ECON 340
Economic Research Methods

Div Bhagia

Lecture 10: Normal Distribution and Z-Score

Random Variables

- *Random variables* take different values under different scenarios.
- Examples: outcome from a coin toss or a die roll, or number of times your wireless network fails before a deadline, etc.
- The likelihood of these scenarios is summarized by the probability distribution.
- Random variables can be *discrete* or *continuous*

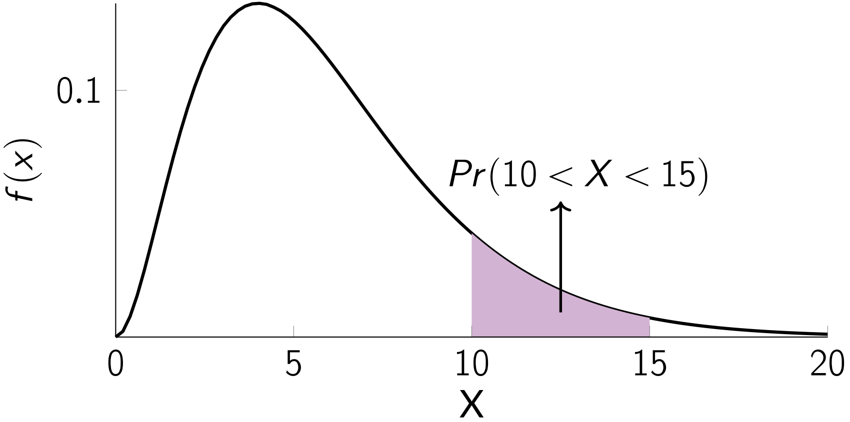
Distribution of a Random Variable

- For a discrete random variable, probability distribution given by the probability of each outcome.

- Continuous random variables summarized by the *probability density function*, where area under the curve gives us the probability of an outcome being in an interval.

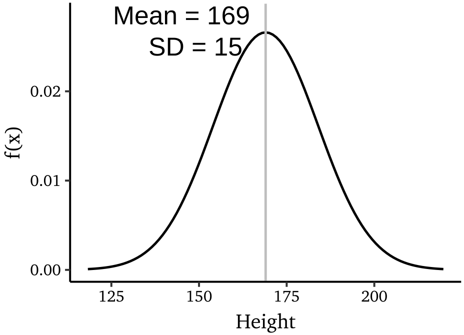
Probability Density Function

The area under the curve tells us the probability of an outcome being in a particular interval.



Normal Distribution

One distribution appears more than others — Normal Distribution



$height \sim N(169, 225)$

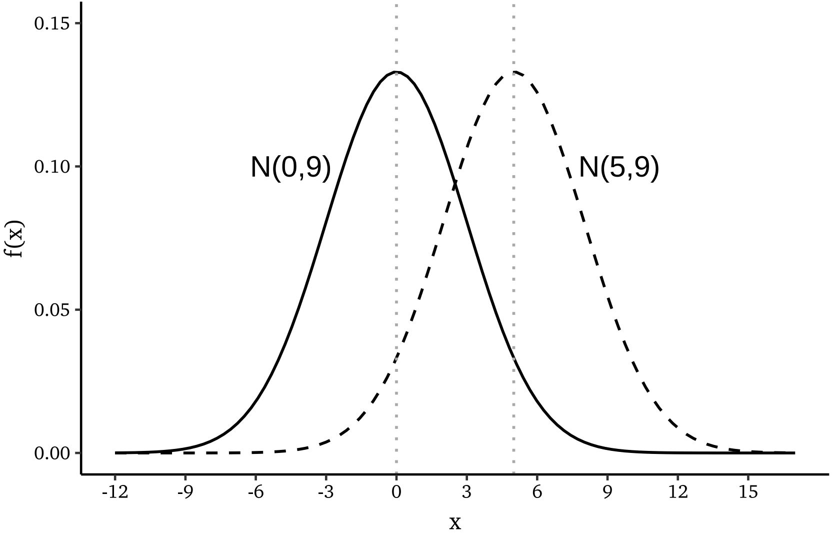
What's special about it?

- Symmetric (no skew, mean=median, bell-shaped)
- Height, birthweight, SAT scores, etc., normally distributed
- Sampling distribution approximately normal

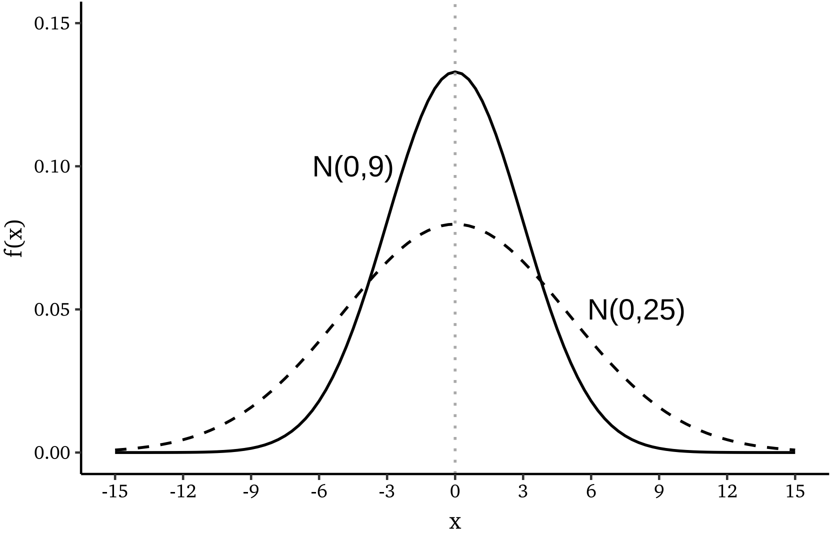
Normal Distribution

- Normal distribution with mean μ and variance σ^2 is expressed as $N(\mu, \sigma^2)$
- So if I write $X \sim N(12, 4)$, it means X is normally distributed with mean 12 and variance 4
- The standard normal distribution is the normal distribution with mean 0 and variance 1, denoted by $N(0, 1)$
- Random variables that have a $N(0, 1)$ distribution are often denoted by Z

Normal Distribution



Normal Distribution



How to find the area under the curve?

- Often interested in finding the probability that a random variable lies in a particular interval
- Cumbersome to take the integral each time
- Since the normal distribution is so commonly used, one can find these probabilities easily for the *standard normal variable*:
$$Z \sim N(0,1)$$
- We can use the standard normal probabilities to get the probabilities for any normally distributed variable

Standardized Random Variables

A random variable can be transformed into a random variable with mean 0 and variance 1 by subtracting its mean and then dividing by its standard deviation, a process called standardization.

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Transformations of Random Variables

- Say, X is a random variable with mean μ_X, variance σ_X^2, and standard deviation σ_X
- If we transform X to create a new random variable

$$Y = a + bX$$

- Then Y is a random variable that has a distribution with the same shape as X and with
  - Mean: $\mu_Y = a + b \cdot \mu_X$
  - Variance: $\sigma_Y^2 = b^2 \cdot \sigma_X^2$
  - SD: $\sigma_Y = b \cdot \sigma_X$

Z-Score

- We can rearrange the terms in the Z-score

$$Z = \frac{X - \mu_X}{\sigma_X} \quad \rightarrow \quad Z = \frac{-\mu_X}{\sigma_X} + \frac{1}{\sigma_X} \cdot X$$

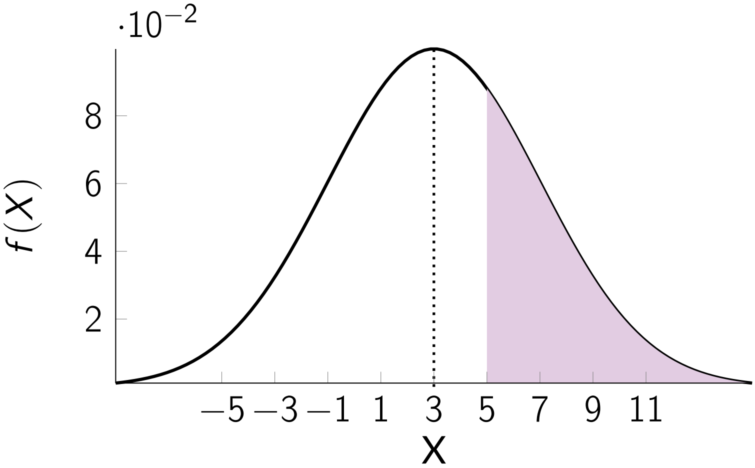
- So here $Z = a + b \cdot X$ with $a = \frac{-\mu_X}{\sigma_X}$ and $b = \frac{1}{\sigma_X}$
- Since Z is a transformed random variable:

$$\mu_Z = a + b \cdot \mu_X = \frac{-\mu_X}{\sigma_X} + \frac{1}{\sigma_X} \cdot \mu_X = 0$$

$$\sigma_Z^2 = b^2 \cdot \sigma_X^2 = \left(\frac{1}{\sigma_X}\right)^2 \cdot \sigma_X^2 = 1$$

How to find the area under the curve?

For example, say $X \sim N(3, 16)$. We want to calculate $Pr(X \geqslant 5)$.



How to find the area under the curve?

Given $X \sim N(3, 16)$,

$$Z = \frac{X - \mu_X}{\sigma_X} = \frac{X - 3}{4} \sim N(0, 1)$$

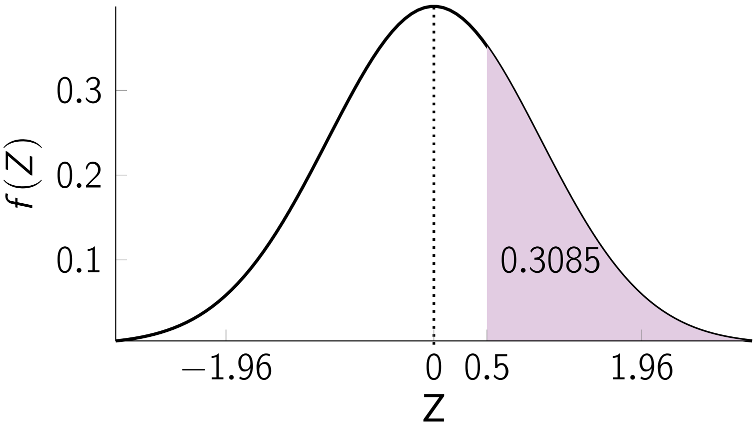
Note that,

$$Pr(X \geqslant 5) = Pr\left(\frac{X - 3}{4} \geqslant \frac{5 - 3}{4}\right) = Pr(Z \geqslant 0.5)$$

We can now refer to the standard normal table and find that $Pr(Z \geqslant 0.5)$

How to find the area under the curve?

Find $Pr(Z \geqslant 0.5)$ from the standard normal table.



Recipe

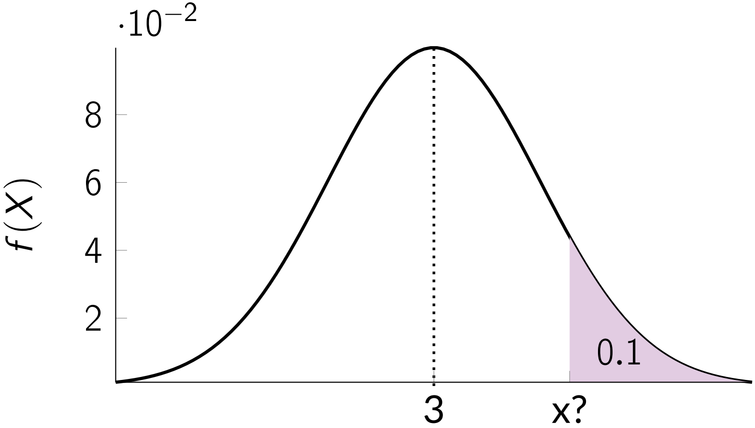
Given $X \sim N(\mu, \sigma^2)$, general recipe to find $Pr(x_0 < X < x_1)$:

- Find $z_0 = (x_0 - \mu)/\sigma$ and $z_1 = (x_1 - \mu)/\sigma$
- Use standard normal table to find $Pr(z_0 < Z < z_1)$

Example. Given $X \sim N(3, 16)$, find $Pr(2 < X < 5)$.

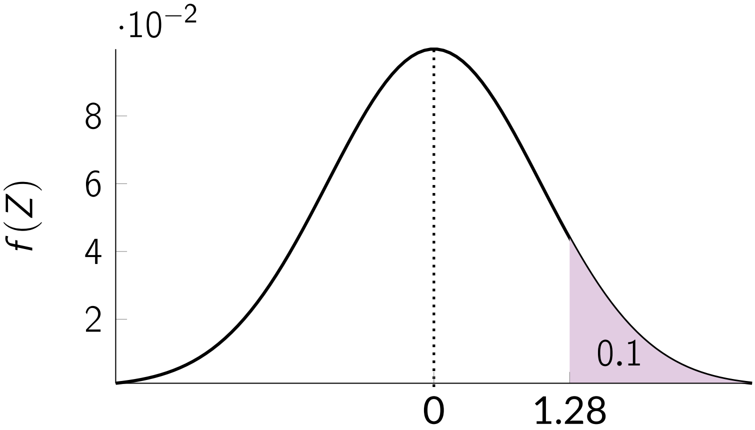
Finding points from area under the curve

For example, say $X \sim N(3, 16)$ and we are given $Pr(X > x) = 0.10$. How to find x?



Finding points from area under the curve

Start by finding z, such that $Pr(Z > z) = 0.1$.



Finding points from area under the curve

Now note that,

$$Z = \frac{X - \mu}{\sigma} \rightarrow X = \mu + Z.\sigma$$

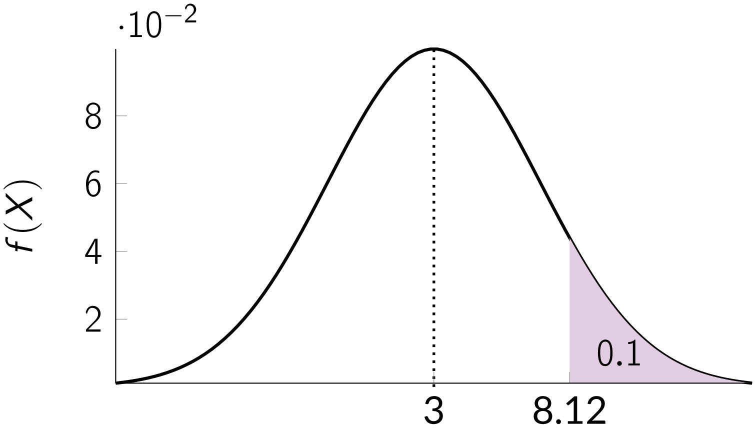
We found $Pr(Z > 1.28) = 0.1$, we can find the corresponding x for 1.28 as follows:

$$3 + 1.28 \times 4 = 8.12$$

So we have that $Pr(X > 8.12) = 0.1$.

Finding points from area under the curve

Transforming z back to $x = 3 + 1.28 \times 4 = 8.12$.



Finding points from area under the curve

- Sometimes we are given $Pr(X < x)$ or $Pr(X > x)$ and we need to find x.
- *Example*: Given $X \sim N(3, 16)$ and $Pr(X > x) = 0.10$, find x.
- Start by finding z, such that $Pr(Z > z) = 0.1$. From the standard normal table, $z = 1.28$.
- Now we just need to convert z to x.
- Since $Z = \frac{X-\mu}{\sigma} \rightarrow X = \mu + Z.\sigma$, so $x = 3 + 1.28 \times 4 = 8.12$