**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.

**Adjusting Figures:**

When looking at difference in gender results some adjustment was needed to make sure both genders had equal samples. I chose to adjust the Galvanic response data over the Perceived Data as the data is derived more scientifically and hence could be best placed when examining if Normal Distribution tests can be applied to the data.

**Data adjustment Male data.**

For Males there were 27 sampled in the control group, and 23 sampled in the Meditation group.We would like an evan split so will remove some samples from the control group.

The stats on the control group for Galvanic Responses are:



The boxplot is:

Chart, box and whisker chart

Description automatically generated

From the boxplot there doesn’t look like there is any outliers. However, the median is less than the mean so there is a small bit of skewness to the right. To get it more normal will remove some numbers up to and including the median.

The resulting stats are:



Chart, box and whisker chart

Description automatically generated

The median is now almost the same as the mean and the boxplot has a more normal appearance. The skewness to the right has reduced. We will proceed with this male data.

**Data adjustment Female data.**

The Female data has 27 in the meditation group V 23 in the control group.

Looking at the galvanic responses we have the following data:



Chart, box and whisker chart

Description automatically generated

The data looks reasonably normal in spread. Mean is slightly larger than median.

No outliers. Took 2 samples from extremes. 2 largest and 2 smallest.

New data gives.



Chart, box and whisker chart

Description automatically generated

Here the mean is closer to the median in value(less skewness) and reasonably normal looking with more data between the first and third quartile.

We will proceed with this data.

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.**

**G1:Stats Male Control Group Galvanic Response:**



**G2:Stats Male Meditation Group Galvanic Response:**



**G3:Stats Male Control Group Perceived Pain Response:**

Text

Description automatically generated with medium confidence

**G4:Stats Male Meditation Group Perceived Pain Response:**

Text

Description automatically generated with low confidence

**G5:Stats Female Control Group Galvanic Response:**

Text

Description automatically generated with medium confidence

**G6:Stats Female Meditation Group Galvanic Response:**

**G7:Stats Female Control Group Perceived Pain Response:**



**G8:Stats Female Med Group Perceived Pain:**

Text

Description automatically generated with medium confidence

**Descriptive Statistics:**

Chart, box and whisker chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart

Description automatically generated

Chart, bar chart

Description automatically generated

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.

95% CI for pop mean of each group:

**Group one: Male Control GR:**

A picture containing text

Description automatically generated







CI 95% is (6.66376,7.52824)

Mean is within that range so we are 95% happy the mean is correct.

**Group Two: Male Med GR:**

A picture containing text

Description automatically generated



Text

Description automatically generated with medium confidence

CI for group 2 is (6.470527,7.181473)

Mean is within that range so we are 95% happy the mean (6.826) is in this range.

**Group 3: Male Perceived Pain Control Group**

Graphical user interface, text, application

Description automatically generated



Text

Description automatically generated with medium confidence

CI 95% is (7.942668,8.491332)

Mean is within that range so we are 95% happy the mean is in this interval.

**Group 4: Male Perceived Pain Med Group**

A picture containing logo

Description automatically generated



Text

Description automatically generated

CI 95% is (7.81319,8.27281)

Mean is within that range so we are 95% happy the mean is in this interval.

**Group 5:Female GR Control Group**

Shape

Description automatically generated with low confidence

Text

Description automatically generated with medium confidence

CI 95% is (6.813555,7.264445)

Mean is within that range so we are 95% happy the mean is in this interval.

**Group 6: Female Med Group GR**

Logo

Description automatically generated



Text

Description automatically generated with medium confidence

CI 95% is (6.296984,6.685016)

Mean is within that range so we are 95% happy the mean is in this interval.

**Group 7: Female Control Group PP**

A picture containing graphical user interface

Description automatically generated



Text

Description automatically generated

CI 95% is (7.407962,8.070038)

Mean is within that range so we are 95% happy the mean is in this interval.

**Group 8: Female Med Group PP**

A picture containing logo

Description automatically generated



A picture containing text

Description automatically generated

CI 95% is (7.357852,7.772148)

Mean is within that range so we are 95% happy the mean is in this interval.

Difference of Means CI – 95%

Table

Description automatically generated

Test One: Male Control GR – Male Med GR

Mean – mean +\_ t\*(SE)