Ch5 Normal Distribution

① $$X \sim N(100, 36)$$

Normal $\quad \mu$(mean) $\quad \sigma^2$(variance)

A continuous random variable X is said to follow a Normal distribution with mean μ and standard deviation σ if its probability density function is

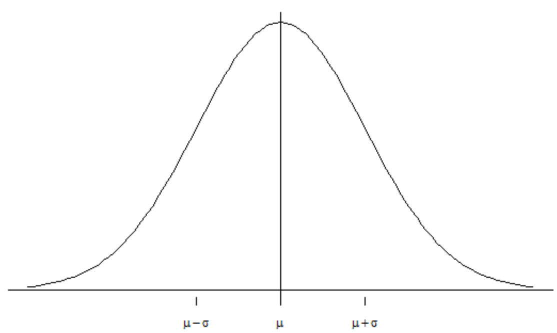
$$\underbrace{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},} \quad -\infty < x < \infty$$

We rarely use this in calculation for STAT 251.

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

The Normal distribution with $\mu = 0$ and $\sigma = 1$ is called a standard Normal distribution. We denote the standard Normal variable by $Z \sim N(0, 1)$.

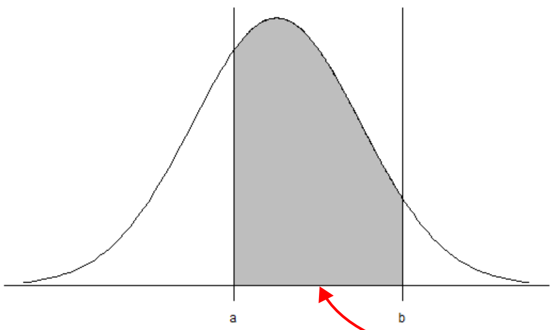
③



Normal distribution is bell-shaped and symmetrical about the line $x = \mu$. The mean, median and mode (where pdf attains max value) coincide.

④ If $Z \sim N(0, 1) \leftarrow$ standard Normal table

then the probability that Z lies between a and b is given by the area



$area = \int_a^b f(x)dx$, but it is hard to evaluate, so we use standard Normal table.

Note: A copy of the standard Normal table is posted on Connect.

Eg

$P(Z < -2.1) =$? 0.0179 from table

$P(Z > -2.12) = 1 - P(Z \leq -2.12) = 1 - 0.0170 = 0.983 \rightarrow$ Note: Instead of F(-2.12), it is customary to use $\Phi(-2.12)$ for Normal distributions

$P(1.0 \leq Z \leq 2.09) = \Phi(2.09) - \Phi(1.0) = 0.9817 - 0.8413 = 0.1404$

Now that you know how to calculate probability given the z-scores, make sure you know how to find z-scores given the probability (as in the next question)?

Find a such that $P(Z < a) = 0.7$

Ans: [$P(Z < 0.525) = 0.7$ from table] $\Rightarrow P(Z < a) = P(Z < 0.525) \Rightarrow a = 0.525$

Most of the time, the question tells us $X \sim N(\underset{\mu}{100}, \underset{\sigma^2}{36})$ and the aim is say to find $P(X \leq 90)$.

How?

We first express in terms of Z and then use the standard Normal table to get the probability.

$$\boxed{Z = \frac{X - \mu}{\sigma}} \leftarrow \textit{Important!}$$

Eg. $X \sim N(10, 25)$. Find $P(X < 13)$.

$$P(X < 13) = P(Z < \tfrac{13-10}{5}) = P(Z < 0.6) = 0.7257 \text{ (from table)}$$

$\sigma = 5$ since $\sigma^2 = 25$

What about $P(X > 13) = 1 - P(X \leq 13) = 1 - 0.7257 = 0.2743$

What about

$$\begin{aligned} P(11 \leq X \leq 13) &= P(\frac{11-10}{5} \leq Z \leq \frac{13-10}{5}) \\ &= \Phi(0.6) - \Phi(0.2) \\ &= 0.7257 - 0.5793 \\ &= 0.1463 \end{aligned}$$

What about $P(X < -5) = P(Z < \frac{-5-10}{5}) = P(Z < -3) = 0.0013$

⑤ All Normal distribution curves satisfy the 68-95-99.7% rule:

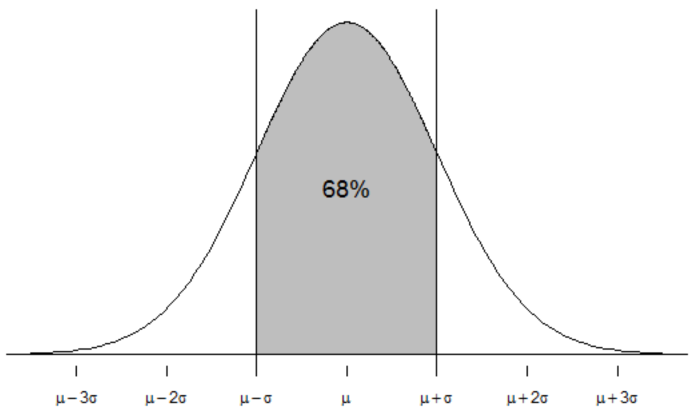
68% of the observations will fall within 1 standard deviation of the mean (that is, $P(\mu - \sigma < X < \mu + \sigma) \approx 0.68$)

approximate

95% of the observations will fall within 2 standard deviations

99.7% of the observations will fall within 3 standard deviations

A picture will help:



area under curve will be 95%

area under curve from $(\mu - 3\sigma, \mu + 3\sigma)$ is approx. 99.7%

Some examples:

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

$$
\begin{aligned}
P(Z < \frac{c-10}{5}) &= P(Z < 1.3) \\
\Rightarrow \frac{c-10}{5} &= 1.3 \\
c &= (1.3 \times 5) + 10 \\
&= 16.5
\end{aligned}
$$

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

$$
\begin{aligned}
P(X < 16) &= 0.05 \\
P(Z < \frac{16-\mu}{2}) &= 0.05 \\
&= P(Z < -1.645) \\
\Rightarrow \frac{16-\mu}{2} &= -1.645 \\
16 - \mu &= -1.645 \times 2 \\
\mu &= 16 + 3.29 \\
&= 19.29
\end{aligned}
$$

⑥ The sums (and differences) of independent Normal variables are also Normal.

Important!

What does it mean?

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

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

Try: what is the distribution of $X - Y$?

$X - Y$ is also Normal.

$$
\begin{aligned}
E(X-Y) &= E(X) - E(Y) \\
&= 10 - 5 = 5 \\
Var(X-Y) &= Var(X) + Var(Y) \qquad [X, Y independent] \\
&= 25 + 16 = 41 \qquad \text{Note} \\
\Rightarrow X - Y &\sim N(5, 41)
\end{aligned}
$$

Example

Given $X \sim N(100, 16)$ and $Y \sim N(90, 20)$. X, Y independent. Find $P(X < Y)$.

Solution: It is important to understand how to solve this simple problem. Hint: bring all the variables to left side.

$$P(X < Y) = P(X - Y < 0)$$

$$X - Y \sim N(10, 16 + 20)$$

[because sum/difference of independent Normal is also Normally distributed.] Must state for exams and assignments

$$\begin{aligned} &\Rightarrow X - Y \sim N(10, 36) \\ P(X < Y) &= P(X - Y < 0) \\ &= P(Z < \frac{0 - 10}{6}) \\ &= P(Z < -1.67) \\ &= 0.0475 \end{aligned}$$

Try it yourself:

What is $P(X + 3Y > 400)$ using X and Y above?

Ans: 0.0162

Another example: $X \sim N(2, 9)$, $\quad Y = 5X + 1$. Find the distribution of Y.

Y is also Normal.

$$E(Y) = E(5X + 1) = 5E(X) + 1 = (5 \times 2) + 1 = 11$$

$$Var(Y) = Var(5X + 1) = 5^2 Var(X) = 5^2 \times 9 = 225$$

$$\Rightarrow Y \sim N(11, 225)$$

⑦ If $X_1, X_2, X_3, \ldots, X_n$ are n independent observations of Normal variable $X \sim N(\mu, \sigma^2)$, then $X_1 + X_2 + X_3 + \cdots + X_n \sim N(n\mu, n\sigma^2)$

Example: Ex. 5.17 in coursenotes (Pg 100).

Let W be the weight of a concrete beam. $W \sim N(31, 0.5^2)$

(a)

$$\begin{aligned} P(30 \leq W \leq 32) &= P(\frac{30-31}{0.5} \leq Z \leq \frac{32-31}{0.5}) \\ &= P(-2 \leq Z \leq 2) \\ &= \Phi(2) - \Phi(-2) \\ &= 0.9772 - 0.0228 \\ &= 0.9544 \end{aligned}$$

(b) It is important for students to think of the sum of 25 randomly selected beams as

$$T = X_1 + X_2 + X_3 + \cdots + X_{25}$$

A student who thinks of T as $25X$ will get the wrong answer!

$$\begin{aligned} T &\sim N(25 \times 31, 25 \times 0.5^2) \\ T &\sim N(775, 6.25) \\ P(T > 795) &= P(Z > \frac{795-775}{\sqrt{6.25}}) \\ &= P(Z > 8) \\ &\approx 0 \end{aligned}$$

⑧ Let $X_1, X_2, X_3, \ldots, X_n$ be random sample of n independent observations of X. If $X \sim N(\mu, \sigma^2)$, then the sample mean $\bar{X}$ is also a Normal distribution and

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

Example:

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:

$$\begin{aligned} X &\sim N(60, 4^2) \\ \bar{X} &\sim N(60, \frac{4^2}{15}) \\ \Rightarrow P(\bar{X} < 58) &= P(Z < \frac{58-60}{\sqrt{4^2/15}}) \\ &= P(Z < -1.936) \\ &= 0.0268 \end{aligned}$$

⑨ Very Important Theorem: Central Limit Theorem

CLT is taught in Chapter 7 but introduced here as it is an appropriate spot:

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n taken from any distribution with mean μ and variance σ^2. Then, for sufficiently large sample n, the sample mean $\bar{X}$ is approximately Normal and

$$\bar{X} \overset{approx.}{\sim} N(\mu, \sigma^2/n)$$

This is known as central limit theorem.

Large sample will usually be of size at least 50 (your notes on ch7 says at least 20 will do).

———————————— End of Chapter 5 Notes ————————————-