

ANOVA-ANALYSIS OF VARAINCE

ERICK@

2024-04

ANOVA-ANALYSIS OF VARAINCE

```
library(tidyverse)
library(forcats)
library(gapminder)
library(report)
```

```
#data
# data("msleep")
df <- msleep
head(df)
```

```
## # A tibble: 6 x 11
##   name      genus vore order conservation sleep_total sleep_rem sleep_cycle awake
##   <chr>    <chr> <chr> <chr> <chr>          <dbl>      <dbl>      <dbl> <dbl>
## 1 Cheetah Acin~ carni Carn~ lc             12.1        NA        NA      11.9
## 2 Owl mo~ Aotus omni Prim~ <NA>          17          1.8        NA       7
## 3 Mounta~ Aplo~ herbi Rode~ nt             14.4        2.4        NA      9.6
## 4 Greate~ Blar~ omni Sori~ lc             14.9        2.3        0.133   9.1
## 5 Cow     Bos  herbi Arti~ domesticated  4           0.7        0.667   20
## 6 Three~~ Brad~ herbi Pilo~ <NA>          14.4        2.2        0.767   9.6
## # i 2 more variables: brainwt <dbl>, bodywt <dbl>
```

```
# dim(df)
colnames(df)
```

```
## [1] "name"      "genus"     "vore"      "order"     "conservation"
## [6] "sleep_total" "sleep_rem" "sleep_cycle" "awake"     "brainwt"
## [11] "bodywt"
```

```
# df %>% pull(name)
```

```
str(df)
```

```
## tibble [83 x 11] (S3: tbl_df/tbl/data.frame)
## $ name      : chr [1:83] "Cheetah" "Owl monkey" "Mountain beaver" "Greater short-tailed shrew" ..
## $ genus     : chr [1:83] "Acinonyx" "Aotus" "Aplodontia" "Blarina" ...
## $ vore      : chr [1:83] "carni" "omni" "herbi" "omni" ...
```

```
## $ order      : chr [1:83] "Carnivora" "Primates" "Rodentia" "Soricomorpha" ...
## $ conservation: chr [1:83] "lc" NA "nt" "lc" ...
## $ sleep_total : num [1:83] 12.1 17 14.4 14.9 4 14.4 8.7 7 10.1 3 ...
## $ sleep_rem   : num [1:83] NA 1.8 2.4 2.3 0.7 2.2 1.4 NA 2.9 NA ...
## $ sleep_cycle : num [1:83] NA NA NA 0.133 0.667 ...
## $ awake       : num [1:83] 11.9 7 9.6 9.1 20 9.6 15.3 17 13.9 21 ...
## $ brainwt     : num [1:83] NA 0.0155 NA 0.00029 0.423 NA NA NA 0.07 0.0982 ...
## $ bodywt      : num [1:83] 50 0.48 1.35 0.019 600 ...
```

```
df1 <- df %>% select(vore, sleep_rem) %>%
  drop_na()
aov_mod <- aov(sleep_rem~vore, data = df1)
summary(aov_mod)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## vore       3  18.93   6.310    4.136 0.0105 *
## Residuals 52   79.33   1.526
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
report(aov_mod)
```

```
## The ANOVA (formula: sleep_rem ~ vore) suggests that:
##
## - The main effect of vore is statistically significant and large (F(3, 52) =
## 4.14, p = 0.011; Eta2 = 0.19, 95% CI [0.03, 1.00])
##
## Effect sizes were labelled following Field's (2013) recommendations.
```

```
#p_value is less than (0.05)
#reject H0(MEANS ARE THE SAME)
```

```
#create a data frame
```

```
gapdata <- gapminder %>%
  filter(year==2007 &
    continent %in% c('Americas', 'Europe', 'Asia')) %>%
  select(continent, lifeExp)
#Take a look at the distribution of means
gapdata %>% group_by(continent) %>%
  summarise(means=mean(lifeExp)) %>%
  arrange(desc(means))
```

```
## # A tibble: 3 x 2
##   continent means
##   <fct>      <dbl>
## 1 Europe    77.6
## 2 Americas  73.6
## 3 Asia     70.7
```

```

# Research Question: Is the life expectancy
# in this three continents equal???
# Hypothesis
# H0:Mean Life Expectancy is the same
# H1:Mean Life Expectancy is not the same

#Observation
#Create ANOVA MODEL
aov_mod1 <- aov(lifeExp~continent, data = gapdata)
summary(aov_mod1)

```

```

##           Df Sum Sq Mean Sq F value    Pr(>F)
## continent    2  755.6    377.8    11.63 3.42e-05 ***
## Residuals   85 2760.3     32.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
report(aov_mod1)
```

```

## The ANOVA (formula: lifeExp ~ continent) suggests that:
##
## - The main effect of continent is statistically significant and large (F(2, 85)
## = 11.63, p < .001; Eta2 = 0.21, 95% CI [0.09, 1.00])
##
## Effect sizes were labelled following Field's (2013) recommendations.

```

```
# Observation:
```

```
# The main effect of continent is statistically significant and large (F(2, 85) =11.63, p < .001; Eta2 = 0.21)
```

```
#is this significance being driven by a particular continent???
```

```
TukeyHSD(aov_mod1) #>% plot()
```

```

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = lifeExp ~ continent, data = gapdata)
##
## $continent
##           diff          lwr          upr      p adj
## Asia-Americas -2.879635 -6.4839802  0.7247099 0.1432634
## Europe-Americas 4.040480  0.3592746  7.7216854 0.0279460
## Europe-Asia     6.920115  3.4909215 10.3493088 0.0000189

```