DATA ACTIVITY 6 – HEALTH DATA

Task 6.1 – Find out the mean, median, and mode of the age variable

```
#Calculate mean, median, and mode of age
mean(Health_Data$age)
median(Health_Data$age)
mode_age <- as.numeric(names(sort(table(Health_Data$age), decreasing=TRUE) [1]))
show(mode_age)
```

```
> mean(Health_Data$age)
[1] 26.51429
> median(Health_Data$age)
[1] 27
> mode_age <- as.numeric(names(sort(table(Health_Data$age), decreasing=TRUE) [1]))
> show(mode_age)
[1] 26
```

Task 6.2 – Find out whether median diastolic blood pressure is same among diabetic and non-diabetic participants.

```
#Find out whether median diastolic blood pressure is same among diabetic and non-diabetic participants
# # Check distribution of dbp
shapiro.test(Health_Data$dbp)
# # Use Mann-Whitney U test
wilcox.test(dbp~diabetes,data=Health_Data)
```

```
> shapiro.test(Health_Data$dbp)

        Shapiro-Wilk normality test

data:  Health_Data$dbp
W = 0.97052, p-value = 0.0002204

> wilcox.test(dbp~diabetes,data=Health_Data)

        Wilcoxon rank sum test with continuity correction

data:  dbp by diabetes
W = 3804.5, p-value = 0.7999
alternative hypothesis: true location shift is not equal to 0
```

The Shapiro-Wilk test was used to determine the distribution of diastolic blood pressure. As the p-value is less than 0.05 (0.0002), the data isn't parametric (normally distributed), meaning that a non-parametric test should be used.

The Mann-Whitney U test was used as the independent variable is two groups independent of each other.

As the p-value calculated with the Mann-Whitney test is larger than 0.05 (0.80), showing that there is no significant difference between the median diastolic blood pressure among diabetic and non-diabetic groups. It can be concluded that diastolic blood pressure is the same amongst both groups.

Task 6.3 – Find out whether systolic BP is different across occupational group.

```
# Find out whether systolic BP is different across occupational group
# # Check distribution of sbp
shapiro.test(Health_Data$sbp)

# # Use Kruskal-Wallis test
kruskal.test(sbp~occupation, data=Health_Data)
```

```
> kruskal.test(sbp~occupation, data=Health_Data)

        Kruskal-Wallis rank sum test

data:  sbp by occupation
Kruskal-Wallis chi-squared = 0.77906, df = 3, p-value = 0.8545
```

The Shapiro-Wilk test was used to determine the distribution of systolic blood pressure. As the p-value is less than 0.05 (3.345e-06), the data isn't parametric (normally distributed), meaning that a non-parametric test should be used.

The Kruskal-Wallis test was used as four groups feature within the independent variable, which is occupation in this scenario, and it extends the Mann-Whitney U test to more than two groups. The p-value returned was larger than 0.05 (0.85). Therefore, no significant difference in the systolic BP between occupational groups.