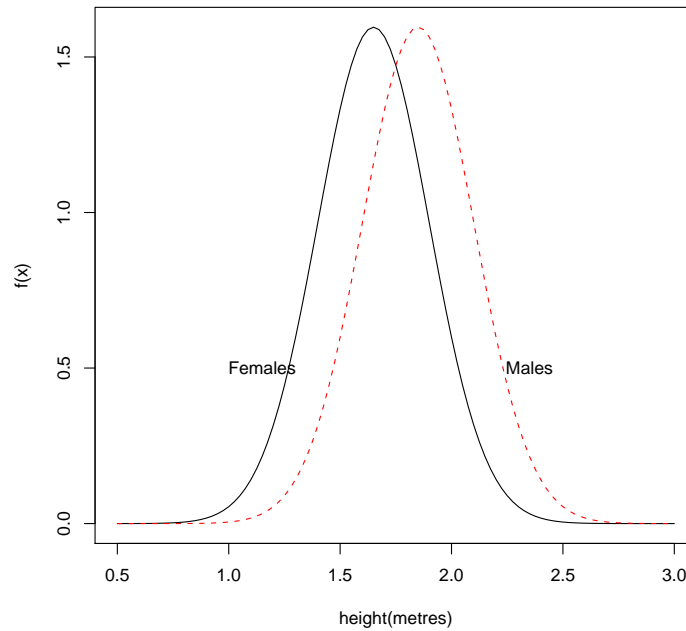


18 THE NORMAL DISTRIBUTION



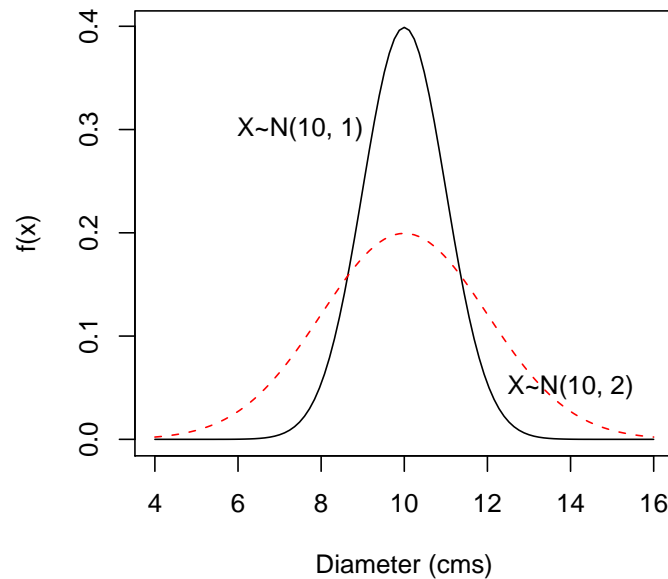
Bell-shaped curves; centred at different points: variation at either side of central points is the same

Height of females and males

$X \sim N(1.65, .25)$ for the females

$X \sim N(1.85, .25)$ for the males

THE NORMAL DISTRIBUTION



Bell-shaped curves, both centered at 10, one high and narrow, the other low and wide.

Components manufactured by two machines which are set to a diameter 10 cms.

One machine produces components with error 1cm:

$$X \sim N(10, 1)$$

The second machine has an error of 2cms:

$$X \sim N(10, 2)$$

The Probability Density Function

The *pdf* of the normal distribution is

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X - \mu)^2}{2\sigma^2}\right) \quad -\infty < X < \infty$$

It can be shown that

$$E(X) = \mu$$

$$V(X) = \sigma^2$$

X is said to be normally distributed with a mean of μ and a variance of σ^2 .

i.e.

$$X \sim N(\mu, \sigma)$$

To plot a normal curve use the function `curve` with `dnorm` as the argument:

```
curve (dnorm(x, 2, .25), from = 1, to = 3)
```

plots a normal curve with a mean of 2 and a standard deviation of .25 in the range 1 to 3.

The Cumulative Distribution Function

$$P(X \leq x) = pnorm(x, \mu, \sigma)$$

Examples:

From previous records it is known that examination results are normally distributed with mean $\mu = 45$ and the standard deviation $\sigma = 4$.

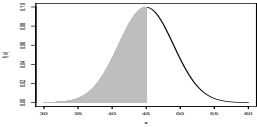
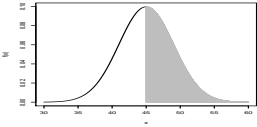
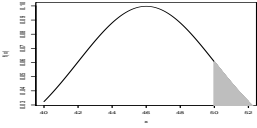
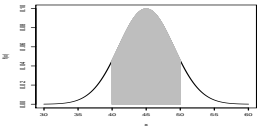
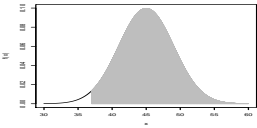
i.e.

$$X \sim N(45, 4)$$

What percentage of students obtain a mark

- larger than 45?
- larger than 45?
- larger than 50?
- between 40 and 50?
- greater than 37?

Table 1: Using *R* to obtain normal probabilities

	<i>R</i> Command	Probability	Area
$P(X < 45)$	<code>pnorm(45, 45, 4)</code>	.5	
$P(X > 45)$	<code>1- pnorm(45, 45, 4)</code>	.5	
$P(X > 50)$	<code>1- pnorm(50, 45, 4)</code>	0.1056498	
$P(40 < X < 50)$	<code>pnorm(50, 45, 4)</code> <code>-pnorm(40, 45, 4)</code>	0.7887005	
$P(X > 37)$	<code>1- pnorm(37, 45, 4)</code>	0.9772499	

Conditional Probabilities

Example

An analogue signal received at a detector (measured in microvolts), is normally distributed with a mean of 100 and a variance of 256;

$$X \sim N(100, 16)$$

What is the probability that the signal will be less than 120 microvolts, given that it is larger than 110 microvolts?

$$\begin{aligned} P(X < 120 | X > 110) &= \frac{P(X < 120 \cap X > 110)}{P(X > 110)} \\ &= \frac{P(110 < X < 120)}{P(X > 110)} \end{aligned}$$

For

$$P(110 < X < 120)$$

```
pnorm(120, 100, 16) - pnorm(110, 100, 16)  
[1] 0.1603358.
```

and

$$P(X > 110). \text{ 1-pnorm(110, 100, 16) = 0.2659855.}$$

Then

$$P(X < 120 | X > 110) = \frac{0.1603358}{0.2659855} = 0.6027988$$

Example

From previous records scores on an aptitude are normally distributed with mean $\mu = 500$ and standard deviation $\sigma = 100$:

$$X \sim N(500, 100)$$

What is the probability that an individual's score exceeds 650, given that it exceeds 500?

$$\begin{aligned} P(X > 650 | X > 500) &= \frac{P(X > 500 \cap X > 650)}{P(X > 500)} \\ &= \frac{P(X > 650)}{P(X > 500)} \end{aligned}$$

In R, `1-pnorm(650, 500, 100)` gives

$$P(X > 650) = 0.0668072$$

and `1-pnorm(500, 500, 100)` gives

$$P(X > 500) = 0.5$$

So

$$\frac{P(X > 650)}{P(X > 500)} = \frac{0.0668072}{.5} = 0.1336144$$

Quantiles

In the the aptitude tests find the cut off point for top 5% best test results.

Recall

$$X \sim N(500, 100)$$

(a) Choose k so that 5% of scores are above and 95% below. i.e.

$$P(X \leq k) = .95$$

the 95 % quantile.

Use *qnorm* in *R*:

```
qnorm(.95, 500, 100)  
[1] 664.4854
```


(b) What is the middle 40%?

Upper 30%, choose k so that

$$P(X > k) = .3$$

or equivalently

$$P(X \leq k) = .7$$

From R

```
qnorm(.70, 500, 100)
```

```
k = 552.44.
```

Lower 30%, choose k so that

$$P(X \leq k) = .3$$

```
qnorm(.30, 500, 100)
```

```
k = 447.5599.
```

(447.5599, 552.44) contains the middle 40% of the scores.

(c) Estimate the interquartile range: middle 50%.

First quartile

```
qnorm(.25, 500, 100)
```

```
[1] 432.551
```

Third quartile

```
qnorm(.75, 500, 100)
```

```
[1] 567.449
```

Interquartile range:(432.5, 567.449).

THE STANDARD NORMAL DISTRIBUTION

$$f(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-Z^2}{2}\right) \quad -\infty < Z < \infty$$

It this case

$$E(Z) = 0$$

$$V(Z) = 1$$

Z is said to be normally distributed with a mean of 0 and a variance of 1.

i.e.

$$Z \sim N(0, 1)$$

Clearly if

$$X \sim N(\mu, \sigma),$$

then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

The *CDF* of the standard normal variable is tabulated.

Area under the standard normal curve from $-\infty$ to z



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6338	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Reading Standard Normal Tables

- $P(Z < 0)$
- $P(Z > 0)$
- $P(Z > 1.25)$
- $P(-1.25 < Z < 1.25)$
- $P(Z > -2)$
- $P(-1.96 < Z < 1.96)$
- $P(Z < 2.58)$

Choose k so that

1. $P(-k < Z < k) = .95$
2. $P(-k < Z < k) = .99$
3. $P(-k < Z < k) = .68$

Examination Results:

$$X \sim N(45, 4)$$

$$Z = \frac{X - 45}{4}.$$

$$\begin{aligned} P(40 < X < 50) &= P\left(\frac{40 - 45}{4} < \frac{X - 45}{4} < \frac{50 - 45}{4}\right) \\ &= P(-1.25 < Z < 1.25) \end{aligned}$$

From tables

$$P(Z < 1.25) = 0.8944$$

Therefore

$$\begin{aligned} P(Z > 1.25) &= 1 - .8944 \\ &= 0.1056 \end{aligned}$$

By symmetry,

$$P(Z < -1.25) = 0.1056$$

Therefore

$$\begin{aligned} P(-1.25 < Z < 1.25) &= P(Z < 1.25) - P(Z < -1.25) \\ &= 0.8944 - 0.1056 \\ &= 0.7888 \end{aligned}$$

Analog Signals: $X \sim N(100, 16)$,

$$P(X < 120 | X > 110) = \frac{P(110 < X < 120)}{P(X > 110)}$$

$$= \frac{P((110 - 100)/16 < Z < (120 - 100)/16)}{P(Z > (110 - 100)/16)}$$

$$= \frac{P(0.625 < Z < 1.25)}{P(Z > .625)}$$

From tables:

$$P(Z < 1.25) = 0.8944$$

$$P(Z < 0.625) \approx 0.7340$$

Therefore

$$\frac{P(0.625 < Z < 1.25)}{P(Z > 0.625)} \approx \frac{0.8944 - .7340}{1 - 0.7340}$$

$$= \frac{0.1640}{0.2660}$$

$$= 0.6030$$