$M^{|}(t) = 5e^t\left(6-5e^t\right)^{-2}$ and $M^{||}(t) = 5e^t\left(6-5e^t\right)^{-2} + 50e^{2t}\left(6-5e^t\right)^{-3}$

$\Rightarrow E(X) = 5e^t\left(6-5e^t\right)^{-2}\Big|_{t-0} = 5$ and $E(Y^2) = \left[5e^t\left(6-5e^t\right)^{-2} + 50e^{2t}\left(6-5e^t\right)^{-3}\right]_{=0} = 55$

$Var(X) = E(X^2) - \mu^2 = 55 - 5^2 = 30$

Exercise

1) The mgf of a r.v Y is given by; a) $M(t) = e^{2t^2+3t}$ b) $M(t) = \exp\left\{\frac{1}{2}\sigma^2t^2 + t\mu\right\}$ Find the mean and variance of Y
2) A r.v X has a gamma distribution with parameters , Find the mgf of X hence obtain the mean and variance of X

3.1 The Mgf of a Sum of Independent Random Variables

The mgf of the sum of n independent random variable is the product of their individual mgf's
The mean (variance) of the sum of n independent random variable is the sum of their individual means (variances).
The mgf about $X = a$ is given by $M_{x,a}(t) = E\left[e^{t(x-a)}\right] = e^{-at}E\left[e^{tx}\right] = e^{-at}M_x(t)$

4 NORMAL DISTRIBUTION

4.1 Introduction

The normal, or Gaussian, distribution is one of the most important distributions in probability theory. It is widely used in statistical inference. One reason for this is that sums of random variables often approximately follow a normal distribution.

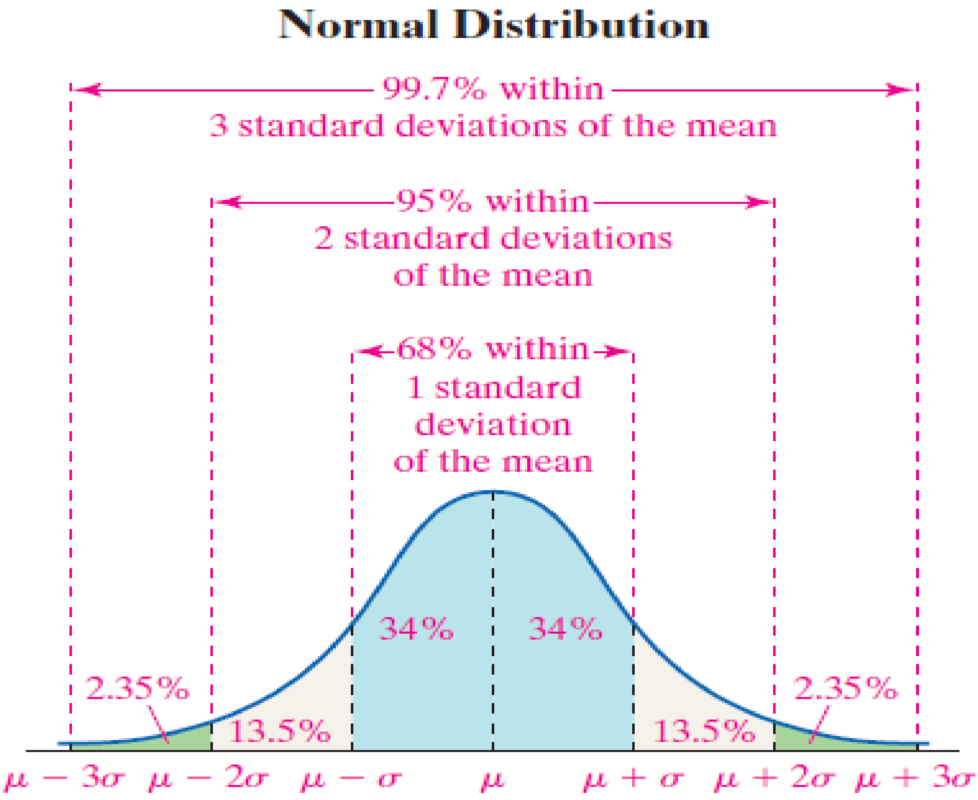
Definition A r.v X has a normal distribution with parameters μ and σ^2, abbreviated $X \sim N\left(\mu, \sigma^2\right)$ if it has probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \text{ for } -\infty < x < \infty \text{ and } \sigma > 0$$

Where μ is the mean and σ is the standard deviation.

4.1.1 Properties of normal distribution

1) The normal distribution curve is bell-shaped and symmetric, about the mean
2) The curve is asymptotic to the horizontal axis at the extremes.
3) The highest point on the normal curve is at the mean, which is also the median and mode.
4) The mean can be any numerical value: negative, zero, or positive
5) The standard deviation determines the width of the curve: larger values result in wider, flatter curves
6) Probabilities for the normal random variable are given by areas under the curve. The total area under the curve is 1 (0.5 to the left of the mean and 0.5 to the right).
7) It has inflection points at $\mu-\sigma$ and $\mu+\sigma$.
8) Empirical Rule:
   a) 68.26% of values of a normal random variable are within ± 1 standard deviation of its mean. ie $P\left(\mu-\sigma \le X \le \mu+\sigma\right) = 0.6826$
   b) 95.44% of values of a normal random variable are within ± 2 standard deviation of its mean. ie $P\left(\mu-2\sigma \le X \le \mu+2\sigma\right) = 0.9544$
   c) 99.72% of values of a normal random variable are within ± 3 standard deviation of its mean. ie $P\left(\mu-3\sigma \le X \le \mu+3\sigma\right) = 0.9972$



4.2 Standard Normal Probability Distribution

A random variable having a normal distribution with a mean of 0 and a varuance of 1 is said to have a **standard normal** probability distribution

Definition The random variable Z is said to have the standard normal distribution if $Z \sim N(0,1)$. Therefore, the density of Z, which is usually denoted $\phi(z)$ is given by;

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}z^2\right\} \text{ for } -\infty < z < \infty$$

The cumulative distribution function of a standard normal random variable is denoted $\Phi(z)$, and is given by

$$\Phi(z) = \int_{-\infty}^{z} \phi(t)dt$$

4.2.1 Computing Normal Probabilities

It is very important to understand how the standardized normal distribution works, so we will spend some time here going over it. There is no simple analytic expression for $\Phi(z)$ in terms of elementary functions. but the values of $\Phi(z)$ has been exhaustively tabulated. This greatly simplifies the task of computing normal probabilities..

Table 1 below reports the cumulative normal probabilities for normally distributed variables in standardized form (i.e. Z-scores). That is, this table reports $P(Z \le z) = \Phi(z)$). For a given value of Z, the table reports what proportion of the distribution lies below that value. For example, $P(Z \le 0) = \Phi(0) = 0.5$; half the area of the standardized normal curve lies to the left of $Z = 0$.

Theorem: It may be useful to keep in mind that

i) $P(Z > z) = 1 - \Phi(z)$ complementary law

ii) $P(Z \le -z) = P(Z \ge z) = 1 - \Phi(z)$ ie due to symmetry

$\Rightarrow \quad \Phi(z) + \Phi(-z) = 1$ Since $P(Z \le z) + P(Z \ge z) = 1$

iii) $P(a \le z \le b) = \Phi(b) - \Phi(a)$

iv) $P(-a \le z \le a) = 2\Phi(a) - 1$ since

$P(-a \le z \le a) = \Phi(a) - \Phi(-a) = \Phi(a) - [1 - \Phi(a)] = 2\Phi(a) - 1$

v) If we now make $\Phi(a)$ the subject, then $\Phi(a) = \frac{1}{2}[1 + P(-a \le z \le a)]$

Table 1

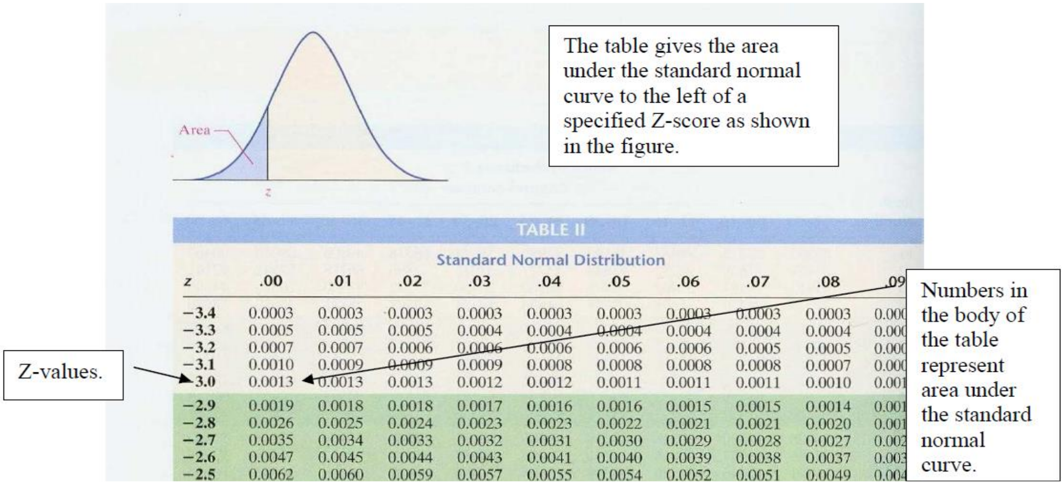


TABLE II

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.000
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.000
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.000
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.000
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.001
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.001
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.001
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.002
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.003
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.004

Example 1

Given $Z \sim N(0,1)$, find;

a) $P(Z \le z)$ if $z = 1.65, -1.65, 1.0, -1.0$

b) $P(Z > z)$ for $z = 1.02, -1.65$

c) $P(0.365 \le z \le 1.75)$

d) $P(-0.696 \le z \le 1.865)$

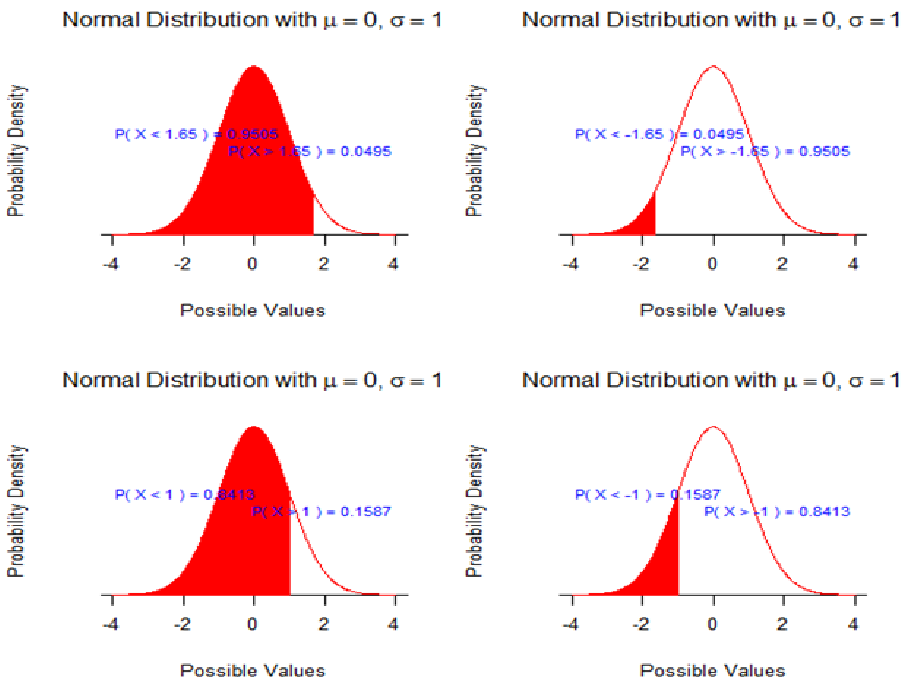
e) $P(-2.345 \le z \le -1.65)$

f) $P(|z| \le 1.43)$

Solution

a) Look up and report the value for $\Phi(z)$ from the standard normal probabilities table

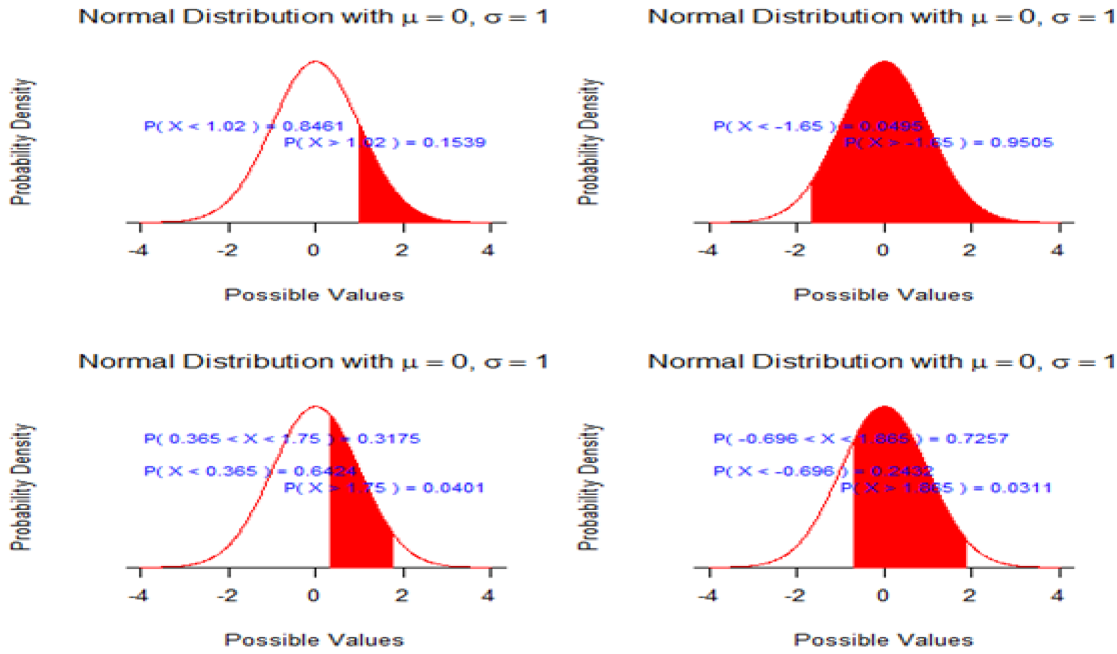
$P(Z \le 1.65) = \Phi(1.65) = 0.9505 \quad \Phi(-1.65) = 0.0495 \quad \Phi(1.0) = 0.8413 \quad \Phi(-1.0) = 0.1587$



b) $P(Z > z) = \Phi(-z)$ Thus $P(Z > 1.02) = \Phi(-1.02) = 0.1515 \quad P(Z > -1.65) = \Phi(1.65) = 0.9505$

c) $P(0.365 \le z \le 1.75) = \Phi(1.75) - \Phi(0.365) = 0.9599 - 0.6350 = 0.3249$

d) $P(-0.696 \le z \le 1.865) = \Phi(1.865) - \Phi(-0.696) = 0.9689 - 0.2432 = 0.3249 = 0.7257$



e) $P(-2.345 \le z \le -1.65) = \Phi(-1.65) - \Phi(-2.345) = 0.0505 - 0.0095 = 0.0410$

f) $P(|z| \le 1.43) = P(-1.43 \le z \le 1.43) = 2\Phi(1.43) - 1 = 2(0.9236) - 1 = 0.8472$

Example 2

If $Z \sim N(0,1)$, find the value of t for which;

a) $P(Z \le t) = = 0.6026, \quad 0.9750, \quad 0.3446$

b) $P(Z > t) == 0.4026, 0.7265, \ 0.5446$

c) $P(-0.28 \le z \le t) = 0.2665$

d) $P(-t \le z \le t) = 0.9972, 0.9505, 0.9750$

Solution

Here we find the probability value in Table I, and report the corresponding value for Z.

a) $\Phi(t) = 0.6026 \Rightarrow t = 0.26 \quad \Phi(t) = 0.950 \Rightarrow t = 1.96 \quad \Phi(t) = 0.3446 \Rightarrow t = -0.40$

b) $P(Z > t) = 0.4026 \Rightarrow \Phi(t) = 0.5974 \Rightarrow t = 0.25$

$P(Z > t) = 0.7265 \Rightarrow \Phi(t) = 0.2735 \Rightarrow t = -0.60$

$P(Z > t) = 0.5446 \Rightarrow \Phi(t) = 0.4554 \Rightarrow t = -0.11$

c) $P(-0.28 \le z \le t) = \Phi(t) - \Phi(-0.28) = 0.2665 \Rightarrow \Phi(t) = 0.3897 + 0.2665 \Rightarrow t = 0.40$

d) $P(-t \le z \le t) = 2\Phi(t) - 1 = 0.9972 \Rightarrow \Phi(t) = 0.9986 \Rightarrow t = 2.99$

$P(-t \le z \le t) = 2\Phi(t) - 1 = 0.9505 \Rightarrow \Phi(t) = 0.9753 \Rightarrow t = 1.96$

$P(-t \le z \le t) = 2\Phi(t) - 1 = 0.9750 \Rightarrow \Phi(t) = 0.9875 \Rightarrow t = 2.24$

Exercise

1..Given $Z \sim N(0,1)$, find;

a) $P(Z \le z)$ if

$z = 1.95, -1.89, 1.074, \ -1.53$

b) $P(Z > z)$ for $z = 1.72, -1.15$

c) $P(0 \le z \le 1.05)$

d) $P(-1.396 \le z \le 1.125)$

e) $P(-1.96 \le z \le -1.65)$

f) $P(|z| \le 2.33)$

2..If $Z \sim N(0,1)$, find the value of z for which;

a) $P(Z \le a) = = 0.973, \quad 0.6693, \quad 0.4634$

b) $\mathrm{P}(Z > a) == 0.3719,\ 0.9545,\ 0.7546$

c) $\mathrm{P}(-1.21 \le z \le t) = 0.6965$

d) $\mathrm{P}(|z| \le t) = 0.9544,\ 0.9905,\ 0.3750$

4.3 The General Normal Density

Consider $Z \sim N(0,1)$ and let $X = \mu + \sigma Z$ for $\sigma > 0$. Then $X \sim N(\mu, \sigma^2)$ But we know that

$f(x) = \frac{1}{\sigma}\phi\left(\frac{X-\mu}{\sigma}\right)$ from which the claim follows. Conversely, if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$. It is also easily shown that the cumulative distribution function satisfies

$$F(x) = \Phi\left(\frac{X-\mu}{\sigma}\right)$$

and so the cumulative probabilities for any normal random variable can be calculated using the tables for the standard normal distribution..

Definition A variable X is said to be standardized if it has been adjusted (or transformed) such that its mean equals 0 and its standard deviation equals 1. Standardization can be accomplished using the formula for a z-score: $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$. The z-score represents the number of standard deviations that a data value is away fromthe m ean.

Let $X \sim N(\mu, \sigma^2)$ then $P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$ where $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

Example 1 A r.v $X \sim N(50, 25)$ compute $P(45 \le z \le 60)$

Solution

$\mu = 50$ and $\sigma = 5 \Rightarrow Z = \frac{x-50}{5} \sim N(0,1)$

$P(45 \le X \le 60) = P\left(\frac{45-50}{5} \le Z \le \frac{60-50}{5}\right) = \Phi(2) - \Phi(-1) = 0.9772 - 0.1587 = 0.8185$

Example 2 Suppose $X \sim N(30, 16)$. Find; a) $P(X < 40)$ b) $P(X > 21)$ c) $P(30 < X < 35)$

Solution

$X \sim N(30,16) \Rightarrow Z = \frac{X-30}{4} \sim N(0,1)$

a) $P(X \le 40) = P\left(Z \le \frac{40-30}{4}\right) = \Phi(2.5) = 0.9938$

b) $P(X > 21) = P\left(Z > \frac{21-30}{4}\right) = P(Z > -2.25) = P(Z \le 2.25) = \Phi(2.25) = 0.9878$

c) $P(30 < X < 35) = P\left(\frac{30-30}{4} \le Z \le \frac{35-30}{4}\right) = P(0 < Z < 1.25) = 0.8944 - 0.5 = 0.3944$

Example 3

The top 5% of applicants (as measured by GRE scores) will receive scholarships. If $GRE \sim N(500, 100^2)$, how high does your GRE score have to be to qualify for a scholarship?

Solution

Let $X = GRE$. We want to find x such that $P(X \ge x) = 0.05$ this is too hard to solve as it stands - so instead, compute $Z = \frac{X-500}{100} \sim N(0,1)$ and find z for the problem,

$$P(Z \ge z) = 1 - \Phi(z) = 0.05 \quad \Rightarrow \quad \Phi(z) = 0.95 \quad \Rightarrow \quad z = 1.645$$

To find the equivalent x, compute $X = \mu + \sigma Z \Rightarrow x = 500 + 100(1.645) = 66.5$

Thus, your GRE score needs to be 665 or higher to qualify for a scholarship.

Example 4

Family income is believed to be normally distributed with a mean of $25000 and a standard deviation on $10000. If the poverty level is $10,000, what percentage of the population lives in poverty? A new tax law is expected to benefit "middle income" families, those with incomes between $20,000 and $30,000. What percentage of the population will benefit from the law?

Solution

Let X = Family income. We want to find P(X ≤ $10,000)., so

$X \sim \mathrm{N}\left(25000, 10000^2\right) \Rightarrow \; Z = \frac{X-25000}{10000} \sim \mathrm{N}(\,0,1)$

$\mathrm{P}(X \le 10{,}000) = \mathrm{P}(Z \le -1.5) = \Phi(-1.5) = 0.0668$.

Hence, a slightly below 7% of the population lives in poverty.

$\mathrm{P}(20{,}000 \le X \le 30{,}000) = \mathrm{P}(-0.5 \le Z \le 0.5) = 2\Phi(0.5) - 1 = 2 \times 0.6915\text{-}1 - 0.383$

Thus, about 38% of the taxpayers will benefit from the new law.

Exercise

1) Suppose $X \sim \mathrm{N}(130, 25)$. Find; a) $P(X < 140)$ b) $P(X > 120)$ c) $P(130 < X < 135)$
2) The random variable X is normally distributed with mean 500 and standard deviation 100. Find; (i) $P(X < 400)$, (ii) $P(X > 620)$ (iii) the 90$^{\text{th}}$ percentile (iv) the lower and upper quartiles. Use graphs with labels to illustrate your answers.
3) A radar unit is used to measure speeds of cars on a motorway. The speeds are normally distributed with a mean of 90 km/hr and a standard deviation of 10 km/hr. What is the probability that a car picked at random is travelling at more than 100 km/hr?
4) For a certain type of computers, the length of time bewteen charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. John owns one of these computers and wants to know the probability that the length of time will be between 50 and 70 hours
5) Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585. Will he be admitted to this university?
6) A large group of students took a test in Physics and the final grades have a mean of 70 and a standard deviation of 10. If we can approximate the distribution of these grades by a normal distribution, what percent of the student; (a) scored higher than 80? (b) should pass the test (grades≥60)? (c) should fail the test (grades<60)?
7) A machine produces bolts which are N(4 0.09) where measurements are in cm. Bolts are measured accurately and any bolt smaller than 3.5 cm or larger than 4.4 cm is rejected. Out of 500 bolts how many would be accepted? Ans 430
8) Suppose IQ ~ N(100,22.5).a woman wants to form an Egghead society which only admits people with the top 1% IQ score. What should she have to set the cut-off in the test to allow this to happen? Ans 134.9
9) A manufacturer does not know the mean and standard deviation of ball bearing he is producing. However a sieving system rejects all the bearings larger than 2.4 cm and those under 1.8 cm in diameter. Out of 1,000 ball bearings, 8% are rejected as too small and 5.5% as too big. What is the mean and standard deviation of the ball bearings produced? Ans mean=2.08 sigma=0.2