Lecture 8: Normal Distribution

Chapter 3.1

Goals for Today

- Define the normal distribution in terms of its parameters
- Review: $\frac{2}{3}$ / 95% / 99.7% rule
- Standardizing normal observations to z-scores

Normal Distribution

From text page 118:

Many variables are nearly normal, but none are exactly normal. Thus the normal distribution, while not perfect for any single problem, is very useful for a variety of problems.

We will use it in data exploration and to solve important problems in statistics.

Normal Distribution

Normal distributions:

1. are symmetric
2. are unimodal and bell-shaped
3. have area under the curve 1

Normal Distribution

A normal curve can be described by two parameters:

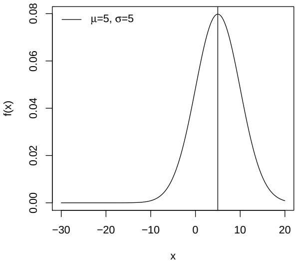
- the mean μ. i.e. the center
- the standard deviation (SD) σ. i.e. the measure of spread

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

Recall these were the population mean and the population SD.

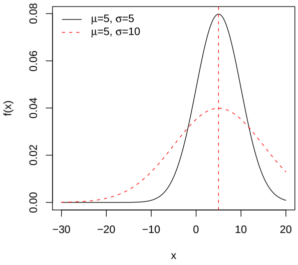
Normal Distribution

μ (mean) specifies the center, σ (standard deviation) the spread.



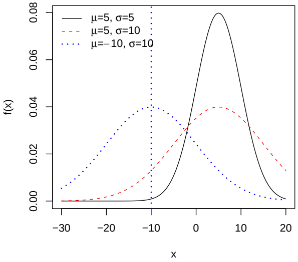
Normal Example

μ (mean) specifies the center, σ (standard deviation) the spread.



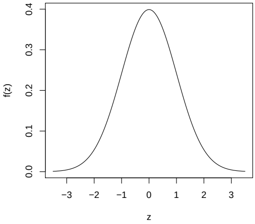
Normal Example

μ (mean) specifies the center, σ (standard deviation) the spread.



Standardized Normal Distribution

If $\mu = 0$ and $\sigma = 1$, this is the standard normal distribution:

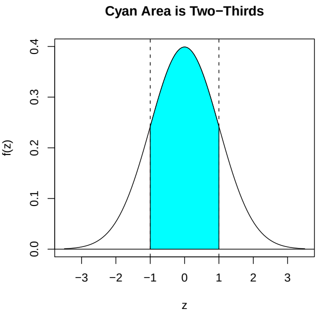


Rules of Thumb

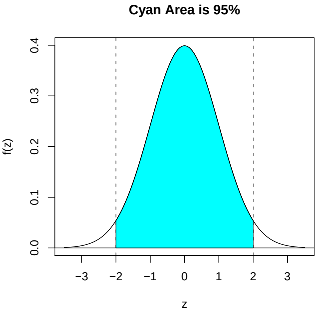
Recall if a distribution is normal, then:

1. Approx. $\frac{2}{3}$'s of the data are within ± 1 SD of the mean
2. Approx. 95% of the data are within ± 2 SD of the mean
3. Also approx. 99.7% of the data are within ± 3 SD of the mean

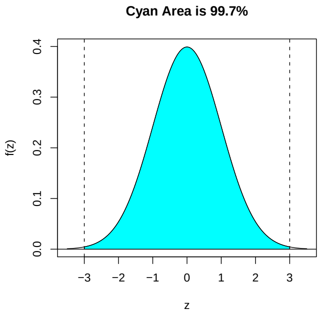
Ex: Standard Normal $\mu = 0$, $\sigma = 1$



Ex: Standard Normal $\mu = 0$, $\sigma = 1$



Ex: Standard Normal $\mu = 0$, $\sigma = 1$

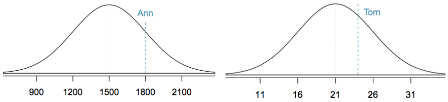


Motivating Example

From text: Say Ann scores 1800 on the SAT and Tom scores 24 on the ACT. Say both tests scores were normally distributed with:

	SAT	ACT
Mean μ	1500	21
SD σ	300	5

Question: Who did relatively better?



z-scores

The z-score AKA standardized observation of an observation x is the number of SD it falls above or below the mean.

The z-score for an observation x that follows a distribution with mean μ and SD σ:

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

z-scores

Why is the z-score $z = \frac{x-\mu}{\sigma}$ called the standardized observation?

1. The observations are centered at μ.
   re-center the x observations to 0 by subtracting μ.
2. The observations have spread σ.
   re-scale the spread of the $x - \mu$ values to be 1 by dividing by σ.

So we can compare observations from any normally distributed data with (μ, σ)

i.e. we've standardized the observations to make them comparable.

Back to Example

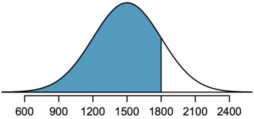
- Ann scored 1800. $z = \frac{1800-1500}{300} = +1$ standard deviation from the mean
- Tom scored 24. $z = \frac{24-21}{5} = +0.6$ standard deviation from the mean

So Ann did relatively better.

Percentiles

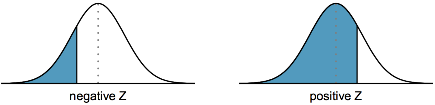
Recall a percentile (%'ile) indicates the value below which a given %'age of observations fall below.

Question: What %'ile is Ann's SAT score of 1800?
i.e. what is the blue shaded area?



Percentiles

Because the total area under the curve is 1, the area to the left of z represents the %'ile of the observation:



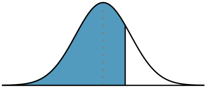
- The blue shaded area on the left plot will be less than 0.5. We have %'iles less than the 50th %'ile.
- The blue shaded area on the right plot will be greater than 0.5. We have %'iles greater than the 50th %'ile.

Normal Probability Table

A normal probability table allows you to:

- identify the %'ile corresponding to a z-score
- or vice versa: the z-score corresponding to a %'ile

The normal probability tables on page 409 represent z-scores and %'iles corresponding to area to the left:



Normal Probability Table

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- Red case: Given a z-score of 0.43. A lookup tells us the area to the left of z=0.43 is 0.6664, i.e. the 66th %'ile
- Blue case: We want the z-score that is the 80th %'ile. Reverse lookup: the closest value on the table is 0.7995, i.e. a z-score of 0.84.

Back to Ann and Tom

- Since Ann had a z-score of 1.0, her %'ile is 0.8413. (1.0 row, 0.00 column)
  i.e. She did better than 84.13% of SAT test takers.
- Since Tom had a z-score of 0.6, his %'ile is 0.7257. (0.6 row, 0.00 column)
  i.e. He did better than 72.57% of ACT test takers

Next Time

Next time we will:

- Re-iterate the motivation for the normal curve.
- Go over examples using z-scores.
- Evaluating the normal approximation.