

NFM2106/NFE2105

Mathematics

Normal distributions and
related concepts

Discrete RVs (previous lecture)

(random) **experiment** = a process by which we observe something **uncertain**

trial = a particular performance of a random experiment

sample space = the set of all possible outcomes of an experiment

RV = a function from the sample space into the set of real numbers

\downarrow
 $X:$

x_1	x_2	x_3	\dots	x_{n-1}	x_n
p_1	p_2	p_3	\dots	p_{n-1}	p_n

$$\sum_{i=1}^n p_i = 1$$

$$p_i \geq 0$$

Sum of discrete RVs

$X:$

x	1	2
p	1/2	1/2

$Y:$

x	1	2	3
p	1/3	1/3	1/3

$X + Y:$

x	2	3	4	5
p	1/6	1/3	1/3	1/6

Product of discrete RVs

$X:$

x	1	2
p	1/2	1/2

$Y:$

x	1	2	3
p	1/3	1/3	1/3

$XY:$

x	1	2	3	4	6
p	1/6	1/3	1/6	1/6	1/6

Expectation value

Let X be a DISCRETE random variable, i.e.

$$X = \{x_1, x_2, x_3, \dots, x_n\}, \quad p_i = P(X = x_i)$$

The **expectation value** (or simply the **mean**) of this R.V. is

$$\mu_X = E(X) = \sum_{i=1}^n p_i x_i$$

For a real function $f: \mathbb{R} \rightarrow \mathbb{R}$ the expectation value of $Y = f(X)$ is given by

$$E(Y) = \sum_{i=1}^n p_i f(x_i)$$

Variance & standard deviation

The **variance** gives the degree of spread of the distribution about the mean

$$\text{Var}(X) = E[(X - \mu)^2]$$

$$\mu = E(X)$$

This can be written as

$$\text{Var}(X) = \sum_{i=1}^n p_i (x_i - \mu)^2$$

Standard deviation:

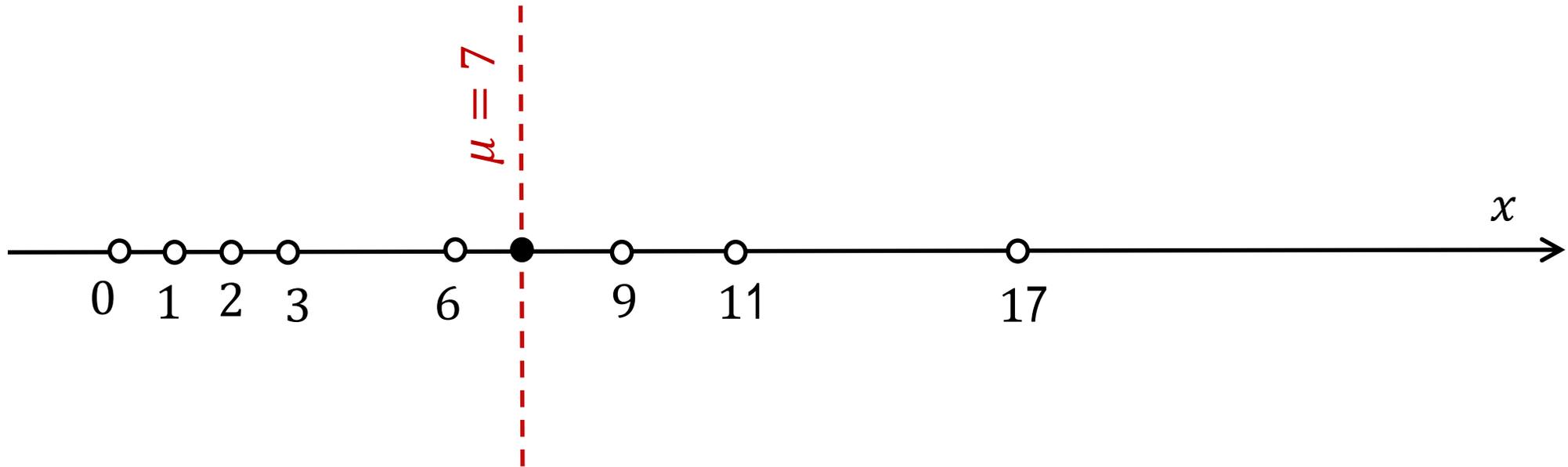
$$\sigma_X = \sqrt{\text{Var}(X)}$$

$$\Rightarrow \sigma_X^2 = \text{Var}(X)$$

Observation

$X:$

x	1	2	3	6	9	11	17
p	1/7	1/7	1/7	1/7	1/7	1/7	1/7



The relationship between σ_X and μ_X

$$\begin{aligned}\underline{\text{Var}(X)} &= \sum_{i=1}^n p_i (x_i^2 - 2x_i\mu + \mu^2) \\ &= \sum_{i=1}^n p_i x_i^2 - 2\mu \sum_{i=1}^n p_i x_i + \mu^2 \sum_{i=1}^n p_i \\ &= E(X^2) - 2\mu^2 + \mu^2 = \underline{E(X^2) - \mu^2} \implies \sigma^2 = E(X^2) - \mu^2\end{aligned}$$

This can be written in a symmetric form that is easy to remember:

$$\sigma^2 = E(X^2) - (E(X))^2$$

OBS.: The standard deviation is σ , **not** σ^2 !!

Example: rolling a single die

Here the possible outcomes or events are rolling one of the numbers $\{1, 2, 3, 4, 5, 6\}$, which all occur with probability $1/6$.

$$\mu = \sum_{i=1}^6 p_i x_i = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} = 3.5$$

$$\begin{aligned} \sigma^2 &= E(X^2) - \mu^2 = \sum_{i=1}^6 p_i x_i^2 - \mu^2 = \frac{1}{6} (1^2 + 2^2 + \dots + 6^2) - \mu^2 \\ &= \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \quad \Rightarrow \quad \sigma = \left(\frac{35}{12}\right)^{1/2} = 1.7078 \dots \end{aligned}$$

Expectation of a sum of RVs

$X, Y =$ discrete random variables

$$E(X + Y) = E(X) + E(Y)$$

The expectation value of a sum of two random variables equals the sum of the expectation values of the two variables

General case: $X_1, X_2, \dots, X_n =$ finite number of discrete RVs

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

Example

$X:$

x	1	2
p	1/2	1/2

$$E(X) = 1.5$$

$Y:$

x	1	2	3
p	1/3	1/3	1/3

$$E(Y) = 2$$

$X + Y:$

x	2	3	4	5
p	1/6	1/3	1/3	1/6

$$E(X + Y) = 3.5$$

Interpretation of the expectation value

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

A special case of this arises when we perform n trials of **the same experiment**. In this case the n R.V.'s X_i are all associated with the same probability distribution; that is, the X_i are **identically distributed** random variables.

For example, the X_i might all refer to the rolling of a die.

$$E(X_1 + X_2 + \cdots + X_n) = nE(X)$$

E.G.

$n = 1000$: the sum of all numbers obtained
in 1000 trials will be approximately $1000 \times 3.5 = 3500$
ing

Continuous Random Variables

A random variable X is **continuous** if its set of possible values is an entire interval of numbers

(e.g., if $A < B$, then any number x between A and B is possible).

E.G.: X $x \in [1,3]$ or $x \in (0, \infty)$

$X =$ “time to failure of an electric light bulb”

or

$Y =$ “outside temperature at noon tomorrow”

Discrete vs. Continuous

Consider the following situation:

Let $x_0 \in [20,30]$, a fixed number.

I pick a real number between 20 and 30, assuming that all real values in the interval $[20, 30]$ are equally likely. Let X be the R.V. corresponding to the number picked.

Sample space: $\Omega = [20,30]$ # outcomes
in $\Omega = \infty$

Let's consider the **event:** $\{X = x_0\}$. # outcomes for
this event = 1

$$\text{Hence } P(X = x_0) = \frac{1}{\infty} = 0 !!$$

Probability distribution

$X =$ continuous RV

A **probability distribution** or **probability density function** (PDF) of X is a function $f(x) \geq 0$ such that for any two numbers a and b :

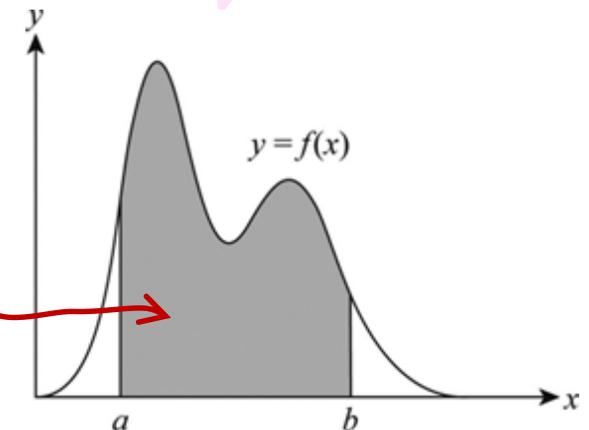
$$P(a < X < b) = \int_a^b f(x) dx$$

$f_X(x)$
same as
 $f(x)$

The event $\{-\infty < X < \infty\}$ is absolutely certain to occur,

$$\Rightarrow P(-\infty < X < \infty) = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} f(x) dx = 1$$



Probability distribution

If X only exists between α and β ,
then the above (normalization) condition
becomes:

$$\int_{\alpha}^{\beta} f(x) dx = 1$$

EXAMPLE:

1). Let X be a continuous random variable with PDF given by

$$f_X(x) = \begin{cases} cx^2 & |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the constant c .

Note that X is zero outside the interval $[-1, 1]$. So the PDF must satisfy
whence $c = 3/2$.

$$\int_{-1}^{+1} cx^2 dx = 1,$$

Probability distribution

EXAMPLE

2). Let X be a continuous random variable with PDF given by

$$f_X(x) = \begin{cases} x^2(2x + b), & 0 < x \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

Find the constant b .

In this case the normalization condition is

$$\int_0^1 x^2(2x + b)dx = 1 \quad \implies \quad b = 3/2$$

Observation

The probabilities corresponding to the events:

$$\{a < X < b\}$$

$$\{a \leq X < b\}$$

$$\{a < X \leq b\}$$

$$\{a \leq X \leq b\}$$

are all **the same**.

That is because in the continuous case any event $\{X = a\}$, etc. has **zero** probability.

The uniform distribution

Let $a < b$ be real numbers.

The **uniform distribution** on $[a, b]$ is a distribution that has constant probability, i.e. all numbers within this interval are “equally likely”.

It is defined by the following particular **PDF**:

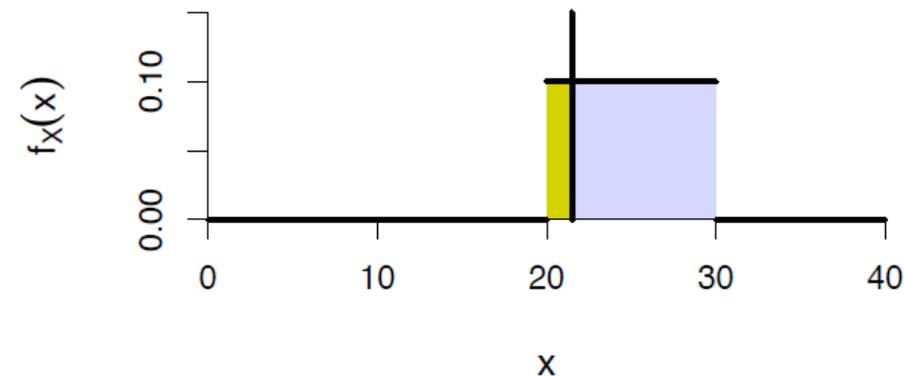
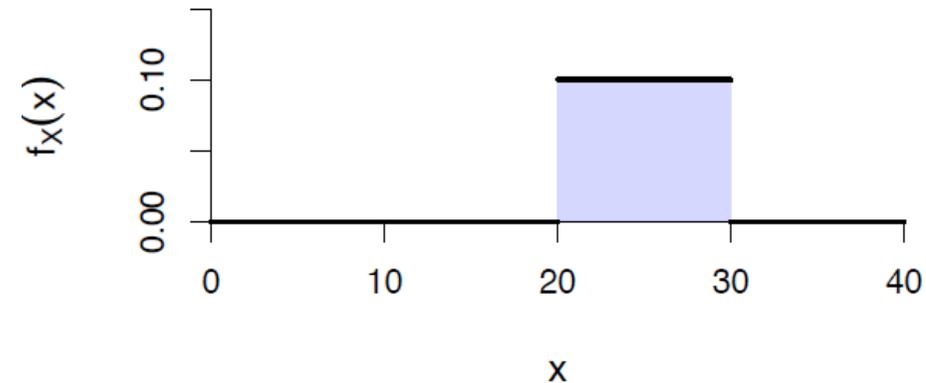
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b; \\ 0, & \text{otherwise} \end{cases}$$

Note that this satisfies the normalization condition trivially, $\int_a^b f_X(u) du = 1$

Back to the earlier eg: let's calculate $P(X \leq 21.5)$

Consider the uniform distribution on $[20,30]$;
the density probability is

$$f_X(x) = \begin{cases} \frac{1}{10}, & \text{if } 20 \leq x \leq 30; \\ 0, & \text{otherwise} \end{cases}$$



$$\begin{aligned} P(X \leq 21.5) &= \int_{20}^{21.5} f_X(u) du = \left[\frac{u}{10} \right]_{20}^{21.5} \\ &= \frac{21.5 - 20}{10} = 0.15 = 15\% \end{aligned}$$

Another example

A man arrives at a bus stop at a random time (that is, with no regard for the scheduled service) to catch the next bus.

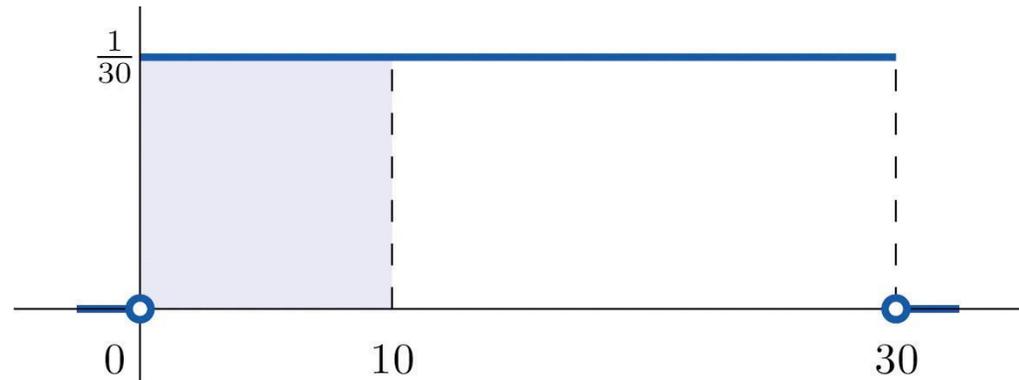
Buses run every 30 minutes without fail, hence the next bus will come any time during the next 30 minutes with evenly distributed probability (a uniform distribution).

Find the probability that a bus will come within the next 10 minutes.

X = waiting time for the bus

The probability sought is $P(0 \leq X \leq 10)$

Final answer $\approx 33.3\%$



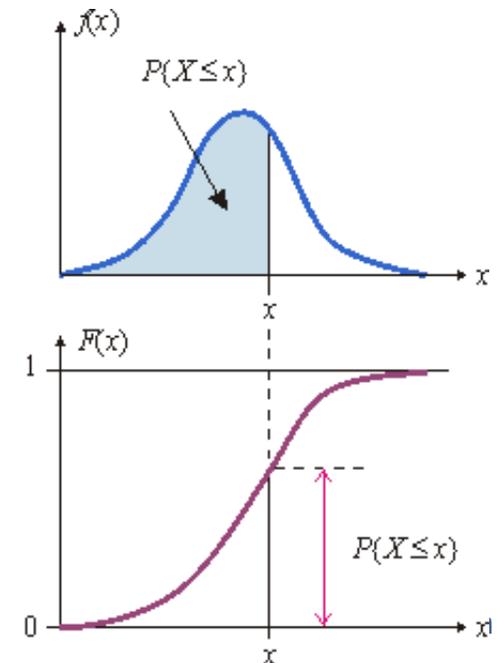
Cumulative Distribution Function (CDF)

$$F(x) = P(X \leq x)$$

$$F(x) = \int_{-\infty}^x f_X(u) du$$

Properties:

- $0 \leq F(x) \leq 1 \quad -\infty < x < \infty$
- $F(-\infty) = 0 \quad F(\infty) = 1$
- $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
- $P(X > x_1) = 1 - F(x_1)$
- $\frac{dF}{dx} = f_X(x)$



Mean, Variance, SDV for continuous RVs

X = continuous R.V.

$\phi: \mathbb{R} \rightarrow \mathbb{R}$ continuous
function

The **mean**: $\mu_X = E(X) = \int_{\alpha}^{\beta} x f_X(x) dx$

$Y = \phi(X)$ is also a continuous R.V. $\implies \mu_Y = E(Y) = \int_{\alpha}^{\beta} \phi(x) f_X(x) dx$

The **variance**: $V(X) \equiv \sigma^2 = \int_{\alpha}^{\beta} (x - \mu_X)^2 f_X(x) dx$ ($\sigma = SDV$)

Easy to show that

$$\sigma^2 = \int_{\alpha}^{\beta} x^2 f_X(x) dx - \mu_X^2$$

Application to uniform distributions

MEAN: $\mu_X = \frac{a + b}{2}$

VARIANCE: $Var(X) \equiv \sigma^2 = \frac{(b - a)^2}{12}$

STANDARD DEVIATION: $\sigma = \frac{b - a}{2\sqrt{3}}$ (note, $\sigma < \mu_X$)

(The above results follow by a trivial direct calculation from the general formulae on the previous slide. **You are invited to check this in your own time**)

The normal probability distribution

NORMAL PROBABILITY DENSITY FUNCTION

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

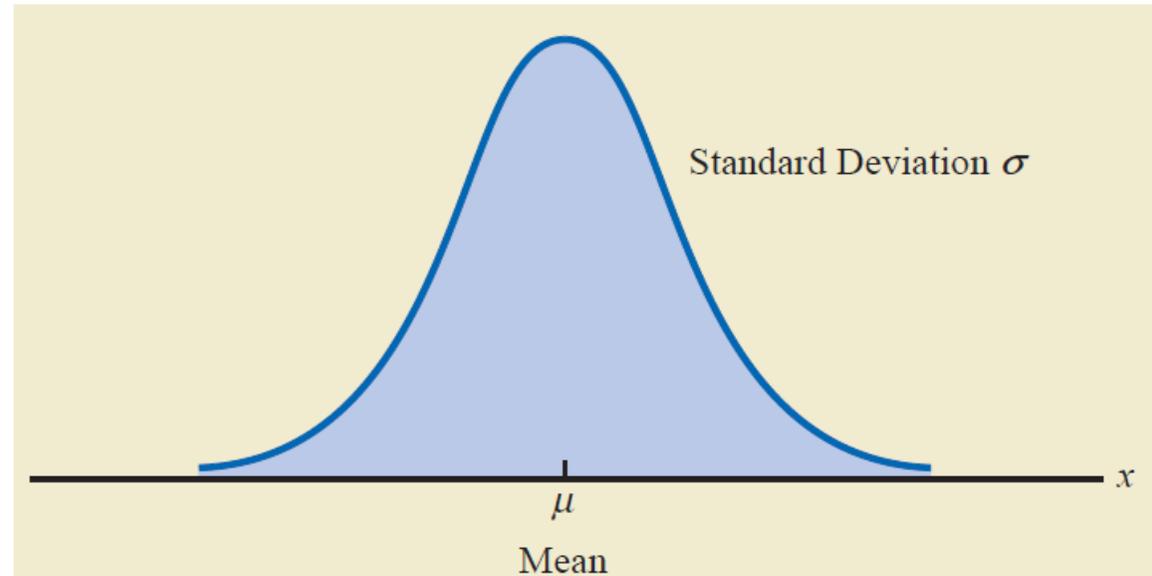
where

μ = mean

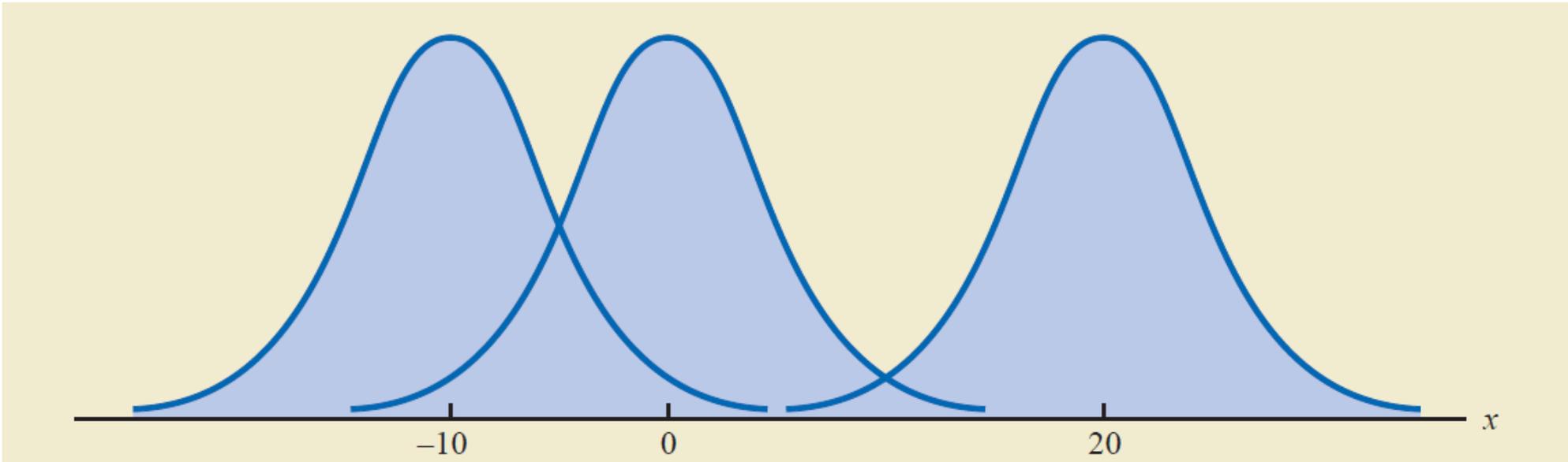
σ = standard deviation

π = 3.14159

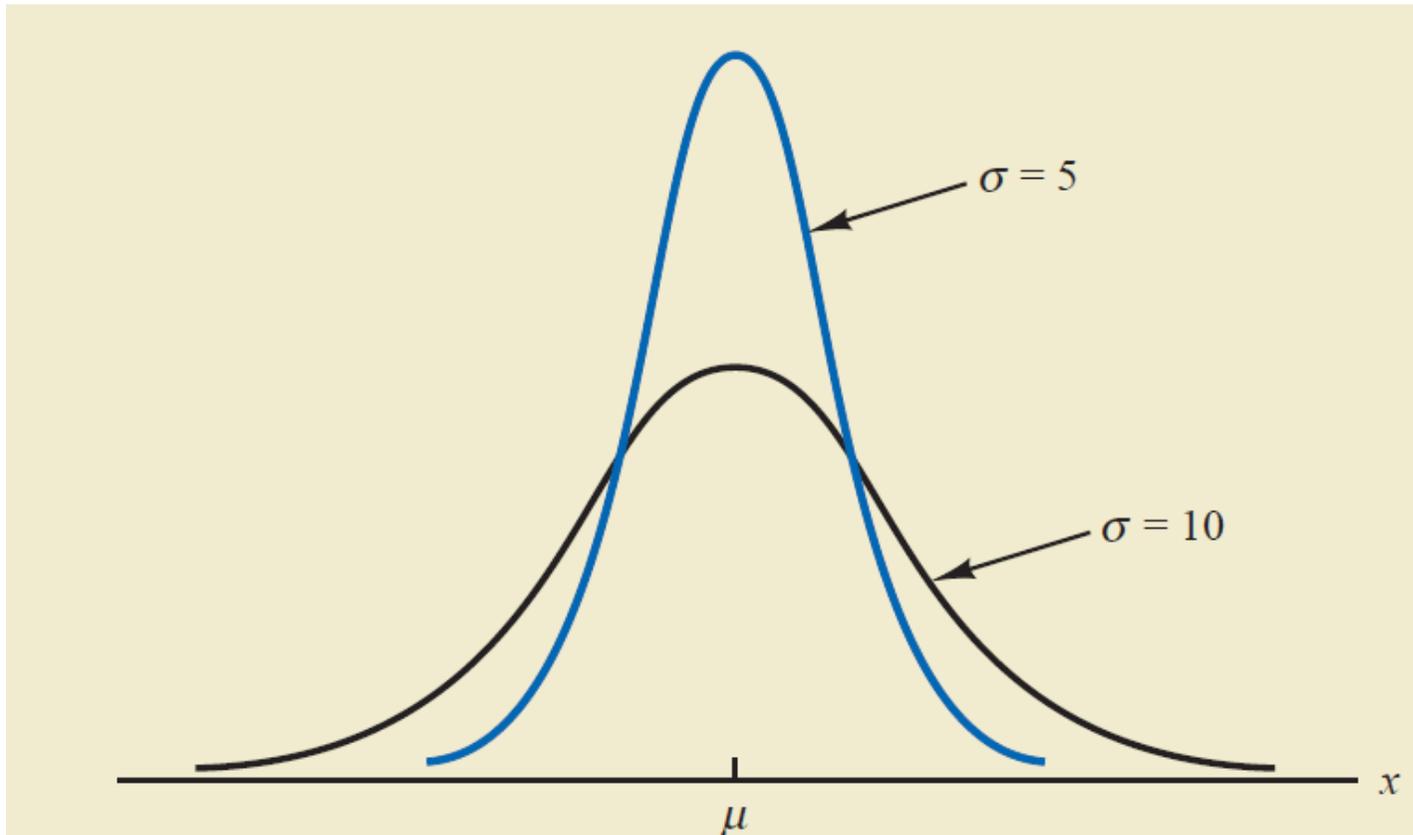
e = 2.71828



Effect of changing the mean:



Effect of changing the standard deviation:



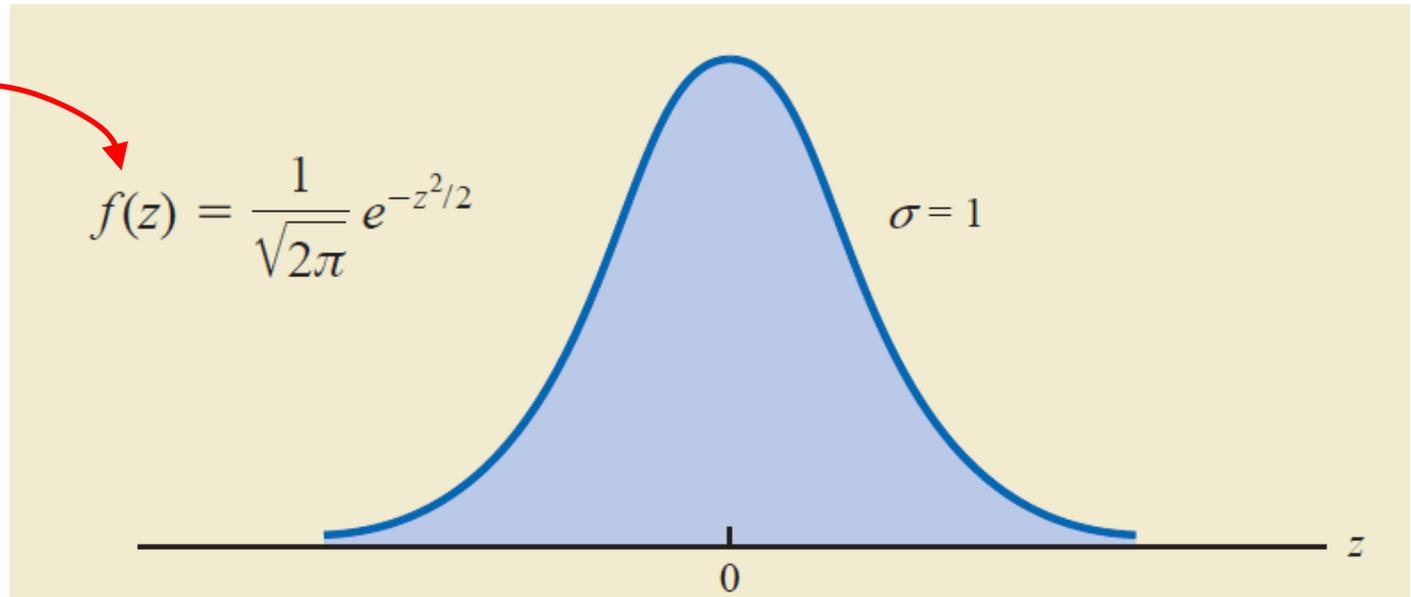
Standard normal probability distribution (SNPD) University of HUDDERSFIELD

A random variable that has a normal distribution with a **mean** of 0 and a **standard deviation** of 1 is said to have a **standard normal probability distribution**. The letter z is commonly used to designate this particular normal random variable.

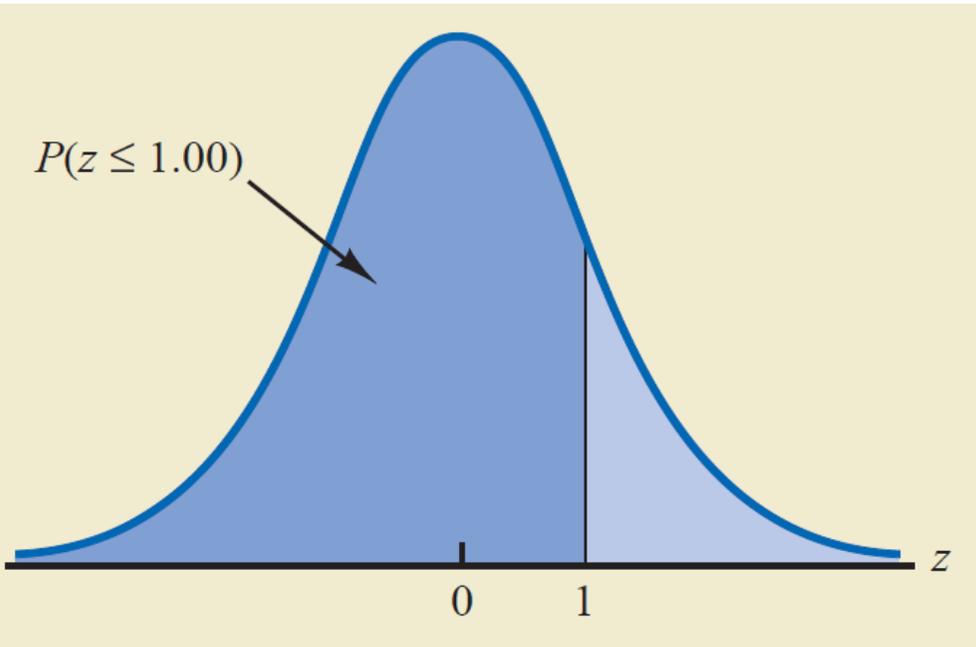
standard normal
density function

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

$\sigma = 1$



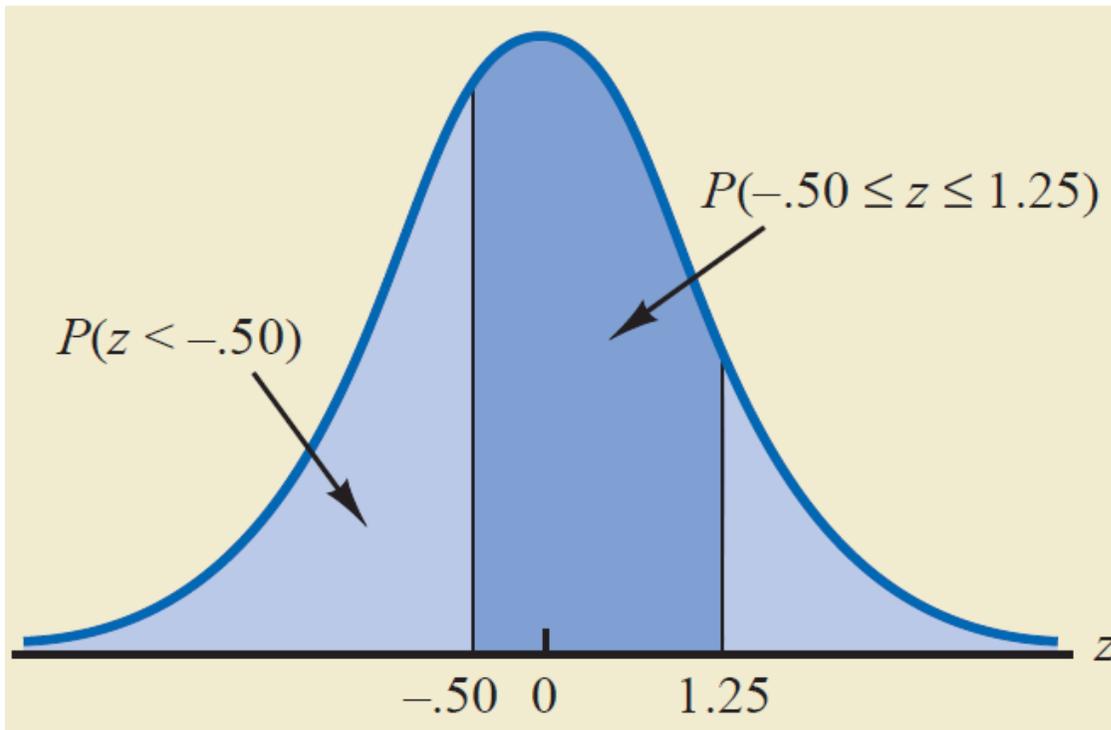
How to use the SNPD:



z	.00	.01	.02
.			
.			
.			
.9	.8159	.8186	.8212
1.0	.8413	.8438	.8461
1.1	.8643	.8665	.8686
1.2	.8849	.8869	.8888
.			
.			
.			

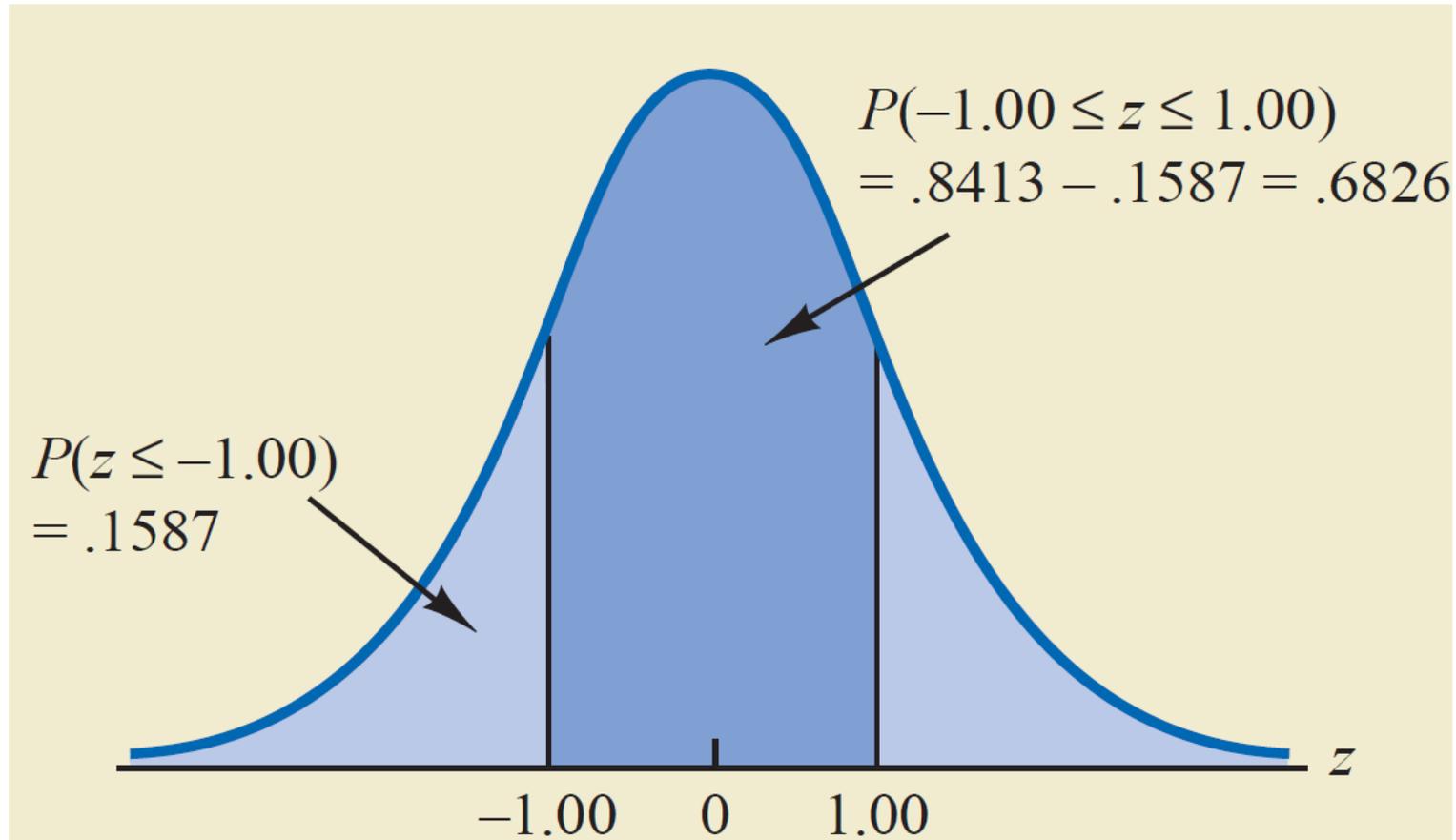
$P(z \leq 1.00)$

How to use the SNPD:

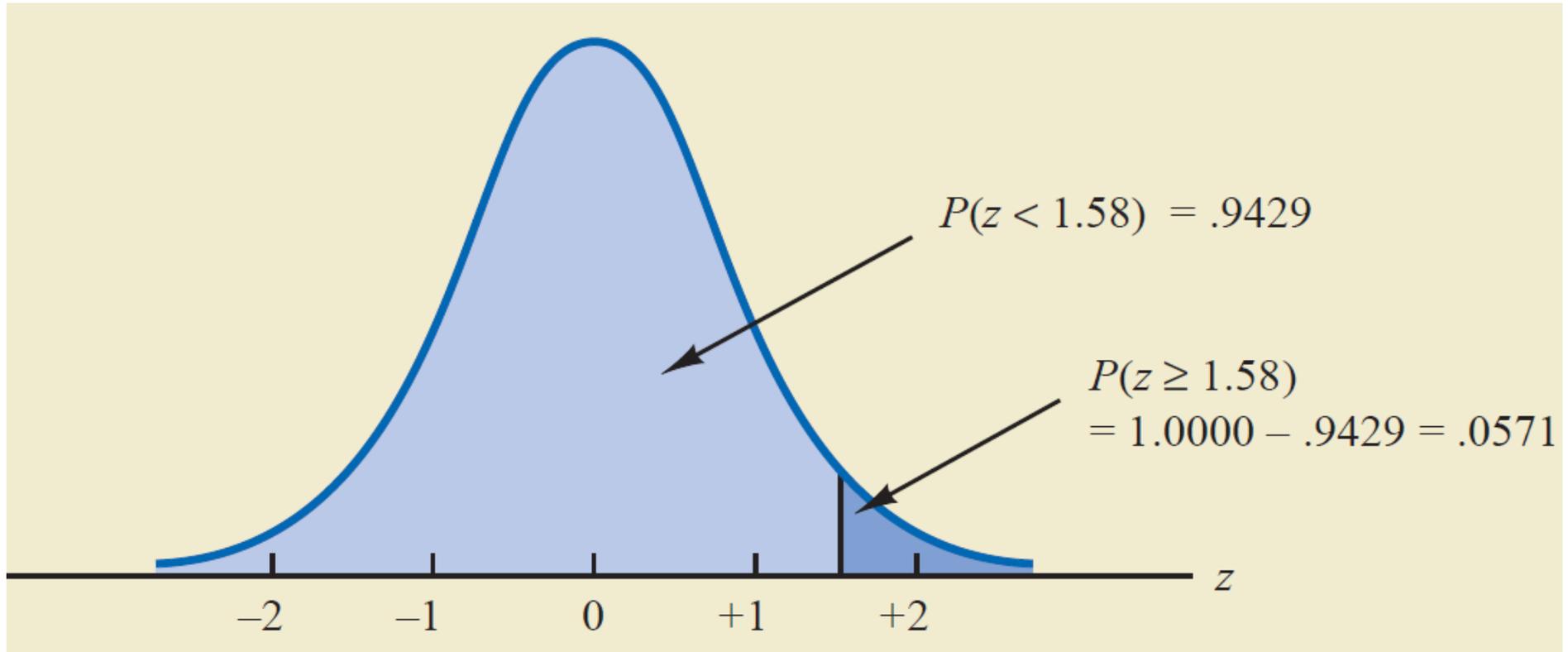


$$\begin{aligned} P(-0.5 < z \leq 1.25) &= P(z \leq 1.25) - P(z \leq -0.5) \\ &= 0.8944 - 0.3085 \\ &= 0.5859 \end{aligned}$$

How to use the SNPD:



How to use the SNPD:

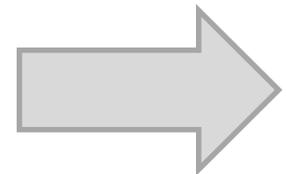


Computing probabilities for any NPD

KEY IDEA:

- **convert** to the **standard** normal random variable
- **use the table** available on Brightspace
-and the observations made on the previous slides

$$Z = \frac{x - \mu}{\sigma}$$



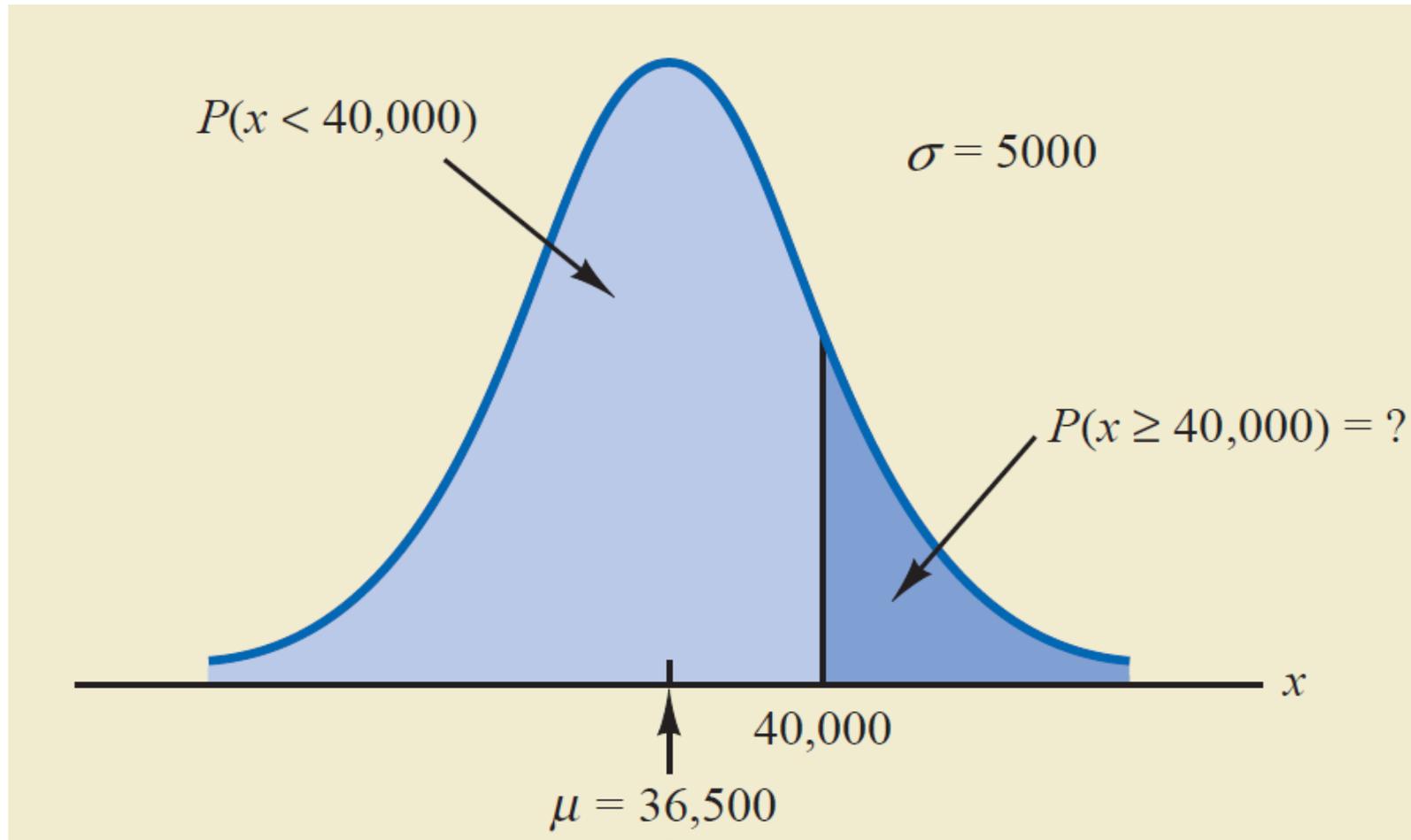
Example

From actual road test, a tyre company estimated that the mean mileage is $\mu = 36,500$ miles, and that the SDV is $\sigma = 5000$.

In addition, the data collected indicates that a normal distribution is a reasonable assumption.

What percentage of the tires can be expected to last more than 40,000 miles?

Example (cont'd)



Example (cont'd)

At $x = 40,000$, we have

$$z = \frac{x - \mu}{\sigma} = \frac{40,000 - 36,500}{5000} = \frac{3500}{5000} = .70$$

$$\begin{aligned} P(x \geq 40,000) &= P(z \geq 0.7) = 1 - P(z < 0.7) \\ &= 1 - 0.7580 \\ &= 0.2420 \end{aligned}$$

We conclude that about **24.2%** of the tyres will exceed 40,000 in mileage.