ANOVA

LANGAT ERICK

2023-08-19

ANOVA

Research Question: Is the life expectancy in **asia, america & Europe different?**

HO;Mean life expectancy is the same

H1:Mean life expectancy is not same

```r
#IMPORT LIBRARIES
library(bookdown)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(patchwork)
library(gapminder)
library(forcats)
```

```r
#data=gapmider
data()
head(gapminder)
```

```
## # A tibble: 6 x 6
##   country     continent  year lifeExp      pop gdpPercap
##   <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Afghanistan Asia       1952    28.8  8425333      779.
## 2 Afghanistan Asia       1957    30.3  9240934      821.
## 3 Afghanistan Asia       1962    32.0 10267083      853.
## 4 Afghanistan Asia       1967    34.0 11537966      836.
## 5 Afghanistan Asia       1972    36.1 13079460      740.
## 6 Afghanistan Asia       1977    38.4 14880372      786.
```

```r
dim(gapminder)
```

```
## [1] 1704    6
```

```r
names(gapminder)
```

```
## [1] "country"   "continent" "year"      "lifeExp"   "pop"       "gdpPercap"
```

```r
tail(gapminder)
```

```
## # A tibble: 6 x 6
##   country  continent  year lifeExp       pop gdpPercap
##   <fct>    <fct>     <int>   <dbl>     <int>     <dbl>
## 1 Zimbabwe Africa     1982    60.4  7636524      789.
## 2 Zimbabwe Africa     1987    62.4  9216418      706.
## 3 Zimbabwe Africa     1992    60.4 10704340      693.
## 4 Zimbabwe Africa     1997    46.8 11404948      792.
## 5 Zimbabwe Africa     2002    40.0 11926563      672.
## 6 Zimbabwe Africa     2007    43.5 12311143      470.
```

```r
view(gapminder)
```

```r
#Create data to work with
gapdata <- gapminder %>% filter(year==2007 &
                                  continent%in% c("Americas", 'Europe', 'Asia'))%>%
              select(continent, lifeExp)

 head(gapdata)
```

```
## # A tibble: 6 x 2
##   continent lifeExp
##   <fct>       <dbl>
## 1 Asia         43.8
## 2 Europe       76.4
## 3 Americas     75.3
## 4 Europe       79.8
## 5 Asia         75.6
## 6 Asia         64.1
```

```r
 # gapdata1 <- gapminder %>%
 #   filter(year==2007 &
 #               continent %in%c("Americas", 'Europe', 'Asia'))%>%
 #       select(continent, lifeExp)
```

```r
#Take a look at the distribution of means
gapdata %>% group_by(continent) %>%  summarise(Mean_life=mean(lifeExp)) %>%
        arrange(Mean_life)
```

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

```r
#ANOVA MODEL
aov_model <- aov(lifeExp~continent,gapdata)
summary(aov_model)
```

```
##             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
```

```r
(aov_model <- aov(lifeExp~continent,gapdata) %>% TukeyHSD())
```

```
##   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
```