Week 06 R Workshop

Elmer V Villanueva

October 14, 2019

New functions and packages

- `dnorm()` draw
- `pnorm()` know cut-pts, calculate prob pnorm(q=4.6, mean=3.4, sd=0.6, lower.tail=FALSE)
- `qnorm()` know prob, calculate cut-pts qnorm(p=0.1, mean=3.4, sd=0.6, lower.tail=FALSE)
- `curve()` draw
- `polygon()` draw

Set your working directory

```
setwd("D:/git/DPH101-xjtlu/Y3/week06_lec_10.14")
```

Case 1: Birth weights in the USA

It is not surprising that babies with low birth weights are at risk for developmental difficulties, but extraordinarily large babies also face a higher than normal risk of medical problems. In the United States the average full term single birth baby has a weight of 3.4 kg with a standard deviation of 0.6 kg. We will assume that the distribution of birth weights is Gaussian.

Question 1

Babies below 2.5 kg in weight are considered to be high risk/low birth weight deliveries. What percentage of births would be in this category?

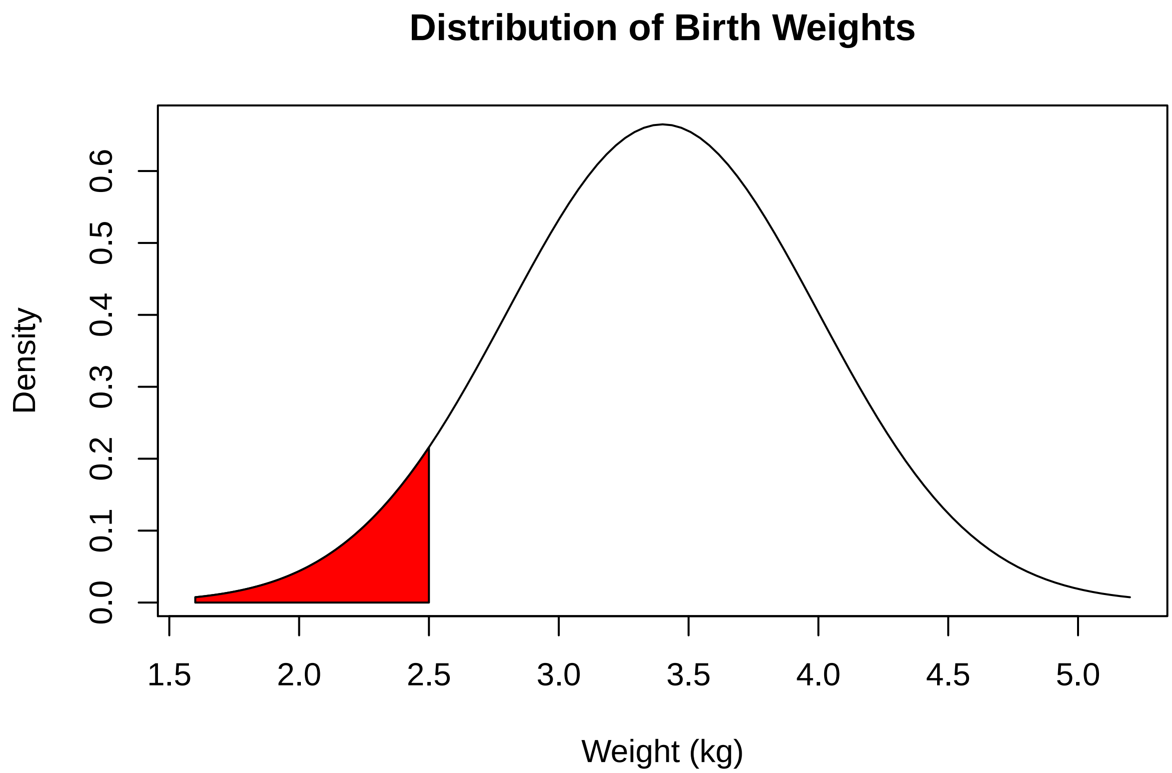
Solution

Given $\mu = 3.4$, $\sigma = 0.6$, find $P(x < 2.5)$.

First, visualise the problem. This takes a two steps. We draw the normal curve. Then, we identify the area of interest.

```
# Plot the normal curve
curve(dnorm(x, mean=3.4, sd=0.6),
      xlim=c(1.6,5.2),
      main="Distribution of Birth Weights",
      xlab="Weight (kg)",
      ylab="Density")

# Share the area of interest
coord.x <- c(1.6,seq(1.6,2.5,0.01),2.5)
coord.y <- c(0,dnorm(seq(1.6,2.5,0.01), mean=3.4, sd=0.6),0)
polygon(coord.x,coord.y,col='red')
```



We can calculate the probability of the shaded area as follows:

```
pnorm(q=2.5, mean=3.4, sd=0.6, lower.tail=TRUE)
```

```
## [1] 0.0668072
```

Question 2

Babies above 4.6 kg in weight are considered to be high risk/high birth weight deliveries. What percentage of births would fall in this category?

Solution

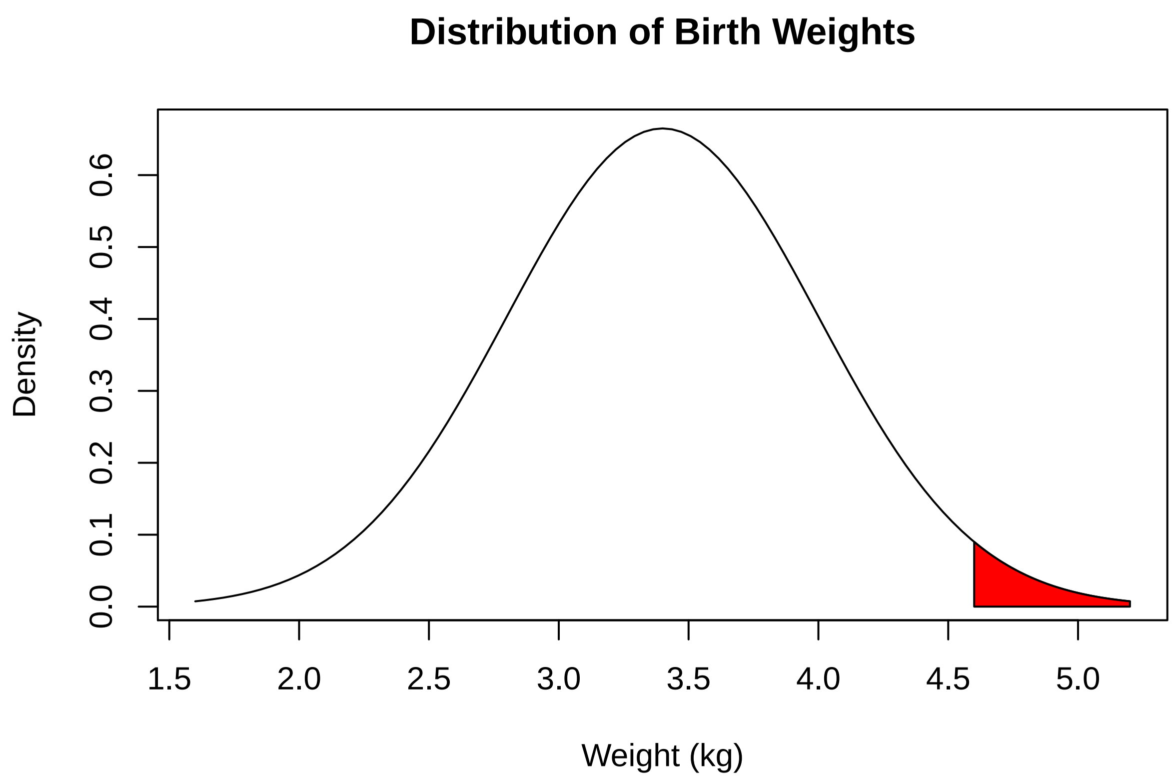
Given $\mu = 3.4$, $\sigma = 0.6$, find $P(x > 4.6)$.

Visualise the problem.

```
curve(dnorm(x, mean=3.4, sd=0.6),
      xlim=c(1.6,5.2),
      main="Distribution of Birth Weights",
      xlab="Weight (kg)",
      ylab="Density")

coord.x <- c(4.6,seq(4.6,5.2,0.01),5.2)
```

```r
coord.y <- c(0,dnorm(seq(4.6,5.2,0.01), mean=3.4, sd=0.6),0)
polygon(coord.x,coord.y,col='red')
```



Calculate the probability

```r
pnorm(q=4.6, mean=3.4, sd=0.6, lower.tail=FALSE)
```

```
## [1] 0.02275013
```

Question 3

Suppose a new study claims that only the middle 80% of the birth weights should be considered normal. What would be the new cut-off points for low and high weight risk deliveries?

Solution

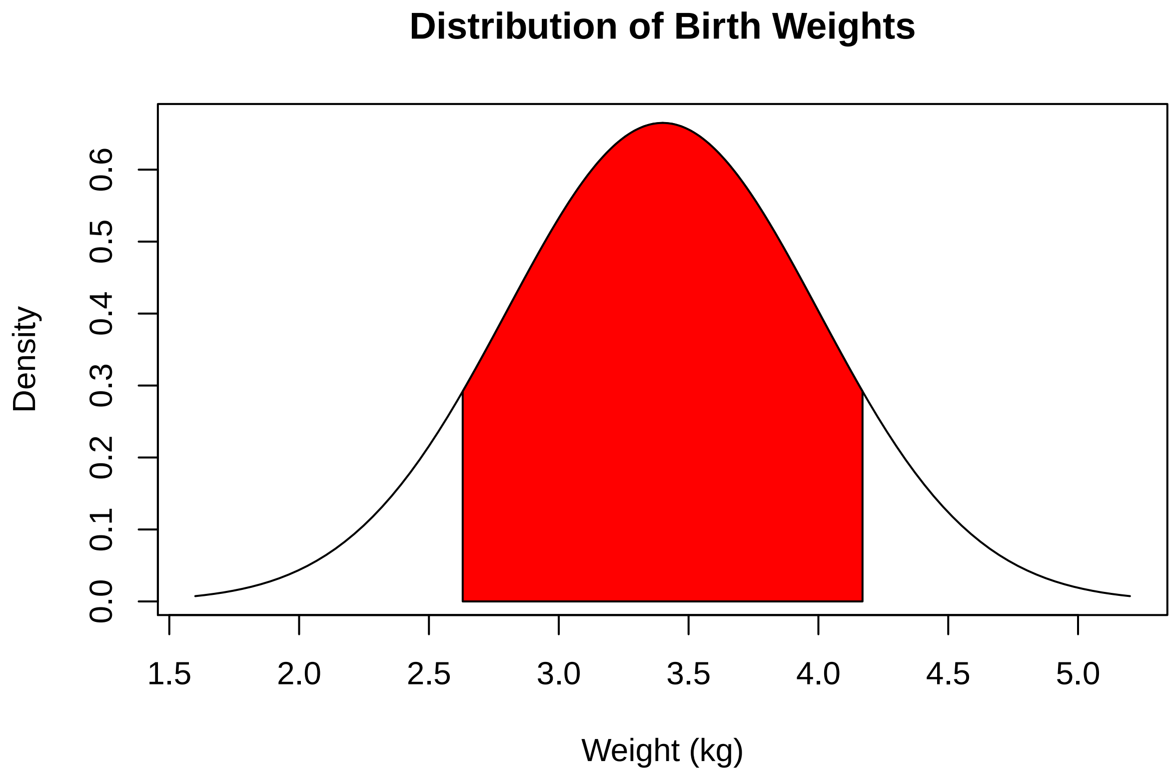
Find the lower and upper bounds of the middle 80%.

Visualise the problem.

```r
curve(dnorm(x, mean=3.4, sd=0.6),
      xlim=c(1.6,5.2),
      main="Distribution of Birth Weights",
      xlab="Weight (kg)",
```

```r
      ylab="Density")

coord.x <- c(2.63,seq(2.63,4.17,0.01),4.17)
coord.y <- c(0,dnorm(seq(2.63,4.17,0.01), mean=3.4, sd=0.6),0)
polygon(coord.x,coord.y,col='red')
```



Calculate the cut-points.

```r
qnorm(p=0.1, mean=3.4, sd=0.6, lower.tail=TRUE)
```

```
## [1] 2.631069
```

```r
qnorm(p=0.1, mean=3.4, sd=0.6, lower.tail=FALSE)
```

```
## [1] 4.168931
```

Case 2: Serum cholesterol

The National Health and Nutrition Examination Survey of 1988-1994 (NHANES III) estimated the mean serum cholesterol level of 183 mg/dl for women aged 20-29 years. The estimated standard deviation was approximately 37 mg/dl. Use these estimates as the mean and standard deviation for the U.S. population.

Question 1

If a simple random sample of size 60 is drawn from this population, what is the mean of the sampling distribution? The standard error?

Solution

The mean of the sampling distribution will be the mean of the population. That is, $\mu_{\bar{x}} = 183\ mg/dl$. The standard error will be the standard deviation of the population divided by the square root of the sample size. That is, $\sigma_{\bar{x}} = 37/\sqrt{60} = 4.8\ mg/dl$

```r
37/sqrt(60)
```

```
## [1] 4.776679
```

Question 2

If a simple random sample of size 60 is drawn from this population, find the probability that the sample mean serum cholesterol level will be between 170 and 195 mg/dl.

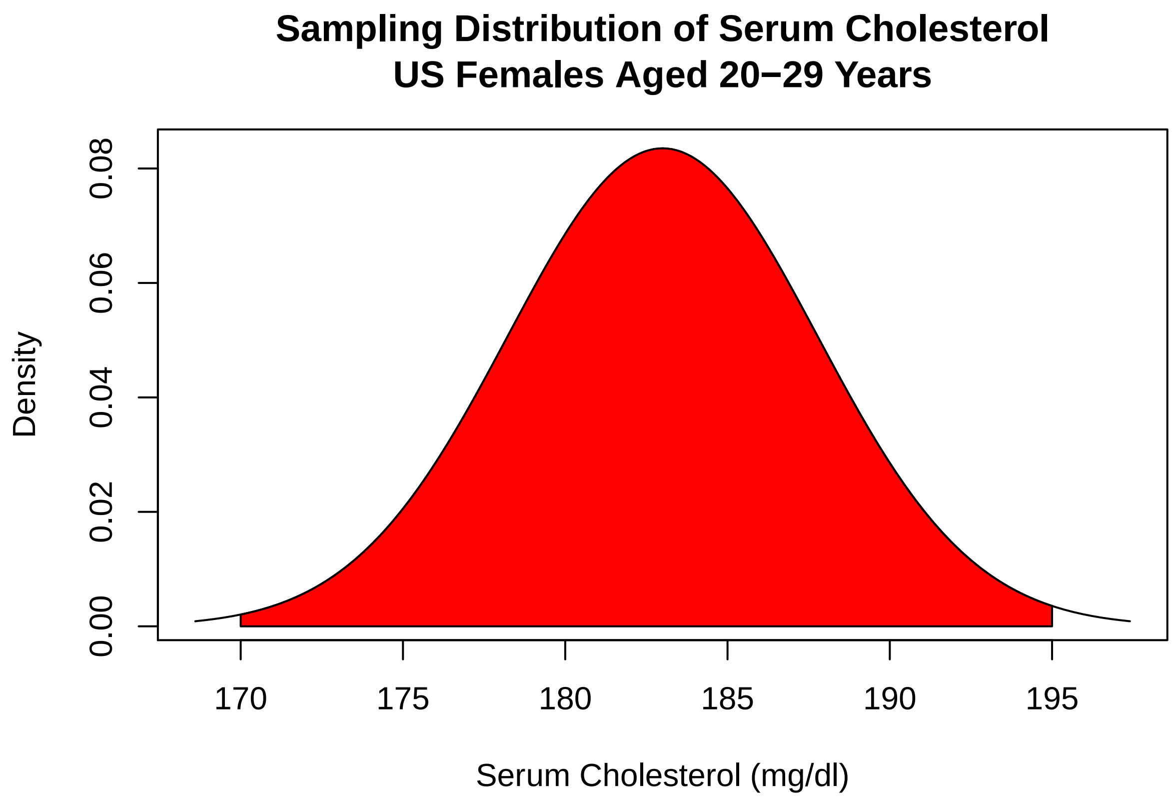
Solution

Given $\mu = 183$, $\sigma = 37$ and $n = 60$, find the probability that the sample mean serum cholesterol is between 170 and 195.

Visualise the problem.

```r
curve(dnorm(x, mean=183, sd=37/sqrt(60)),
      xlim=c(168.6,197.4),
      main="Sampling Distribution of Serum Cholesterol\nUS Females Aged 20-29 Years",
      xlab="Serum Cholesterol (mg/dl)",
      ylab="Density")

coord.x <- c(170,seq(170,195,0.01),195)
coord.y <- c(0,dnorm(seq(170,195,0.01), mean=183, sd=37/sqrt(60)),0)
polygon(coord.x,coord.y,col='red')
```



Calculate the probability.

```r
pnorm(q=195, mean=183, sd=37/sqrt(60), lower.tail=TRUE) -
    pnorm(q=170, mean=183, sd=37/sqrt(60), lower.tail=TRUE)
```

```
## [1] 0.9907523
```

THE END