a healthy eating intervention in improving self-reported intention to eat healthily in patients diagnosed with clinical obesity.

Study Design: a randomised controlled **between-subjects** trial in which clinically obese patients **aged 18-40** are randomly allocated to either a daily motivational text-messaging intervention over a six week period or to receive standard treatment (attendance at a weekly support group).

Main outcome measure for determining sample size: self-reported intention to eat healthy (rated on mean scores to a validated intentions to eat healthy questionnaire, scale range 0-30).

In order to conduct this study, the researchers wish to establish how many participants are needed in order to have a **90% power** of detecting a clinically significant difference in intentions to eat healthy.

The clinicians involved in this research project consider a difference in intentions to eat healthy score of **9 points** between the two groups at the 5% level of significance as being a **clinically** significant difference.

Based on previous literature, the researchers are expecting the **standard deviation** of intention to eat healthy in those in the intervention group to be **17.99**.

Sample size Questions:

1. In the space below, state the formula that should be used to calculate sample size in Altman's nomogram for the above study (2 marks).

Independent samples t-test

$$\frac{\delta}{\sigma}$$

2. Using the correct formula for Altman's Nomogram, calculate the standardised difference for the above study and then use Altman's nomogram to estimate the required sample size – show your workings below (4 marks).

$\delta = 9$ scale points and $\sigma = 17.99$ intention to eat healthy standard deviation.

Therefore, $9/17.99 = 0.50$.

On Altman's Nomogram, the line connecting a standardized difference of 0.50 and a power of 90% cuts the sample axis on the 0.05 alpha level at approximately 160.

160 divided by the number of groups (2) = $160/2 = 80$.

3. Write out your Power Statement below (4 marks).

“Eighty participants would be required in each group (total of 160) to have a 90% chance of detecting a clinically significant difference in means of 9 scale points (SD=17.99) at the 5% level of significance using the Independent-samples t-test.”

Task Two: Data screening and cleaning (10 marks).

Using the SPSS datasheet and information on study design and measures provided on Moodle, you are required to use SPSS in order to demonstrate your ability to carry out the following tasks.

Using the SPSS datasheet titled “**Screening datasheet**” from the above (hypothetical) study you should:

1. Screen the data for any errors within the **categorical** variables data, correct those errors in the SPSS input to “99” and write in the space below details of those errors (1 mark)

Group, Gender, and Marital Status should have been screened here as the categorical variables.

Errors: Participant ID 321 had a gender score of 222 – gender could only be classed between one and four in this study.

2. Screen the data for any errors within the **continuous** variables data, correct those errors in the SPSS input to “99” and write in the space below details of those errors (1 mark)

Age and Intention to Eat Healthy score at time two should have been screened here as the continuous variables.

Errors:

Participant ID 225 had an age of 5436

Apart from being exceptionally high, based on the study information above, this randomised control trial was only conducted on participants aged between 18-40 years. Therefore, this age must be an error and was changed to 99 Missing data.

Participant ID 114 had an intention to eat healthy score of 274

Based on the study information provided above, the self-reported intention to eat healthy was rated on a validated intentions to eat healthy questionnaire, with a scale range from 0-30. Therefore, the above score must be an error as it falls outside of this range and was consequently changed to 99 Missing data.

3. Run descriptives on both the categorical and continuous data variables to check for any missing data, and write in the space below the **ID number(s)** of the participant(s) and **the variable(s)** in which there is missing data (2 marks)

From the output, in terms of categorical descriptives, there is one missing data for the gender variable for participant ID 321.

From the output, in terms of continuous descriptives, there is one missing data for the age variable for participant ID 225. There are also ten missing data for the intentions to eat healthy score at time two. One of these came from an error [ID 114], while the other nine chose not to respond. These were: ID 20, ID 56, ID 104, ID 105, ID 137, ID 160, ID 248, ID 306, ID 384.

4. Run the correct analysis for assessing the normality of the intentions data and provide a brief interpretation of the output as per APA guidelines (4 marks).

In total, the normality analysis for the intentions to eat healthy data was conducted on 390 participants. The 5% trimmed mean (23.06) was similar to the overall mean (23.10), suggesting that extreme scores were not having much of an influence on the data. The skewness value of 0.037 is relatively close to zero, suggesting little skew of the data, while the kurtosis value of 1.503 suggests a slight peaking of the distribution (Leptokurtic). The results of the Kolmogorov-Smirnov and Shapiro-Wilk Test of normality are both significant, suggesting a violation of the assumption of normality. This is confirmed by the histogram, which shows the majority of scores at the higher end of the scale. It is also confirmed by the Normal Q-Q plot, which shows some values deviating away from the expected value, thereby not forming a straight line. The Detrended Normal Q-Q Plot also shows one value deviating away from the straight line.

5. Check for any outliers in the data and write in the space below the **ID number(s)** and the **specific measure(s)** in which those participants are highlighted as outliers (2 marks).

Outliers = Following inspection of the histogram, stem and leaf plot, Q-Q plots, and boxplot of the data, participant ID 275 has been highlighted as an outlier with an intention to eat healthily score of 2. Although this score is within the scale range of the questionnaire (0-30), it has fallen well below the scores from all other participants and is therefore considered as an outlier for this particular piece of research and would be removed from the data.

Task Three: Analysis and effect size (10 marks)

Using the study information and SPSS datasheet on Moodle titled “**Analysis datasheet**” (which has already been screened and cleaned), you should do the following:

1. Run the correct **descriptive** analyses on the data in SPSS and in the space below write out the interpretation of the output as you would see in a Psychology report (4 marks).

From the output, we can see for the categorical descriptives there are 200 participants in each of the two groups (50% in each). There are 119 males (29.8%), 186 females (46.5%), 67 other (16.8%), and 27 (6.8%) who would prefer not to say. There is also now

one missing data for the gender variable (.3%). In terms of marital status, 115 participants are single (28.7%), 182 are either married or living with a partner (45.5%), 70 are separated or divorced (17.5%), and 33 are widowed (8.3%).

From the output, we can see for the continuous descriptives that 399 participants have provided information on their age. Participants' ages range from 19 to 40, with the mean age of the sample being 28 and a standard deviation of 6.73. Three-hundred and eighty-nine participants completed the intention to eat healthy scale, with scores ranging from 17 to 30. The mean on this scale was 23, with a standard deviation of 3.61.

2. Run the correct **inferential** analyses on the data in SPSS and in the space below write out the interpretation of the output as you would see in a Psychology report (4 marks).

The results from an independent samples t-test on intention to eat healthy between the text-messaging intervention group and the standard treatment group revealed a statistically significant difference $t(361.88) = 14.21, p < 0.001$, with those in the text-messaging intervention group reporting significantly higher intentions to eat healthy ($M = 25.28; SD = 3.28$) compared to the standard treatment group ($M = 21.05; SD = 2.54$). However, based on the outcome from Levene's test for equality of variances, the assumption of equality of variances for the current data was violated; therefore, the reported values have been obtained from the equal variances not assumed row in SPSS.

3. In the space below state whether the difference in scores is considered **clinically** significant? Justify your answer based on the study information provided (2 marks).

Although there was a statistically significant difference in the magnitude of the difference in the means (mean difference = 4.23, 95%), with narrow confidence intervals CI: 3.65-4.82, and an extremely large effect size – Cohen's $d = 2.93$; the current data would not be considered as clinically significant as the researchers on this project consider a difference in intentions score of 9 scale points between the two groups at the 5% level of significance as being a clinically significant difference. As mentioned above, the mean difference in this data was only 4.82, thereby falling nearly four scale points below what the researchers would accept as a clinically significant effect.