# Quantiles

*S520*

*September 20, 2016*

## Quantiles

A CDF answers the question: Given a value of the random variable, what is the "less than or equal to" probability? But we could also ask the opposite question: Given a "less than or equal to" probability, what is the corresponding value of the random variable.

Formally, let $X$ be a random variable and choose $\alpha$ such that $0 < \alpha < 1$. If $q$ is such that

$$P(X < q) \leq \alpha$$

and

$$P(X > q) \geq 1 - \alpha$$

then $q$ is an $\alpha$-quantile of $X$.

For continuous random variables, the $\alpha$-quantile exists for all $\alpha$ (though it may not be unique; see below.) We can find a general expression for quantiles by inverting the CDF: set $F(q) = \alpha$ and solve for $q$.

**Example.** A Uniform$(a, b)$ random variable has CDF

$$F(y) = \begin{cases} 0 & y < a \\ \frac{y-a}{b-a} & a \leq y < b \\ 1 & y \geq b \end{cases}$$

We're only worried about the middle part.

$$\begin{aligned} F(q) &= \alpha \\ \frac{q-a}{b-a} &= \alpha \\ q &= a + \alpha(b-a) \end{aligned}$$

## Population median

The **population median**, denoted $q_2$, is the 0.5-quantile of a random variable:

$$\begin{aligned} P(X < q) = 0.5 \\ P(X > q) = 1 - \alpha \end{aligned}$$

However, a distribution may not have just one 0.5-quantile: it may have an interval of values that all satisfy the definition of 0.5-quantile. (This can happen when there is a "gap" in the PDF.) If this is the case, take the midpoint of the interval of all values that satisfy the definition of 0.5-quantile.

For a symmetric distribution, the median is the same as the expected value.

## Quartiles and IQR

- The 0.25-quantile is called the **first quartile** ($q_1$).
- The 0.5-quantile is called the **second quartile** ($q_2$).
- The 0.75-quantile is called the **third quartile** ($q_3$).

The **interquartile range**, or **IQR**, is $q_3 - q_1$. Like the standard deviation, it's a measure of spread on the same scale as the original random variable.

## qnorm()

The R function `qnorm()` is used to find quantiles of the Normal. By default, the function uses a standard normal; otherwise you need to specify mean and standard deviation.

```
# Median of the standard normal
qnorm(0.5)
```

```
## [1] 0
```

```
# Find the IQR
qnorm(0.75) - qnorm(0.25)
```

```
## [1] 1.34898
```

```
# Let X be Normal(0, 10^2)
# Find the IQR
qnorm(0.75, mean=0, sd=10) - qnorm(0.25, mean=0, sd=10)
```

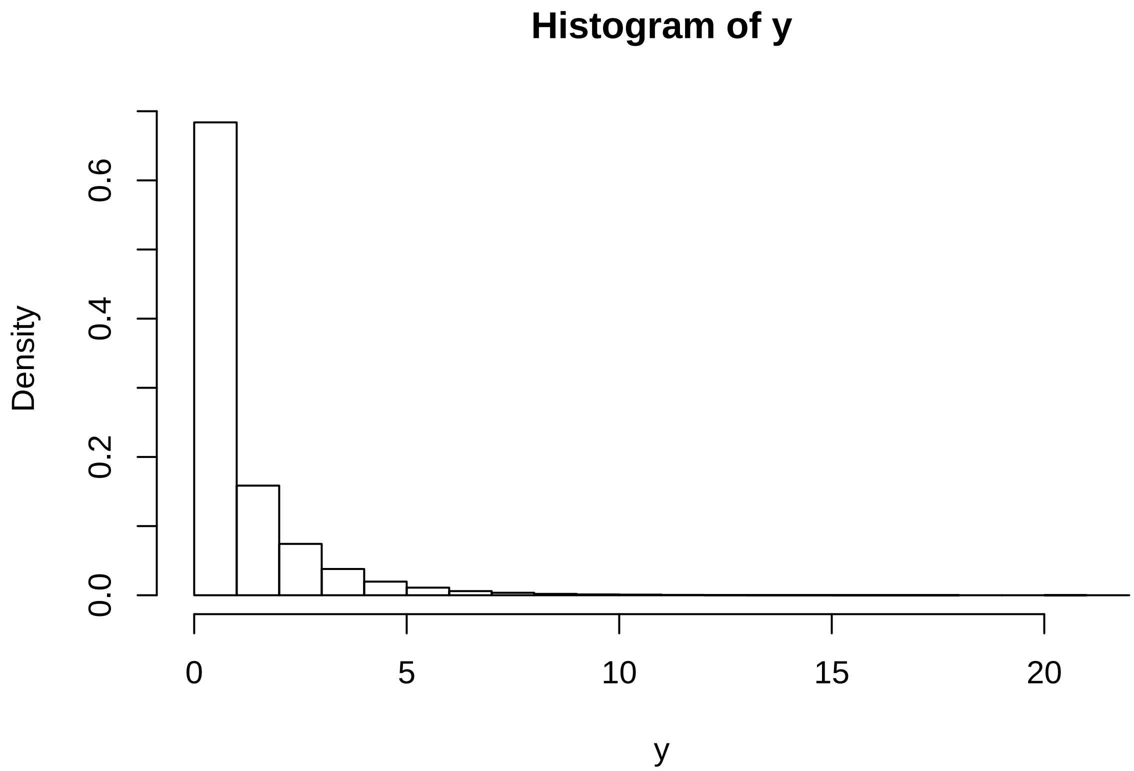
```
## [1] 13.4898
```

### rnorm() and simulation

The function `rnorm()` generates random independent realizations of a Normal random variable. We can use these in **simulations** to study problems that are hard to solve analytically.

**Example.** Let $Z$ be std normal. Let $Y = Z^2$. What does the PDF of $Y$ look like? What's the expected value of $Z$?

```
z = rnorm(100000)
y = z^2
# Draw a histogram
hist(y, prob=TRUE)
```




```r
mean(y)
```

```
## [1] 1.001134
```

### Least squares

I don't have anything interesting to say about least squares, so just read Trosset p. 148.