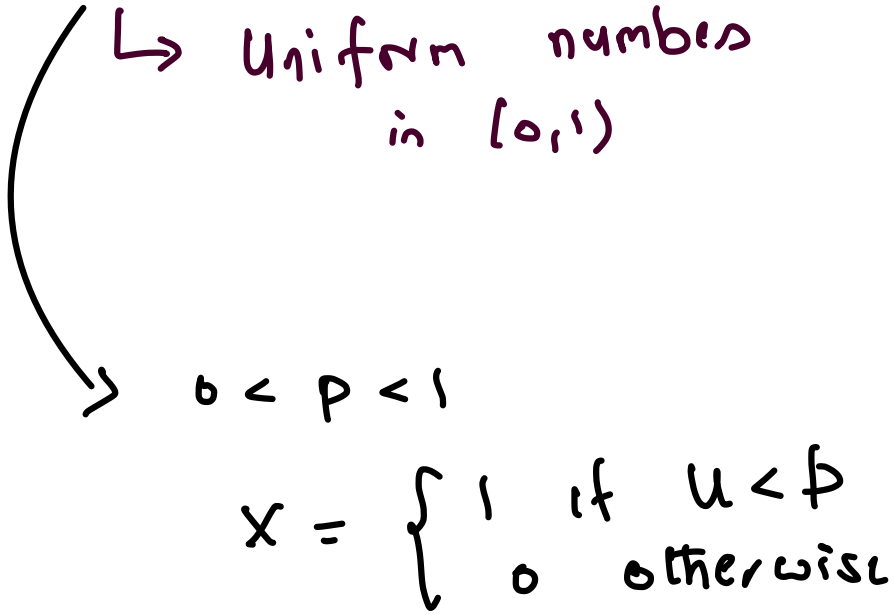
Recall

- runif(1) := u
  - ↳ Uniform numbers in [0,1)



Simulating Random Variables

in R
- (inbuilt) &
- from U(0,1)

Goal: To do this for any F

(any distribution)

Simulating Uniform Random Variables - U(0,1)

Assume - Coored

```
> set.seed(2021)
> runif(3)
[1] 0.4512674 0.7837798 0.7096822
> set.seed(2021)
> runif(3)
[1] 0.4512674 0.7837798 0.7096822
> runif(3)
[1] 0.3817443 0.6363238 0.7013460
```

Seed a random number algorithm uses

Same #'s given same seed

Different

Simulating Bernoulli (p) Distribution

```
> p = 0.5;
> U = runif(1, min = 0, max = 1);
> X = as.numeric(U < p);
> X
[1] 0
```

{ Uniform (0,1) 1- sample

{ TRUE FALSE

{ 1 0

← output of the program

$X \overset{d}{=}$ Bernoulli(p)

Ex:- Generate $x_1, x_2, \ldots, x_{10}$; $x_i \overset{d}{=}$ Bernoulli(p)

Simulating Binomial(n, p) Distribution

```
> p = 0.2;
> n = 50;
> U = runif(n, min = 0, max = 1);
> X = sum(U < p);
> X
[1] 9
```

Generates 50 U[0,1), #s

$$X = \sum_{i=1}^{50} 1(U_i < p)$$

Fact: $1 \le i \le 50$

$X_i \overset{d}{=}$ Bernoulli(p) & independent

$$X = \sum_{i=1}^{50} X_i$$

$$\Downarrow$$

$X \overset{d}{=}$ Binomial$(50, p)$

$X \overset{d}{=}$ Binomial$(50, p)$

Simulating $\{p_i\}_{i=1}^{n}$ Distribution

Ex: $n = \infty$, $0 < p_i < 1$

```
> p1=0.35
> p2=0.15
> p3=0.4
> p4=0.1
> P = c(p1,p1+p2, p1+p2+p3, p1+p2+p3+p4)
> X = c(1, 2, 3, 4)
> counter = 1
> set.seed(2021)
> r = runif(1, min = 0, max = 1);
> while(r > P[counter])
+ counter = counter + 1
> end
> X[counter]
```

u

} - find the interval j:

Sample from X

Fact: $\text{Range}(X) = \{1, \dots, n\}$

$\mathbb{P}(X = i) = p_i$

$$X = \begin{cases} 1 & u < p_1 \\ 2 & p_1 \le u < p_1 + p_2 \\ \vdots & \\ i & \sum_{k=1}^{i-1} p_k \le u < \sum_{k=1}^{i} p_k \\ \vdots & \\ n & \sum_{k=1}^{n-1} p_k \le u \le 1 \end{cases}$$

$u \overset{d}{=} U(0,1)$

$$\sum_{k=1}^{j-1} p_k \le u < \sum_{k=1}^{j} p_k$$

Simulating $\{p_i\}_{i=1}^{n}$ Distribution

```
> p1=0.35
> p2=0.15
> p3=0.4
> p4=0.1
> P = c(p1,p1+p2, p1+p2+p3, p1+p2+p3+p4)
> X = c(1, 2, 3, 4)
> counter = 1;
> set.seed(2021)
> r = runif(1, min = 0, max = 1);
> for (value in P[counter]) {
+ if(r > P[counter]){
+   counter = counter +1
+   }
+ }
> X[counter]
[1] 2
```

Using for loop

instead of

while

Ex: • $n = \infty$, $0 < p_i < 1$

• Range $(X) = S$

S - Countable.

Note: $X = i \Leftrightarrow F(i-1) \leq U < F(i)$

where $F(x) = \mathbb{P}(X \leq x)$

Formal Idea to generate sample from continuous r.v.

Suppose

- $U \overset{d}{=}$ Uniform $(0,1)$ r.v.
- X is a random variable with a strictly increasing

$$F_X : \mathbb{R} \to [0,1]$$

distribution function

$\Rightarrow$ • $W = F_X^{-1}(U)$ has the same distribution as X

$$\mathbb{P}(X \le x) = F_X(x)$$

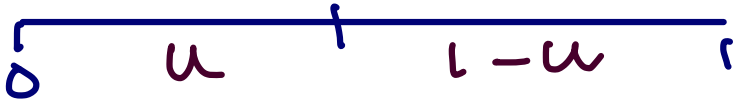
$$= \mathbb{P}(U \le F_X(x))$$

F ↑ strictly

$$= \mathbb{P}(X \ge F_X^{-1}(U))$$

Simulating Exponential (λ)Distribution

$$F_X^{-1}(u) = -\ln(1-u)$$




```
> U = runif(1, min = 0, max = 1);
> lambda = 1
> X = -(1/lambda)*log(U)
```

$X \overset{d}{=} \text{Exp}(\lambda)$

Symmetry of Uniform(0,1)

$X \overset{d}{=}$ Exponential (λ)

$$F(x) = \int_0^x \lambda e^{-\lambda y}\, dy$$

$$= 1 - e^{-\lambda x} \quad \text{for } x > 0$$

$F(x) = 0$ for $x \le 0$

$X_i \overset{d}{=} Exp(\lambda) \qquad 1 \le i \le n$

and X_i's are independent

$$X = \sum_{i=1}^{n} X_i$$

$$X \overset{d}{=} Gamma(n, \lambda)$$

$$f_X(x) = \begin{cases} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} & x > 0 \\ 0 & \text{o.w.} \end{cases}$$

$$F_X(x) = \int_0^x f_X(y)\, dy$$

(maybe difficult to invert!)

Simulating Gamma(n, λ) Distribution

```
> n = 20;
> lambda = 10;
> X = (-1/lambda)*sum(log(runif(n,min=0,max=1)))
```

$[\log(U_1), \log(U_2) \dots \log(U_n)]$

$$\sum_{i=1}^{n} -\frac{1}{\lambda} \log(U_i)$$

$$X = \sum_{i=1}^{n} x_i \; ; \; x_i \overset{d}{=} \text{Exp}(\lambda) \text{ \& independent}$$

Expectation

Let $X : S \to T$ be a discrete random variable (so T is countable). Then the expected value (or average) of X is written as $E[X]$ and is given by

$$E[X] = \sum_{t \in T} t \cdot P(X = t)$$

provided that the sum converges absolutely. In this case we say that X has "finite expectation".

- If the sum diverges to $\pm\infty$ we say the random variable has infinite expectation.
- If the sum diverges, but not to infinity, we say the expected value is undefined.

X - continuous random variable with probability density function

$$f: \mathbb{R} \to [0, \infty)$$

& piecewise continuous.

$$\mathbb{P}(X \in A) = \int_A f(x)\, dx$$

A - nice set

e.g. (a, b)

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx \quad < \infty$$

Riemann integral

$$\int_{-\infty}^{\infty} |x|\, f(x)\, dx \quad < \infty$$

Properties of Expected Value

$E[X]$ – measure of "centre" of r.v. X

Suppose that X and Y are discrete random variables, both with finite expected value and both defined on the same sample space S. If a and b are real numbers then

$$E[aX] = aE[X];$$

$$E[X + Y] = E[X] + E[Y]; \quad \text{and}$$

$$E[aX + bY] = aE[X] + bE[Y].$$

If $X \geq 0$ then $E[X] \geq 0$.

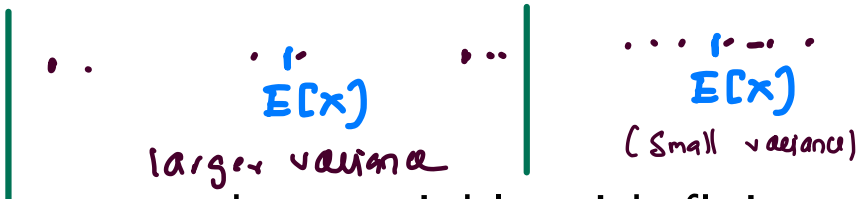
$E[X-Y] = E[X] - E[Y]$

$X \leq Y$

$\Rightarrow E[X] \leq E[Y]$

Variance

$$E[h(X)] = \begin{cases} \sum_{t \in T} h(t)\, P(X=t) \\ \int_{\mathbb{R}} h(t) f(t)\, dt \end{cases}$$



Let $X : S \xrightarrow{\mathbb{R}} T$ be a discrete random variable with finite expected value. Then the variance of the random variable is written as $Var[X]$ and is defined as

$$Var[X] = E[(X - E[X])^2] = \sum_{t \in T} (t - E[X])^2 \; P(X = t)$$

- The standard deviation of X is written as $SD[X]$ and is defined as

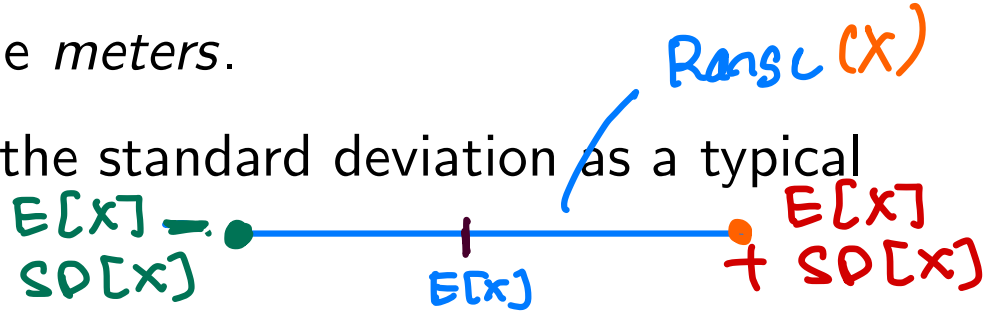
$$SD[X] = \sqrt{Var[X]}$$

– Spread in same units as X

- If $(t - E[X])^2$ is large & $P(X=t)$ is high $\Rightarrow$ higher variance

Interpretation of Variance

- If X has a high probability of being far away from $E[X]$ the variance will tend to be large, while if X is very near $E[X]$ with high probability the variance will tend to be small.
- If we were to associate units with the random variable X (say *meters*), then the units of $Var[X]$ would be *meters*^{2} and the units of $SD[X]$ would be *meters*.
- Informally we will view the standard deviation as a typical distance from average.



- It is possible that $Var[X]$ and $SD[X]$ could be infinite even if $E[X]$ is finite – meaning that the random variable has a clear average, but is so spread out that any finite number underestimates the typical distance of the random variable from its average.

Standardising Random Variables

- A standardized random variable X is one for which

$$E[X] = 0 \quad \text{and} \quad Var[X] = 1.$$

- Let X be a discrete random variable with finite expected value and finite, non-zero variance. Then $Z = \frac{X - E[X]}{SD[X]}$ is a standardized random variable.

$$E[Z] = E\left[\frac{X - E[X]}{SD[X]}\right] = \frac{1}{SD[X]}\left[E[X] - E[X]\right] = 0$$

$$Var[Z] = E\left[Z - E[Z]\right]^2 = \frac{E\left[X - E[X]\right]^2}{SD[X]^2} = 1$$

Normal Distribution: PDF

$Z \sim N(0,1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

You can calculate the values of the normal density function using the the `dnorm` command.

```
> dnorm(0)
[1] 0.3989423
> dnorm(1)
[1] 0.2419707
> dnorm(0, mean=4, sd=3)
[1] 0.05467002
```

$f(0)$

$f(1)$

$g(0)$

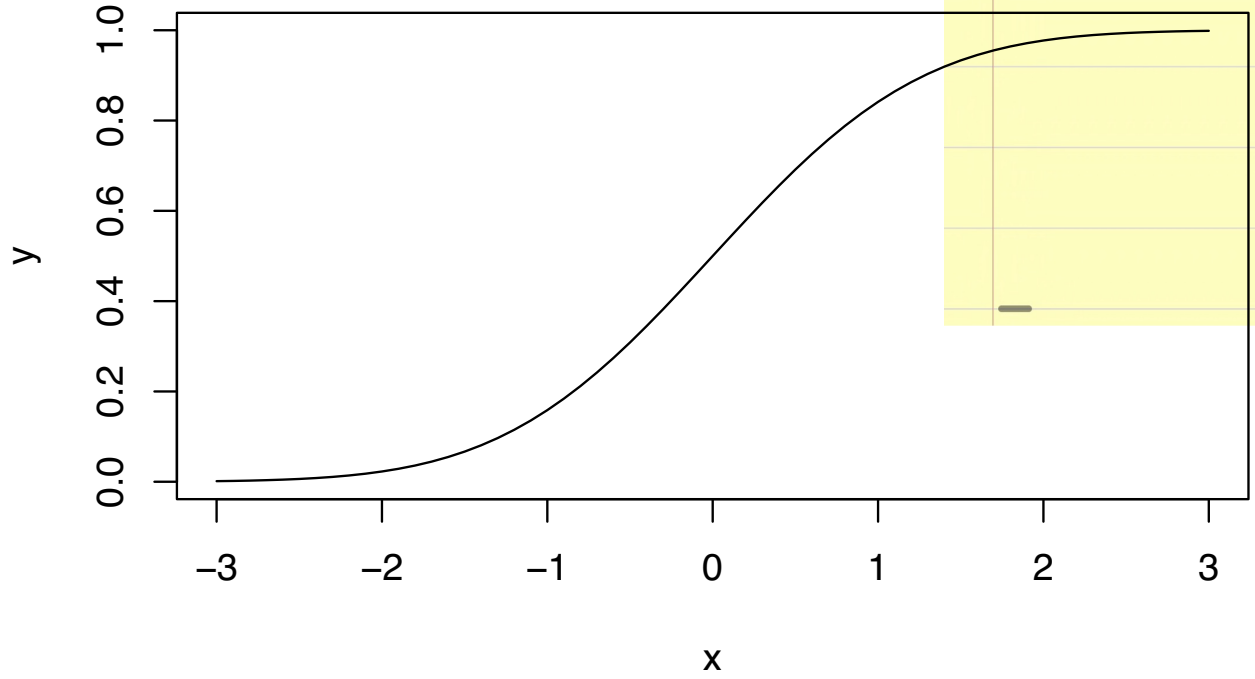
$$g(x) = \frac{1}{\sqrt{2\pi}\,3} e^{-\frac{(x-4)^2}{2\,3^2}}$$

Normal Distribution: CDF

You can calculate the values of the cummulative distribution function of the normal using the the `pnorm` command.

```
> pnorm(0)
> pnorm(1)
> x = seq(-3,3, by=0.1); y = pnorm(x) ;plot(x,y, type="l")
```

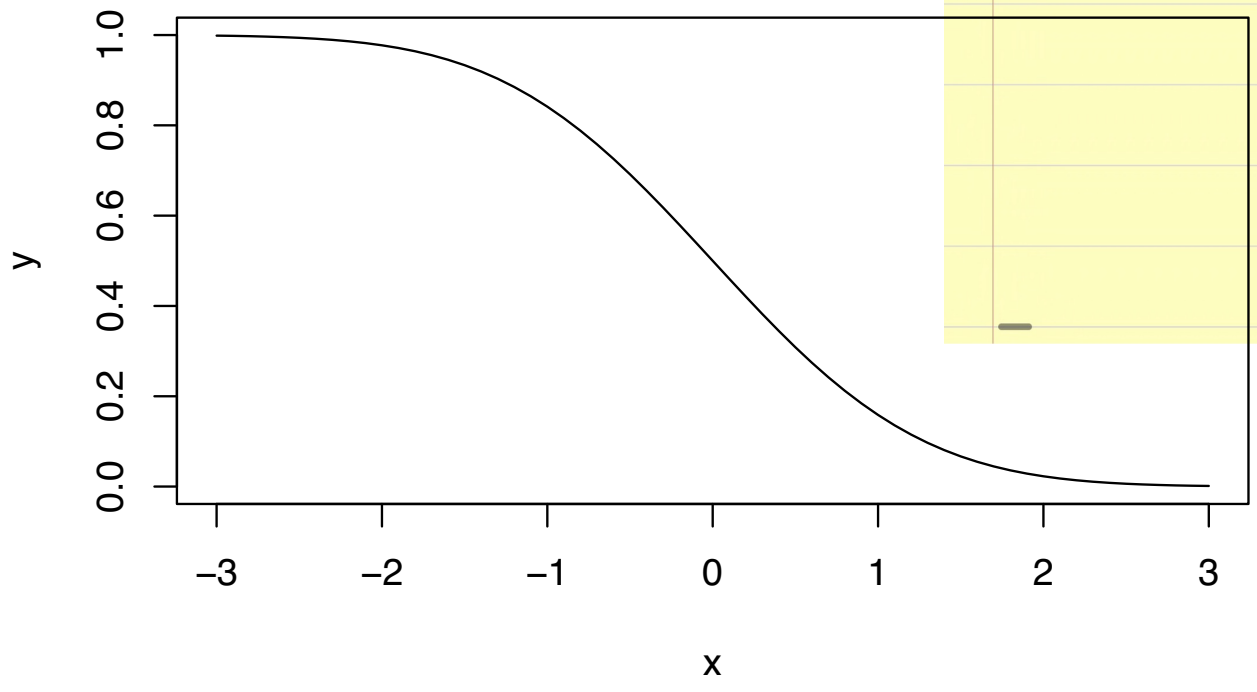
$F(x) = \mathbb{P}(Z \leq x)$



Normal Distribution: Tail Probabilities

```
> pnorm(0, lower.tail=FALSE)
> pnorm(1, lower.tail=FALSE)
> x = seq(-3,3, by=0.1); y = pnorm(x, lower.tail=FALSE)
> plot(x,y, type="l")
```

$\mathbb{P}(Z \geq x)$



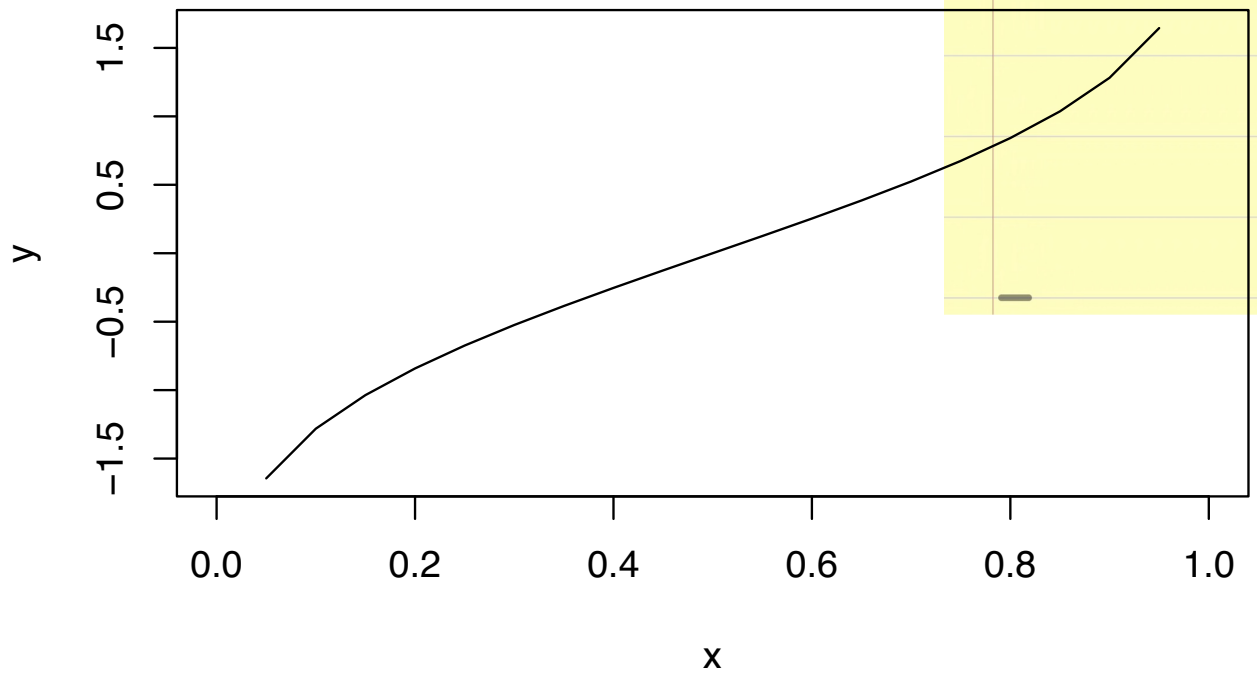
Normal Distribution: quantiles

```
> qnorm(0.68); qnorm(0.95);qnorm(0.997)
[1] 0.4676988
[1] 1.644854
[1] 2.747781
> x = seq(0,1, by=0.05); y = qnorm(x);plot(x,y, type="l")
```

$P(Z \leq 0.4676988) = 0.68$

$P(Z \leq 1.64) = 0.95$

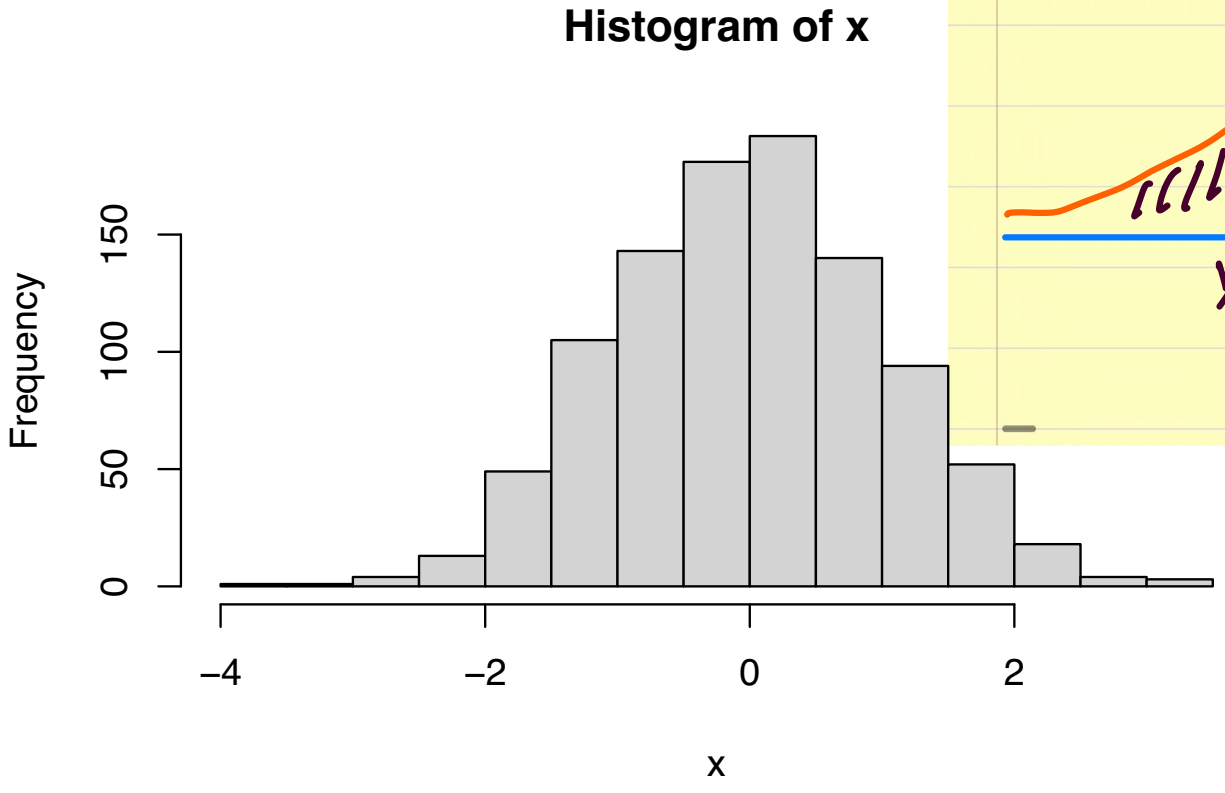
$P(Z \leq qnorm(\alpha)) = \alpha$

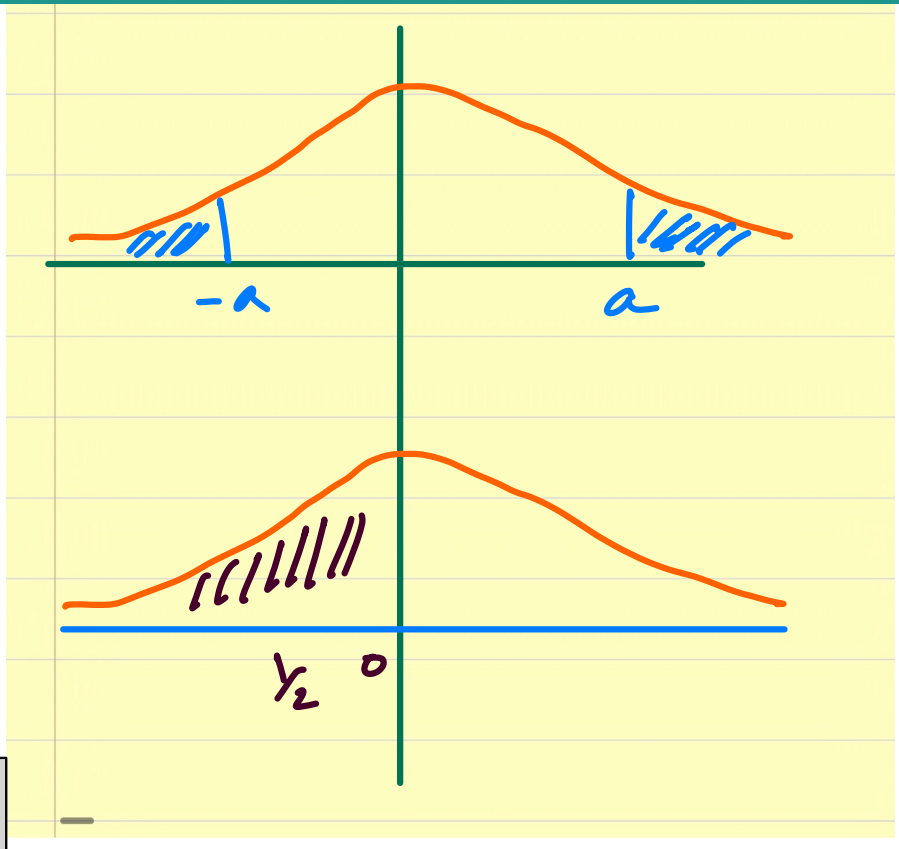


Normal Distribution: samples

```
> x=rnorm(1000)

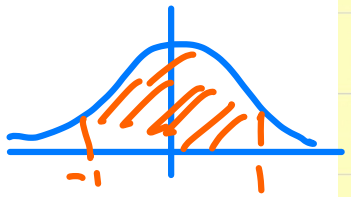
> hist(x)
```





Normal Distribution: Key Probabilities

$\mathbb{P}(|Z| < 1)$



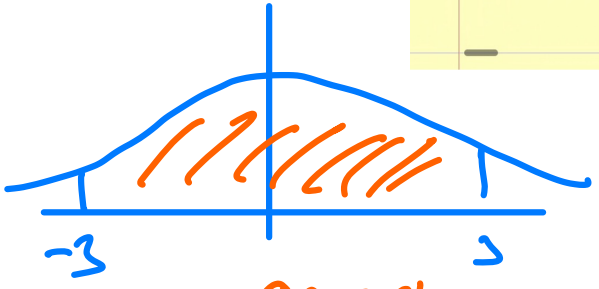
68%

```
> pnorm(1) - pnorm(-1) # within one standard deviation
[1] 0.6826895
> pnorm(2) - pnorm(-2) # within two standard deviation
[1] 0.9544997
> pnorm(3) - pnorm(-3) # within three standard deviation
[1] 0.9973002
```

"$(\mu - 3\sigma, \mu + 3\sigma)$"

Effective range of $N(\mu, \sigma^2)$

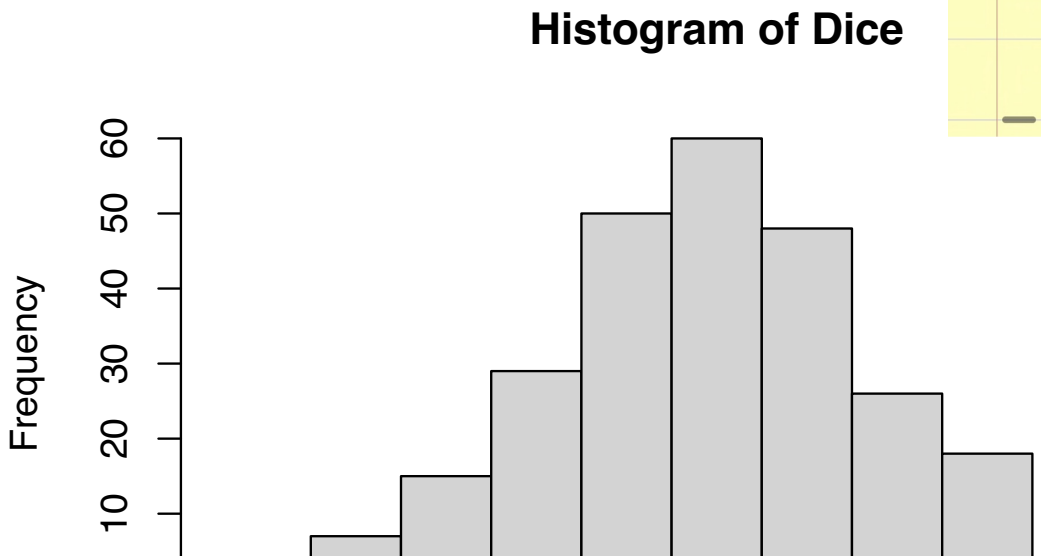
$\mathbb{P}(|Z| < 3)$



99.7%

Are the Sum of Rolls normal ?

```
> DiceR = read.csv("Dice.csv", header=T)
> names(DiceR)= "Sum"
> DiceR$Sum = as.numeric(DiceR$Sum)
> Dice = na.omit(DiceR$Sum)
> hist(Dice)
```



Are the Sum of Rolls normal ?

```
> summary(Dice)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   7.00   15.00   17.50   17.23   20.00   27.00
> sd(Dice)
[1] 3.670899
> cs = (Dice-mean(Dice))/(sd(Dice))
```

– standardised

Are the Dice normal ?

```
> mean(cs)
[1] 2.261974e-16
> sd(cs)
[1] 1
> onesdcs = cs[cs >-1& cs <1] # within one sd
> twosdcs = cs[cs >-2& cs <2] # within two sd
> threesdcs = cs[cs >-3& cs <3]# within three sd
```

Are the Dice normal ?

```
> length(onesdcs)/length(cs)
[1] 0.6653846
> length(twosdcs)/length(cs)
[1] 0.9576923
> length(threesdcs)/length(cs)
[1] 1
```

Check 1:

"$(\mu - k\sigma, \mu + k\sigma)$"

$k = 1, 2, 3$

This is called the 68-95-99.7 first check to see if data is normal or not.

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

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

Another check for normal are *Kurtosis* and *Skewness* coefficients.

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

≈ 0 for normal ≈ 3

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.

Are the Dice normal ?

```
> library(moments)
> Kurtosis = kurtosis(Dice)
> Kurtosis
[1] 3.002322
> Skewness = skewness(Dice)
> Skewness
[1] -0.2548971
```

Skewness < 0

Are the Dice normal ?

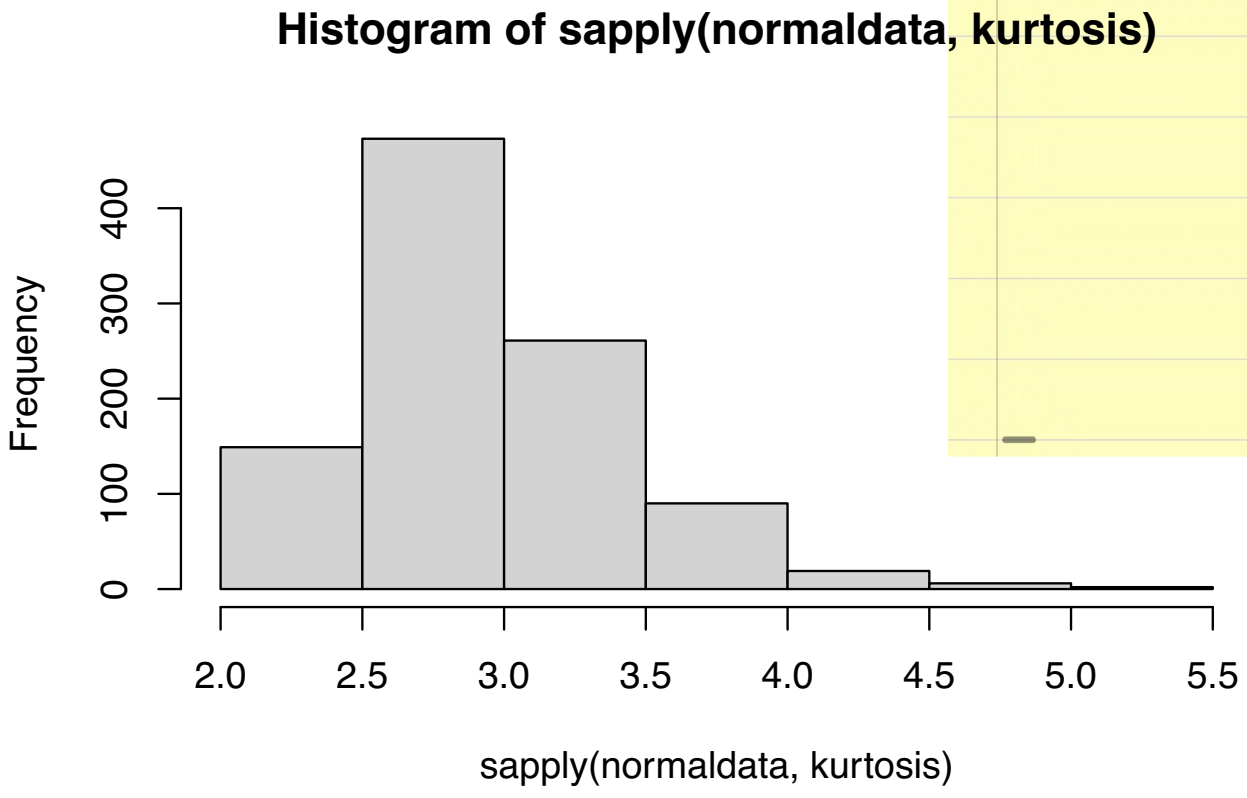
does 1000 Trials → Generates 97 samples of N(0,1)

```
> normaldata = replicate(1000, rnorm(97,0,1),
+ simplify=FALSE)
> library(moments)
> mean(sapply(normaldata, kurtosis))
[1] 2.95045
> mean(sapply(normaldata, skewness))
[1] 0.00127284
```

1000... kurtosis

Are the Dice normal ?

```
> normaldata = replicate(1000, rnorm(97,0,1),
+ simplify=FALSE)
> library(moments)
> hist(sapply(normaldata, kurtosis))
```



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.

$\frac{k}{n+1} = P(Z \leq z_{\frac{k}{n+1}})$

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

Normal Distribution: Quantiles

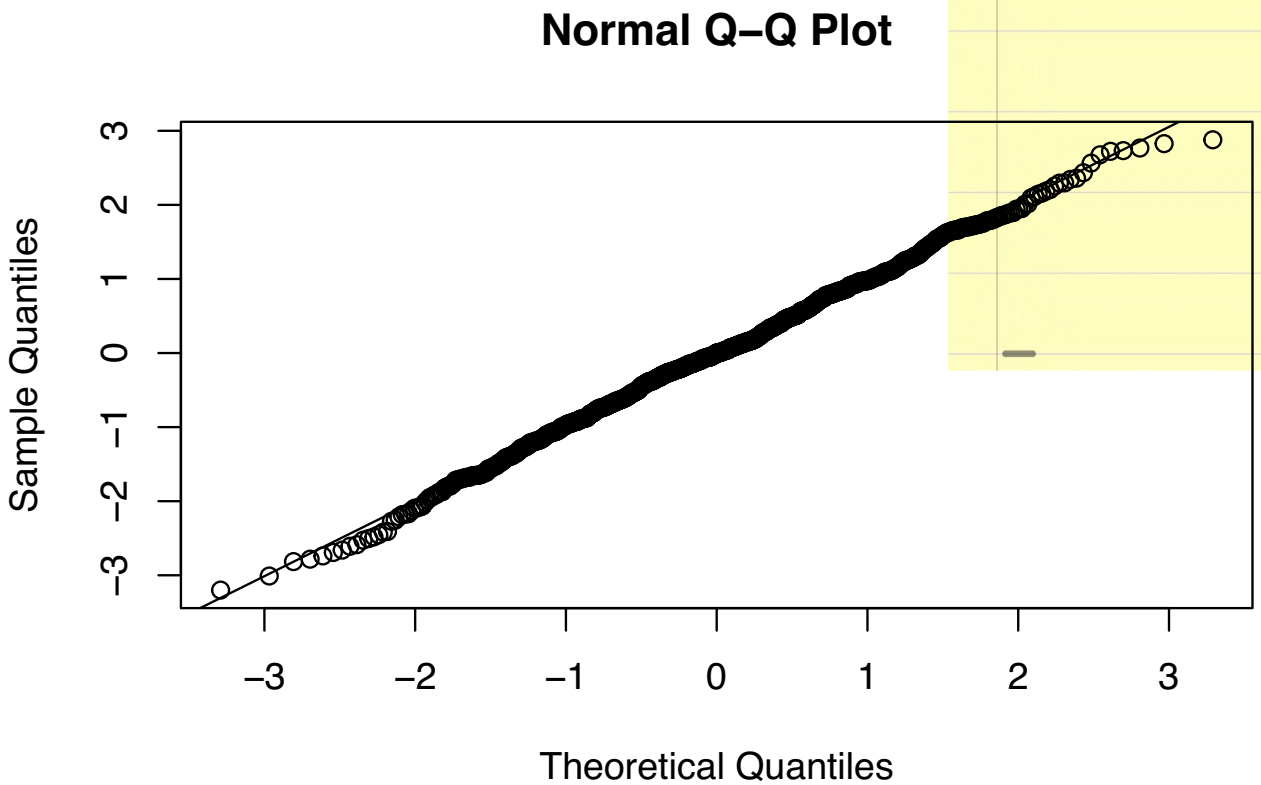
Another check whether data is normal is to plot

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

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

Normal Distribution: Sample Normal-Quantile plot

```
> x=rnorm(1000)
> qqnorm(x)
> qqline(x)
```



Dice Normal-Quantile-plot

```
> qqnorm(Dice)
> qqline(Dice)
```

