

ACCA7022 – Choosing the Right Statistical Test (Activity)

Scenario 1 – Life Expectancy: Asia vs Europe (2007)

Variables:

Continent → Categorical (Asia, Europe)

Life Expectancy → Numerical (continuous)

Appropriate Test:

Independent Samples t-test

Justification:

The independent t-test compares the mean of a numerical variable between two independent groups.

R Code:

```
library(gapminder)
```

```
library(dplyr)
```

```
data_2007 <- gapminder %>%
```

```
  filter(year == 2007, continent %in% c("Europe", "Asia"))
```

```
t.test(lifeExp ~ continent, data = data_2007)
```

Interpretation:

If $p < 0.05$, the life expectancy between Asia and Europe differs significantly.

Scenario 2 – GDP per Capita by Continent

Variables:

Continent → Categorical (multiple groups)

GDP per capita → Numerical

Appropriate Test:

One-Way ANOVA

Justification:

ANOVA compares the mean of a numerical variable across three or more groups.

R Code:

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

Interpretation:

If $p < 0.05$, at least one continent has significantly different GDP per capita.

Scenario 3 – Relationship Between GDP and Life Expectancy

Variables:

GDP per capita → Numerical

Life expectancy → Numerical

Appropriate Test:

Pearson Correlation

Justification:

Pearson correlation measures the linear relationship between two numerical variables.

R Code:

```
cor(gapminder$gdpPercap, gapminder$lifeExp, method = "pearson")
```

Interpretation:

A value close to +1 indicates strong positive correlation between GDP and life expectancy.

Scenario 4 – Sleep Duration Across Age Groups

Variables:

Age group → Categorical

Sleep hours → Numerical

Appropriate Test:

One-Way ANOVA

Justification:

ANOVA compares the mean sleep hours across multiple age groups.

R Code:

```
anova_model <- aov(hours ~ age_group, data = sleep_data)
summary(anova_model)
```

Interpretation:

If $p < 0.05$, at least one age group has significantly different sleep duration.