Ch. 5 - The Normal Distribution

1 Definition and Properties

Normal Distribution:

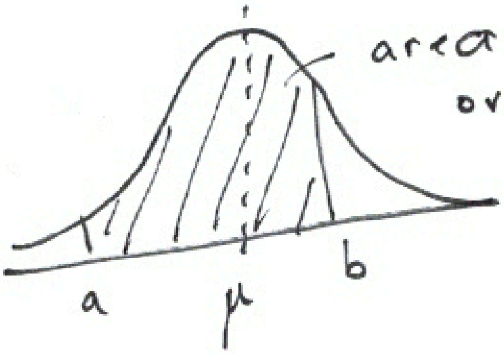
- If a continuous random variable X is said to follow a normal distribution with mean μ and standard deviation σ, we write

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

The normal distribution has a continuous density, which is summarized by two **parameters** (numerical characteristics of a model). The location of the distribution is determined by the value of the mean (μ) and spread is determined by the standard deviation (σ). The density function for a normal random variable is

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

(Note: we rarely use this in calculations for STAT251)



area under the normal curve over [a, b] gives percentage of observations that have values between a and b.

2 Finding Probabilities

Areas under the normal density can be found using: computer software and statistical tables. Statistical tables only give areas under the $N(0,1)$ curve. (Statistical table posted on course website)

Example 2. Scores on a standard IQ test follow approximately the Normal model with mean $\mu = 110$ and standard deviation $\sigma = 25$.

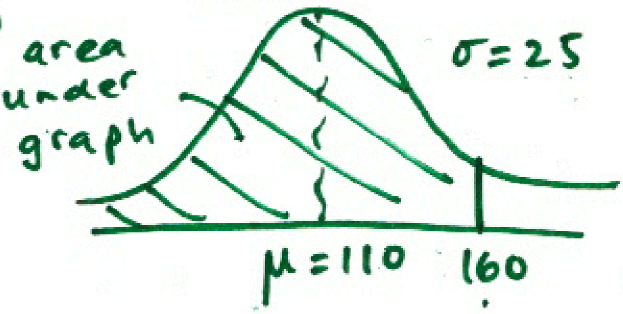
(a) What percentage of people have IQ scores below 160?

(b) What percentage of people have scores between 90 and 120?

(c) How high is the IQ such that only 0.15% of people fall above?

<u>Solution:</u> Let X rep. IQ score.

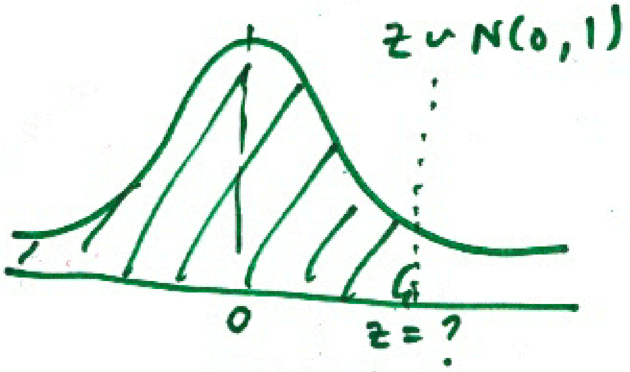
a)



$P(X < 160) = ?$

$$z = \frac{x-\mu}{\sigma} = \frac{160-110}{25}$$

$$= 2$$



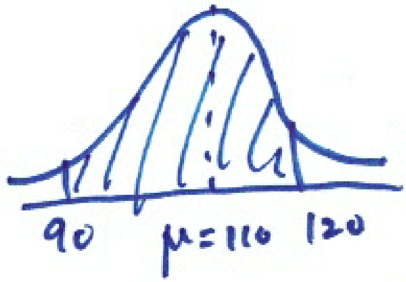
$P(X < 160) =$

$$P\left(\frac{X-\mu}{\sigma} < \frac{160-110}{25}\right)$$

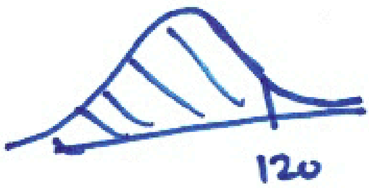
$$= P(Z < 2)$$

$$= 0.9772$$ (look at table)

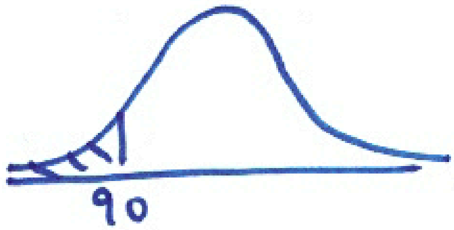
b) $P(90 < X < 120) = ?$



$=$



$-$



$$P(90 < X < 120) = P(X < 120) - P(X < 90)$$

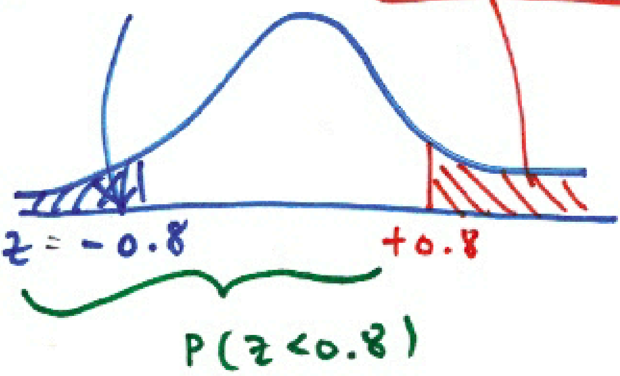
$$= P\left(Z < \frac{120-110}{25}\right) - P\left(Z < \frac{90-110}{25}\right)$$

$$= P(Z < 0.4) - P(Z < -0.8)$$

$P(Z < 0.4) = \underline{0.6554}$ (look in z table)

$P(Z < -0.8) = 1 - P(Z < 0.8) = 1 - 0.7881$

$= \underline{\underline{0.2119}}$



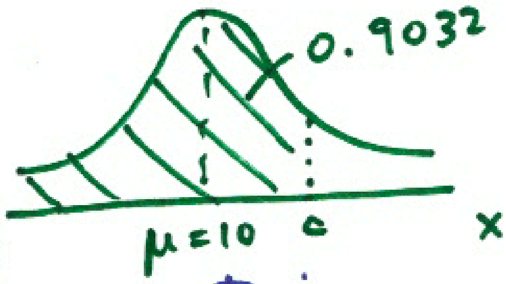
$P(90 < X < 120) = 0.6554 - 0.2119 = 0.4435$

c)

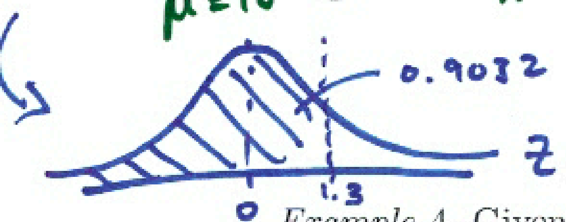
$= \sigma^2$

Example 3. Given $X \sim N(10, 25)$, find c such that $P(X < c) =$ 0.9032

Solution:



$$P(Z < \frac{c-10}{5}) = P(Z < 1.3)$$
$$\Longrightarrow \frac{c-10}{5} = 1.3$$
$$c = (1.3 \times 5) + 10 = 16.5$$

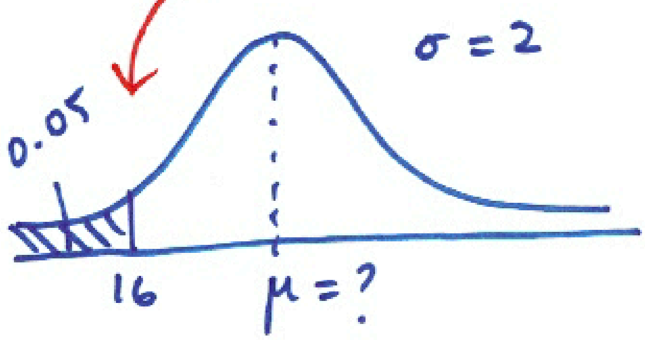


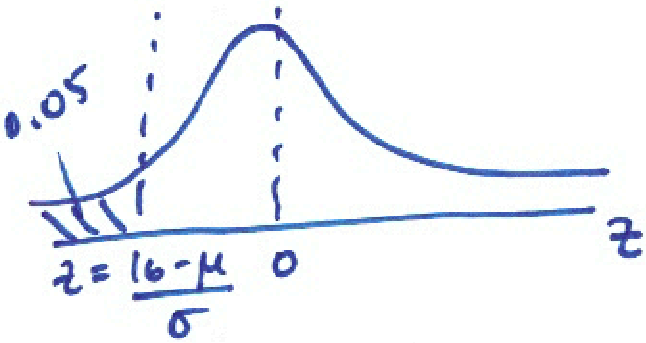
$\sigma^2 = 4 \to \sigma = 2$

Example 4. Given $X \sim N(\mu, 4)$ and $P(X < 16) = 0.05$. Find μ.

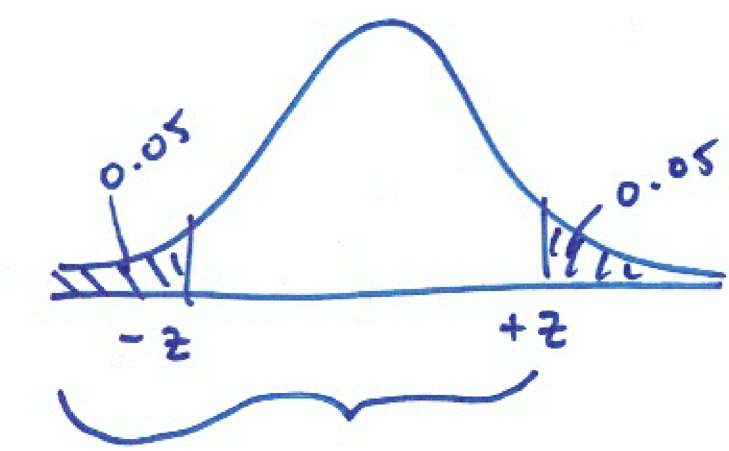
Solution:

$$P(X < 16) = P(\frac{X-\mu}{\sigma} < \frac{16-\mu}{2}) = P(Z < \frac{16-\mu}{2}) = 0.05$$
$$\Longrightarrow \frac{16-\mu}{2} = -1.645$$
$$\mu = (1.645 \times 2) + 16 = 19.29$$





look up 1 - 0.05 = 0.95



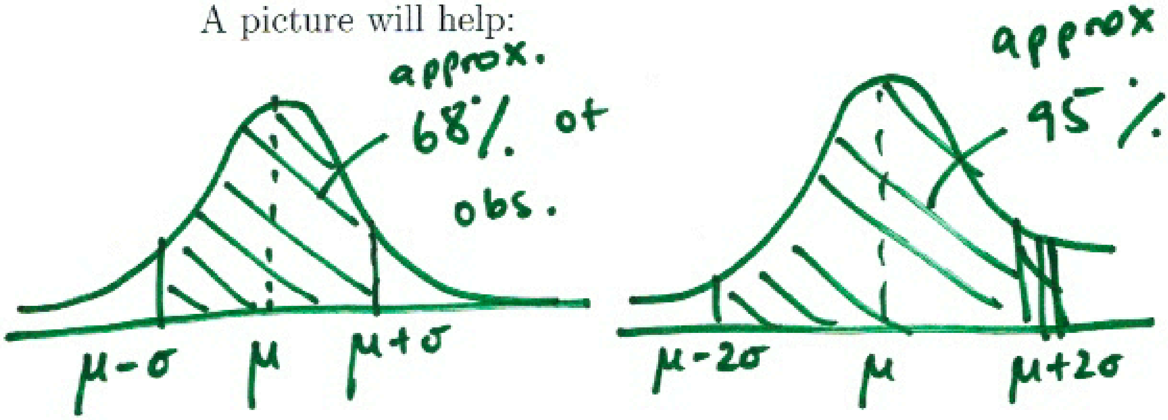
68-95-99.7 Rule:

A useful rule for the normal distribution is the **68-95-99.7** or **empirical rule**:

If $X \sim N(\mu, \sigma)$ then approximately,

- 68% of the observations will fall within 1 standard deviation of the mean
- 95% of the observations will fall within 2 standard deviations of the mean
- 99.7% of the observations will fall within 3 standard deviations of the mean

A picture will help:

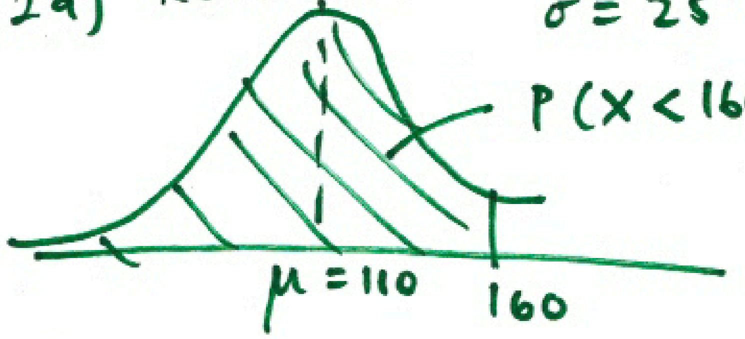


Let's revisit example 2(a) and verify we get approximately the same answer using this method...

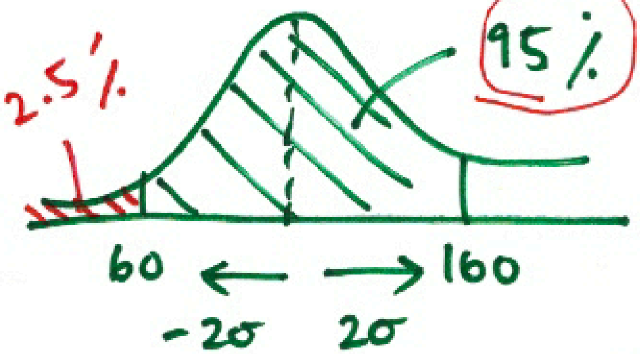
2a) Revisited

$\sigma = 25$

$P(X < 160) = ?$



$$z = \frac{160 - 110}{25} = 2$$



approx 95% by the 68-95-99.7% rule

So $95\% + 2.5\%$

$= 97.5\%$

approx. 97.5% of values fall below 160

3 Independent Normal Samples

The sums (and differences) of independent normal variables are also normal

- If $X \sim N(10, 25)$ and $Y \sim N(5, 16)$ are independent, then $X + Y$ is also normal.

$$\begin{aligned} E(X+Y) &= E(X) + E(Y) \\ &= 10 + 5 = 15 \\ Var(X+Y) &= Var(X) + Var(Y) \qquad X,\ Y \ \underline{\text{independent}} \\ &= 25 + 16 = 41 \\ \Longrightarrow\ X + Y &\sim N(15, 41) \end{aligned}$$

$E(X-Y) = E(X) - E(Y) = 5$

What about $X - Y$?
$\Longrightarrow\ X - Y \sim N(5, 41)$

$Var(X-Y) = Var(X) + Var(Y)$

- If $X_1, X_2, ..., X_n$ are n independent observations of normal variable $X \sim N(\mu, \sigma^2)$, then

$$X_1 + X_2 + ... + X_n \sim N(n\mu, n\sigma^2)$$

$$\begin{aligned} E(X_1 + \ldots + X_n) &= E(X_1) + E(X_2) + \ldots + E(X_n) \\ &= \mu + \mu + \ldots + \mu \\ &= n\mu \\ Var(X_1 + \ldots + X_n) &= n\,Var(X) \end{aligned}$$

Example 5. A certain company manufactures stereo systems. The times required to pack the stereos can be described by a Normal model with mean of 9 minutes and standard deviation of 1.5 minutes. The times for the boxing stage can be modelled as Normal with a mean of 6 minutes and standard deviation of 1 minute. Assume that the two packing times are independent.

(a) What is the probability that packing two consecutive systems takes over 20 minutes?

(b) What percentage of the stereo systems take longer to pack than to box? Assume that the packing and boxing times are independent.

Solution:

Let P_1 = time pack 1st system.

P_2 = time pack 2nd system.

$T = P_1 + P_2$ total time to pack two systems

→ we are told both r.v ~ Normal

→ times independent.

$P_i \sim N(\underline{9}, 1.5^2) \qquad i = 1, 2$

$T \sim N(?, ?)$

$E(T) = E(P_1 + P_2) = E(P_1) + E(P_2)$

$= 9 + 9 = \underline{18}$ minutes.

$Var(T) = Var(P_1 + P_2) = Var(P_1) + Var(P_2)$

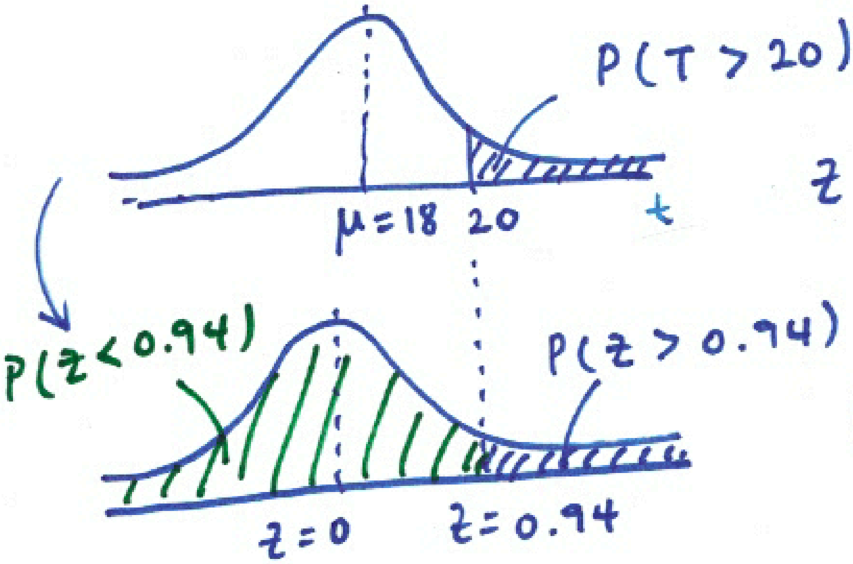
times independent

$= 1.5^2 + 1.5^2 = 4.50$

$SD(T) = \sqrt{4.50} = 2.12$ min.

$T \sim N(18, 2.12^2)$

$P(T > 20) = ?$



$$z = \frac{20 - 18}{2.12} = 0.94$$

$$P(T > 20) = P\left(\frac{T - \mu}{\sigma} > \frac{20-18}{2.12}\right)$$
$$= P(z > 0.94)$$
$$= 1 - P(z < 0.94)$$
$$= 1 - 0.8264$$
$$= 0.1736$$

There is a 17.36% that it will take over 20 min to pack two systems.

b) we want to estimate the percentage of stereo systems that take longer to pack than to box.

P = time packing a system

B = time boxing a system.

D = P − B difference in times to pack and box.

$P(P > B) = P(P - B > 0) = P(D > 0)$

$D \sim N(?, ?)$

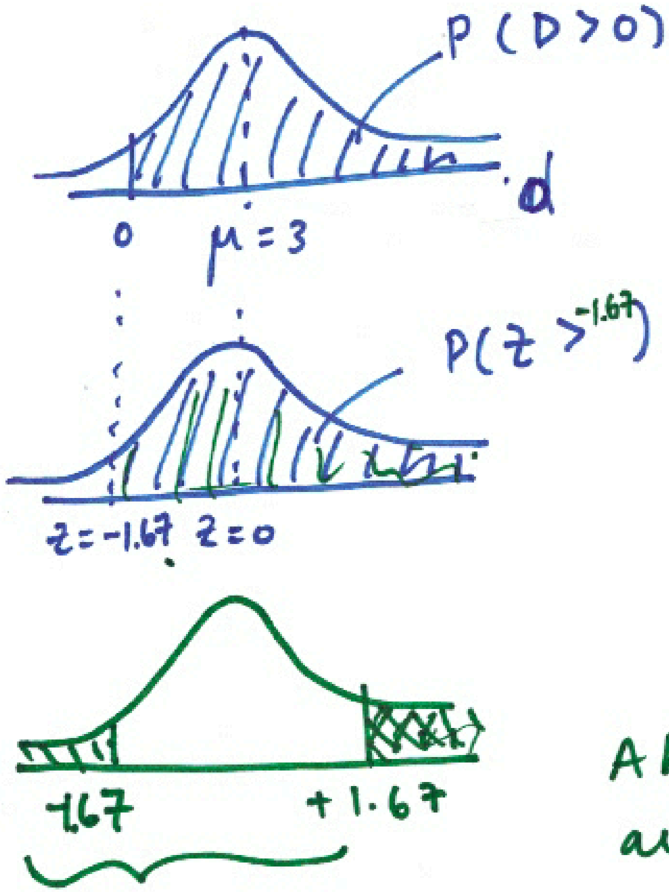
$E(D) = E(P - B) = E(P) - E(B) = 9 - 6 = 3$ min.

$Var(D) = Var(P - B) = Var(P) + Var(B)$

$= 1.5^2 + 1^2 = 3.25$

$SD(D) = \sqrt{3.25} = 1.8$ minutes.

$D \sim N(3, 1.80^2)$



$P(D > 0) = P\left(z > \frac{0-3}{1.8}\right)$

$= P(z > -1.67)$

$= 1 - P(z < -1.67)$

$= 1 - (1 - P(z < 1.67))$

$= P(z < 1.67)$

$= 0.9525$

About 95.25% of all stereo systems require more time for packing than boxing.

The Sample Mean

Let $X_1, X_2, ..., X_n$ be a random sample of n independent observations If $X_i \sim N(\mu, \sigma^2)$, then the sample mean $\bar{X}$ also follows a normal distribution and

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

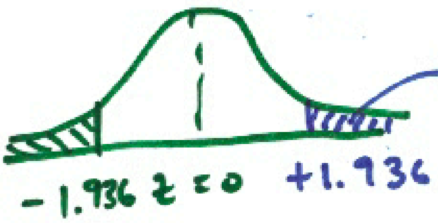
Example 6. A random sample of size 15 is taken from a normal distribution with mean 60 and standard deviation 4. Find the probability that the mean of the sample is less than 58.

Solution:

$$X \sim N(60, 4^2) \qquad \bar{X} = \frac{X_1 + \dots + X_n}{n}$$

$$\bar{X} \sim N(60, \frac{4^2}{15})$$

$$\Rightarrow P(\bar{X} < 58) = P\left(Z < \frac{58 - 60}{\sqrt{4^2/15}}\right)$$



$$= P(Z < -1.936)$$

$$= 1 - P(Z < 1.936)$$

look in table

$$= 1 - 0.9732$$

$$= 0.0268$$

$$E(\bar{X}) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n}\left[E(X_1) + \ldots + E(X_n)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \ldots + \mu\right]$$

$$= \frac{1}{n}\left[n\mu\right] = \mu$$

$$Var(\bar{X}) = Var\left(\frac{X_1 + \ldots + X_n}{n}\right)$$

$$= \frac{1}{n^2}\left[Var(X_1) + \ldots + Var(X_n)\right] \quad \text{indep.}$$

$$= \frac{1}{n^2}\left(\sigma^2 + \ldots + \sigma^2\right] = \frac{1}{n^2} n \sigma^2$$

$$= \frac{\sigma^2}{n}$$

$$\therefore \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$