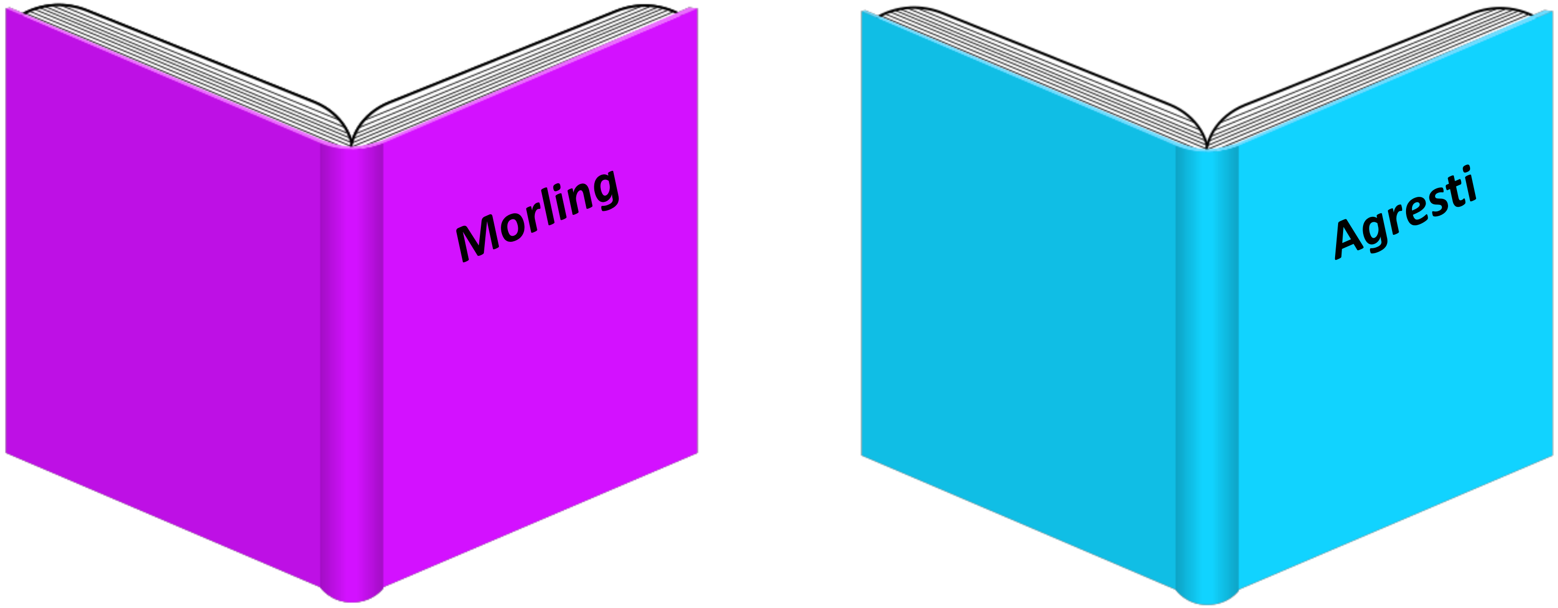


Research Methods and Statistics

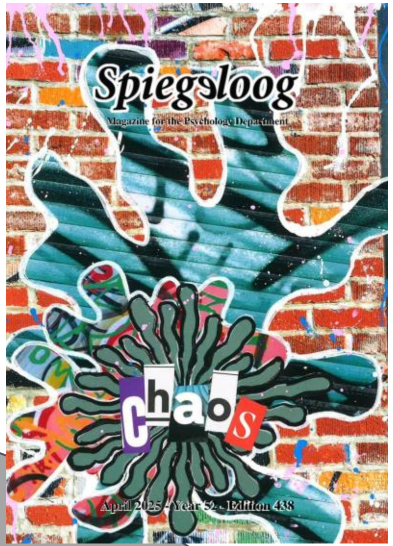
Lecture 9: Probability Distributions

Riet van Bork

Pictures source: pixabay.org

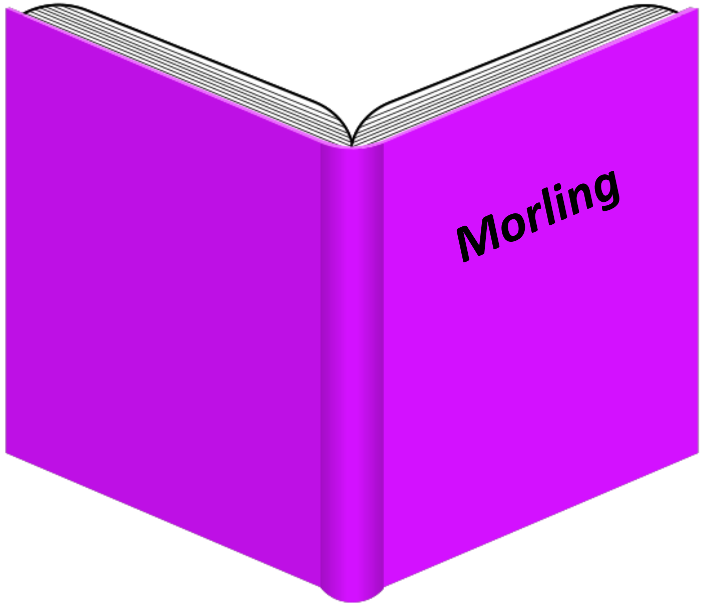
Spiegeloog

Tijdschrift voor de Afdeling Psychologie



spiegeloog-fmg@uva.nl

Wednesdays Room G3.07



Research Methods and Statistics

Lecture 9: Probability Distributions

Riet van Bork

Pictures source: pixabay.org

Clash of books?

- Last year, observer effect, observer bias, and reactivity (terms used in the RMS book) differed from similar concepts in the introductory psychology book.
- In RMS, there is a distinction between observer effect and observer bias:
    - Observer effect (or ‘expectancy effect’): the observer influences the subject’s behavior
    - Observer bias: the observer’s expectation influences their *interpretation* of behavior
- Now there is a new book for introductory psychology. Is it consistent with the definitions in the RMS book?
- In case there are inconsistencies, let me know. And for the RMS exam, base your answers on the RMS reading material

Today

- **Random variables**

- Probability distributions
  - Discrete variables
    - Three different probability calculations
    - Mean and variability
  - Continuous variables

Two specific distributions:

- Normal distribution (possible for continuous variables)
  - Mean and variability
  - Three different probability calculations

- Binomial distribution (possible for discrete variables)
  - General formulae
  - Mean and variability

Random variables

Random variable: A numerical measure of the outcome of a random phenomenon

"Effectiveness of therapy depends on number of close relationships"

Variable: "effective": 0 (no), 1 (yes)
Variable: "number of close relationships": 1, 2, 3,.. Etc (positive integer)

Random variables

Random variable: A numerical measure of the outcome of a random phenomenon

Often, randomness results from the use of random sampling to gather the data

Random phenomena:

Randomly drawing a patient and evaluating whether the treatment is effective → Variable "effective": 0 (no), 1 (yes)

Randomly drawing a student and taking their height → Variable "height": number

It is called a variable because the values vary over something (in this case over people)

Randomly drawing a student and asking their hobby → Variable "hobby": 0 ("tennis"), 1 ("shopping"), 2 ("netflix"), etc.

Randomly drawing a student and counting their number of siblings → Variable "number of siblings": positive integer

Random variables

Random variable: A numerical measure of the outcome of a random phenomenon

Random phenomena:

Randomly drawing a patient and seeing whether the treatment is effective	→ Variable "effective": 0 (no), 1 (yes)	Categorical
Randomly drawing a student and taking their height	→ Variable "height": number	Continuous
Randomly drawing a student and asking their hobby	→ Variable "hobby": 0 ("tennis"), 1 ("shopping"), 2 ("netflix"), etc.	Categorical
Randomly drawing a student and counting their number of siblings	→ Variable "number of siblings": positive integer	Discrete

Today

- Random variables

- Probability distributions
  - **Discrete variables**
    - Three different probability calculations
    - Mean and variability
  - Continuous variables

Two specific distributions:

- Normal distribution (possible for continuous variables)
  - Mean and variability
  - Three different probability calculations

- Binomial distribution (possible for discrete variables)
  - General formulae
  - Mean and variability

Example

Nine key symptoms (“A” criteria) of Major Depression according to DSM IV:

1. depressed mood
2. loss of interest
3. appetite or weight change
4. Insomnia or hypersomnia
5. psychomotor agitation or retardation
6. fatigue
7. feelings of worthlessness or guilt
8. diminished ability to think or concentrate or indecisiveness
9. thoughts of death or suicide

Example

- If we select a random person living in the Netherlands. How many of these symptoms will that person have?

- → Number of symptoms is a **discrete** random variable
- Possible values: 0, 1, 2, 3, ..., 9

- But not all values are equally probable
  - Probability of “9 symptoms” is much smaller than probabiliy of “0 symptoms”

Image source: pixabay.com

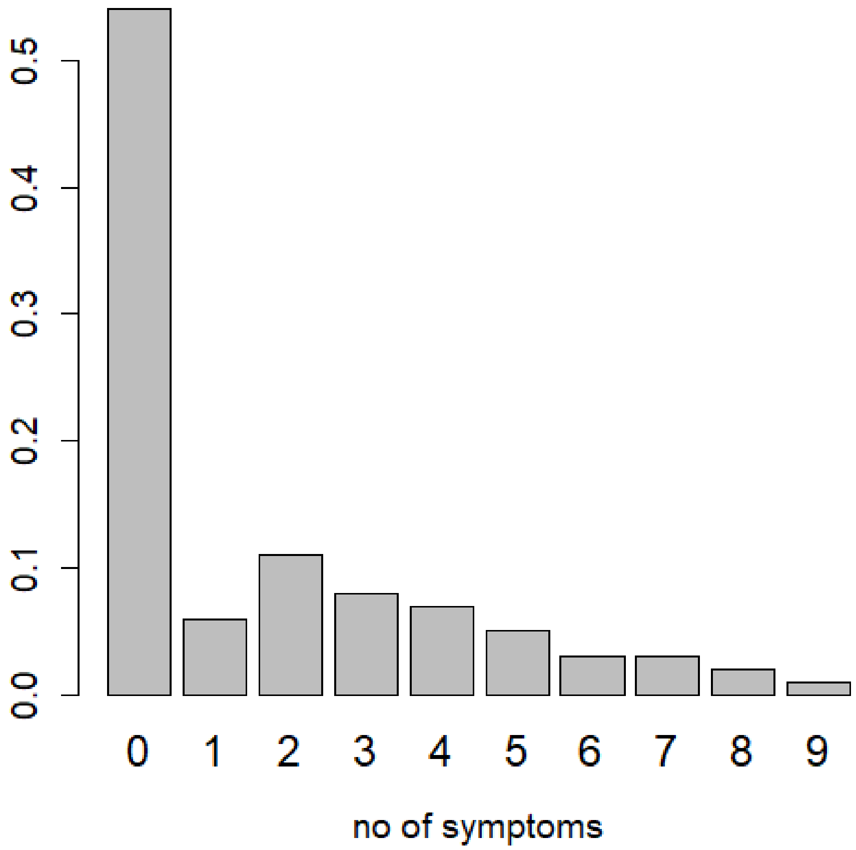
- X = number of symptoms
- Values: x = 0, x = 1, x = 2, x = 3, x = 4, x = 5, x = 6, x = 7, x = 8, x = 9

x	P(x)
0	0.54
1	0.06
2	0.11
3	0.08
4	0.07
5	0.05
6	0.03
7	0.03
8	0.02
9	0.01

Probability distribution

Sum of these probabilities?

Total probability is always 1!!
(i.e., 0.54+0.06+…+0.01 = 1)



Probability distribution: An overview of how probable each value of a random variable is

Today

- Random variables

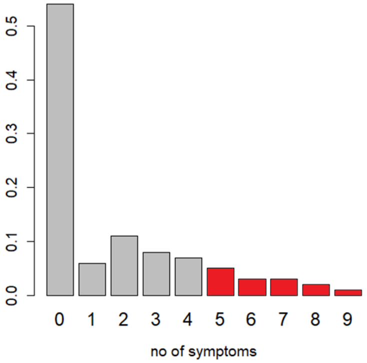
- Probability distributions
  - Discrete variables
    - **Three different probability calculations**
    - Mean and variability
  - Continuous variables

Two specific distributions:

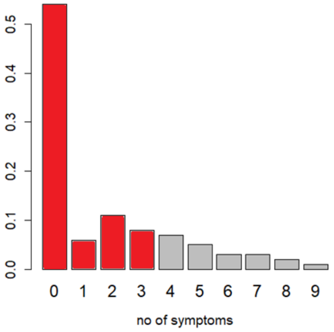
- Normal distribution (possible for continuous variables)
  - Mean and variability
  - Three different probability calculations

- Binomial distribution (possible for discrete variables)
  - General formulae
  - Mean and variability

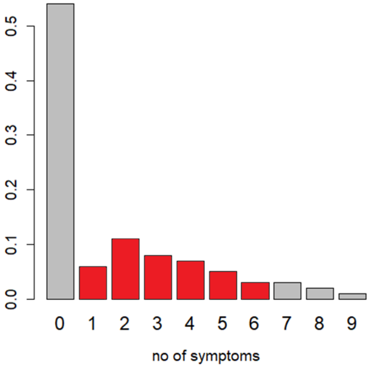
Three different probability calculations



1) Probability of 5 or more symptoms
- $P(X \geq 5)$



2) Probability of 3 or fewer symptoms
- $P(X \leq 3)$



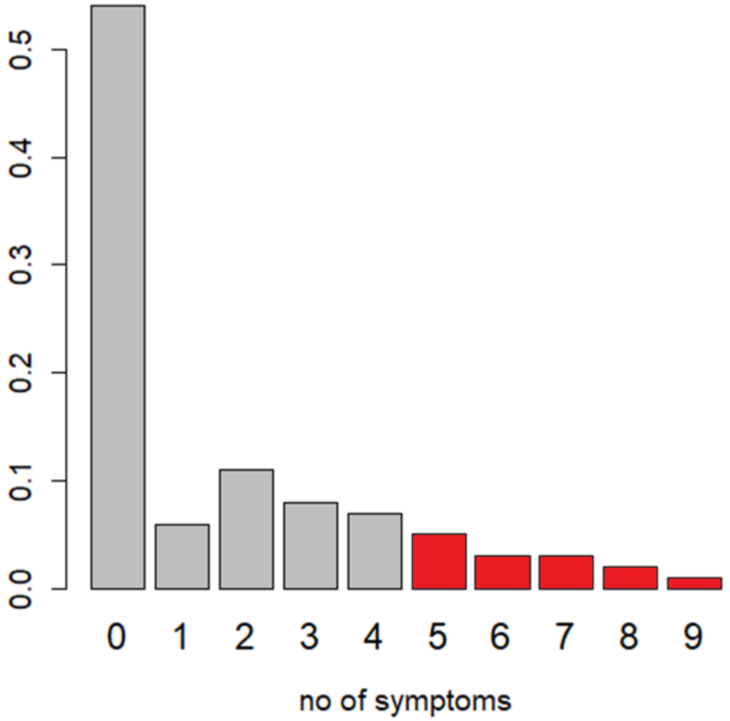
3) Probability of at least 1 and at most 6 symptoms
- $P(1 \leq X \leq 6)$

- You are diagnozed as “depressed” if you have 5 or more symptoms
- X = number of symptoms
- → P(depression) = P(5 or more) = P(X ≥ 5) = **P(5) + P(6) + P(7) + P(8) + P(9)**

= **0.05 + 0.03 + 0.03 + 0.02 + 0.01** = **0.14**

We can simply add them up because it is OR 5 OR 6 OR 7 OR 8 OR 9. And they are disjoint

x	P(x)
0	0.54
1	0.06
2	0.11
3	0.08
4	0.07
5	**0.05**
6	**0.03**
7	**0.03**
8	**0.02**
9	**0.01**

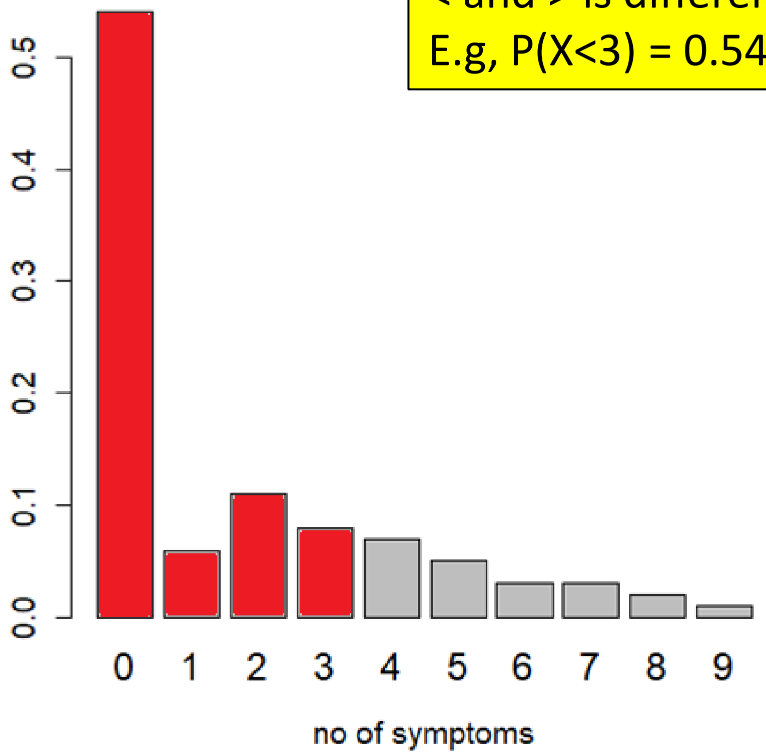


- What is the probability of 3 or fewer symptoms
- X = number of symptoms
- → P(depression) = P(3 or fewer) = P(X ≤ 3) = **P(0) + P(1) + P(2) + P(3)**

= **0.54 + 0.06 + 0.11 + 0.08 = 0.79**

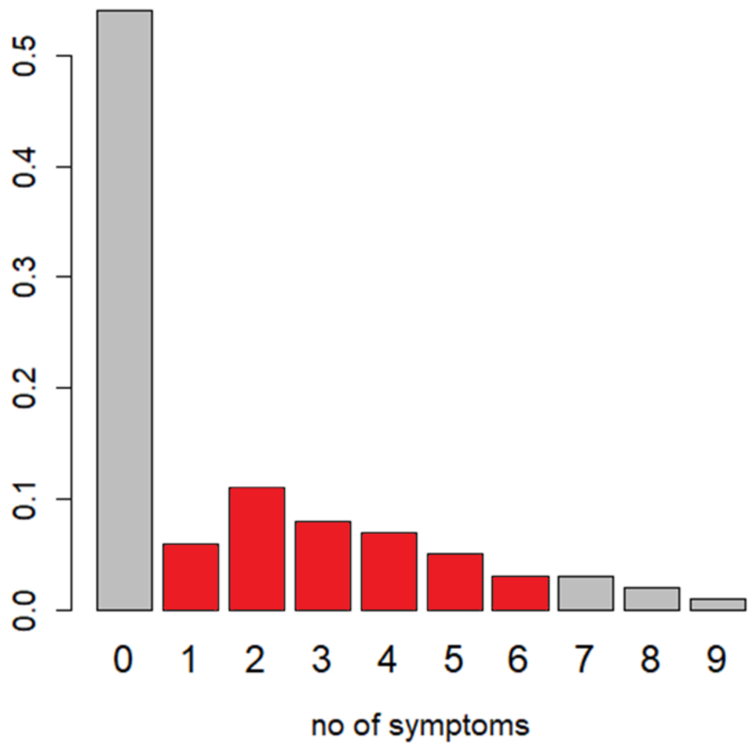
Be careful as for discrete distributions < and > is different from ≤ and ≥
E.g, P(X<3) = 0.54 + 0.06 + 0.11 = 0.71

x	P(x)
0	**0.54**
1	**0.06**
2	**0.11**
3	**0.08**
4	0.07
5	0.05
6	0.03
7	0.03
8	0.02
9	0.01



- What is probability of number of symptoms between 1 and 6?
- X = number of symptoms
- → $P(X=1 \text{ OR } X=2 \ldots \text{ OR } X=6) = P(1 \leq X \leq 6) =$ **$P(1) + P(2) + P(3) + \ldots + P(6)$**
- → $P(1 \leq X \leq 6) =$ **$0.06 + 0.11 + 0.08 + 0.07 + 0.05 + 0.03 = 0.4$**

x	P(x)
0	0.54
1	**0.06**
2	**0.11**
3	**0.08**
4	**0.07**
5	**0.05**
6	**0.03**
7	0.03
8	0.02
9	0.01



Today

- Random variables

- Probability distributions
  - Discrete variables
    - Three different probability calculations
    - **Mean and variability**
  - Continuous variables

Two specific distributions:

- Normal distribution (possible for continuous variables)
  - Mean and variability
  - Three different probability calculations

- Binomial distribution (possible for discrete variables)
  - General formulae
  - Mean and variability

Mean (or expected value), μ

$\mu = \Sigma x P(x)$

- a weighted average
- **don't** do: (0+1+2+3+4+5+6+7+8+9)/10 = 4.5

x	P(x)
0	**0.54**
1	**0.06**
2	**0.11**
3	**0.08**
4	**0.07**
5	**0.05**
6	**0.03**
7	**0.03**
8	**0.02**
9	**0.01**

μ =

0 x P(0) + 1 x P(1) + 2 x P(2) + 3 x P(3) + 4 x P(4) + 5 x P(5) + 6 x P(6) + 7 x P(7) + 8 x P(8) + 9 x P(9)

= 0.00 + 0.06 + 0.22 + 0.24 + 0.28 + 0.25 + 0.18 + 0.21 + 0.16 + 0.09 =

= 1.69

Standard deviation, σ

- Indicates the amount of variability or dispersion in the population
- If you randomly sample from the population, σ describes how far, on average, the sampled value will be from from the mean of the distribution
- No formula yet, you'll get a formula for a specific

distribution later this lecture

x	P(x)
0	0.54
1	0.06
2	0.11
3	0.08
4	0.07
5	0.05
6	0.03
7	0.03
8	0.02
9	0.01

Today

- Random variables

- Probability distributions
  - Discrete variables
    - Three different probability calculations
    - Mean and variability
  - **Continuous variables**

Two specific distributions:

- Normal distribution (possible for continuous variables)
  - Mean and variability
  - Three different probability calculations

- Binomial distribution (possible for discrete variables)
  - General formulae
  - Mean and variability

Continuous variables

- Continuous variables that can take any decimal value
  - e.g., time to solve an arithmetic item (10.71 s, 13.83 s, etc)
  - e.g., heigth (171.19 cm, 156.53 cm, etc)
  - e.g., weight babies (3462.83 gr, 4362.17 gr, etc)

Picture source: pxabay.com

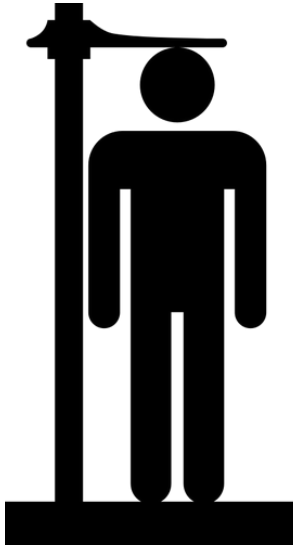
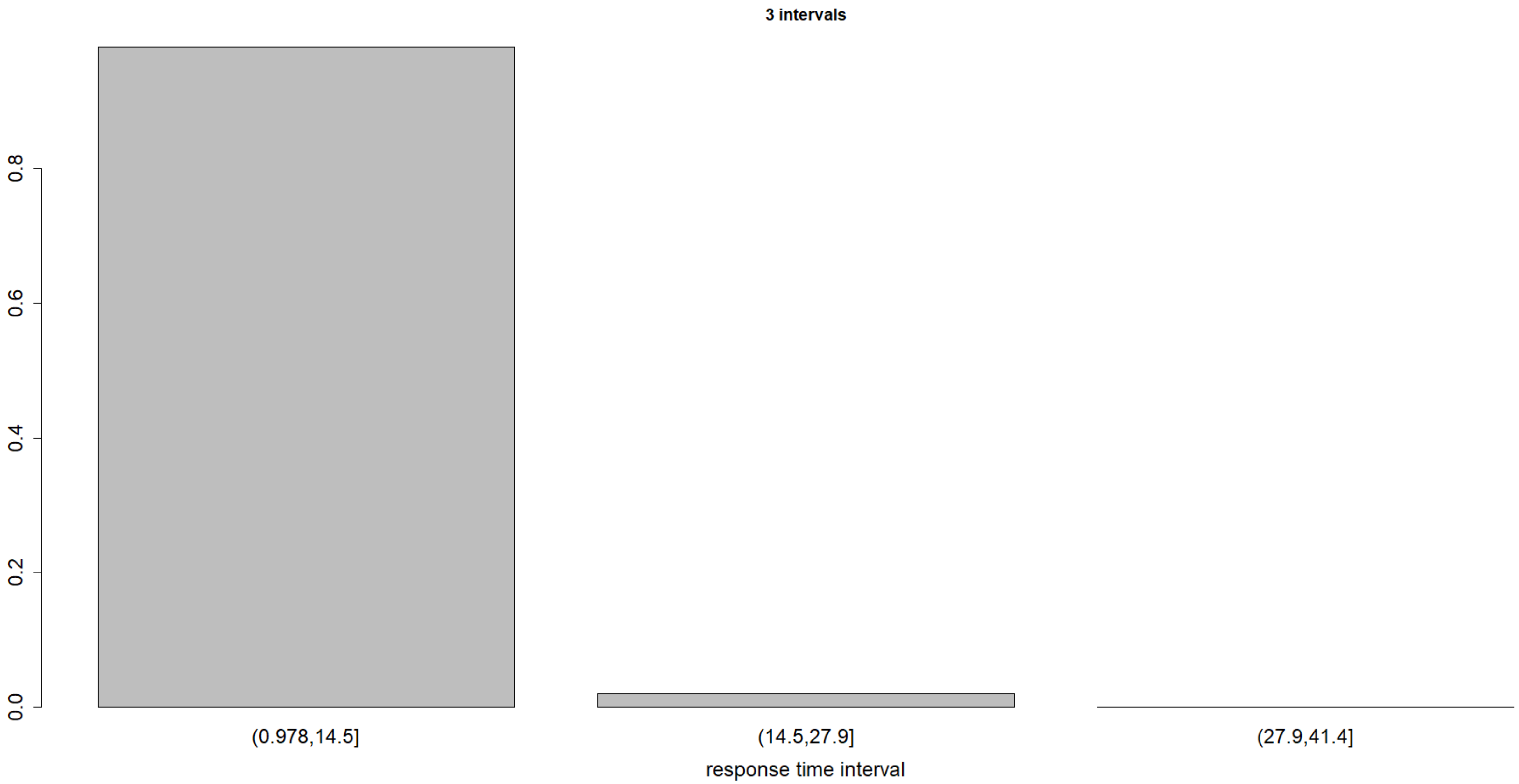
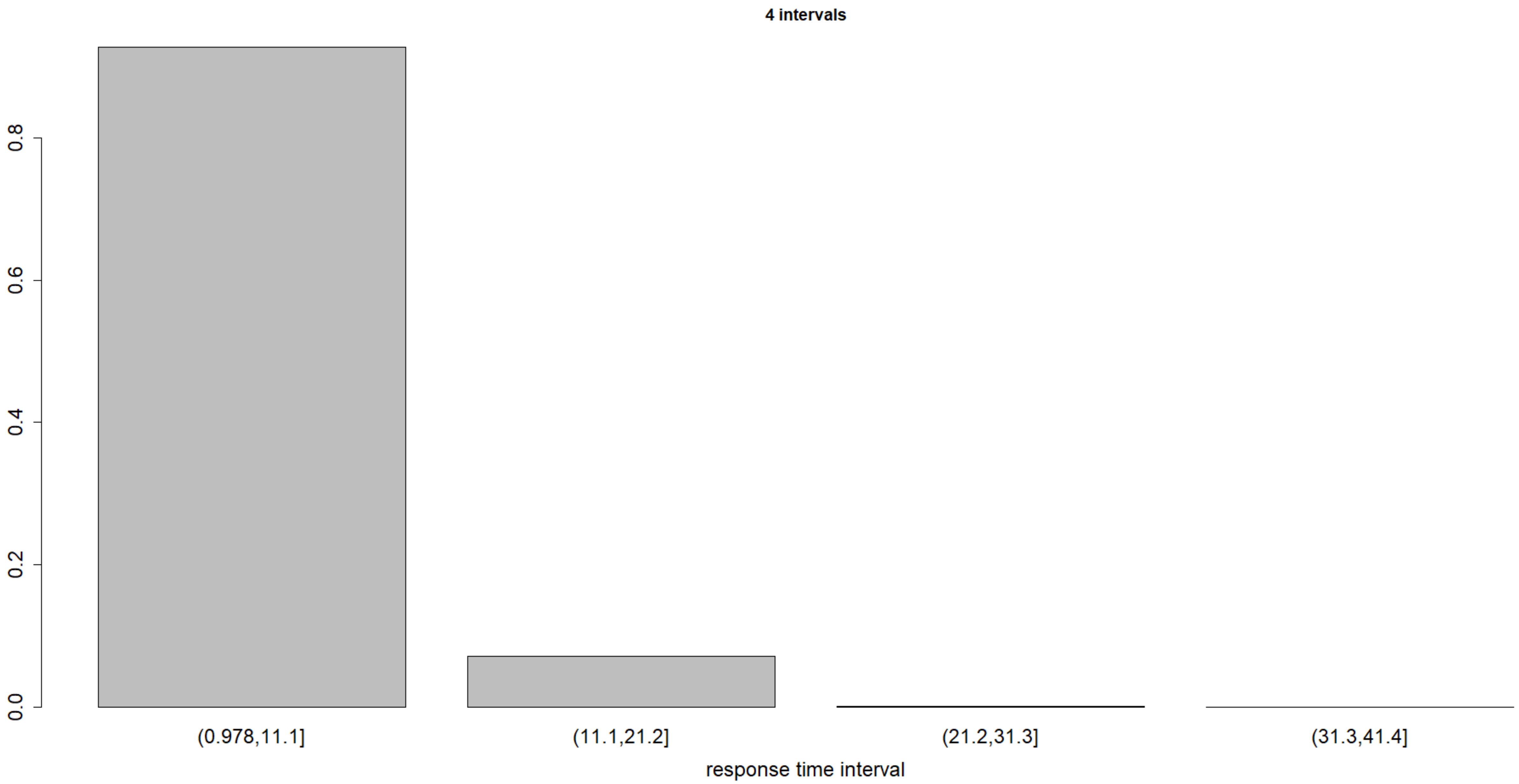


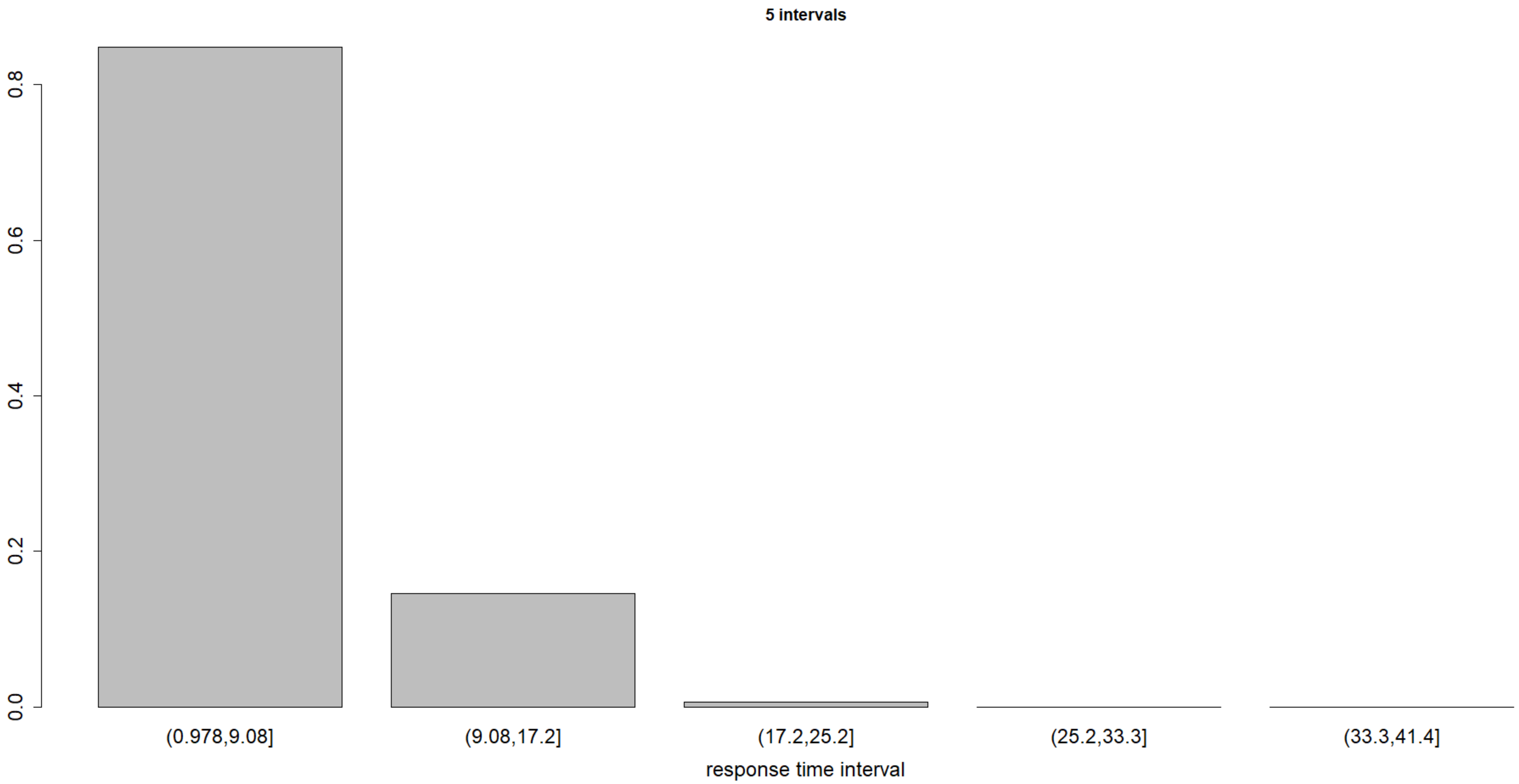
Photo source: https://publicdomainvectors.org/en/free-clipart/Height-measurement/64875.html

Example continuous variable: Time to solve an arithmetic item

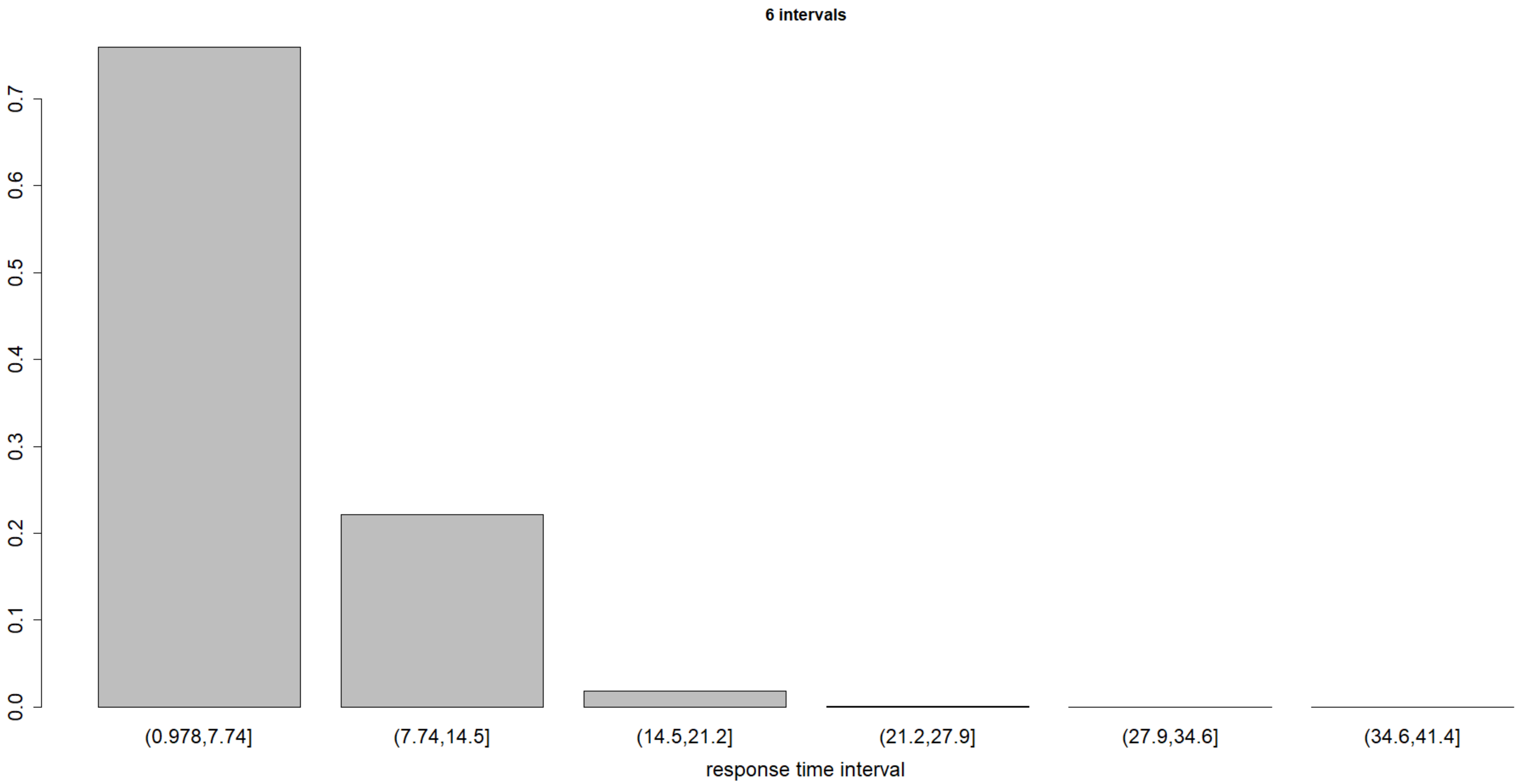
Subject	Response time (s)
1	13.46649
2	2.662474
3	8.177177
4	1.832033
5	5.550982
6	4.835395
7	3.124351
…	….
1000	5.9045371

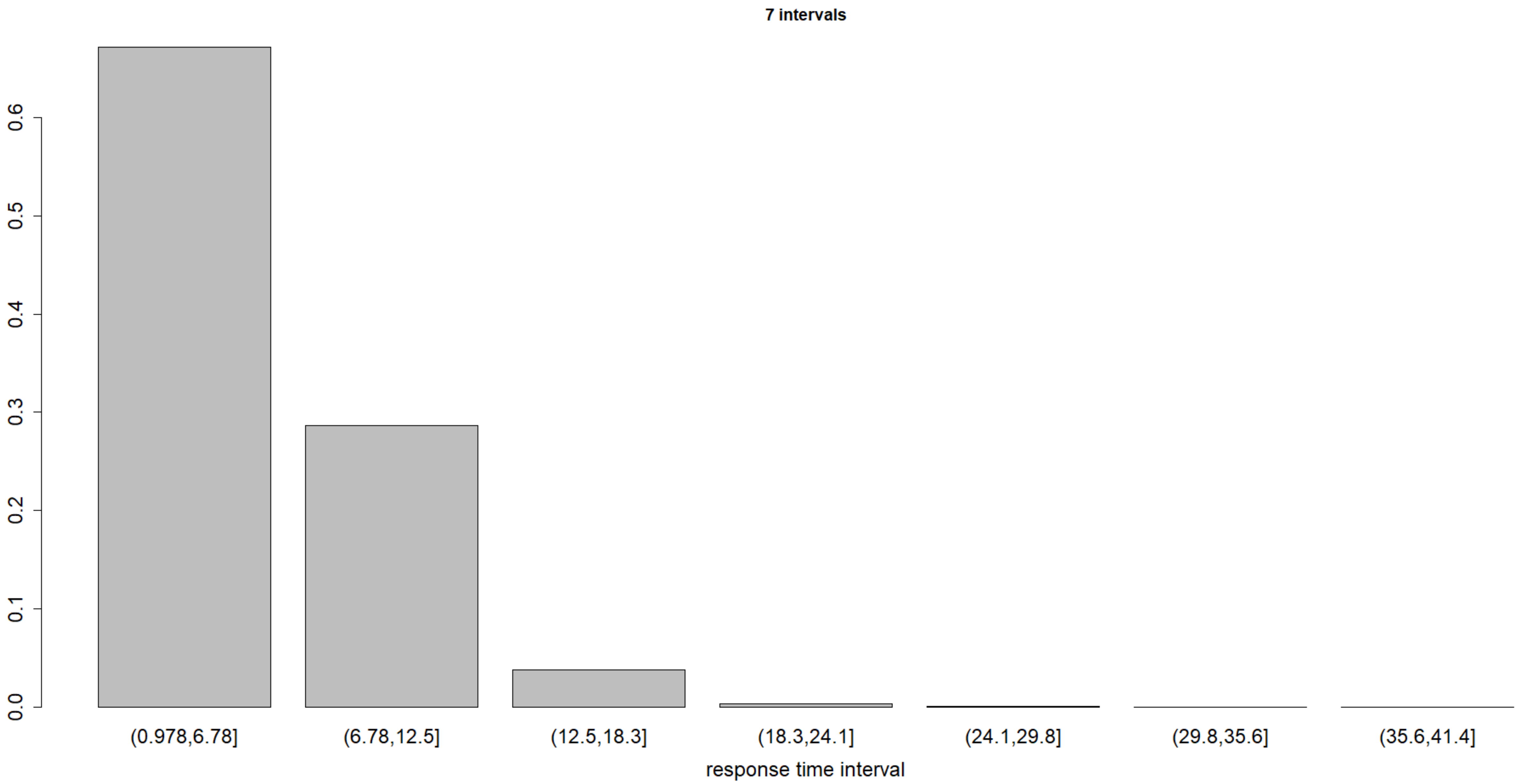

3 intervals
0.0
0.2
0.4
0.6
0.8
(0.978,14.5]
(14.5,27.9]
(27.9,41.4]
response time interval

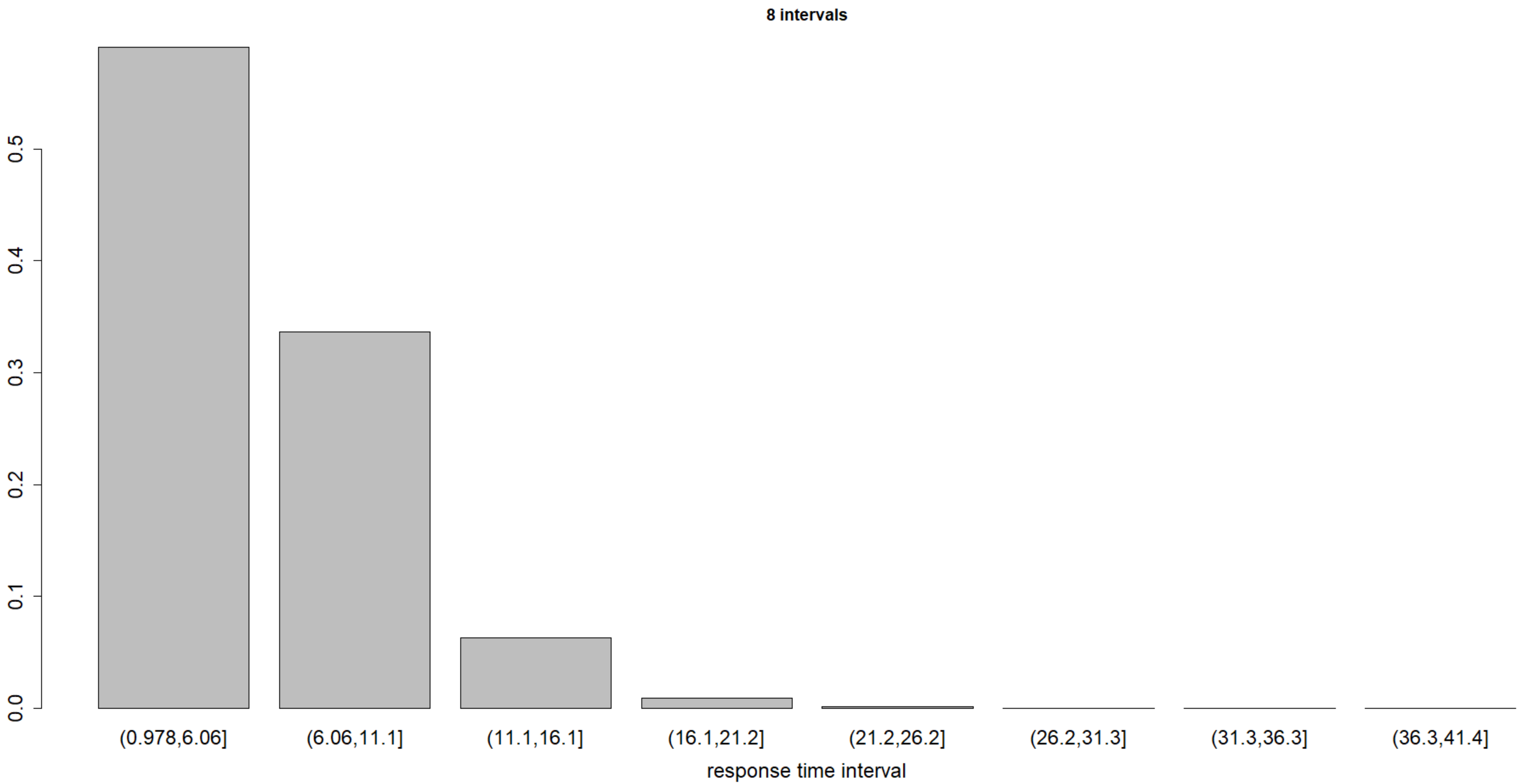

4 intervals
0.0
0.2
0.4
0.6
0.8
(0.978,11.1]
(11.1,21.2]
(21.2,31.3]
(31.3,41.4]
response time interval

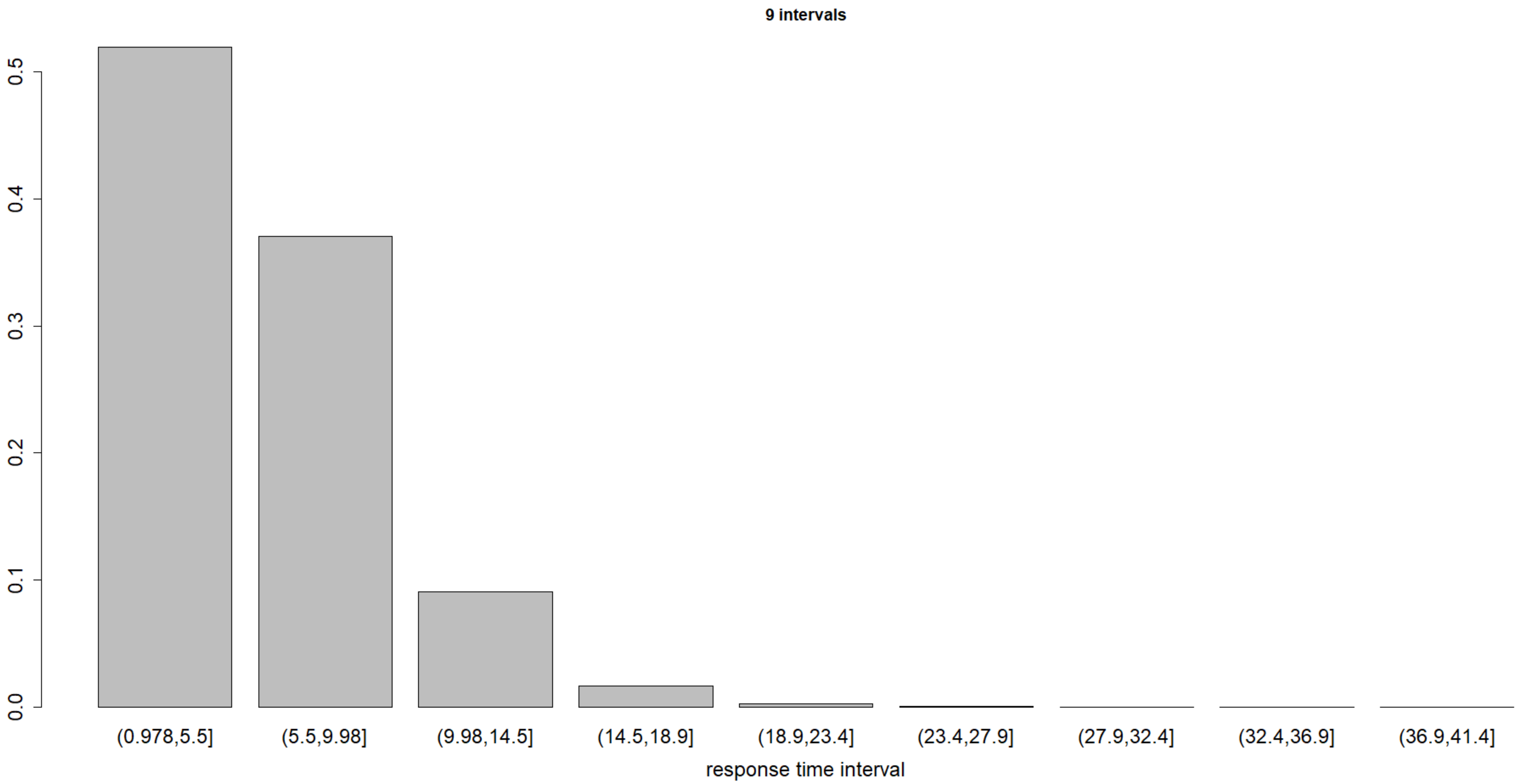


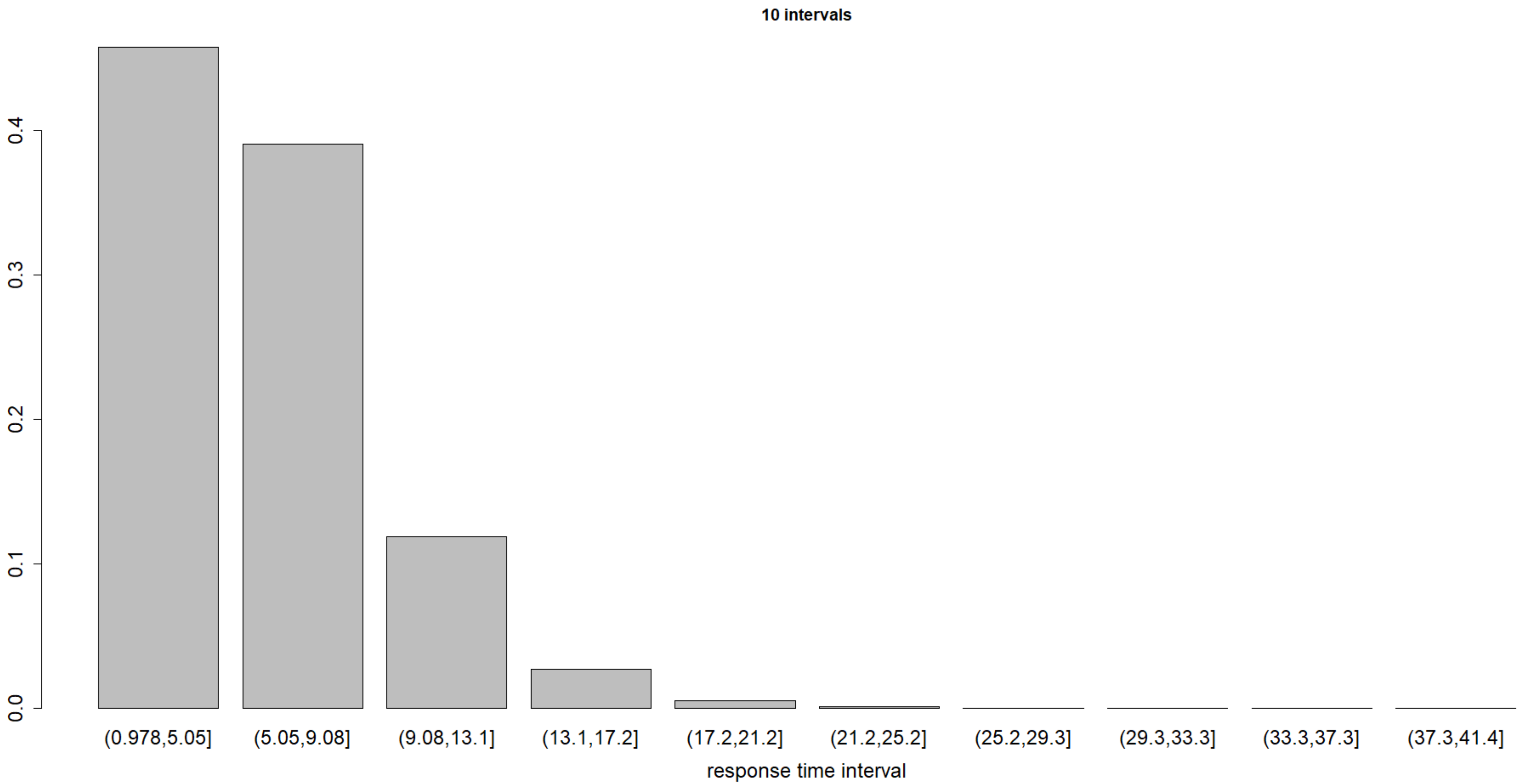
5 intervals
0.0
0.2
0.4
0.6
0.8
(0.978,9.08]
(9.08,17.2]
(17.2,25.2]
(25.2,33.3]
(33.3,41.4]
response time interval

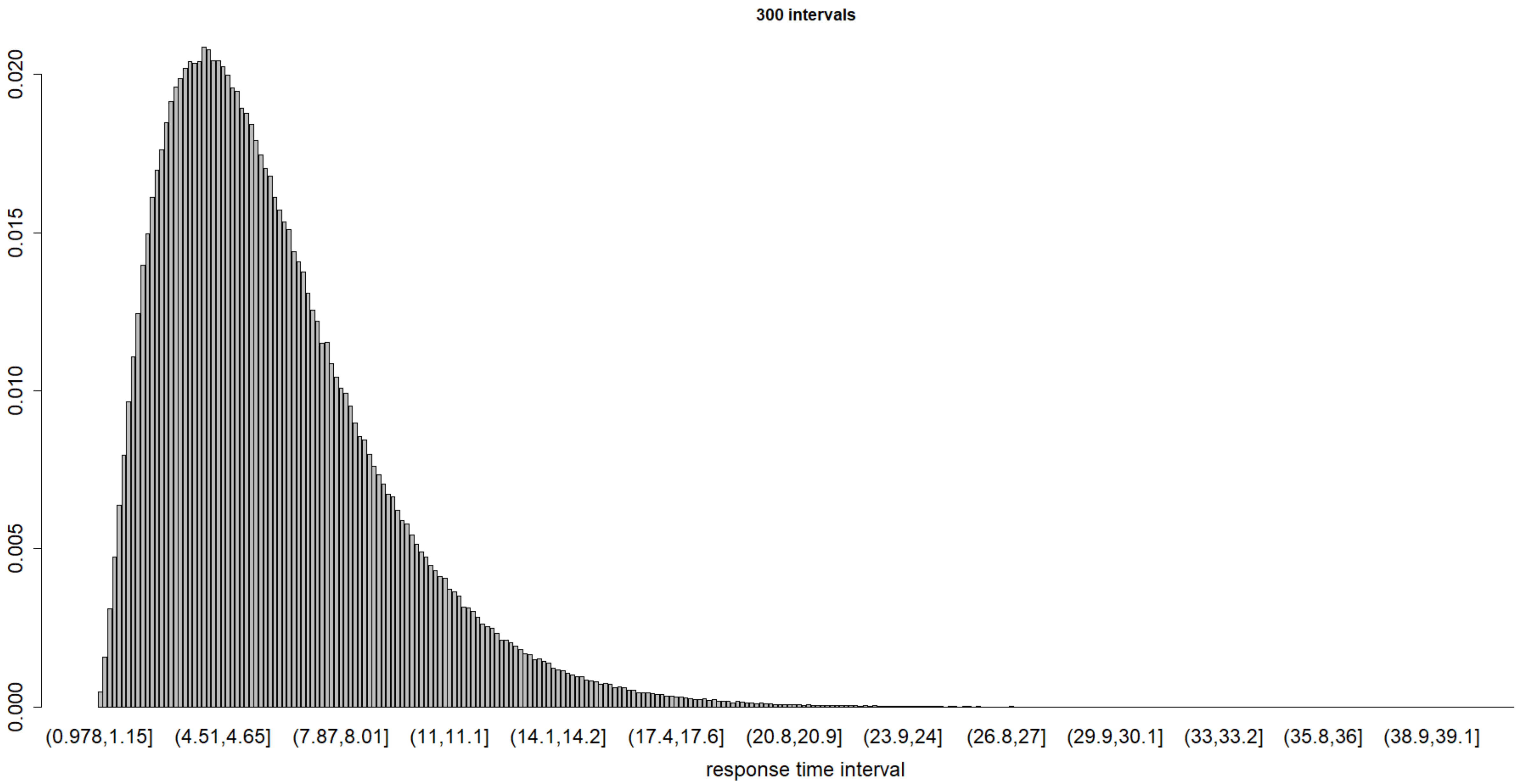

6 intervals
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
(0.978,7.74]
(7.74,14.5]
(14.5,21.2]
(21.2,27.9]
(27.9,34.6]
(34.6,41.4]
response time interval

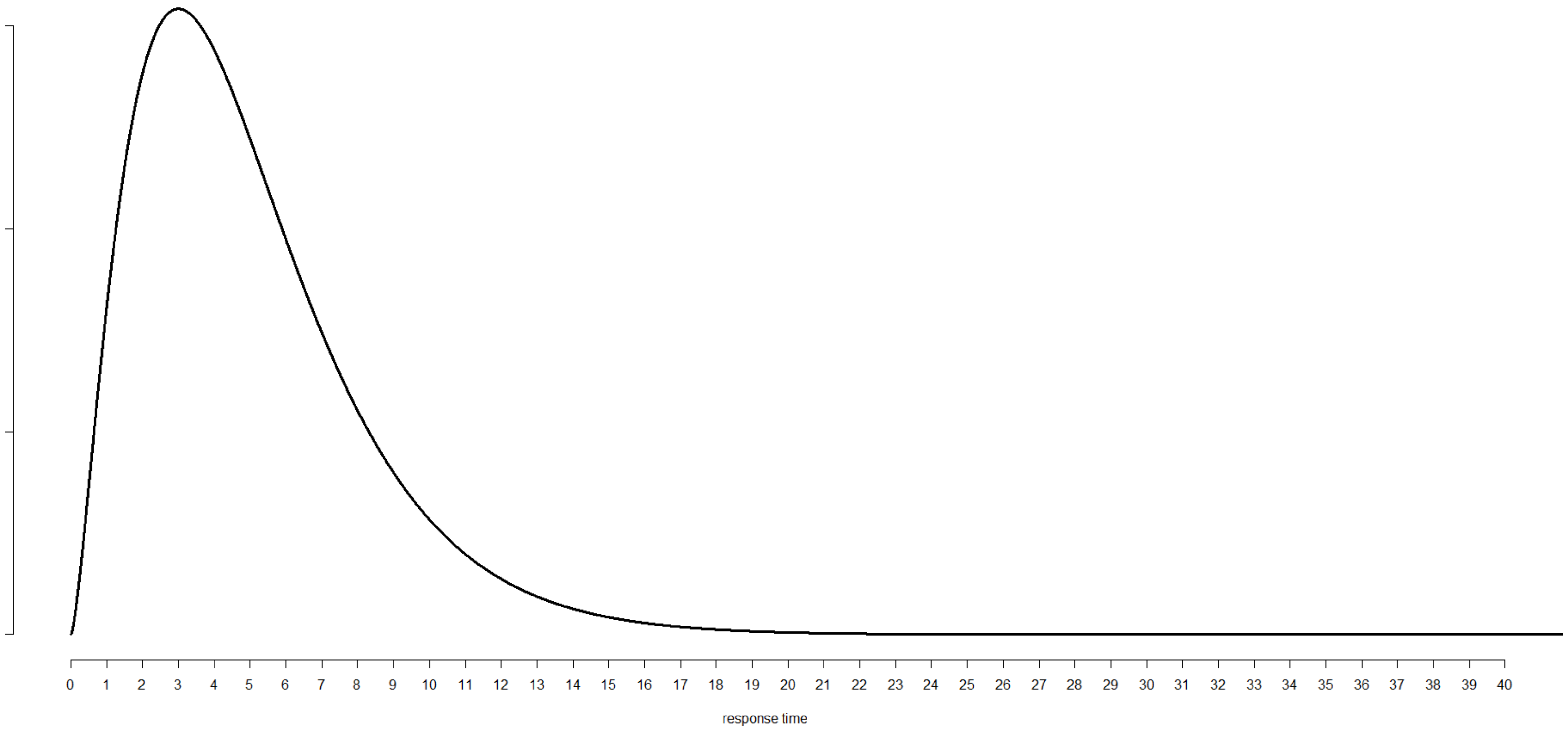

7 intervals
0.0
0.1
0.2
0.3
0.4
0.5
0.6
(0.978,6.78]
(6.78,12.5]
(12.5,18.3]
(18.3,24.1]
(24.1,29.8]
(29.8,35.6]
(35.6,41.4]
response time interval

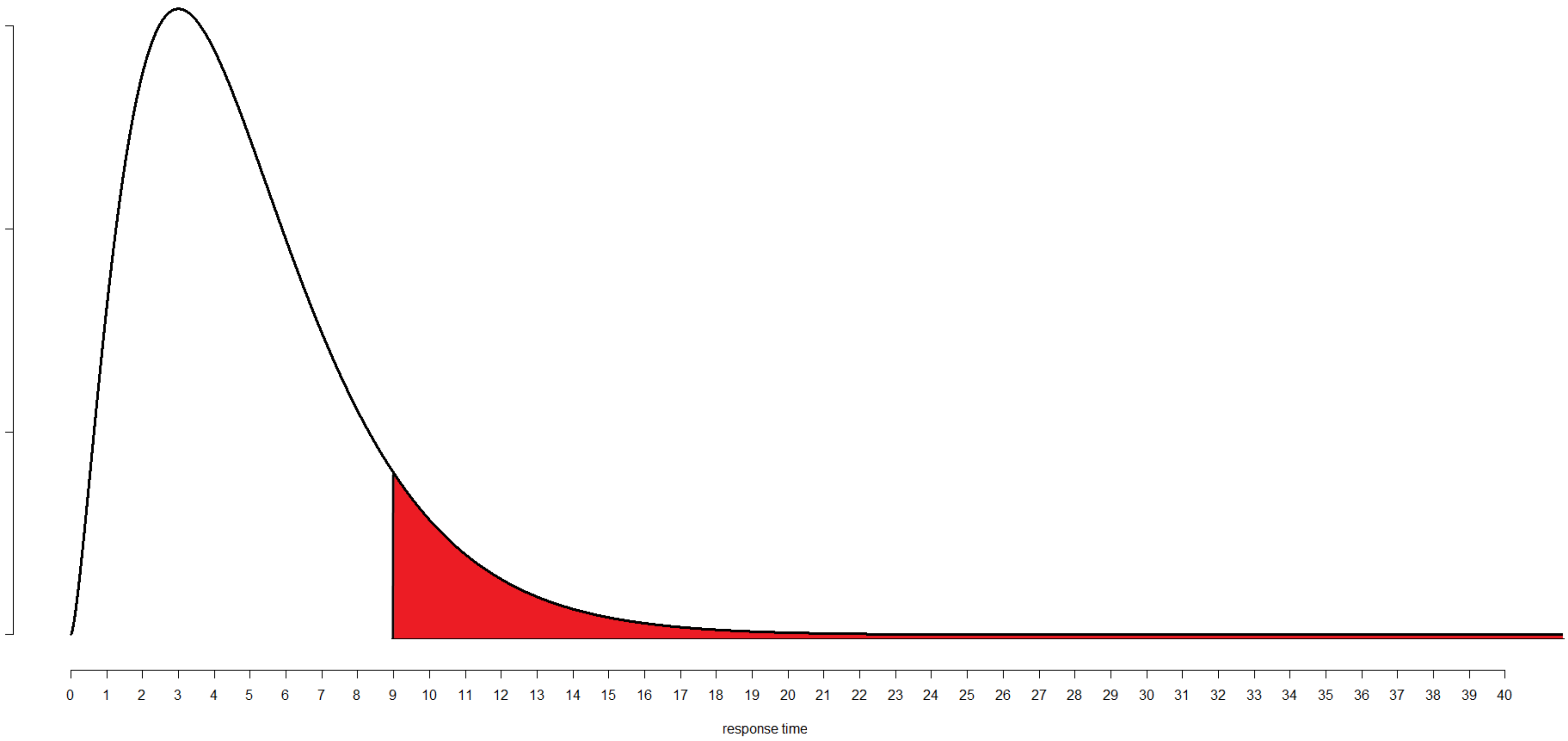

8 intervals
0.0
0.1
0.2
0.3
0.4
0.5
(0.978,6.06]
(6.06,11.1]
(11.1,16.1]
(16.1,21.2]
(21.2,26.2]
(26.2,31.3]
(31.3,36.3]
(36.3,41.4]
response time interval

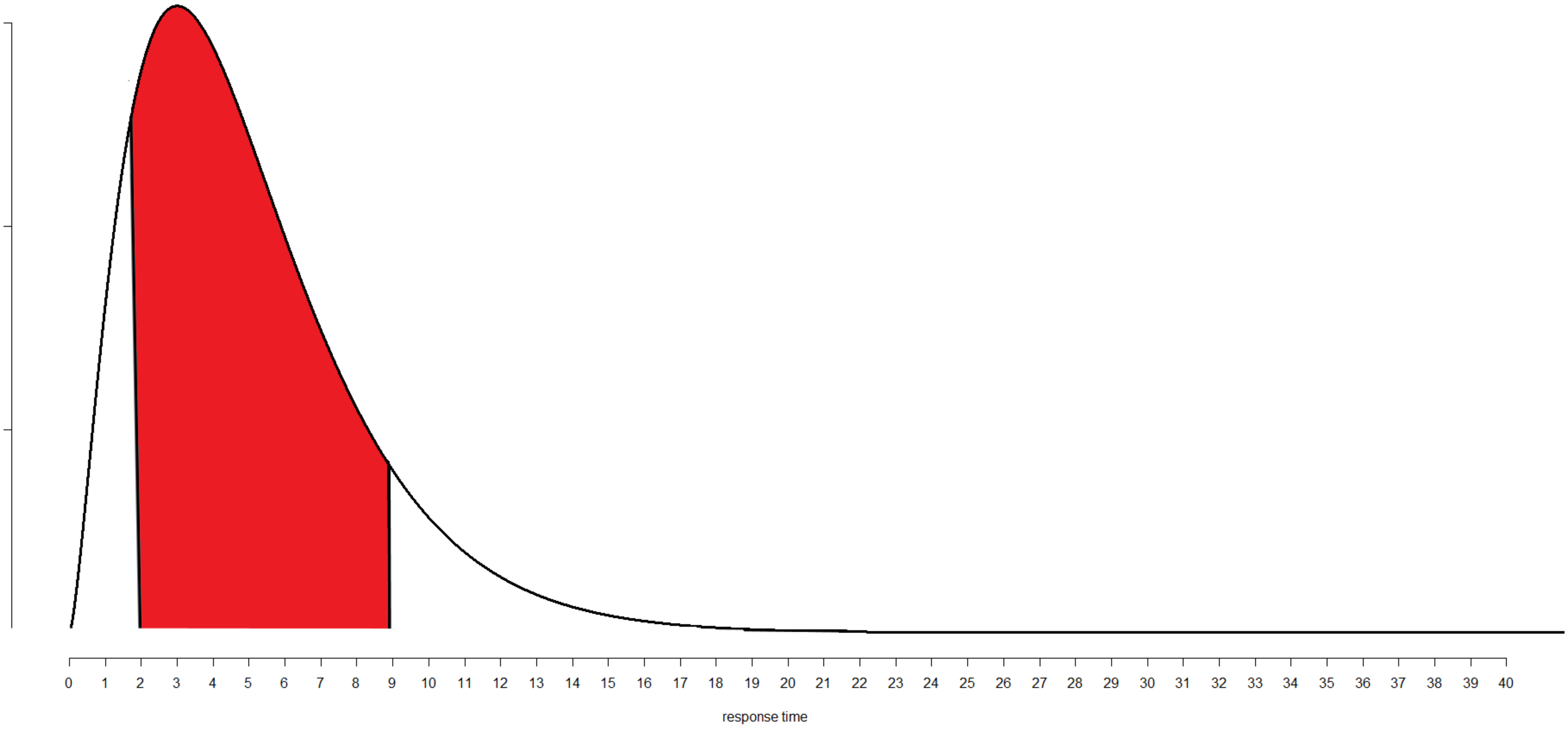

9 intervals
0.0
0.1
0.2
0.3
0.4
0.5
(0.978,5.5]
(5.5,9.98]
(9.98,14.5]
(14.5,18.9]
(18.9,23.4]
(23.4,27.9]
(27.9,32.4]
(32.4,36.9]
(36.9,41.4]
response time interval


10 intervals
0.0
0.1
0.2
0.3
0.4
(0.978,5.05]
(5.05,9.08]
(9.08,13.1]
(13.1,17.2]
(17.2,21.2]
(21.2,25.2]
(25.2,29.3]
(29.3,33.3]
(33.3,37.3]
(37.3,41.4]
response time interval


300 intervals
0.000
0.005
0.010
0.015
0.020
(0.978,1.15]
(4.51,4.65]
(7.87,8.01]
(11,11.1]
(14.1,14.2]
(17.4,17.6]
(20.8,20.9]
(23.9,24]
(26.8,27]
(29.9,30.1]
(33,33.2]
(35.8,36]
(38.9,39.1]
response time interval

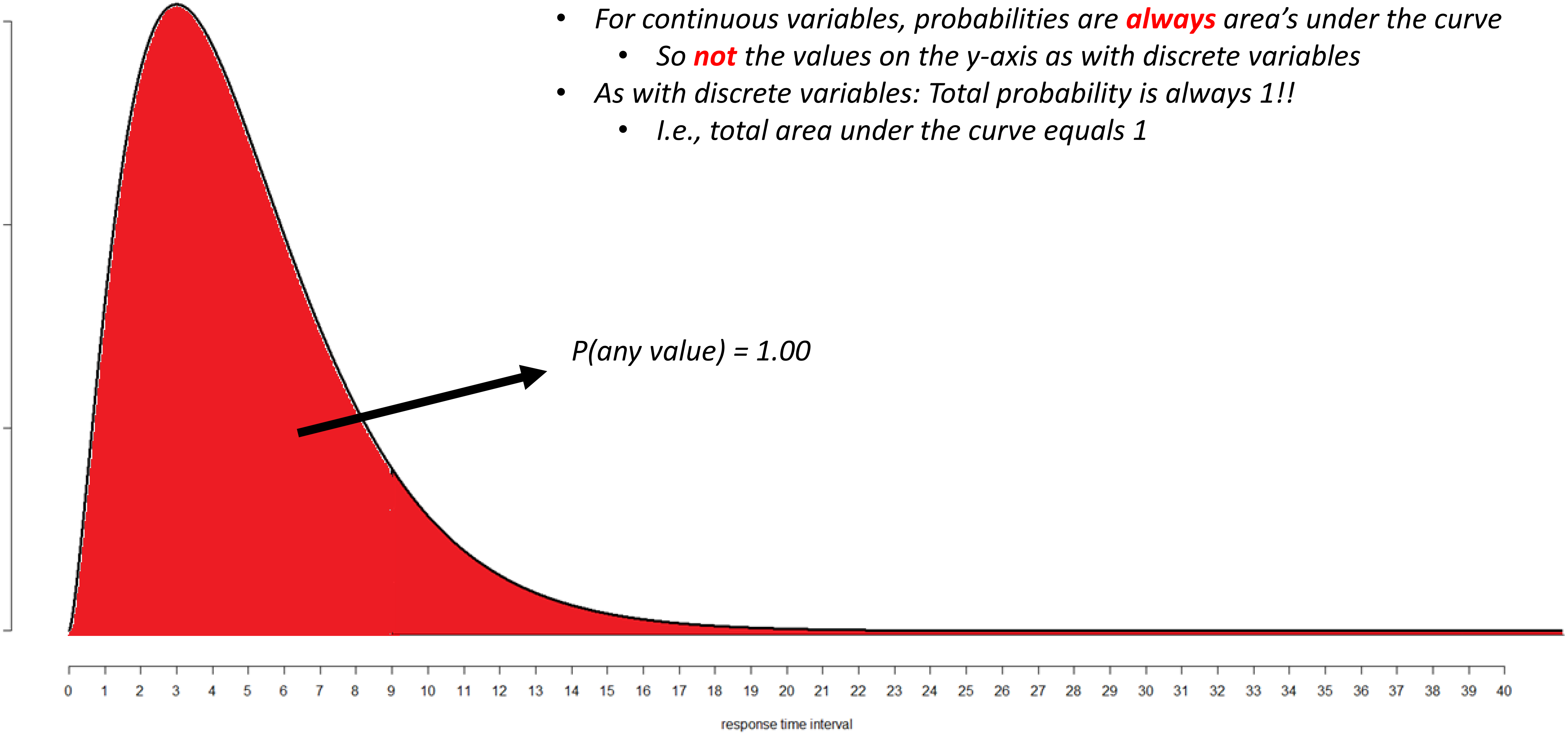

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
response time


0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
response time


0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
response time

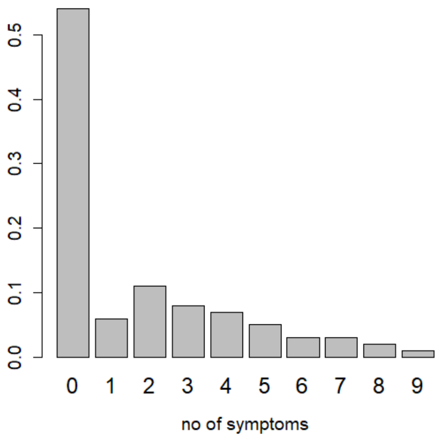
Some notes:

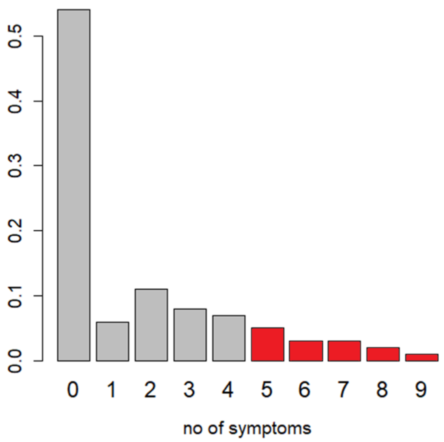
- *For continuous variables, probabilities are* ***always*** *area's under the curve*
    - *So* ***not*** *the values on the y-axis as with discrete variables*
- *As with discrete variables: Total probability is always 1!!*
    - *I.e., total area under the curve equals 1*

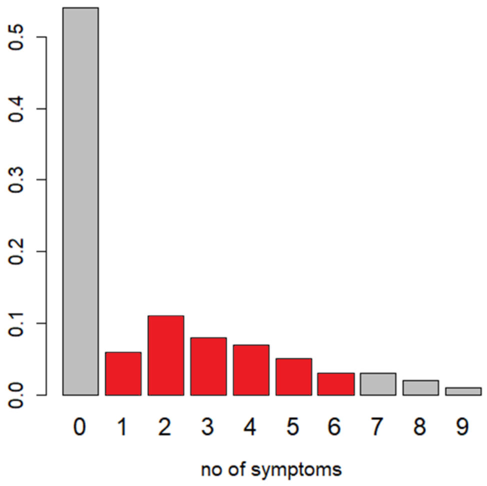


To summarize:

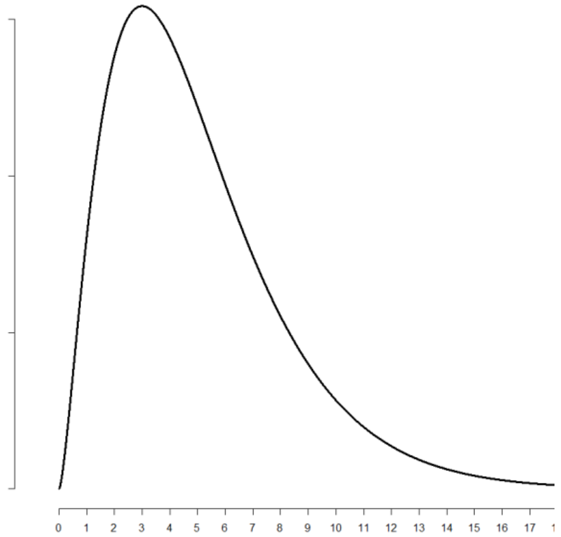
Discrete Distributions

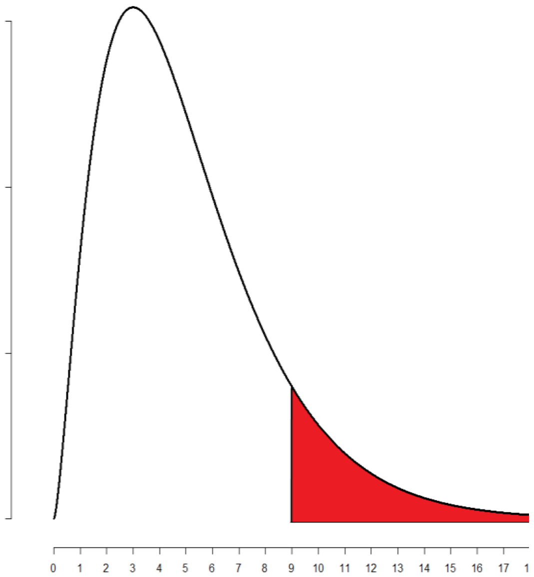


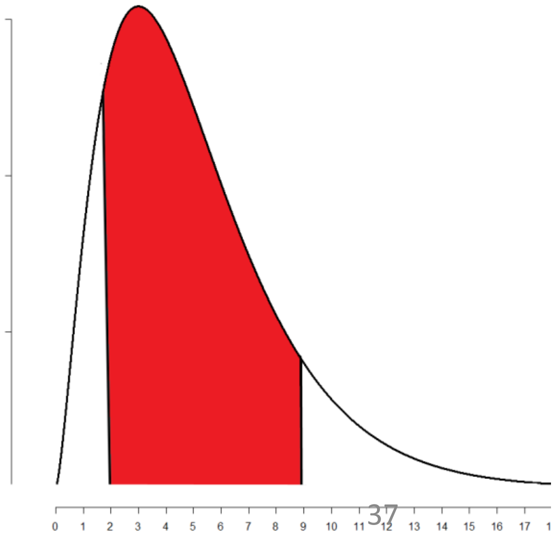




Continuous Distributions







Today

- Random variables

- Probability distributions
  - Discrete variables
    - Three different probability calculations
    - Mean and variability
  - Continuous variables

Two specific distributions:

- Normal distribution (possible for continuous variables)
  - **Mean and variability**
  - Three different probability calculations

- Binomial distribution (possible for discrete variables)
  - General formulae
  - Mean and variability

The normal distribution

"Bell-shaped curve"

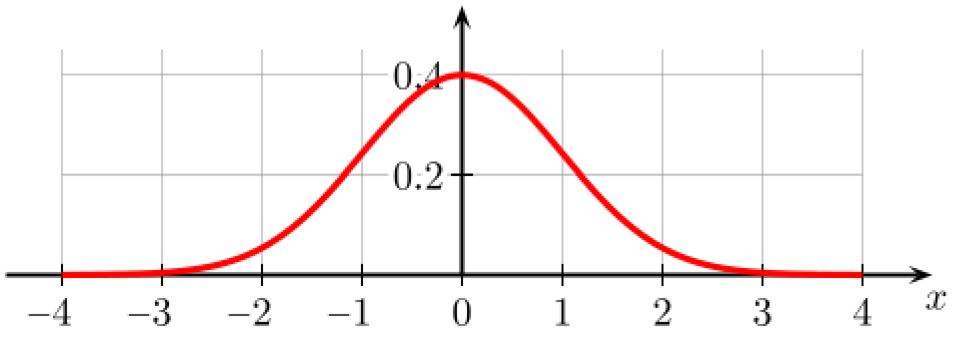


Photo source: https://commons.wikimedia.org/wiki/File:Normal_distribution.svg

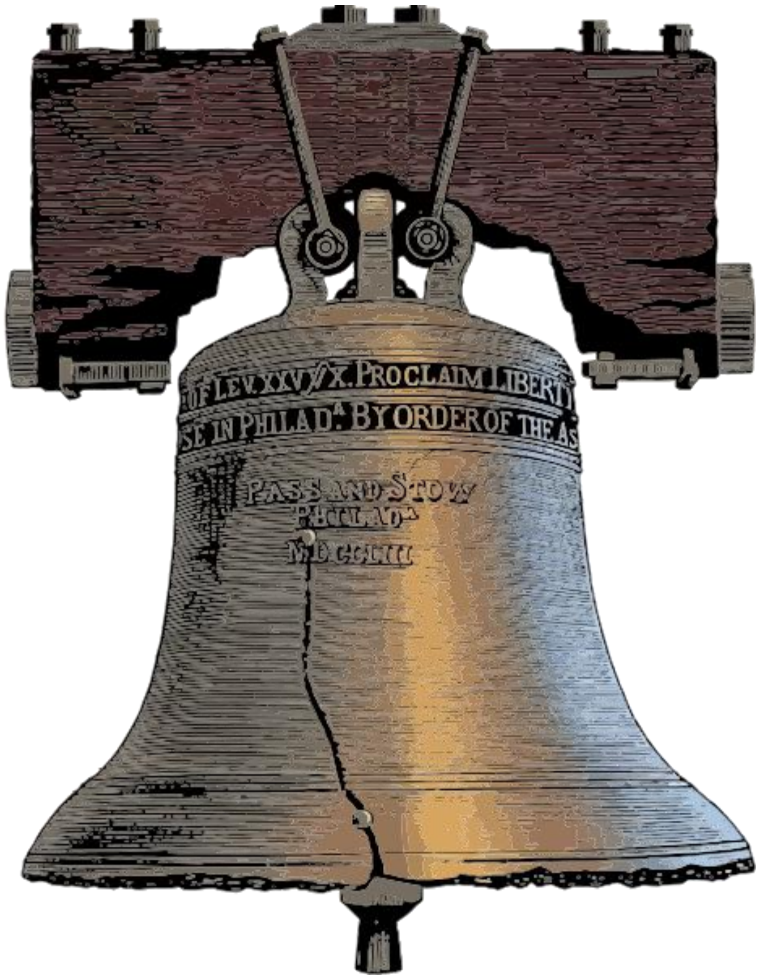
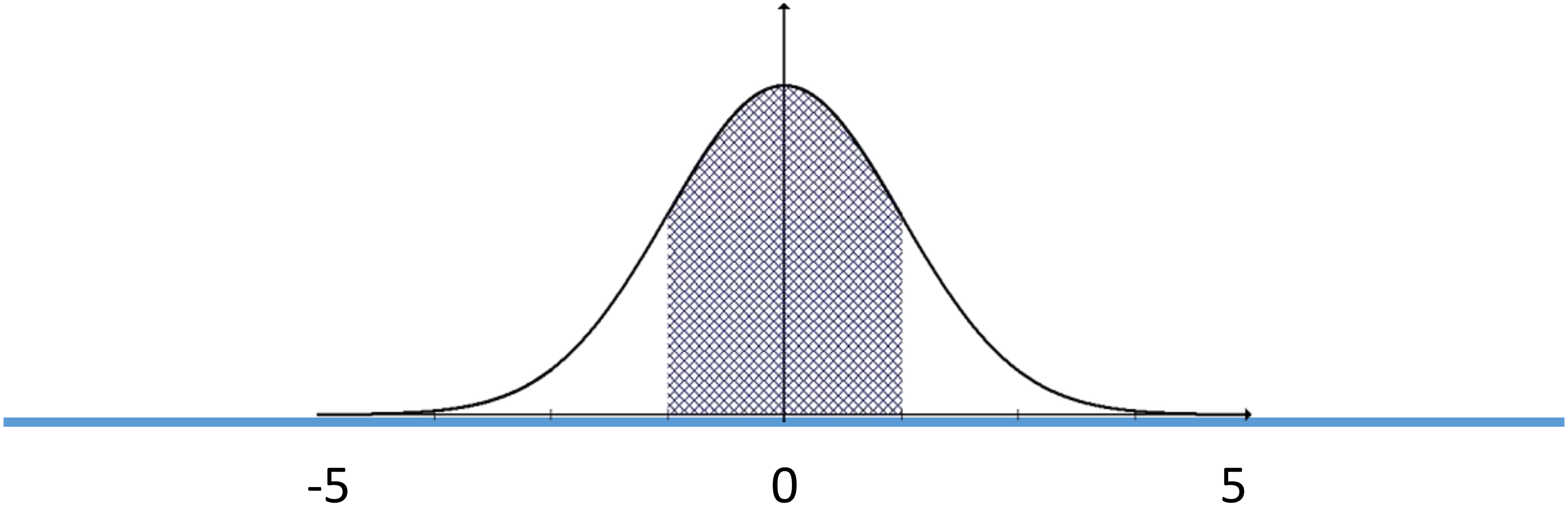
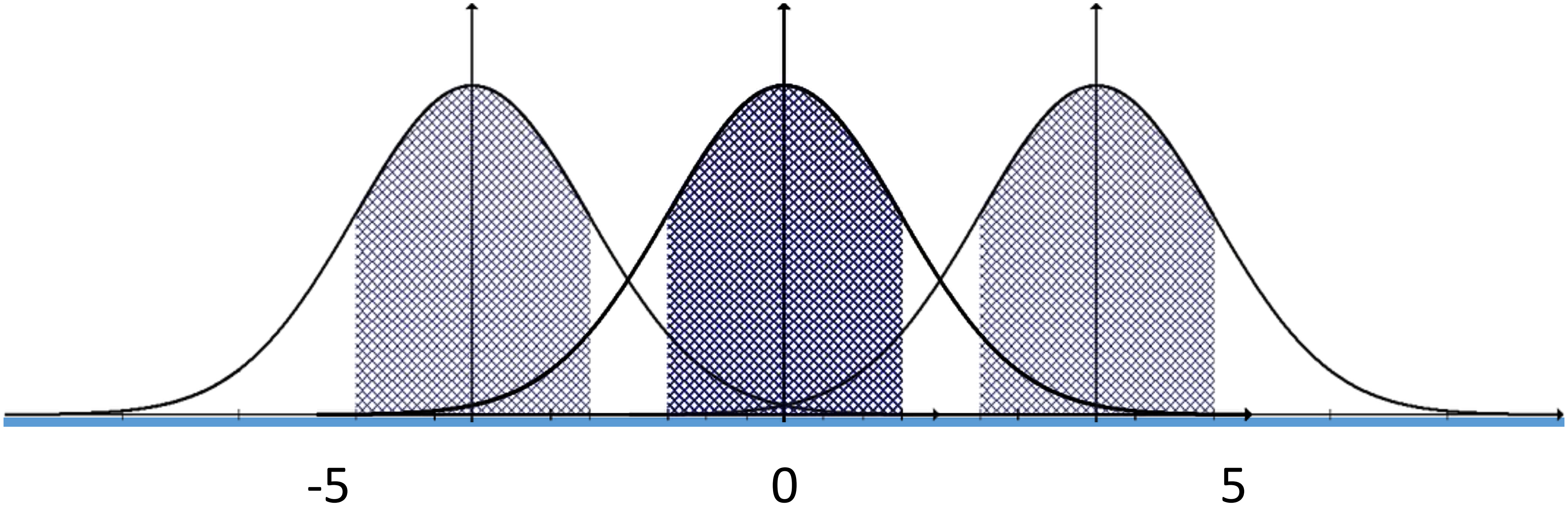
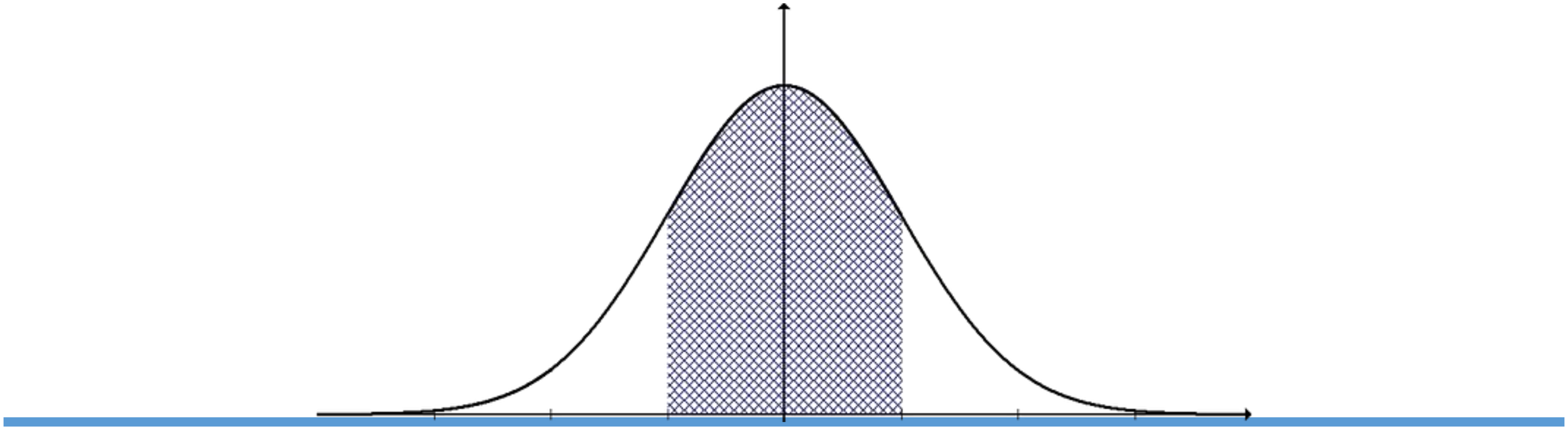


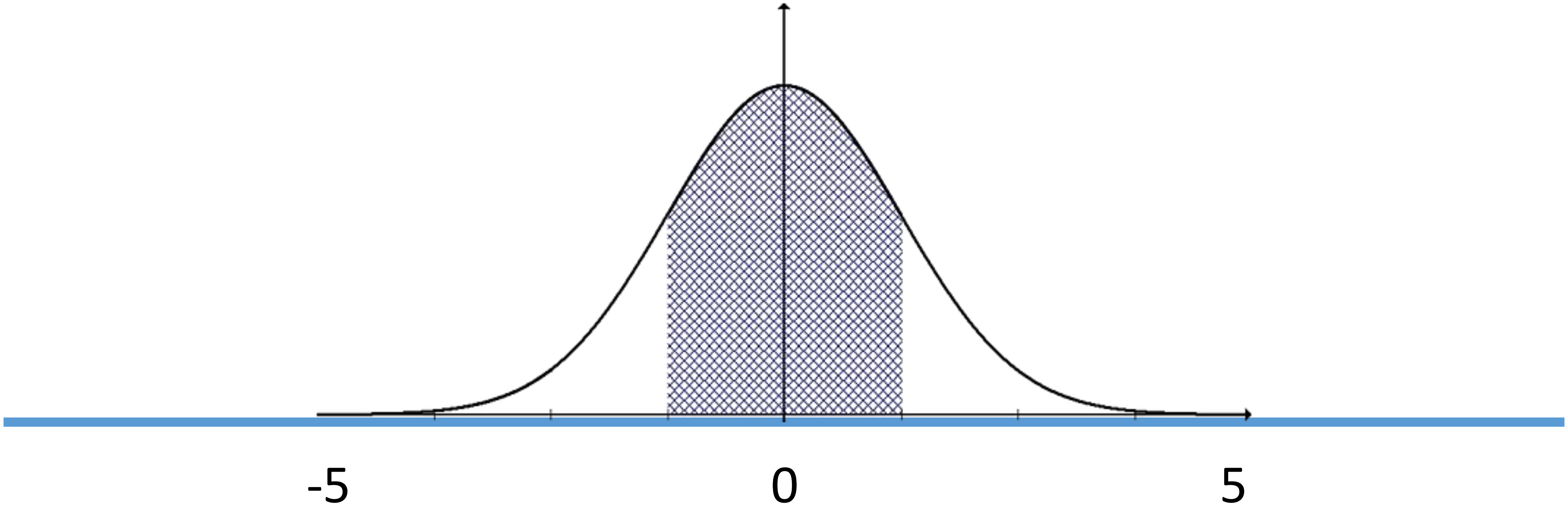
Photo source: https://commons.wikimedia.org/wiki/File:Liberty_Bell_icon.svg

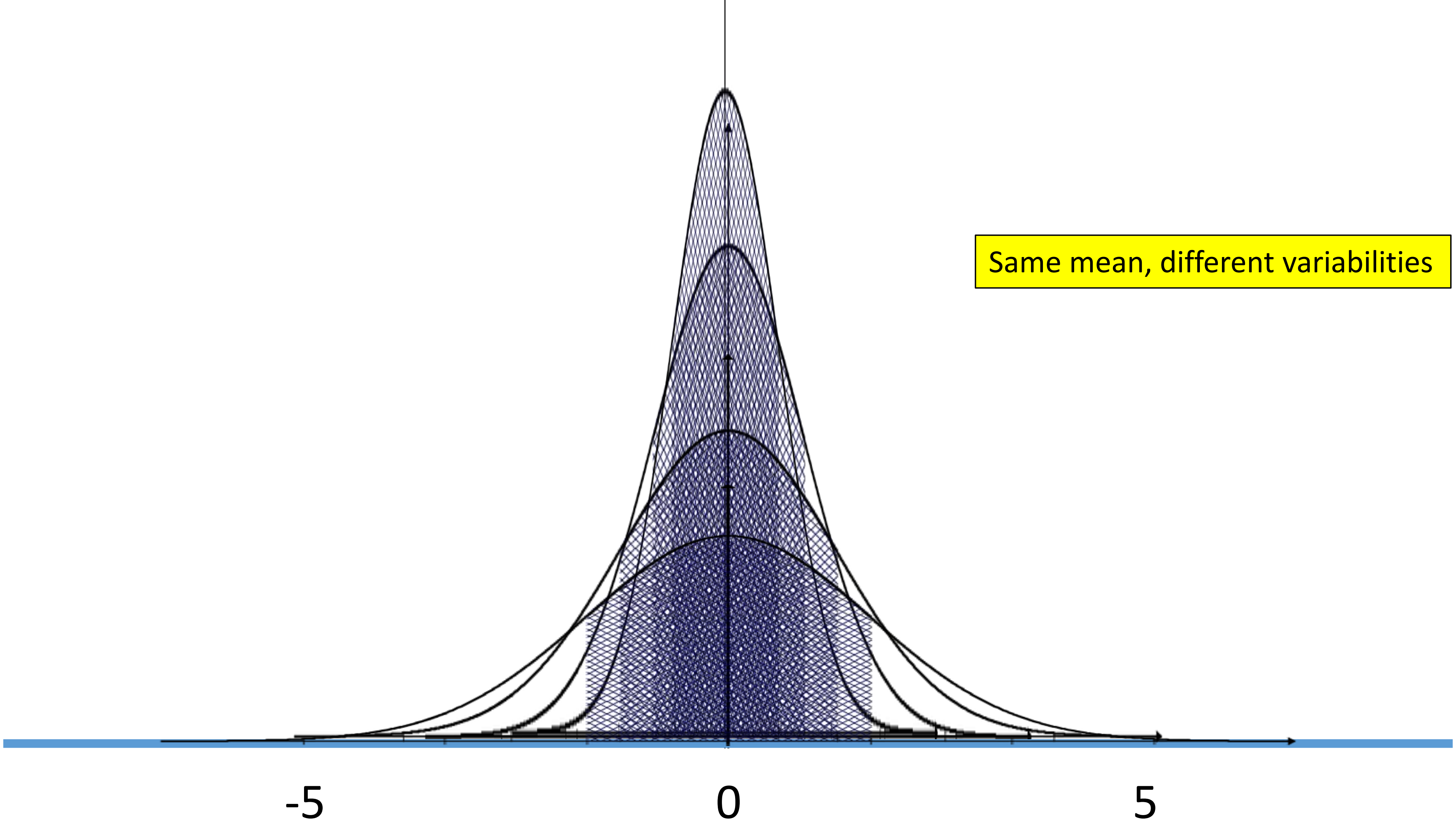

-5
0
5


-5
0
5

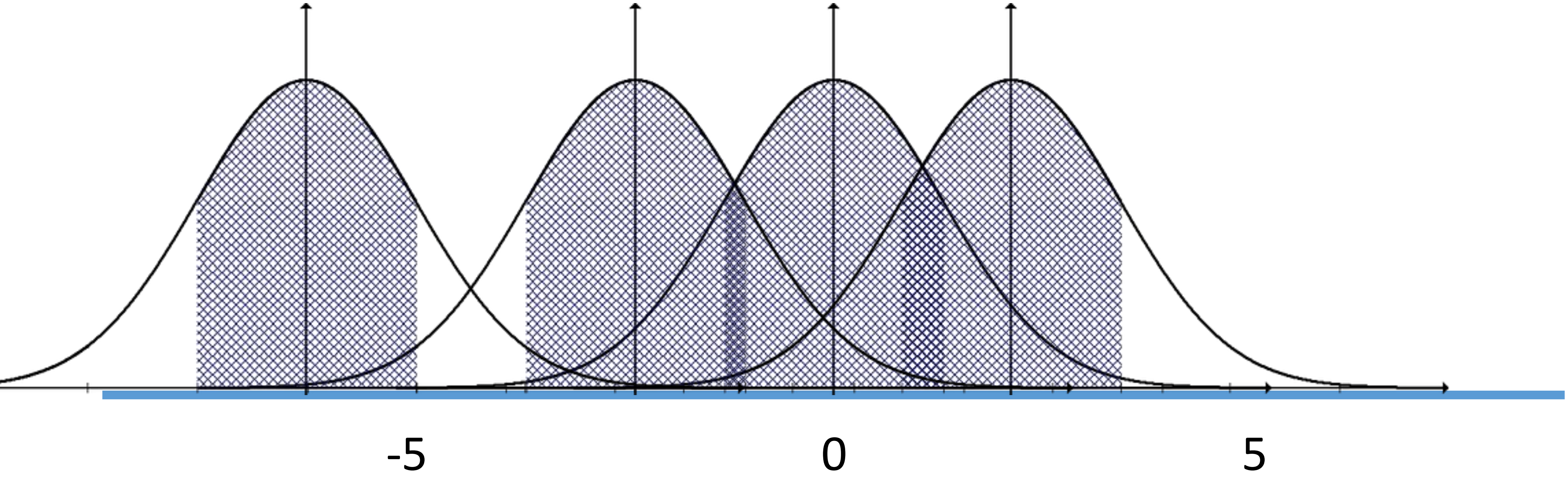


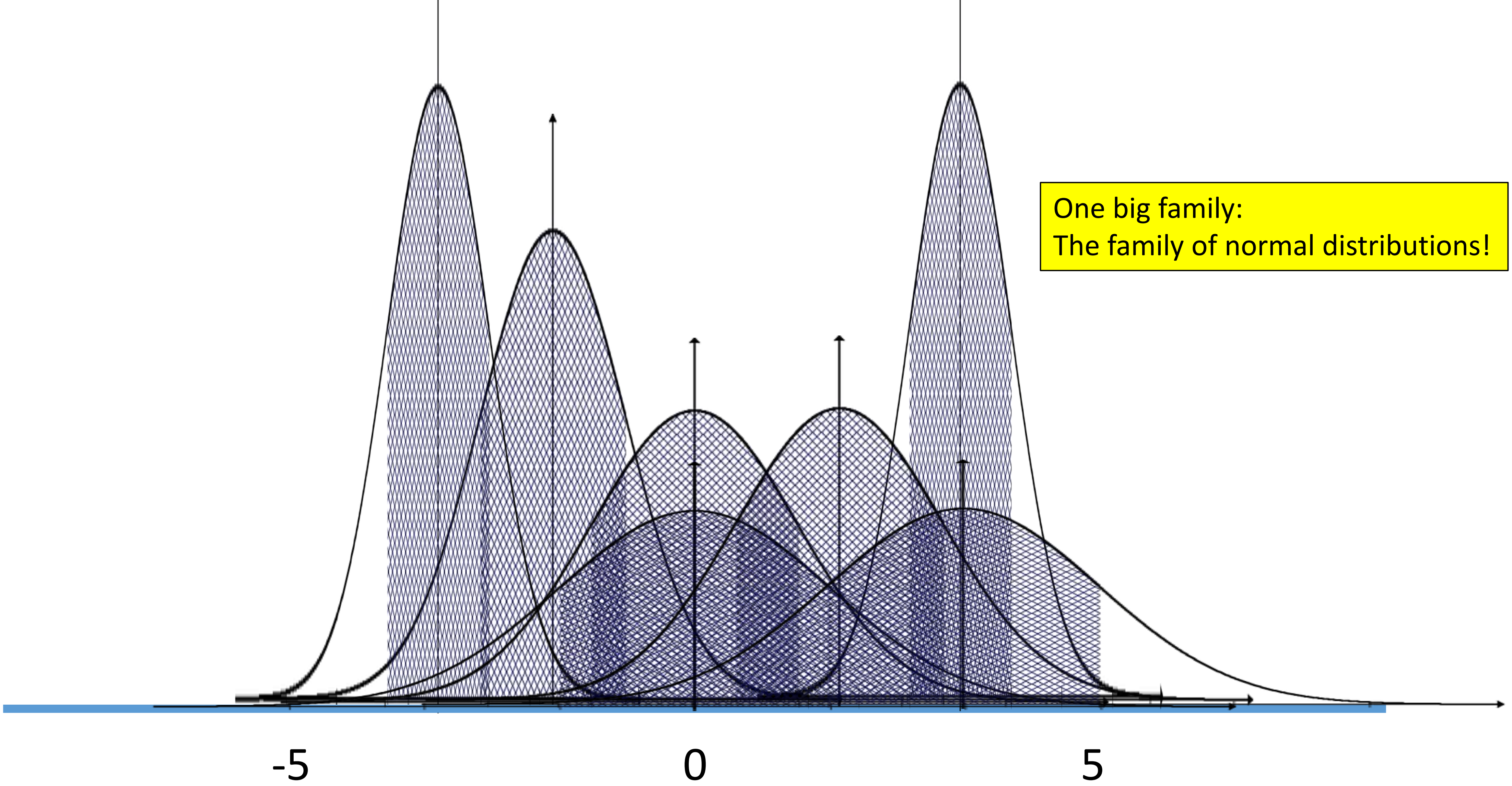
-5 0 5


-5
0
5


Same mean, different variabilities
-5
0
5

Same variability, different means

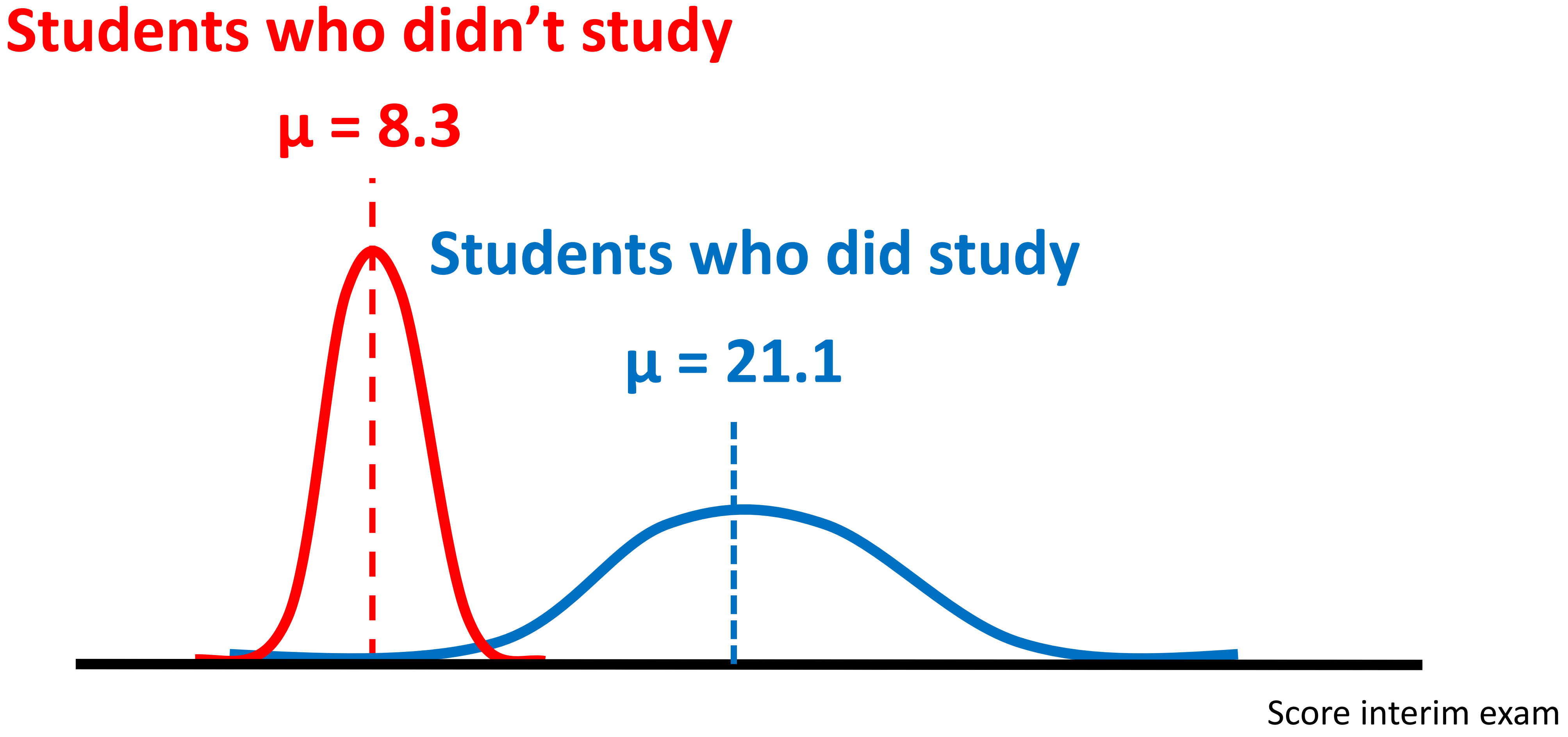




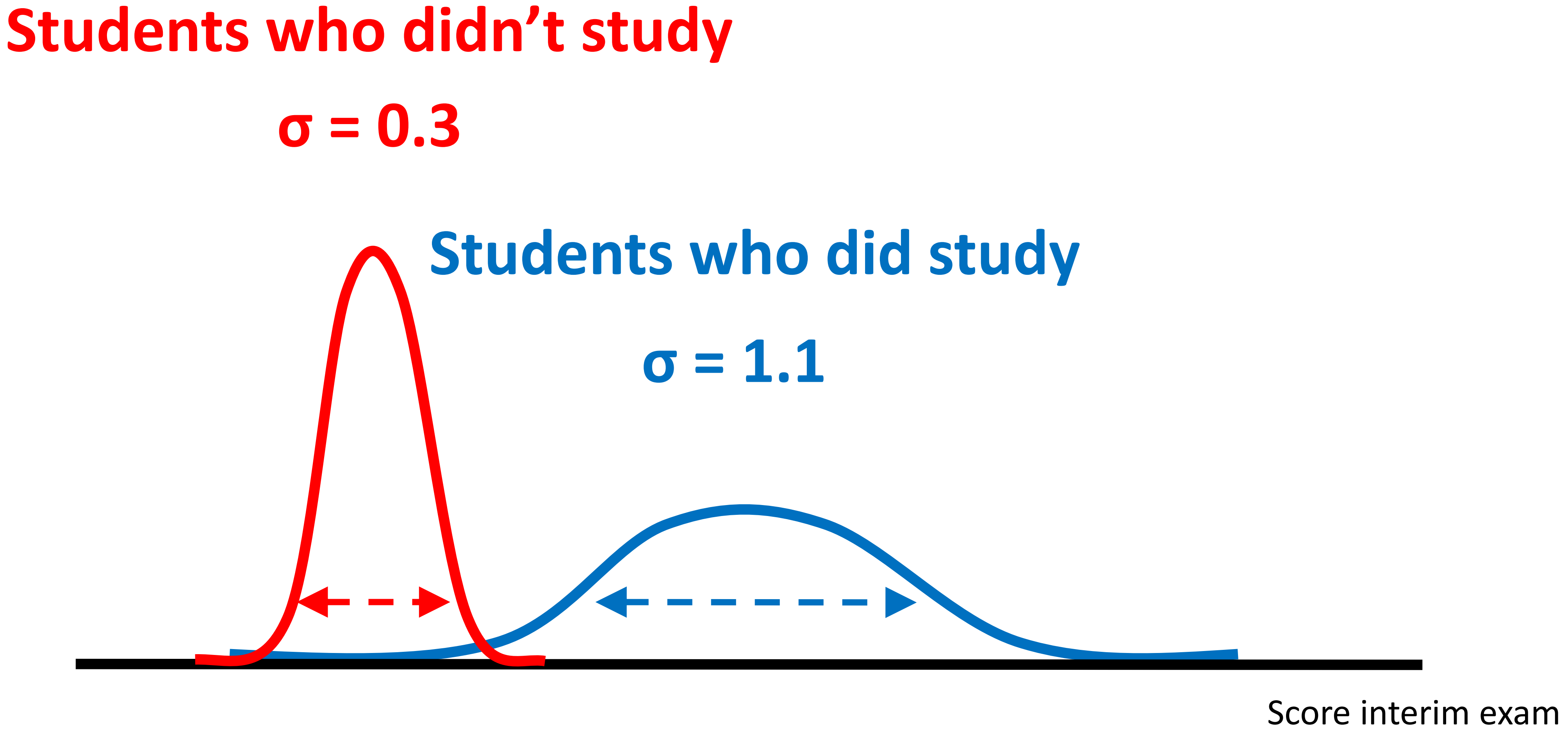
One big family:
The family of normal distributions!

The normal distribution is characterized by 2 parameters: the mean and Standard Deviation. This means that if you know the values of these two parameters (e.g., mean = 2 and SD = 1) then you identify one single curve from the family.

The normal distribution: mean μ



The normal distribution: standard deviations σ



Today

- Random variables

- Probability distributions
  - Discrete variables
    - Three different probability calculations
    - Mean and variability
  - Continuous variables

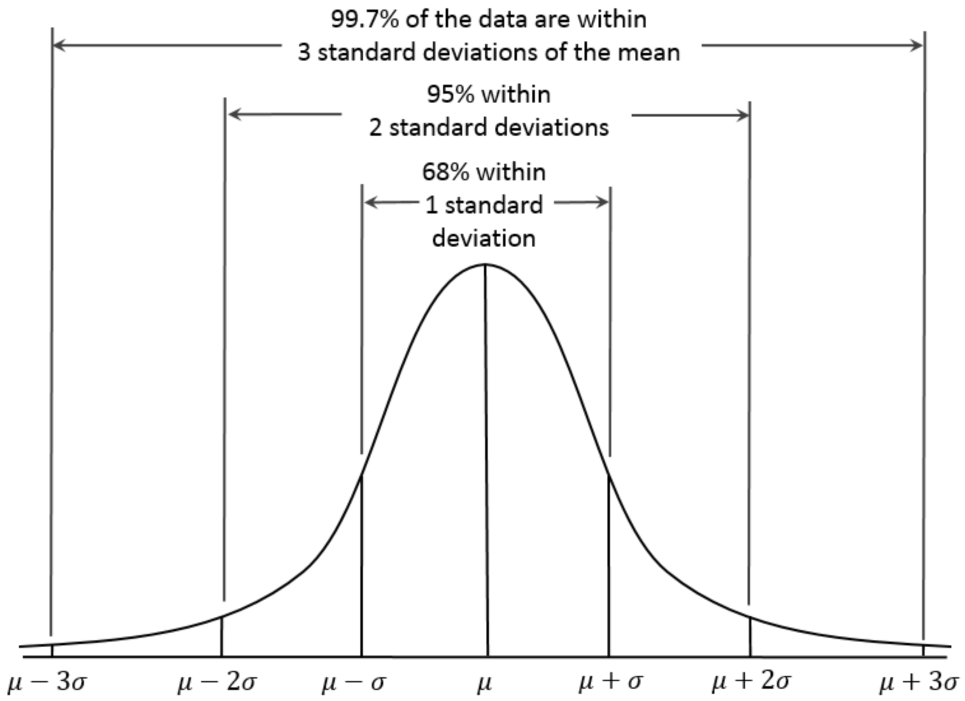
- Specific distribution for continuous variables: Normal distribution
  - Mean and variability
  - **Three different probability calculations**

- Specific distribution for discrete variables: Binomial distribution
  - General formulae
  - Mean and variability

Probability calculation

Empirical rule (section 2.4; section 6.2)

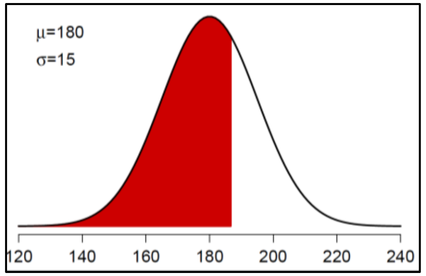
- Can be used for rough probability calculation



- Now we focus on precise calculations

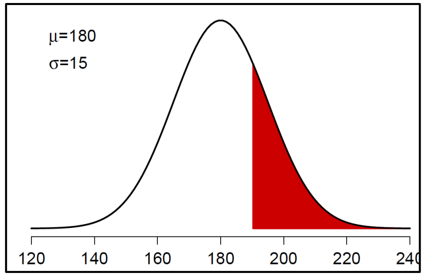
Photo source: https://commons.wikimedia.org/wiki/File:Empirical_Rule.PNG

Three different probability calculations

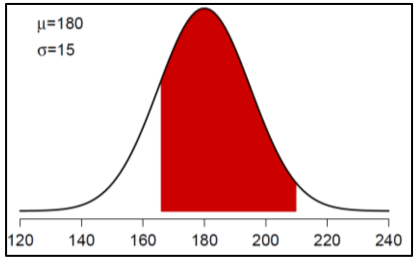


1) Probability of height smaller than 187.5 cm
- $P(X < 187.5)$

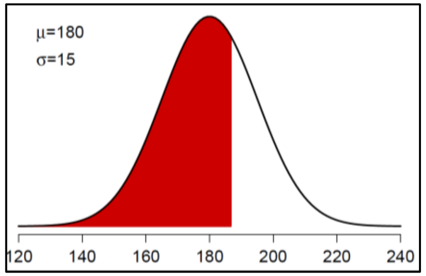
In continuous distributions, we use $<$ and $>$ (not $\leq$ and $\geq$) because the result is the same e.g., $P(X<190) = P(X\leq190)$. This is not the case for discrete variables!



2) Probability of height larger than 190 cm
- $P(X > 190)$



3) Probability of height between 165 cm and 210 cm
- $P(165 < X < 210)$



Situation I: Lower tail probability

Step 1: Find *z*

- In Chapter 2 (and lecture 4 by Johnny)

$$z = \frac{\text{observation} - \text{mean}}{\text{standard deviation}}$$

→z = the number of standard deviations that a given observation falls from the mean

- If we apply this to the normal distribution, we get:

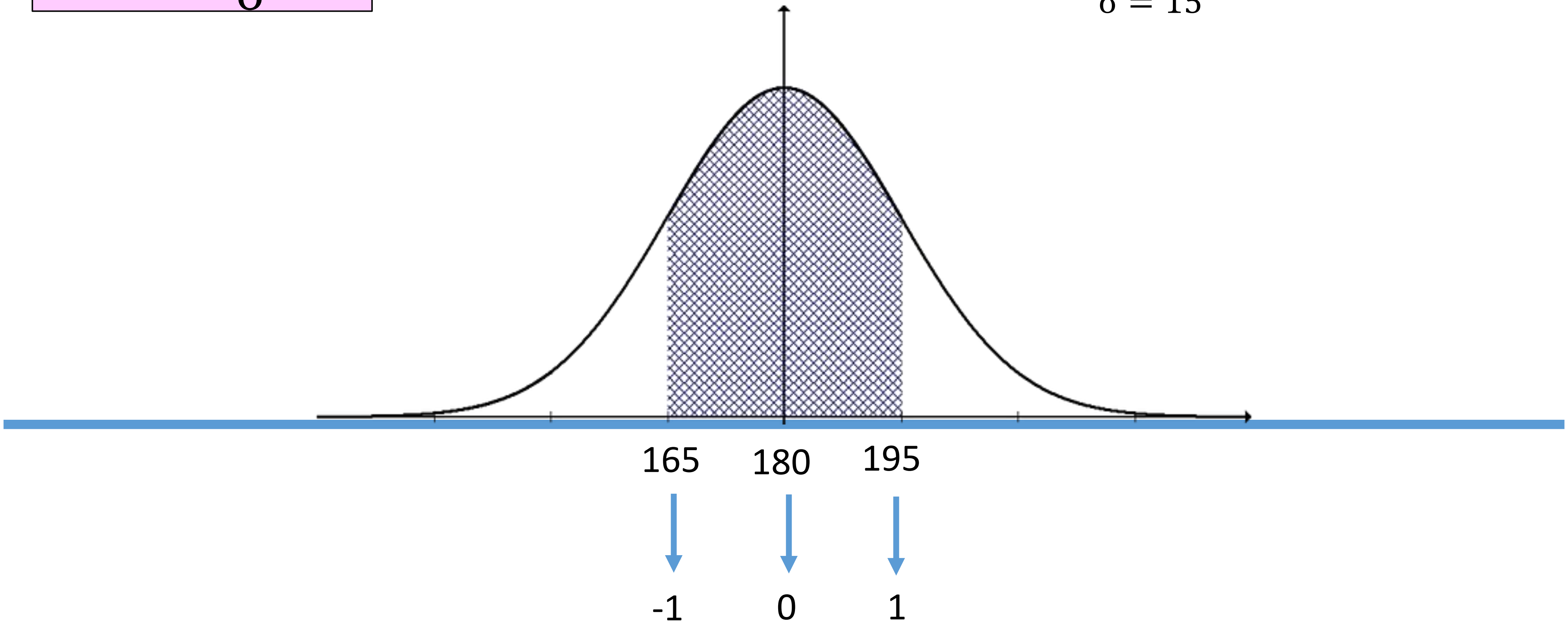
$$z = \frac{\text{x} - \mu}{\sigma}$$

where x is an observation (e.g., someone’s IQ score, someone’s depression score, etc)

Step 1: Find *z*

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

$\mu = 180$
$\sigma = 15$

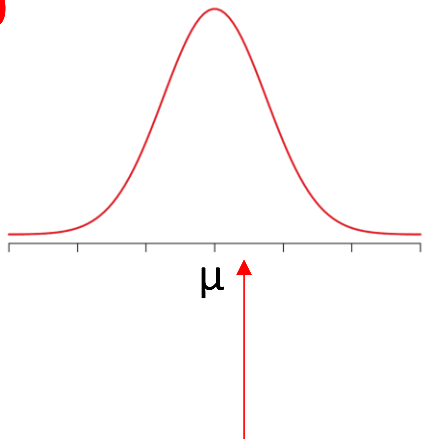


Step 1: Find *z*

The question was: what is $P(X < 187.5)$?

Original variable (e.g., depression score)

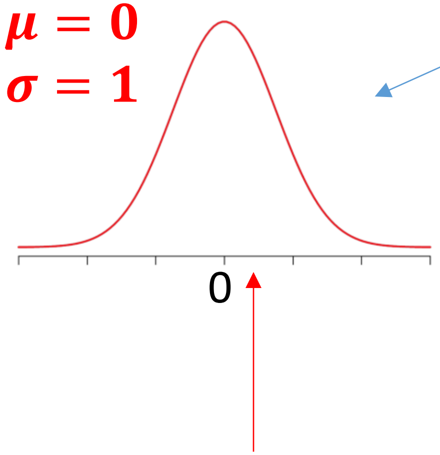
$\boldsymbol{\mu = 180}$
$\boldsymbol{\sigma = 15}$



$$x = 187.5$$

Corresponding z-score

$\boldsymbol{\mu = 0}$
$\boldsymbol{\sigma = 1}$

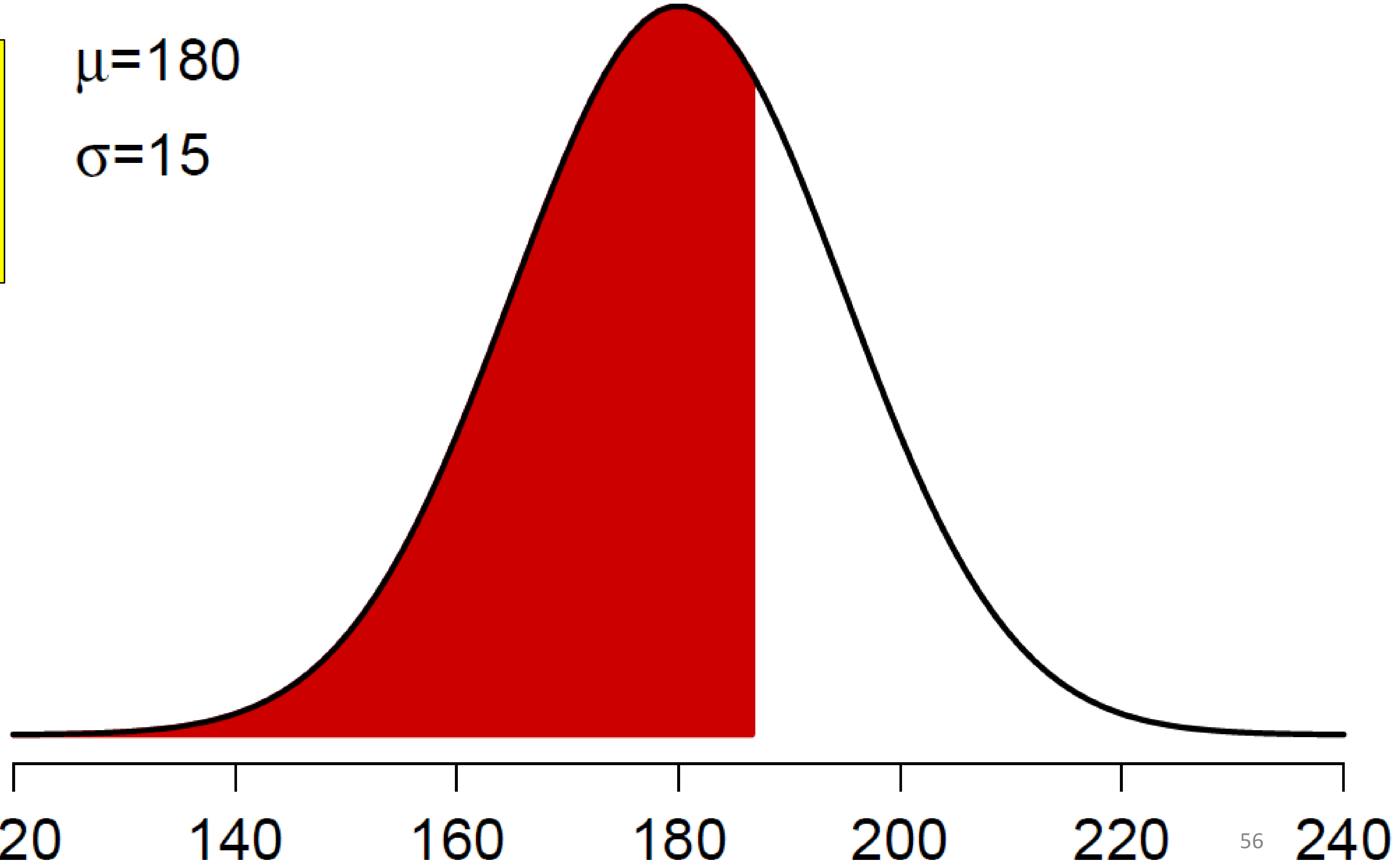


$$z = \frac{187.5 - 180}{15} = 0.5$$

The normal distribution that is characterized by $\mu = 0$ and $\sigma = 1$ is called the 'standard normal distribution'

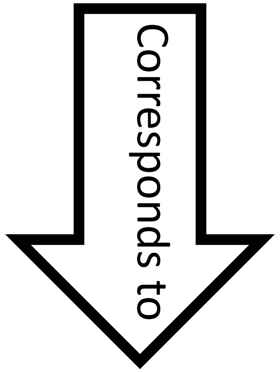
Step 1: Find *z*

Probability of a male with height 187.5 cm or shorter?
i.e., P(X<187.5) ?

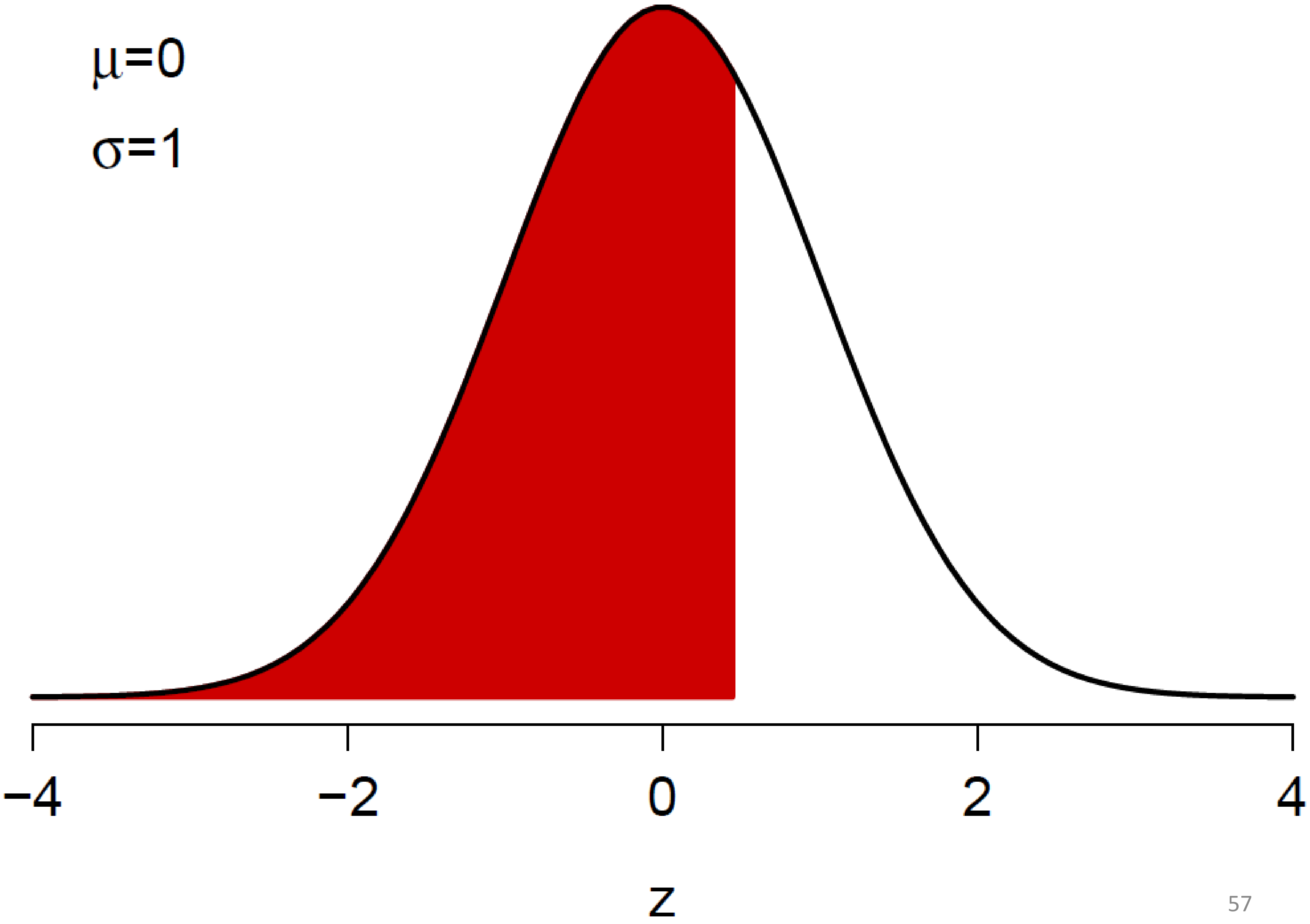


Step 2: Find the probability for z

Probability of a male with height 187.5 cm or shorter?
i.e., P(X<187.5) ?



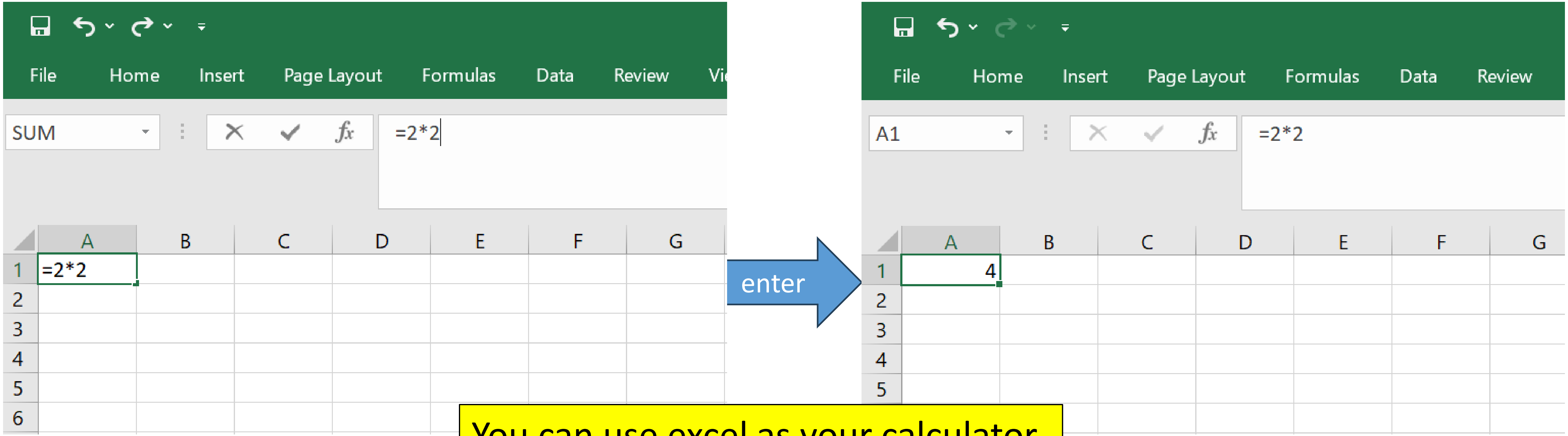
Probability of z smaller than 0.5
i.e., $P(z < 0.5)$



Step 2: Find the probability for z

For more info on the use of Excel see:
Canvas -> Module -> Microsoft Excel
Some Excel commands will also be included in formula sheet

Welcome Excel!!



You can use excel as your calculator

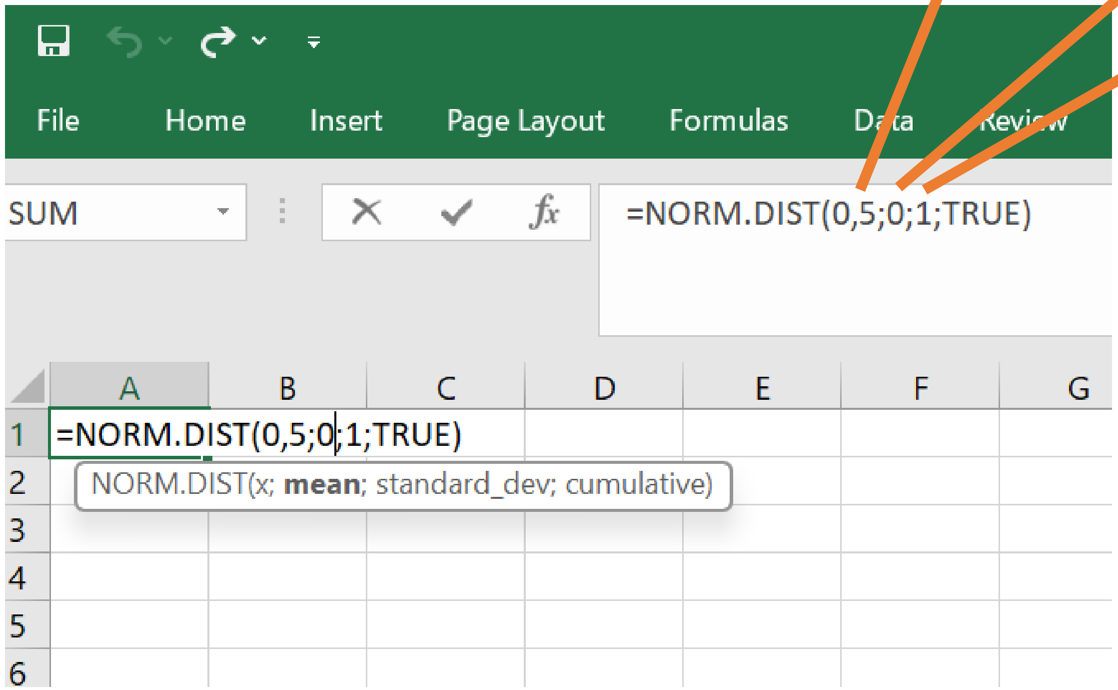
Step 2: Find the probability for z

What we want: Probability of z smaller than 0.5
i.e., P(z < 0.5)

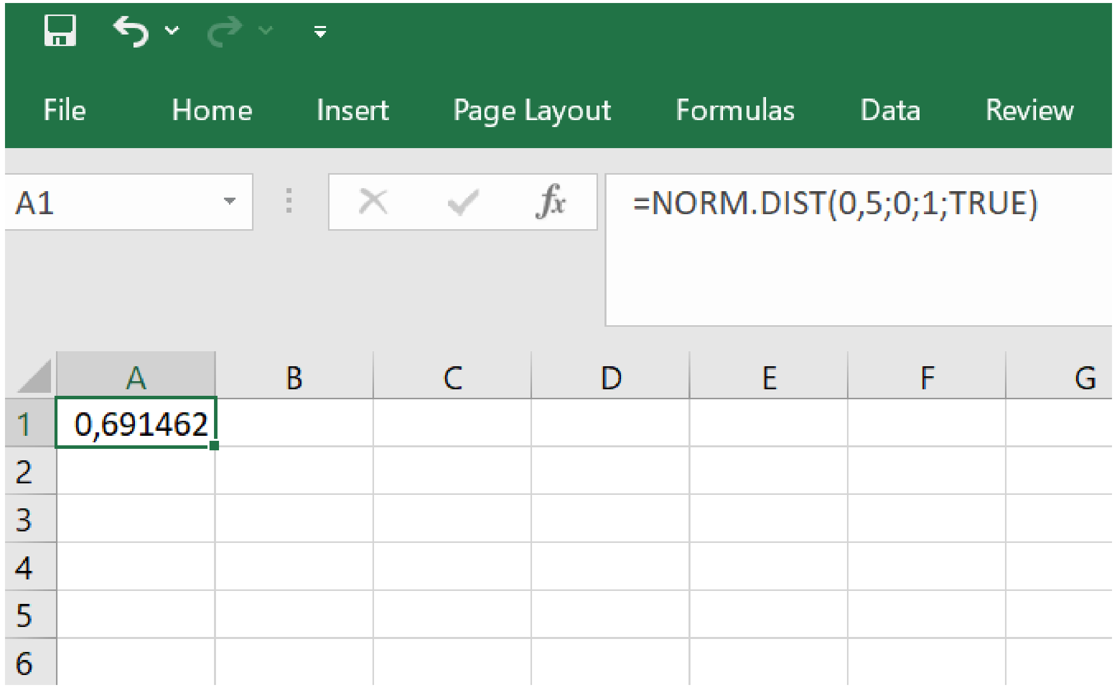
Welcome Excel!!

= NORM.DIST(Z-value ; Mean = 0 ; SD = 1 ; TRUE)

Gives you probability mass LEFT of that z-value (i.e., smaller than z)



enter

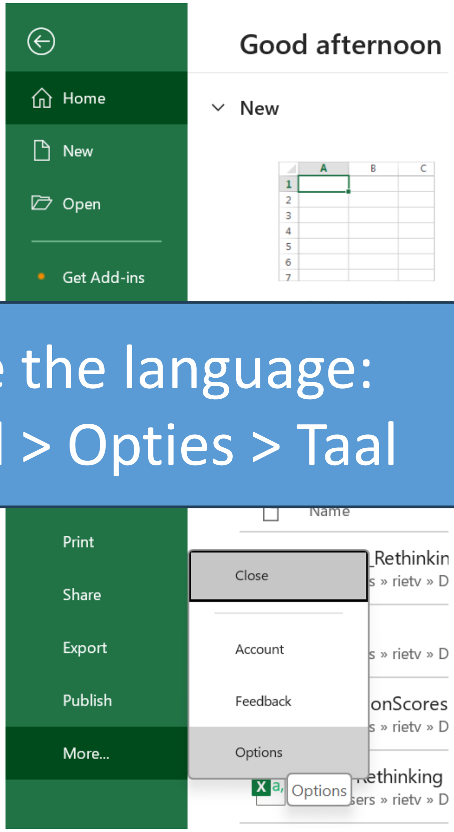


Step 2: Find the probability for z

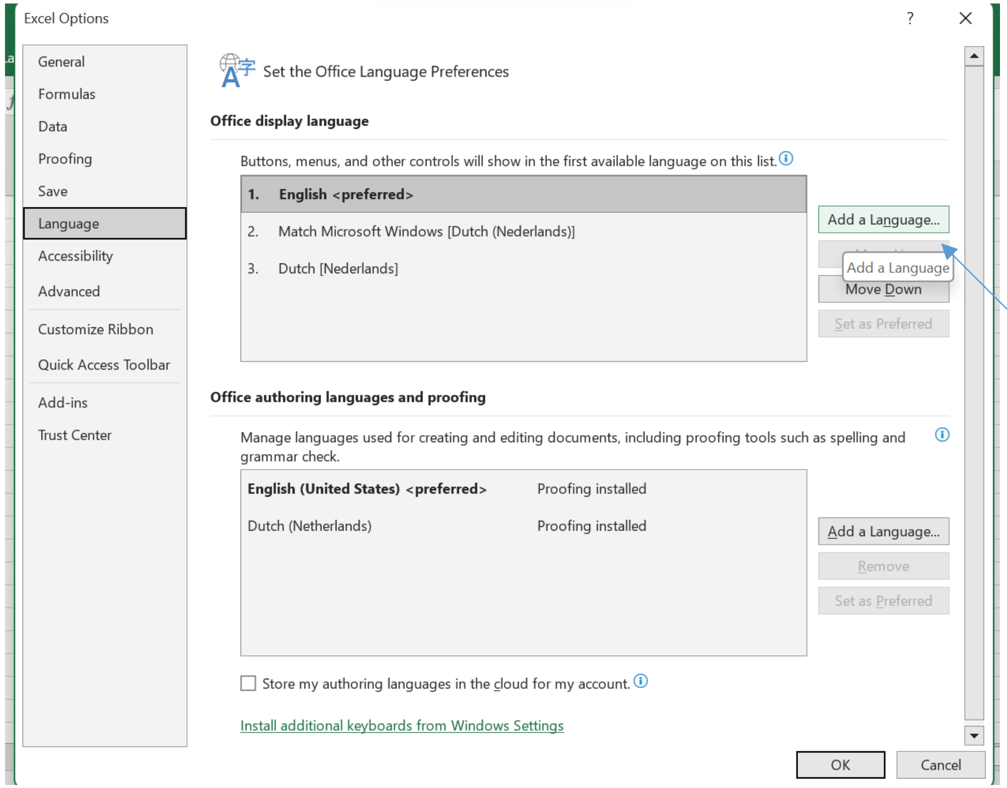
What we want: Probability of z smaller than 0.5
i.e., P(z < 0.5)

Dutch Excel?

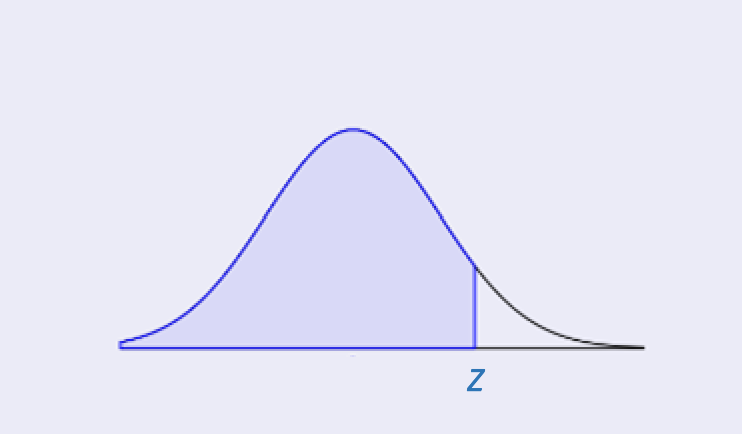
= NORM.VERD(Z-value ; Mean = 0 ; SD = 1 ; WAAR)



change the language:
Bestand > Opties > Taal



Step 2: Find the probability for z



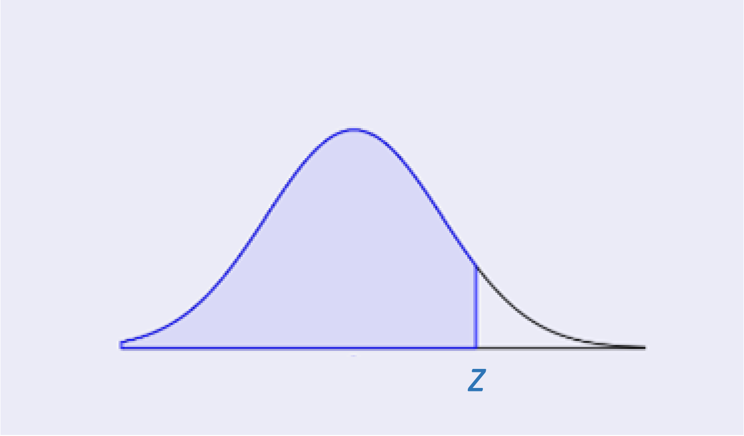
When you are still practicing with excel,
you can check your answers with Appendix A of the book:

	.00	.01	.02	.03	.04	...
...	...	...	...	...	...	...
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	...
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	...
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	...
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	...
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	...
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	...
...	...	...	...	...	...	...

Probability of z
smaller than 0.5
i.e., $P(z < 0.5)$

So: $P(z < 0.5) = 0.6915 = P(height < 178.5)$

Step 2: Find the probability for z



When you are still practicing with excel,
you can check your answers with Appendix A of the book:

	.00	.01	.02	.03	.04	. . .
. . .	. . .	. . .	. . .	. . .	. . .	. . .
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	. . .
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	. . .
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	. . .
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	. . .
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	. . .
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	. . .
. . .	. . .	. . .	. . .	. . .	. . .	. . .

Probability of z smaller than 0.53 i.e., P(z < 0.53)?

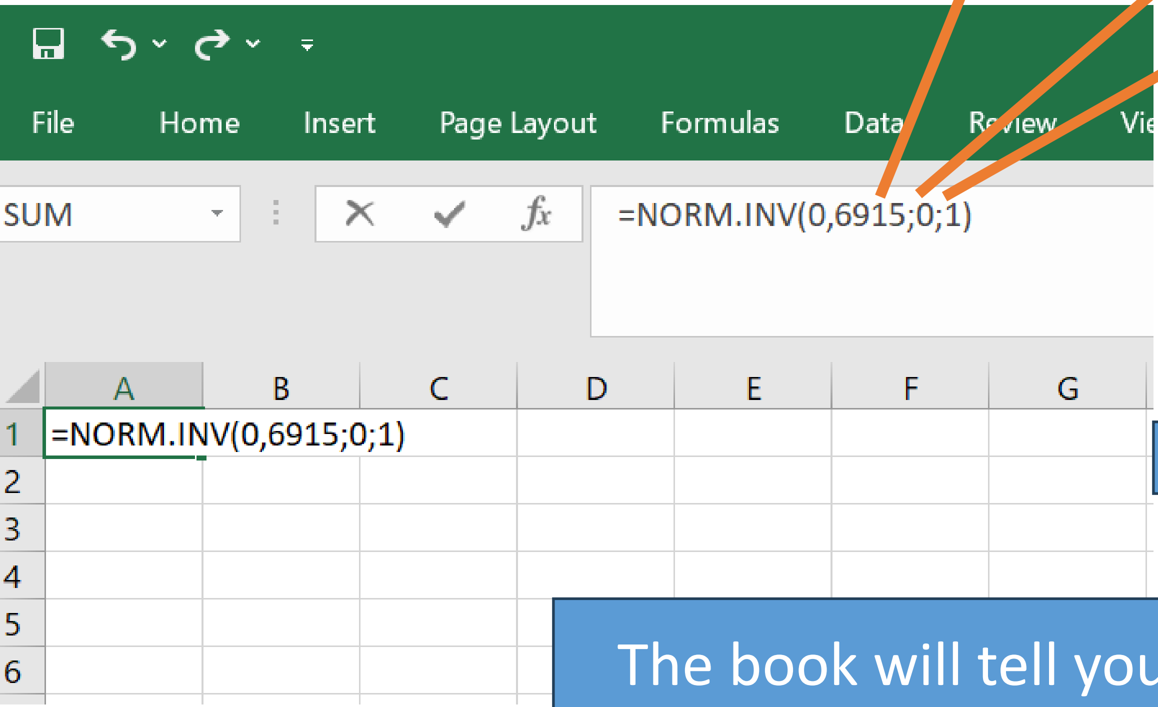
So: P(z < 0.53) = 0.7019

Inverse: Find the z for a probability

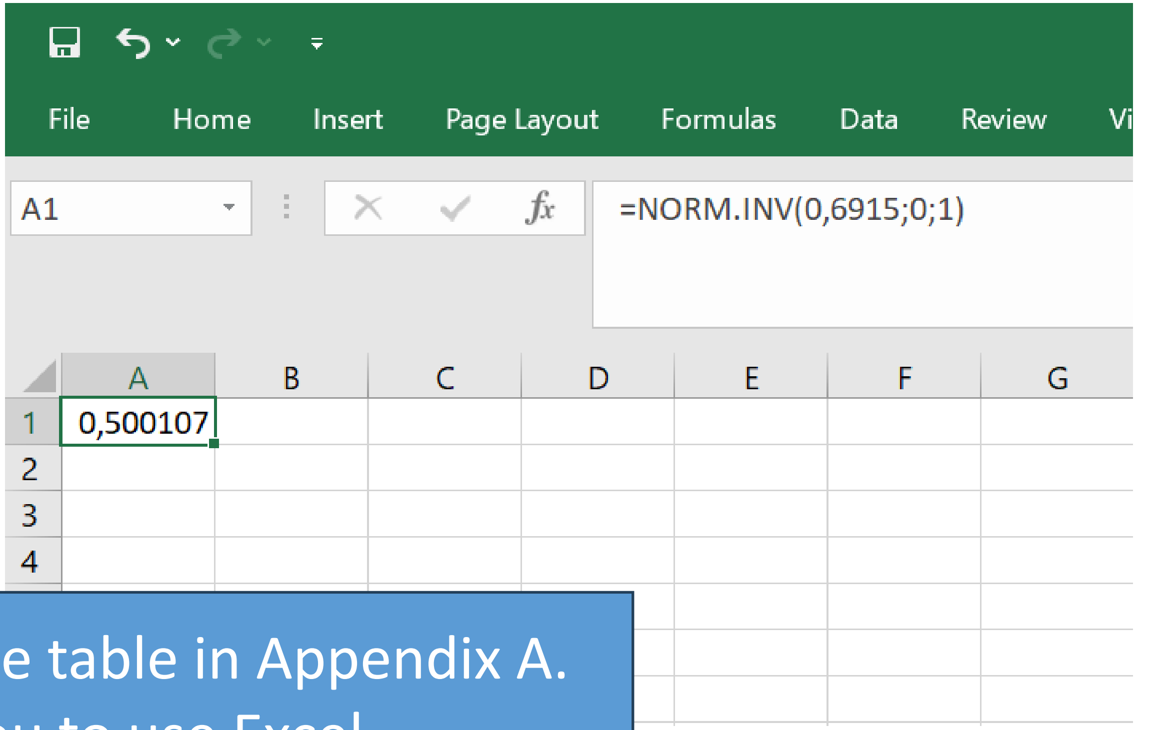
Suppose we want to know: What value for z has a lower-tail probability of .6915 ?

Welcome Excel!!

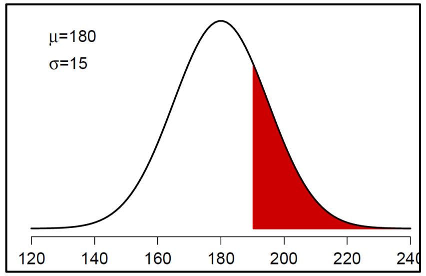
= NORM.INV(probability ; Mean = 0 ; SD = 1)



enter



The book will tell you to use the table in Appendix A. However, we expect you to use Excel

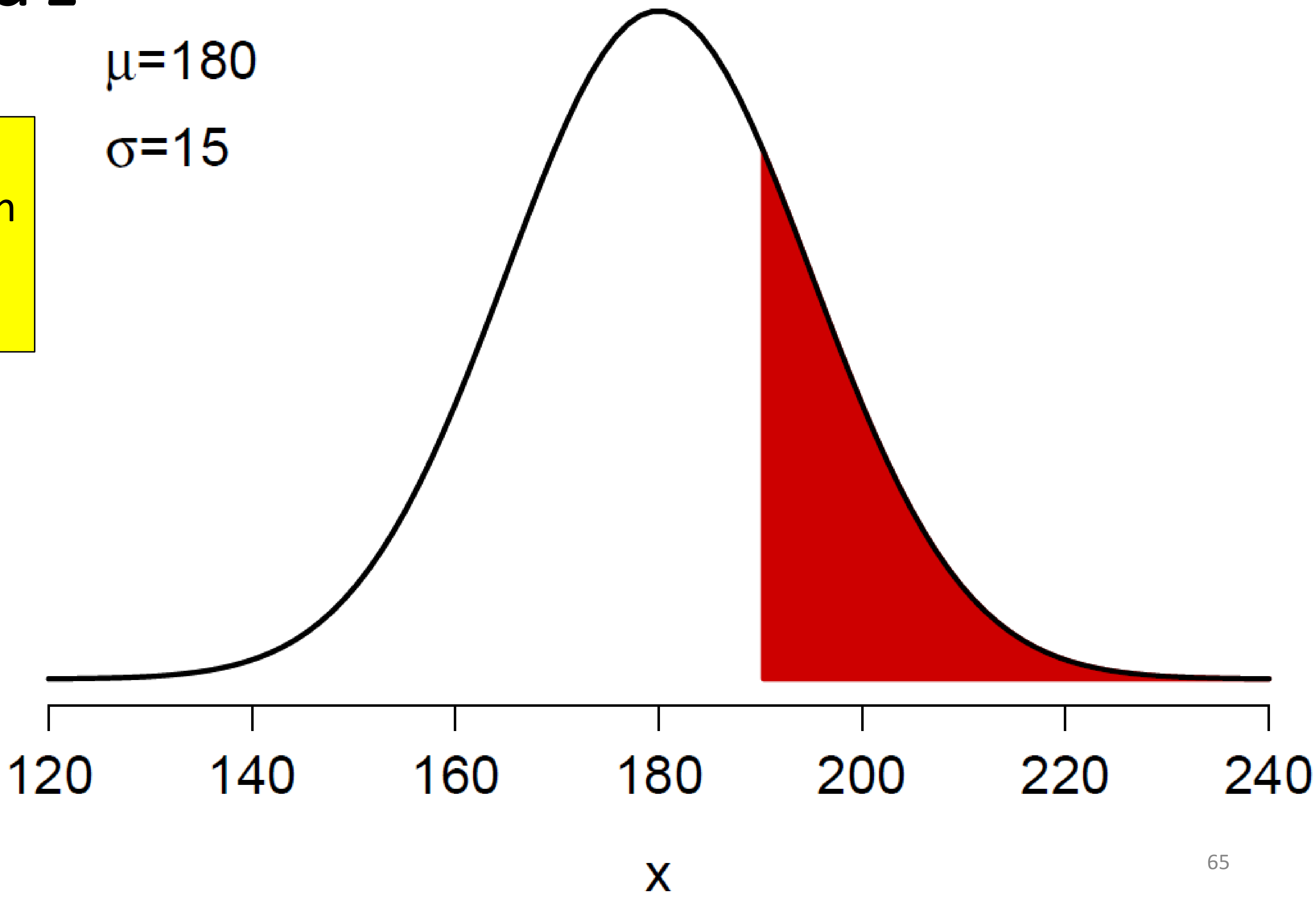


Situation II: Upper tail probability

Step 1: Find *z*

Probability of a male with height larger than 190 cm ?
i.e., P(X>190) ?

$$z = \frac{190-180}{15} = 0.67$$



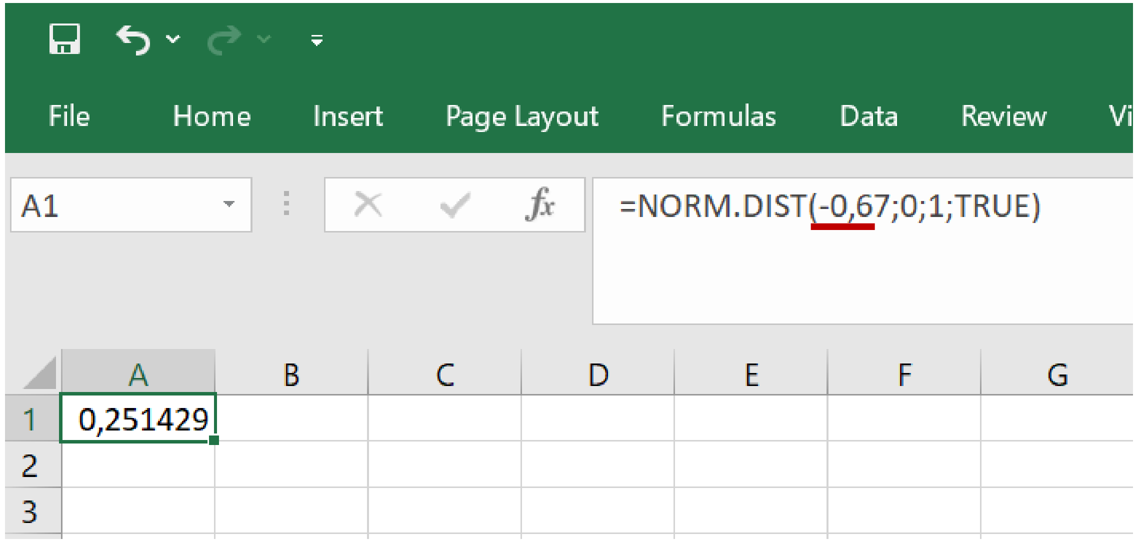
Step 2: Find the probability for z

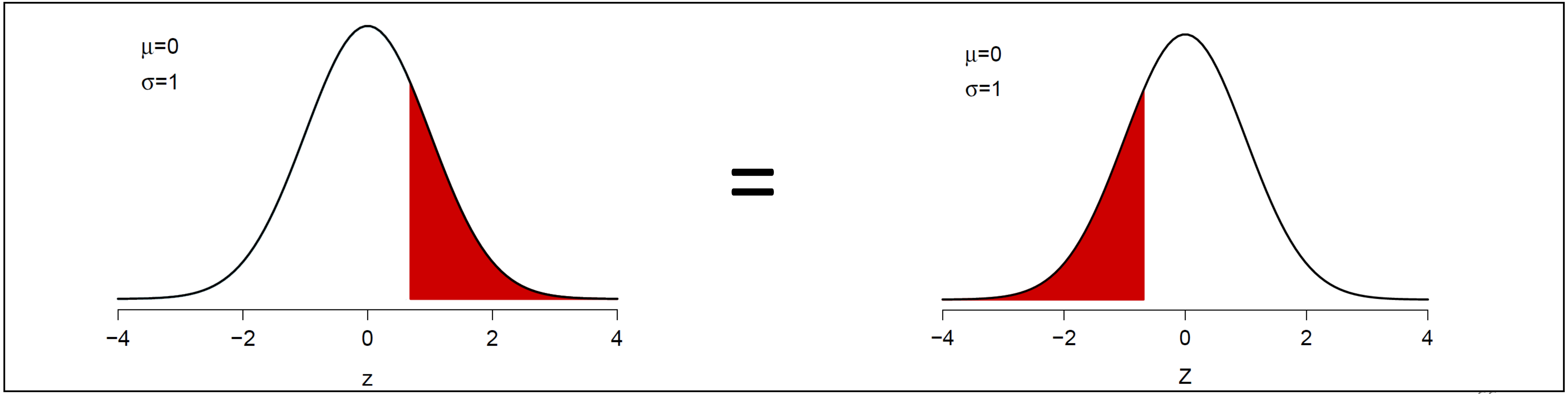
Two approaches:

A) Symmetry:

- P(z > 0.67) = P (z < -0.67)

So: P(z > 0.67) = P(z < -0.67) = 0.2514





Step 2: Find the probability for z

Two approaches:

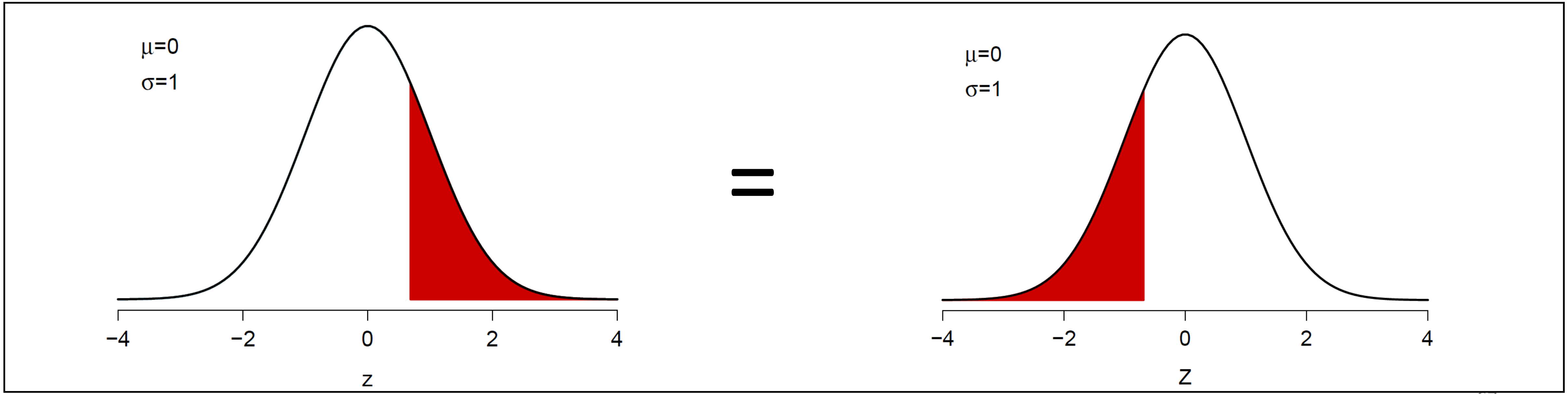
A) Symmetry:

- P(z > 0.67) = P (z < -0.67)

So: P(z > 0.67) = P(z < -0.67) = 0.2514

You can double check in Appendix A

	...	0.05	0.06	0.07	0.08	0.09
...	...	...	...		...	...
-0.8	...	0.1977	0.1949	[illegible]	0.1894	0.1867
-0.7	...	0.2266	0.2236	[illegible]	0.2177	0.2148
-0.6		0.2578	[illegible]	0.2514	0.2483	0.2451
-0.5	...	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	...	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	...	0.3632	0.3594	0.3557	0.3520	0.3483
...	...	...	...	...	...	...



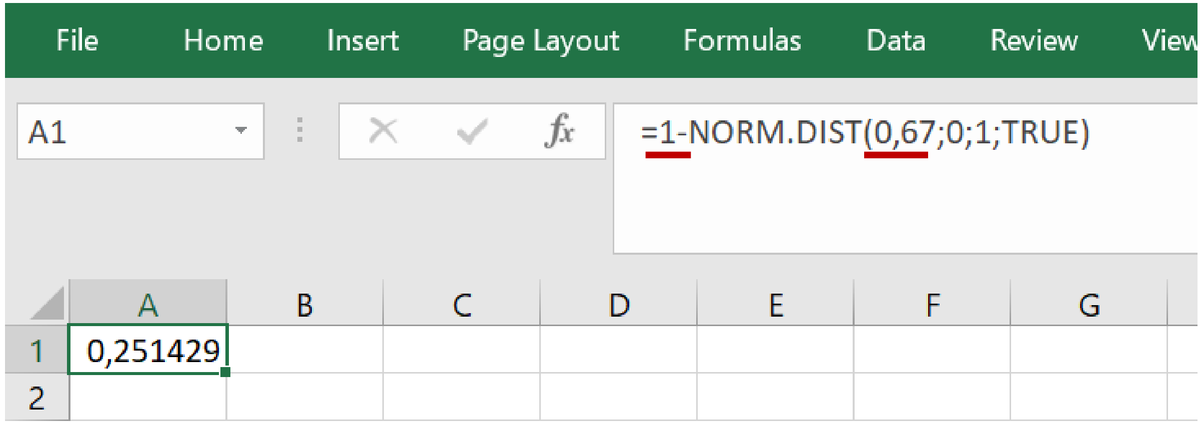
Step 2: Find the probability for z

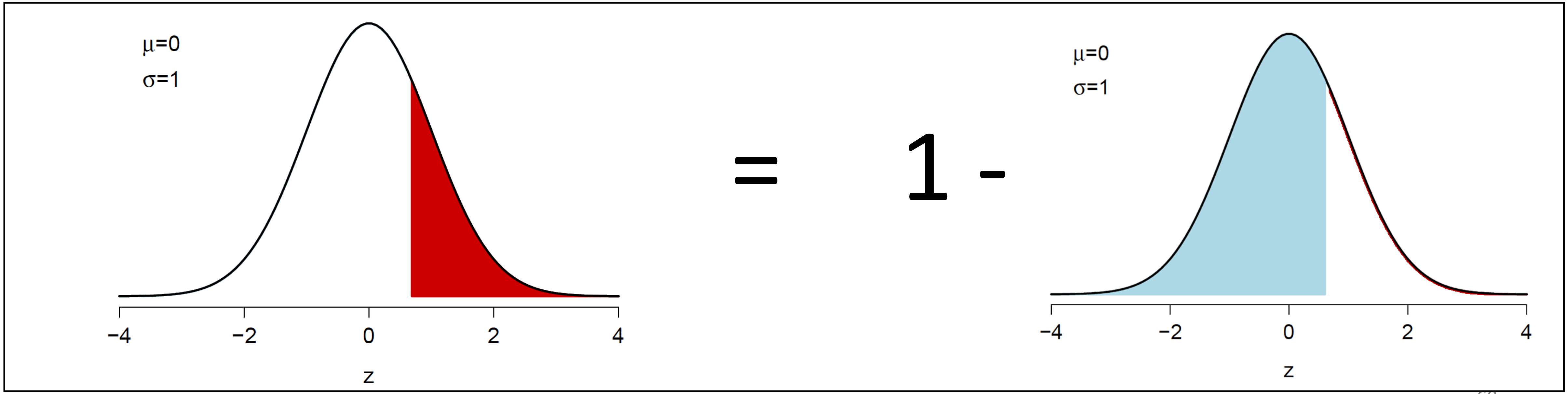
Two approaches:

B) Total probability is always 1:

- P(z > 0.67) = 1 - P (z < 0.67)

So: P(z > 0.67) = 1 - P(z < 0.67) = 1 - 0.7486 = 0.2514





Step 2: Find the probability for z

Two approaches:

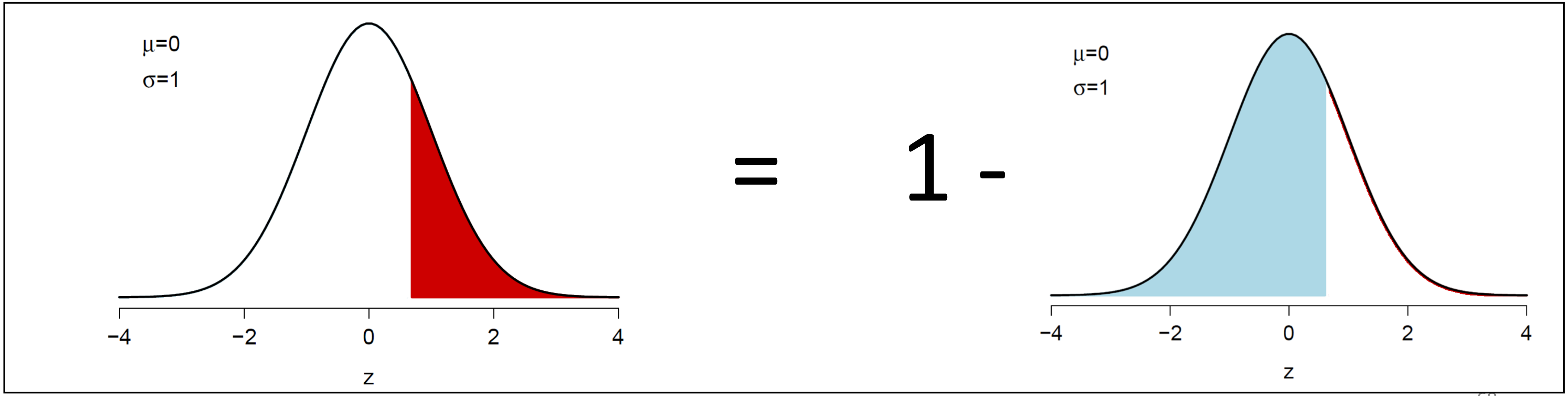
B) Total probability is always 1:

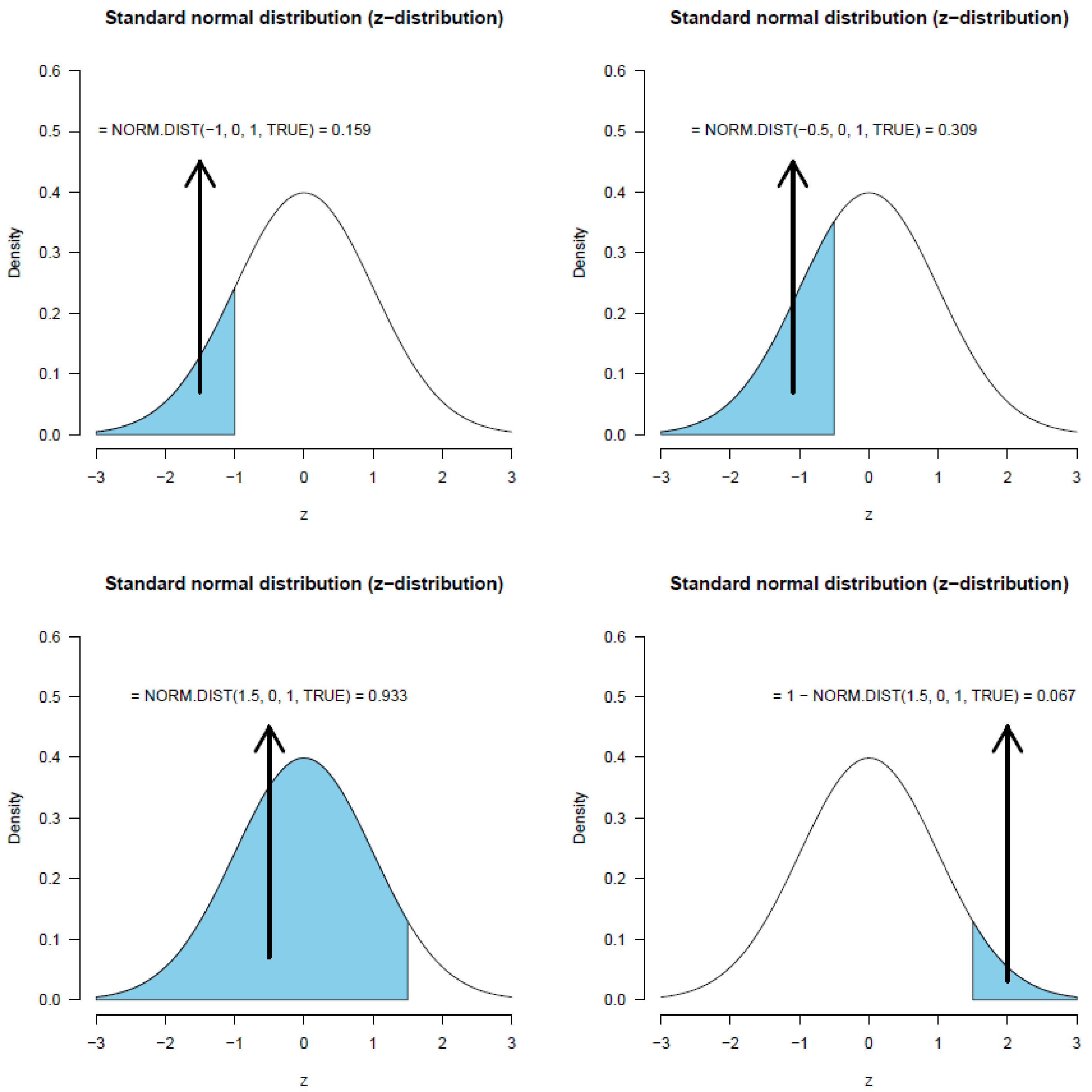
- P(z > 0.67) = 1 - P (z < 0.67)

So: P(z > 0.67) = 1 - P(z < 0.67) = 1 - 0.7486 = 0.2514

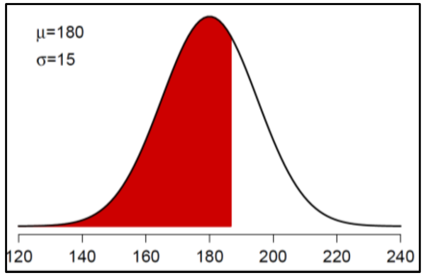
You can check with Appendix A

	...	0.05	0.06	0.07	0.08	0.09
...	...	...	...	...	...	...
0.3	...	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	...	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	...	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	...	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	...	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	...	0.8023	0.8051	0.8078	0.8106	0.8133
...	...	...	...	...	...	...



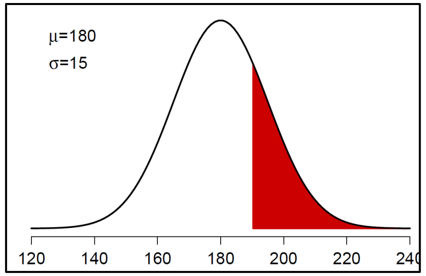

Standard normal distribution (z−distribution)
= NORM.DIST(−1, 0, 1, TRUE) = 0.159
Standard normal distribution (z−distribution)
= NORM.DIST(−0.5, 0, 1, TRUE) = 0.309
Standard normal distribution (z−distribution)
= NORM.DIST(1.5, 0, 1, TRUE) = 0.933
Standard normal distribution (z−distribution)
= 1 − NORM.DIST(1.5, 0, 1, TRUE) = 0.067
Density
z
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−3
−2
−1
0
1
2
3

Three different probability calculations



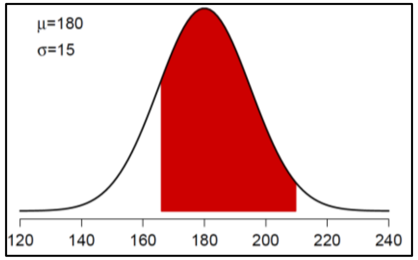
1) Probability of height smaller than 187.5 cm

- P(X < 187.5)



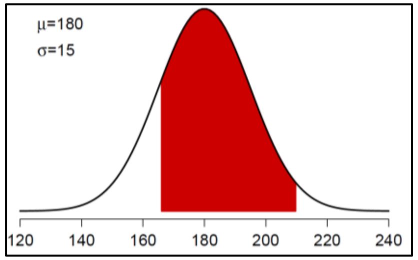
2) Probability of height larger than 190 cm

- P(X > 190)



3) Probability of height between 165 cm and 210 cm

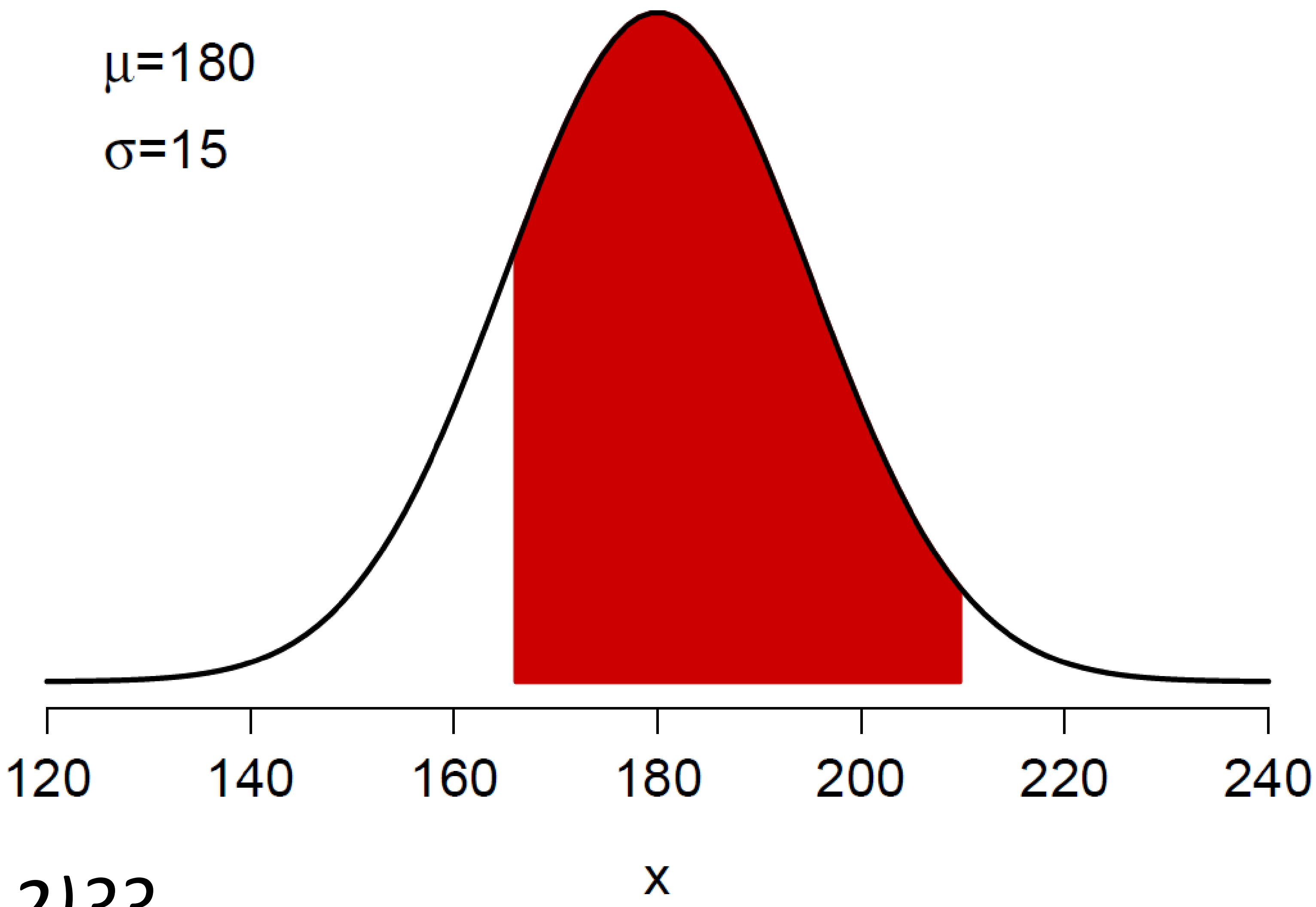
- P(165 < X < 210)



Situation III: Interval probability

Step 1: Find *z*

Probability of a male with height between 165 and 210 cm ?
i.e., P(165<X<210) ?



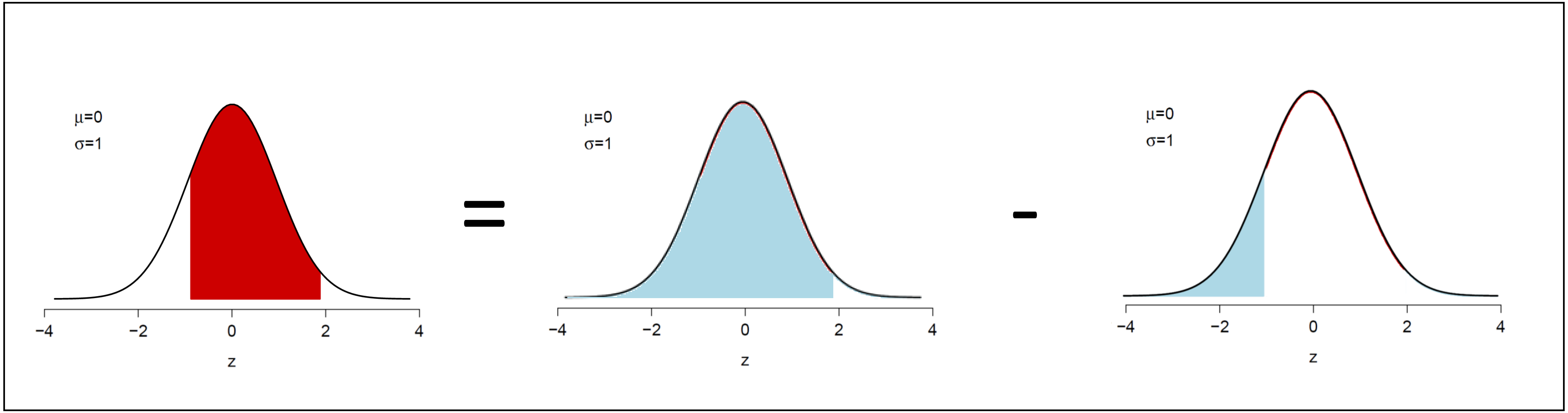
$$z = \frac{165-180}{15} = -1$$

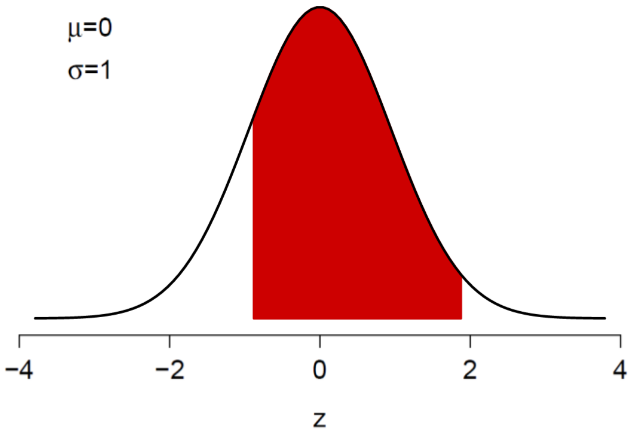
$$z = \frac{210-180}{15} = 2$$

i.e., what is P(-1 < z < 2)??

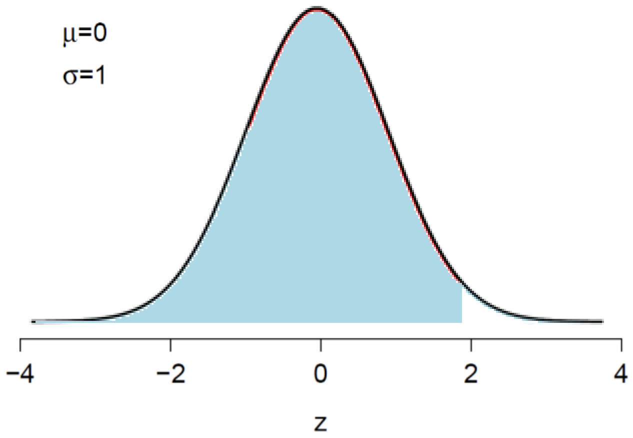
Step 2: Find the probability for z

- $P(-1 < z < 2) = P(z < 2) - P(z < -1)$

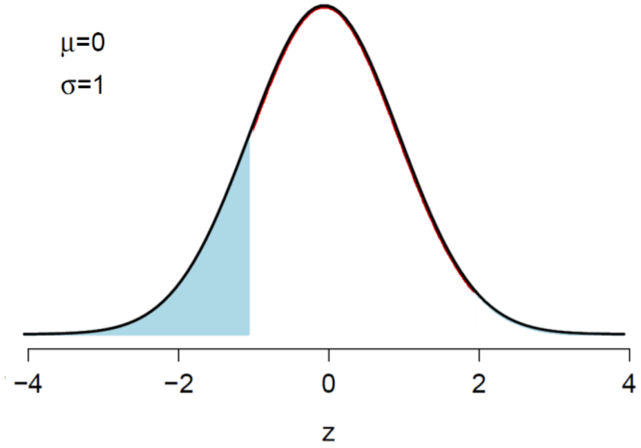


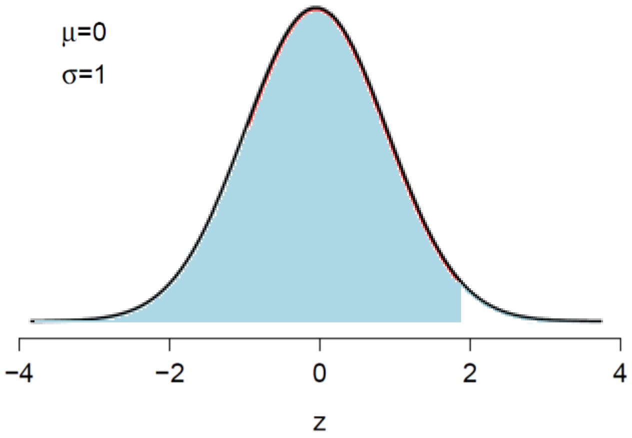


=



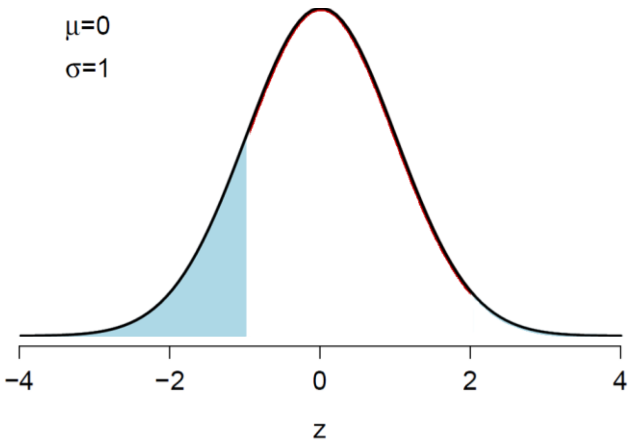
-





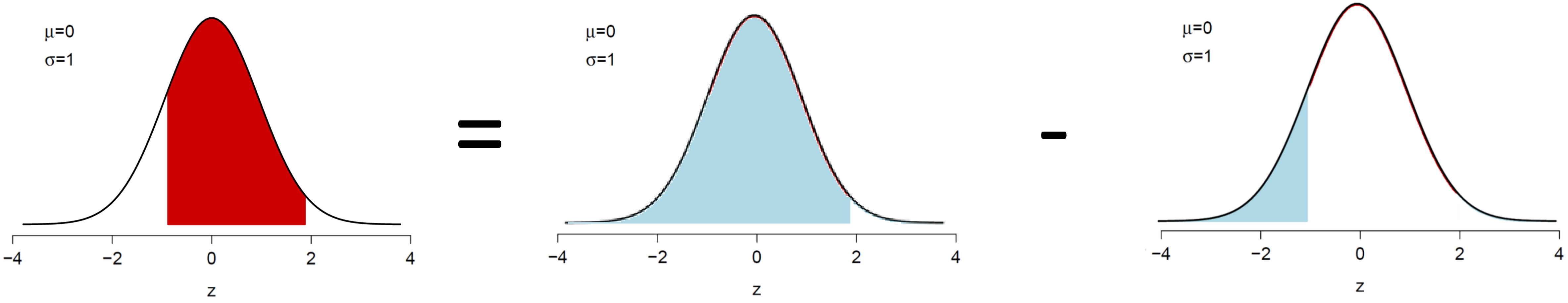
$$\mathbf{z = \frac{210-180}{15} = 2}$$

P(z < 2) = 0.9772



$$\mathbf{z = \frac{165-180}{15} = -1}$$

P(z < -1) = 0.1587



P(-1 < z < 2) = 0.9772 - 0.1587

= 0.8185

So: Probability of sampling someone with height between 165 cm and 210 cm is 0.82

Today

- Random variables

- Probability distributions
    - Discrete variables
        - Three different probability calculations
        - Mean and variability
    - Continuous variables

Two specific distributions:

- Normal distribution (possible for continuous variables)
    - Mean and variability
    - Three different probability calculations

- Binomial distribution (possible for discrete variables)
    - **General formulae**
    - Mean and variability

Binomial distribution

- A probability distribution for a specific discrete variable
- The specific discrete variable: number of 'successes' in *n* independent trials
- This discrete variable is called the *random binomial variable*
- Each trial has binary outcome (e.g., 'correct' & 'incorrect', or 'heads' & 'tails')

- Let's consider an example!

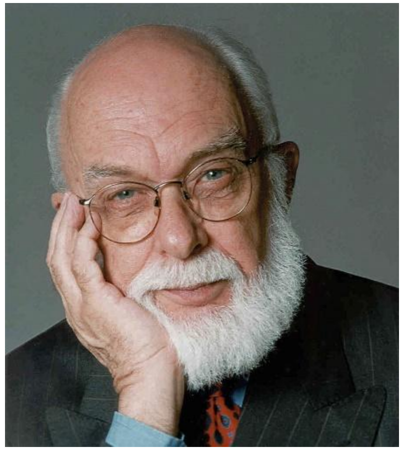
Source image: pixabay.com

One-million dollar paranormal challence by James Randi

Derek Ogilvie

https://nl.wikipedia.org/wiki/Derek_Ogilvie

Year	Challenger	Purported Ability	Test	Results	Notes
2007	Derek Ogilvie	Mediumship	Identify which one out of ten toys is being used by a child at a particular time.	Failed	
2009	Connie Sonne	Dowsing (Pendulum)	Identify playing cards in sealed envelope.	Failed[30]	
2010	Anita Ikonen	Medical dowsing	Determine by observation which of five subjects was missing a kidney.	Failed	Billed as "demonstration" not "test"
2011	No challenger available				
2012	Andrew Needles	Performance-enhancing bracelet	Distinguish participants wearing real product significant number of times.	Failed	
2013	Brahim Addoun	Remote viewing	Remotely identify 3 of 20 objects.	Failed	
2014	Fei Wang	Sending energy through his hand that can be felt by another person.	The energy should be felt by the target person correctly eight out of nine times.	Failed	
2015	No claimant – "Demonstration test"	Ethernet cables that are claimed to be "directional".	Volunteers were played sound twice and were asked to determine which cable had the highest sound quality.	Failed	



James Randi

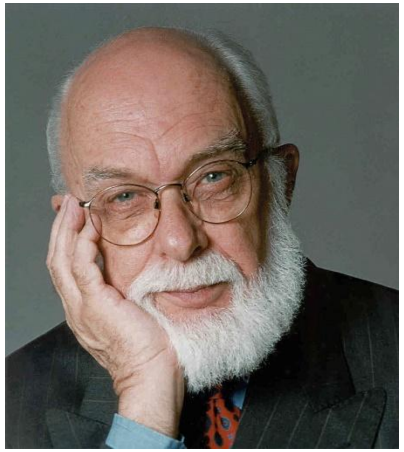
Picture: https://nl.wikipedia.org/wiki/James_Randi

https://en.wikipedia.org/wiki/One_Million_Dollar_Paranormal_Challenge

Example: 1-Million dollar challenge by James Randi

Derek Ogilvie

https://nl.wikipedia.org/wiki/Derek_Ogilvie



James Randi

Picture: https://nl.wikipedia.org/wiki/James_Randi

- One trial has a **binary** outcome (correct / false):

- Let's pretend that Derek had 5 choices in each trial
  - If guessing, probability of correct = 0.2 (1 out of 5; 1/5) and probability of incorrect = 1 – 0.2 = 0.8
- And pretend that Derek had 4 trials
  - n = 4

outcome	#correct	probability
{C, C, C, C}	4	

outcome	#correct	probability
{C, C, C, C}	4	
{C, C, C, F}	3	

outcome	#correct	probability
{C, C, C, C}	4	
{C, C, C, F}	3	
{C, C, F, C}	3	

outcome	#correct	probability
{C, C, C, C}	4	
{C, C, C, F}	3	
{C, C, F, C}	3	
{C, F, C, C}	3	

outcome	#correct	probability
{C, C, C, C}	4	
{C, C, C, F}	3	
{C, C, F, C}	3	
{C, F, C, C}	3	
{F, C, C, C}	3	
{C, C, F, F}	2	
{C, F, F, C}	2	
{C, F, C, F}	2	
{F, C, F, C}	2	
{F, F, C, C}	2	
{F, C, C, F}	2	
{C, F, F, F}	1	
{F, C, F, F}	1	
{F, F, C, F}	1	
{F, F, F, C}	1	
{F, F, F, F}	0	

Recap from Chapter 5

→Trials are independent

Thus: P(A and B) = P(A) × P(B)

→e.g., P(girl AND glasses) = P(girl) × P(glasses)

i.e., P({C, C, F, F})= P(C) × P(C) × P(F) × P(F)

P(C) = 0.2 (probability of a correct guess, 1 out of 5)

P(F) = 1 – 0.2 = 0.8 (complement rule)

P({C, C, F, F})= P(C) × P(C) × P(F) × P(F) = 0.2 × 0.2 × 0.8 × 0.8 = 0.0256

outcome	#correct	probability
{C, C, C, C}	4	
{C, C, C, F}	3	
{C, C, F, C}	3	
{C, F, C, C}	3	
{F, C, C, C}	3	
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	
{C, F, C, F}	2	
{F, C, F, C}	2	
{F, F, C, C}	2	
{F, C, C, F}	2	
{C, F, F, F}	1	
{F, C, F, F}	1	
{F, F, C, F}	1	
{F, F, F, C}	1	
{F, F, F, F}	0	

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	
{C, C, F, C}	3	
{C, F, C, C}	3	
{F, C, C, C}	3	
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	
{C, F, C, F}	2	
{F, C, F, C}	2	
{F, F, C, C}	2	
{F, C, C, F}	2	
{C, F, F, F}	1	
{F, C, F, F}	1	
{F, F, C, F}	1	
{F, F, F, C}	1	
{F, F, F, F}	0	

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	
{C, F, C, C}	3	
{F, C, C, C}	3	
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	
{C, F, C, F}	2	
{F, C, F, C}	2	
{F, F, C, C}	2	
{F, C, C, F}	2	
{C, F, F, F}	1	
{F, C, F, F}	1	
{F, F, C, F}	1	
{F, F, F, C}	1	
{F, F, F, F}	0	

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	
{F, C, C, C}	3	
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	
{C, F, C, F}	2	
{F, C, F, C}	2	
{F, F, C, C}	2	
{F, C, C, F}	2	
{C, F, F, F}	1	
{F, C, F, F}	1	
{F, F, C, F}	1	
{F, F, F, C}	1	
{F, F, F, F}	0	

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	0.2 × 0.8 × 0.2 × 0.2 = 0.0064
{F, C, C, C}	3	
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	
{C, F, C, F}	2	
{F, C, F, C}	2	
{F, F, C, C}	2	
{F, C, C, F}	2	
{C, F, F, F}	1	
{F, C, F, F}	1	
{F, F, C, F}	1	
{F, F, F, C}	1	
{F, F, F, F}	0	

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	0.2 × 0.8 × 0.2 × 0.2 = 0.0064
{F, C, C, C}	3	0.8 × 0.2 × 0.2 × 0.2 = 0.0064
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	
{C, F, C, F}	2	
{F, C, F, C}	2	
{F, F, C, C}	2	
{F, C, C, F}	2	
{C, F, F, F}	1	
{F, C, F, F}	1	
{F, F, C, F}	1	
{F, F, F, C}	1	
{F, F, F, F}	0	

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	0.2 × 0.8 × 0.2 × 0.2 = 0.0064
{F, C, C, C}	3	0.8 × 0.2 × 0.2 × 0.2 = 0.0064
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	0.2 × 0.8 × 0.8 × 0.2 = 0.0256
{C, F, C, F}	2	0.2 × 0.8 × 0.2 × 0.8 = 0.0256
{F, C, F, C}	2	0.8 × 0.2 × 0.8 × 0.2 = 0.0256
{F, F, C, C}	2	0.8 × 0.8 × 0.2 × 0.2 = 0.0256
{F, C, C, F}	2	0.8 × 0.2 × 0.2 × 0.8 = 0.0256
{C, F, F, F}	1	0.2 × 0.8 × 0.8 × 0.8 = 0.1024
{F, C, F, F}	1	0.8 × 0.2 × 0.8 × 0.8 = 0.1024
{F, F, C, F}	1	0.8 × 0.8 × 0.2 × 0.8 = 0.1024
{F, F, F, C}	1	0.8 × 0.8 × 0.8 × 0.2 = 0.1024
{F, F, F, F}	0	0.8 × 0.8 × 0.8 × 0.8 = 0.4096

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	0.2 × 0.8 × 0.2 × 0.2 = 0.0064
{F, C, C, C}	3	0.8 × 0.2 × 0.2 × 0.2 = 0.0064
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	0.2 × 0.8 × 0.8 × 0.2 = 0.0256
{C, F, C, F}	2	0.2 × 0.8 × 0.2 × 0.8 = 0.0256
{F, C, F, C}	2	0.8 × 0.2 × 0.8 × 0.2 = 0.0256
{F, F, C, C}	2	0.8 × 0.8 × 0.2 × 0.2 = 0.0256
{F, C, C, F}	2	0.8 × 0.2 × 0.2 × 0.8 = 0.0256
{C, F, F, F}	1	0.2 × 0.8 × 0.8 × 0.8 = 0.1024
{F, C, F, F}	1	0.8 × 0.2 × 0.8 × 0.8 = 0.1024
{F, F, C, F}	1	0.8 × 0.8 × 0.2 × 0.8 = 0.1024
{F, F, F, C}	1	0.8 × 0.8 × 0.8 × 0.2 = 0.1024
{F, F, F, F}	0	0.8 × 0.8 × 0.8 × 0.8 = 0.4096

X: Number correct

What is the probability of 3 correct?

- P(X = 3)

- = P({C, C, C, F} OR {C, C, F, C} OR {C, F, C, C} OR {F, C, C, C})

  = 0.0064 + 0.0064 + 0.0064 + 0.0064

  = 4 × 0.0064 = 0.0256

Simply adding up because disjoint

Number of ways you can guess 3 items correctly

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	0.2 × 0.8 × 0.2 × 0.2 = 0.0064
{F, C, C, C}	3	0.8 × 0.2 × 0.2 × 0.2 = 0.0064
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	0.2 × 0.8 × 0.8 × 0.2 = 0.0256
{C, F, C, F}	2	0.2 × 0.8 × 0.2 × 0.8 = 0.0256
{F, C, F, C}	2	0.8 × 0.2 × 0.8 × 0.2 = 0.0256
{F, F, C, C}	2	0.8 × 0.8 × 0.2 × 0.2 = 0.0256
{F, C, C, F}	2	0.8 × 0.2 × 0.2 × 0.8 = 0.0256
{C, F, F, F}	1	0.2 × 0.8 × 0.8 × 0.8 = 0.1024
{F, C, F, F}	1	0.8 × 0.2 × 0.8 × 0.8 = 0.1024
{F, F, C, F}	1	0.8 × 0.8 × 0.2 × 0.8 = 0.1024
{F, F, F, C}	1	0.8 × 0.8 × 0.8 × 0.2 = 0.1024
{F, F, F, F}	0	0.8 × 0.8 × 0.8 × 0.8 = 0.4096

X: Number correct

What is the probability of 2 correct?

- P(X = 2)

- = 0.0256 + 0.0256 + 0.0256 + 0.0256 + 0.0256 + 0.0256

= 6 × 0.0256 = 0.1536

Number of ways you can guess 2 items correctly

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	0.2 × 0.8 × 0.2 × 0.2 = 0.0064
{F, C, C, C}	3	0.8 × 0.2 × 0.2 × 0.2 = 0.0064
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	0.2 × 0.8 × 0.8 × 0.2 = 0.0256
{C, F, C, F}	2	0.2 × 0.8 × 0.2 × 0.8 = 0.0256
{F, C, F, C}	2	0.8 × 0.2 × 0.8 × 0.2 = 0.0256
{F, F, C, C}	2	0.8 × 0.8 × 0.2 × 0.2 = 0.0256
{F, C, C, F}	2	0.8 × 0.2 × 0.2 × 0.8 = 0.0256
{C, F, F, F}	1	0.2 × 0.8 × 0.8 × 0.8 = 0.1024
{F, C, F, F}	1	0.8 × 0.2 × 0.8 × 0.8 = 0.1024
{F, F, C, F}	1	0.8 × 0.8 × 0.2 × 0.8 = 0.1024
{F, F, F, C}	1	0.8 × 0.8 × 0.8 × 0.2 = 0.1024
{F, F, F, F}	0	0.8 × 0.8 × 0.8 × 0.8 = 0.4096

General formula

- We have *n* trials of which the outcome is binary ('success' and 'no success')
- The probability of success in a trial is *p*
- The probability of no succes in a trial is *1 - p*

Factorial in excel:
x! = fact(x)

Then: the probability of *x* successes is given by:

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

! is the factorial function:
*5! = 5*4*3*2*1 = 120*
*3! = 3*2*1 = 6*
1! = 1
0! = 1

Number of ways in which you get x successes in n trials

Probability

How many permutations (ways to order) are there of 0 numbers?
One! (doing nothing)

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	0.2 × 0.8 × 0.2 × 0.2 = 0.0064
{F, C, C, C}	3	0.8 × 0.2 × 0.2 × 0.2 = 0.0064
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	0.2 × 0.8 × 0.8 × 0.2 = 0.0256
{C, F, C, F}	2	0.2 × 0.8 × 0.2 × 0.8 = 0.0256
{F, C, F, C}	2	0.8 × 0.2 × 0.8 × 0.2 = 0.0256
{F, F, C, C}	2	0.8 × 0.8 × 0.2 × 0.2 = 0.0256
{F, C, C, F}	2	0.8 × 0.2 × 0.2 × 0.8 = 0.0256
{C, F, F, F}	1	0.2 × 0.8 × 0.8 × 0.8 = 0.1024
{F, C, F, F}	1	0.8 × 0.2 × 0.8 × 0.8 = 0.1024
{F, F, C, F}	1	0.8 × 0.8 × 0.2 × 0.8 = 0.1024
{F, F, F, C}	1	0.8 × 0.8 × 0.8 × 0.2 = 0.1024
{F, F, F, F}	0	0.8 × 0.8 × 0.8 × 0.8 = 0.4096

General formula

- We have *n* trials of which the outcome is binary ('success' and 'no success')
- The probability of success in a trial is *p*
- The probability of no succes in a trial is *1 - p*

Then: the probability of *x* successes is given by:

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

For this example about Ogilvie

- P(4) = 1 × 0.20^4 × 0.80^0 = 0.0016
- P(3) = 4 × 0.20^3 × 0.80^1 = 0.0256
- P(2) = 6 × 0.20^2 × 0.80^2 = 0.1536
- P(1) = 4 × 0.20^1 × 0.80^3 = 0.4096
- P(0) = 1 × 0.20^0 × 0.80^4 = 0.4096

Today

- Random variables

- Probability distributions
  - Discrete variables
    - Three different probability calculations
    - Mean and variability
  - Continuous variables

Two specific distributions:

- Normal distribution (possible for continuous variables)
  - Mean and variability
  - Three different probability calculations

- Binomial distribution (possible for discrete variables)
  - General formulae
  - **Mean and variability**

outcome	#correct	probability
{C, C, C, C}	4	0.2 × 0.2 × 0.2 × 0.2 = 0.0016
{C, C, C, F}	3	0.2 × 0.2 × 0.2 × 0.8 = 0.0064
{C, C, F, C}	3	0.2 × 0.2 × 0.8 × 0.2 = 0.0064
{C, F, C, C}	3	0.2 × 0.8 × 0.2 × 0.2 = 0.0064
{F, C, C, C}	3	0.8 × 0.2 × 0.2 × 0.2 = 0.0064
{C, C, F, F}	2	0.0256 (see previous slide)
{C, F, F, C}	2	0.2 × 0.8 × 0.8 × 0.2 = 0.0256
{C, F, C, F}	2	0.2 × 0.8 × 0.2 × 0.8 = 0.0256
{F, C, F, C}	2	0.8 × 0.2 × 0.8 × 0.2 = 0.0256
{F, F, C, C}	2	0.8 × 0.8 × 0.2 × 0.2 = 0.0256
{F, C, C, F}	2	0.8 × 0.2 × 0.2 × 0.8 = 0.0256
{C, F, F, F}	1	0.2 × 0.8 × 0.8 × 0.8 = 0.1024
{F, C, F, F}	1	0.8 × 0.2 × 0.8 × 0.8 = 0.1024
{F, F, C, F}	1	0.8 × 0.8 × 0.2 × 0.8 = 0.1024
{F, F, F, C}	1	0.8 × 0.8 × 0.8 × 0.2 = 0.1024
{F, F, F, F}	0	0.8 × 0.8 × 0.8 × 0.8 = 0.4096

General formula

- We have *n* trials of which the outcome is binary ('success' and 'no success')
- The probability of success in a trial is *p*
- The probability of no succes in a trial is *1 - p*

Then: the probability of *x* successes is given by:

$$P(x) = \frac{n!}{x!\,(n-x)!}p^x(1-p)^{n-x}$$

Mean

μ = *np*

Standard deviation

σ = $\sqrt{np(1-p)}$

For the binomial, the more general formula for the mean of discrete variables ($\mu = \Sigma x P(x)$) simplifies to: np

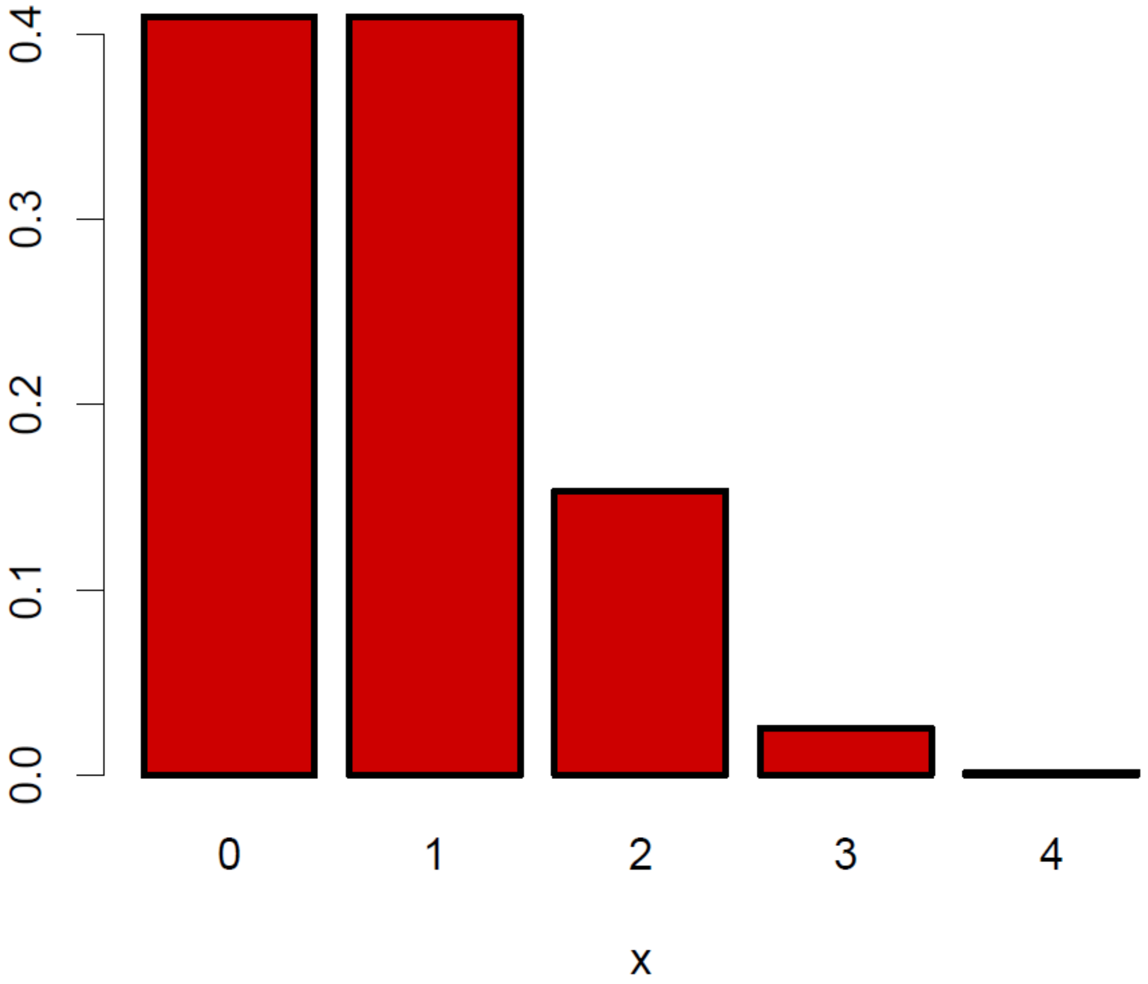
$$\mu = np = 4 \times 0.2 = 0.8$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{4 \times 0.2 \times (1-0.2)} = 0.8$$

So, suppose Derek guesses and n=4, p=0.2, then this is what we can expect:

Derek Ogilvie

https://nl.wikipedia.org/wiki/Derek_Ogilvie



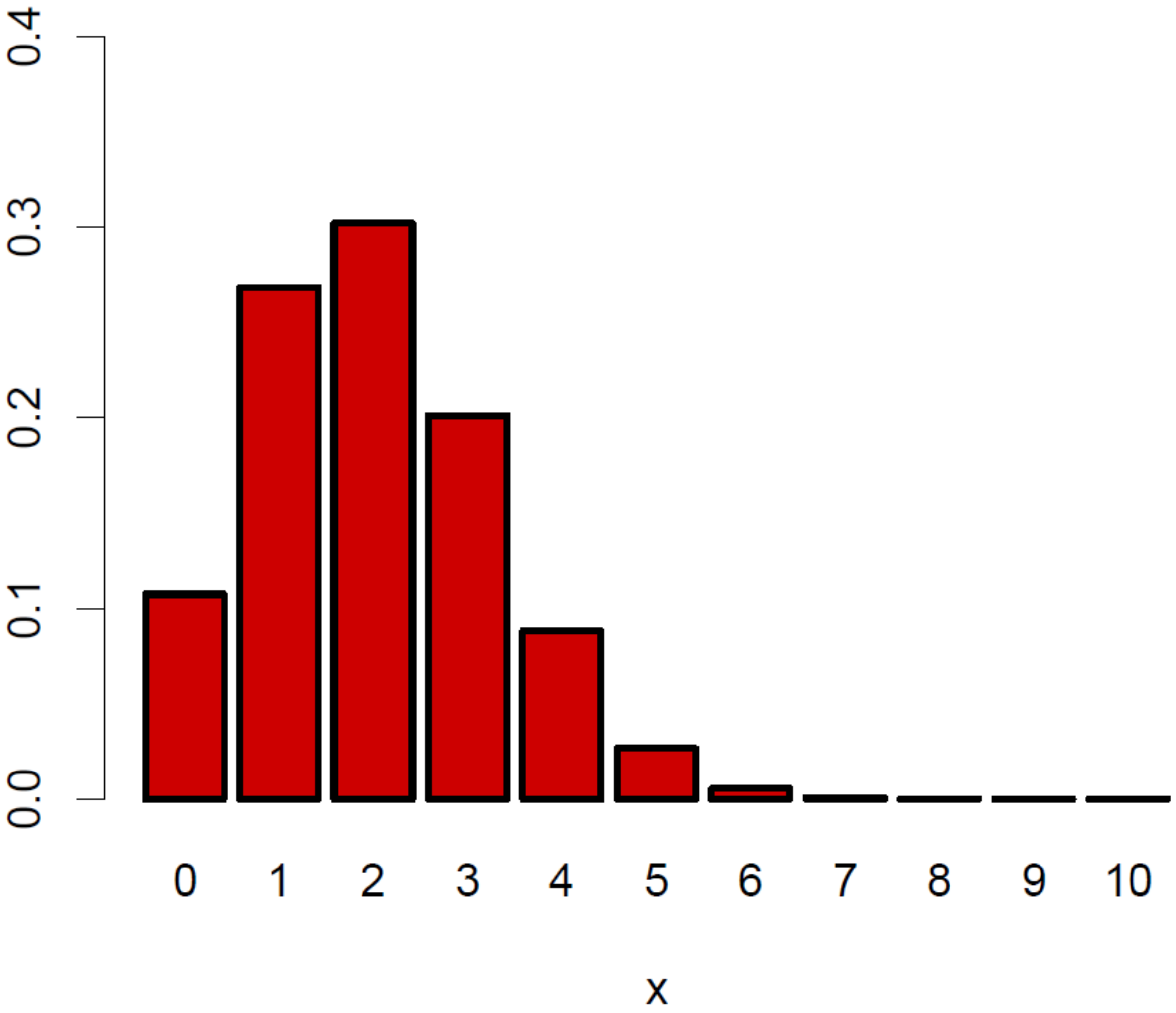
$$\mu = np = 10 \times 0.2 = 2$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{10 \times 0.2 \times (1-0.2)} = 1.26$$

Increasing the trials..
n=10, p=0.2, then this is what we can expect:

Derek Ogilvie

https://nl.wikipedia.org/wiki/Derek_Ogilvie



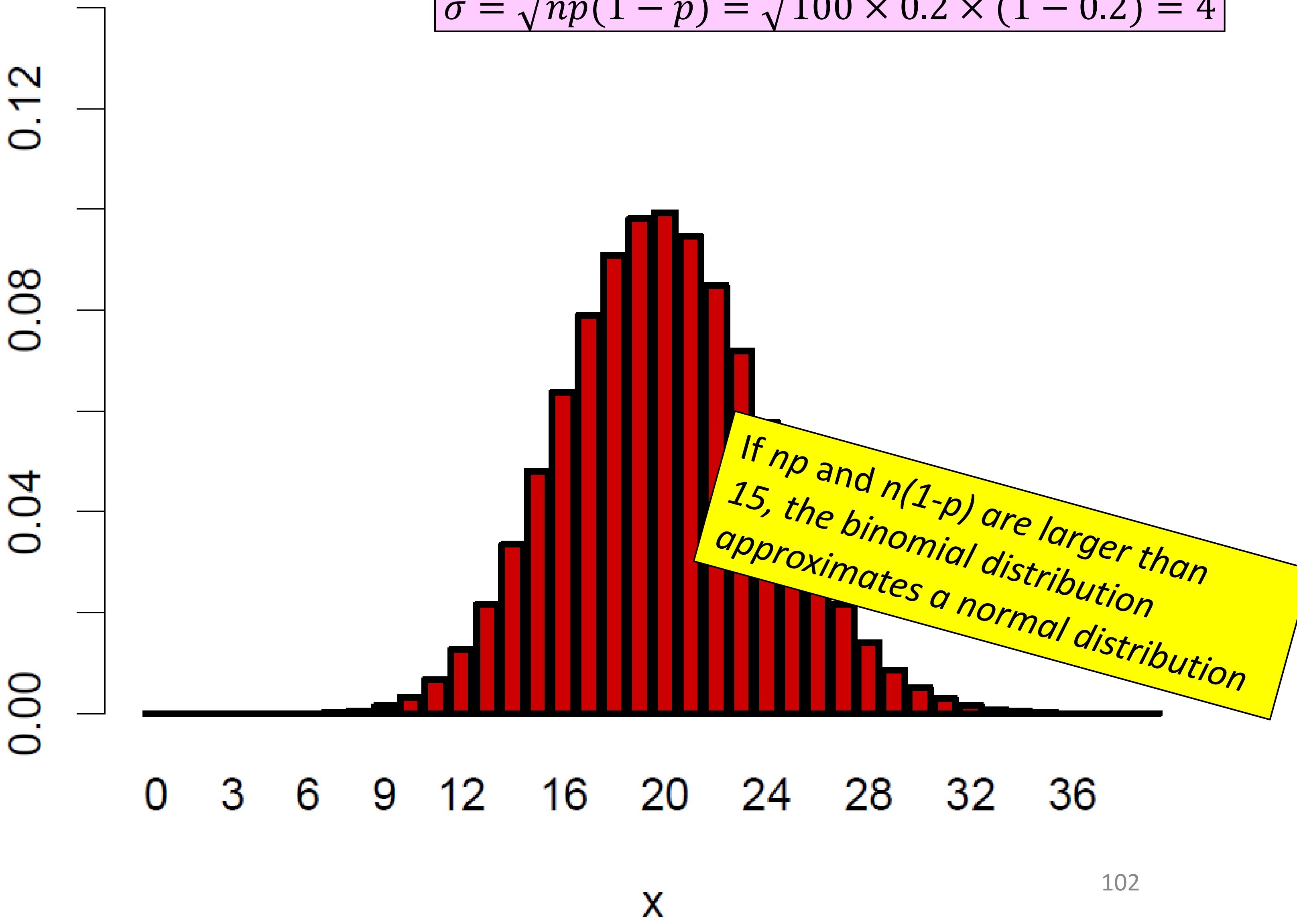
Increasing the trials..
n=100, p=0.2, then this is what we can expect:

$$\mu = np = 100 \times 0.2 = 20$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.2 \times (1-0.2)} = 4$$

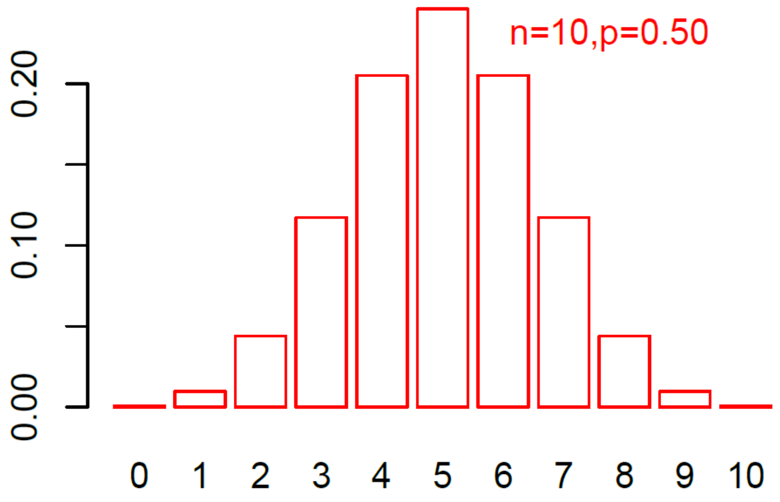
Derek Ogilvie

https://nl.wikipedia.org/wiki/Derek_Ogilvie



Thus:

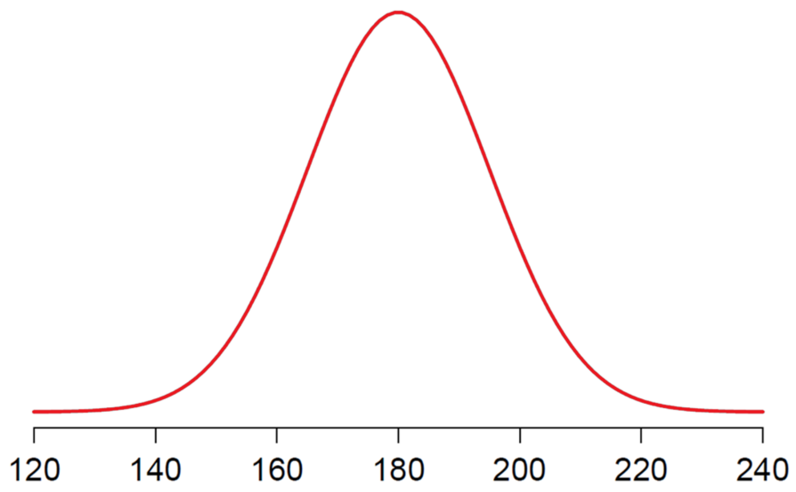
Binomial distribution:



μ = *np*

σ = $\sqrt{np(1-p)}$

Normal distribution:



μ

σ

Today

- The normal distribution is a specific distribution that can hold for continuous variables. In case a continuous variable is normally distributed, you can use z-values to calculate probabilities of intervals of the distribution.
- The binomial random variable, X (number of successes in n independent trials), is an example of a discrete variable. The distribution of this variable is the binomial distribution.
- The binomial distribution is just like the normal distribution characterized by two parameters. For the normal distribution this is μ (the mean) and σ (the SD) and for the binomial distribution this is p (the probability of a success) and n (the number of trials).
- With sufficient trials (np>15 and n(1-p)>15) the binomial distribution will approximate a normal distribution, such that the tricks learned for the normal distribution can be applied there as well.

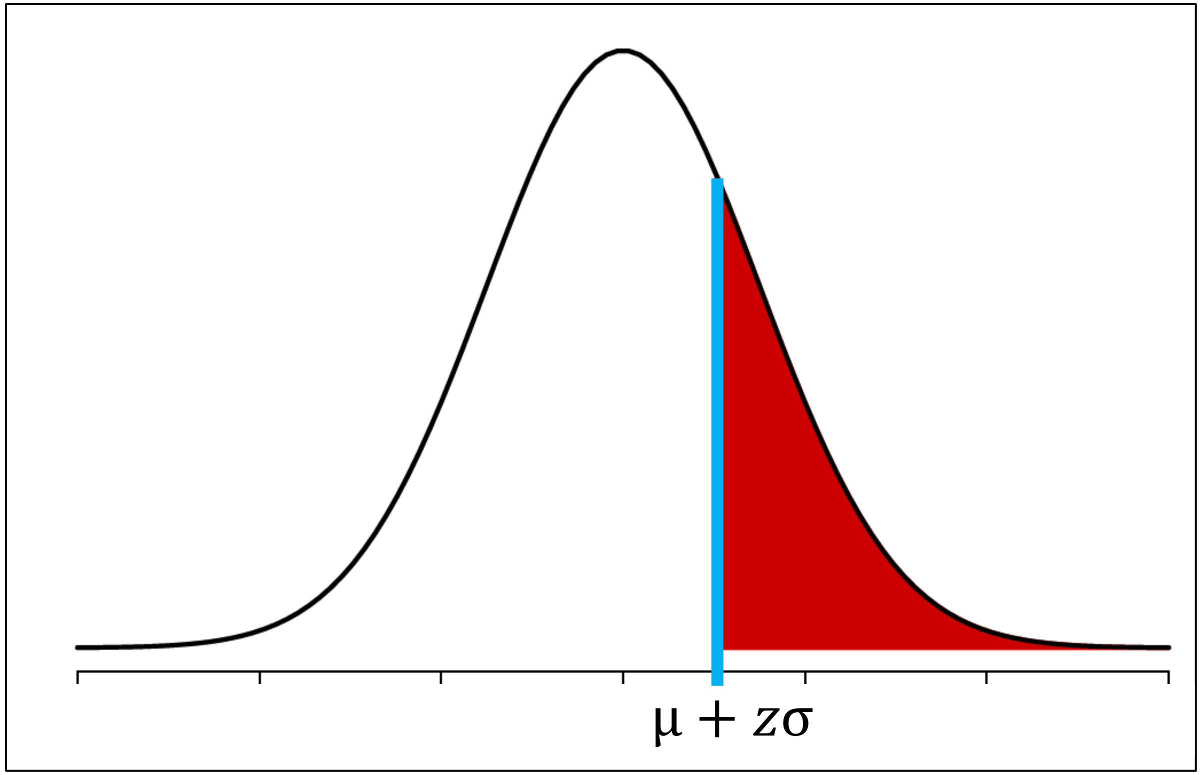
Practice exam question

Consider the normal distribution below, what is the z-score at the blue line given that the red area has a probability mass of 0.3?

a) 0.5244

b) 0.7580

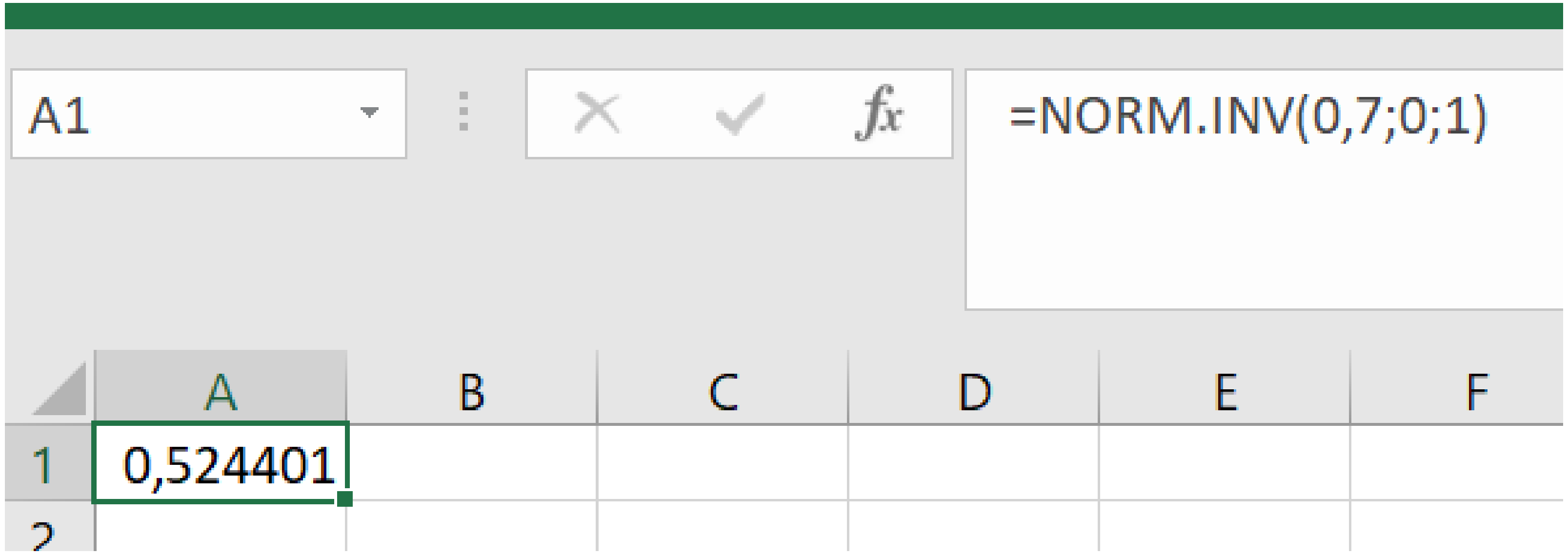
c) 0.6179



Answer

Step 1: The probability on the left side of the z-value is 1 - 0.3 = 0.7

Step 2: use excel to find the z-value associated with this left-tail probability:



The answer:

Z = 0.5244