JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

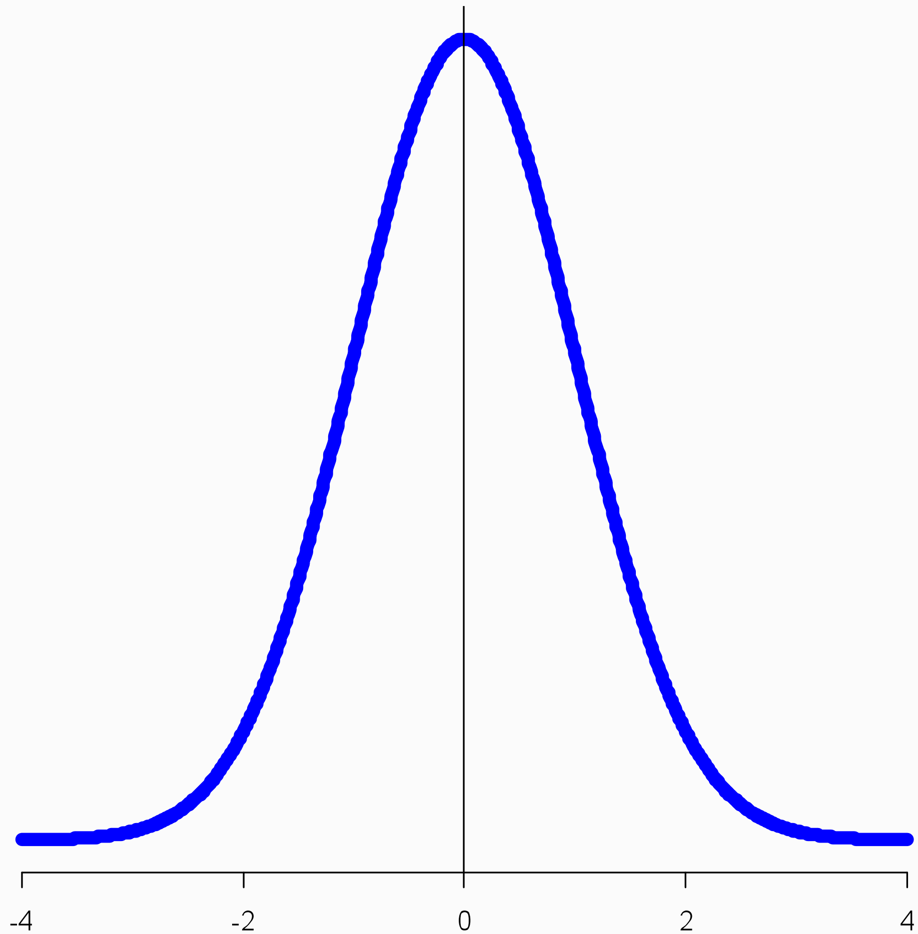
JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section B

Variability in the Normal Distribution: Calculating Normal Scores

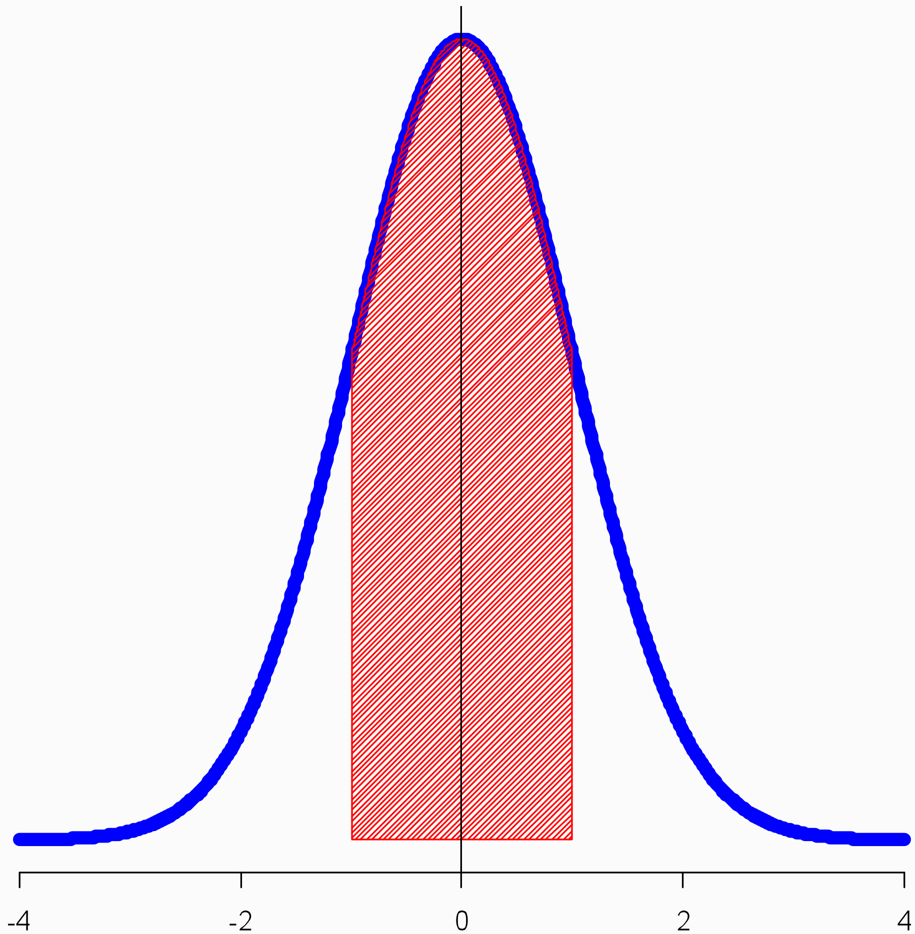
The Standard Normal Distribution

- The standard normal distribution has a mean of 0, and standard deviation of 1



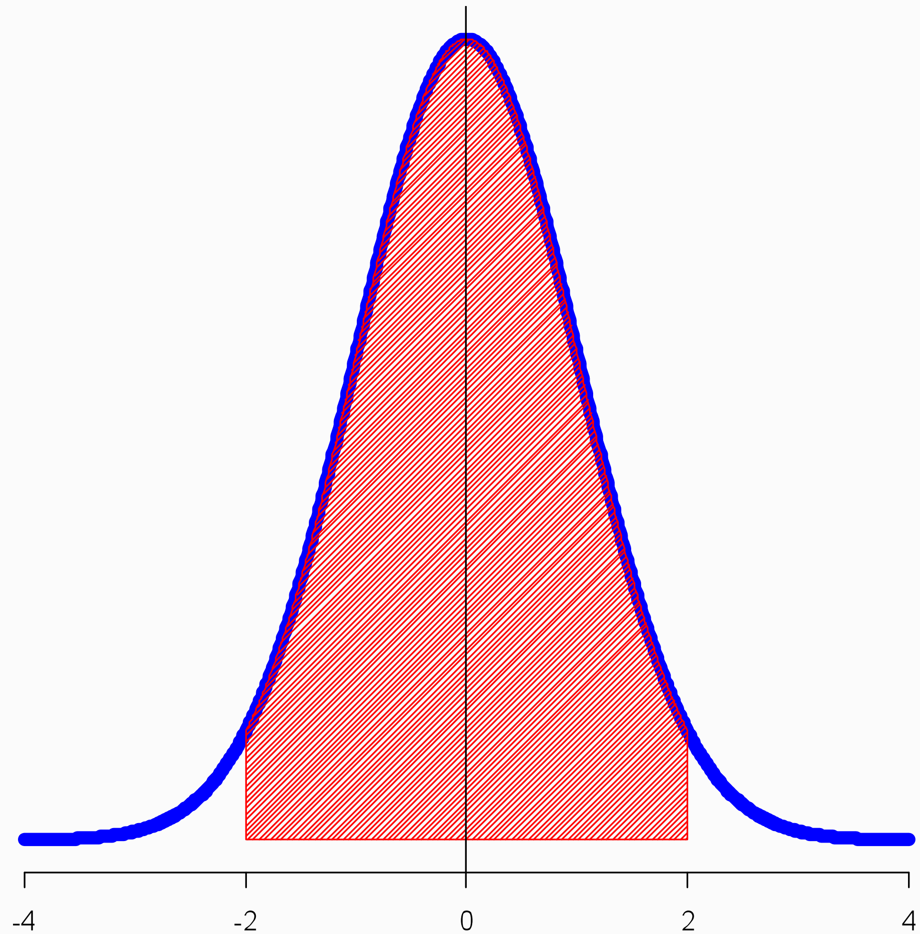
The 68-95-99.7 Rule for the Normal Distribution

- 68% of the observations fall within one standard deviation of the mean



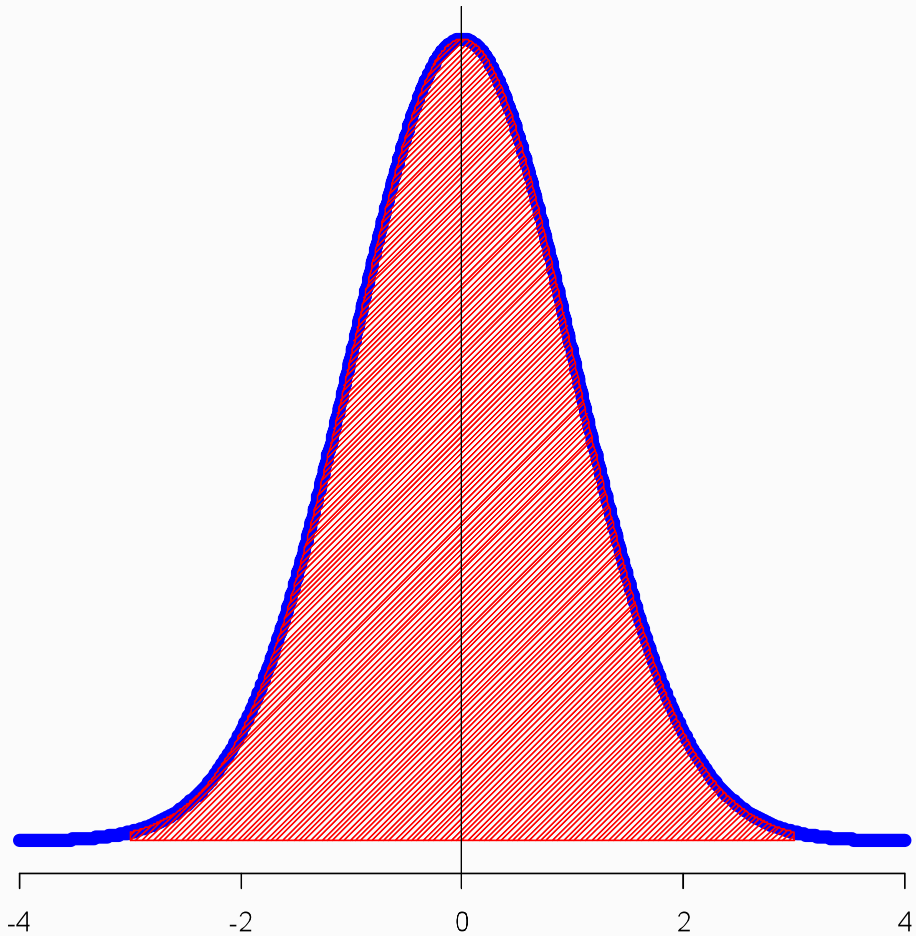
The 68-95-99.7 Rule for the Normal Distribution

- 95% of the observations fall within two standard deviations of the mean (truthfully, within 1.96)

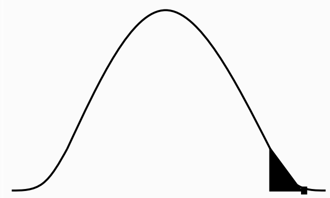
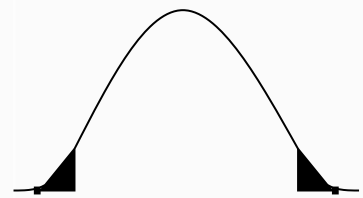


The 68-95-99.7 Rule for the Normal Distribution

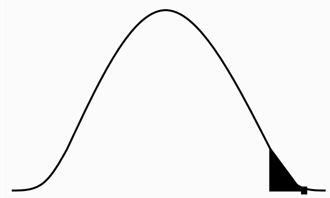
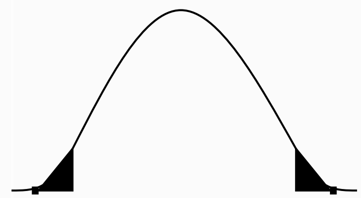
- 99.7% of the observations fall within three standard deviations of the mean



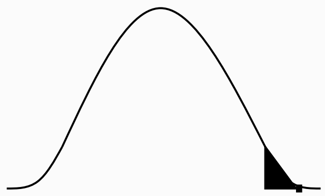
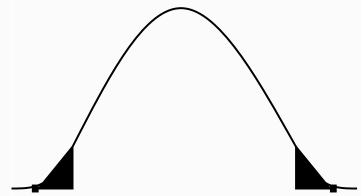
Fraction of Observations under Standard Normal

	Within Z SDs of the mean	**More than Z SDs above the mean**	**More than Z SDs above or below the mean**
Z			
1.0	68.27%	15.87%	31.73%
2.0	95.45%	2.28%	4.55%
2.5	98.76%	0.62%	1.24%
3.0	99.73%	0.13%	0.27%

Fraction of Observations under Standard Normal

	Within Z SDs of the mean	More than Z SDs above the mean	More than Z SDs above or below the mean
Z			
1.0	68.27%	15.87%	31.73%
2.0	95.45%	2.28%	4.55%
2.5	98.76%	0.62%	1.24%
3.0	99.73%	0.13%	0.27%

Fraction of Observations under Standard Normal

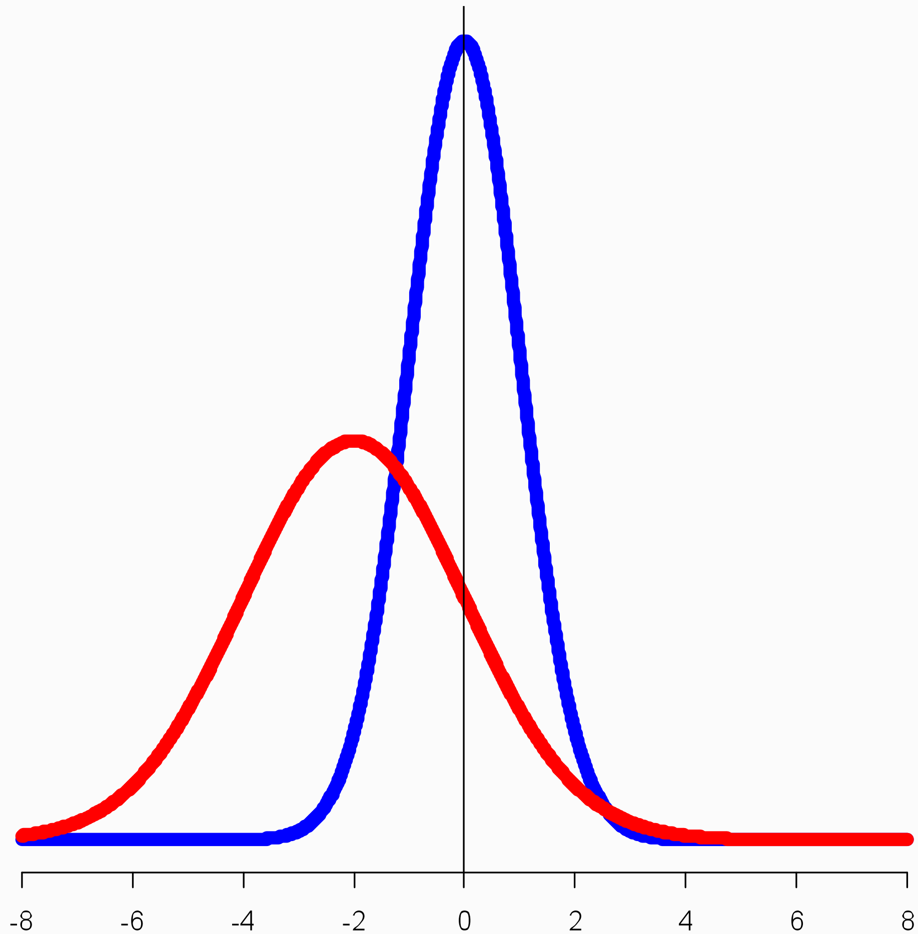
	Within Z SDs of the mean	More than Z SDs above the mean	More than Z SDs above or below the mean
Z			
1.0	68.27%	15.87%	31.73%
2.0	95.45%	2.28%	4.55%
2.5	98.76%	0.62%	1.24%
3.0	99.73%	0.13%	0.27%

The 68-95-99.7 Rule for the Normal Distribution

- What about other normal distributions with other means and standard deviations?
- Same exact properties apply
- In fact, any normal distribution with any mean and standard deviation can be transformed to a standard normal curve

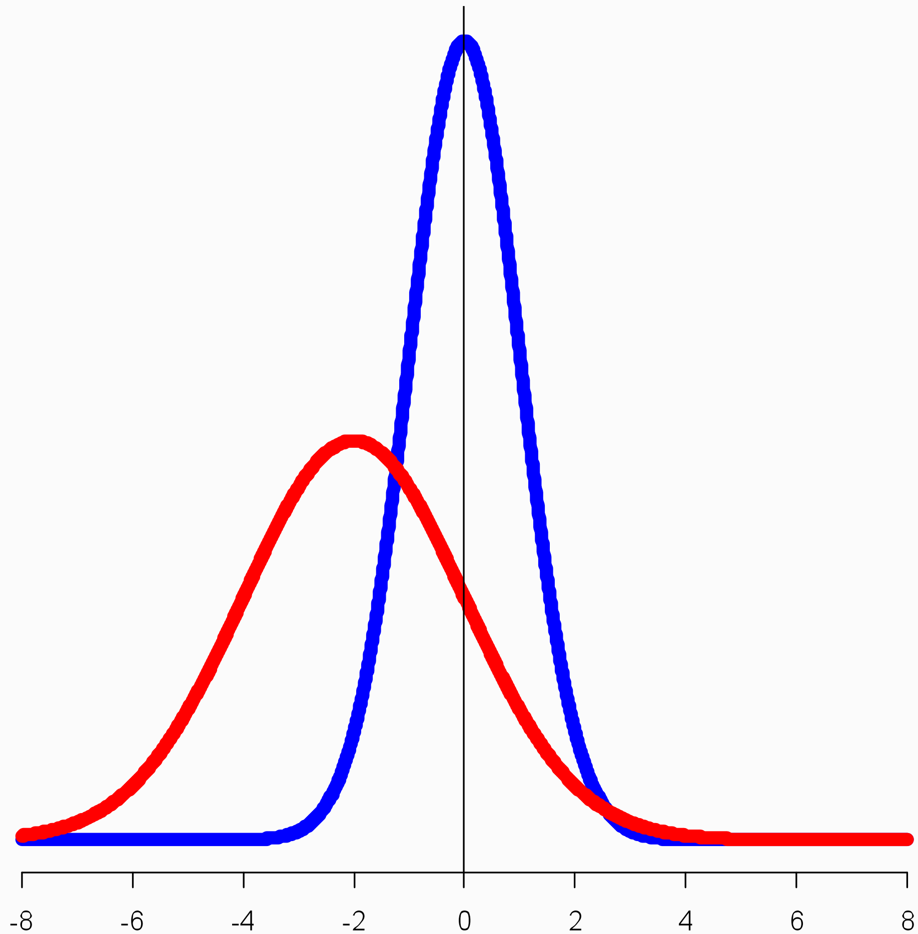
Transforming to Standard Normal

- The standard normal curve (blue) and another normal with mean -2, and standard deviation 2



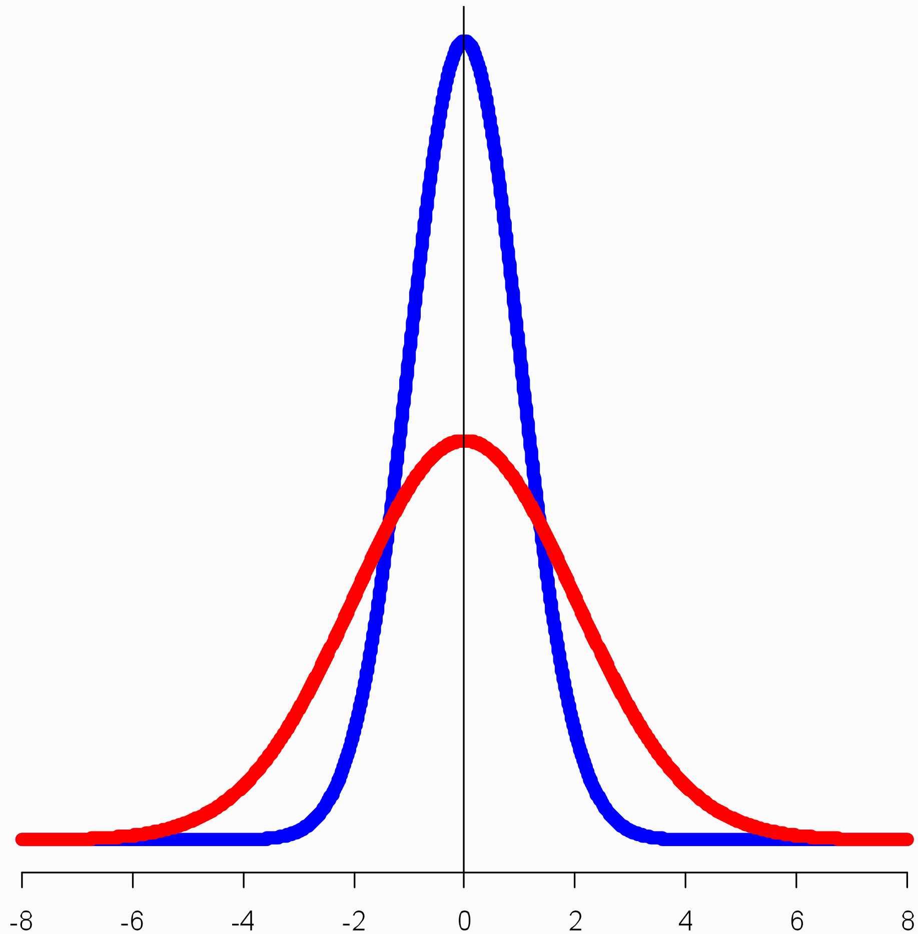
Transforming to Standard Normal

- To center at zero, subtract of mean of -2 from each observation under the red curve



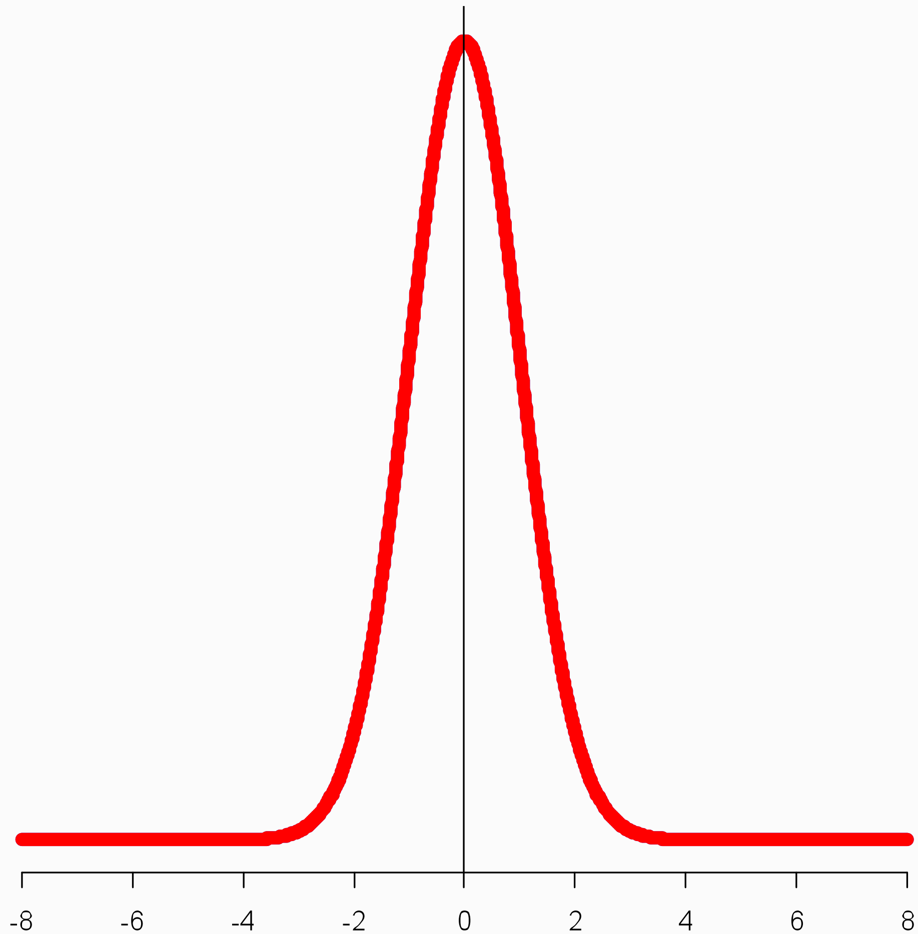
Transforming to Standard Normal

- To “change shape” (i.e., change spread; i.e., standard deviation) divide each “new observation” by standard deviation of 2



Transforming to Standard Normal

- To "change shape" (i.e., change spread; i.e., standard deviation) divide each "new observation" by standard deviation of 2

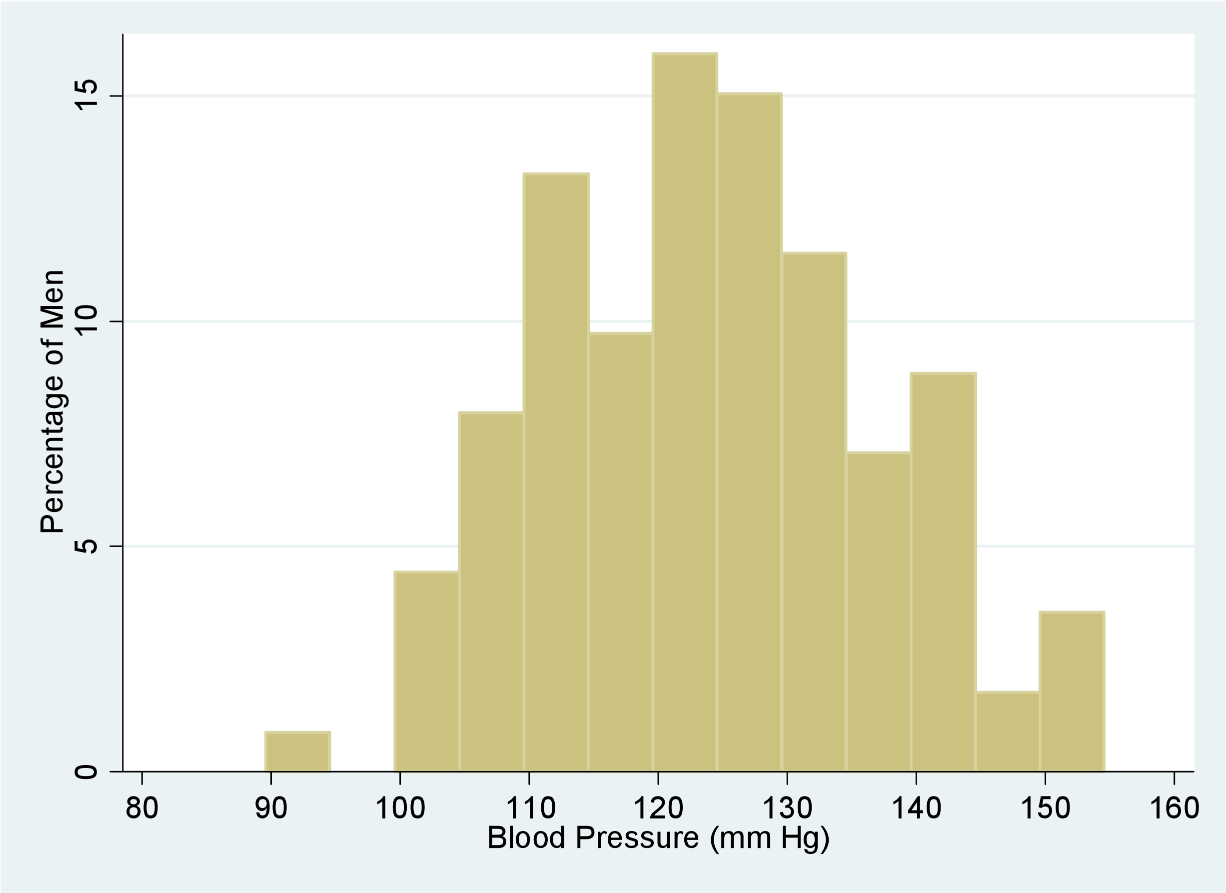


Transforming to Standard Normal

- This process is called standardizing or computing z-scores
- A z-score can be computed for any observation from any normal curve
- A z-score measures the distance of any observation from its distribution's mean in units of standard deviation
- This z-score can help asses where the observations fall relative to the rest of the observations in the distribution
- z-score computed by: $z = \dfrac{observation\text{ - }mean}{standard\ deviation}$

Example 1: Blood Pressure in Males

- Histogram of BP values for random sample of 113 men suggest BP measurements approximated by a normal distribution



Example 1: Blood Pressure in Males

- Data in Stata

```
. list bp in 1/10

     +-------+
     | bp    |
     |-------|
  1. |    89 |
  2. |    99 |
  3. |   101 |
  4. |   101 |
  5. |   103 |
     |-------|
  6. |   103 |
  7. |   104 |
  8. |   105 |
  9. |   106 |
 10. |   106 |
     +-------+
```

Example 1: Blood Pressure in Males

- Summarize command gives sample mean and standard deviation

```
. summarize bp

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       bp    |       113    123.5929    12.86512         89        152
```

Example 1: Blood Pressure in Males

- Summarize command gives sample mean and standard deviation (and sample size, minimum and maximum values)

```
. summarize bp

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      bp     |       113    123.5929    12.86512         89        152
```

$$\bar{x} = 123.6\ mmHg;\ \ s = 12.9\ mmHg$$

Example 1: Blood Pressure in Males

- Using the sample data, let's estimate the range of blood pressure values for "most" (95%) of men in the population

- For normally distributed data, 95% will fall within 2 sds of the mean

$$\bar{x} \pm 2s$$

$$123.6 \ \pm \ 2 \times 12.9$$

$$(97.8, 149.4)$$

- Again, this is just an estimate using the best guesses from the sample for mean and sd of the population

Example 1: Blood Pressure in Males

- Suppose a man comes into my clinic, gets his blood pressure measured, and wants to know how he compares to all men

- His blood pressure is 130 mmHg

- What percentage of men have blood pressures greater than 130 mmHg?

- Translate to z-score $z = \frac{130 - 123.6}{12.9} \approx 0.5$

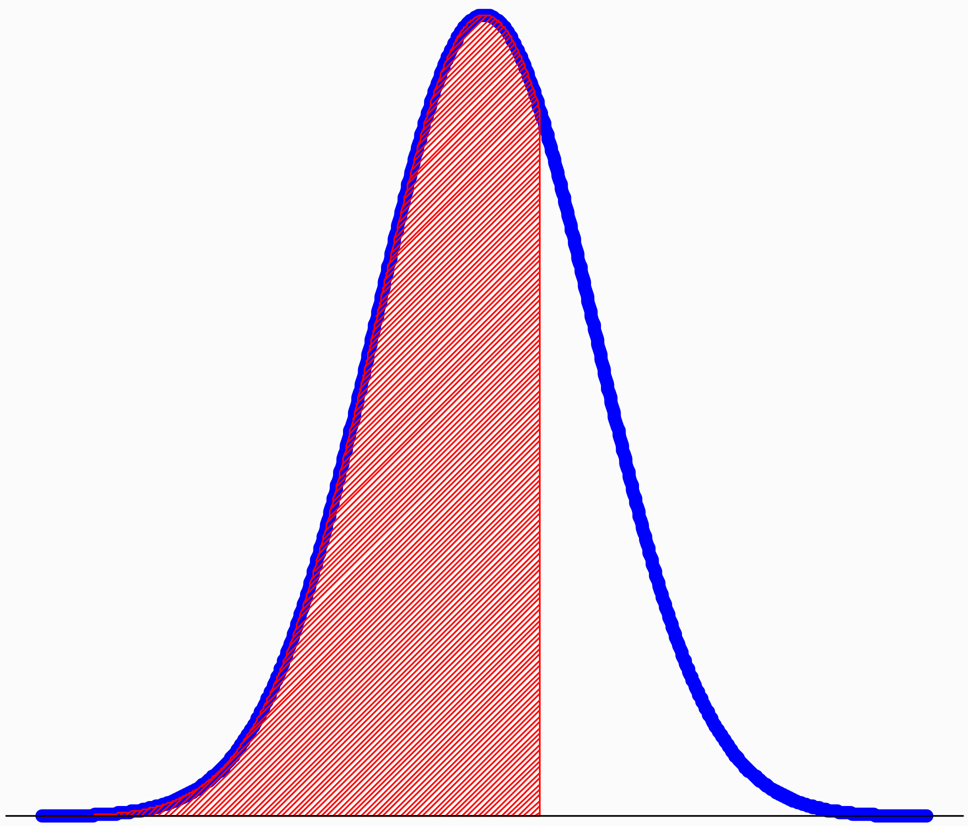
- Question akin to "what percentage of observations under a standard normal curve are 0.5 sds or more above the mean in value?"

Example 1: Blood Pressure in Males

- Could look this up in a normal table (more extensive tables can be found in the back of any stats book or by searching online)
- Could also use normal function in Stata

Example 1: Blood Pressure in Males

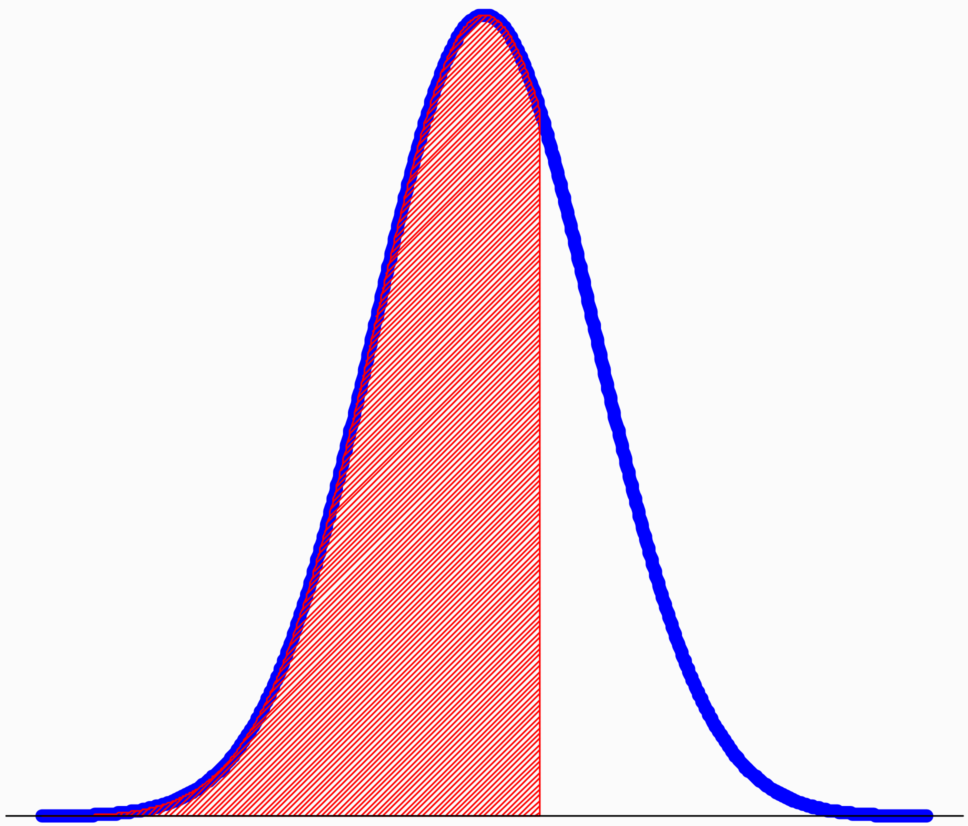
- Typing *display normal(z)* at command line gives proportion of observation less than z standard deviations from mean:



Example 1: Blood Pressure in Males

- For z = 0.5, roughly 69% percent of observations fall below .5 sds from mean

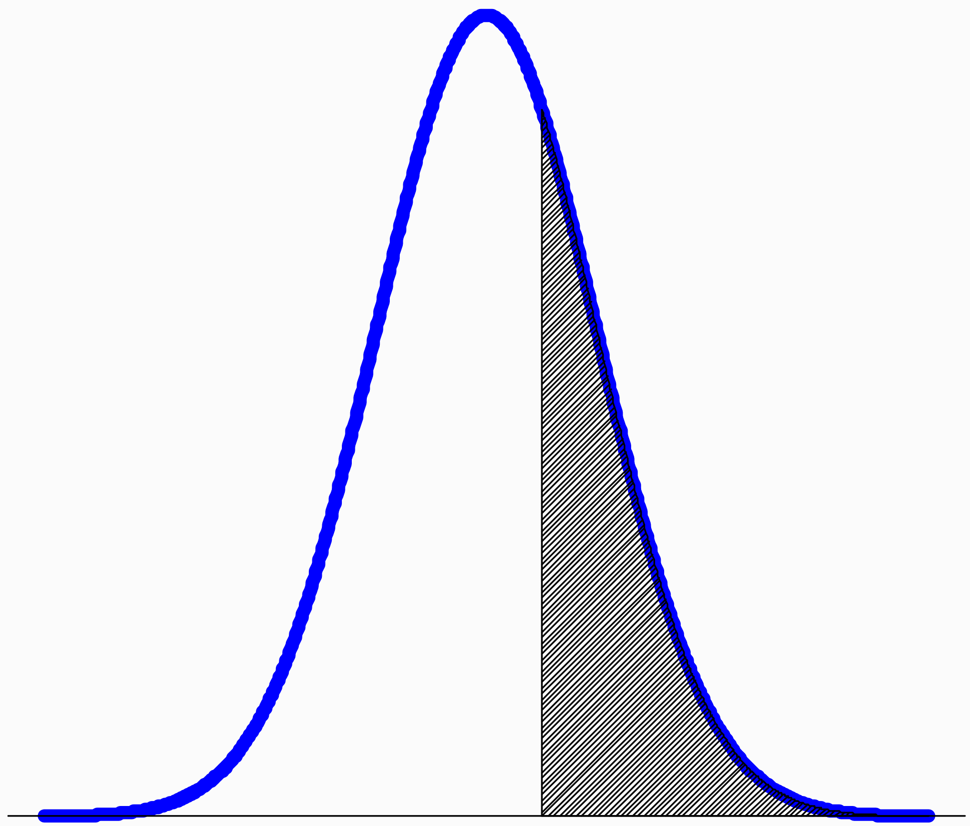
```
. display normal(.5)

.69146246
```



Example 1: Blood Pressure in Males

- For z = 0.5, roughly 100%-69% = 31% of observations fall above .5 sds from mean

```
. display 1- normal(.5)
.30853754
```



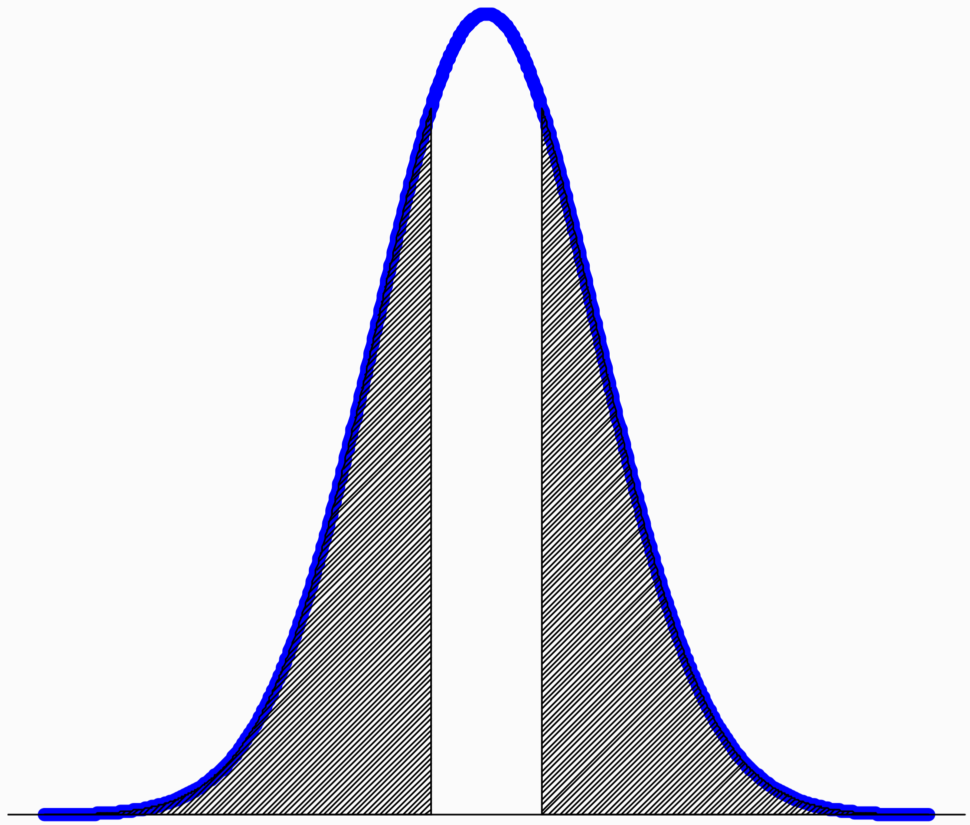
Example 1: Blood Pressure in Males

- So approximately 31% of all men have blood pressures greater than our subject with a blood pressure of 130

- What percentage of men have blood pressures more extreme, i.e. farther than .5 sds from the mean of all men in either direction?

Example 1: Blood Pressure in Males

- What we want



Example 1: Blood Pressure in Males

- By symmetry of normal curve, 31% of observations are above .5 sd, and 31% below -.5 sd

- So a total of 62% is farther than .5 sds from mean in either direction