Tutorial 6 - R

Adam Shen

February 25, 2019

Question 7.2.18, Page 293

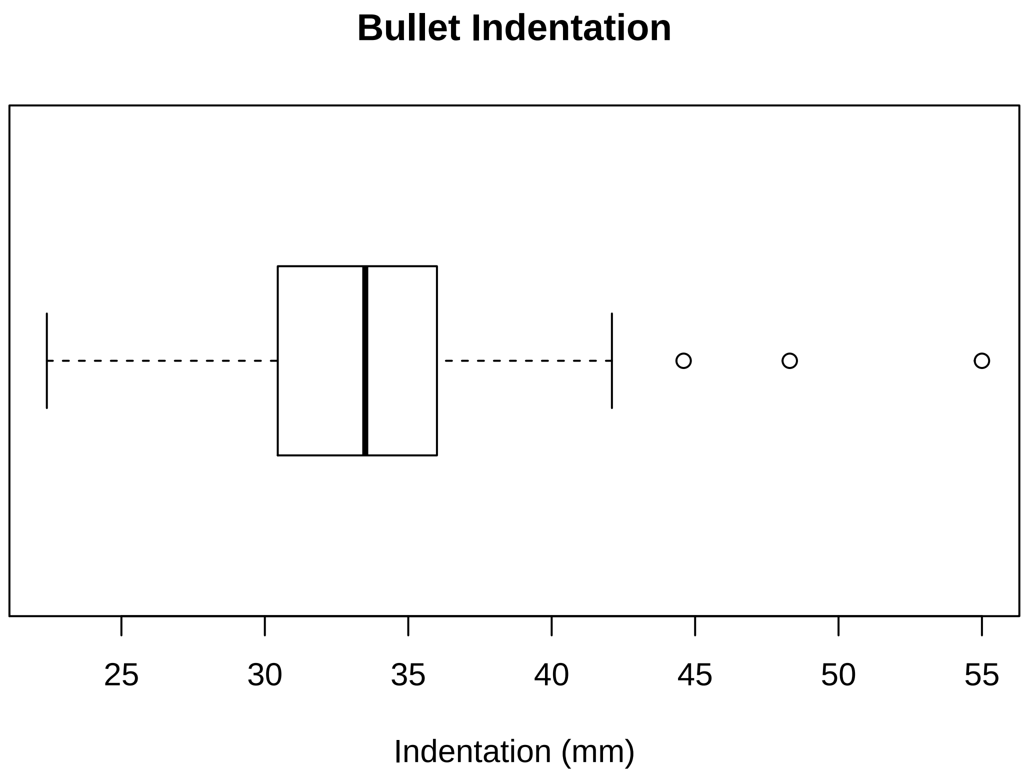
The U.S. Army commissioned a study to assess how deeply a bullet penetrates ceramic body armor. In the standard test, a cylindrical clay model is layered under the armor vest. A projectile is then fired, causing an indentation in the clay. The deepest impression in the clay is measured as an indication of survivability of someone wearing the armor. Measurements were obtained in millimetres.

Load the Data

```
indent <- c(22.4, 23.6, 24.0, 24.9, 25.5, 25.6,
            25.8, 26.1, 26.4, 26.7, 27.4, 27.6,
            28.3, 29.0, 29.1, 29.6, 29.7, 29.8,
            29.9, 30.0, 30.4, 30.5, 30.7, 30.7,
            31.0, 31.0, 31.4, 31.6, 31.7, 31.9,
            31.9, 32.0, 32.1, 32.4, 32.5, 32.5,
            32.6, 32.9, 33.1, 33.3, 33.5, 33.5,
            33.5, 33.5, 33.6, 33.6, 33.8, 33.9,
            34.1, 34.2, 34.6, 34.6, 35.0, 35.2,
            35.2, 35.4, 35.4, 35.4, 35.5, 35.7,
            35.8, 36.0, 36.0, 36.0, 36.1, 36.1,
            36.2, 36.4, 36.6, 37.0, 37.4, 37.5,
            37.5, 38.0, 38.7, 38.8, 39.8, 41.0,
            42.0, 42.1, 44.6, 48.3, 55.0)
```

Create a boxplot

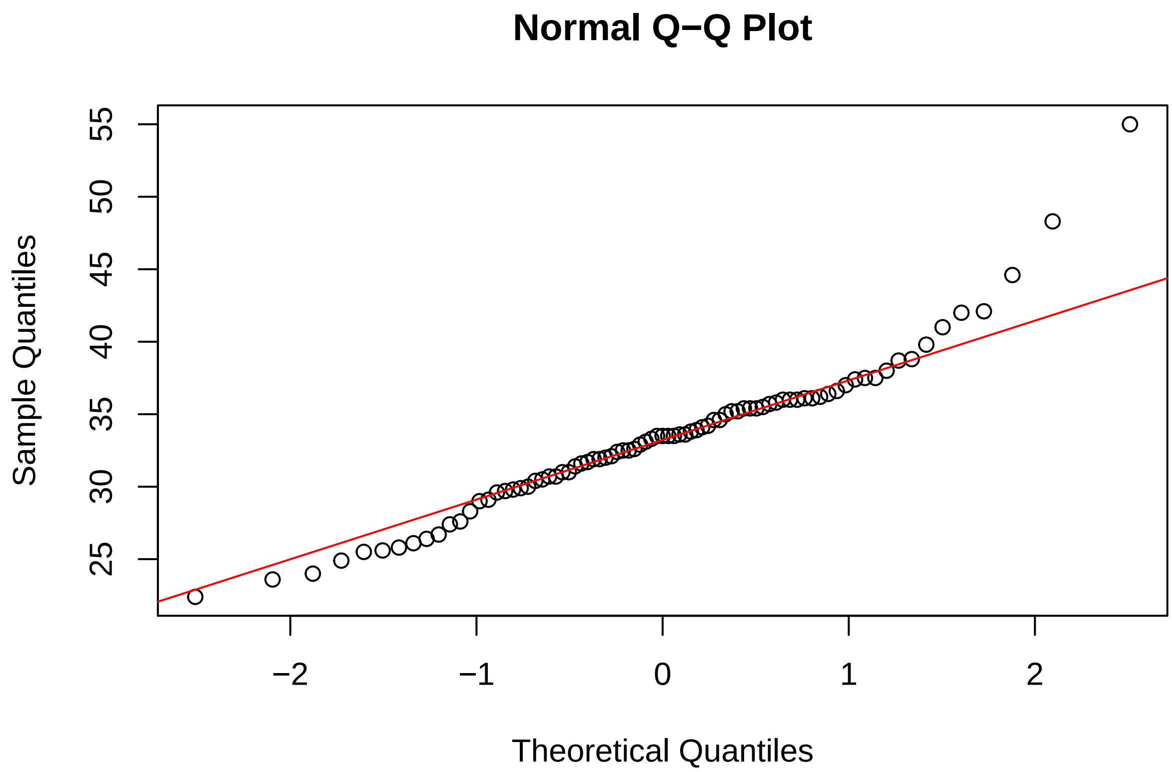
```
boxplot(indent, horizontal=TRUE, main="Bullet Indentation", xlab="Indentation (mm)")
```



Is it plausible that the data is normally distributed?

We can check this by constructing a qqplot (quantile-quantile plot). If the data was truly from a normal distribution, we expect the points to form a straight line.

```
qqnorm(indent)
qqline(indent, col="red")
```



Is a normal distribution assumption needed in order to calculate a confidence interval or bound for the true average depth μ using the foregoing data? Explain.

Values required to calculate bounds/intervals

```
(xbar <- mean(indent))
```

```
## [1] 33.36988
```

```
(s <- sd(indent))
```

```
## [1] 5.268307
```

```
(n <- length(indent))
```

```
## [1] 83
```

Calculate a 95% Confidence Interval for the Population Mean

For a 95% confidence interval, $\alpha = 0.05$. Recall that the formula for the large sample confidence interval is:

$$\left(\overline{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \quad \overline{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right)$$

where $1 - \frac{\alpha}{2}$ is the area to the left of z.

```r
# Lower
(lwr <- xbar - qnorm(0.975)*s/sqrt(n))
```

```
## [1] 32.23649
```

```r
xbar + qnorm(0.025)*s/sqrt(n)
```

```
## [1] 32.23649
```

```r
# Upper
(upr <- xbar + qnorm(0.975)*s/sqrt(n))
```

```
## [1] 34.50327
```

```r
# The CI
cbind(lwr,upr)
```

```
##           lwr      upr
## [1,] 32.23649 34.50327
```

An approximate 95% confidence interval for the population mean is $(32.24, 34.50)$.

Note that `-qnorm(0.975)` = `qnorm(0.025)` due to the symmetry of the normal distribution.

Calculate a 99% Upper Confidence Bound for the Population Mean

This is a one-sided interval so we need $z_{1-\alpha}$, **NOT** $z_{1-\frac{\alpha}{2}}$!! The formula for an upper confidence bound for the population mean is:

$$\mu < \overline{x} + z_{1-\alpha} \cdot \frac{s}{\sqrt{n}}$$

where $1 - \alpha$ is the area to the left of z.

```r
xbar + qnorm(0.99)*s/sqrt(n)
```

```
## [1] 34.71514
```

An approximate 99% upper confidence bound for the population mean is $\mu < 34.72$.

Question 7.3.33, Page 302

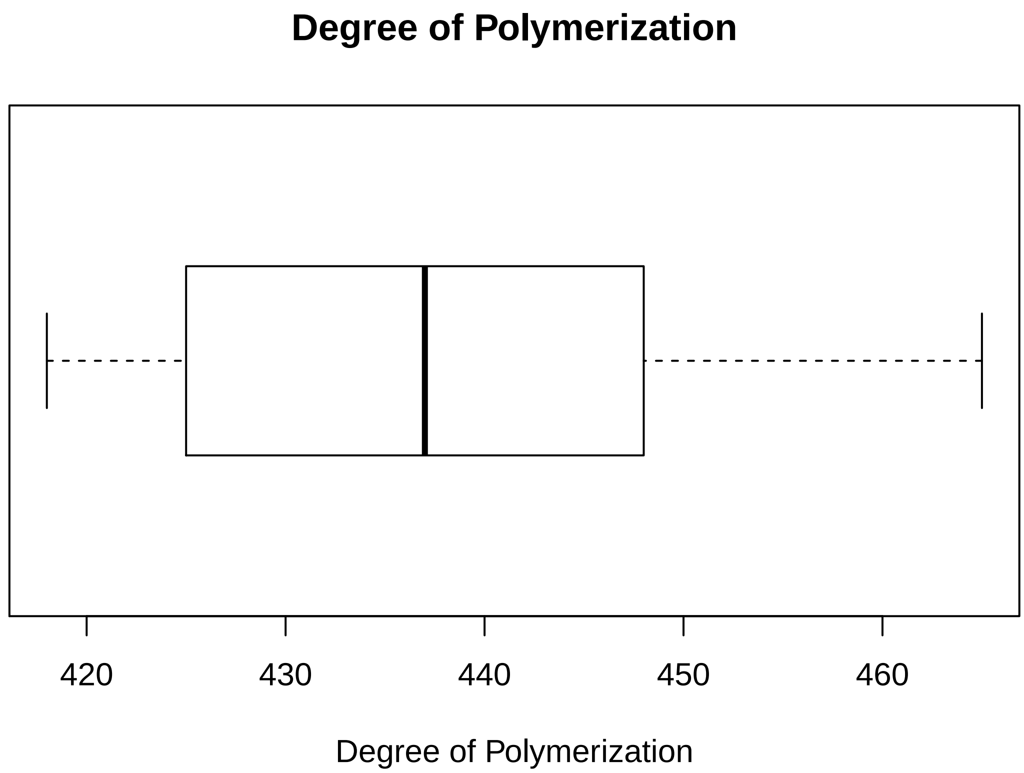
The following data contains observations of degree of polymerization for paper specimens for which viscosity times concentration fell in a certain middle range.

Load the Data

```r
dd <- c(418, 421, 421, 422, 425, 427, 431,
        434, 437, 439, 446, 447, 448, 453,
        454, 463, 465)
```

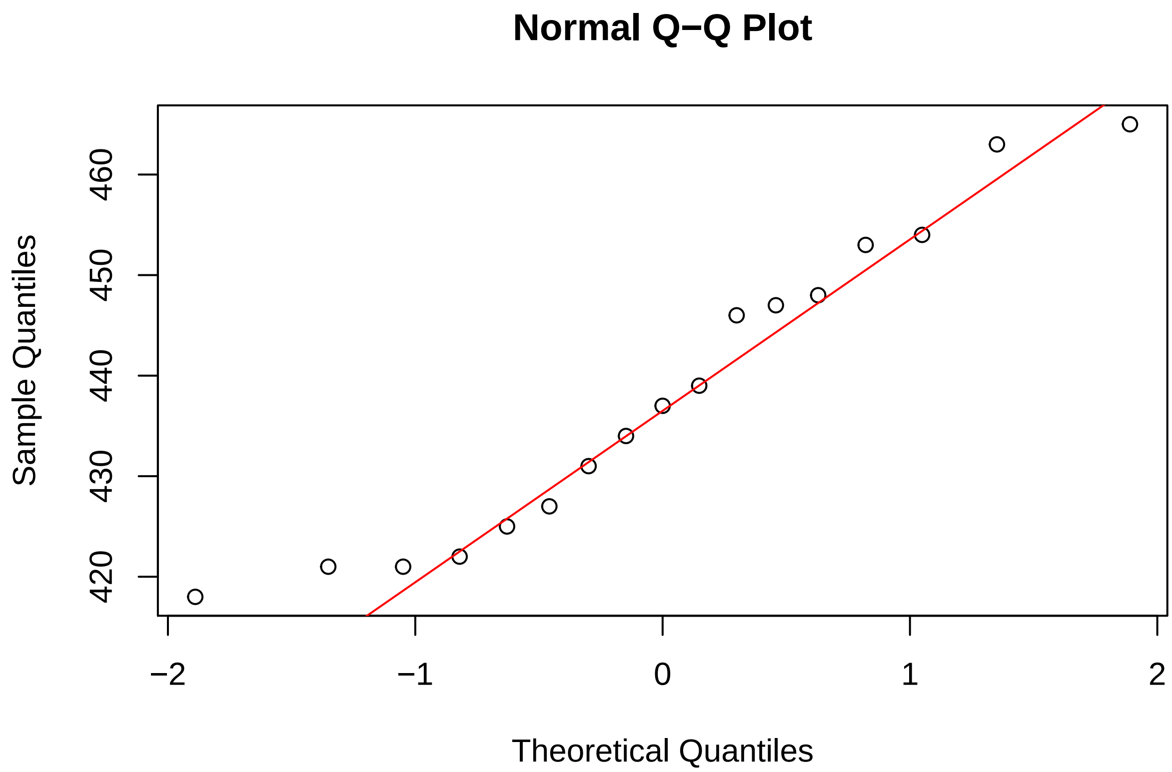
Create a boxplot

```r
boxplot(dd, horizontal=TRUE, main="Degree of Polymerization", xlab="Degree of Polymerization")
```



Is it plausible that the sample is from a normal distribution?

```
qqnorm(dd)
qqline(dd, col="red")
```



Calculate a 95% Confidence Interval

```
(xbar <- mean(dd))
```

```
## [1] 438.2941
```

```
(s <- sd(dd))
```

```
## [1] 15.14416
```

```
(n <- length(dd))
```

```
## [1] 17
```

The formula for a 95% CI is:

$$\overline{x} - t_{1-\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}}, \quad \overline{x} + t_{1-\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}}$$

where $1 - \frac{\alpha}{2}$ is still the area to the left (this is the default in R; your textbook uses area to the right - be careful of this)

```
xbar + c(-1,1)*qt(0.975, df=n-1)*s/sqrt(n)
```

```
## [1] 430.5077 446.0805
```

An exact 95% CI for the true average degree of polymerimerization is (430.51, 446.08).

Are 440 and 450 plausible values?

Since 440 lies within the above interval, it is a plausible value. Since 450 does not lie within the above interval, it is not a plausible value.