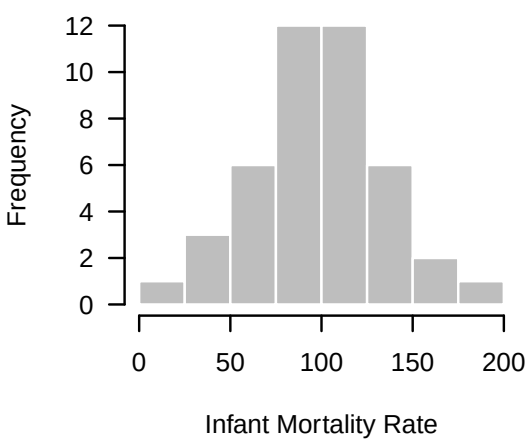
The normal distribution

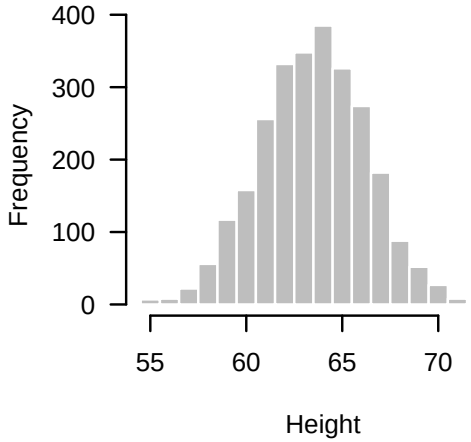
Patrick Breheny

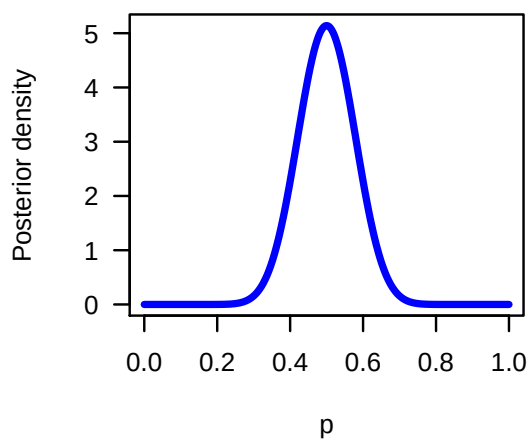
September 29

A common histogram shape

A histograms of infant mortality rates in Africa, heights from a sample of adult women in the U.S., and a Bayesian posterior for a binomial outcome with 20 successes/20 failures:





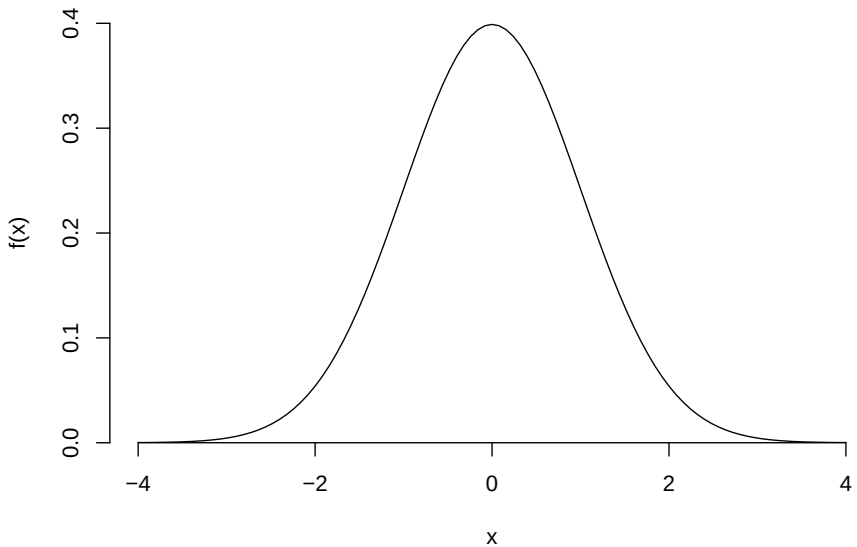


Three extremely different contexts, and yet all three have essentially the same shape

The normal curve

All three distributions (and, of course, many, many more) are well described by the following equation:

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



This distribution is known as the *normal distribution*; other names for it include the Gaussian distribution (after Gauss, one of the first to describe it mathematically) and the bell curve (because it looks

Features of the normal curve

Note that

- The normal curve is symmetric around $x = 0$
- The normal curve is always positive
- The normal curve drops rapidly down near zero as x moves away from 0

Transforming the data/distribution

- Of course, the normal distribution doesn't look *exactly* like the distributions on the first slide – it's centered at zero and the others aren't
- There are large differences in how spread out the distributions are
- There are two ways to fix this: we can either transform the data, or transform the normal distribution itself
- Both are important and widely used in statistics, so we'll discuss each approach

Standardization

- Let's discuss transforming the data first, and take height as an example
- In the women's height data, one woman measured 66.0 inches tall
- Because the average height of the women was 63.5 inches, another way of describing her height is to say that she was 2.5 inches above average
- Furthermore, because the standard deviation was 2.75 inches, yet another way of describing her height is to say that she was 0.91 standard deviations above average

Standardization (cont'd)

- This idea of taking a variable and converting it into SDs away from the mean is known as *standardization*, and can be expressed mathematically as:

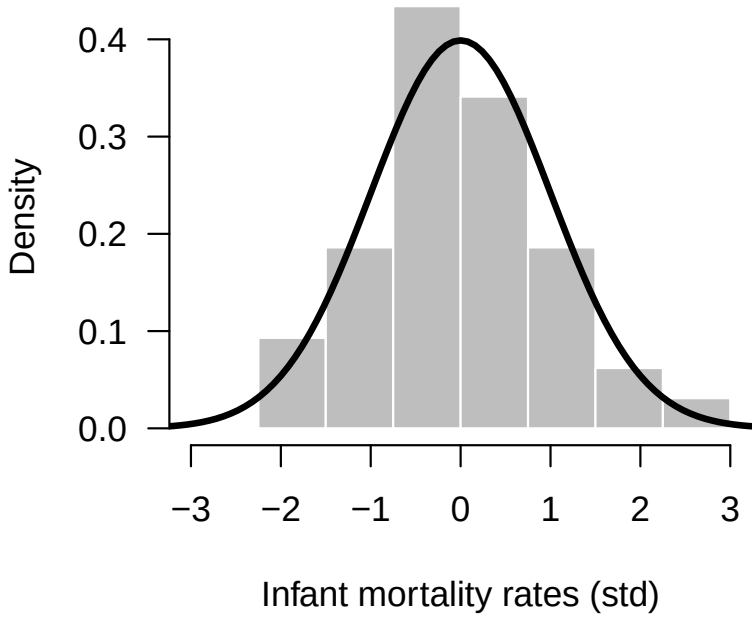
$$z_i = \frac{x_i - \bar{x}}{\mathrm{SD}_x},$$

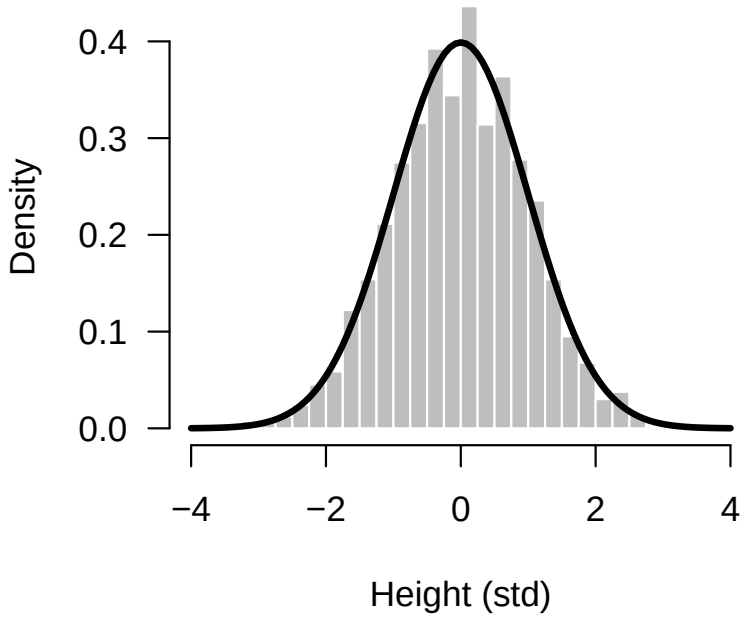
  where $\bar{x}$ and SD_x are the mean and standard deviation of x
- One virtue of standardizing a variable is interpretability:
  - If someone tells you that the concentration of urea in your blood is 50 mg/dL, that likely means nothing to you
  - On the other hand, if you are told that the concentration of urea in your blood is 4 standard deviations above average, you can immediately recognize this as a very high value

More on standardization

- If you standardize all of the observations in your sample, the resulting variable (i.e., z on the previous slide) will have mean 0 and standard deviation 1
- Standardization therefore brings all variables onto a common scale – regardless of whether the heights were originally measured in inches, centimeters, or miles, the standardized heights will be identical
- For many kinds of data, it would be illogical if the results depended on the scale of measurement, so typically, we can analyze standardized data without loss of generality

Standardization in action





Data whose density/histogram looks like the normal curve are said to be "normally distributed" or to "follow a normal distribution"

Transforming densities

- It's also useful to be able to transform the normal distribution itself
- However, we need to be a little careful when we start stretching and compressing probability densities to ensure that the actual probabilities are correctly preserved
- For example, let's suppose Z is a random variable and we want to create a new random variable X given by

$$X = \sigma Z + \mu;$$

i.e., by shifting Z μ units and stretching it out by a factor of σ

Transforming densities (cont'd)

- What happens when we just plug $(x - \mu)/\sigma$ into the density function for z?
- We run into the problem that $f((x - \mu)/\sigma)$ integrates to σ, not 1
- However, this is an easy fix; we can simply divide the density function of X by σ to guarantee that the function integrates to 1:

$$g(x) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

and is therefore a valid probability density for all values of μ and σ

Location-scale family of normal distributions

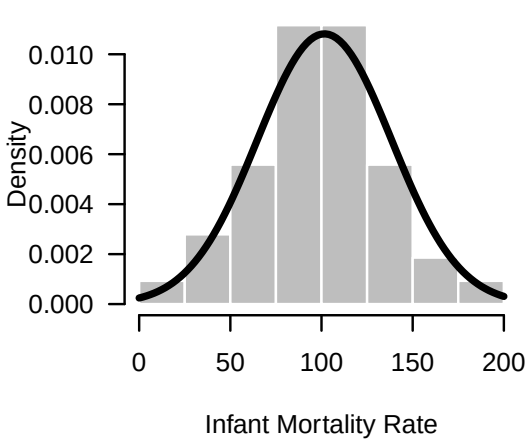
- Let's extend this logic to the normal distribution, shifting it over by μ units and rescaling it by a factor of σ
- The density function of $X = \sigma Z + \mu$, where Z has the normal distribution given on slide 3, is given by

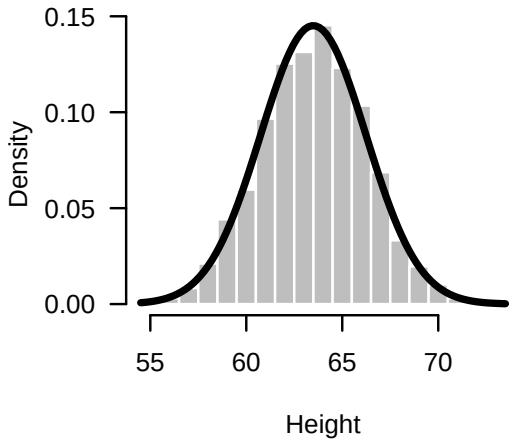
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

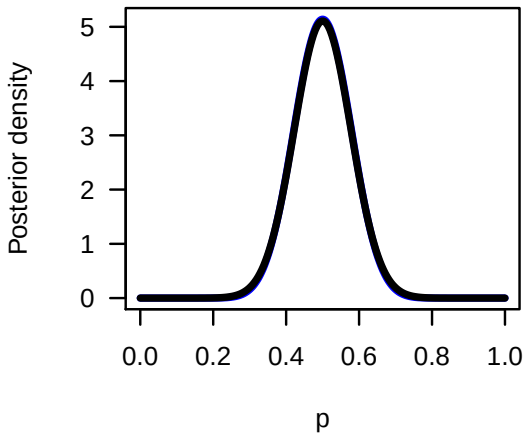
- This is the density function for a family of normal distributions with location scale μ and scale parameter σ
- The special case with $\mu = 0$ and $\sigma = 1$ that we encountered originally is known as the *standard normal* distribution

The location-scale family in action

Plugging in the appropriate means/SDs for the location and scale parameters:







Expected values

- Now is a good time to introduce one final concept that is fundamental to random variables and probability theory, that of the *expected value*
- The expected value generalizes the idea of the sample mean to a distribution
- The expected value of a discrete random variable X is defined by

$$\mathrm{E}(X) = \sum xf(x)$$

- The expected value of a continuous random variable X is defined by

$$\mathrm{E}(X) = \int xf(x)dx$$

Linearity of the expectation operator

- **Theorem:** Let X be a random variable and a, b denote constants. Then

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$$

- This linearity property is extremely convenient and makes expected values very easy to work with – much more so than working with probability distributions directly
- It is worth noting, however, that expectations cannot, in general, be moved inside of functions:

$$\mathrm{E}(f(X)) \neq f(\mathrm{E}(X))$$

in general, the linear case above being one of the rare cases where this holds

Expected value of a normal distribution

- **Theorem:** Let X follow a standard normal distribution. Then $\mathrm{E}(X) = 0$.
- **Theorem:** Let $X \sim \mathrm{N}(\mu, \sigma)$. Then $\mathrm{E}(X) = \mu$.

Moments

- The expected values of various powers of X are called its *moments*
- The nth moment of X is $\mathrm{E}(X^n)$
- The nth central moment of X is $\mathrm{E}\{(X-\mu)^n\}$, where $\mu = \mathrm{E}(X)$

Variance

- The second central moment is of particular interest, and is called the *variance*:

$$\mathrm{Var}(X) = \sum (x-\mu)^2 f(x) \qquad \text{(discrete)}$$

$$\mathrm{Var}(X) = \int (x-\mu)^2 f(x) \qquad \text{(continuous)}$$

- **Theorem:** Let X be a random variable and a, b denote constants. Then

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$$

- **Theorem:** For any random variable X,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$$

Variance of a normal distribution

- **Theorem:** Let X follow a standard normal distribution. Then $\mathrm{Var}(X) = 1$.
- **Theorem:** Let $X \sim \mathrm{N}(\mu, \sigma)$. Then $\mathrm{Var}(X) = \sigma^2$.

Reconstructing data with just two numbers

- One of the reasons that the normal distribution is so convenient to work with is that it only has two parameters, and those parameters are directly connected to the mean and standard deviation
- So, as you may recall, we said earlier in the course that the mean and standard deviation provide a two-number summary of a histogram; we can now make this remark a little more concrete
- To a large extent, anything we want to know about the data, we can determine by approximating the real distribution of the data by the normal distribution
- This approach is called the *normal approximation*

NHANES adult women

- The data set we will work with on these examples is the NHANES sample of the heights of 2,649 adult women
- The mean height is 63.5 inches
- The standard deviation of height is 2.75 inches

Estimating probabilities: Example # 1

- Suppose we want to estimate the percent of women who are under 5 feet tall
- We could take one of two equivalent approaches:
  - Take the CDF of the normal distribution with $\mu = 63.5$, $\sigma = 2.75$
  - Transform 5 feet (60 inches) to $(60 - 63.5)/2.75 = -1.27$ and take the standard normal CDF of -1.27
- $P(X < 60) = P(Z < -1.27) = 10.2\%$ (and either way, this is not an integral you can easily calculate without a computer)
- In the actual sample, 282 out of 2,649 women were under 5 feet tall, which comes out to 10.6%

Estimating probabilities: Example # 2

- Another example: suppose we want to estimate the percent of women who are between 5'3 and 5'6 (63 and 66 inches)
- Again, either $F(66) - F(63)$ based on $\mathrm{N}(63.5, 2.75)$ or $\Phi(0.91) - \Phi(-0.18)$; ϕ is commonly used to denote the pdf of the standard normal distribution and Φ its CDF
- Using the normal distribution, the probability of falling in this region is 39.0%
- In the actual data set, 1,029 out of 2,649 women were between 5'3 and 5'6: 38.8%

Approximating percentiles: Example

- Suppose instead that we wished to find the 75th percentile of these women's heights
- Again we could take one of two equivalent approaches:
  - Take the inverse CDF of 0.75 for the normal distribution with $\mu = 63.5$, $\sigma = 2.75$
  - Take $\Phi^{-1}(0.75)$, then transform using $63.5 + 2.75\Phi^{-1}(0.75)$
- Using the normal distribution, the 75th quantile is 65.35 inches
- For the actual data, the 75th percentile is 65.39 inches

The broad applicability of the normal approximation

- These examples are by no means special: the distribution of many random variables are very closely approximated by the normal distribution (we will discuss *why* this happens next time)
- When it happens, this is pretty remarkable and powerful – for variables with approximately normal distributions, the mean and standard deviation essentially tell us everything we need to know about the data; other summary statistics and graphics are redundant

Caution

- Other variables, however, are not approximated by the normal distribution well, and give misleading or nonsensical results when you apply the normal approximation to them
- For example, the value 0 lies 1.63 standard deviations below the mean infant mortality rate for Europe
- The normal approximation therefore predicts a probability that 5.1% of the countries in Europe will have negative infant mortality rates

Caution (cont'd)

- As another example, the normal distribution will always predict the median to lie 0 standard deviations above the mean
- *i.e.*, it will always predict that the median equals the mean
- As we have seen, however, the mean and median can differ greatly when distributions are skewed
- For example, according to the U.S. census bureau, the mean income in the United States is $66,570, while the median income is $48,201

Summary

- The distribution of many random variables are very closely approximated by the normal distribution:

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

- Expected value and variance of a continuous distribution:

$$\mathrm{E}(X) = \int x f(x) dx$$

$$\mathrm{Var}(X) = \int (x-\mu)^2 f(x)$$

- $\mathrm{E}(aX+b) = a\mathrm{E}(X) + b$
- $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$
- Know how to calculate quantiles for the normal distribution and use them to approximate other distributions