Mini-Project duo group 7

Venkata Bapanapalli

Karthik Meyyappan

Contributions: Both members equally contributed to analytically solve and implement the code of the given two questions.

Problem 1

1. Chart, box and whisker chart

   Description automatically generated

Chart, scatter chart

Description automatically generatedThe general distribution of the body temperatures of females, on average is higher than that of the males. The female body temperatures have some outliers which indicates that there is high variability in the data.

The above plots indicate that both sets of data have normal distribution. Due to the outliers presented earlier the female body temperatures seem to show indications of deviating from normal distribution but we can assume normal distribution if we do not factor in the outliers.

Hypotheses:

H0:

H1:

Calculating the confidence interval, we get (-0.54, -0.04) with the p-value being 0.02. Furthermore, the value 0 is not within the confidence interval and the p-value is less than 0.05, we reject the null hypothesis and can conclude that the average temperatures of males and females is not equal.

1. Chart, box and whisker chart

   Description automatically generated

Like body temperatures, the heart rates of females are on average higher than that of males. Females also seem to have higher variability as the hearts rates have higher range. However, unlike body temperatures the heart rates do not have outliers even with high variability.

Chart, scatter chart

Description automatically generatedThe distributions of heart rates for both males and females are approximately normal. No abnormalities are found with the data; hence we have strong confidence of these distributions being normal.

Hypotheses:

H0:

H1:

Calculating the confidence interval, we get (-3.24, 1.67) with the p-value being 0.53. Furthermore, the value 0 is within the confidence interval and p-value is greater than 0.05, so we accept the null hypothesis that the mean heartrates of both genders males and females is approximately equal.

1. Chart, scatter chart

   Description automatically generated

Chart, scatter chart

Description automatically generated

The correlation can be seen using a regression line from a scatter plot. scatter plots for both genders indicate that the correlation between heart rates and body temperatures exists but its weak. It can also be determined that the correlation is positive which can be determined from the positive regression line. Using a correlation function we determined that the correlation values are 0.2 and 0.29 for male and female data, respectively. The correlation for both genders can be concluded to be relatively weak due to the small correlation values.

Problem 2

1. Coverage probabilities:
   * Z-interval: 0.9142
   * Bootstrap interval: 0.8996
2. Summary of results for all other combinations of n and lambda

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Z-Proportions | L = 0.01 | L = 0.1 | L = 1 | L = 10 |
| N = 5 | 0.8136 | 0.816 | 0.8164 | 0.8114 |
| N = 10 | 0.8678 | 0.874 | 0.8716 | 0.863 |
| N = 30 | 0.9146 | 0.9212 | 0.916 | 0.9174 |
| N = 100 | 0.93 | 0.9376 | 0.9392 | 0.9364 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| B-Proportions | L = 0.01 | L = 0.1 | L = 1 | L = 10 |
| N = 5 | 0.8994 | 0.896 | 0.8962 | 0.8902 |
| N = 10 | 0.9262 | 0.9226 | 0.9156 | 0.9244 |
| N = 30 | 0.9346 | 0.9376 | 0.9392 | 0.939 |
| N = 100 | 0.9436 | 0.9434 | 0.95 | 0.944 |

A picture containing text, screenshot, computer

Description automatically generated

Figure : Red represents Z proportions and Blue represents bootstrap proportions.

A picture containing text, computer, screenshot

Description automatically generated

Figure : Red represents Z proportions and Blue represents bootstrap proportions.

1. From the first set graphs it can be observed that the coverage probabilities are not dependent on lambda as they do not change drastically with the change in lambda. Moreover, it can be observed that the coverage probabilities obtained from the bootstrap method is higher than the z-interval method. From the second set of graphs, we can conclude that the coverage probabilities are dependent on n. When n is large the coverage probabilities are for the bootstrap method when compared to the z-interval method. When comparing all the graphs it can be concluded that the bootstrap method is more accurate even at lower n values compared to the z-interval method. Hence, the bootstrap method is preferred.
2. For lambda at 0.1 it can be concluded that conclusions in part c are valid from the tables above.