Mini Project#5

Sreedeep Reddy Konkayala, Sai Nanda Kishan Edimadakala

**1.Consider the data stored in bodytemp-heartrate.csv on eLearning, containing measurements of body temperature and heart rate for 65 male (gender = 1) and 65 female (gender = 2) subjects.**

1. **Do males and females differ in mean body temperature? Answer this question by performing an appropriate analysis of the data, including an exploratory analysis.**

Ans. By observing the box plot below we can see that, the 1st ,2nd (median) & 3rd quartiles of female body temperature is slightly higher than the male body temperatures so, the mean body temperature of Females can be Higher than that of Males. And the Female values have a lot of outliers so we cannot assume equal variances.

Chart, diagram, box and whisker chart

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedNow, by Observing the q-q plots

From the above Q-Q plots we can assume that the distributions of the male and female body temperature values as approximately normal.

Let µm be the population mean & m be the sample mean of body temperatures of Males, µf the population mean & f be the sample mean of female body temperatures

Now we take the Null hypothesis H0 : difference between means=0 =>m-f=0

Alternative hypothesis H1 : difference between the means0 => m-f0

As we can see that the samples here have unequal variances and are to be considered as independent samples coming from an approximately normal distributions. We use T distribution with Satterthwaite’s Approximation, to get the confidence interval.

The Confidence interval Observed from R,

Text, letter

Description automatically generated

The Confidence Interval (-0.53964856, -0.03881298).

The value of the p is 0.02394.

Since, the value of P is less than 0.05 and the 0 does not lie in the Confidence interval. We can reject the null hypothesis and conclude that the mean body temperatures of males and females are not equal.

**B. Do males and females differ in mean heart rate? Answer this question by performing an appropriate analysis of the data, including an exploratory analysis.**

Ans. By observing the boxplots of the male and female heart rate below, we can see that the 1st Quartile Q1 of Female heart rate is less than the Male Quartile Q1 heart rate. While the median & Q3 of female heart rate is greater than male median & Q3 heart rate. The females values seem to be more stretched with variability.

Chart, box and whisker chart

Description automatically generated

Now let’s look at the Q-Q plots:

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

From the above Q-Q plots we can assume that the distributions of the male and female Heart rate values as approximately normal.

Let µm be the population mean & m be the sample mean of Heart rate of Males, µf  the population mean & f be the sample mean of female Heart Rates

Now we take the,

Null hypothesis H0 : difference between means=0 =>m-f=0

Alternative hypothesis H1 : difference between the means0 => m-f0

As we can see that the samples here have unequal variances and are to be considered as independent samples coming from an approximately normal distributions. We use T distribution with Satterthwaite’s Approximation, to get the confidence interval.

The Confidence interval Observed from R,

Text, letter

Description automatically generated

The confidence interval from the t.test is = (-3.243732, 1.674501)

The value p is = 0.5287.

Since the value of p is greater than 0.05 and the 0 lies in the confidence interval, we accept the Null Hypothesis and can conclude that, Mean Heart rates of Males and females are Equal.

C. Is there a linear relationship between body temperature and heart rate? Does this relationship depend on gender? Answer these questions by performing an appropriate analysis of the data, including an exploratory analysis.

Ans. Now lets look at the Scatter plots with plotted regression line that represents the linear relationship between the heart rate and body temperature.

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

As we can observe that the regression line drawn has a slope that is greater than 0, this implies a positive correlation between the body temperature and the heart rate. But based on the graphs we can also assume that the correlation between them is very weak.

We can get the correlation between them by using the function cor().

Text

Description automatically generated

The value of Correlation between the male body Temperature and Heart rate is = 0.1955894

The value of Correlation between the Female body Temperature and Heart rate is = 0.2869312

From the above results of correlation, we can conclude that the relationship between Body Temperature and Heart rate is Weak. And since the value of correlation is higher for females than males we can also conclude that this relationship also depends on the gender.

Q2.

* (a).For a given setting, compute Monte Carlo estimates of coverage probabilities of the two intervals by simulating appropriate data, using them to construct the two confidence intervals, and repeating the process 5000 times.

To compute Monte Carlo estimates of coverage probabilities, We draw different samples and compute the confidence intervals for each. Then the coverage probabilities can be figured out by dividing the total number of times the mean value lies inside the confidence interval by total number of samples.

Now we use to methods to find the confidence interval: Z Interval and the Other is Percentile Boot Strap.





* (b) Repeat (a) for the remaining combinations of (n, λ). Present an appropriate summary of the results.

A picture containing text, receipt

Description automatically generated

Graphs (i)

Chart, line chart

Description automatically generated

Here, the red line denotes the z proportions, and the blue line denotes the bootstrap proportions. The values are plotted for lambda while keep n fixed.

Graphs(ii)

Chart, line chart

Description automatically generated

Here, the red line denotes the z proportions, and the blue line denotes the bootstrap proportions. The values are plotted for n while keep lambda fixed.

* (c) From the above graphs (ii), we observe that the graphs have minor differences for varying 𝜆. We can deduce that the coverage probabilities don’t depend on 𝜆. The coverage probabilities from bootstrap method are higher than those of z-interval method. From graphs (i), we see observable changes for fixed n. We can conclude that the coverage probabilities depend on n.

For z-interval with a large sample (n=100), the coverage probabilities are as accurate as the coverage probabilities from bootstrap method. The coverage probabilities for the bootstrap method are higher starting from n=30. From the graph, the coverage probabilities from bootstrap method are higher for every combination of (n, 𝜆) compared to the z-interval method for large sample. We can thus choose the bootstrap method as it is more appropriate for smaller sample sizes.

* (d)

A picture containing text, receipt

Description automatically generated

From the above accuracy matrix, We can Infer that that,

The coverage probability for bootstrap is for n = 5 lambda = 0.1 is 0.8984.

The coverage probability for large sample z for n = 5 lambda = 0.1 is 0.8070.

The coverage probability for bootstrap is for n = 10 lambda = 0.1 is 0.9218.

The coverage probability for large sample z for n = 10 lambda = 0.1 is 0.8638.

The coverage probability for bootstrap is for n = 30 lambda = 0.1 is 0.9374.

The coverage probability for large sample z for n = 30 lambda 0.1 is 0.9212.

The coverage probability for bootstrap is for n = 100 lambda = 0.1 is 0.9424.

The coverage probability for large sample z for n = 100 lambda = 0.1 is 0.9400.

Therefore the conclusions obtained in (c) hold a specific values of lambda. In this case lambda = 0.1