The Normal Distribution
Part 2: Standardization and Percentiles

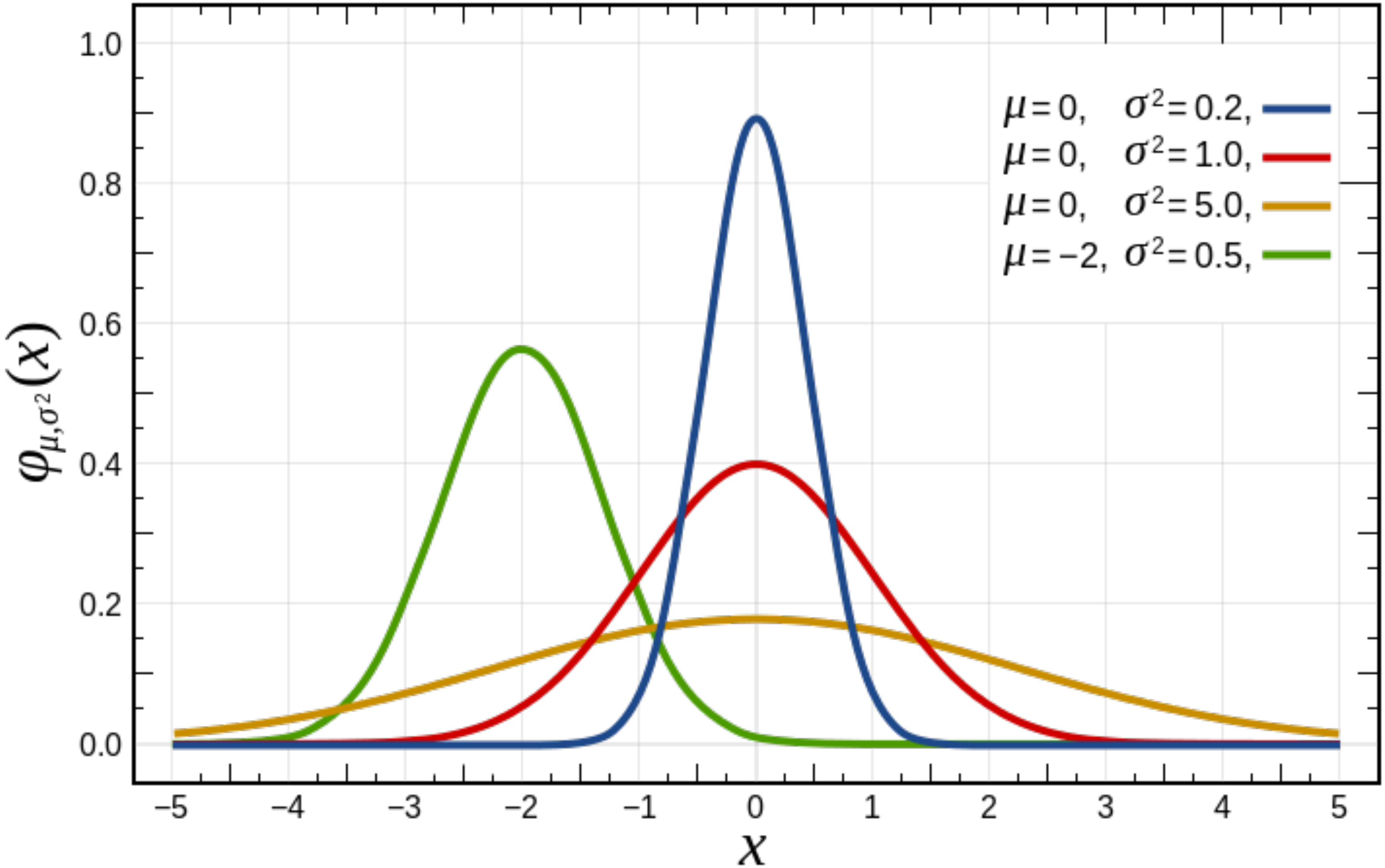
INFO-1301, Quantitative Reasoning 1

University of Colorado Boulder

March 20, 2017

Prof. Michael Paul

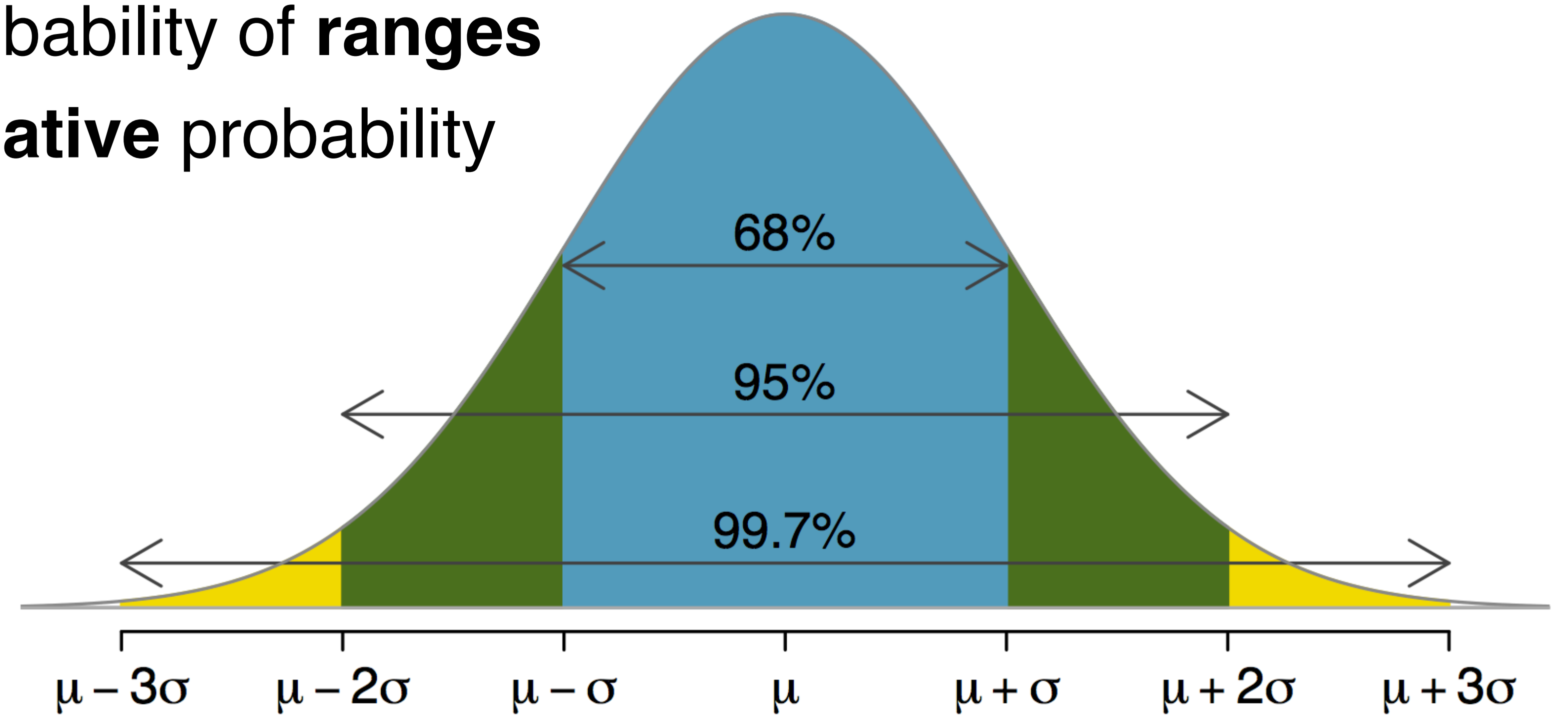
Normal Distribution



What can we do with this?

If the normal distribution is a good approximation, then we can use the math of the probability density to answer questions about the data:

- Probability of **ranges**
- **Relative** probability



Data Standardization

If your data is approximately normal, a useful way to describe a value is by how many standard deviations it is from the mean

- “0.5 standard deviations below average”
- “2.2 standard deviations above average”

Data Standardization

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

- Called the “Z-score”
- If Z(x) is negative, x is to the left of the mean.
- If Z(x) = 0, x is the mean.
- If Z(x) is positive, x is to the right of the mean.
- The absolute value of Z(x) is the number of standard deviations x is above or below the mean.

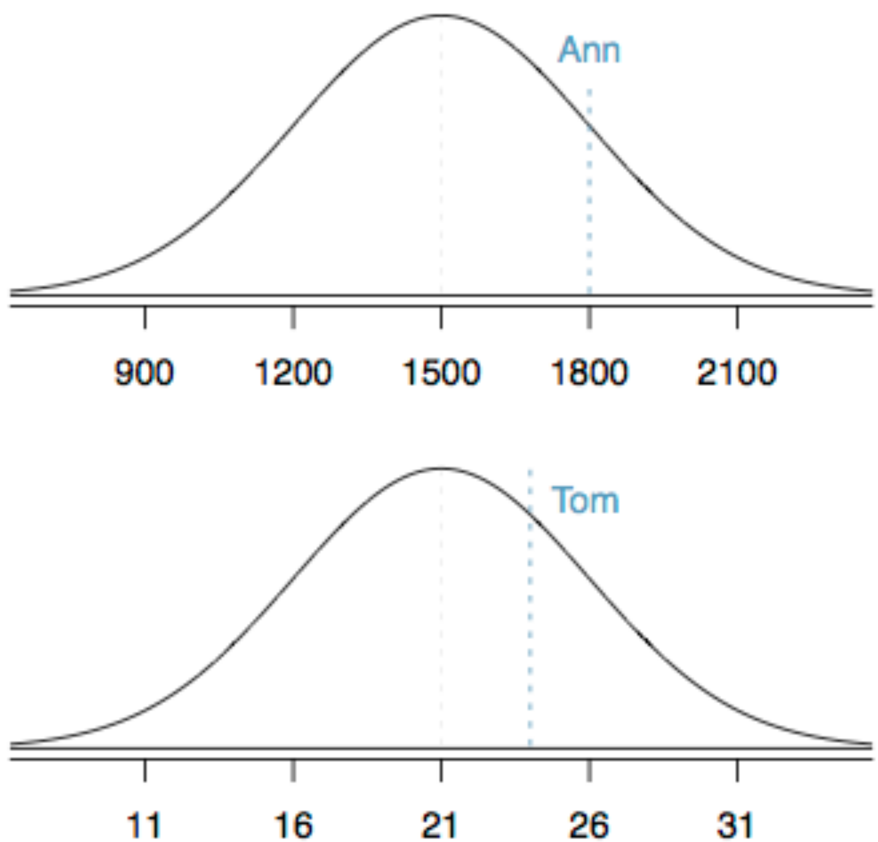
Data Standardization

Z-scores allow you to compare data values that are on different scales

- The normal distribution of the SAT is given by N(1500, 300), while the normal distribution of the ACT is given by N(21, 5).
- Ann gets an 1800 on the SAT, while Tom gets a 24 on the ACT. Which student did better on the college entrance exam?
  - Compare their Z-scores

Data Standardization

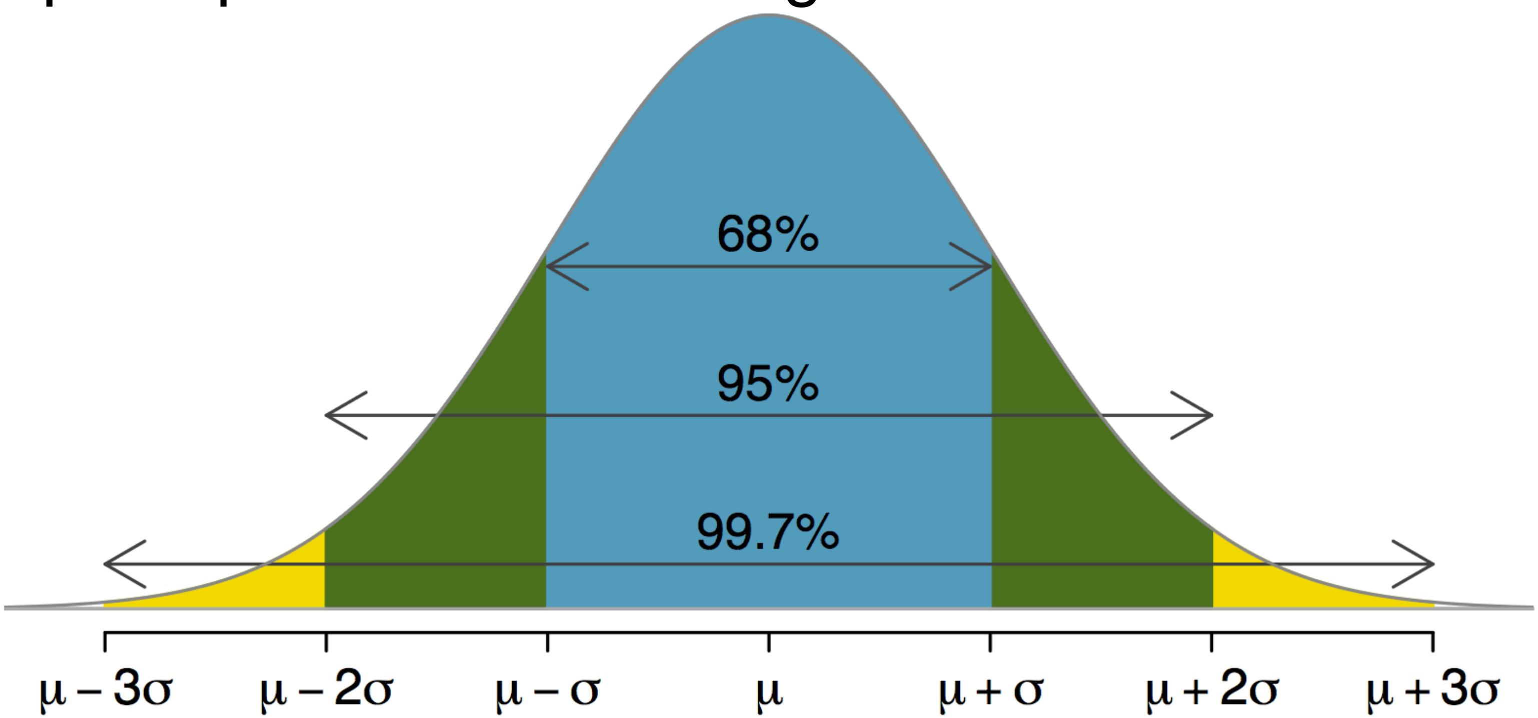
Z-scores allow you to compare data values that are on different scales



Percentiles and Ranges

The normal distribution can tell you the probability that data falls within a certain range

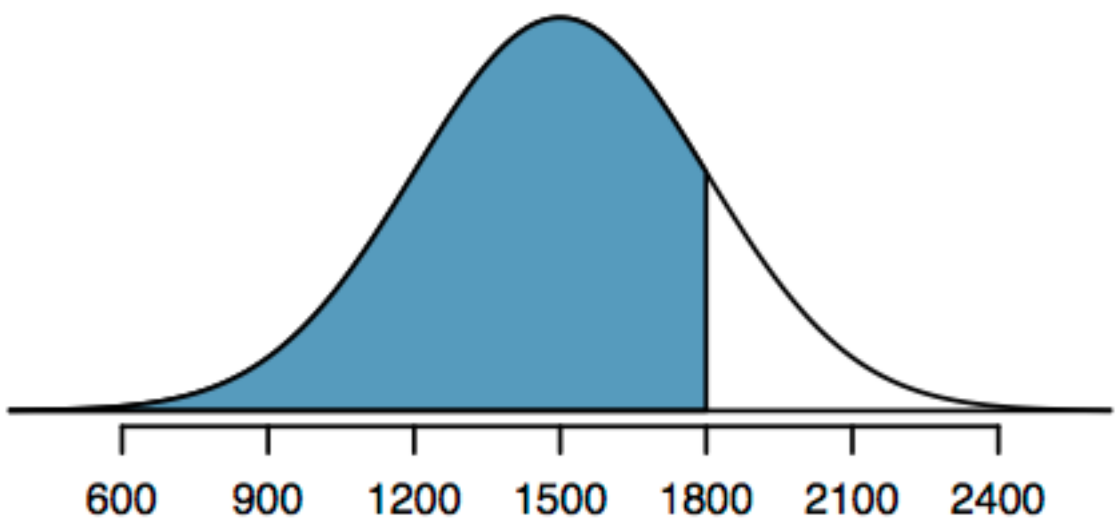
- If you convert your data to Z-scores, you can look up the probabilities of ranges for a standard normal



Normal Probability Table

A standard normal probability table can tell you the percentile of a given Z-score

- That is: what % of data is below a certain Z-score
- This corresponds to the **area** under the normal curve



Normal Probability Table

- Appendix B.1
- Page 427-429

Z	Second decimal place of 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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 3.8: A section of the normal probability table. The percentile for a normal random variable with $Z = 0.43$ has been *highlighted*, and the percentile closest to 0.8000 has also been **highlighted**.

Normal Probability Table

Special case: Z-score of 0 is the 50th percentile

(Remember: 50th percentile is the same as the median. Because normal distributions are symmetric, the mean and median are the same.)

Percentiles and Ranges

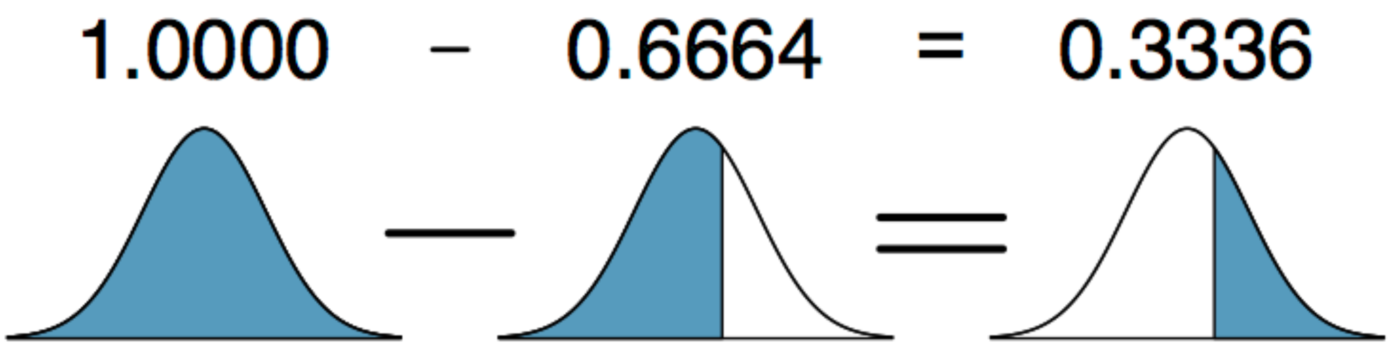
Use the normal probability table to answer questions about percentiles and the probability of a range of values (after converting to z-scores)

- The normal distribution of the SAT is given by N(1500, 300), while the normal distribution of the ACT is given by N(21, 5).
- Ann gets an 1800 on the SAT, while Tom gets a 24 on the ACT. What is Ann’s percentile? What is Tom’s percentile?

Percentiles and Ranges

The percentiles from the table tell you the probability that data is below the value

- Need to do some arithmetic if you want the probability above a value or between two values



Percentiles and Ranges

The percentiles from the table tell you the probability that data is below the value

- Need to do some arithmetic if you want the probability above a value or between two values

Percentiles and Ranges

Head lengths of brushtail possums follow a nearly normal distribution with mean 92.6 mm and standard deviation 3.6 mm.

What is the probability that a randomly selected possum has a head length larger than 93mm?

Larger than 100?

Larger than 110?

Percentiles and Ranges

What percentage of data is within 1 standard deviation of the mean?

Within 2 standard deviations?

Within 3 standard deviations?

68-95-99 rule