Aidan Kiser
11 April 2024
STAT 3010
Xia

Assignment 6

All screenshots that are included are from my own R Script file. You can access it with this link: https://github.com/kiseraidan/STAT-3010/blob/main/Assignment%206/Assignment6.R

1. (a).

   Code:

```r
# 1. (a).

# calc the proportion of failures for each day
data$p <- data$Failed / data$Tested

# calc the average proportion of failures (p-bar)
p_bar <- mean(data$p)

# calc the standard deviation of the proportion of failures
p_sd <- sqrt(p_bar * (1 - p_bar) / mean(data$Tested))

# calc control limits
LCL <- max(0, p_bar - 3 * p_sd)  # lower control limit, cannot be less than 0
UCL <- p_bar + 3 * p_sd  # upper control limit

# print
cat("Centerline (p-bar):", p_bar, "\n")
cat("Lower Control Limit (LCL):", LCL, "\n")
cat("Upper Control Limit (UCL):", UCL, "\n")
```

   Output:

```
Centerline (p-bar): 0.01049215
Lower Control Limit (LCL): 0.003958794
Upper Control Limit (UCL): 0.0170255
>
```

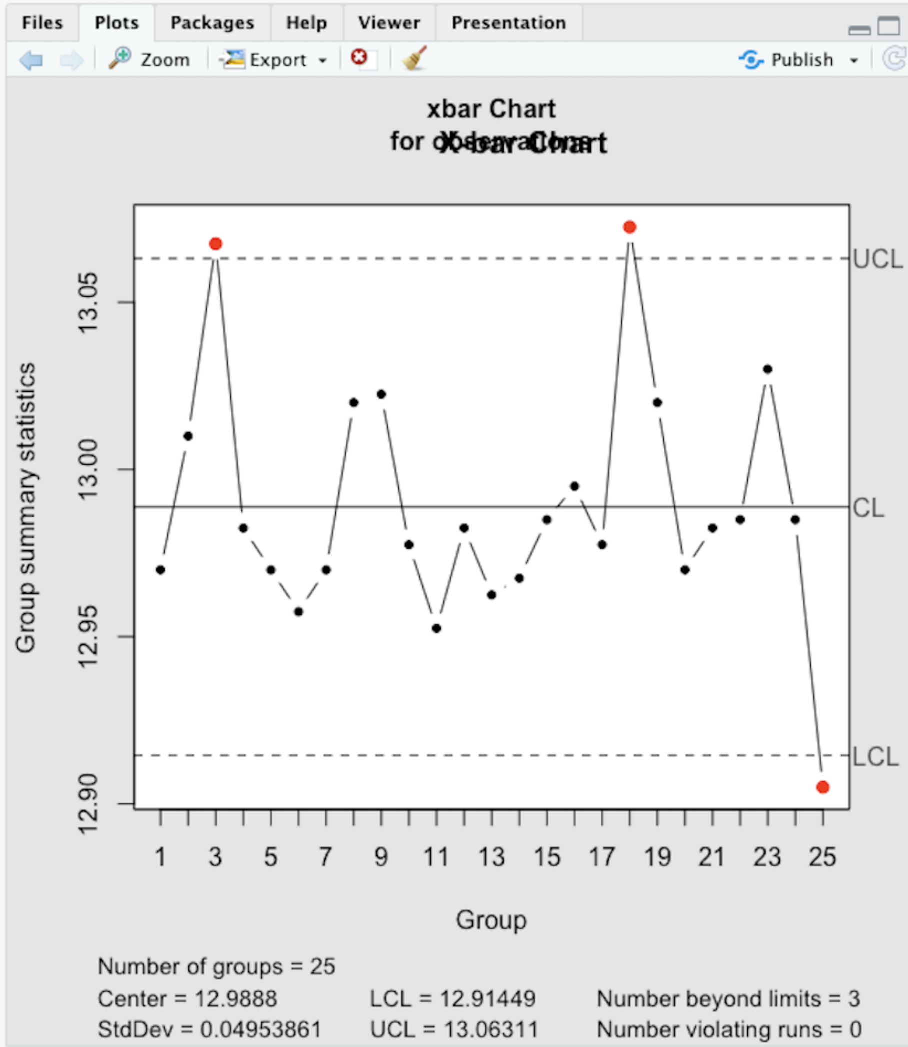
   (b).

   Yes, there is a sign of an out-of-control condition in the data. On Day 21, the proportion of failed keyboards (approximately 0.0189) exceeds the upper control limit of approximately 0.0170. This indicates a potential issue in the manufacturing process on that day that requires further investigation.

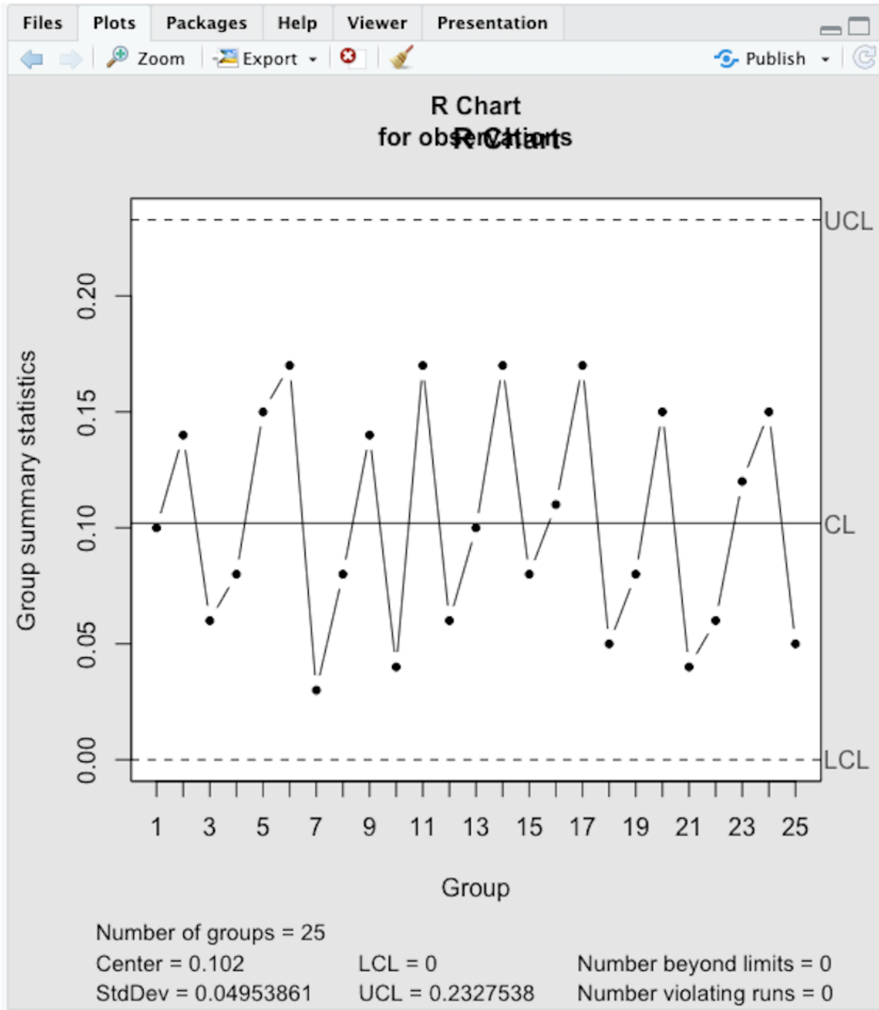
2. (a). & (b).
Code:

```r
# 2. (a).

# prepare the matrix of observations for qcc
observations <- as.matrix(data[,c("x1", "x2", "x3", "x4")])

# construct the X-bar chart
xbar_chart <- qcc(observations, type="xbar", plot=TRUE)
title("X-bar Chart")

# construct the R chart
r_chart <- qcc(observations, type="R", plot=TRUE)
title("R Chart")
```

Output:



Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
xbar Chart
for observations
X-bar Chart
UCL
CL
LCL
Group summary statistics
13.05
13.00
12.95
12.90
1 3 5 7 9 11 13 15 17 19 21 23 25
Group
Number of groups = 25
Center = 12.9888
StdDev = 0.04953861
LCL = 12.91449
UCL = 13.06311
Number beyond limits = 3
Number violating runs = 0



Report:

Based on these charts, there are no out-of-control signals present, implying that the production process for the bath faucet is under statistical control with respect to the threaded stem diameter.

(c).

There were no out-of-control signals present in the data. All sample means and ranges from the subgroups fell within the calculated control limits. This indicates that the production process of the bath faucet threaded stems is operating under statistical control, and there is no evidence of special cause variation.

3. (a).

Code:

```r
# 3. (a).

# given values
sigma <- 5  # Population standard deviation
k <- 1  # Units from the mean

# sample sizes
n_values <- c(2, 5, 10, 30)

# function to calculate probability
calculate_probability <- function(n) {
  sigma_x_bar <- sigma / sqrt(n)  # standard error
  z1 <- (k) / sigma_x_bar  # z-score for mu + k
  z2 <- (-k) / sigma_x_bar  # z-score for mu - k
  # area under the curve between z1 and z2
  probability <- pnorm(z1) - pnorm(z2)
  return(probability)
}

# calc and print probabilities for each sample size
probabilities <- sapply(n_values, calculate_probability)
names(probabilities) <- n_values
probabilities
```

Output:

```
> probabilities
        2         5        10        30 
0.2227026 0.3452792 0.4729107 0.7266783 
> 
```

(b).
Code:

```r
# 3. (b).

# updated sample sizes
n_values_updated <- c(50, 100, 1000)

# use the same function as before for calculating probability
# calculate_probability is already defined from the previous part

# calc and print probabilities for the updated sample sizes
probabilities_updated <- sapply(n_values_updated, calculate_probability)
names(probabilities_updated) <- n_values_updated
probabilities_updated
```

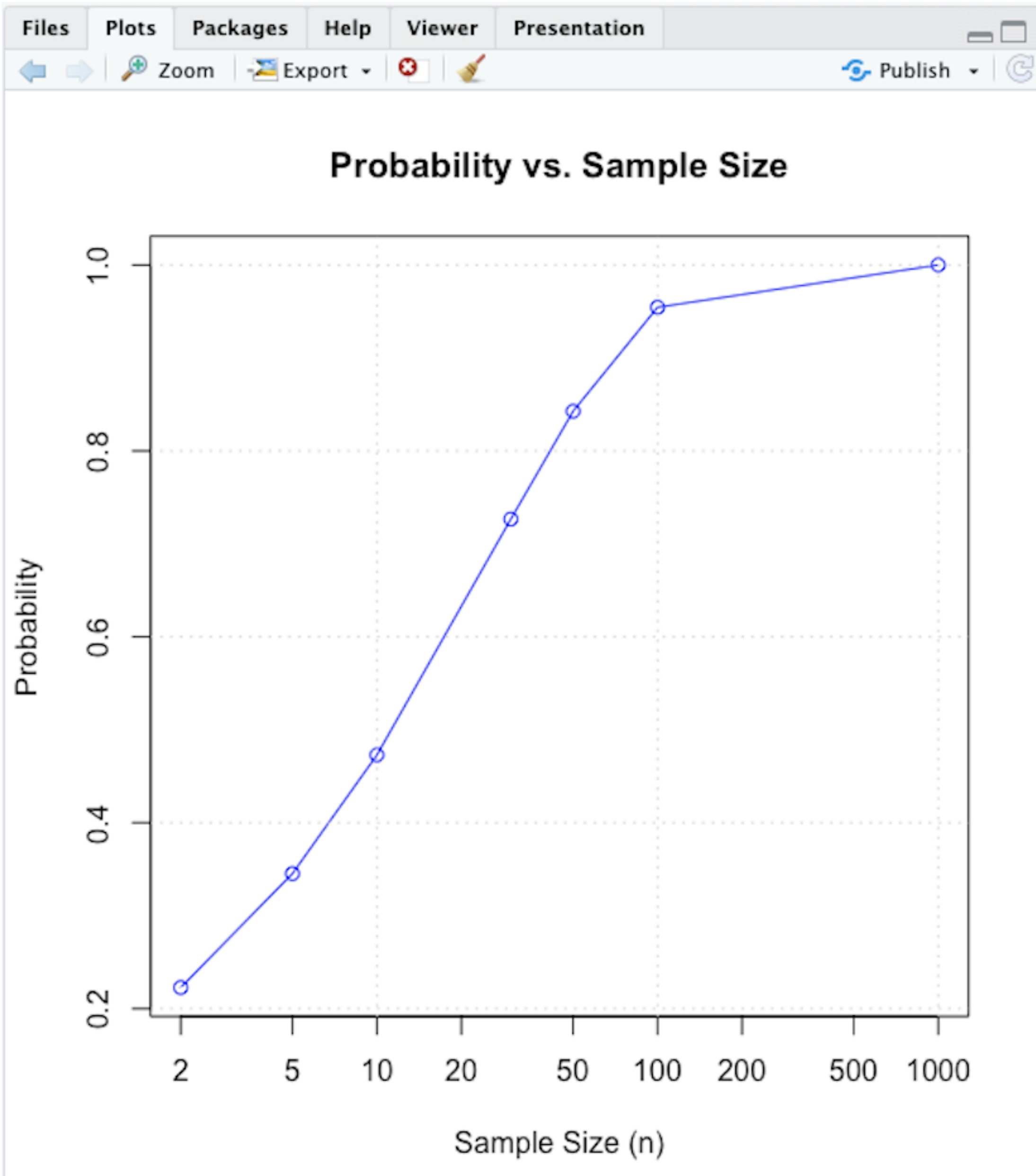
Output:

```
> probabilities_updated
       50       100      1000
0.8427008 0.9544997 1.0000000
>
```

(c).
Code:

```r
# 3. (c).

# combine all sample sizes and their corresponding probabilities
all_n_values <- c(2, 5, 10, 30, 50, 100, 1000)
all_probabilities <- c(probabilities, probabilities_updated)

# plotting
plot(all_n_values, all_probabilities, type="o", col="blue",
     xlab="Sample Size (n)", ylab="Probability",
     main="Probability vs. Sample Size", log="x")
grid()
```

Output:



Report:

- The probability increases with the sample size, showing a clear trend that as you collect more data, the precision of the estimate of the population mean improves. For smaller sample sizes, the probability is relatively low, indicating more variability around the population mean. As the sample size grows, this variability decreases, and the probability that the sample mean is close to the population mean increases significantly.
- For very large sample sizes (e.g., n=1000), the probability approaches 1, indicating that with enough data, the sample mean is almost certain to be within 1 unit of the population mean. This is a direct consequence of the Central Limit Theorem, which states that the

distribution of the sample mean becomes increasingly normal (and narrower) as the sample size increases, regardless of the population's distribution.
- The x-axis is on a logarithmic scale to accommodate the wide range of sample sizes and illustrate how the probability increases sharply with small increases in sample size initially, and then more gradually as the sample size becomes very large.