Q7 (Unmentioned in Question Paper)

Niraj Raj Kharel (16)

2024-05-31

Question 7 (Unmentioned in Question Paper)

```
# Question 7 (Not mentioned)
data("airquality")

# Assigning airquality dataframe to a new variable
aq <- airquality
str(aq)
```

```
## 'data.frame':    153 obs. of  6 variables:
##  $ Ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...
##  $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
##  $ Wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
##  $ Temp   : int  67 72 74 62 56 66 65 59 61 69 ...
##  $ Month  : int  5 5 5 5 5 5 5 5 5 5 ...
##  $ Day    : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
# Question a
# Test of normality
# Perform Shapiro-Wilk test for temperature
result <- tapply(aq$Temp, aq$Month, shapiro.test)

# Print the results
# The p-value is not significant and this implies that the temp variable
# with month does follow normal distribution
print(result)
```

```
## $`5`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.94771, p-value = 0.1349
##
##
## $`6`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.97158, p-value = 0.5832
##
##
```

```
## $`7`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.94579, p-value = 0.1194
##
##
## $`8`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.96391, p-value = 0.3688
##
##
## $`9`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.9513, p-value = 0.1831
```

```r
# Question b
# Convert Month to a factor
aq$Month <- factor(aq$Month)

# Perform Bartlett's test for homogeneity of variances
bartlett_result <- bartlett.test(Temp ~ Month, data = aq)

# Print Bartlett's test result
print(bartlett_result)
```

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

```r
#Here it can be seen that the variances are not equal.

# Question c
#In the above scenario, Bartlett's test indicates that the variances of the
# "Wind" variable are not approximately equal across different months.

#Fit one way anova model
# Convert Month to a factor
aq$Month <- factor(aq$Month)

anova_model <- aov(Wind ~ Month, data = aq)
summary(anova_model)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## Month         4  164.3   41.07   3.529 0.00879 **
## Residuals   148 1722.3   11.64
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Question d

# Perform Tukey's HSD test
tukey_result <- TukeyHSD(anova_model)

# Print the Tukey HSD test result
print(tukey_result)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Wind ~ Month, data = aq)
##
## $Month
##            diff       lwr        upr     p adj
## 6-5 -1.35591398 -3.768713  1.0568846 0.5305524
## 7-5 -2.68064516 -5.073585 -0.2877054 0.0197174
## 8-5 -2.82903226 -5.221972 -0.4360925 0.0117066
## 9-5 -1.44258065 -3.855379  0.9702179 0.4674045
## 7-6 -1.32473118 -3.737530  1.0880674 0.5535894
## 8-6 -1.47311828 -3.885917  0.9396803 0.4456532
## 9-6 -0.08666667 -2.519162  2.3458285 0.9999786
## 8-7 -0.14838710 -2.541327  2.2445527 0.9998052
## 9-7  1.23806452 -1.174734  3.6508631 0.6176733
## 9-8  1.38645161 -1.026347  3.7992502 0.5081147
```

```r
# Here it makes a pairwise comparison across months for temperature variable.
# We can see that the month pairs 7-5 and 8-5 are significant
# where there is the difference in mean temperatures.
```