



# Statistics: lecture 2

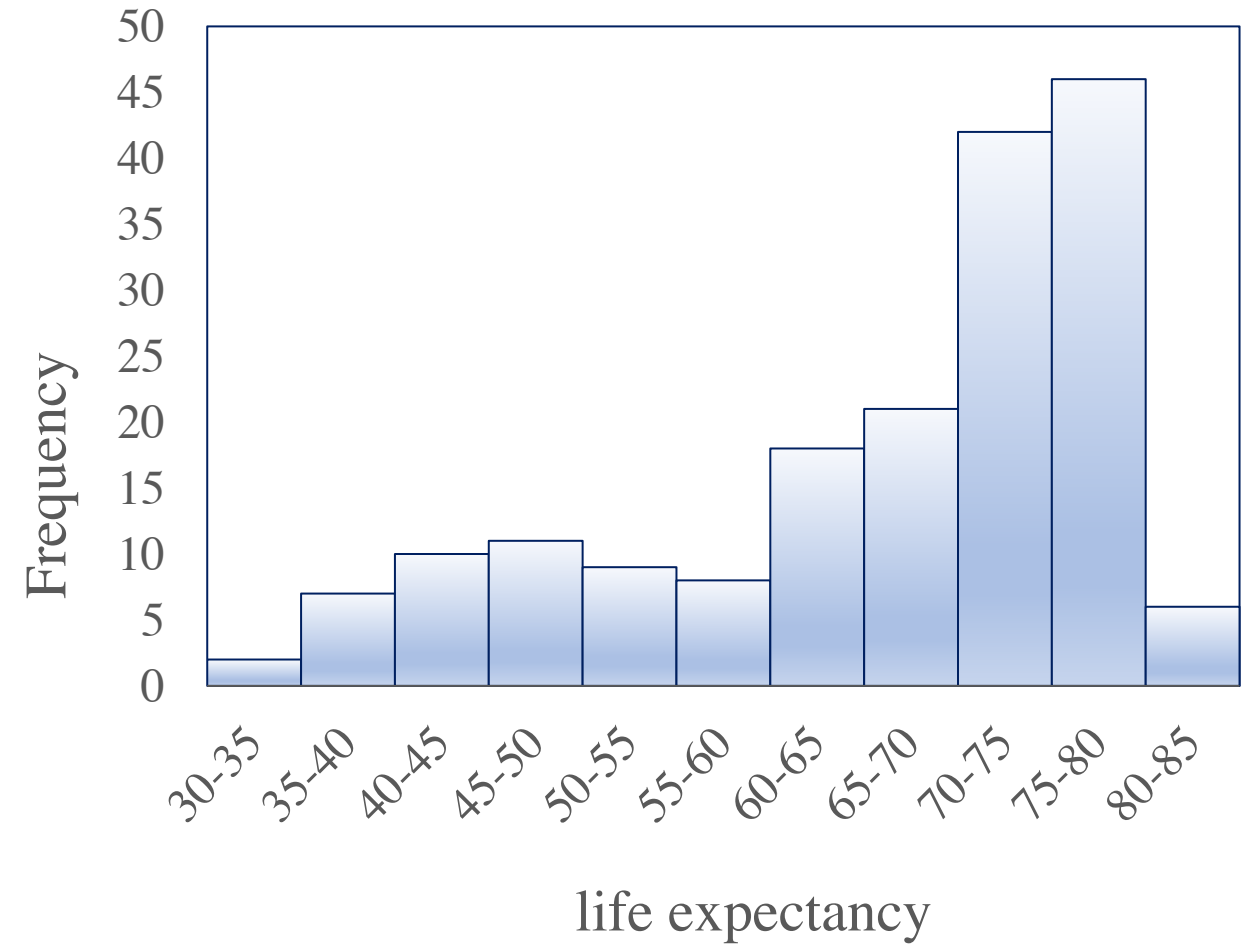
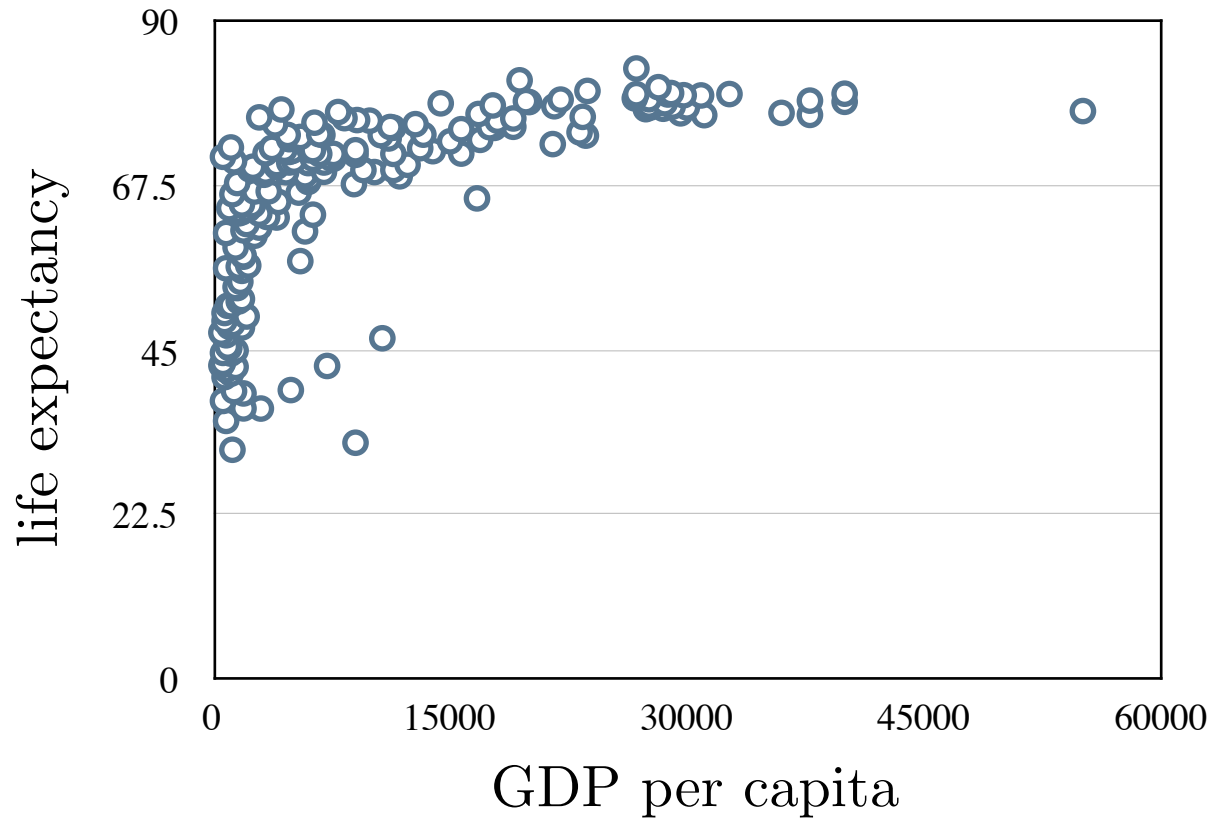
**Naima Hammoud**

Feb 16, 2017

# Distributions

- A **distribution** of a population (or of sampled data from the population) is a representation that shows all the possible data outcomes and their frequency (how often they occur ).
- For any distribution of numerical data, we can calculate the mean, median, and standard deviation (or variance)

# Distributions



# Distributions

- The mean of a distribution is simply the arithmetic average

Example: Consider a list of the number of books bought by 10 different students in 2016

8, 6, 16, 12, 13, 14, 12, 5, 9, 10

The mean, which we call  $\bar{x}$ , is:

$$\bar{x} = \frac{8 + 6 + 16 + 12 + 13 + 14 + 12 + 5 + 9 + 10}{10} = 10.5$$



# Distributions

- The median of a distribution is the midpoint of the data  
Example:

8, 6, 16, 12, 13, 14, 12, 5, 9, 10

5, 6, 8, 9, 10, 12, 12, 13, 14, 16



$$\text{median} = \frac{10 + 12}{2} = 11$$

# Distributions

- The variance  $s^2$ , is the average of the squared deviations of the data from the mean

Example:

8, 6, 16, 12, 13, 14, 12, 5, 9, 10

$$\bar{x} = 10.5$$

$$s^2 = \frac{(8 - 10.5)^2 + (6 - 10.5)^2 + (16 - 10.5)^2 + \cdots + (5 - 10.5)^2 + (9 - 10.5)^2 + (10 - 10.5)^2}{9}$$
$$= 12.5$$

- The standard deviation (SD or  $s$ ), is the square root of the variance

$$s = \sqrt{12.5} \approx 3.5$$

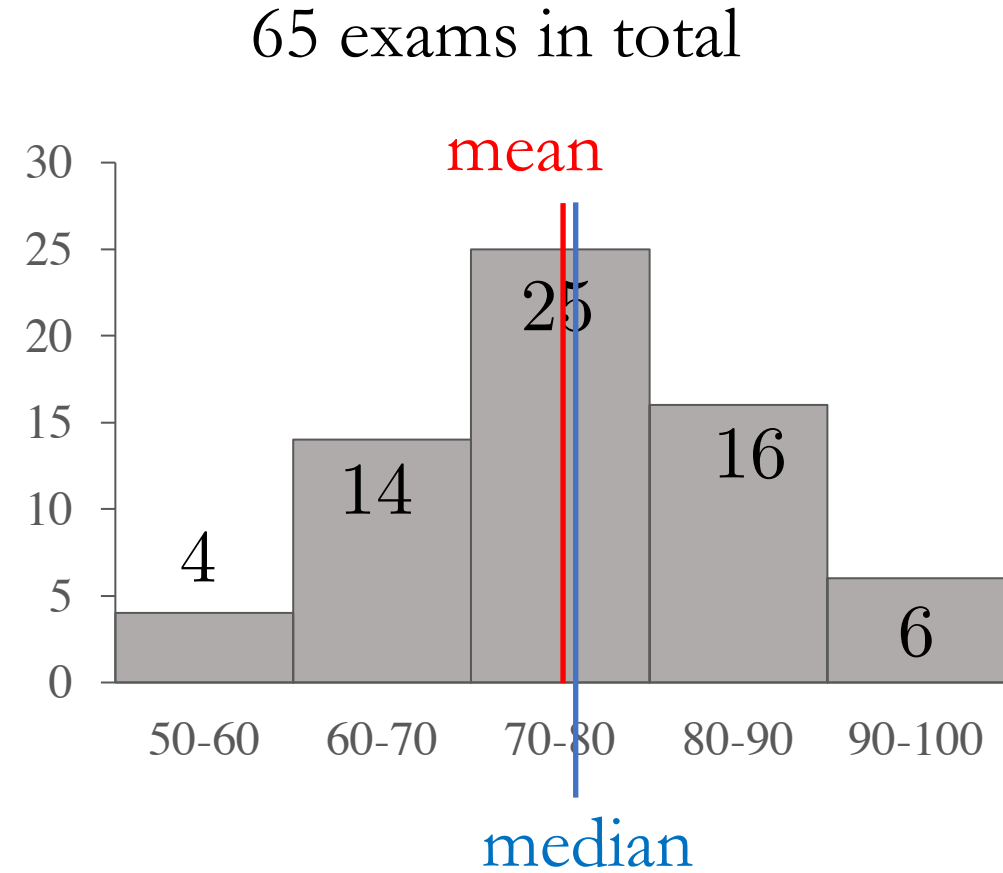
Think of SD as the error

# Distributions

Suppose I give you a midterm which is of average difficulty, the distribution of the scores should be more or less symmetric

Where is the median?

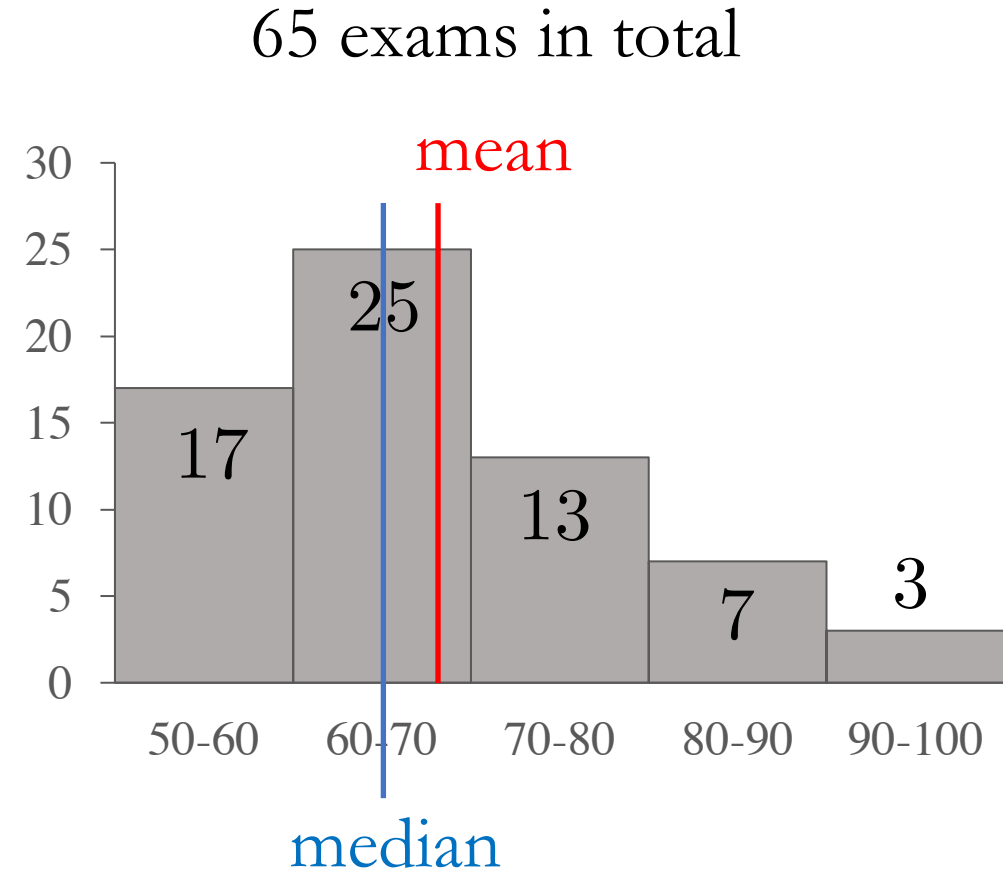
It should be the 33<sup>rd</sup> test score once the grades are arranged in order



# Distributions

What if I make the exam really difficult?

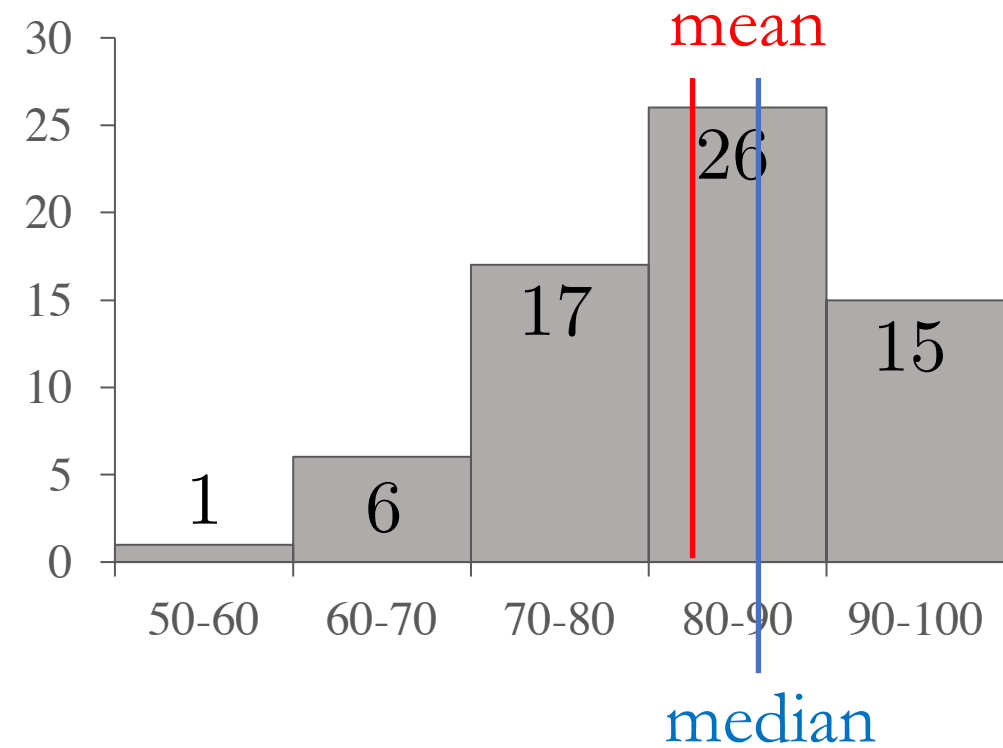
The distribution will be right skewed with the mean larger than the median



# Distributions

What if I make the exam really easy?

The distribution will be left skewed with the mean smaller than the median

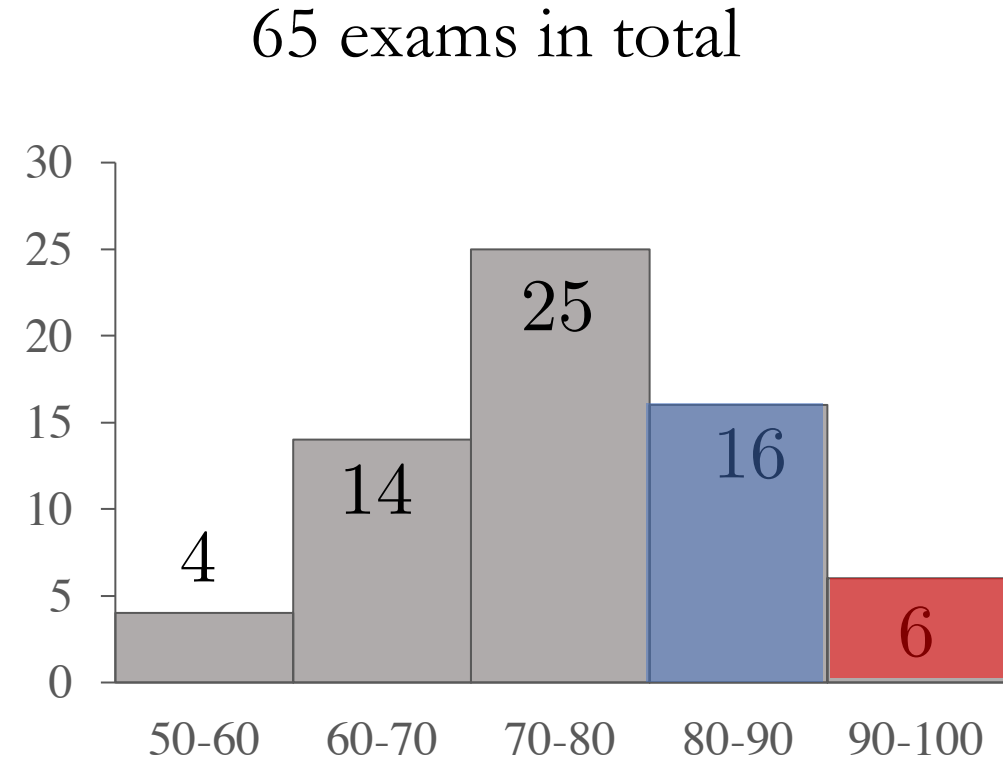


# Distributions

Suppose I give a reasonable exam with an average of 74.

What is the probability that a student chosen at random has scored above 80?

$$P(\text{score} > 80) = \frac{16 + 6}{65} \approx 0.34$$

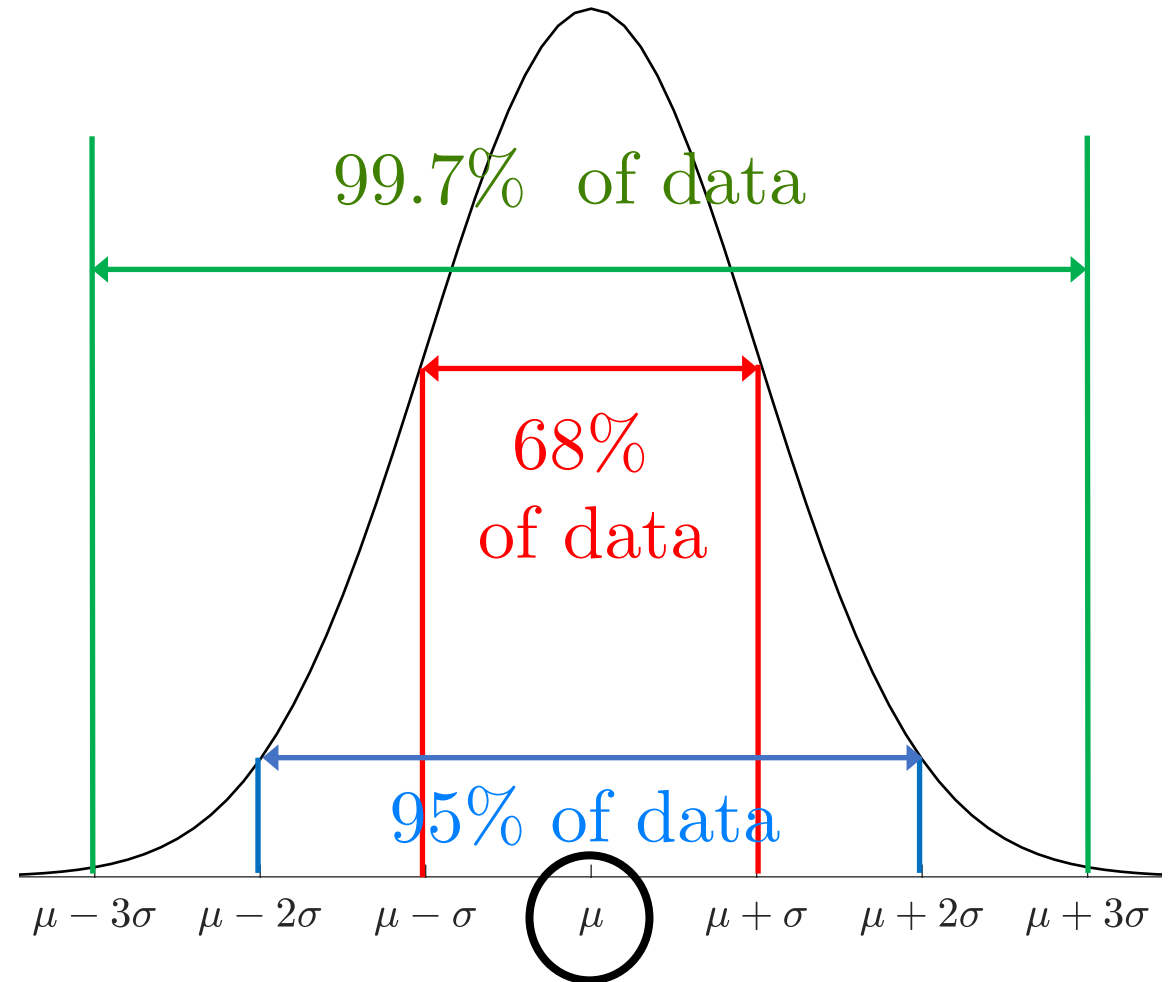


# The normal distribution

The normal distribution is shaped like a bell, and it is symmetric about the mean

normal distribution with mean  $\mu$  and standard deviation  $\sigma$

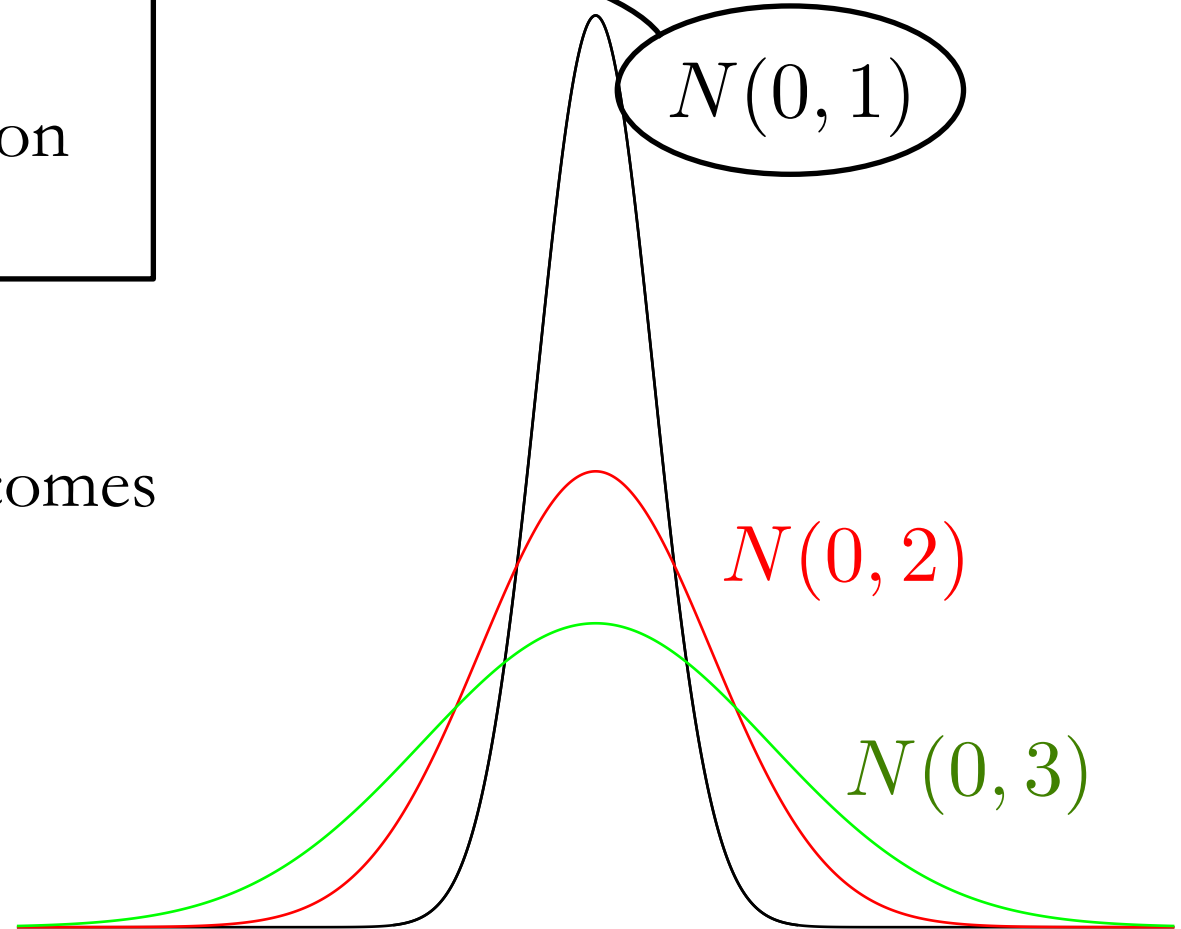
$$N(\mu, \sigma)$$



# The normal distribution

We will refer to this as the  
**standard normal** distribution  
Or the *z-curve*

As  $\sigma$  increases, the distribution becomes  
more spread out around the mean

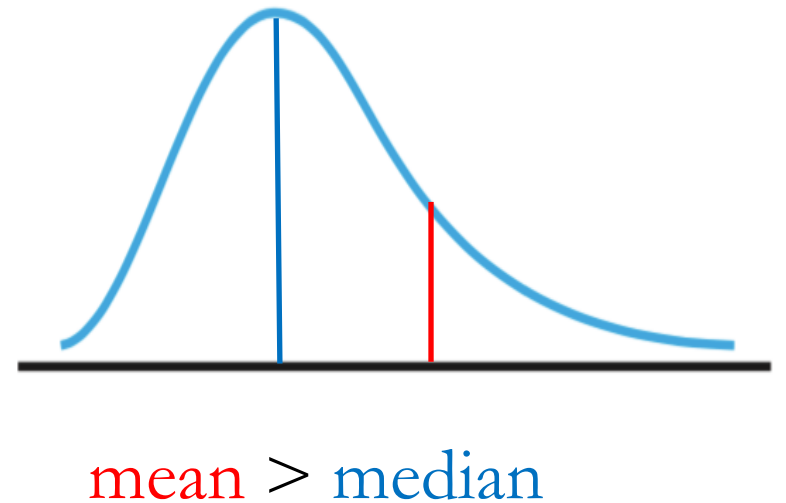




# Example 1

Among all American adults, the average (mean) number of books read or listened to in the past year is 12 and the median (midpoint) number is 5—in other words, half of adults read more than 5 books and half read fewer.

How do you think the distribution looks like?



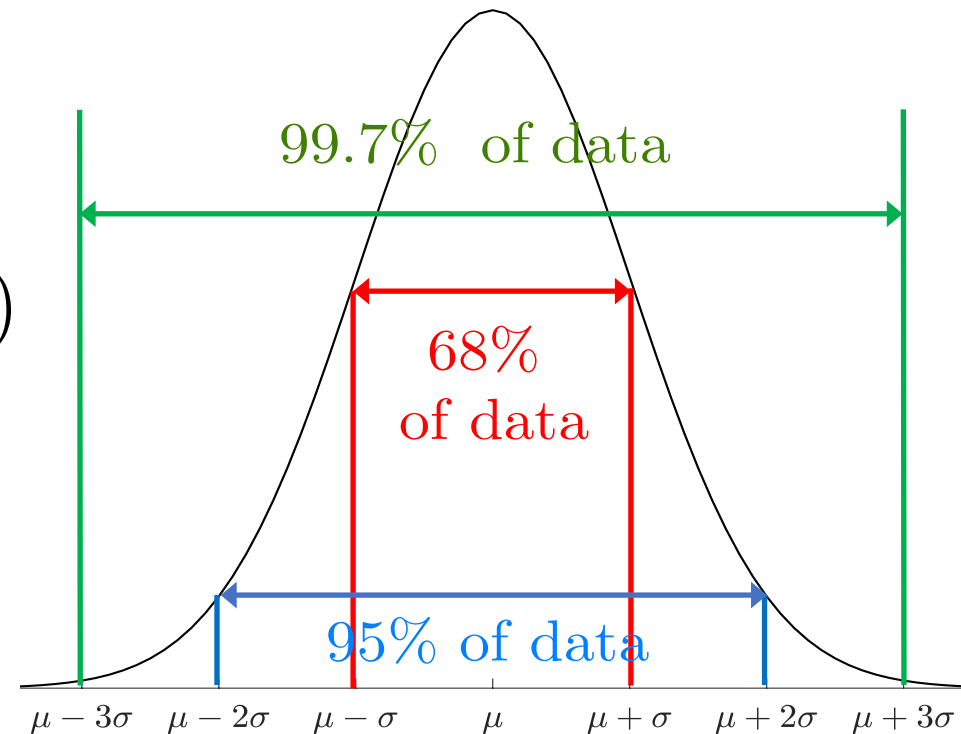
# Example 2

The average height, weight, IQ scores are all normally distributed. For example, the average height of American men is 70 inches (5'10") with a standard deviation of 4 inches.

What is the range of weights of American men?

$$70 \pm (3 \times 4) = 70 \pm 12 = (58, 82)$$

So 99.7% of American men are between 58 and 82 inches.



# Example 3

Amy and Colin are taking the same course, but with two different professors. On the midterm exam (which was different for each class), Amy scored 88, whereas Colin scored 95. The grades for both exams were normally distributed with:

Colin's midterm  $\approx N(77, 7)$

Amy's midterm  $\approx N(72, 6)$

Who do you think did better?

# Example 3

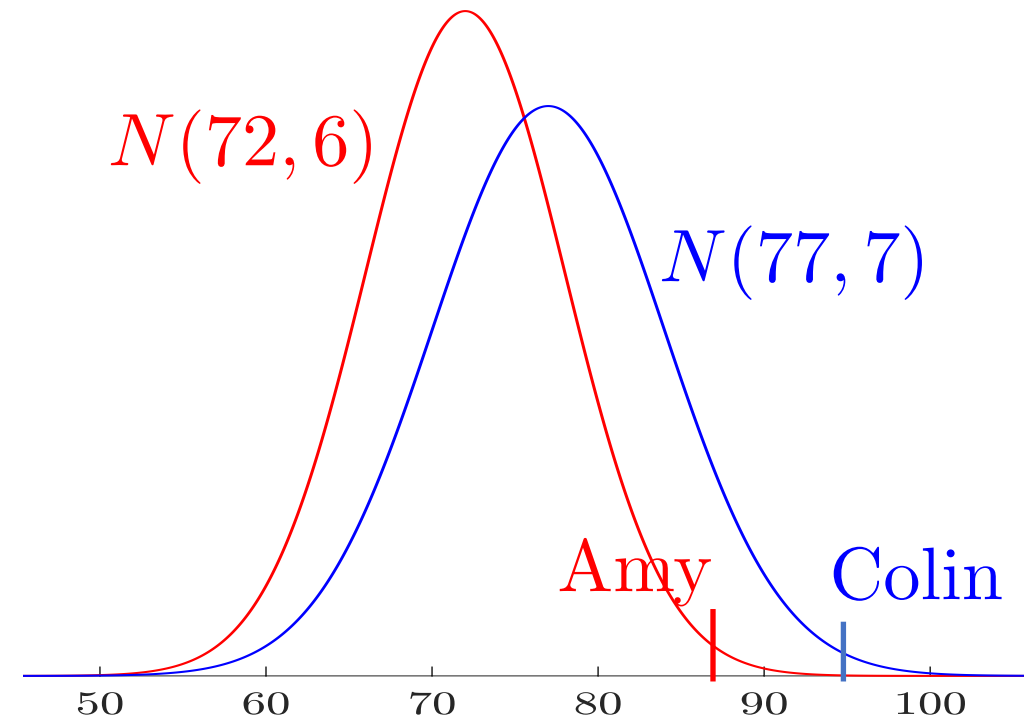
Amy and Colin are taking the same course, but with two different professors. On the midterm exam (which was different for each class), Amy scored 88, whereas Colin scored 95. The grades for both exams were normally distributed with:

Colin's midterm  $\approx N(77, 7)$

Amy's midterm  $\approx N(72, 6)$

Who do you think did better?

This is literally like comparing apples to oranges!!

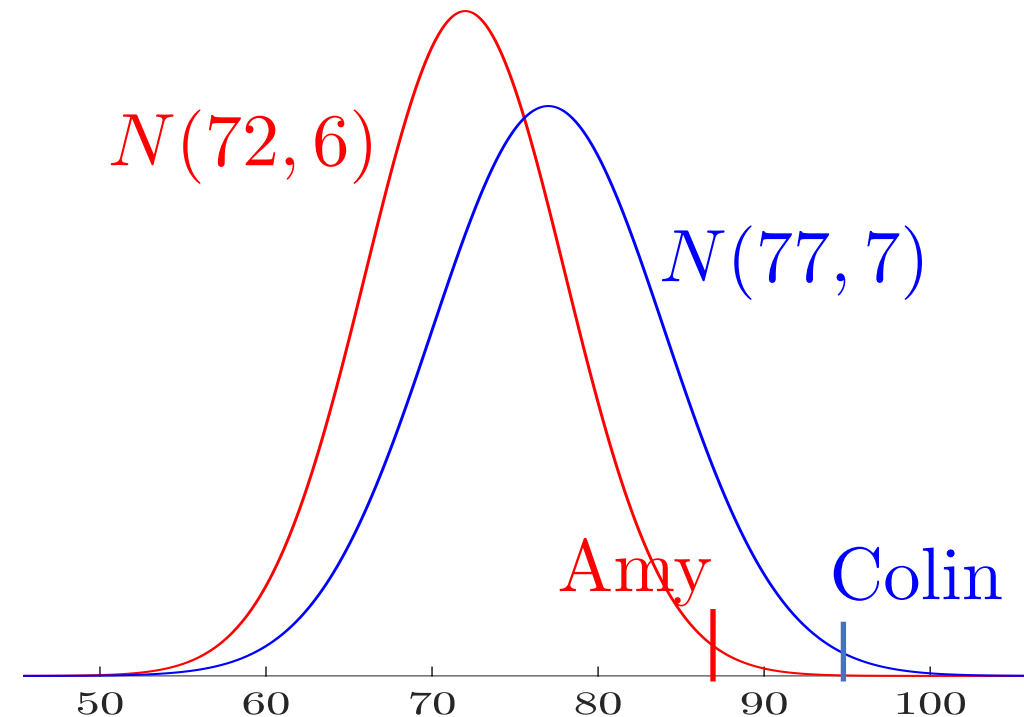


# Example 3

Now we must find a way to compare Amy's grade to Colin's

This is called standardizing, or finding the ***z-score***

Basically, we want to see where Amy and Colin's scores fall on a standard normal



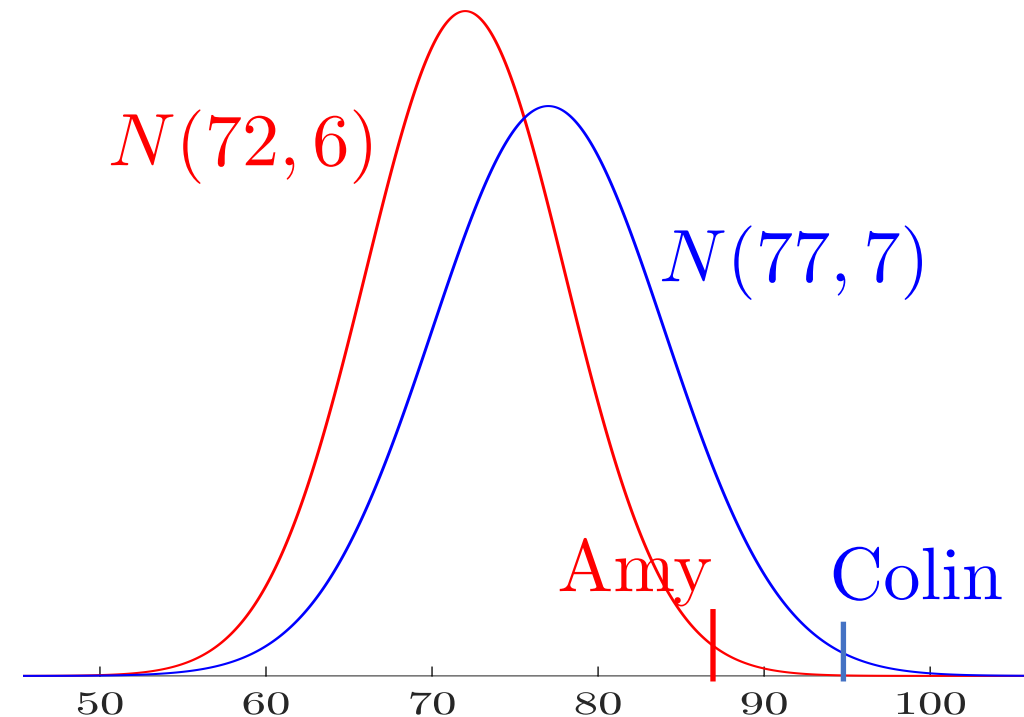
# Example 3

Remember that we called the standard normal the ***z-curve***

So, a ***z-score*** is finding where the grades would fall on a standard normal

$$\text{Amy : } \frac{88 - 72}{6} \approx 2.67$$

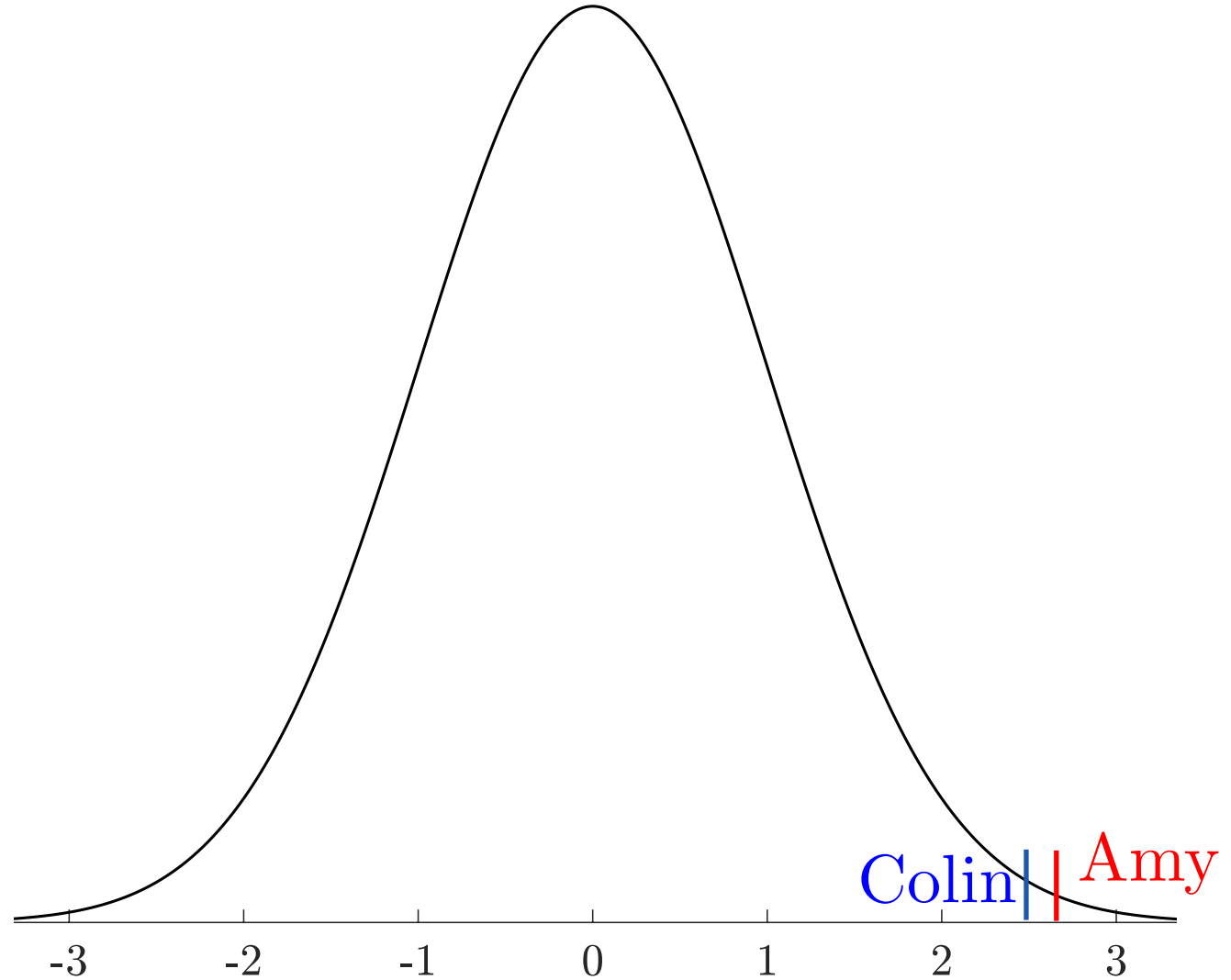
$$\text{Colin : } \frac{95 - 77}{7} \approx 2.57$$



# Example 3

$$\text{Amy : } \frac{88 - 72}{6} \approx 2.67$$

$$\text{Colin : } \frac{95 - 77}{7} \approx 2.57$$



# Standardizing

- As in the example, to standardize from a normal distribution  $N(\mu, \sigma)$  to the standard normal  $N(0, 1)$  we have to find the *z-score*

$$Z = \frac{\text{observation} - \mu}{\sigma}$$

- If the observation is 1 SD above the normal, then  $Z=1$
- Recall that  $Z=2$  means we are two SDs above the normal, and that's already where 95% of the data should be, so anything above that is very hard to obtain.
- We can find the *z-score* for a distribution of any shape



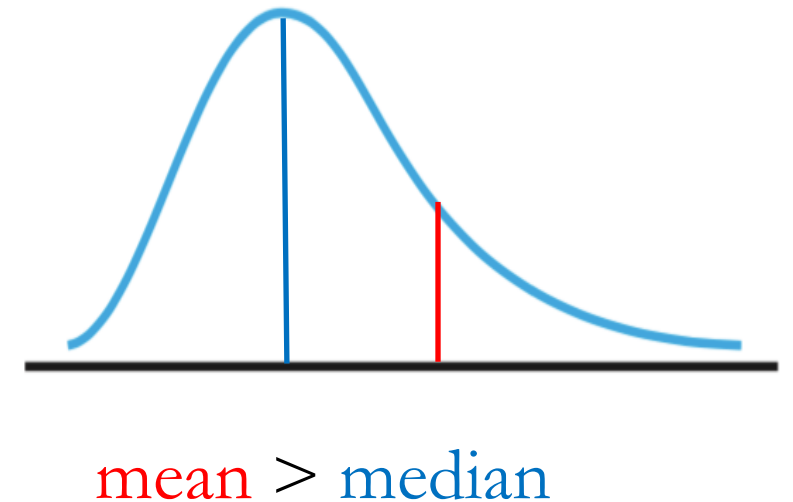
# Back to Example 1

Among all American adults, the average (mean) number of books read or listened to in the past year is 12 and the median (midpoint) number is 5—in other words, half of adults read more than 5 books and half read fewer.

Suppose the SD is 3, then how many books is considered quite high?

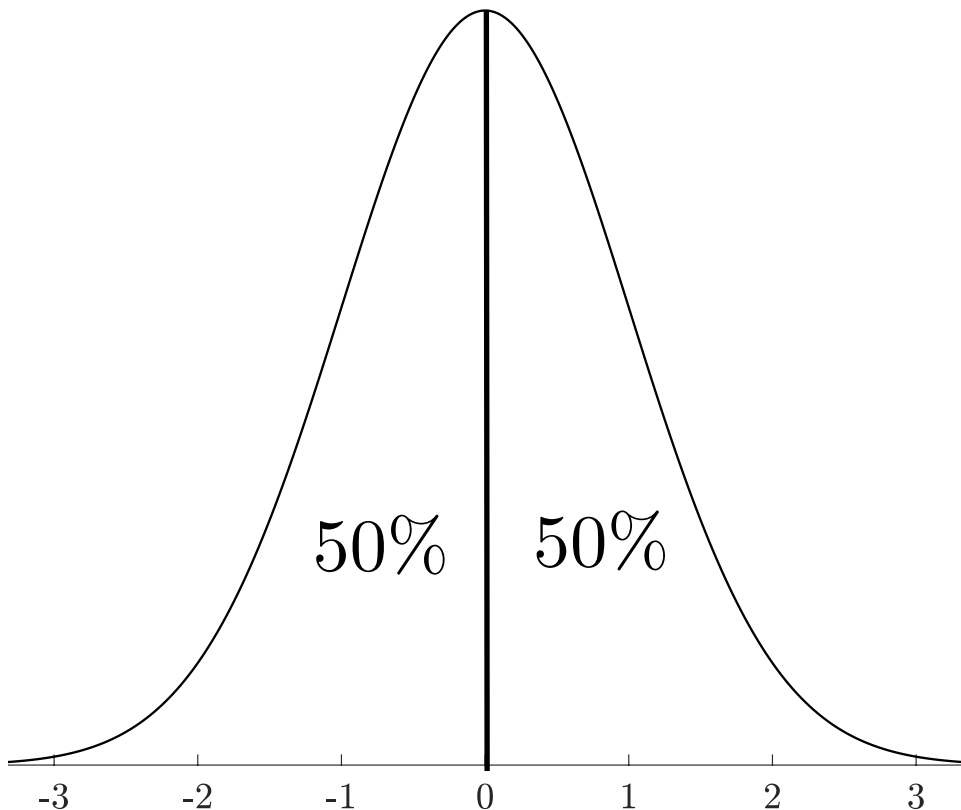
$$2 = \frac{\text{observation} - 12}{3}$$

$$\text{observation} = 18$$



# Percentiles

We have mentioned the median, which is the 50<sup>th</sup> percentile of the data, i.e. 50% of the data lies above the median and 50% below it



If we take a random variable  $X$  from this distribution, then

$$P(X > \text{median}) = 0.5 \text{ (top 50\%)}$$

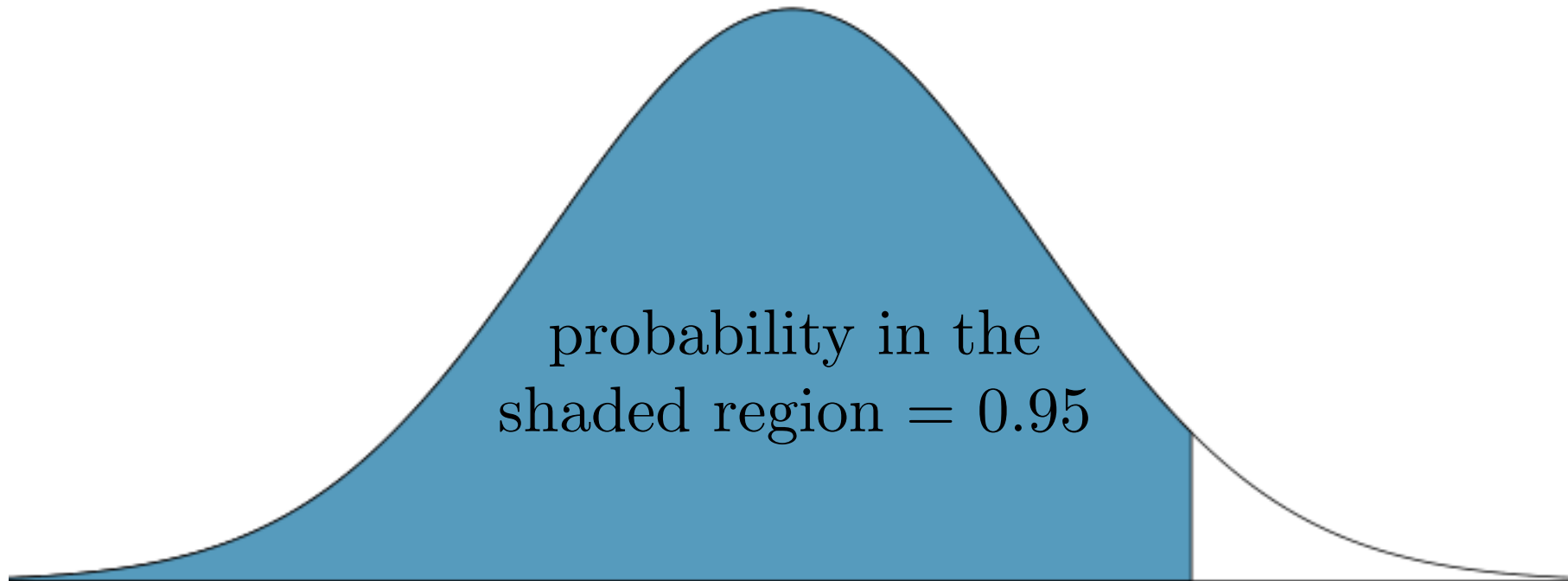
$$P(X < \text{median}) = 0.5 \text{ (bottom 50\%)}$$

# Percentiles

In general, we will refer to a percentile either as the bottom % or the top %.

For example, if we are drawing a variable  $X$  from a normal distribution, then the probability that  $X$  is in the bottom 95<sup>th</sup> percentile is:

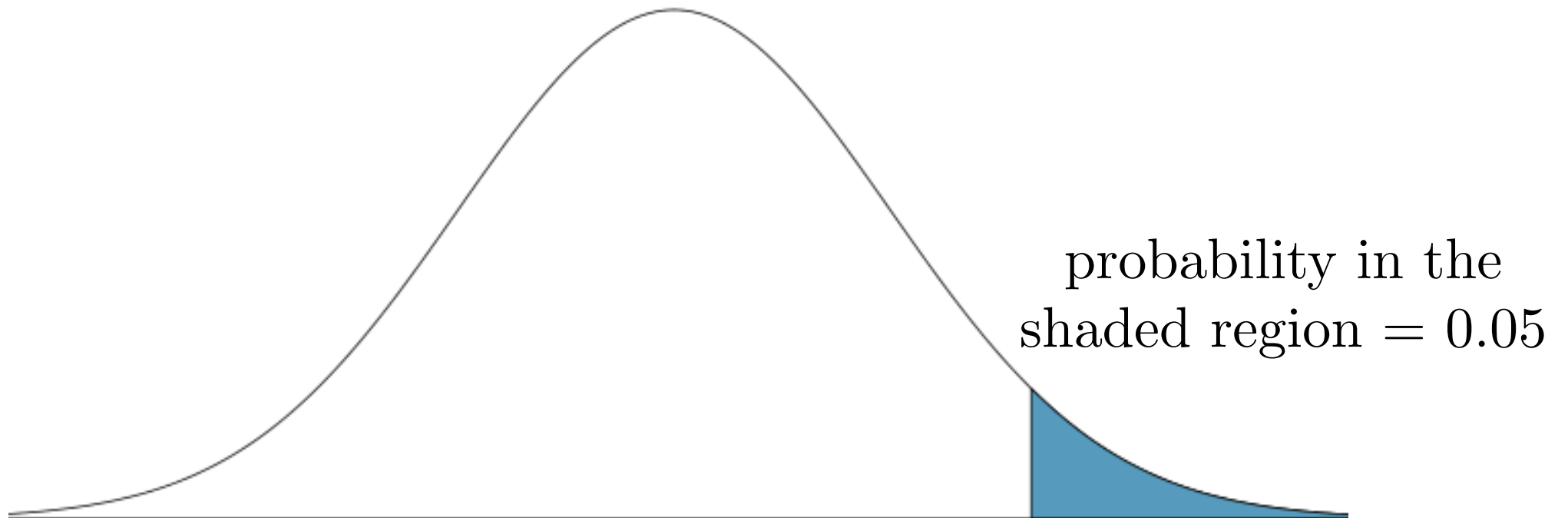
$$P(X < 95^{\text{th}} \text{ percentile}) = 0.95$$



# Percentiles

The probability that  $X$  is in the top 95<sup>th</sup> percentile is:

$$P(X > 95^{\text{th}} \text{ percentile}) = 1 - P(X < 95^{\text{th}} \text{ percentile}) = 0.05$$



# Back to Example 3

Amy scored 88, whereas Colin scored 95. The grades for both exams were normally distributed with:

Colin's midterm  $\approx N(77, 7)$

Amy's midterm  $\approx N(72, 6)$

What are Amy's and Colin's percentile scores?

$$\text{Amy's } z\text{-score : } Z = \frac{88 - 72}{6} \approx 2.67 \rightarrow P(Z < 2.67)$$

$$\text{Colin's } z\text{-score : } Z = \frac{95 - 77}{7} \approx 2.57 \rightarrow P(Z < 2.57)$$

[illegible]

$$P(Z < 2.57)$$

$$P(Z < 2.67)$$

[illegible]

$$P(Z < 2.57)$$

$$P(Z < 2.67)$$



[illegible]

$$P(Z < 2.57)$$

$$P(Z < 2.67)$$



[illegible]

$$P(Z < 2.57)$$

$$P(Z < 2.67)$$

$$P(Z < 2.67)$$

[illegible]

[illegible]

$$P(Z < 2.57) = 0.9949$$

$$P(Z < 2.67) = 0.9962$$



$$P(Z < 0.78)$$

[illegible]

$$P(Z < 0.78)$$

[illegible]

$$P(Z < 0.78)$$

[illegible]



$$P(Z < 0.78) = 0.7823$$

[illegible]

$$P(Z > 1.25)$$

[illegible]



$$P(Z < 0.78) = 0.7823$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

$$P(Z > 1.25) = 1 - P(Z < 1.25)$$

$$P(Z < 0.78) = 0.7823$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

$$P(Z > 1.25) = 1 - P(Z < 1.25)$$

$$P(Z < 0.78) = 0.7823$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

$$P(Z > 1.25) = 1 - P(Z < 1.25)$$



$$P(Z < 0.78) = 0.7823$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

$$P(Z > 1.25) = 1 - P(Z < 1.25)$$

$$= 1 - 0.8944$$

$$= 0.1056$$

# Example

The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in helping to avoid rear-end collisions. The article **“Fast-Rise Brake Lamp as a Collision-Prevention Device”** (*Ergonomics*, 1993: 391–395) suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled with a normal distribution having mean value 1.25 sec and standard deviation of 0.46 sec. What is the probability that reaction time is less than 1.75 sec? Let  $X$  denote reaction time.

$$P(X \leq 1.75)$$

$$P(X - 1.25 \leq 1.75 - 1.25)$$

$$P\left(\frac{X - 1.25}{0.46} \leq \frac{1.75 - 1.25}{0.46}\right)$$

$$P(Z \leq 1.09)$$

$$Z = \frac{X - 1.25}{0.46}$$

$$P(Z \leq 1.09)$$

[illegible]



[illegible]

$$P(Z \leq 1.09)$$

[illegible]

$$P(Z \leq 1.09)$$



[illegible]

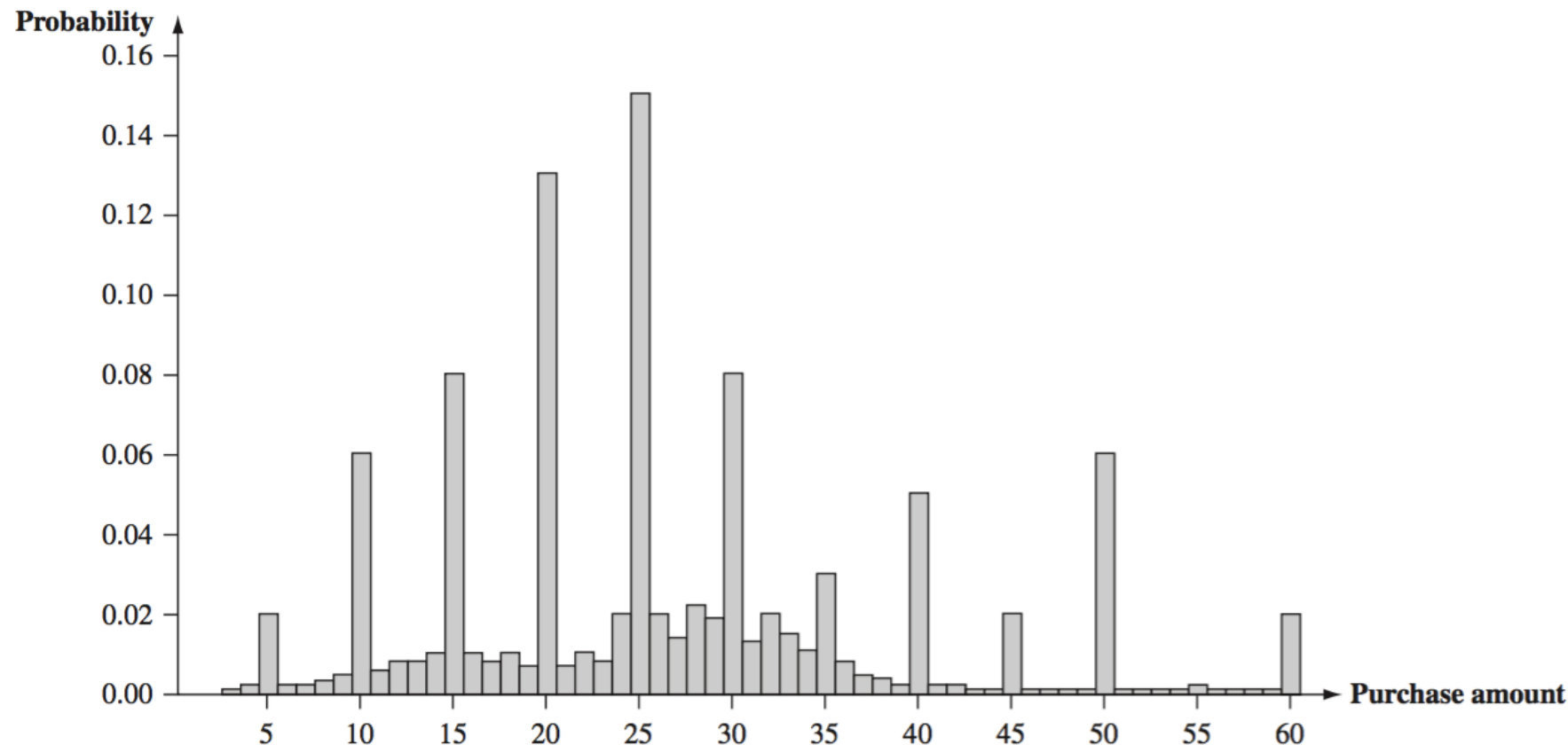
$$P(Z \leq 1.09) = 0.8621$$

# Why is the normal distribution so important?

- There aren't real data that resemble an exact normal distribution. However, there are many that are nearly normal. Example: the average height of women in the US
- Some distributions become nearly normal when we take a large number of samples. Example: the binomial distribution (this will be discussed briefly via an example)
- The most important thing about the normal distribution is how sample means are distributed. We will start with an example to explain what is meant by this statement.

# Example

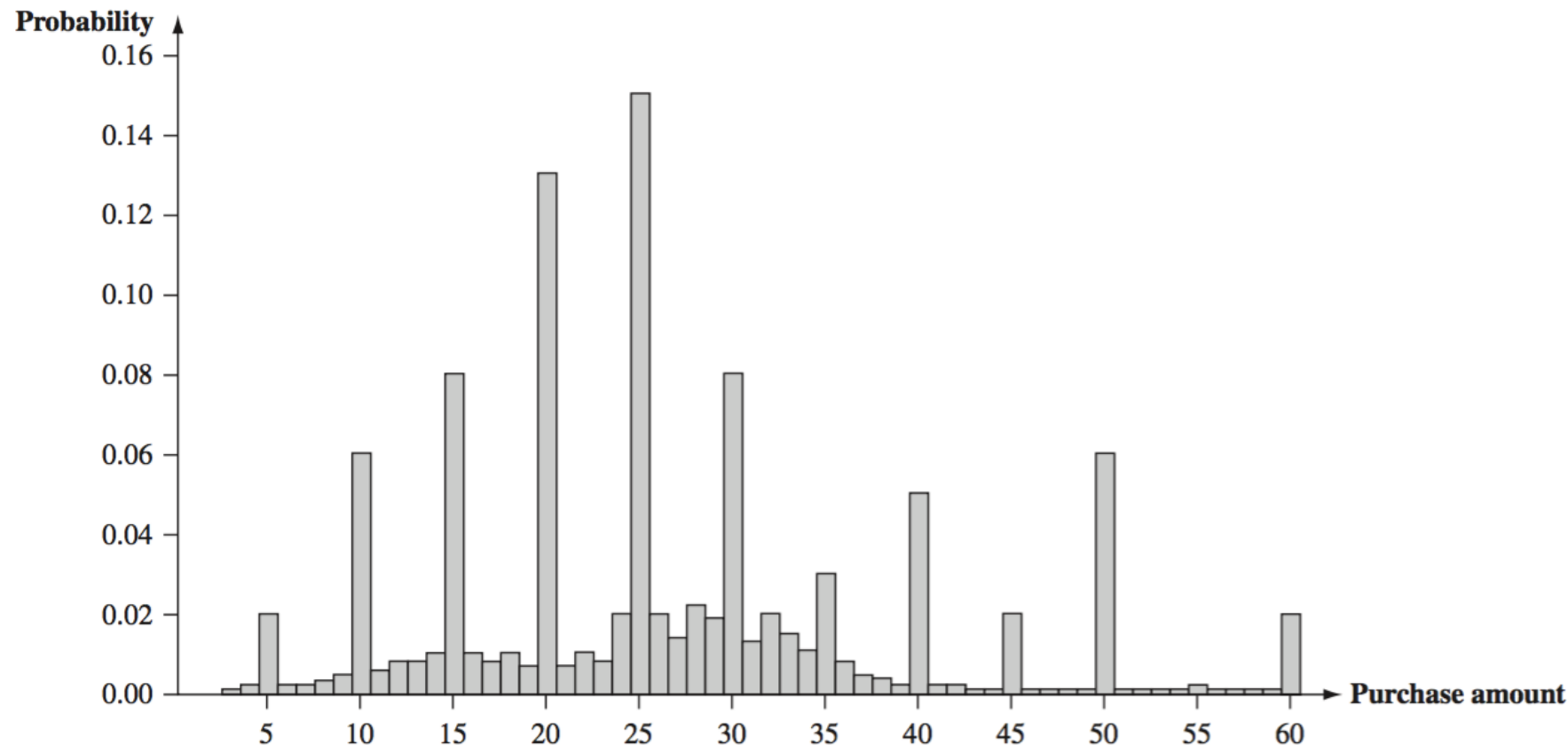
Consider the distribution shown for the amount purchased (rounded to the nearest dollar) by a randomly selected customer at a particular gas. The distribution will be our population distribution (or the underlying distribution).



# Example

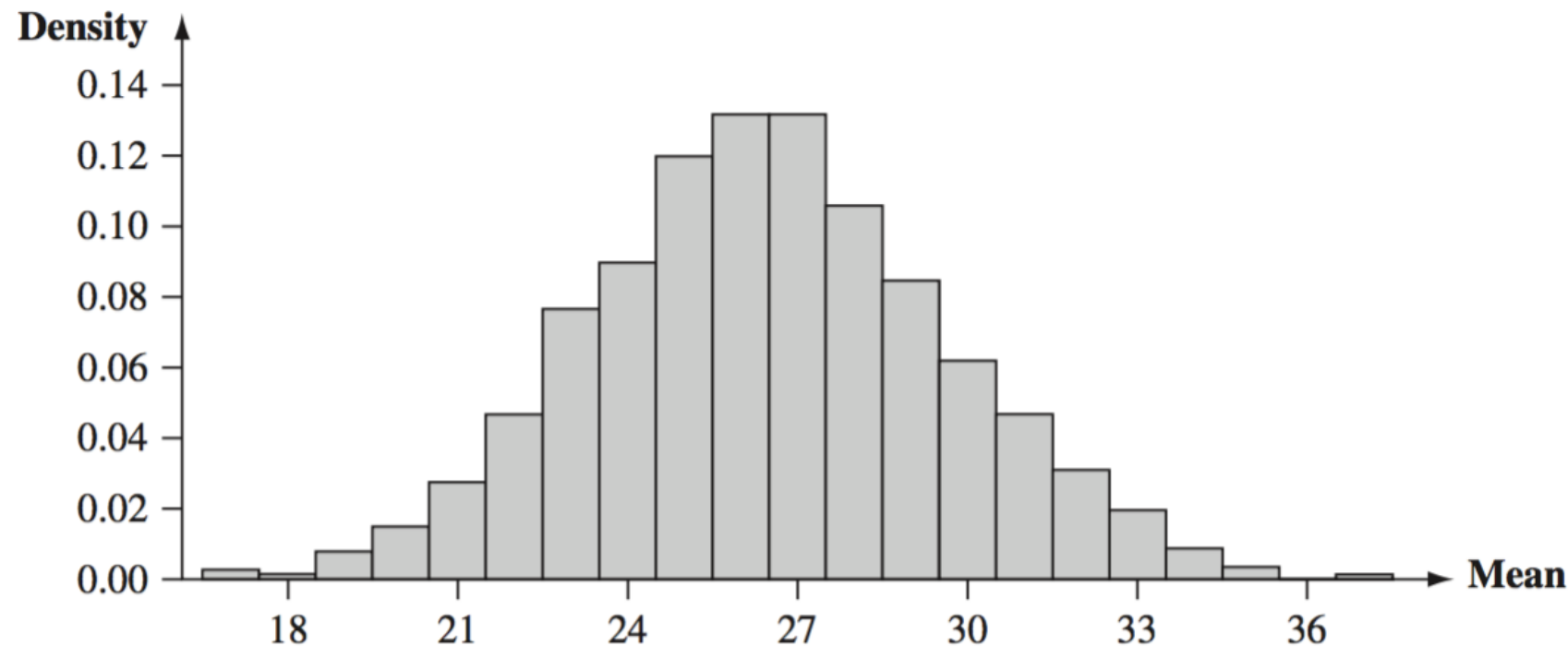
We ask a computer program to select 1000 samples from this distribution, where each sample has 15 observations. Then we ask the program to calculate the mean of each sample. For example, sample 1 mean will be denoted by  $\bar{x}_1$ .

Then we plot all the means and see how their distribution looks like.



# Example

It turns out that no matter what the underlying distribution is, if we take many many samples, the **sample means** will be **normally distributed**. This is called the central limit theorem (CLT), and it the most important theorem in statistics.



# Central Limit Theorem (CLT)

Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution with mean  $\mu$  and variance  $\sigma^2$ . Then if  $n$  is sufficiently large,  $\bar{X}$  has approximately a normal distribution with

$$\bar{x} = \mu \quad \text{and} \quad s^2 = \frac{\sigma^2}{n}$$

The larger the value of  $n$ , the better the approximation.

# Conditions for the CLT

- Sampled observations must be independent.
- If sampling without replacement,  $n < 10\%$  of population
- The underlying population distribution is either normal, or it is skewed, the sample size must be large (rule of thumb:  $n > 30$ )

# The binomial distribution

Recall the binomial coefficients

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- I have 5 elements, how many 0-element subsets can I create?  $\binom{5}{0} = \frac{5!}{0!(5-0)!} = 1$
- How many 1-element subsets?  $\binom{5}{1} = \frac{5!}{1!(5-1)!} = 5$
- How many 2-element subsets?  $\binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$
- How many 3-element subsets?  $\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$
- How many 4-element subsets?  $\binom{5}{4} = \frac{5!}{4!(5-4)!} = 5$
- How many 5-element subsets?  $\binom{5}{5} = \frac{5!}{5!(5-5)!} = 1$

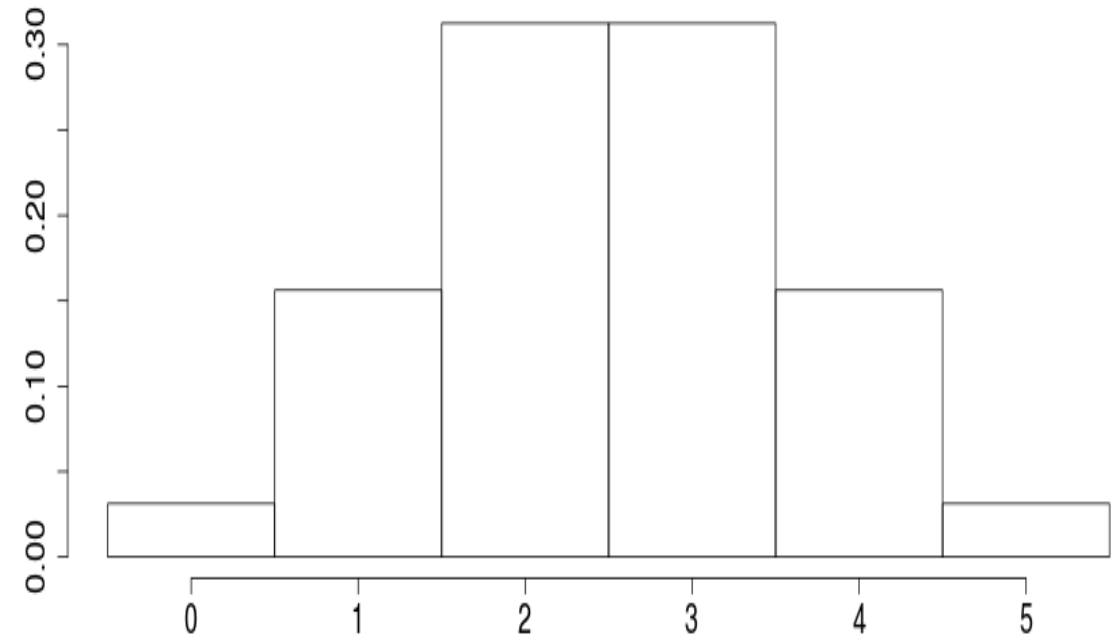
Total = 32



# The binomial distribution

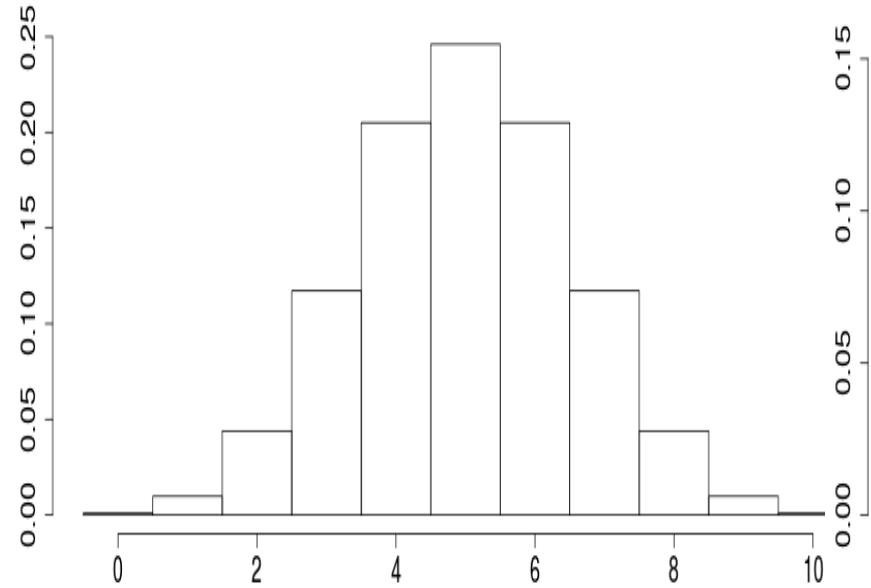
- $P(0\text{-element subsets}) = 1/32$
- $P(1\text{-element subsets}) = 5/32$
- $P(2\text{-element subsets}) = 10/32$
- $P(3\text{-element subsets}) = 10/32$
- $P(4\text{-element subsets}) = 5/32$
- $P(5\text{-element subsets}) = 1/32$

5 elements

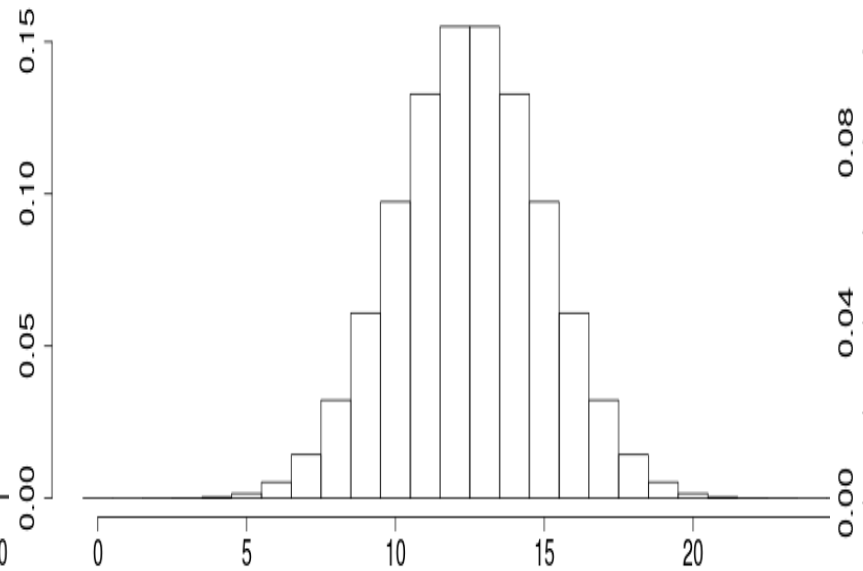


# The binomial distribution

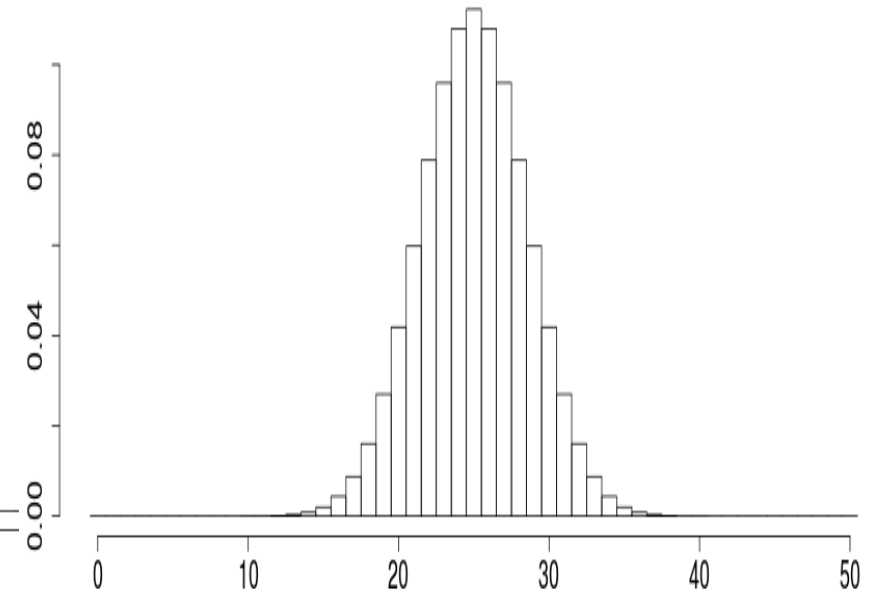
10 elements



25 elements



50 elements



# Simpson's Paradox

University of Alexandria: Thracians got angry at the Admissions Office, because they found out that 4400 Dacians had applied and 4400 Thracians had also applied, but 3280 Dacians were admitted while only 1120 Thracians were admitted.

Nationality	Applicants	Admitted	Percent Admitted
Dacian	4400	3280	74.55 %
Thracian	4400	1120	25.45 %

# Simpson's Paradox

Someone pointed out to the Thracians that there were two academic divisions in University of Alexandria, the Trivium (Grammar, Logic, and Rhetoric), and the Quadrivium (Arithmetic, Geometry, Astronomy, and Music).

Nationality	Triv. Applicants	Triv. Acceptances	Triv. % Admitted
Dacians	4000	3200	80 %
Thracians	400	320	80 %
Nationality	Quadriv. Applicants	Quadriv. Acceptances	Quadriv. % Admitted
Dacians	400	80	20 %
Thracians	4000	800	20 %

# Simpson's Paradox

Then a clever Dacian realized that the Trivium has two departments, one handling Grammar and Logic (G&L) and the other handling Rhetoric.

## Trivium Programs

Nationality	G&L Applicants	G&L Acceptances	G&L % Admitted
Dacians	3600	3060	85 %
Thracians	100	95	95 %
Nationality	Rhetoric Applicants	Rhetoric Acceptances	Rhetoric % Admitted
Dacians	400	140	35 %
Thracians	300	225	75 %

# Simpson's Paradox

Likewise, the Quadrivium is split into two departments: Arithmetic, Geometry, and Astronomy, and the Department of Music.

Quadrivium Programs			
Nationality	AGA Applicants	AGA Acceptances	AGA % Admitted
Dacians	300	75	25 %
Thracians	400	260	65 %
Nationality	Music Applicants	Music Acceptances	Music % Admitted
Dacians	100	5	5 %
Thracians	3600	540	15 %