Handout #4 (Step 2)

André Ramolivaz

2025-02-17

Introduction

Analysis of the 30 water-fill time measurement and check if they are approximately normally distributed. Step included::

- Load the data into R and compute summary statistics (mean, standard deviation).
- Plot a histogram with an overlaid normal curve.
- Create the Q–Q plot.
- Run normality tests (Shapiro–Wilk).

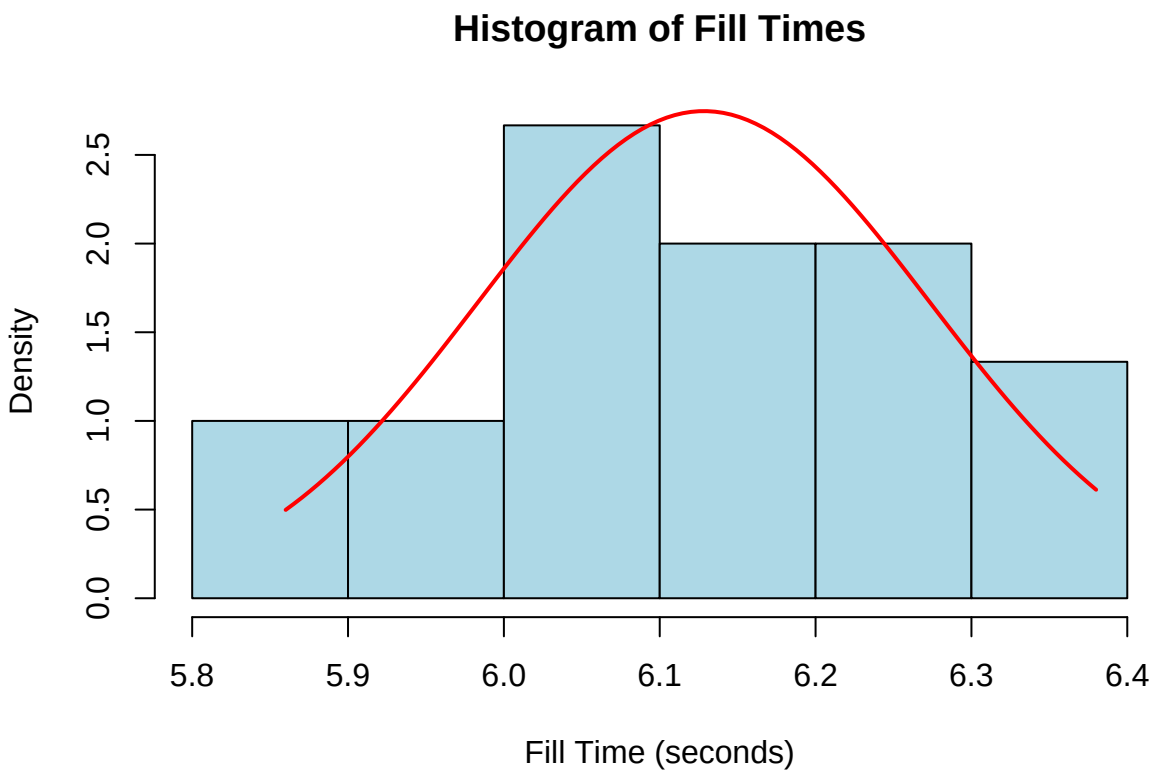
```
# 2. Load the Data
fill_times <- c(
  6.16, 6.15, 6.18, 6.28, 6.17,
  5.88, 5.98, 5.86, 6.07, 5.96,
  6.08, 6.38, 6.26, 6.07, 6.10,
  6.17, 6.22, 6.29, 6.23, 5.99,
  6.02, 6.34, 6.01, 6.31, 6.16,
  6.03, 6.34, 6.25, 6.02, 5.89
)
# 3. Compute summary statistics
mean_fill <- mean(fill_times)
sd_fill <- sd(fill_times)

cat("Sample Mean:", mean_fill, "\n")
```

```
## Sample Mean: 6.128333
```

```
cat("Sample Standard Deviation:", sd_fill, "\n")
```

```
## Sample Standard Deviation: 0.1452248
```

```
# Histogram
hist(fill_times,
     prob = TRUE,
     col = "lightblue",
     main = "Histogram of Fill Times",
     xlab = "Fill Time (seconds)")

# Normal curve overlay
curve(dnorm(x, mean = mean_fill, sd = sd_fill),
      from = min(fill_times), to = max(fill_times),
      col = "red", lwd = 2, add = TRUE)
```




```r
par(mfrow = c(1, 2))

# Sort data
z <- sort(fill_times)
n <- length(z)

# Empirical probabilities (using i/(n+1) to avoid 0 and 1)
p_emp <- (1:n) / (n + 1)

# Theoretical probabilities from the Normal CDF
p_theory <- pnorm(z, mean = mean_fill, sd = sd_fill)

# Plot the P-P graph: p_theory on the x-axis, p_emp on the y-axis
plot(
  p_theory, p_emp,
  xlab = "Theoretical Probabilities (Normal)",
  ylab = "Empirical Probabilities (Data)",
  pch  = 19,
  col  = "blue",
  main = "P-P Plot for Fill Times"
)
```

```
## Warning in title(...): conversion failure on 'P-P Plot for Fill Times' in
## 'mbcsToSbcs': dot substituted for <e2>

## Warning in title(...): conversion failure on 'P-P Plot for Fill Times' in
## 'mbcsToSbcs': dot substituted for <80>

## Warning in title(...): conversion failure on 'P-P Plot for Fill Times' in
## 'mbcsToSbcs': dot substituted for <93>
```

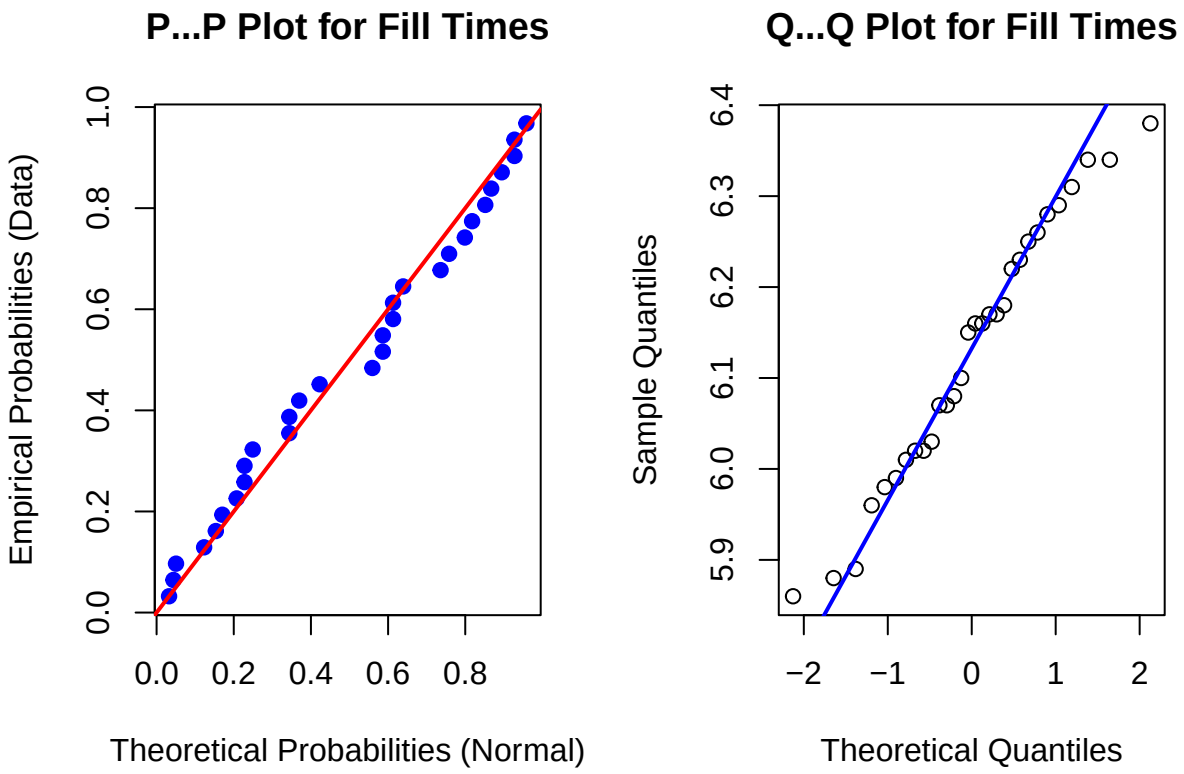
```r
abline(0, 1, col = "red", lwd = 2)

# 5. Q-Q Plot to assess normality
qqnorm(fill_times,
       main = "Q-Q Plot for Fill Times")
```

```
## Warning in title(...): conversion failure on 'Q-Q Plot for Fill Times' in
## 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in title(...): conversion failure on 'Q-Q Plot for Fill Times' in
## 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in title(...): conversion failure on 'Q-Q Plot for Fill Times' in
## 'mbcsToSbcs': dot substituted for <93>
```

```r
qqline(fill_times, col = "blue", lwd = 2)
```




```r
# 6. Normality Tests

# Shapiro-Wilk test
shapiro_result <- shapiro.test(fill_times)
cat("Shapiro-Wilk Test p-value:", shapiro_result$p.value, "\n")
```

```
## Shapiro-Wilk Test p-value: 0.523527
```

```r
# (Optional) Kolmogorov-Smirnov (using standardized data)
ks_result <- ks.test(scale(fill_times), "pnorm", mean=0, sd=1)
```

```
## Warning in ks.test.default(scale(fill_times), "pnorm", mean = 0, sd = 1): ties
## should not be present for the Kolmogorov-Smirnov test
```

```r
cat("K-S Test p-value:", ks_result$p.value, "\n")
```

```
## K-S Test p-value: 0.959033
```

```r
#Additional tests
```