Recall : (X,Y) have a joint density

- $f: \mathbb{R}^2 \to \mathbb{R}_+$ if

$$\mathbb{P}(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du$$

- X, Y are independent then

$$f(x,y) = f_X(x)\, f_Y(y)$$

- Transformation of random variables

$$Y = g(X)$$

- $F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y)$ ... to known X
- differentiate $F_Y(\,)$ to get p.d.f of Y

- X, Y are two independent random variables with pdf $f_X(\cdot)$ and $f_Y(\cdot)$

$Z = X + Y$ – Q: Find distribution of Z.

Step1 :- Find $F(z) = \mathbb{P}(Z \le z)$

$$F(z) = \mathbb{P}(Z \le z)$$

$$= \mathbb{P}(X + Y \le z)$$

$$= \iint_{\{(x,y):\, x+y \le z\}} f(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x)\, f_Y(y)\, dy\, dx$$

$$F(z) = \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_X(x) \, f_Y(u-x) \, dx \right) du$$

Step 2: Differentiate $F(\cdot)$

$$f_Z(z) = F'(z)$$

(fundamental Theorem)

$$= \int_{-\infty}^{\infty} f_X(x) \, f_Y(z-x) \, dx$$

∴ X, Y are independent with p.d.f $f_X(\cdot)$ and $f_Y(\cdot)$ $Z = X + Y$

then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \, f_Y(z-x) \, dx$$

Statistics

Data – work with real data

Probability – study of models for (random) experiments when the model is fully known.

Statistics – Model is not fully known. One tries to infer about the unknown aspects of the model based on observed outcomes of experiment

Given: $X_1, X_2, \ldots X_n, \ldots$ i.i.d observations

Kurtosis and Skewness

Let $\{x_i : 1 \leq i \leq n\}$ be given data set. Then the mean and Standard Deviation, sd, are given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Kurtosis and *Skewness* coefficients.

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3, \qquad \text{Kurtosis} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4.$$

Skewness ↓ Symmetry of distribution

Kurtosis ↓ Peak of distribution

$\bar{x}$ - Sample mean is an estimate for true mean of the underlying distribution

σ_x^2 - Sample variance is an estimate for true variance of the underlying distribution

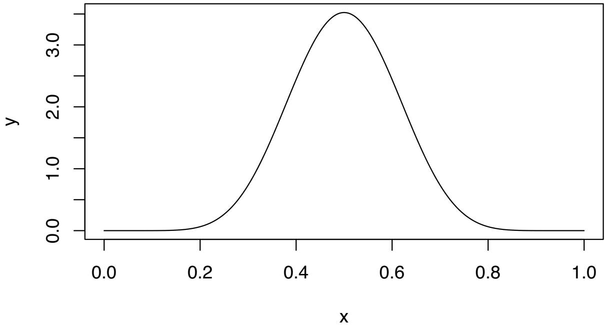
Symmetric, Skewed- Left and Right

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3.$$

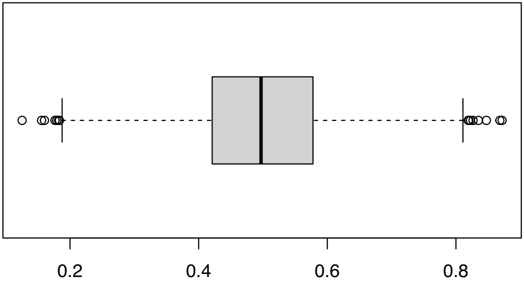
- Skewness is a measure of symmetry.
- Negative skewness will imply that the mean $<$ median $<$ mode, and the data distribution is left-skewed.

  Positive skewness will imply that the mean $>$ median $>$ mode, and the data distribution is right-skewed.

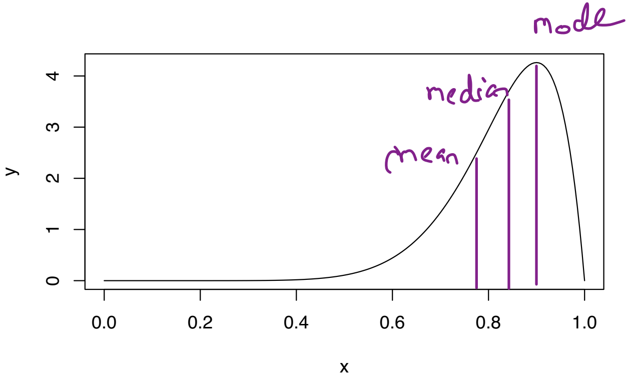
Symmetric



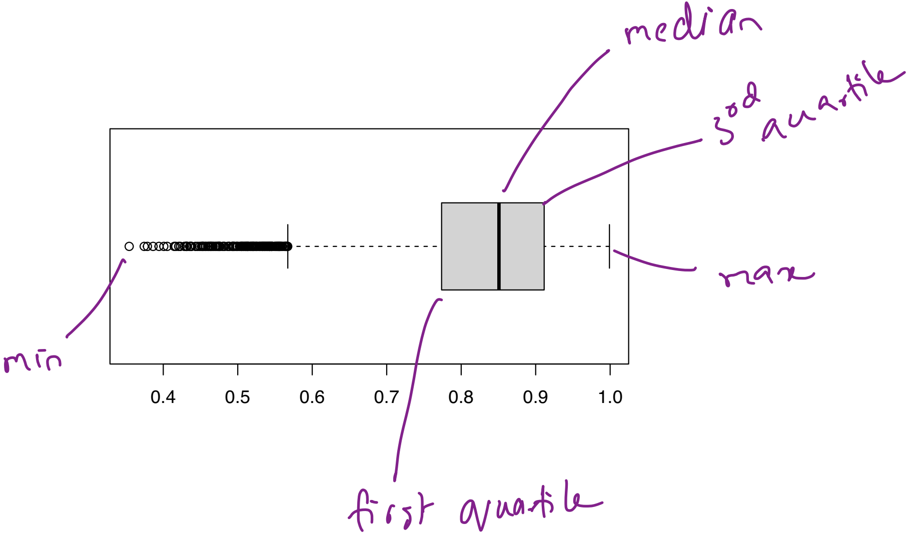
Symmetric



Left Skewed

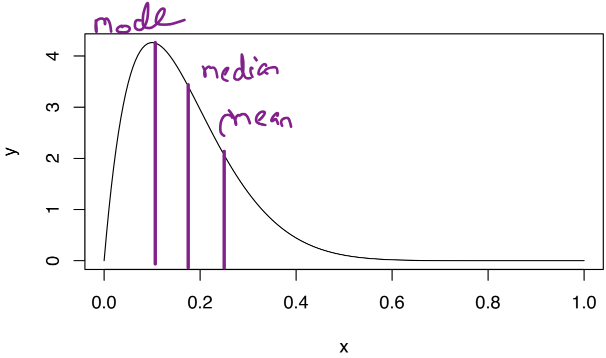


Left Skewed

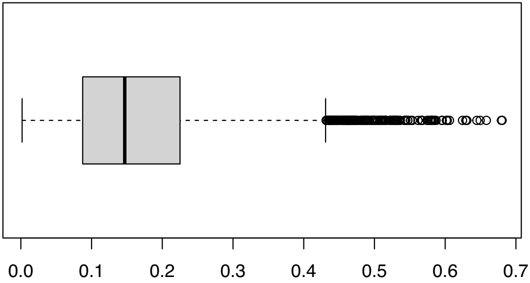


median
3rd quartile
max
min
first quartile
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Right-Skewed



Right-Skewed



Normal Distribution: Kurtosis and Skewness

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3, \qquad \text{Kurtosis} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4.$$

Skewness 0 indicates that the distribution is symmetric

Kurtosis is a measure of the peak of the distribution.

For standard normal Kurtosis is 3.

If Kurtosis and Skewness coefficients are *far* from 3 and 0 we can conclude that the data is not normal.

Normal Distribution: Quantiles

If Z is standard Normal random variable then the α-th quantile of Z is denoted by z_α where

$$P(Z \leq z_\alpha) = \alpha, \qquad 0 < \alpha < 1.$$

Let $\{x_i : 1 \leq i \leq n\}$ be given data set then we can order them to get

$$(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$$

we view $x_(k)$ as the $\frac{k}{n+1}$ sample quantile.

Normal Distribution: Quantiles

Q - Q plots

Another check whether data is normal is to plot

$$\left(z_{\frac{k}{n+1}}, x_{(k)}\right)$$

if the plot is a straight line then it indicates that data is normal

Check: if data is normal or not

(1) – Do 68-95-99.7 first check to see if data is like normal or not.

Compute

(2) – $$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3, \qquad \text{Kurtosis} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4.$$

Skewness 0 indicates that the distribution is symmetric

Kurtosis is a measure of the peak of the distribution.

For standard normal Kurtosis is 3.

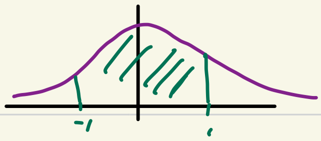
If Kurtosis and Skewness coefficients are *far* from 3 and 0 we can conclude that the data is not normal.

(3) – Final check whether data is normal is to plot

$$(z_{\frac{k}{n+1}}, x_{(k)})$$

if the plot is a straight line then it indicates that data is normal

$Z \sim N(0,1)$



$\mathbb{P}(-1 \le Z \le 1) \cong 0.68$

$\mathbb{P}(-2 \le Z \le 2) \cong 0.95$

$\mathbb{P}(-3 \le Z \le 3) \cong 0.99$