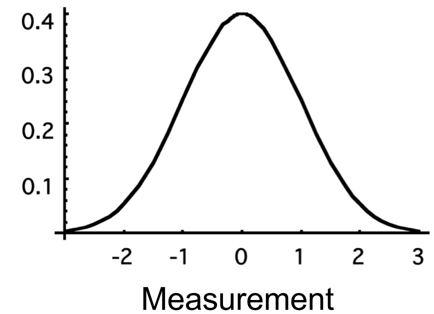
The Normal distribution

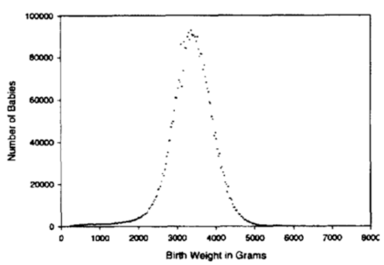
Chapter 10

Normal distribution

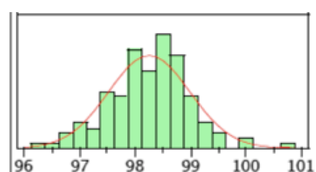
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The normal distribution is very common in nature

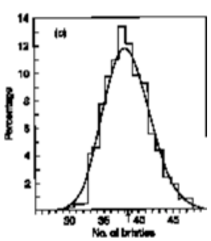


Human birth weight

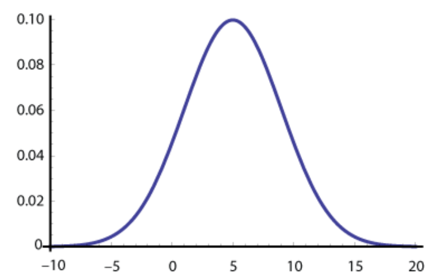


Human body temperature

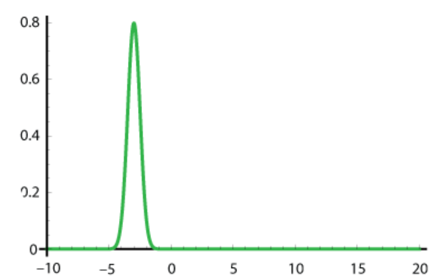
Number of bristles on a *Drosophila abdomen*



A normal distribution is fully described by its mean and standard deviation

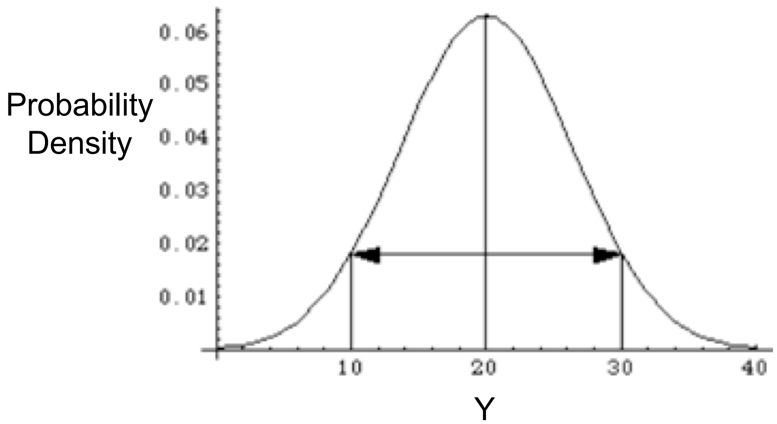


$\mu = 5$; $\sigma = 4$



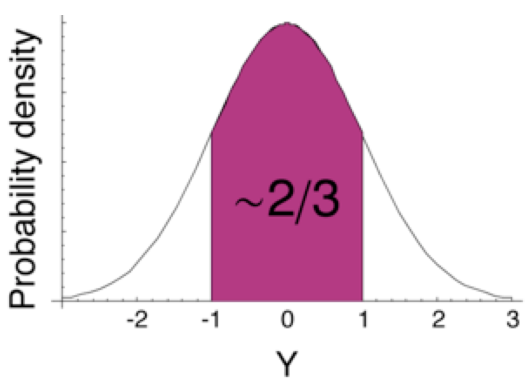
$\mu = -3$; $\sigma = 1/2$

A normal distribution is symmetric around its mean

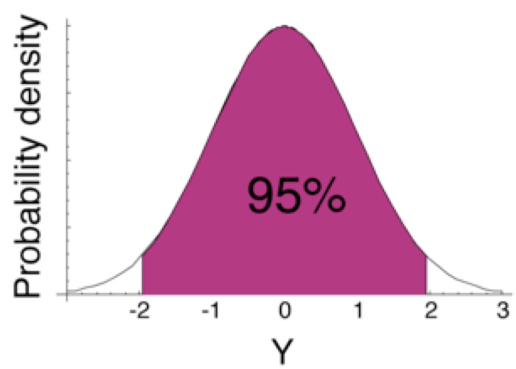


With a normal distribution, the mean, median and mode are all the same.

About 2/3 of random draws from a normal distribution are within one standard deviation of the mean



About 95% of random draws from a normal distribution are within two standard deviations of the mean

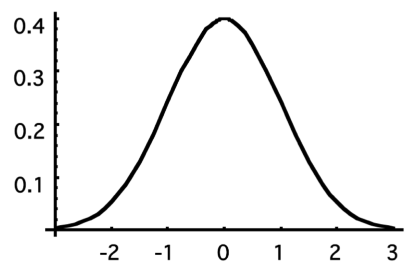


(Really, it's 1.96 SD.)

Standard normal distribution

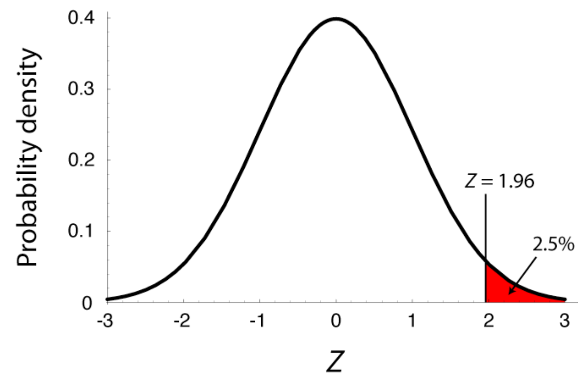
Mean is zero. ($\mu = 0$)

Standard deviation is one. ($\sigma = 1$)



Standard normal table

Gives the probability of getting a random draw from a standard normal distribution greater than a given value



Standard normal table: $Z = 1.96$

First two digits of *a.bc*	Second digit after decimal (*c*)									
	0	1	2	3	4	5	6	7	8	9
1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	$\boxed{0.02500}$	0.02442	0.02385	0.02330
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426

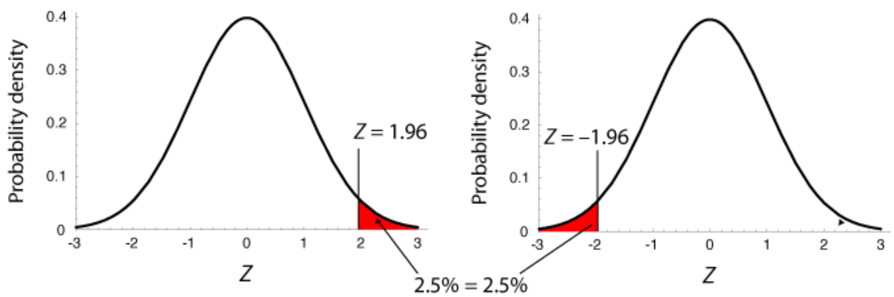
Using R to find area under curve

```
pnorm(1.96, mean = 0, sd = 1, lower.tail = FALSE)

[1] 0.0249979
```

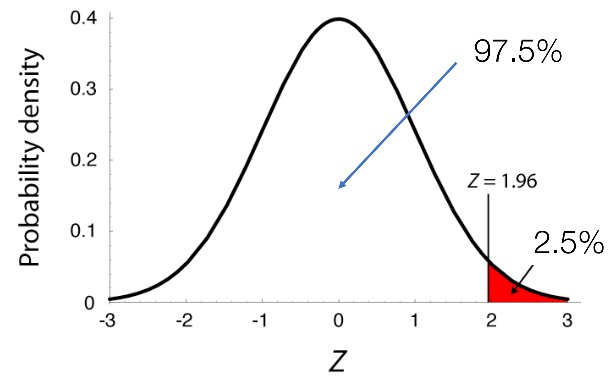
Standard normal is symmetric, so...

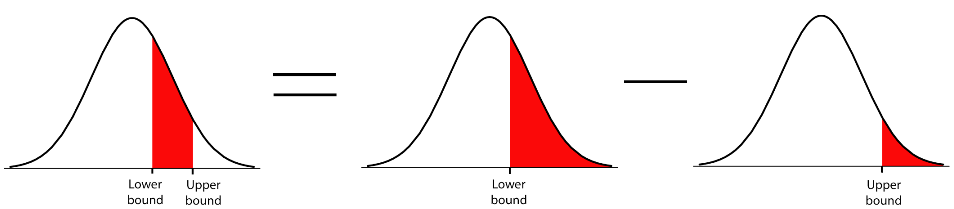
$$\Pr[Z > x] = \Pr[Z < -x]$$



Remember the total area under the curve is equal to 1

$$\Pr[Z < x] = 1 - \Pr[Z > x]$$





What about other normal distributions?

All normal distributions are shaped alike, just with different means and variances

Any normal distribution can be converted to a standard normal distribution, by

$$Z = \frac{Y - \mu}{\sigma}$$

Z is called a "standard normal deviate."

$$Z = \frac{Y - \mu}{\sigma}$$

Z tells us how many standard deviations *Y* is from the mean

The probability of getting a value greater than Y is the same as the probability of getting a value greater than *Z* from a standard normal distribution.

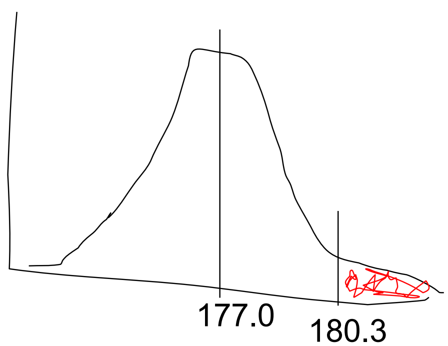
Example: British spies

MI5 says a man has to be shorter than 180.3 cm tall to be a spy.

Mean height of British men is 177.0cm, with standard deviation 7.1cm, with a normal distribution.

What proportion of British men are excluded from a career as a spy by this height criteria?

Draw a rough sketch of the question



μ = 177.0cm
σ = 7.1cm
Y = 180.3

Pr[*height > 180.3*]

$$Z = \frac{Y - \mu}{\sigma}$$

$$Z = \frac{180.3 - 177.0}{7.1}$$

$$Z = 0.46$$

Part of the standard normal table

	x.x0	x.x1	x.x2	.x3	x.x4	x.x5	x.x6	x.x7	x.x8	x.x9
0.0	0.5	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.4721	0.46812	0.46414
0.1	0.46017	0.4562	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
0.3	0.38209	0.37828	0.37448	0.3707	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
0.4	0.34458	0.3409	0.33724	0.3336	0.32997	0.32636	$\boxed{0.32276}$	0.31918	0.31561	0.31207
0.5	0.30854	0.30503	0.30153	0.29806	0.2946	0.29116	0.28774	0.28434	0.28096	0.2776

Pr[Z > 0.46] = 0.32276,
so Pr[*height* > 180.3] = 0.32276

Sample means are normally distributed

(If the variable itself is normally distributed.)

The mean of the sample means is μ.

The standard deviation of the sample means is $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$

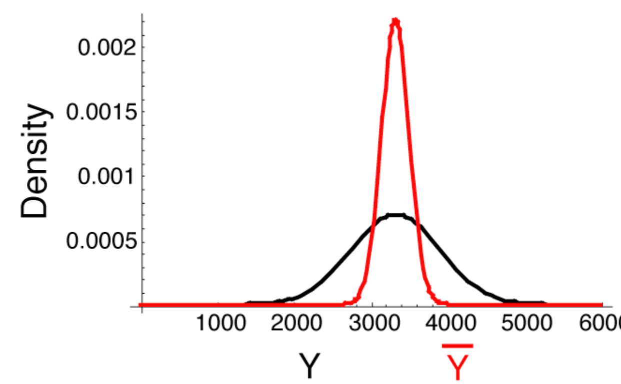
Standard error

The standard error of an estimate of a mean is the standard deviation of the distribution of sample means

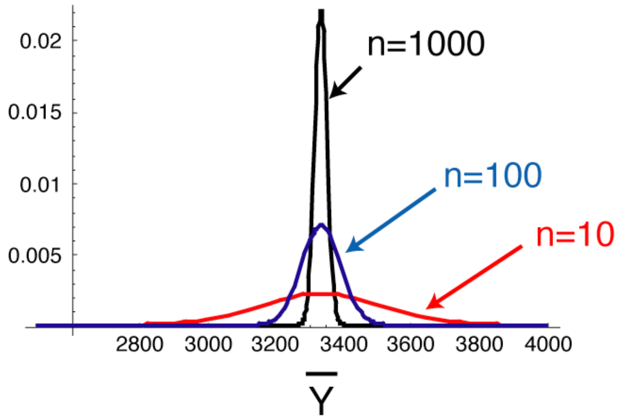
$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$

We can approximate this by $SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$

Distribution of means of samples with *n* =10



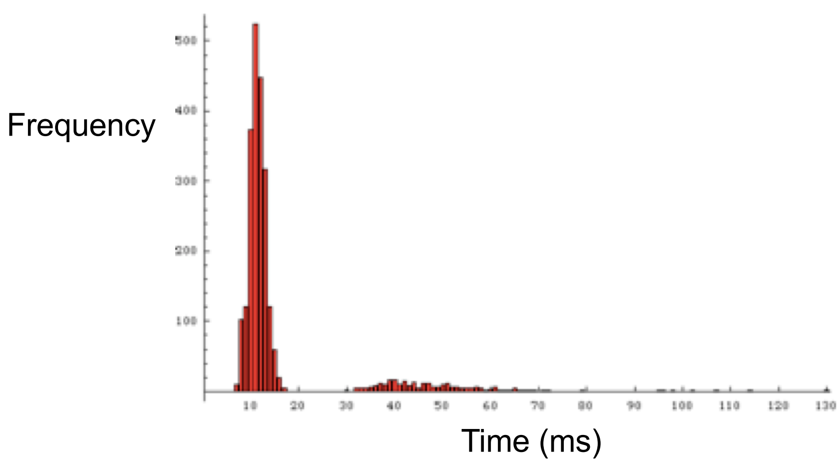
Larger samples equal smaller standard errors

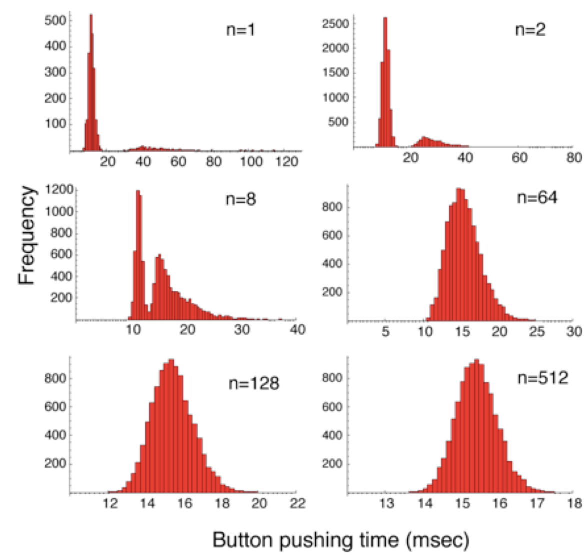


Central limit theorem

The sum or mean of a large number of measurements randomly sampled from *any* population is approximately normally distributed.

Button pushing times





https://www.zoology.ubc.ca/~whitlock/Kingfisher/CLT.htm