

Week8 Activity

2025-07-01

```
# Load necessary libraries
```

```
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)
```

```
library(car)
```

```
## Loading required package: carData
```

```
##
```

```
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##   recode
```

```
# Load the gapminder dataset
```

```
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.
```

```
# Filter data for year 1962
```

```
data_1962 <- gapminder %>% filter(year == 1962)
```

```
table(data_1962$continent)
```

```
##
```

```
##   Africa Americas      Asia  Europe Oceania
```

```
##     52        25       33      30       2
```

```
# Check normality for each continent with sufficient data points
```

```
normality_tests <- data_1962 %>%
```

```

group_by(continent) %>%
filter(n() >= 3) %>% # Ensure there are at least 3 data points
summarise(
  shapiro_stat = shapiro.test(lifeExp)$statistic,
  shapiro_p = shapiro.test(lifeExp)$p.value
)

print(normality_tests)

```

```

## # A tibble: 4 x 3
##   continent shapiro_stat shapiro_p
##   <fct>      <dbl>      <dbl>
## 1 Africa      0.975    0.328
## 2 Americas    0.950    0.248
## 3 Asia        0.967    0.395
## 4 Europe      0.821    0.000163

```

```

# Get unique continent names
continents <- unique(data_1962$continent)

# Create all pairwise combinations
continent_pairs <- combn(continents, 2, simplify = FALSE)

# Run variance test for each pair and store results
var_test_results <- lapply(continent_pairs, function(pair) {
  subset_data <- data_1962 %>% filter(continent %in% pair)
  test <- var.test(lifeExp ~ continent, data = subset_data)

  data.frame(
    group1 = pair[1],
    group2 = pair[2],
    p_value = test$p.value,
    F_value = test$statistic
  )
})

# Combine results into a single data frame
var_test_results_df <- do.call(rbind, var_test_results)
print(var_test_results_df)

```

```

##   group1 group2      p_value      F_value
## F     Asia  Europe 2.250897e-05    5.2099937
## F1    Asia  Africa 1.024489e-03    0.3579239
## F2    Asia Americas 4.697745e-01    0.7497579
## F3    Asia Oceania 3.533847e-02 2007.1761023
## F4   Europe  Africa 7.383272e-02    1.8647812
## F5   Europe Americas 6.319719e-04    3.9062338
## F6   Europe Oceania 8.056723e-02 385.2549940
## F7   Africa Americas 2.691533e-02    0.4773860
## F8   Africa Oceania 5.923120e-02 718.4162612
## F9 Americas Oceania 4.070466e-02 1504.8960688

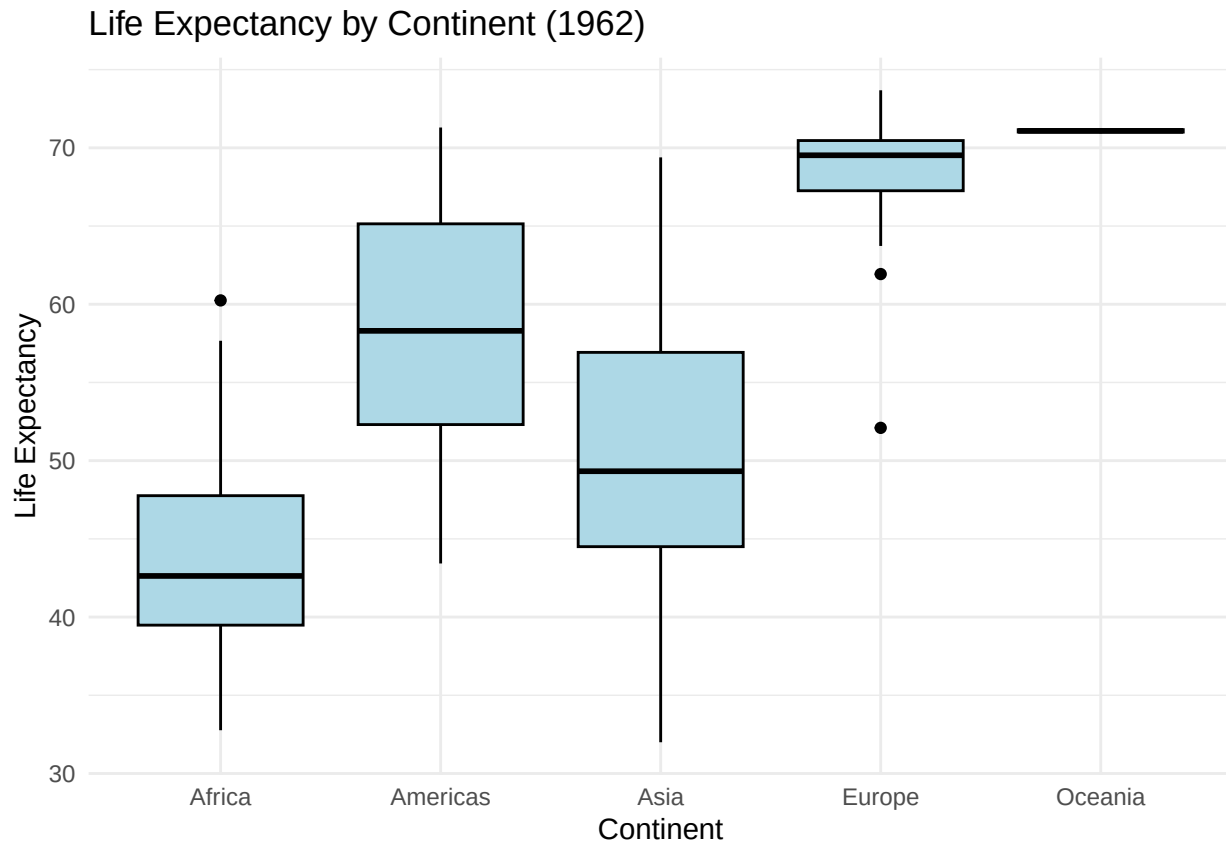
```

```

# Boxplot for visual inspection
ggplot(data_1962, aes(x = continent, y = lifeExp)) +
  geom_boxplot(fill = "lightblue", color = "black") +

```

```
theme_minimal() +
labs(title = "Life Expectancy by Continent (1962)",
     y = "Life Expectancy", x = "Continent")
```



```
# Run t-test for each pair and store results
t_test_results <- lapply(continent_pairs, function(pair) {
  subset_data <- data_1962 %>% filter(continent %in% pair)
  test <- t.test(lifeExp ~ continent, data = subset_data, var.equal = TRUE)

  data.frame(
    group1 = pair[1],
    group2 = pair[2],
    p_value = test$p.value,
    t_statistic = test$statistic
  )
})

# Combine results into a single data frame
t_test_results_df <- do.call(rbind, t_test_results)
print(t_test_results_df)
```

```
##      group1  group2      p_value t_statistic
## t      Asia   Europe 2.446593e-12 -8.7319181
## t1     Asia   Africa 5.758492e-06 -4.8471828
## t2     Asia  Americas 7.422546e-03  2.7783083
## t3     Asia   Oceania 9.085993e-03 -2.7720206
## t4     Europe  Africa 1.268470e-33 -20.5271382
```

##	t5	Europe	Americas	5.051792e-07	-5.7189270
##	t6	Europe	Oceania	4.164285e-01	-0.8240254
##	t7	Africa	Americas	1.071337e-13	-9.0751819
##	t8	Africa	Oceania	1.980860e-08	-6.6221967
##	t9	Americas	Oceania	4.872892e-02	-2.0720225