Chapter 5

Normal population distribution $X \sim N(\mu, \sigma^2)$

5.1 Definition of a standard Normal distribution, $Z \sim N(0,1)$

If the random variable Z has a standard Normal population distribution,

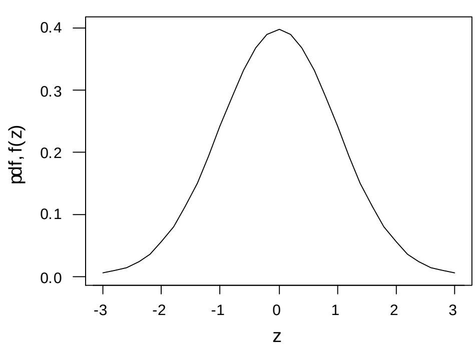
i.e., $Z \sim N(0,1)$ then its pdf and cdf are given by

pdf of Z $$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad -\infty < z < \infty$$

cdf of Z $$F_z(z) = \Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \qquad -\infty < z < \infty$$

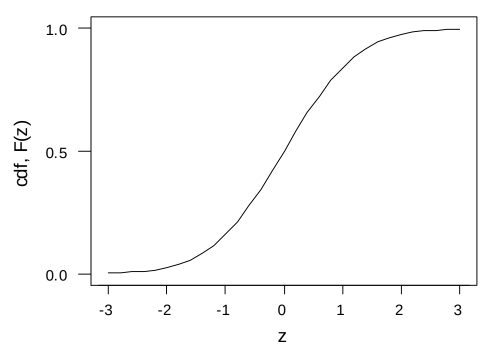
Graph of the population pdf and cdf for $Z \sim N(0,1)$

$f_Z(z)$



z

$F_Z(z) = \Phi(z)$



z

5.2 Definition of a general Normal distribution, $X \sim N(\mu, \sigma^2)$

If the random variable X has a standard Normal population distribution,

i.e., $X \sim N(\mu, \sigma^2)$ then its pdf and cdf are given by

pdf of X

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

cdf of X

$$F_X(x) = p(X \le x) = p(Z \le \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma}) \qquad -\infty < x < \infty$$

A Normal distribution is often a good model for many naturally occurring continuous random variables, e.g. lengths, weights, heights, temperatures, etc.

Note if $X \sim N(\mu, \sigma^2)$

Then

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

Examples of Normal distribution

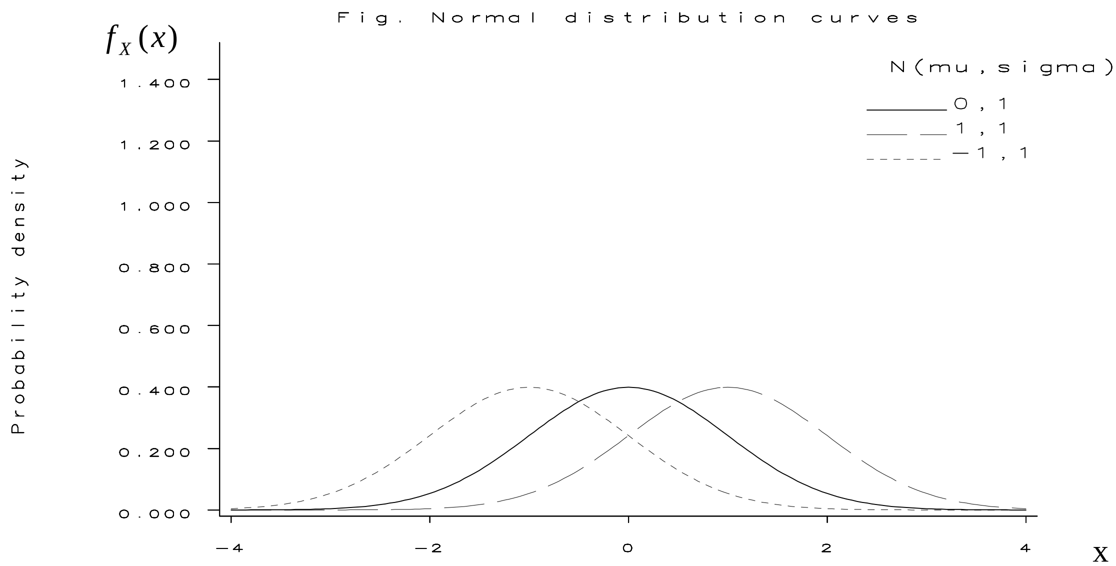


Figure : Normal distributions with the same variance but different means**.**

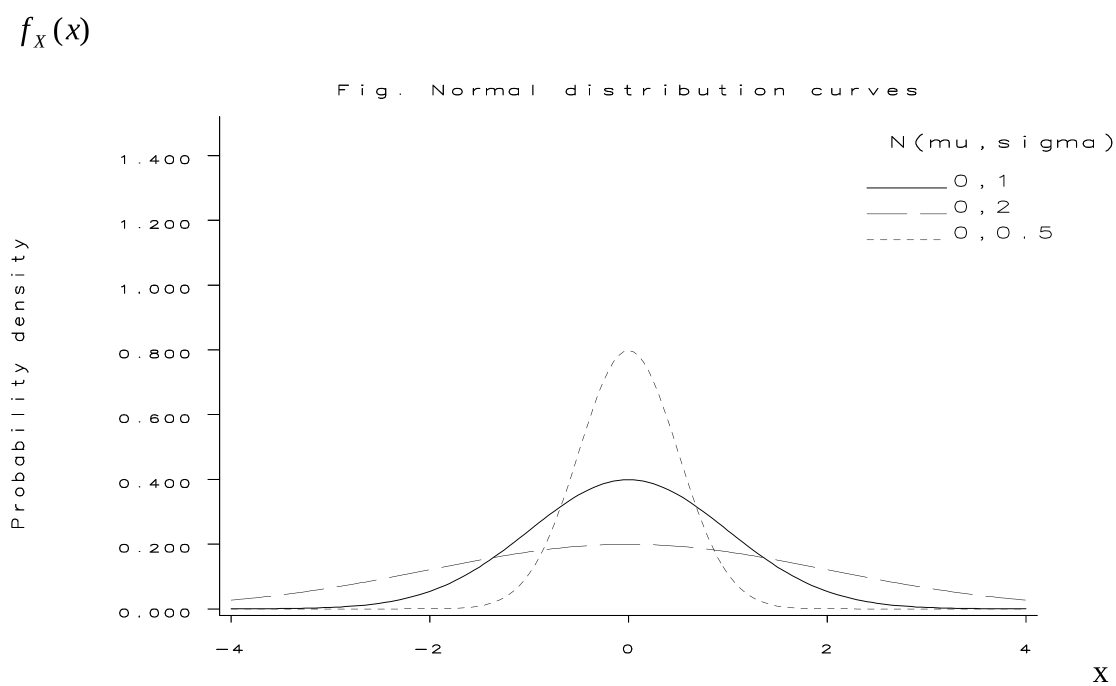


Figure : Normal distributions with the same mean but different variances**.**

5.3 Population summary measures for a Normal random variable

$$X \sim N(\mu, \sigma^2)$$

pdf	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	for $-\infty < x < \infty$

The following can be derived from the *population pdf* :

cdf	$F_X(x) = \Phi(\frac{x-\mu}{\sigma})$
mean	$\mu_X = \mu$
standard deviation	$\sigma_X = \sigma$
median,m	$m = \mu$
lower quartile, Q_1	$Q_1 = \mu - z_{0.25}\sigma = \mu - 0.67\sigma$
upper quartile, Q_3	$Q_3 = \mu + z_{0.25}\sigma = \mu + 0.67\sigma$
semi-interquartile range,	$SIR = \frac{Q_3 - Q_1}{2} = 0.67\sigma$
skessness,skew	*skew=0*
kurtosis, kurt	*kurt=0*

Derivation of the population lower quartile, Q_1 **for** $X \sim N(\mu, \sigma^2)$

$$F_X(Q_1) = 0.25 \Rightarrow p(X \le Q_1) = 0.25$$

$$\Rightarrow p(\frac{x-\mu}{\sigma} \le \frac{Q_1 - \mu}{\sigma}) = 0.25$$

$$\Rightarrow p(Z \le \frac{Q_1 - \mu}{\sigma}) = 0.25$$

$$\Rightarrow \frac{Q_1 - \mu}{\sigma} = -z_{0.25}$$

$$\Rightarrow Q_1 = \mu - z_{0.25}\sigma$$

Similarly Q_3 can be derived.

Derivation of a central 95% interval for $X \sim N(\mu, \sigma^2)$

$$p(c_1 < X < c_2) = 0.95 \Rightarrow p(\frac{c_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{c_2 - \mu}{\sigma}) = 0.95$$

$$\Rightarrow p(\frac{c_1 - \mu}{\sigma} < Z < \frac{c_2 - \mu}{\sigma}) = 0.95$$

$$\Rightarrow \frac{c_1 - \mu}{\sigma} = -z_{0.025} \text{ and } \frac{c_2 - \mu}{\sigma} = -z_{0.025}$$

$$\Rightarrow c_1 - \mu - z_{0.025}\sigma \text{ and } \Rightarrow c_2 - \mu - z_{0.025}\sigma$$

Hence a 95% interval for X = $(\mu - z_{0.025}\sigma, \mu + z_{0.025}\sigma)$,

A general $100(1-\alpha)\%$ **interval for X is given by** $(\mu - z_{\alpha/2}\sigma, \mu + z_{\alpha/2}\sigma)$,

5.4 Example

5.4.1 Data

The skull breadths of 84 ancient Etruscan males and 70 modern Italian males were measured :

Etruscan male skull breadths :

141	148	132	138	154	142	150	146	155	158	150	140
147	148	144	150	149	145	149	158	143	141	144	144
126	140	144	142	141	140	145	135	147	146	141	136
140	146	142	137	148	154	137	139	143	140	131	143
141	149	148	135	148	152	143	144	141	143	147	146
150	132	142	142	143	153	149	146	149	138	142	149
142	137	134	144	146	147	140	142	140	137	152	145

Italian male skull breadths :

133	138	130	138	134	127	128	138	136	131	126	120
124	132	132	125	139	127	133	136	121	131	125	130
129	125	136	131	132	127	129	132	116	134	125	128
139	132	130	132	128	139	135	133	128	130	130	143
144	137	140	136	135	126	139	131	133	138	133	137
140	130	137	134	130	148	135	138	135	138		

[From : Barnicot, N.A. and Brothwell, D.R. (1959) in 'Medical Biology and Etruscan Origins' and Hand et al. 'Small data sets' p121, ETRUSCAN.DAT, data set 155]

MINITAB

> Stat > Basic Statistics > Display Descriptive Statistics

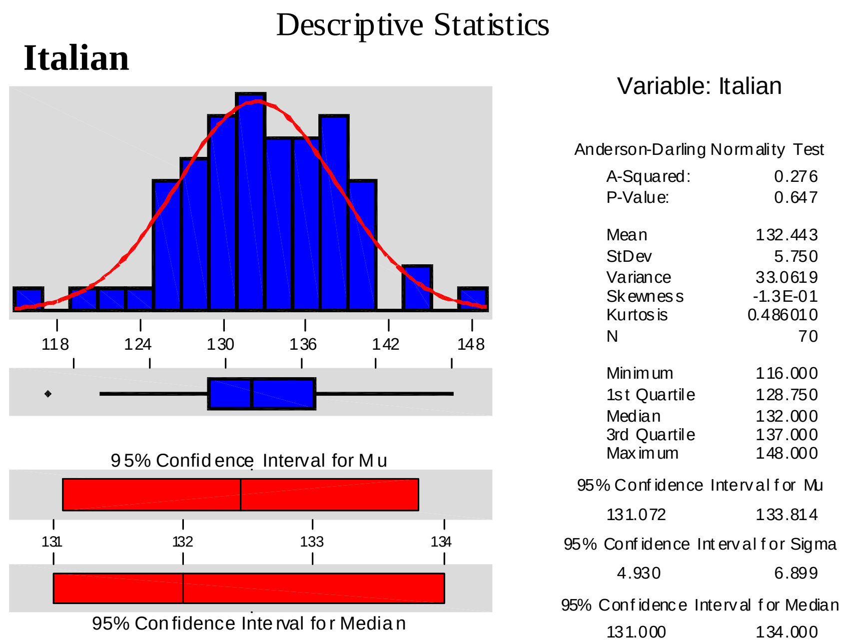
| Variables **'C1 C2'** |

> Graphs | O Graphical Summary | OK

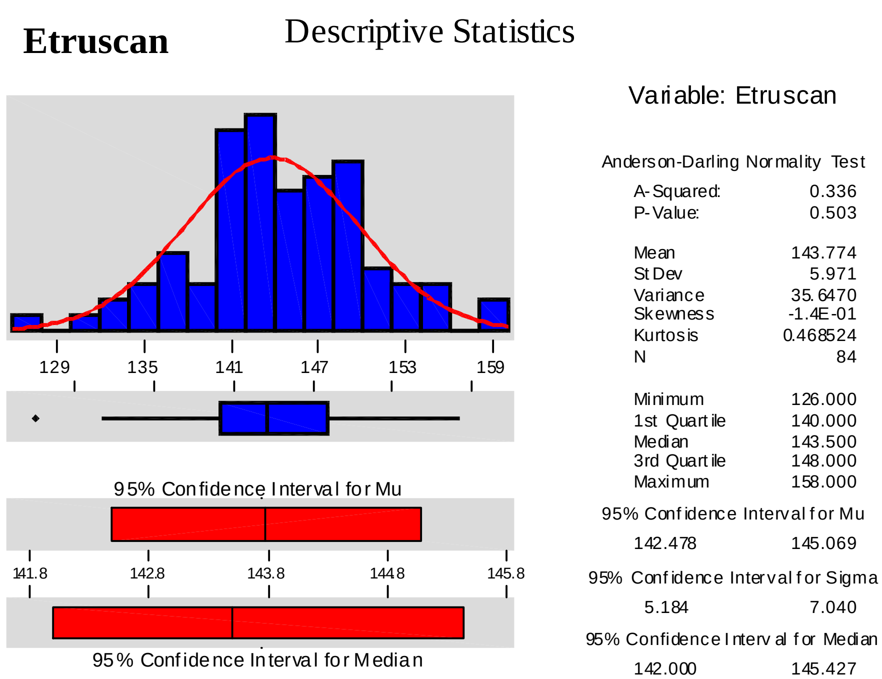
> OK

5.4.2 A comparison of skull breadths of ancient Etruscan and modern Italian with fitted Normal pdf's superimposed on the sample pdf's.

Italian



Etruscan



5.4.3 Comparison of the samples with Normal population distributions using Probaility Plots

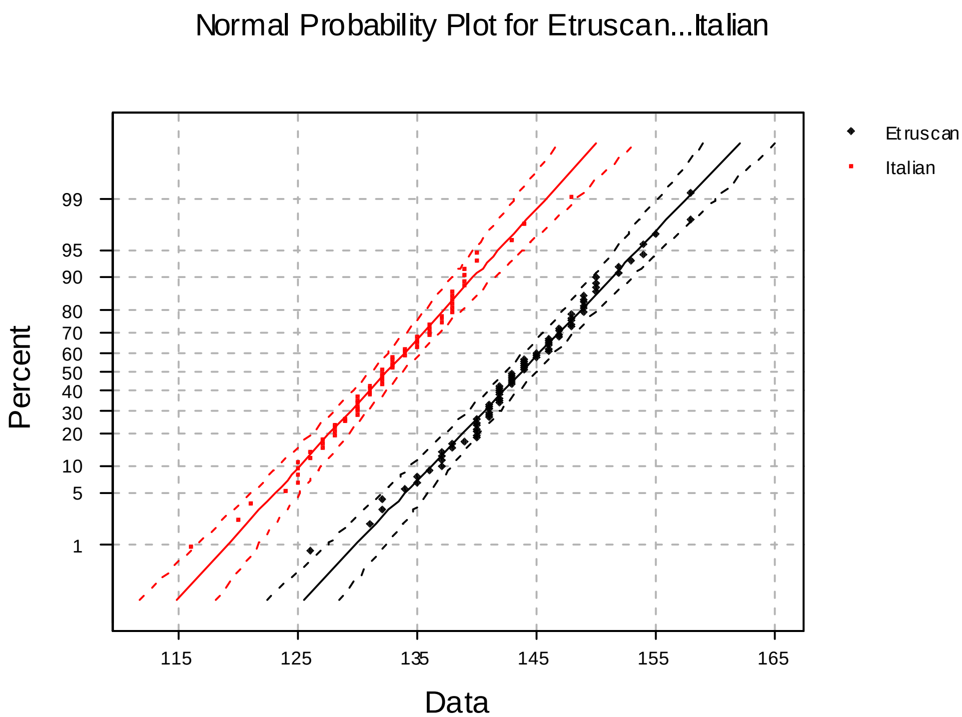
MINITAB

> Graph > Probability Plot

| Variables **'C1 C2'** |

> Distribution > Normal | OK

> OK



If the Normal model is OK, then the points plotted should roughly follow the middle line, lying within the two dashed bands. Here the Normal model appears very good for both the Italian and Etruscan skull breadths.

The fitted Normal models used in both the fitted Normal pdf curve and the Normal probability plot are fixed to have the same mean and standard deviation as the corresponding samples.

5.4.4 Compare the summary measures for the sample with those of the model

The fitted Normal models $X \sim N(\mu, \sigma^2)$ have the same mean and standard deviation as the corresponding samples.

	Italian		Etruscan	
Summary measure	Sample	Normal Model $X \sim N(132.4, 5.8^2)$	Sample	Normal Model $X \sim N(143.8, 6^2)$
Mean	132.4	132.4	143.8	143.8
Standard deviation	5.8	5.8	6	6
Median	132	132.4	143.5	143.8
Lower Quartile	128.75	128.5	140	139.8
Upper Quartile	137	136.3	148	147.8
SIR	4.1	3.9	4	4
Skewness, Skew	-0.1	0	-0.14	0
Kurtosis, Kurt	0.49	0	0.47	0

5.4.5 Conclusion

The Normal model for the data is very good for both Italian and Etruscan skull breadths from the following :

i) the comparison of the sample and Normal model pdfs
ii) the Normal probability plots
iii) the comparison of the sample and Normal model summary measures

5.4.5 Applying the model

The model could now be used to obtain estimated answers to questions relating to random variable X, the skull breadth of Etruscan males measure in mm, assuming that

$$X \sim N(143.8, 6^2)$$

Q1 Estimate the probability that the skull breadth is less than 140mm

$$p(X \le 140) = p\left(\frac{X-\mu}{\sigma} \le \frac{140-\mu}{\sigma}\right) = p\left(Z \le \frac{140-143.8}{6}\right) = p(Z \le -0.633) = 0.264$$

Q2 Estimate the probability that the skull breadth is more than 150mm

$$p(X \le 150) = p\left(\frac{X-\mu}{\sigma} \le \frac{150-\mu}{\sigma}\right) = p\left(Z \le \frac{150-143.8}{6}\right) = p(Z \le 1.033) = 1-0.151 = 0.849$$

$$p(X > 150) = 1 - p(X \le 150) = 1 - 0.849 = 0.151$$

Q3 Estimate the probability that the skull breadth is between 140 and 150 mm

$$p(140 < X < 150) = p(X \le 150) - p(X \le 140) = 0.849 - 0.264 = 0.585$$

Q4 Estimate the lower quartile Q_1 of the skull breadth

$F_X(Q_1) = p(X \le Q_1) = 0.25$, and hence

$Q_1 = \mu - z_{0.25}\sigma = \mu - 0.67\sigma \approx$ 143.8-0.67*6 = 139.8

Q5 Find an approximate 95% interval for the skull breadth

A 95% interval for X is given by

$(\mu - z_{0.025}\sigma, \mu + z_{0.025}\sigma) \approx$ (143.8-1.96*6 , 143.6+1.96*6) = (132.0 , 155.4)

5.5 MINITAB commands for calculating pdf, cdf and inverse cdf values

5.5.1 Calculating cdf $F_X(x) = p(X \le x) = p$ using MINITAB

Q1 To show that $p(X \le 140) = F_X(140) = 0.264$ when $X \sim N(143.8, 6^2)$

> Calc > Probability Distributions > Normal

⦿ Cumulative Probability | Mean **'143.8'** | Standard Deviation **'6'** | Input constant **'140'** | OK

Cumulative Distribution Function

```
Normal with mean = 143.800 and standard deviation = 6.00000

        x      P( X <= x)
 140.0000         0.2633
```

5.5.2 Calculating inverse cdf $x = F_X^{-1}(p)$ using MINITAB

Q2 To show that the lower quartile of X is $Q_1 = 139.8$ when

$F_X(Q_1) = p(X \le Q_1) = 0.25$ so that $Q_1 = F_X^{-1}(0.25)$

> Calc > Probability Distributions > Normal

⦿ Inverse Cumulative | Mean **'143.8'** | Standard Deviation **'6'** | Input constant **'0.5'** | OK

Inverse Cumulative Distribution Function

```
Normal with mean = 143.800 and standard deviation = 6.00000

 P( X <= x)          x
    0.2500      139.7531
```

5.5.3 Calculating pdf $f_X(x)$ using MINITAB

Q3 To show that the pdf $f_X(145) = 0.0652 = 0.0652$ when $X \sim N(143.8, 6^2)$

> Calc > Probability Distributions > Normal

⦿ Probability Density | Mean **'143.8'** | Standard Deviation **'6'** | Input constant **'145'** | OK

Probability Density Function

```
Normal with mean = 143.800 and standard deviation = 6.00000

        x       P( X = x)
 145.0000         0.0652
```