ECO220Y

Continuous Probability Distributions: Normal

Readings: Chapter 9, section 9.10

Fall 2011

Lecture 8 Part 2

Normal Density Function

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

where $-\infty < x < \infty$

μ and σ are parameters of this distribution

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

Example: mean 0 and s.d. 1

$$f(x) = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-0}{1})^2}$$

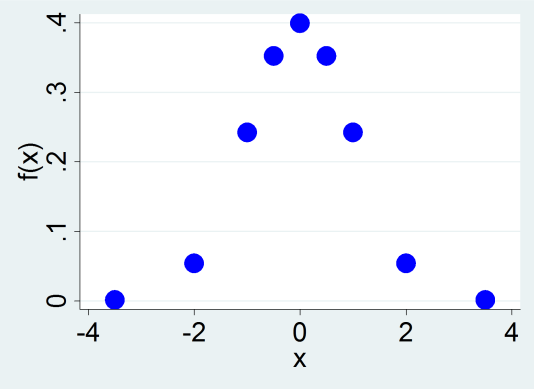
$$f(x) = \frac{1}{1\sqrt{2\pi}} e^{-0.5x^2}$$

$$f(x) = 0.40 \frac{1}{e^{0.5x^2}}$$

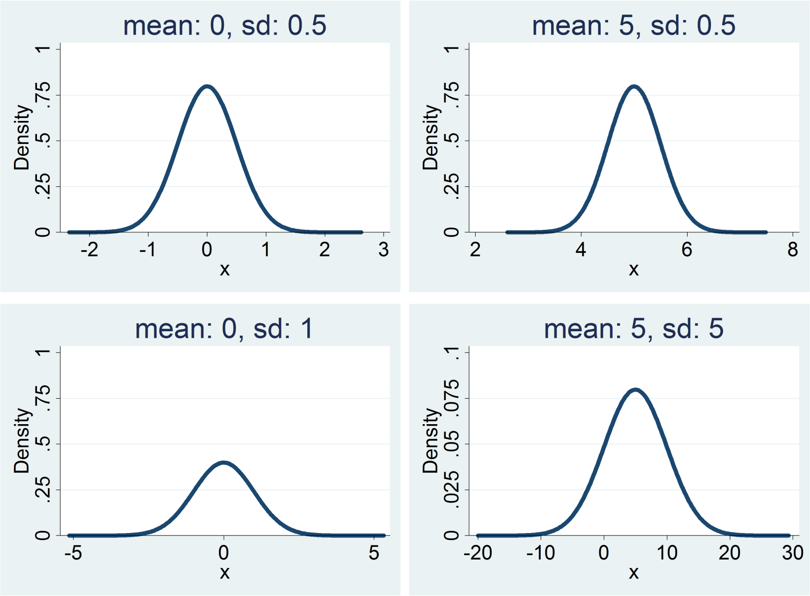
$$f(x) = \frac{0.40}{1.65^{x^2}}$$

x	f(x)
-3.5	0.001
-2	0.054
-1	0.242
-0.5	0.352
0	0.399
0.5	0.352
1	0.242
2	0.054
3.5	0.001

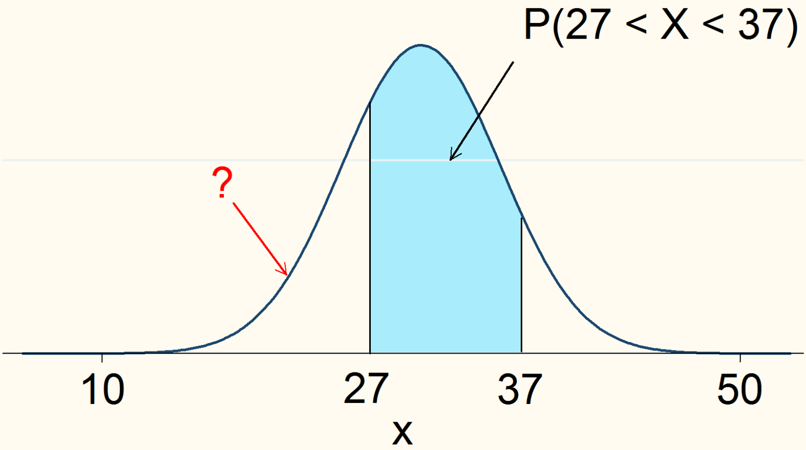
Example: mean 0 and s.d. 1



x	f(x)
-3.5	0.001
-2	0.054
-1	0.242
-0.5	0.352
0	0.399
0.5	0.352
1	0.242
2	0.054
3.5	0.001


mean: 0, sd: 0.5
Density
0
.25
.5
.75
1
-2
-1
0
1
2
3
x
mean: 5, sd: 0.5
Density
0
.25
.5
.75
1
2
4
6
8
x
mean: 0, sd: 1
Density
0
.25
.5
.75
1
-5
0
5
x
mean: 5, sd: 5
Density
0
.025
.05
.075
.1
-20
-10
0
10
20
30
x

Area is Probability



Finding Normal Probabilities

- Probability is area under the bell curve
- Finding the area is tricky even with calculus
- Can use tables and software to find normal probabilities
- Algorithm for finding normal probabilities:
    1. Standardize
    2. Read the table

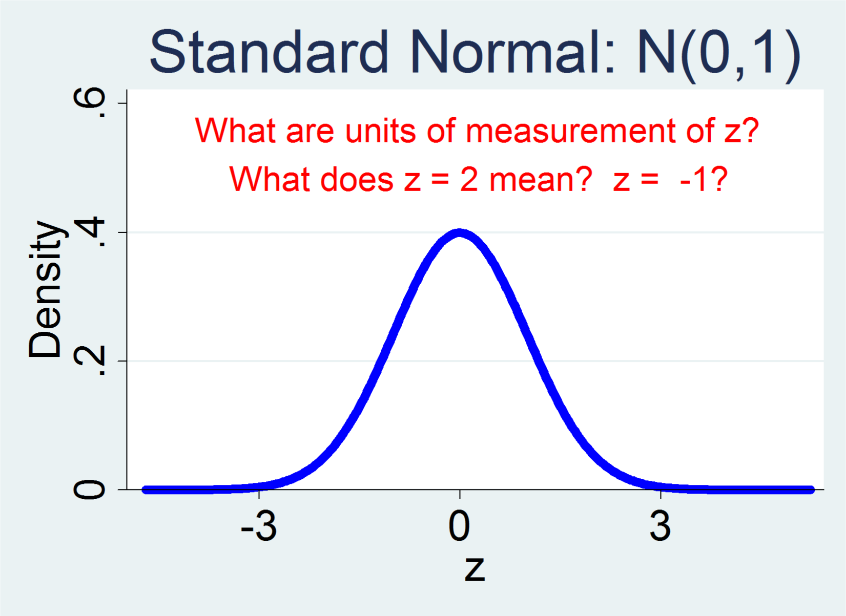
Standard Normal: Z ∼ N(0,1)

If X ∼ N(μ, σ), then to get standard normal random variable Z with mean 0 and s.d. 1:

$$\boxed{z = \frac{X - \mu}{\sigma}}$$

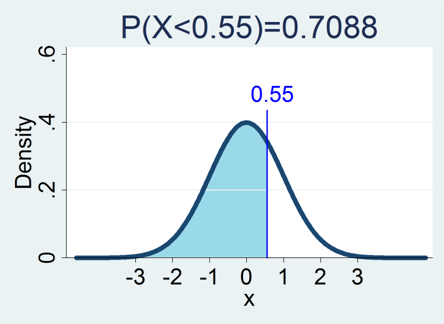
Recall: standardization is a linear transformation:

$$z = \frac{X - \mu}{\sigma} \longrightarrow z = \underbrace{\frac{-\mu}{\sigma}}_{a} + \underbrace{\frac{1}{\sigma}}_{b} x$$


Standard Normal: N(0,1)
What are units of measurement of z?
What does z = 2 mean? z = -1?
Density
0
.2
.4
.6
-3
0
3
z

Finding Probabilities with Standard Normal Table

Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088



Finding Probabilities

Need to find:

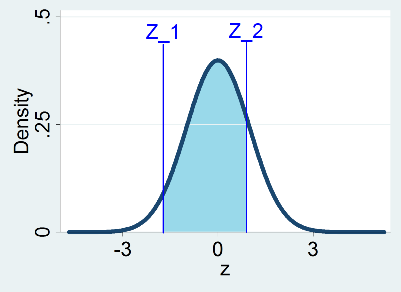
$$P(x_1 < X < x_2)$$

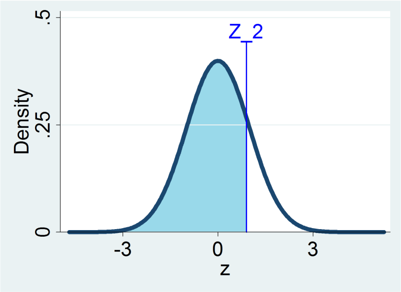
We can standardize X and find $P(z_1 < Z < z_2)$ from table:

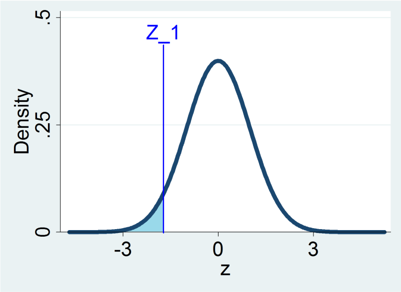
$$P\left(\frac{x_1-\mu}{\sigma} < \underbrace{\frac{X-\mu}{\sigma}}_{Z} < \frac{x_2-\mu}{\sigma}\right)$$

Answer is exactly the same. Why?

Finding Range Probabilities







Finding Probabilities

Here is the intuition:

X - the percentage return on investment - is distributed normally with a mean of 10 percent and a s.d. of 10 percent, or $X \sim N(10, 10^2)$.

What is the probability that return will be greater than 20 percent, or $X > 20$?

$$\boxed{X > 20} \implies \boxed{X - \mu > 20 - \mu} \implies \boxed{\tfrac{X-\mu}{\sigma} > \tfrac{20-\mu}{\sigma}}$$

Therefore,

$$P(X > 20) = P\left[\underbrace{\frac{X - \mu}{\sigma}}_{Z} > \tfrac{20-\mu}{\sigma}\right]$$

Example

The amount of time devoted to studying statistics each week by students is normally distributed with a mean of 7.5 hours and a s.d. of 2.1 hours.

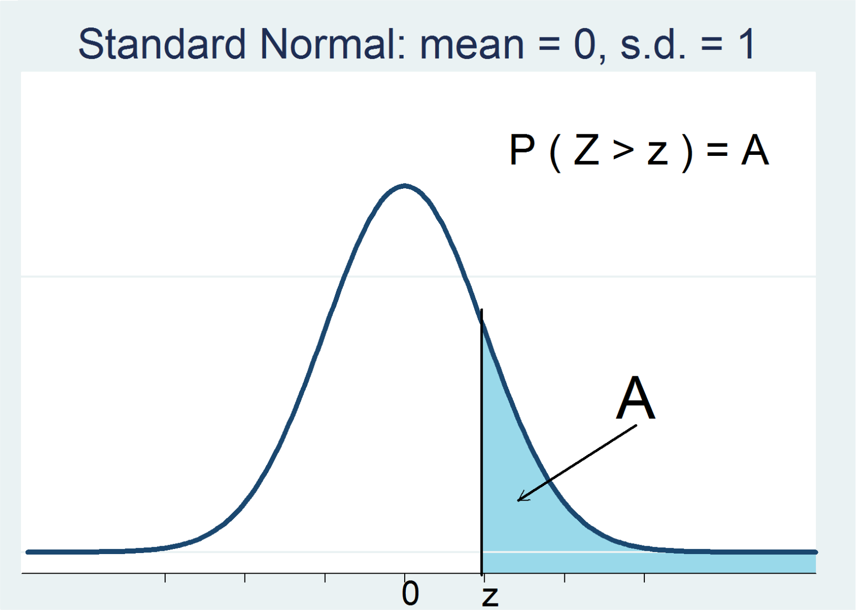
• What proportion of students study for more than 10 hours per week: $P(X > 10) = ?$

$$P(X > 10) = P(Z > \frac{10-7.5}{2.1}) = P(Z > 1.19) = 0.117$$

• Probability that a randomly selected student spends between 7 and 9 hours studying: $P(7 < X < 9) = ?$

$$P(7 < X < 9) = P(\frac{7-7.5}{2.1} < Z < \frac{9-7.5}{2.1}) = P(-0.24 < Z < 0.71)$$
$$= P(Z < 0.71) - P(Z < -0.24) = 0.3559$$

Find z_A such that P $(Z > z_A) = A$



Example Cont'd

• The amount of time spent studying X$\sim$ N(7.5, 2.1^2)

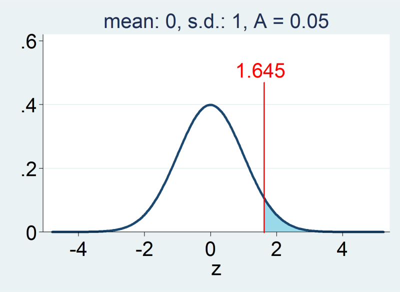
• If a student is in the top 5%, what amount of time does he/she spend studying? $P(X > x_A) = 0.05$. Need to find $x_A = ?$

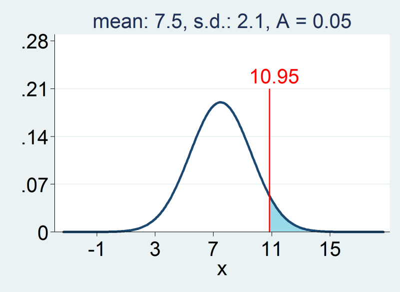
•$P(X > x_A) = 0.05 \longrightarrow P(Z > \frac{x_A - 7.5}{2.1}) = 0.05 \longrightarrow P(Z > z_A) = 0.05$

• From table $z_{0.05} = 1.645$

• Un-standardize z_A to get back x_A

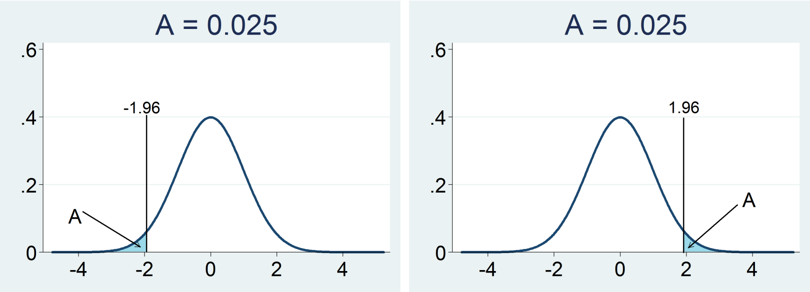
•$\frac{x_A - 7.5}{2.1} = 1.645 \longrightarrow x_A = 1.645 * 2.1 + 7.5 = 10.95$ hours





$$Z_A = \frac{X_A - \mu}{\sigma} \longrightarrow X_A = Z_A * \sigma + \mu$$

Symmetry of Normal Distribution



$$P(Z < -z_A) = P(Z > z_A)$$

Summary: Normal Distribution

- Normal distribution is symmetric and bell-shaped
- Values are clustered around the mean. Recall Empirical Rule:
  - about 68.3% within 1 s.d. of mean
  - about 95.4% within 2 s.d. of mean
  - about 99.7% within 3 s.d. of mean
- Parameters: μ and σ
- Unbounded support: $(-\infty, \infty)$

Normal Approximation to Binomial Distribution

$X \sim B(100, 0.2)$

$P(X = 20) =$

$C_{20}^{100}(0.2)^{20}(0.8)^{80} =$

$\frac{100!}{20!80!}(0.2)^{20}(0.8)^{80} =$

$= 0.0993$

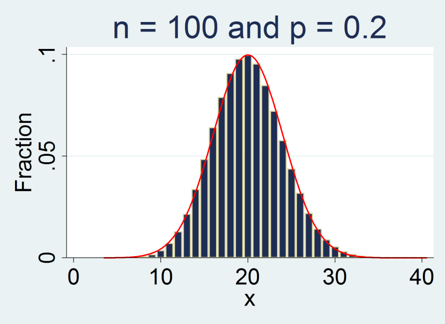
$X \sim N(20, 16)$

$P(X = 20) =$

$P(19.5 < X < 20.5) =$

$P(-0.125 < Z < 0.125)$

$= 0.0995$



Rule of Thumb for Binomial Distribution

To determine if Normal distribution is a good approximation for the Binomial:

- Check if the entire interval lies between 0 and n, where interval is given by:

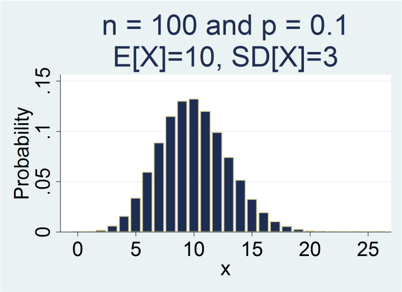
$$np \pm 3\sqrt{np(1-p)}$$

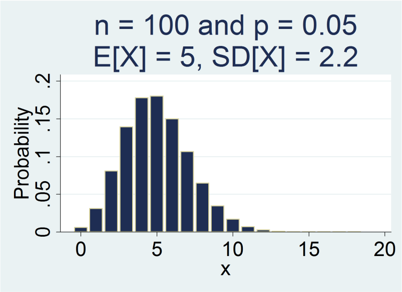
What is the concept behind the rule of thumb for Binomial distribution?

Alternative Rule of Thumb

- At least two alternative rules of thumb for approximation
  1. $np > 5$ and $n(1-p) > 5$
  2. $np > 10$ and $n(1-p) > 10$
  3. Obviously, if $\boxed{2}$ holds, then $\boxed{1}$ holds
- Use the one you find more intuitive and convenient

Use Normal Approximation?





$10 \pm 3 * 3 \rightarrow (1, 19)$

Within (0,100)

$5 \pm 3 * 2.2 \rightarrow (-1.6, 11.6)$

Not within (0,100)