Normal Distributions

MATH 107: Finite Mathematics

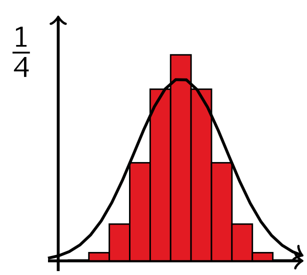
University of Louisville

April 2, 2014

Getting patterns from aggregated data

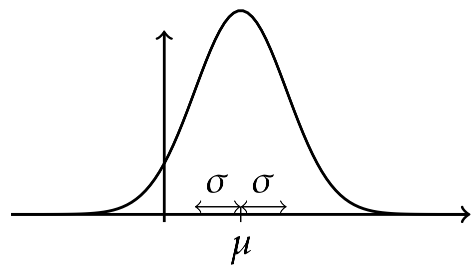
A random variable can have *any* distribution. Add up a bunch of random variables, though, and the story is different.

For instance, we might compute the probability distribution of heads over 10 coin-flips:



This is "mound-shaped"; as it turns out, aggregated data is always shaped like this particular mound!

The normal distribution



This shape, to which all aggregated data tends, is called the "normal" or "Gaussian" distribution. It has the awful equation:

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

π here is the beloved constant (about 3.141592); σ and μ describe the positioning of the curve.
μ is the *mean*: the center and highest point of the curve.
σ is the *standard deviation*, which quantifies spread.

Normally distributed features

Many everyday features are normally distributed.

Some normally distributed population features

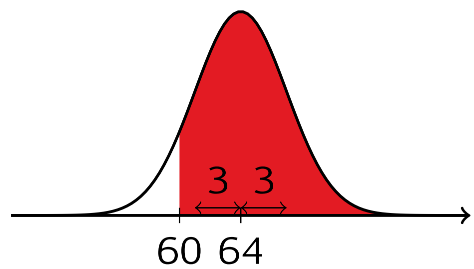
- Adult height in women: $\mu = 64$ inches, $\sigma = 3$ inches.
- IQ: $\mu = 100$, $\sigma = 15$.

Based on these, if we understand normal distributions, we can work out probabilities in the population!

Finding population stats

A simple statistical question

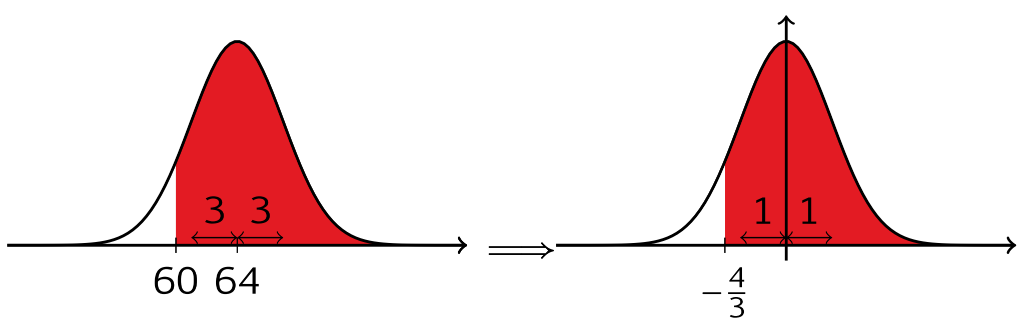
What percentage of women are more than 60 inches tall?



This curve describes the distribution of women's heights.

We want to know what proportion of the population is in the shaded area.

Standardizing curves



Having to know a lot about *every* normal curve would be difficult.

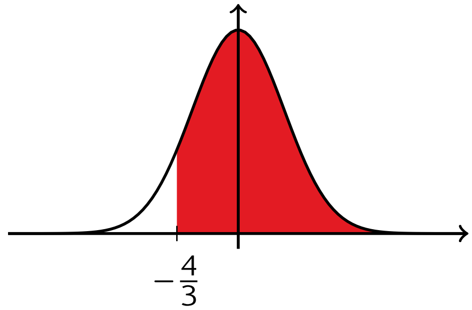
Instead, we look at a *standard* normal curve with $\mu = 0$ and $\sigma = 1$.

We map 60 on the old curve to a position on the new curve called a z-value:

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

So now we want to know what's to the right of $-\frac{4}{3}$ on the *standard normal curve*.

The bad news

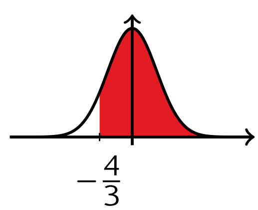


There is no easy way to determine this area!

There are complicated computations which have been done in advance.

Table 1 in Appendix C has a list of these computations.

Using a table

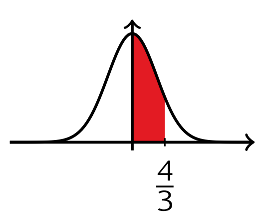


We're interested in what happens at $-\frac{4}{3} \approx -1.33$.

We look up 1.33 in our table:

	$\cdots$	.03	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1.3	$\cdots$	0.4082	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

That's the area between 0 and 1.33 on the graph.



This is not quite what we want, but it's close; we add $\frac{1}{2}$ to get the real thing: $0.4082 + 0.5 = 90.82\%$

General approach

To find the probability of a given range in a normal curve:

- Translate both ends of the range to z-values: $z = \frac{x-\mu}{\sigma}$.
- Look up the absolute value of each end in a table, taking $\frac{1}{2}$ as the answer when an end is open.
- Reverse the sign on either end which has negative z-value.
- Subtract the two table-results.

Examples

Characterization of IQ

IQs are normally distributed around 100 with a standard deviation of 15.

How likely is an IQ of between 110 and 130?

$$x_1 = 110 \to z_1 = \frac{110-100}{15} \approx 0.66 \to p_1 = 0.2454$$

$$x_2 = 130 \to z_2 = \frac{130-100}{15} = 2.00 \to p_2 \to 0.4772$$

so the likelihood is $0.4772 - 0.2454 = 23.18\%$.

Examples

Characterization of IQ

IQs are normally distributed around 100 with a standard deviation of 15.

How likely is an IQ of less than 120?

$$x_1 = -\infty \to p_1 = -0.5$$

$$x_2 = 150 \to z_2 = \frac{120 - 100}{15} \approx 1.33 \to p_2 \to 0.4082$$

so the likelihood is $0.4082 - (-0.5) = 90.82\%$.

Yet Another Example

Manufacturing

Your factory produces washers with an interior diameter of 15mm; due to manufacturing variance the washers actually have normally distributed diameter with standard deviation of 0.14mm. For high-precision applications, these washers need to have interior diameter between 14.9mm and 15.05mm. What proportion of our washers are suitable for high-precision applications?

$$x_1 = 14.90 \to z_1 = \frac{14.90 - 15}{0.14} \approx -0.71 \to p_1 = -0.2611$$

$$x_2 = 15.05 \to z_2 = \frac{15.05 - 15}{0.14} \approx 0.36 \to p_2 \to 0.1406$$

so $0.1406 - (-0.2611) = 40.17\%$ of our washers will be sufficiently precise.