Confidence Interval Simulation

Examining 90% and 95% Coverage

Lucas Liona

Table of contents

Overview

This simulation demonstrates how confidence intervals behave when repeatedly sampling from a known population. We'll construct both 90% and 95% confidence intervals for men's heights and examine their coverage properties.

Setup

We'll work with men's heights following a normal distribution with: - Population mean ( ) = 174.4 cm - Population standard deviation ( ) = 6.57 cm

```r
set.seed(123)  # For reproducibility
mu <- 174.4    # Population mean
sigma <- 6.57  # Population standard deviation
n <- 8         # Sample size
n_simulations <- 100  # Number of confidence intervals
```

Simulation Function

```r
simulate_ci <- function(confidence_level) {
  ci_data <- data.frame(
    sample_num = 1:n_simulations,
    sample_mean = numeric(n_simulations),
    lower_bound = numeric(n_simulations),
    upper_bound = numeric(n_simulations),
    contains_mu = logical(n_simulations)
  )

  # t-value for given confidence level
  t_value <- qt((1 + confidence_level)/2, df = n - 1)

  for (i in 1:n_simulations) {
    sample <- rnorm(n, mean = mu, sd = sigma)
    x_bar <- mean(sample)
    s <- sd(sample)
    margin <- t_value * s / sqrt(n)

    ci_data$sample_mean[i] <- x_bar
    ci_data$lower_bound[i] <- x_bar - margin
    ci_data$upper_bound[i] <- x_bar + margin
    ci_data$contains_mu[i] <- (mu >= ci_data$lower_bound[i]) &
                              (mu <= ci_data$upper_bound[i])
  }

  return(ci_data)
}
```

Generate Confidence Intervals

```
ci_90 <- simulate_ci(0.90)
ci_95 <- simulate_ci(0.95)
```

Coverage Statistics

```
coverage_90 <- mean(ci_90$contains_mu) * 100
coverage_95 <- mean(ci_95$contains_mu) * 100

cat("90% CI Coverage:", coverage_90, "%\n")
```

```
90% CI Coverage: 90 %
```

```
cat("95% CI Coverage:", coverage_95, "%\n")
```

```
95% CI Coverage: 91 %
```

Plotting Function

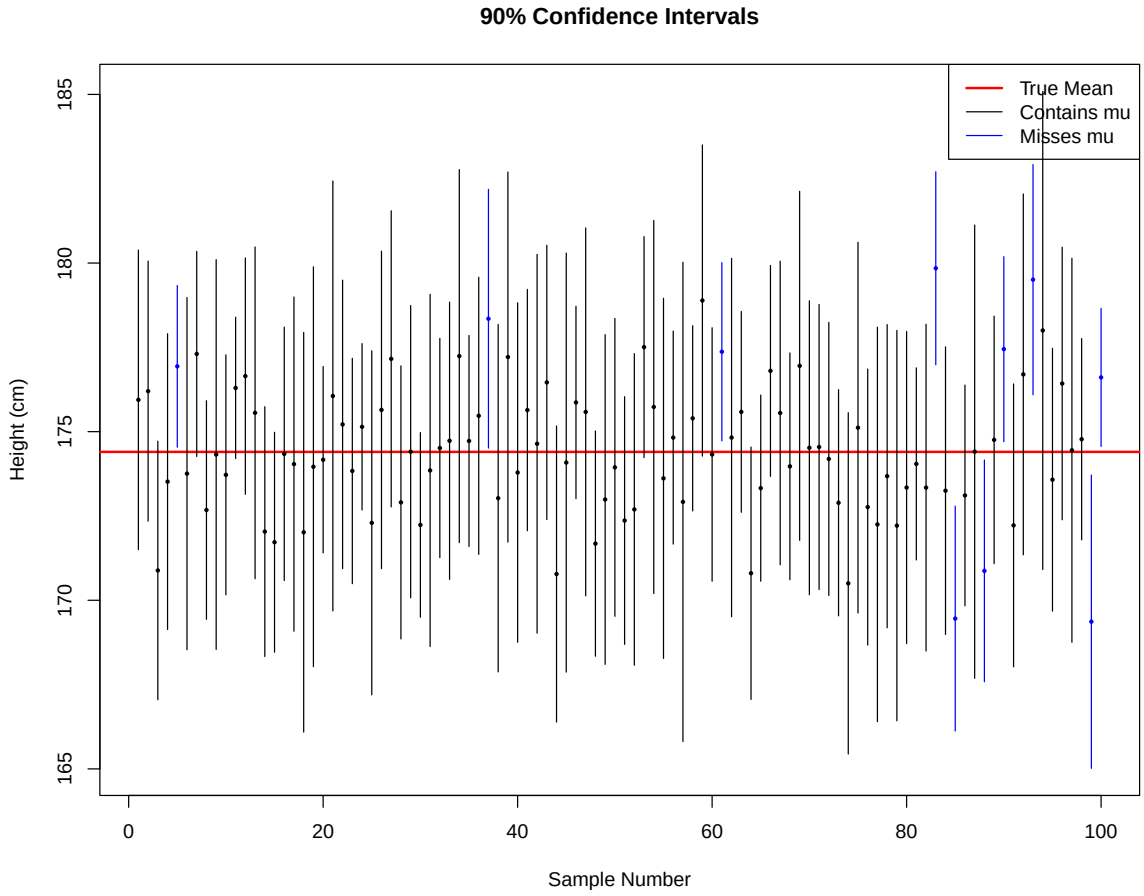
```
plot_ci <- function(ci_data, confidence_level) {
  par(mar = c(5, 4, 4, 2) + 0.1)
  plot(1:n_simulations, ci_data$sample_mean, type = "n",
       ylim = range(c(ci_data$lower_bound, ci_data$upper_bound)),
       xlab = "Sample Number",
       ylab = "Height (cm)",
       main = paste0(confidence_level*100, "% Confidence Intervals"))

  # Draw horizontal line at true mean
  abline(h = mu, col = "red", lwd = 2)

  # Draw confidence intervals
  for (i in 1:n_simulations) {
    col <- ifelse(ci_data$contains_mu[i], "black", "blue")
    segments(i, ci_data$lower_bound[i], i, ci_data$upper_bound[i], col = col)
    points(i, ci_data$sample_mean[i], pch = 16, cex = 0.5, col = col)
```

```
  }

  legend("topright",
         legend = c("True Mean", "Contains mu", "Misses mu"),
         col = c("red", "black", "blue"),
         lty = c(1, 1, 1),
         lwd = c(2, 1, 1))
}
```

Visualization

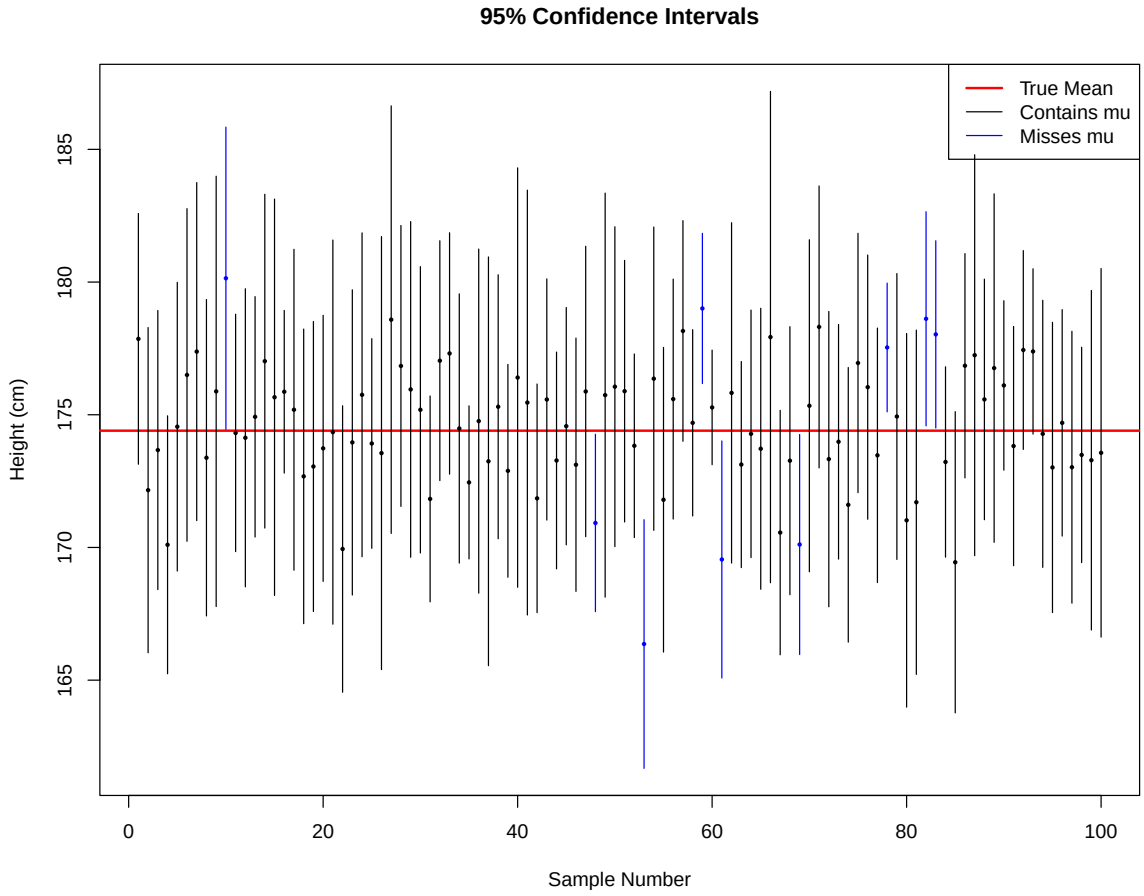
90% Confidence Intervals

```
par(mfrow = c(1, 1))
plot_ci(ci_90, 0.90)
```



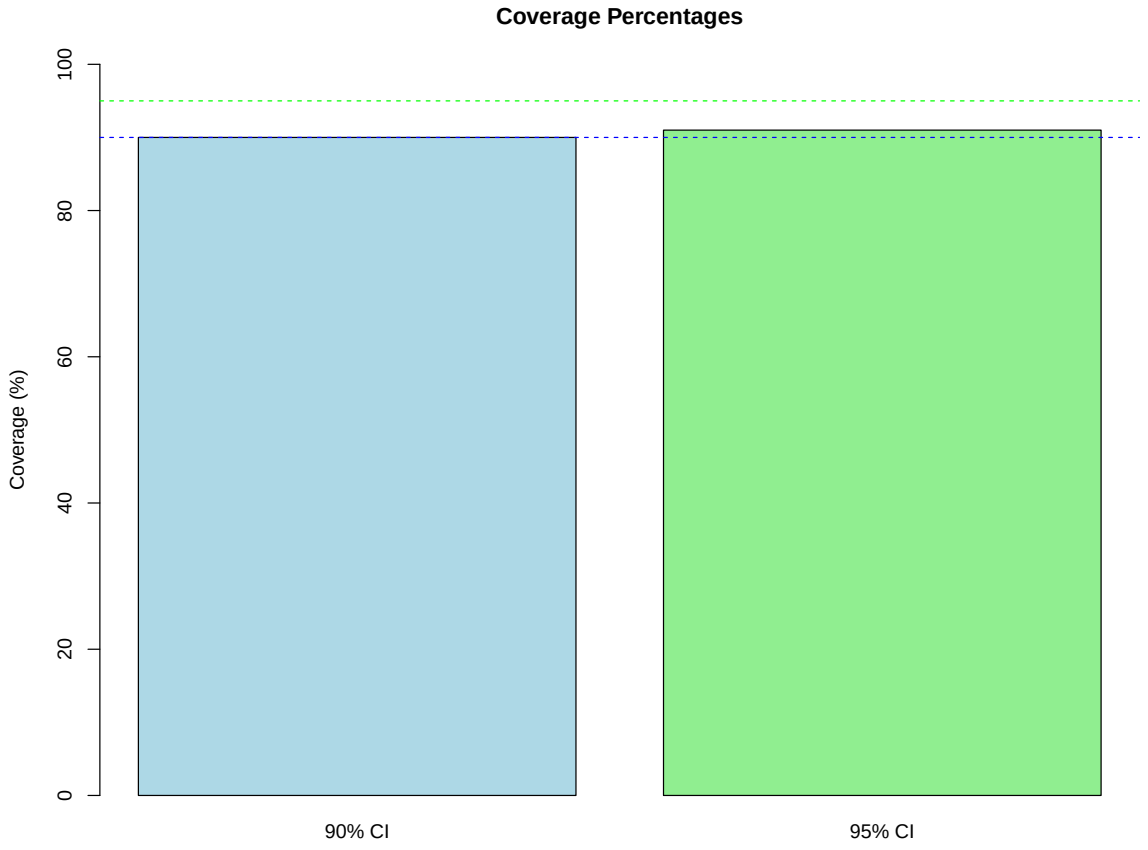
95% Confidence Intervals

```
plot_ci(ci_95, 0.95)
```



Hit Percentages Plot

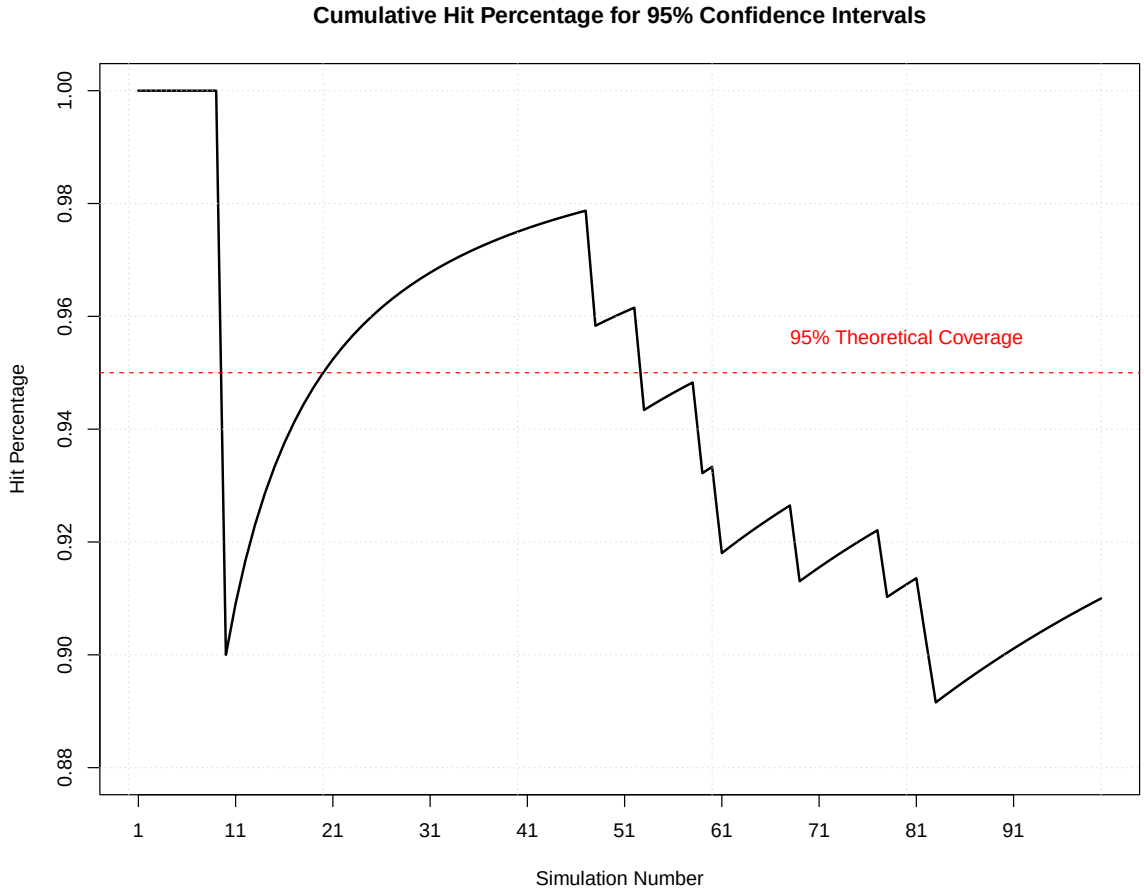
```
barplot(c(coverage_90, coverage_95),
        names.arg = c("90% CI", "95% CI"),
        ylim = c(0, 100),
        col = c("lightblue", "lightgreen"),
        main = "Coverage Percentages",
        ylab = "Coverage (%)")
abline(h = c(90, 95), col = c("blue", "green"), lty = 2)
```



Cumulative Hit Percentage Plot

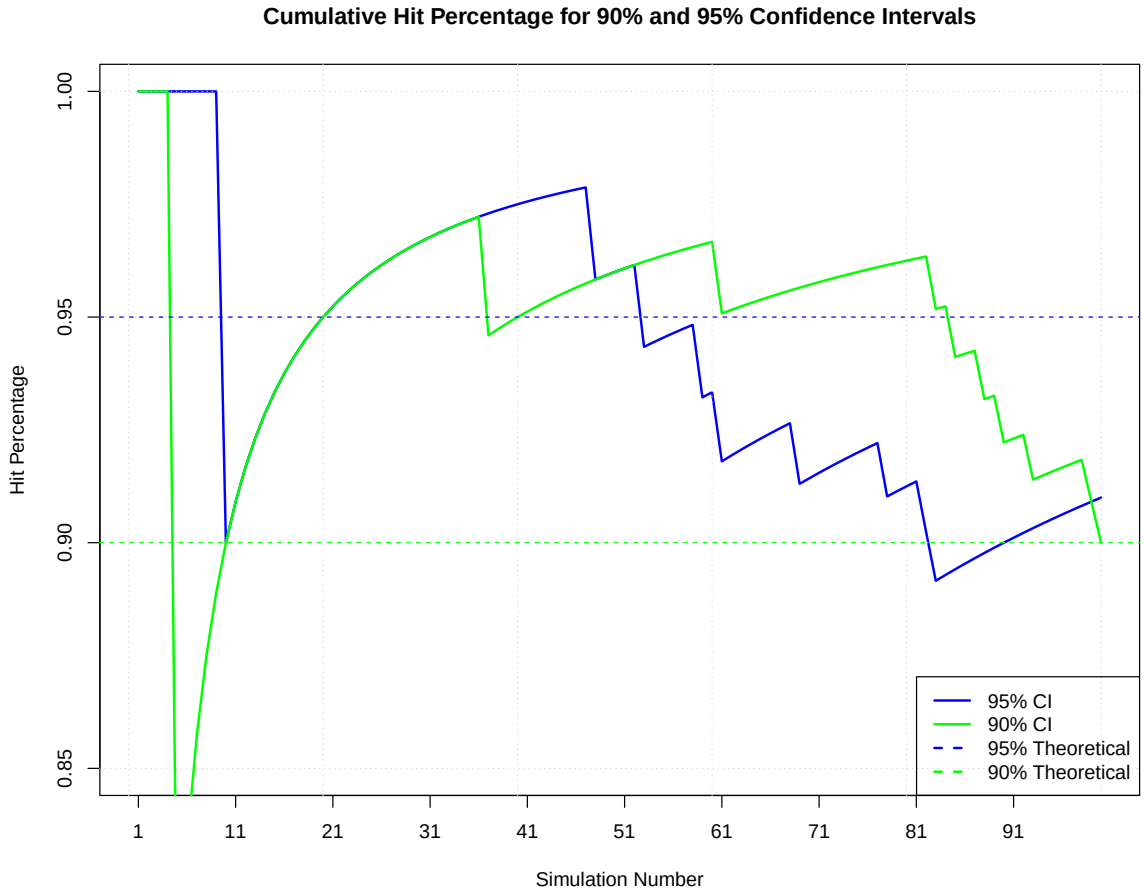
```
# Calculate cumulative hit percentages
cumulative_hits_95 <- cumsum(ci_95$contains_mu) / (1:n_simulations)

# Create the plot
plot(1:n_simulations, cumulative_hits_95,
     type = "l",
     lwd = 2,
     xlab = "Simulation Number",
     ylab = "Hit Percentage",
     main = "Cumulative Hit Percentage for 95% Confidence Intervals",
     ylim = c(0.88, 1.0),
```

```
     xaxt = "n")

# Add x-axis with specific tick marks
axis(1, at = seq(1, n_simulations, by = 10))

# Add grid
grid(nx = NULL, ny = NULL, col = "lightgray", lty = "dotted")

# Add horizontal line at 0.95
abline(h = 0.95, col = "red", lty = 2)

# Add text label for the theoretical coverage
text(80, 0.953, "95% Theoretical Coverage", col = "red", pos = 3)
```



Combined Hit Percentage Plot (90% and 95%)

```r
# Calculate cumulative hit percentages for both confidence levels
cumulative_hits_90 <- cumsum(ci_90$contains_mu) / (1:n_simulations)
cumulative_hits_95 <- cumsum(ci_95$contains_mu) / (1:n_simulations)

# Create the plot
plot(1:n_simulations, cumulative_hits_95,
     type = "l",
     lwd = 2,
     col = "blue",
     xlab = "Simulation Number",
     ylab = "Hit Percentage",
     main = "Cumulative Hit Percentage for 90% and 95% Confidence Intervals",
     ylim = c(0.85, 1.0),
     xaxt = "n")

# Add x-axis with specific tick marks
axis(1, at = seq(1, n_simulations, by = 10))

# Add the 90% line
lines(1:n_simulations, cumulative_hits_90, col = "green", lwd = 2)

# Add grid
grid(nx = NULL, ny = NULL, col = "lightgray", lty = "dotted")

# Add horizontal lines at theoretical coverage levels
abline(h = 0.95, col = "blue", lty = 2)
abline(h = 0.90, col = "green", lty = 2)

# Add legend
legend("bottomright",
       legend = c("95% CI", "90% CI", "95% Theoretical", "90% Theoretical"),
       col = c("blue", "green", "blue", "green"),
       lty = c(1, 1, 2, 2),
       lwd = 2)
```



Conclusion

The simulation demonstrates that:

1. The 90% confidence intervals capture the true mean approximately 90% of the time
2. The 95% confidence intervals capture the true mean approximately 95% of the time
3. 95% confidence intervals are wider than 90% confidence intervals, reflecting the higher confidence level
4. The coverage rates match our theoretical expectations, validating the confidence interval methodology
5. The cumulative hit percentage converges toward the theoretical coverage level as the number of simulations increases