Dr Isabella Gollini

Probability and Statistics

Solutions 4

1. (i)

$$\mu \equiv E(X) = \int_0^\theta \frac{2x^2}{\theta^2} dx = \left[\frac{2x^3}{3\theta^2}\right]_0^\theta = \frac{2}{3}\frac{\theta^3}{\theta^2} = \frac{2}{3}\theta \ .$$

$$E(X^2) = \int_0^\theta \frac{2x^3}{\theta^2} dx = \left[\frac{2x^4}{4\theta^2}\right]_0^\theta = \frac{1}{2}\frac{\theta^4}{\theta^2} = \frac{1}{2}\theta^2 \ .$$

Hence

$$\sigma^2 = \frac{1}{2}\theta^2 - \left(\frac{2}{3}\theta\right)^2 = \frac{1}{18}\theta^2 \ .$$

(ii)

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(u)\, du \\ &= \begin{cases} 0 & (x \le 0) \\ \int_0^x 2u\, du/\theta^2 = x^2/\theta^2 & (0 \le x \le \theta) \\ 1 & (x \ge \theta) \end{cases} \end{aligned}$$

2. The corresponding standardized r.v. is $Z \equiv (X - 20)/2 \sim N(0, 1)$.

(i)

$$\Pr(X < 22) = \Pr\left(\frac{X - 20}{2} < \frac{22 - 20}{2}\right) = \Phi(1) = 0.8413$$

You can use any of the following `R` commands:

```
pnorm(22, mean = 20, sd = 2)

## [1] 0.8413447

pnorm(1)

## [1] 0.8413447
```

(ii)

$$\begin{aligned} \Pr(X < 19) &= \Pr\left(\frac{X - 20}{2} < \frac{19 - 20}{2}\right) \\ &= \Phi(-0.5) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085 \end{aligned}$$

You can use any of the following `R` commands:

```r
pnorm(19, mean = 20, sd = 2)
## [1] 0.3085375
pnorm(-0.5)
## [1] 0.3085375
1 - pnorm(0.5)
## [1] 0.3085375
```

(iii)

$$
\begin{aligned}
\Pr(X > 16) &= \Pr\left(\frac{X-20}{2} > \frac{16-20}{2}\right) \\
&= 1 - \Phi(-2) = \Phi(2) = 0.9772
\end{aligned}
$$

You can use any of the following `R` commands:

```r
pnorm(16, mean = 20, sd = 2, lower.tail = FALSE)
## [1] 0.9772499
1 - pnorm(16, mean = 20, sd = 2)
## [1] 0.9772499
1 - pnorm(-2)
## [1] 0.9772499
```

(iv)

$$
\begin{aligned}
\Pr(17 < X < 22) &= \Pr\left(\frac{17-20}{2} < \frac{X-20}{2} < \frac{22-20}{2}\right) \\
&= \Phi(1) - \Phi(-1.5) = \Phi(1) + \Phi(1.5) - 1 \\
&= 0.8413 + 0.9332 - 1 = 0.7745
\end{aligned}
$$

You can use any of the following `R` commands:

```r
pnorm(22, mean = 20, sd = 2) - pnorm(17, mean = 20, sd = 2)
## [1] 0.7745375
pnorm(1) - pnorm(-1.5)
## [1] 0.7745375
pnorm(1) + pnorm(1.5) - 1
## [1] 0.7745375
```

3.

$$\bar{X} \sim N\left(20, \frac{16}{36}\right) = N\left(20, \frac{4}{9}\right)$$

The corresponding standardized r.v. is

$$Z \equiv \frac{\bar{X} - 20}{2/3} \sim N(0, 1).$$

Hence

$$\begin{aligned}
\Pr(19 < \bar{X} < 21) &= \Pr\left(\frac{19 - 20}{2/3} < Z < \frac{21 - 20}{2/3}\right) \\
&= \Pr(-1.5 < Z < 1.5) \\
&= \Phi(1.5) - \Phi(-1.5) \ = \ \Phi(1.5) - (1 - \Phi(1.5)) \\
&= 2\Phi(1.5) - 1 \ = \ (2 \times 0.9332) - 1 \ = \ 0.866.
\end{aligned}$$

4.

$$\bar{X} \sim N\left(3320, \frac{660^2}{225}\right) = N\left(3320, 44^2\right)$$

The corresponding standardized r.v. is

$$Z \equiv \frac{\bar{X} - 3320}{44} \sim N(0, 1).$$

Hence

$$\begin{aligned}
\Pr(3223 < \bar{X} < 3407) &= \Pr\left(\frac{3223 - 3320}{44} < Z < \frac{3407 - 3320}{44}\right) \\
&= \Pr(-97/44 < Z < 87/44) \\
&= \Phi(1.977) - \Phi(-2.205) \ = \ \Phi(1.977) + \Phi(2.205) - 1 \\
&= 0.9760 + 0.9863 - 1 \ = \ 0.962.
\end{aligned}$$

5. (i) Using Table 5 of *Lindley and Scott*,

$$q_3 = x(25) = 0.6745.$$

Or, using `R`:

```
q3 <- qnorm(1 - 25 / 100)
```

By the symmetry of the distribution, $q_1 = -0.6745$. Hence the interquartile range is $q_3 - q_1 = 2q_3 = 1.349$.

(ii) $4q_3 = 4 \times 0.6745 = 2.698$. Hence, using interpolation in Table 4,

$$\Pr(|Z| > 4q_3) = \Pr(|Z| > 2.698) = 2(1-\Phi(2.698)) = 2(1-0.99651) = 0.00698.$$

```
2 * (1 - pnorm(4 * q3))

## [1] 0.006976603
```

(iii) If the data are from the standard normal distribution, the above probability, 0.007 to three decimal places, is approximately the probability that a randomly chosen data point falls beyond the maximum permitted extent of the whiskers, i.e., one and half times the interquartile range beyond the quartiles. So points beyond the whiskers may be regarded as outliers.

By a simple extension, this argument holds for any normal distribution.

6. The appropriate `R` code is:

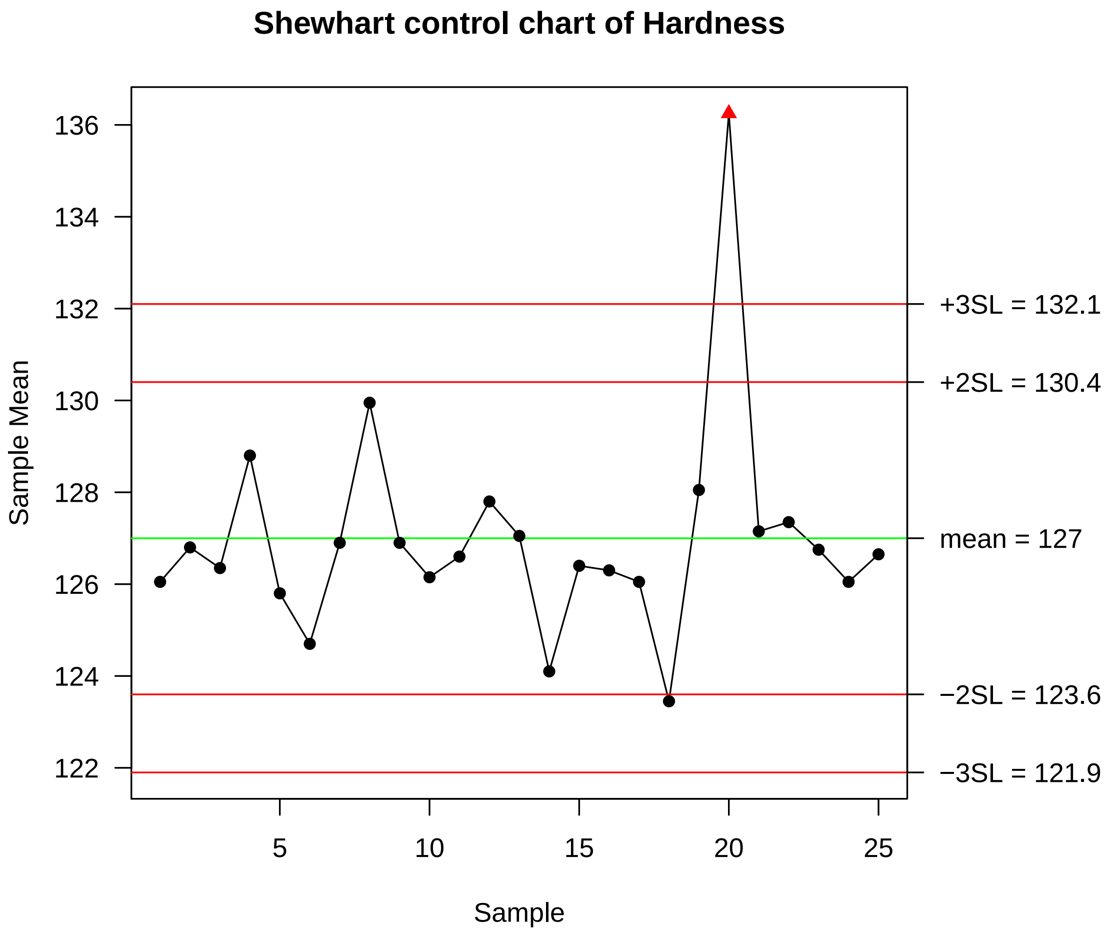
```
hardness <- read.csv("Hardness.csv", header = FALSE)
colnames(hardness) <- c("Sample", "Hardness")
head(hardness)

##   Sample Hardness
## 1      1    125.8
## 2      1    128.4
## 3      1    129.0
## 4      1    121.0
## 5      2    125.2
## 6      2    127.0

tail(hardness)

##     Sample Hardness
## 95      24    130.0
## 96      24    122.8
## 97      25    129.2
## 98      25    126.2
## 99      25    128.0
## 100     25    123.2
```

```
Xbar <- aggregate(hardness[,2], by = list(hardness[,1]), mean)
Xbar

##    Group.1      x
## 1        1 126.05
## 2        2 126.80
## 3        3 126.35
## 4        4 128.80
## 5        5 125.80
## 6        6 124.70
## 7        7 126.90
## 8        8 129.95
## 9        9 126.90
## 10      10 126.15
## 11      11 126.60
## 12      12 127.80
## 13      13 127.05
## 14      14 124.10
## 15      15 126.40
## 16      16 126.30
## 17      17 126.05
## 18      18 123.45
## 19      19 128.05
## 20      20 136.25
## 21      21 127.15
## 22      22 127.35
## 23      23 126.75
## 24      24 126.05
## 25      25 126.65

mean <- 127
sd <- 3.4
n <- 4
plot_title <- "Shewhart control chart of Hardness"
```

This is the same code of the one in the file “ControlCharts.txt” available on Moodle (That is the same code of the last two grey boxes of Chapter 7, page 9.).

```
sample <- Xbar[,1]
xbar <- Xbar[,2]
SL3 <- mean + c(-3, 3) * sd / sqrt(n)
SL2 <- mean + c(-2, 2) * sd / sqrt(n)
outliers <- (xbar > max(SL3)) | (xbar < min(SL3))
sample[outliers]

## [1] 20
```

```
par(mar = c(4, 4, 4, 7))
  plot(sample, xbar, type = "l",
       xlab = "Sample",
       ylab = "Sample Mean",
       main = plot_title,
       ylim = range(xbar, SL3),
       las = 1)
  abline(h = mean, col = "green")
  abline(h = SL2, col = "red")
  abline(h = SL3, col = "red")
  points(sample, xbar, pch = 16 + outliers, col = 1 + outliers)
  axis(4,
   at = c(SL3, mean, SL2),
   label = paste(c("-3SL =", "+3SL =", "mean =", "-2SL =", "+2SL ="),
   round(c(SL3, mean, SL2), 2), sep = " "),
las = 1)
```



The mean for Sample 18 lies just below the lower warning line, but the only really noteworthy point is that the mean for Sample 20 lies way above the upper action line and so would be of serious concern.