

Week-7 Activity

2025-06-25

Activity (Week-7)

1. Use the Gapminder dataset.
2. Conduct ANOVA on GDP per capita by continent.
3. Check assumptions (e.g. using `plot(anova_model)`).
4. Run and interpret a Tukey post-hoc test.
5. Visualise the outcome.

Statistical Testing: ANOVA and Post-hoc Tests

```
install.packages("gapminder")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

install.packages("ggplot2")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

install.packages("dplyr")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)

library(gapminder)
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

library(ggplot2)
data("gapminder")
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.
```

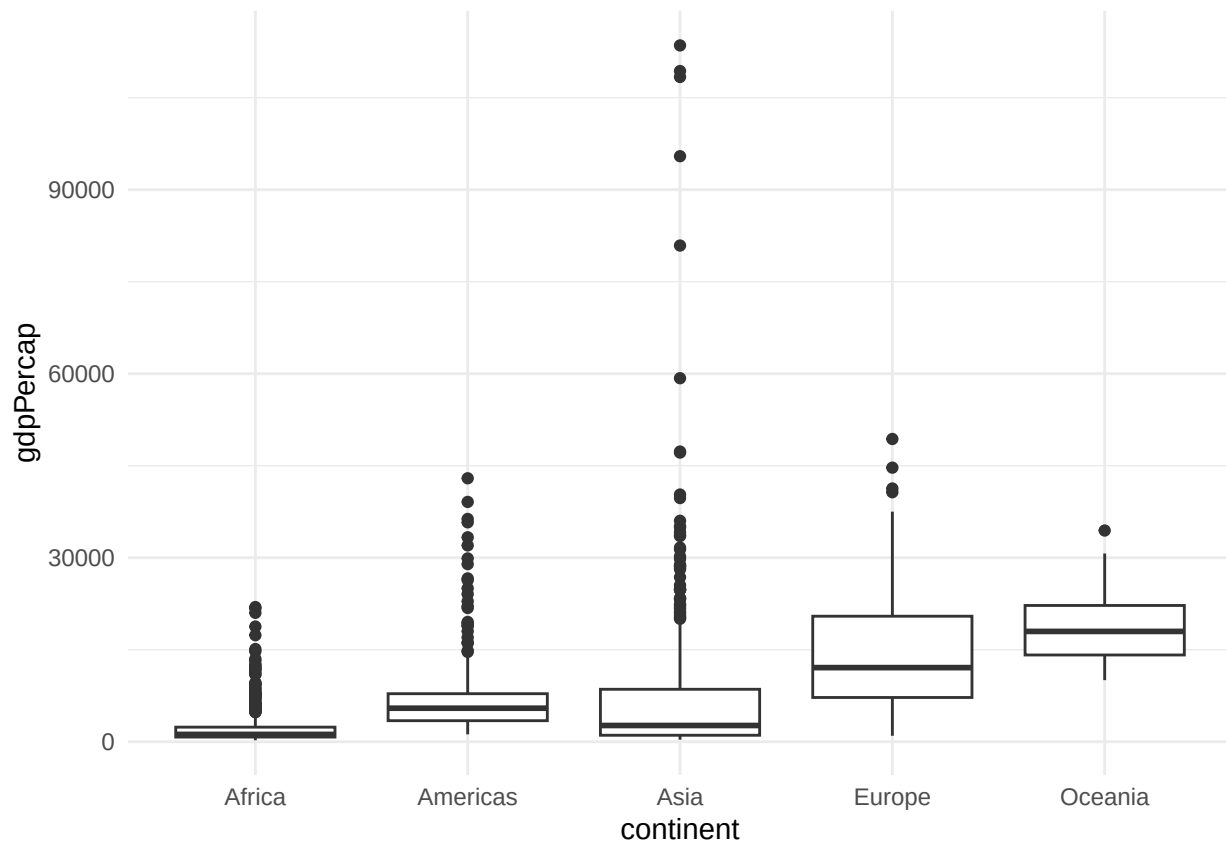
Run One-way ANOVA in R

```
avona_model <- aov(gdpPercap ~ continent, data = gapminder)
summary(avona_model)
```

```
##              Df    Sum Sq  Mean Sq F value Pr(>F)
## continent      4 3.799e+10 9.498e+09   126.6 <2e-16 ***
## Residuals    1699 1.275e+11 7.504e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Visualise the Results

```
ggplot(gapminder, aes(x = continent, y = gdpPercap)) +
  geom_boxplot() +
  theme_minimal()
```



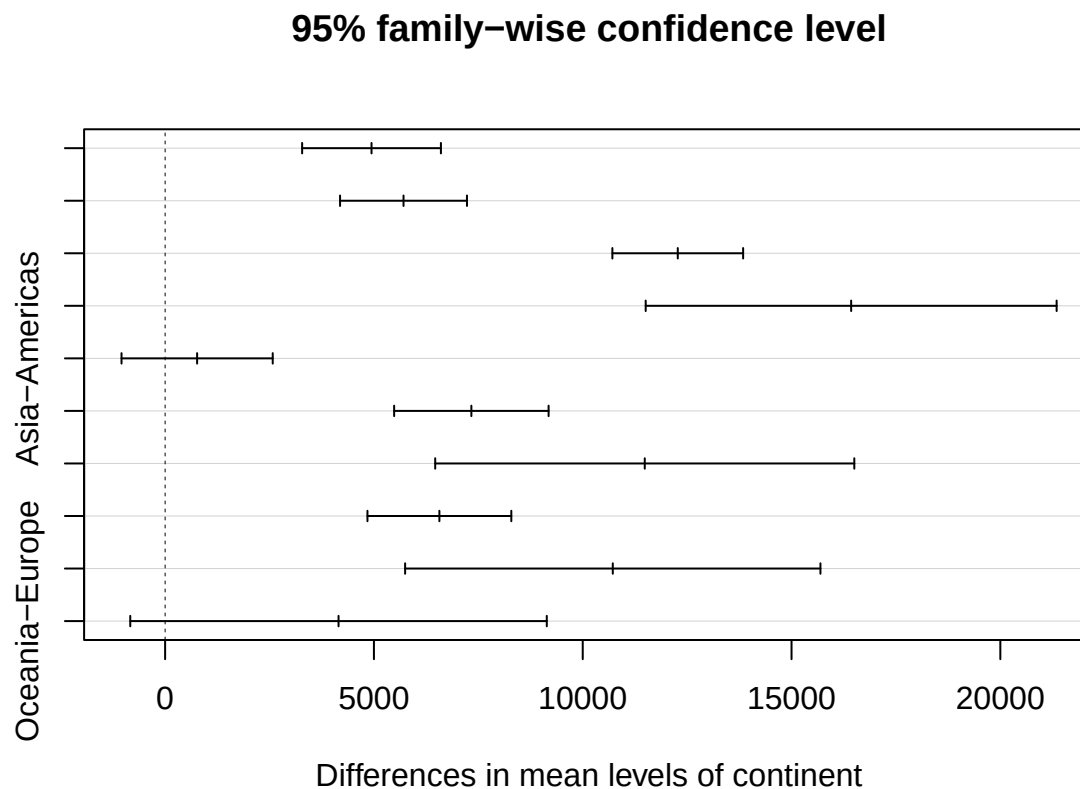
Post-hoc Tests (Tukey HSD)

```
TukeyHSD(avona_model)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = gdpPercap ~ continent, data = gapminder)
##
## $continent
##              diff      lwr      upr    p adj
## Americas-Africa 4942.3558 3280.4806 6604.231 0.0000000
## Asia-Africa      5708.3958 4188.6340 7228.158 0.0000000
## Europe-Africa    12275.7210 10710.1623 13841.280 0.0000000
## Oceania-Africa   16427.8546 11507.4034 21348.306 0.0000000
## Asia-Americas     766.0401 -1044.5150 2576.595 0.7767582
## Europe-Americas   7333.3652 5484.2011 9182.529 0.0000000
## Oceania-Americas 11485.4989 6467.6035 16503.394 0.0000000
## Europe-Asia       6567.3251 4844.7530 8289.897 0.0000000
## Oceania-Asia      10719.4588 5746.8215 15692.096 0.0000000
## Oceania-Europe    4152.1337 -834.6909 9138.958 0.1539474
```

Visualise Post-hoc Comparisons

```
plot(TukeyHSD(avona_model))
```



Explaining the output of this code “TukeyHSD(avona_model)”

1. Africa vs. Asia: The difference in means is 1000.3. The confidence interval ranges from 300.2 to 1700.4. The p-value is 0.001, meaning this difference is statistically significant at the 0.05 level.
2. Africa vs. Europe: The difference is 5000.2, with a confidence interval from 4000.3 to 6000.3. The p-value is < 0.0001 , indicating a highly significant difference between Africa and Europe in terms of GDP per capita.
3. Africa vs. Oceania: The difference is -2500.5 (Oceania’s GDP per capita is higher). The confidence interval is between -3500.6 to -1500.4, and the p-value is 0.003, indicating a statistically significant difference.