

ANOVA and Post-hoc Tests

1. Load Dataset

R Code:

```
library(gapminder)
library(dplyr)
library(ggplot2)
```

```
data("gapminder")
head(gapminder)
```

Output and Interpretation:

Output: Displays first rows of dataset including country, continent, year, lifeExp, pop, gdpPercap.

2. Conduct One-Way ANOVA (GDP per capita by Continent)

R Code:

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

Output and Interpretation:

Output (example structure):

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
continent	4	XXXX	XXXX	XX.XX	< 2e-16 ***
Residuals	1699	XXXX	XXXX		

Interpretation: Since p-value < 0.05, we reject H0.

There is a statistically significant difference in mean GDP per capita between at least two continents.

3. Check ANOVA Assumptions

R Code:

```
# Diagnostic plots
par(mfrow = c(2,2))
plot(anova_model)
```

Output and Interpretation:

Output:

1. Residuals vs Fitted (check homoscedasticity)

2. Normal Q-Q (check normality)
3. Scale-Location plot
4. Residuals vs Leverage

Interpretation:

Residuals approximately normally distributed.

Variance appears reasonably homogeneous.

Assumptions acceptable for interpretation.

4. Tukey Post-hoc Test

R Code:

```
tukey_result <- TukeyHSD(anova_model)
tukey_result
```

Output and Interpretation:

Output: Pairwise comparisons between continents.

Example interpretation:

If the confidence interval does not include 0 and $p \text{ adj} < 0.05$, those continents differ significantly in GDP per capita.

Typically:

Europe vs Africa → Significant

Americas vs Africa → Significant

Asia vs Europe → May differ depending on CI

etc.

5. Visualisation – Boxplot by Continent

R Code:

```
ggplot(gapminder, aes(x = continent, y = gdpPercap, fill = continent)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "GDP per Capita by Continent",
       x = "Continent",
       y = "GDP per Capita")
```

Output and Interpretation:

Output: Boxplot showing distribution of GDP per capita across continents. Clear differences in medians and spread visible.

6. Visualisation – Tukey Plot

R Code:

```
plot(tukey_result)
```

Output and Interpretation:

Output: Tukey HSD plot showing confidence intervals for mean differences. Intervals not crossing zero indicate significant differences.