

Week-6 Activity

Dev

2025-06-24

```
# Activity 1: Effect of Sample Size on Sample Mean Distribution
install.packages("gapminder", repos='http://cran.us.r-project.org', quiet=TRUE)
install.packages("dplyr", repos='http://cran.us.r-project.org', quiet=TRUE)
library(gapminder)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

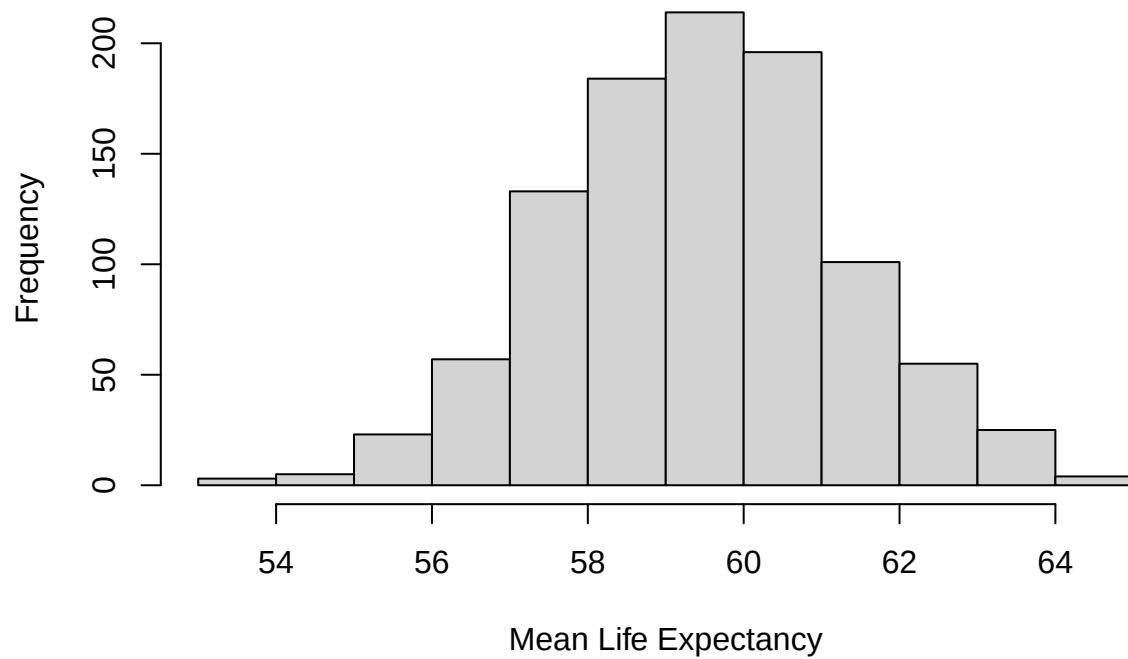
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

data("gapminder")
life_data <- gapminder$lifeExp

# Generate 1000 means from samples of size 50 (changed from 100)
sample_means <- replicate(1000, mean(sample(life_data, size = 50, replace = TRUE)))

hist(sample_means,
      main = "Sample Means (n=50)",
      xlab = "Mean Life Expectancy")
```

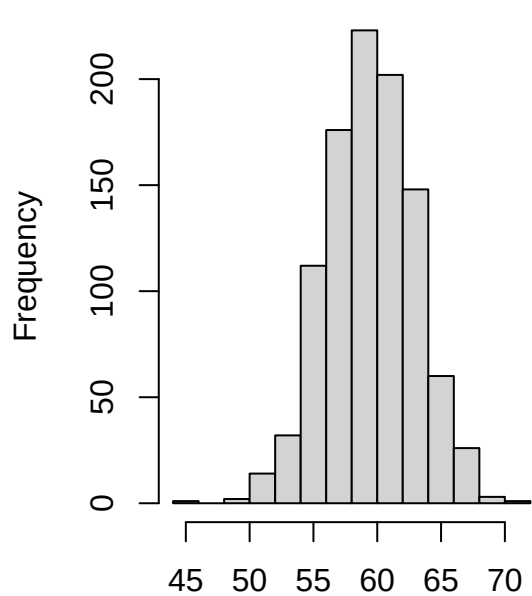
Sample Means (n=50)



```
means_15 <- replicate(1000, mean(sample(life_data, 15, replace = TRUE)))
means_75 <- replicate(1000, mean(sample(life_data, 75, replace = TRUE)))

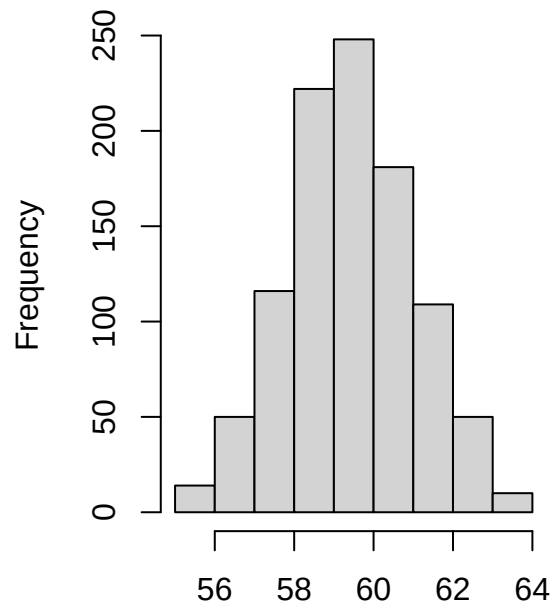
par(mfrow = c(1,2))
hist(means_15, main = "Sample Size 15")
hist(means_75, main = "Sample Size 75")
```

Sample Size 15



means_15

Sample Size 75



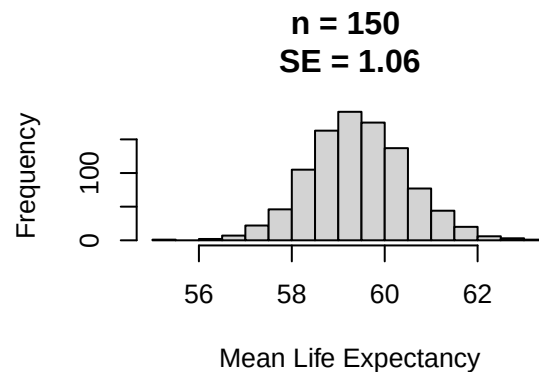
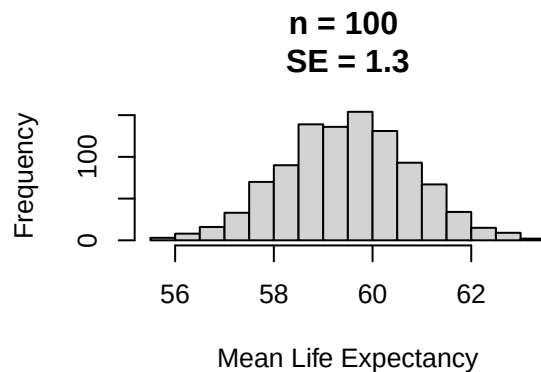
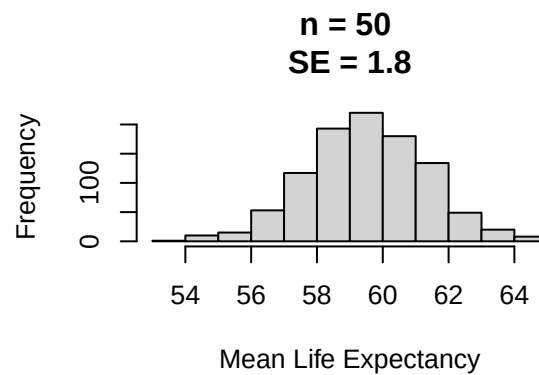
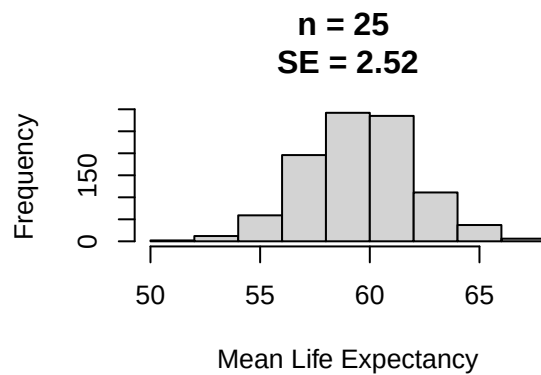
means_75

```
par(mfrow = c(1,1))

simulate_means <- function(n, reps = 1000) {
  replicate(reps, mean(sample(life_data, size = n, replace = TRUE)))
}

sample_sizes <- c(25, 50, 100, 150) # Adjusted sizes

par(mfrow = c(2,2))
for (n in sample_sizes) {
  sample_dist <- simulate_means(n)
  se <- sd(sample_dist)
  hist(sample_dist,
        main = paste("n =", n, "\nSE =", round(se, 2)),
        xlab = "Mean Life Expectancy")
}
```



```
par(mfrow = c(1,1))
```

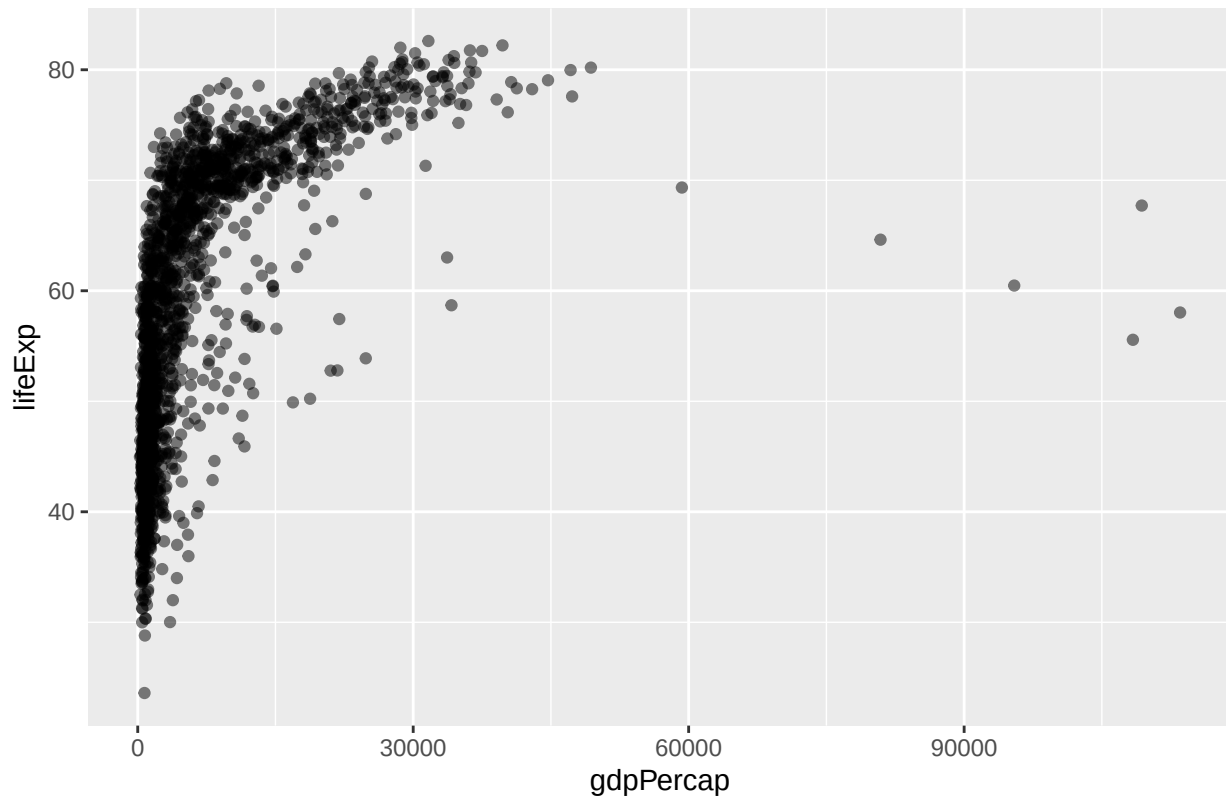
```
install.packages("ggplot2", repos='http://cran.us.r-project.org', quiet=TRUE)
library(ggplot2)
```

```
summary(gapminder)
```

```
##      country      continent      year      lifeExp
## Afghanistan: 12 Africa :624 Min. :1952 Min. :23.60
## Albania : 12 Americas:300 1st Qu.:1966 1st Qu.:48.20
## Algeria : 12 Asia :396 Median :1980 Median :60.71
## Angola : 12 Europe :360 Mean :1980 Mean :59.47
## Argentina : 12 Oceania : 24 3rd Qu.:1993 3rd Qu.:70.85
## Australia : 12 Max. :2007 Max. :82.60
## (Other) :1632
##      pop      gdpPercap
## Min. :6.001e+04 Min. : 241.2
## 1st Qu.:2.794e+06 1st Qu.: 1202.1
## Median :7.024e+06 Median : 3531.8
## Mean :2.960e+07 Mean : 7215.3
## 3rd Qu.:1.959e+07 3rd Qu.: 9325.5
## Max. :1.319e+09 Max. :113523.1
##
```

```
ggplot(gapminder, aes(x = gdpPercap, y = lifeExp)) +
  geom_point(alpha = 0.5) +
  labs(title = "GDP per Capita vs Life Expectancy")
```

GDP per Capita vs Life Expectancy



```
pearson_corr <- cor(gapminder$lifeExp, gapminder$gdpPercap, method = "pearson")
spearman_corr <- cor(gapminder$lifeExp, gapminder$gdpPercap, method = "spearman")

pearson_corr

## [1] 0.5837062

spearman_corr

## [1] 0.8264712

install.packages("ggpubr", repos='http://cran.us.r-project.org', quiet=TRUE)
library(ggpubr)

ggscatter(gapminder, x = "gdpPercap", y = "lifeExp", add = "reg.line", conf.int = TRUE)
```

