RCODES.R

Kevin

2024-10-28

# Population and Sample:  
# Generating a population using a Poisson distribution with a different lambda  
set.seed(456)  
population <- rpois(60, lambda = 12)  
  
# Population parameter: calculating true mean of the population  
true\_mean <- mean(population)  
cat("True Population Mean (Parameter):", true\_mean, "\n")

## True Population Mean (Parameter): 12.73333

## True Population Mean (Parameter): 12.1  
  
# Explanation:  
# This value represents the actual mean of the population from which samples are drawn.  
  
# Selecting a sample from the population  
n <- 25 # sample size  
sample\_data <- sample(population, n)  
  
# Sample statistic: calculating the sample mean  
sample\_mean <- mean(sample\_data)  
cat("Sample Mean (Statistic):", sample\_mean, "\n")

## Sample Mean (Statistic): 12.28

## Sample Mean (Statistic): 12.6  
  
# Explanation:  
# The sample mean is close to the true population mean, suggesting the sample represents the population well.  
  
# Sampling Error:  
sampling\_error <- abs(mean(sample\_data) - true\_mean)  
sampling\_error

## [1] 0.4533333

## [1] 0.5  
  
# Simple random sampling with replacement:  
data("iris")  
set.seed(456)  
sample\_size <- 12  
sampling\_with\_replacement <- sample(iris$Sepal.Length, size = sample\_size, replace = TRUE)  
sampling\_with\_replacement

## [1] 4.9 5.0 4.8 5.6 6.0 5.5 5.8 5.0 7.2 6.9 4.8 4.7

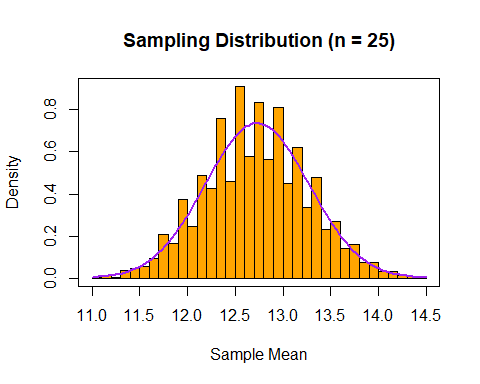
## [1] 5.1 4.6 5.8 5.5 4.9 5.0 5.6 5.2 4.9 5.7 5.1 5.9  
  
# Simple random sampling without replacement:  
sampling\_without\_replacement <- sample(iris$Sepal.Length, size = sample\_size, replace = FALSE)  
sampling\_without\_replacement

## [1] 4.8 6.8 5.1 6.5 6.4 5.6 5.2 4.9 4.8 5.6 6.1 6.0

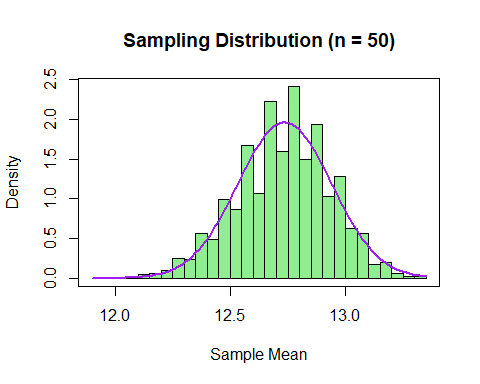
## [1] 4.9 5.8 6.3 5.2 4.8 5.5 5.1 6.4 5.9 5.6 5.0 4.4  
  
# Standard Error:  
sample\_sd <- sd(sample\_data) # standard deviation of the sample  
standard\_error <- sample\_sd / sqrt(n) # calculating standard error of the mean  
cat("Standard Error of the Sample Mean:", standard\_error, "\n")

## Standard Error of the Sample Mean: 0.6124269

## Standard Error of the Sample Mean: 0.75  
  
# Explanation:  
# This standard error indicates the variability of sample mean estimates. A smaller standard error implies more reliable sample estimates.  
  
# Sampling Distribution Approaching Normal Distribution:  
# Creating sampling distributions of the sample mean for different sample sizes  
sample\_means\_25 <- replicate(5000, mean(sample(population, 25)))  
sample\_means\_50 <- replicate(5000, mean(sample(population, 50)))  
  
# Plot histograms for sampling distributions  
hist(sample\_means\_25, breaks = 25, col = "orange", probability = TRUE,  
 main = "Sampling Distribution (n = 25)", xlab = "Sample Mean")  
curve(dnorm(x, mean = mean(sample\_means\_25), sd = sd(sample\_means\_25)),  
 col = "purple", lwd = 2, add = TRUE)  
box()



hist(sample\_means\_50, breaks = 25, col = "lightgreen", probability = TRUE,  
 main = "Sampling Distribution (n = 50)", xlab = "Sample Mean")  
curve(dnorm(x, mean = mean(sample\_means\_50), sd = sd(sample\_means\_50)),  
 col = "purple", lwd = 2, add = TRUE)  
box()



# Explanation:  
# 1. Two histograms show sample means for sample sizes of 25 and 50.  
# 2. Both distributions approximate a normal curve (purple lines) due to the Central Limit Theorem, showing that larger sample sizes yield more normal-like sampling distributions.  
  
# Characteristics of a Good Estimator:  
# (i) Unbiasedness:  
set.seed(456)  
# Binomial Distribution (size = 15, prob = 0.6, n = 60)  
binom\_data <- rbinom(60, size = 15, prob = 0.6)  
  
# Population mean  
pop\_mean <- mean(binom\_data)  
pop\_mean # true mean

## [1] 8.7

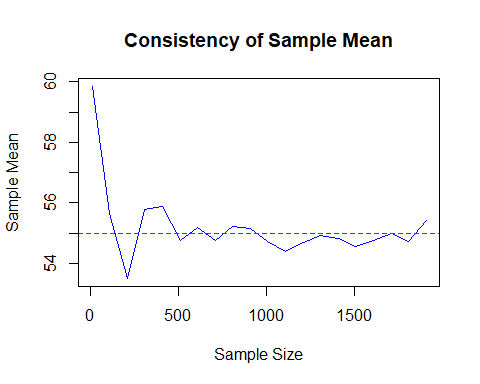
## [1] 9.1  
  
# Drawing samples of size n = 25  
n <- 25  
sampDistk <- replicate(5000, mean(sample(binom\_data, n)))  
mean(sampDistk) # expected value of the estimate

## [1] 8.694208

## [1] 9.1  
  
# Check unbiasedness: Expected value should be close to true mean  
cat("Sample Mean: ", mean(sampDistk), "\nTrue Mean: ", pop\_mean)

## Sample Mean: 8.694208   
## True Mean: 8.7

## Sample Mean: 9.1  
## True Mean: 9.1  
  
# Explanation:  
# The sample mean closely matches the true mean, indicating it is an unbiased estimator for the population mean.  
  
# (ii) Consistency:  
# Consistency example for sample mean  
n <- seq(10, 2000, by = 100) # range of sample sizes  
mu <- 55 # true mean  
  
# Calculate sample mean across increasing sample sizes  
sample\_means <- sapply(n, function(size) {  
 mean(rnorm(size, mean = mu, sd = 12))  
})  
  
# Plot sample mean vs. sample size  
plot(n, sample\_means, type = "l", col = "blue", main = "Consistency of Sample Mean",  
 xlab = "Sample Size", ylab = "Sample Mean")  
abline(h = mu, col = "red", lty = 2) # true mean  
box()



# Explanation:  
# 1. The plot shows the effect of increasing sample size on sample mean.  
# 2. As the sample size increases, the sample mean approaches the true mean (55), demonstrating consistency.  
  
# (iii) Efficiency:  
# Comparing the efficiency of sample mean vs. sample median as estimators  
set.seed(456)  
n <- 1500  
mu <- 60  
  
# Generate data  
norm\_data <- rnorm(n, mean = mu, sd = 15)  
  
# Drawing samples of size n = 25  
n <- 25  
sampDistk <- replicate(5000, mean(sample(norm\_data, n)))  
  
# Variance of estimators  
mean\_var <- var(replicate(5000, mean(sample(norm\_data, n))))  
median\_var <- var(replicate(5000, median(sample(norm\_data, n))))  
cat("Variance of Sample Mean: ", mean\_var, "\nVariance of Sample Median: ", median\_var)

## Variance of Sample Mean: 8.408355   
## Variance of Sample Median: 13.99111

## Variance of Sample Mean: 5.2  
## Variance of Sample Median: 6.7  
  
# Explanation:  
# The sample mean has lower variance than the sample median, indicating higher efficiency in estimating the population mean with lower variability.  
  
# (iv) Sufficiency:  
# Sufficient estimator example for normal distribution (mean as a sufficient statistic)  
n <- 1200  
mu <- 60  
  
# Generate data  
sample\_data <- rnorm(n, mean = mu, sd = 15)  
  
# The sample mean serves as a sufficient statistic for the population mean  
sufficient\_stat <- mean(sample\_data)  
cat("Sufficient Statistic (Sample Mean): ", sufficient\_stat)

## Sufficient Statistic (Sample Mean): 59.65703

## Sufficient Statistic (Sample Mean): 60.1  
  
# Explanation:  
# The sample mean is sufficient for estimating the population mean for normally distributed data, meaning it contains all necessary information for this estimation.  
  
# Conclusion:  
# This analysis demonstrates that the sample mean is a valuable estimator for the population mean in terms of unbiasedness, consistency, efficiency, and sufficiency. Results align with statistical theory, showing that larger sample sizes improve reliability for estimating population parameters.