**Group Project Report**

**Subject:**

Data Analytics

**Students:**

Patrick Curran

Steven Dignam

**Introduction.**

We received a batch of data based on patients who attended a dentist.

Each patient was tested for pain using two different tests. One test was the patient giving a verbal perceived pain after treatment. The other was using scientific measures called Galvanic Responses.

So we can see this would be paired data as each score is related to the same patient. But there was variation between these scores.

A sample of 100 patients was taken from the population. This sample was split into 2 groups. One group of 50 received meditation. The other group of 50 did not get meditation.

The data analysis will then be concerned with whether patients who do meditation will have reduced pain from the treatments.

1. Determine whether the data provided is appropriate for the test(s) available and that any analysis is achievable.

We want to have a reasonable Normal Distribution to apply our standard tests. These data satisfies this as sample size > 30 and there are no significant outliers. Box and Whisker charts were completed on the data sets to show there were no significant outliers. This provides information that the data can be taken as independent.

We can see from the bar charts that the data has a reasonable normal distribution appearance.

We had to make some assumptions.

* **Dental Work:** We assume dentist treatment was similar to all in sample.
* **Age.** Age was not recorded so we assume the age spread is similar to each testing pot to rule out the possibility that age is a factor on pain recorded.
* **Gender split:**Though each data is split equally, within the groups the male and female divide is not equal. For example there was more females (27) in the sample who meditated against not (23). Alternatively, there was less males (23) in the sample who meditated than not (27).

Testing only male and female data we can see there is some differences. See graphs where mean scores for male pain are higher than females. However, we will proceed based on the fact that the numbers samples is taken to be random and the difference in male and female scores could be due to other factors such as age.

1. Formulate a hypothesis test to be used to compare the effectiveness of the two approaches (control, meditation) used during dental surgery.

**Test One: Meditation V Control (All): GR**

* H0: µdiff = 0. There is no difference in the average pain recorded between the Meditation group and the control group in the Galvanic Response category.
* HA: µdiff 6= 0. There is a difference between the pain recorded from those who did meditation and those who were in control group.

**Test One: Meditation V Control: GR Male**

* H0: µdiff = 0. There is no difference in the average Galvanic Response pain recorded for Males between the Meditation group and the control group.
* HA: µdiff 6= 0. There is a difference in the average Galvanic Response pain recorded for Males between the Meditation group and the control group.

**Test Two: Meditation V Control: PP Male**

* H0: µdiff = 0. There is no difference in the average Perceived pain recorded for Males between the Meditation group and the control group.
* HA: µdiff 6= 0. There is a difference in the average Perceived pain recorded for Males between the Meditation group and the control group.

**Test Three: Meditation V Control: GR Female**

* H0: µdiff = 0. There is no difference in the average Galvanic Response pain recorded for Females between the Meditation group and the control group.
* HA: µdiff 6= 0. There is a difference in the average Galvanic Response pain recorded for Females between the Meditation group and the control group.

**Test Four: Meditation V Control: PP Female**

* H0: µdiff = 0. There is no difference in the average Perceived pain recorded for Females between the Meditation group and the control group
* HA: µdiff 6= 0. There is a difference in the average Perceived pain recorded for Males between the Meditation group and the control group.

1. **Analyse the data to provide the hypothesis testing conclusion.**

Do we include CI data here?

Just mean

1. Provide descriptive statistics (graphs and tables) of the data.

**Test One: Meditation V Control (All): GR**

1. Chart, histogram

   Description automatically generatedChart, box and whisker chart

   Description automatically generated

Spread of original Data. Can see is skewed slightly to the right. Has two outlier either side.

Looking at data though mean and median or more or less the dame and Standard deviation is low

**The summary of all the data is as follows:**

Min. 1st Qu. Median Mean 3rd Qu. Max.

5.000 6.300 6.800 6.847 7.300 8.900



Standard deviation is low. Median is close to Mean.

Spread fairly Normal.

**Control Data Only:**

Chart, histogram

Description automatically generated Chart, box and whisker chart

Description automatically generated

The data for control look reasonable normal.

The mean is close to the median and the standard deviation is low. What is high SD?

6 Sigma Rule

**Meditation Only:**

Chart, histogram

Description automatically generated Chart, box and whisker chart

Description automatically generated

The boxplot looks skewed to the right. However, there are few response variables, 7,8,9. Also the mean is close to the median and the standard deviation is low.

**Test One: Meditation V Control: GR Male**

* H0: µdiff = 0. There is no difference in the average Galvanic Response pain recorded for Males between the Meditation group and the control group.
* HA: µdiff 6= 0. There is a difference in the average Galvanic Response pain recorded for Males between the Meditation group and the control group.

1. Determine the 95% confidence interval for the population mean of each group, and the 95% confidence interval for the difference between the means of the two groups.