Today's plan:

- Section 4.2: Normal Distribution

Characteristics of a data set:

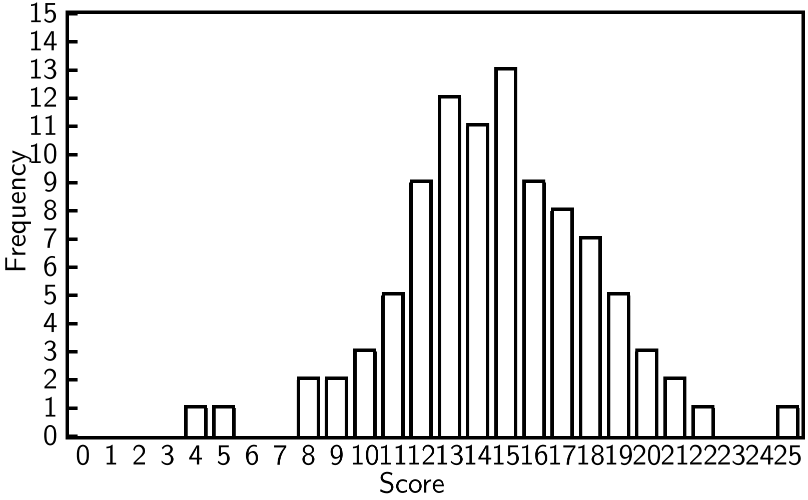
- mean
- median
- standard deviation
- five-number summary

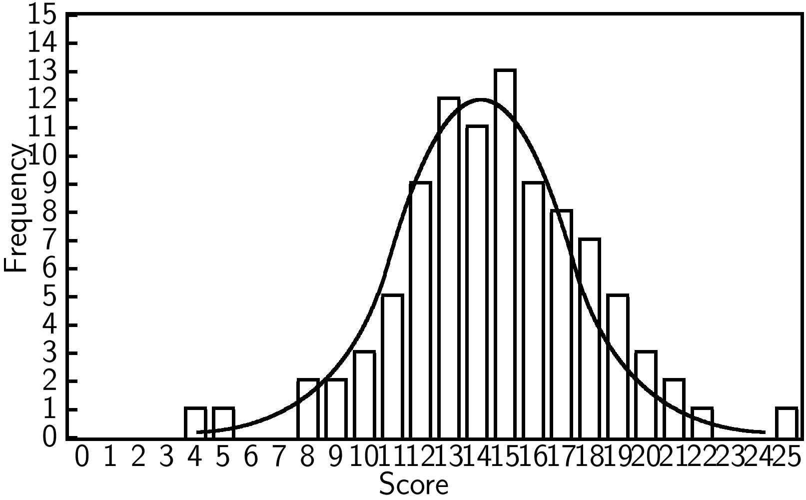
Characteristics of a data set:

- mean
- median
- standard deviation
- five-number summary

These don't always tell you everything though.

Bar graph for scores of Math 109 quiz:


Frequency
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
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
Score



It's somewhat close to a bell-shaped curve.

Definition

- We call that bell-shaped curve a **normal curve**.

Definition

- We call that bell-shaped curve a **normal curve**.
- A data set whose distribution is a normal curve has a **normal distribution**.

The quiz scores in Math 109 have a **near-normal** distribution:

- it's close, but not super close to the normal curve

The quiz scores in Math 109 have a **near-normal** distribution:

- it's close, but not super close to the normal curve
- Note the outliers

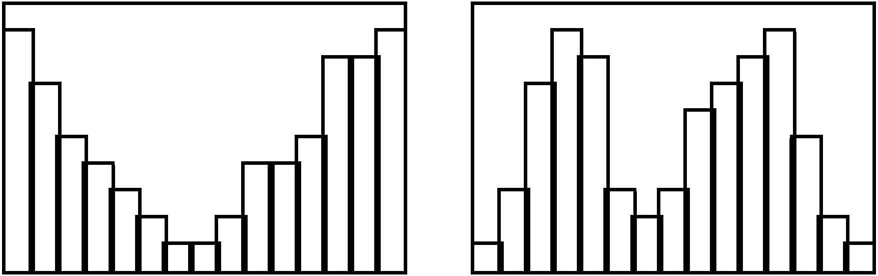
The quiz scores in Math 109 have a **near-normal** distribution:

- it's close, but not super close to the normal curve
- Note the outliers
- the spike at 15

The quiz scores in Math 109 have a **near-normal** distribution:

- it's close, but not super close to the normal curve
- Note the outliers
- the spike at 15
- the whole picture is a little skewed to the right

Non-normal distributions:

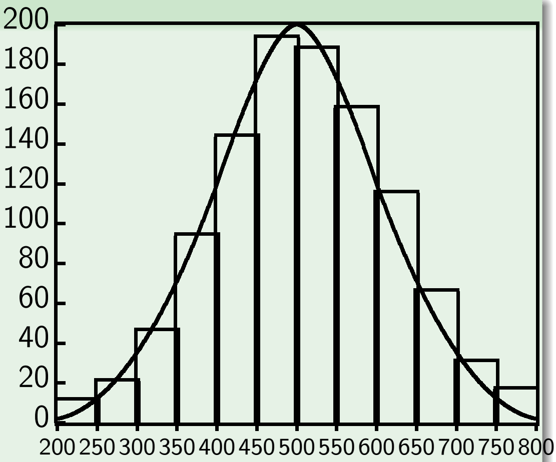


- The first one has an inverted bell shape.
- The second one is called a bimodal distribution.

Lots of data sets do have normal or near-normal distributions though, so we're interested in them.

Example

Score	Freq.
[200, 250)	11,393
[250, 300)	20,874
[300, 350)	46,583
[350, 400)	94,037
[400, 450)	143,630
[450, 500)	193,470
[500, 550)	187,773
[550, 600)	158,138
[600, 650)	115,401
[650, 700)	66,066
[700, 750)	30,503
[750, 800]	16,857



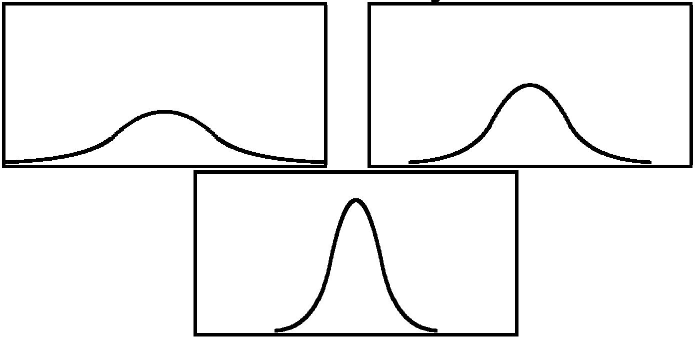
This is a normal distribution. (SAT scores)

Section 4.2.2: Properties of Normal Distributions

- Normal distributions **approximate** real life distributions

- Normal distributions **approximate** real life distributions
- Knowing the properties of normal curves, we can approximate the properties of near-normal data sets.

Normal curves can vary:



but they all have common properties.

Symmetry A normal curve is symmetric about the vertical line at the mean μ. This is called its **axis of symmetry**.

Concavity A normal curve has a concave up region on the left, a concave down region in the middle, and a concave up region on the right.

Inflection points The points where the curve changes concavity are called inflection points.

Inflection points The points where the curve changes concavity are called inflection points.

- There are two inflection points, each one standard deviation from the center.

68-95-99.7 Principle Tells where the data points should be:

68-95-99.7 Principle Tells where the data points should be:

- 68% of the data set is located under the concave down part of the curve. That's within σ of μ.

68-95-99.7 Principle Tells where the data points should be:

- 68% of the data set is located under the concave down part of the curve. That's within σ of μ.
- 95% of the data points are within 2 σ's of μ.

68-95-99.7 Principle Tells where the data points should be:

- 68% of the data set is located under the concave down part of the curve. That's within σ of μ.
- 95% of the data points are within 2 σ's of μ.
- 99.7% of the data points are within 3 σ's of μ.

Quartiles

$Q1 = \mu - 0.675\sigma$ and

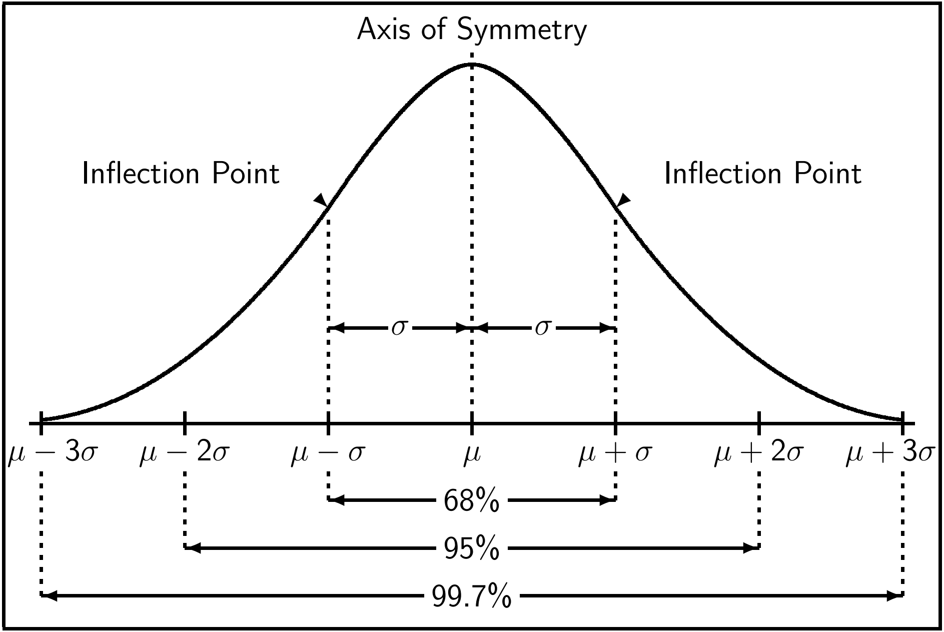
$$Q3 = \mu + 0.675\sigma$$

Quartiles

$Q1 = \mu - 0.675\sigma$ and

$$Q3 = \mu + 0.675\sigma$$

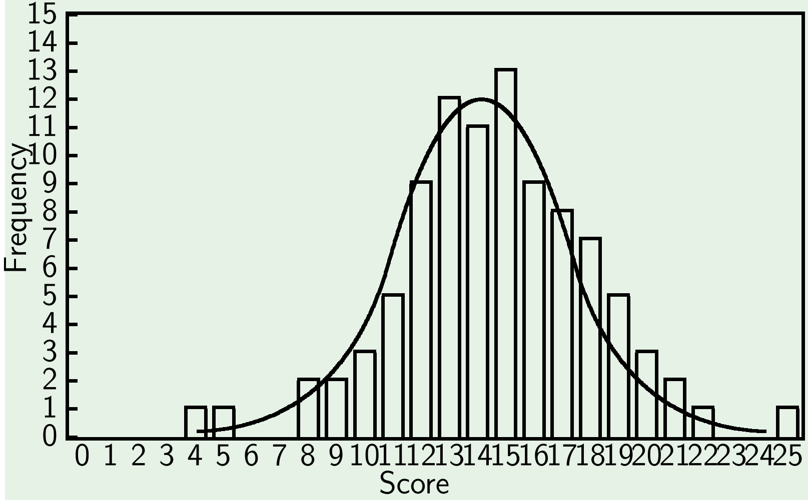
So 50% of data points are within 0.675σ of μ.


Axis of Symmetry
Inflection Point
Inflection Point
σ
σ
$\mu - 3\sigma$
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$
$\mu + 3\sigma$
68%
95%
99.7%

For near-normal distributions, the above properties should be taken as approximations.

Example

Recall Math 109 quiz:



score	4	5	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	25
freq.	1	1	2	2	3	5	9	12	11	13	9	8	7	5	3	2	1	1
cum fr.	1	2	4	6	9	14	23	35	46	59	68	76	83	88	91	93	94	95

We found earlier:

- $\mu = 14.64$

score	4	5	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	25
freq.	1	1	2	2	3	5	9	12	11	13	9	8	7	5	3	2	1	1
cum fr.	1	2	4	6	9	14	23	35	46	59	68	76	83	88	91	93	94	95

We found earlier:

- $\mu = 14.64$
- $\sigma = 3.35$

score	4	5	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	25
freq.	1	1	2	2	3	5	9	12	11	13	9	8	7	5	3	2	1	1
cum fr.	1	2	4	6	9	14	23	35	46	59	68	76	83	88	91	93	94	95

We found earlier:

- $\mu = 14.64$
- $\sigma = 3.35$
- 62 out of 95 scores are within one σ from μ

$$\frac{62}{95} \cdot 100\% = 65.3\%$$

score	4	5	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	25
freq.	1	1	2	2	3	5	9	12	11	13	9	8	7	5	3	2	1	1
cum fr.	1	2	4	6	9	14	23	35	46	59	68	76	83	88	91	93	94	95

We found earlier:

- $\mu = 14.64$
- $\sigma = 3.35$
- 62 out of 95 scores are within one σ from μ

$$\frac{62}{95} \cdot 100\% = 65.3\%$$

which is close to the ideal 68%

- number of scores within 2 σ's of μ:

- number of scores within 2 σ's of μ:

$$\mu - 2\sigma = 7.94 \quad \text{and} \quad \mu + 2\sigma = 21.34,$$

- number of scores within 2 σ's of μ:

$$\mu - 2\sigma = 7.94 \quad \text{and} \quad \mu + 2\sigma = 21.34,$$

so 91 out of 95 (i.e., 95.8%)

- number of scores within 2 σ's of μ:

  $\mu - 2\sigma = 7.94 \quad \text{and} \quad \mu + 2\sigma = 21.34,$

  so 91 out of 95 (i.e., 95.8%), close to the ideal 95%.

- number of scores within 2 σ's of μ:

  $$\mu - 2\sigma = 7.94 \quad \text{and} \quad \mu + 2\sigma = 21.34,$$

  so 91 out of 95 (i.e., 95.8%), close to the ideal 95%.
- Finally, 93 out of 95 (i.e., 97.9%) points are within 3 σ's of μ.

- number of scores within 2 σ's of μ:

  $$\mu - 2\sigma = 7.94 \quad \text{and} \quad \mu + 2\sigma = 21.34,$$

  so 91 out of 95 (i.e., 95.8%), close to the ideal 95%.
- Finally, 93 out of 95 (i.e., 97.9%) points are within 3 σ's of μ.
  All of this is close to the ideal 68-95-99.7 principle.

We can also estimate the quartiles for the Math 109 quizzes:

We can also estimate the quartiles for the Math 109 quizzes:

- $Q1 \approx \mu - 0.675\sigma = 12.38$
- $Q3 \approx \mu + 0.675\sigma = 16.90$

(Faster than actually calculating them.)

Conclusion: If you know that the distribution of data points is normal, it is enough to know just 2 numbers to describe the data set completely:

Conclusion: If you know that the distribution of data points is normal, it is enough to know just 2 numbers to describe the data set completely:

- Mean μ

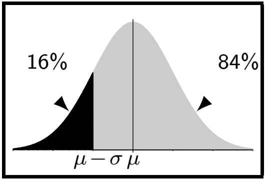
Conclusion: If you know that the distribution of data points is normal, it is enough to know just 2 numbers to describe the data set completely:

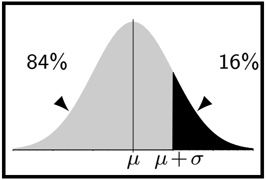
- Mean μ
- Standard deviation σ

Section 4.2.3: Using the 68-95-99.7 Principle

- Each tail outside one σ contains

$$\frac{100\% - 68\%}{2} = 16\%$$





This means that a random point has:

This means that a random point has:

- 16% chance of being below $\mu - \sigma$

This means that a random point has:

- 16% chance of being below $\mu - \sigma$
- 84% chance of being above $\mu - \sigma$

This means that a random point has:

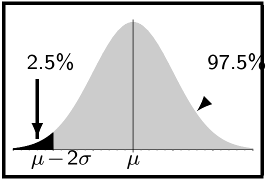
- 16% chance of being below $\mu - \sigma$
- 84% chance of being above $\mu - \sigma$

$$\mathrm{pr}(x < \mu - \sigma) = 0.16 \quad \text{and}$$

$$\mathrm{pr}(x \geq \mu - \sigma) = 0.84.$$

Similarly, each tail outside 2σ contains

$$\frac{100\% - 95\%}{2} = 2.5\%$$



Example

The weights of 6-month-old baby boys in the U.S. have a near-normal distribution with

- $\mu = 17.25$ lbs
- $\sigma = 2$ lbs

Example

Questions:

- **(a)** What can we say about babies whose weight is 19.25 lb.?

Example

Questions:

- **(a)** What can we say about babies whose weight is 19.25 lb.?
- **(b)** If we pick a random baby, what's the probability he's less than 13.25 lb.?

Solution

(a) *The baby whose weight is 19.25 lb.*

Solution

(a) *The baby whose weight is 19.25 lb.*

- *is certainly above the average*

Solution

(a) *The baby whose weight is 19.25 lb.*

- *is certainly above the average*
- *Moreover, his weight*

$$19.25 = 17.25 + 2 = \mu + \sigma$$

is precisely one standard deviation above the average.

Solution

Thus

- *84% of babies weigh less than him*

Solution

Thus

- *84% of babies weigh less than him*
- *16% of babies weigh more than him*

We say that his weight is in the 84th **percentile**.

Solution

(b) *A weight of 13.25 lb. or less places a baby at least 2 standard deviations below the average.*

$$13.25 = 17.25 - 2 \cdot 2 = \mu - 2\sigma$$

Solution

(b) *A weight of 13.25 lb. or less places a baby at least 2 standard deviations below the average.*

$$13.25 = 17.25 - 2 \cdot 2 = \mu - 2\sigma$$

The probability for that is 2.5%.

Next Time:

- Section 4.3: Data Collection
- Section 4.3.1: Population v. Sample