Question_No_7

Utsab Bhattarai

2024-05-31

Do the followings in R studio using "airquality" data set of R script to knit PDF output:

```
data("airquality")
```

a. Perform goodness-of-fit test on Temp variable to check it it follows normal distribution or not.

```
library(stats)
shapiro_test <- shapiro.test(airquality$Temp)
shapiro_test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  airquality$Temp
## W = 0.97617, p-value = 0.009319
```

Interpretation: The Shapiro-Wilk test for normality checks if the Temp variable follows a normal distribution. Here, the p-value (0.009319) is less than 0.05, thus we reject the null hypothesis that the data follows a normal distribution.

b. Perform goodness-of-fit test on Temp variable by Month variable to check if the variances of mpg are equal or not on am variable categories.

```
bartlett_test <- bartlett.test(Temp ~ factor(Month), data = airquality)
bartlett_test
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Temp by factor(Month)
## Bartlett's K-squared = 12.023, df = 4, p-value = 0.01718
```

Interpretation: Bartlett Test checks if the variances of Temp are equal across the different months. Here, the p-value (0.01718) is less than 0.05, we reject the null hypothesis that the variances are equal.

c. Discuss which independent sample must be used to compare "Temp" variable by "Month" variable categories based on the results obtained above.

Interpretation: As the Temp data follows a normal distribution and the variances are equal, we use ANOVA.

d. Perform the best independent sample statistical test for this data now and interpret the results carefully.

```
anova_test <- aov(Temp ~ factor(Month), data = airquality)
summary(anova_test)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## factor(Month)   4   7061  1765.3   39.85 <2e-16 ***
## Residuals     148   6557    44.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation: ANOVA checks if there are any statistically significant differences in the means of Temp across the different months. Here, the p-value is less than 0.05, we reject the null hypothesis that the means are equal.