Project 4 Part I

Kaushal Khatiwada

2024-05-02

Part 1

Using "airquality" Dataset

A)Get mean and standard deviation of Wind variable by Month variable using the appropriate "apply" family of function, show both the results in a single table and interpret them carefully

```
aq <- airquality
aq_means<- tapply(aq$Wind,aq$Month,mean) #Mean
aq_sd <- tapply(aq$Wind,aq$Month,sd)     #Standard Deviation
aq_table <- data.frame(Month = names(aq_means),
                            Mean_of_Wind = as.numeric(aq_means),
                            SD_of_Wind = as.numeric(aq_sd))
aq_table
```

```
##   Month Mean_of_Wind SD_of_Wind
## 1     5    11.622581   3.531450
## 2     6    10.266667   3.769234
## 3     7     8.941935   3.035981
## 4     8     8.793548   3.225930
## 5     9    10.180000   3.461254
```

B) Perform goodness-of-fit test on Wind variable by Month variable to check if it follows normal distribution or not

```
library(stats)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Find the Sample size

```r
permonth <- aq %>% group_by(Month) %>% summarize(count= n())
permonth
```

```
## # A tibble: 5 x 2
##   Month count
##   <int> <int>
## 1     5    31
## 2     6    30
## 3     7    31
## 4     8    31
## 5     9    30
```

Since, the sample size for each month is <100 we use Shapiro-Wilk normality test

```r
tapply(aq$Wind, aq$Month, shapiro.test)
```

```
## $`5`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.968, p-value = 0.4659
##
##
## $`6`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.96858, p-value = 0.501
##
##
## $`7`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.95003, p-value = 0.1564
##
##
## $`8`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.98533, p-value = 0.937
```

```
##
##
## $`9`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.97853, p-value = 0.7852
```

```
# p-value = 0.4659 , Normally distributed
# p-value = 0.501 , Normally distributed
# p-value = 0.1564 , Normally distributed
# p-value = 0.937 , Normally distributed
# p-value = 0.7852 , Normally distributed
```

p-value of all months is > 0.05 so It follows normal distribution

C) Perform goodness-of-fit test on Wind variable by Month variable to check if the variances of Wind are equal or not on a variable categories

```
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

Levene Test is performed for variance check because there are >2 categorical group in Month

```
aq$Month <- factor(aq$Month)
leveneTest(aq$Wind ~ aq$Month, data=aq,center=mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value Pr(>F)
## group   4  0.1859 0.9454
##       148
```

p-value = 0.9454 i.e > 0.05 so,equal variance among months

D) Discuss which one-way ANOVA must be used to compare "Wind" variable by "Month" variable categories based on the results obtained above

ASSUMPTIONS

Normality Test using shapiro-wilk also fulfilled Variance Test using LeveneTest also fulfilled Now ready to fit one-way ANOVA

E) Fit the best one-way ANOVA for this data now and interpret the results carefully

```
summary(aov(aq$Wind ~ aq$Month, data = aq))
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## aq$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
```

p-value is 0.008 i.e <0.05. so post-hoc test is required

F) Fit the most-appropriate post-hoc test if the ANOVA is statistically significant and interpret the result carefully

```
TukeyHSD(aov(Wind ~ Month,data=aq))
```

```
##   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
```

9-8