

PS_ Probability and Hypothesis Testing for TAs

Pablo E. Gutiérrez-Fonseca

2023-08-12

R practice.

Load the water pollution data into R.

```
## Species Mammalian.Size.group BodyWt BrainWt
## Length:62 Length:62 Min. : 0.005 Min. : 0.14
## Class :character Class :character 1st Qu.: 0.600 1st Qu.: 4.25
## Mode :character Mode :character Median : 3.342 Median : 17.25
## Mean : 198.790 Mean : 283.13
## 3rd Qu.: 48.202 3rd Qu.: 166.00
## Max. :6654.000 Max. :5712.00
##
## NonDreaming Dreaming TotalSleep LifeSpan
## Min. : 2.100 Min. :0.000 Min. : 2.60 Min. : 2.000
## 1st Qu.: 6.100 1st Qu.:0.900 1st Qu.: 8.05 1st Qu.: 6.625
## Median : 8.300 Median :1.800 Median :10.45 Median : 15.100
## Mean : 8.541 Mean :1.941 Mean :10.53 Mean : 19.878
## 3rd Qu.:11.000 3rd Qu.:2.500 3rd Qu.:13.20 3rd Qu.: 27.750
## Max. :17.900 Max. :6.600 Max. :19.90 Max. :100.000
## NA's :13 NA's :11 NA's :4 NA's :4
## Gestation Predation Exposure Danger
## Min. : 12.00 Min. :1.000 Min. :1.000 Min. :1.000
## 1st Qu.: 35.75 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:1.000
## Median : 79.00 Median :3.000 Median :2.000 Median :2.000
## Mean :142.35 Mean :2.871 Mean :2.419 Mean :2.613
## 3rd Qu.:207.50 3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000
## Max. :645.00 Max. :5.000 Max. :5.000 Max. :5.000
## NA's :4
```

1. Based on the summary results in JMP, mammals spend most of their time in which type of sleep phase (dreaming or non-dreaming)?

```
mean(df_mammals$NonDreaming, na.rm = TRUE)
```

```
## [1] 8.540816
```

```
mean(df_mammals$Dreaming, na.rm = TRUE)
```

```
## [1] 1.941176
```

2. Which type of sleep phase has the highest variability across the species included here (dreaming or non-dreaming)?
Non-dreaming

```
sd(df_mammals$NonDreaming, na.rm = TRUE)
```

```
## [1] 3.744046
```

```
sd(df_mammals$Dreaming, na.rm = TRUE)
```

```
## [1] 1.445016
```

3. Enter the p-value for the goodness of fit test for the **Dreaming variable** (NOTE...just enter the number (no letters, symbols, equal signs etc)...also be careful of decimal places.

```
shapiro.test(df_mammals$Dreaming)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: df_mammals$Dreaming  
## W = 0.87556, p-value = 7.067e-05
```

4. Based on this goodness of fit test, is the **dreaming variable** normally distributed? **No**
5. Enter the new p-value for the goodness of fit test for the **Dreaming variable** when the outliers are excluded (NOTE...just enter the number (no letters, symbols, equal signs etc)...also be careful of decimal places.

```
quartiles <- quantile(na.omit(df_mammals$Dreaming), probs = c(0.25, 0.75))  
IQR <- IQR(na.omit(df_mammals$Dreaming))  
Lower <- quartiles[1] - 1.5*IQR  
Upper <- quartiles[2] + 1.5*IQR  
  
new_Dreaming <- subset(df_mammals$Dreaming, df_mammals$Dreaming > Lower & df_mammals$Dreaming < Upper)  
  
shapiro.test(new_Dreaming)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  new_Dreaming  
## W = 0.95543, p-value = 0.06599
```