Using the Normal

Mark Muldoon

Departments of Mathematics and

Optometry & Neuroscience

UMIST

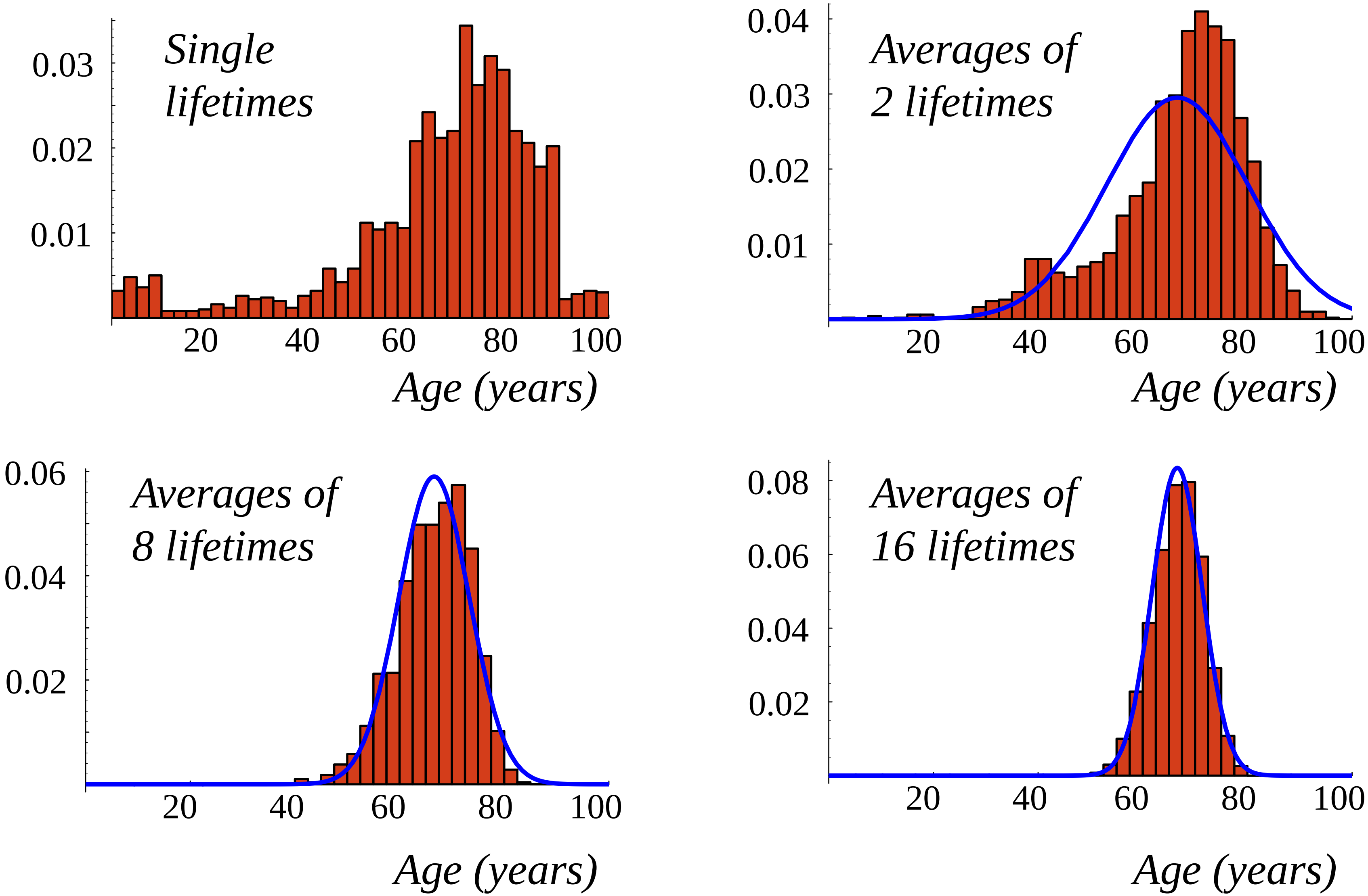
http://www.ma.umist.ac.uk/mrm/Teaching/2P1/

Overview

Today we'll discuss ways to use the normal distribution

- **Why is the normal famous?** review/clarification from last lecture
- **Changing units:** talking about the weather and continuous probability
- **Properties of the normal:** changing μ and σ for the normal
- **Integrals:** doing them without pain
- **Using the normal:** characterizing extremity and approximating the binomial distribution.

Why is the normal famous?



Averaging makes distributions more normal

Upper Left:

- Generate 2,000 random lifetimes

$$\{t_1, t_2, \ldots, t_{2000}\}$$

- Make histogram

Upper Right:

- Generate 4,000 random lifetimes, average successive pairs

$$\{\frac{(t_1 + t_2)}{2}, \frac{(t_2 + t_3)}{2}, \ldots\}$$

- Make histogram of averages

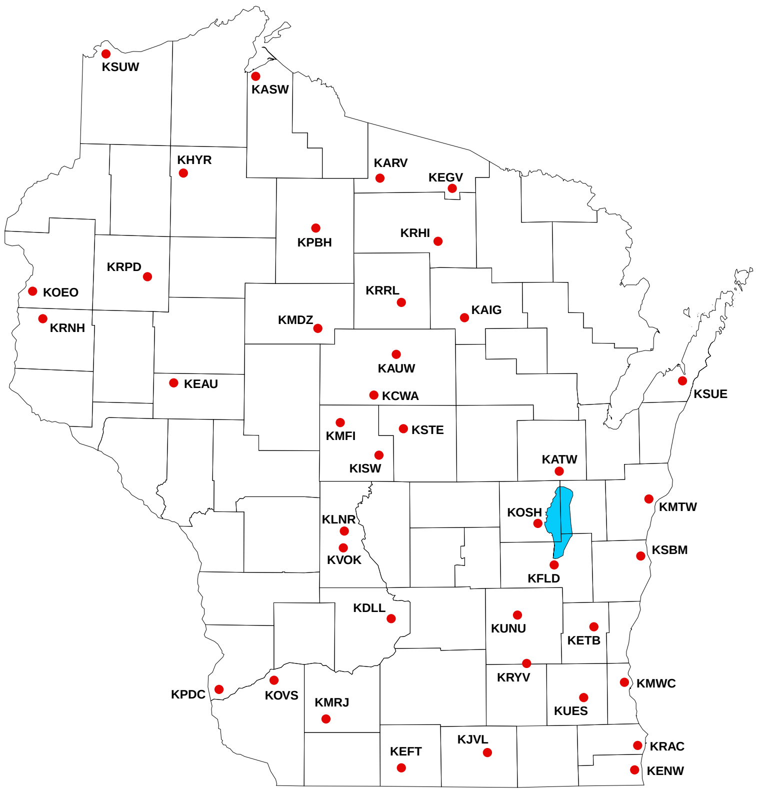
Lower Left:

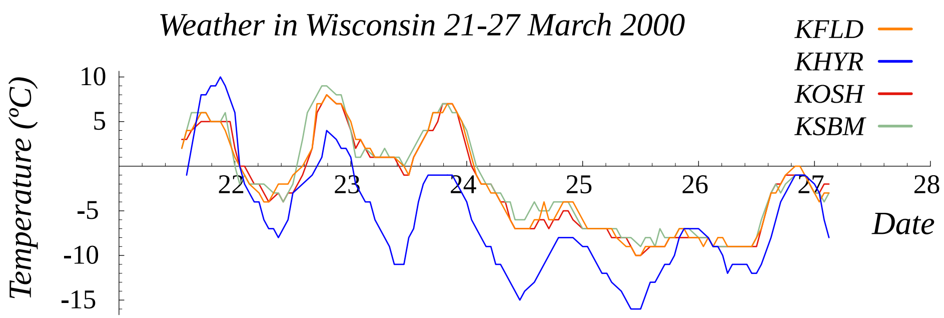
- Generate 16,000 lifetimes, average blocks of 8.
- Make histogram of averages

Lower Right:

- Generate 32,000 lifetimes, average blocks of 16.
- Make histogram of averages

Cold Wisconsin nights





Hourly weather data are freely available from

ftp://tgftp.nws.noaa.gov/data/observations/metar/stations

Changing units

Data are in Centigrade, but my intuition about comfort/coldness is all in terms of temperatures in Fahrenheit

$$T_F = (9/5) \times T_C + 32$$

Change of units is a combination of two processes:

a) **scaling:** multiplication by 1.8

b) **translation:** addition of 32

Changing units: expectations

Think of late-March temperature as a random variable:

Q1: If μ_C and σ_C are the mean and std. deviation of the temperatures as recorded in Centigrade, what are the corresponding stats for Fahrenheit measurements?

A1:

Changing units: expectations

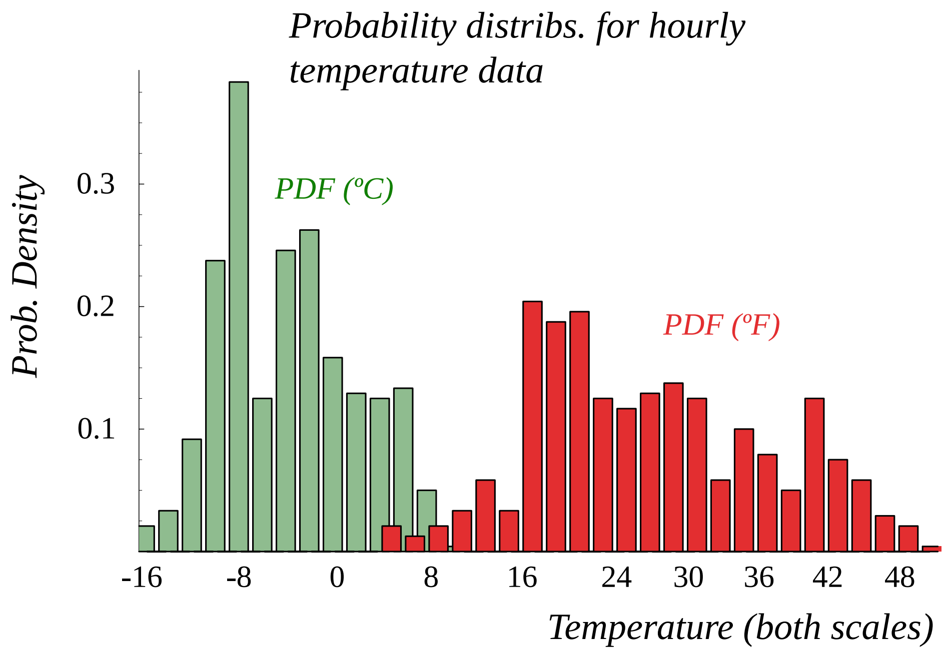
Think of late-March temperature as a random variable:

Q1: If μ_C and σ_C are the mean and std. deviation of the temperatures as recorded in Centigrade, what are the corresponding stats for Fahrenheit measurements?

A1: Scaling affects both the mean and the standard deviation, but translation affects only the mean:

$$\begin{aligned} \mu_F &= (9/5)\mu_C + 32 \\ \sigma_F &= (9/5)\sigma_C \end{aligned}$$

Two PDFs for Wisconsin

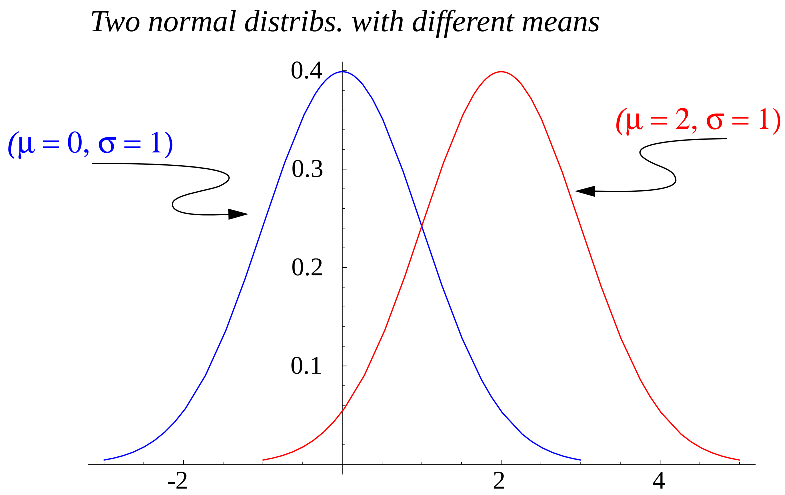


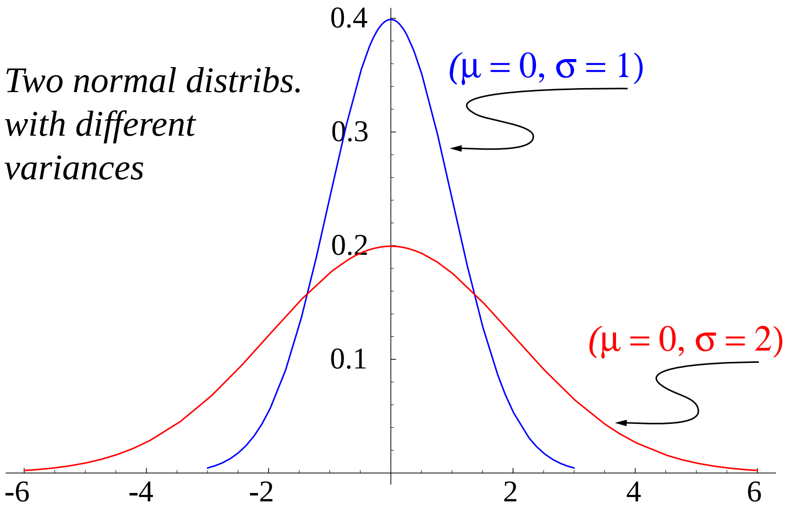
Fahrenheit histogram is:

- shifted to the right (by 32 degrees);
- both wider and flatter (by factor of (9/5)).

Shifting and scaling the normal

The normal distribution has the property that shifted and scaled versions of it are still normal (though with different means and variances):





Check: shifting and scaling

If x is a normally-distributed random variable with mean μ_x and standard deviation σ_x, then a scaled, translated random variable $y = 2x + 5$

a) a normal distribution centered on $\mu_y = 2\mu_x$;

b) a normal distribution centered on $\mu_y = 2\mu_x + 5$;

c) a binomial distribution;

d) variance $\sigma_y^2 = 4\sigma_x^2$;

e) variance $\sigma_y^2 = s2\sigma_x^2 + 5$.

Answer:

Check: shifting and scaling

If x is a normally-distributed random variable with mean μ_x and standard deviation σ_x, then a scaled, translated random variable $y = 2x + 5$

a) a normal distribution centered on $\mu_y = 2\mu_x$;

b) a normal distribution centered on $\mu_y = 2\mu_x + 5$;

c) a binomial distribution;

d) variance $\sigma_y^2 = 4\sigma_x^2$;

e) variance $\sigma_y^2 = s2\sigma_x^2 + 5$.

Answer: *Only items (b) and (d) are true.*

Tables for the normal

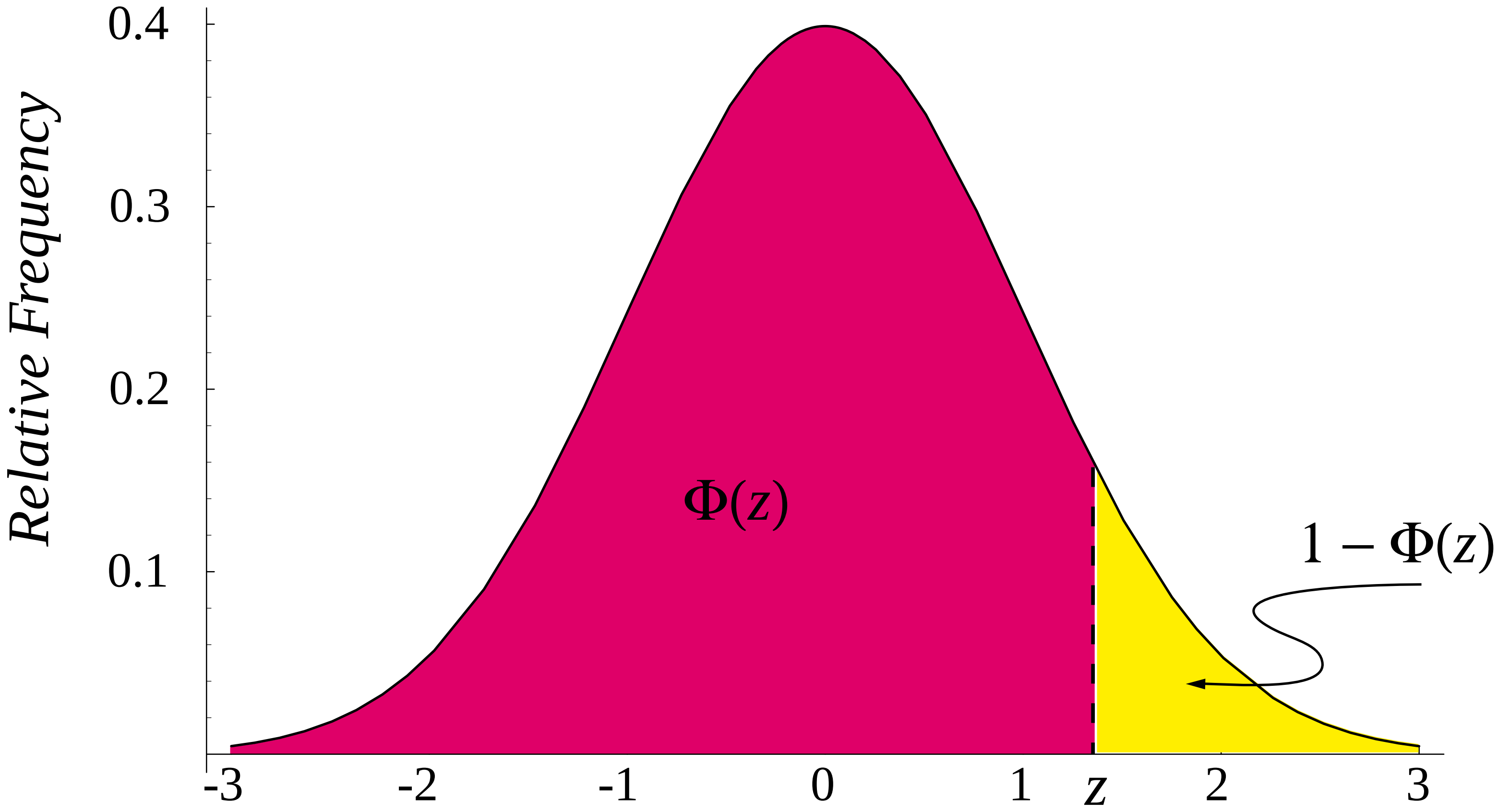
The main point about the shifting and scaling that we have been discussing is that it means that we never have to do any integrals to use the normal distribution. Today we will learn to use the attached table, which gives $1 - \Phi(z)$ where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2}\, dz$$

is the probability of drawing a value $\leq z$ from a standard normal distribution.

Tables of the upper tail

A table of $1 - \Phi(z)$ is attached to these notes.



Application: testing extremity

Acme Salts aims to produce contact lens solutions with tightly controlled salinity: the 500 ml bottles contain a mean of 5 gm of NaCl with a standard deviation of 0.05 gm. Bottles are rejected if their salt content lies outside the band 4.92 to 5.08. Assuming the salinity of bottles to be normally distributed:

a) Out of 100 bottles, how many would you expect to be rejected?

b) How many rejections would you expect if the acceptable range narrowed to 4.98–5.03?
Answer: around 62 bottles—check this for yourself.

Salinity: using the standard normal

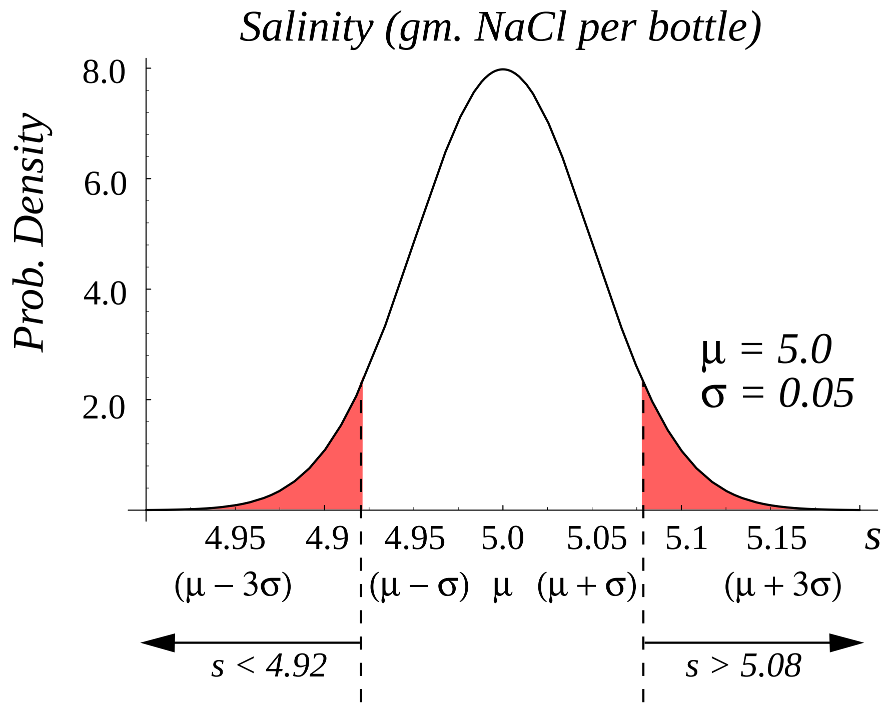
As the mean salinity is 5 gm. per bottle and the standard deviation is 0.05, the scaled, translated random variable z given by

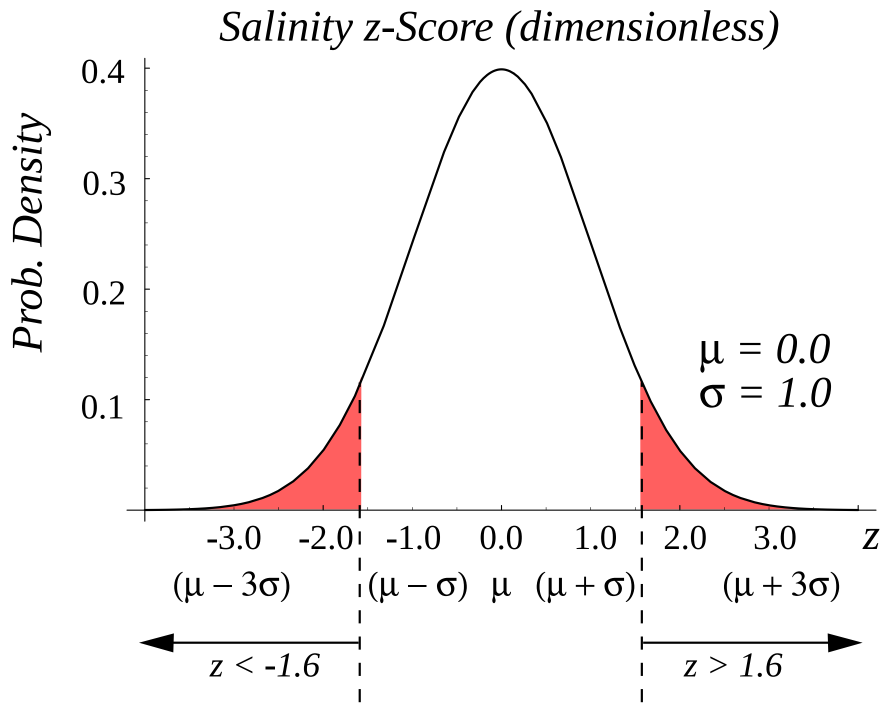
$$z = \frac{s - \text{(mean salinity)}}{\text{(standard deviation of salinity)}} = \frac{(s-5)}{0.05}$$

has a standard normal distribution ($\mu = 0$, $\sigma = 1$).

This is an example of a z-*score*.

Salinity: both distributions





Notice that the vertical scale changes in concert with the horizontal scale: this is so that the area beneath the bell-curve remains equal to 1.0.

Salinity: part (a)

- Allowed range is 4.92 to 5.08
- z-values corresponding to limits are -1.6 (for 4.92) and 1.6 (for 5.08).
- z has standard normal distribution:

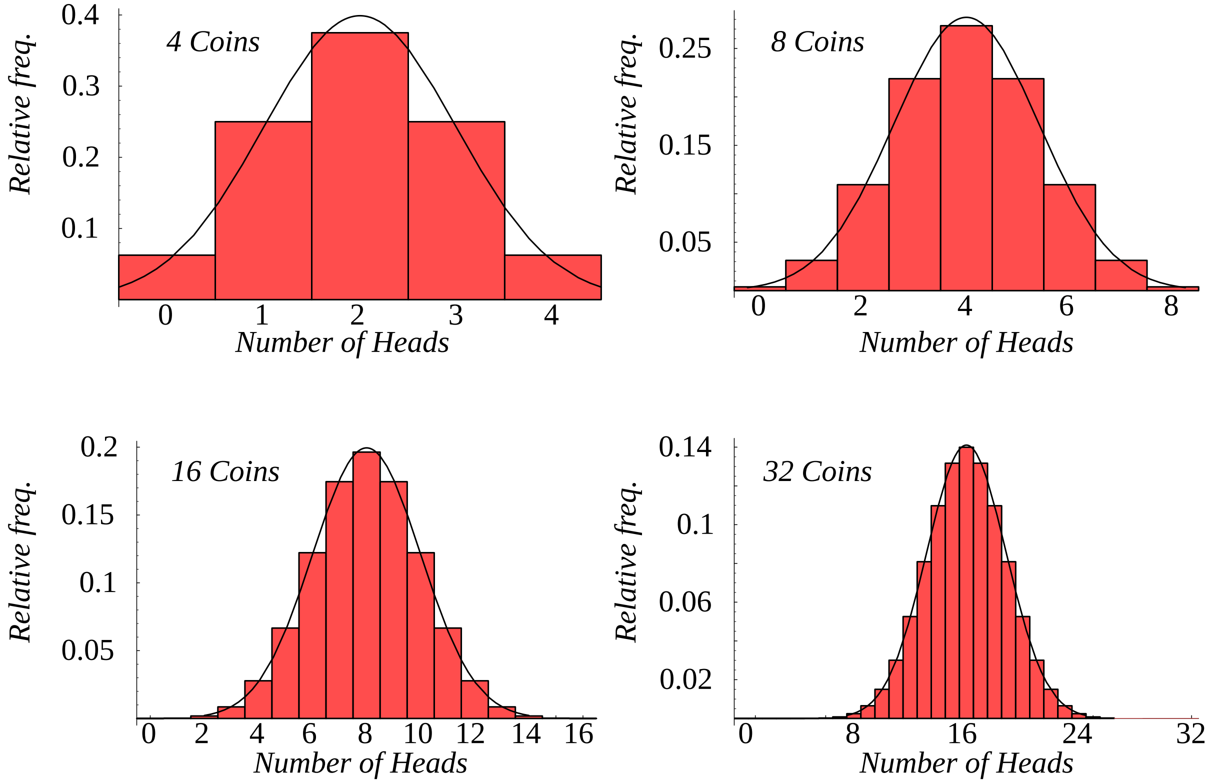
$$
\begin{aligned}
P((s < 4.92) \text{ or } (s > 5.08)\,) = \\
= \quad & P((z < -1.6) \text{ or } (z > 1.6)) \\
= \quad & P(z < -1.6) + P(z > 1.6) \\
\approx \quad & 0.0548 + 0.0548 \; = \; 0.1096
\end{aligned}
$$

- Expected number of rejections is

$$100 \times P(\text{ unacceptable } s) = 100 \times 0.1096 \approx 11 \text{ bottles.}$$

From binomial to normal

For large numbers of trials the binomial distribution is well-approximated by a normal distribution with $\mu = pN$ and $\sigma = \sqrt{p(1-p)N}$:



Approx. is good when both pN and $(1-p)N$ exceed 5.

Application: approximating the binomial

Suppose you are interested in studying a electrical fault known from previous work to be occur in 20 out of every 100 five-year old models of a certain car made by Rover. What is the probability of finding 220 or more examples of the fault in a sample of 1000 cars?

Faulty Rovers

Convert “220 examples in a sample of 1000” into a z-score:

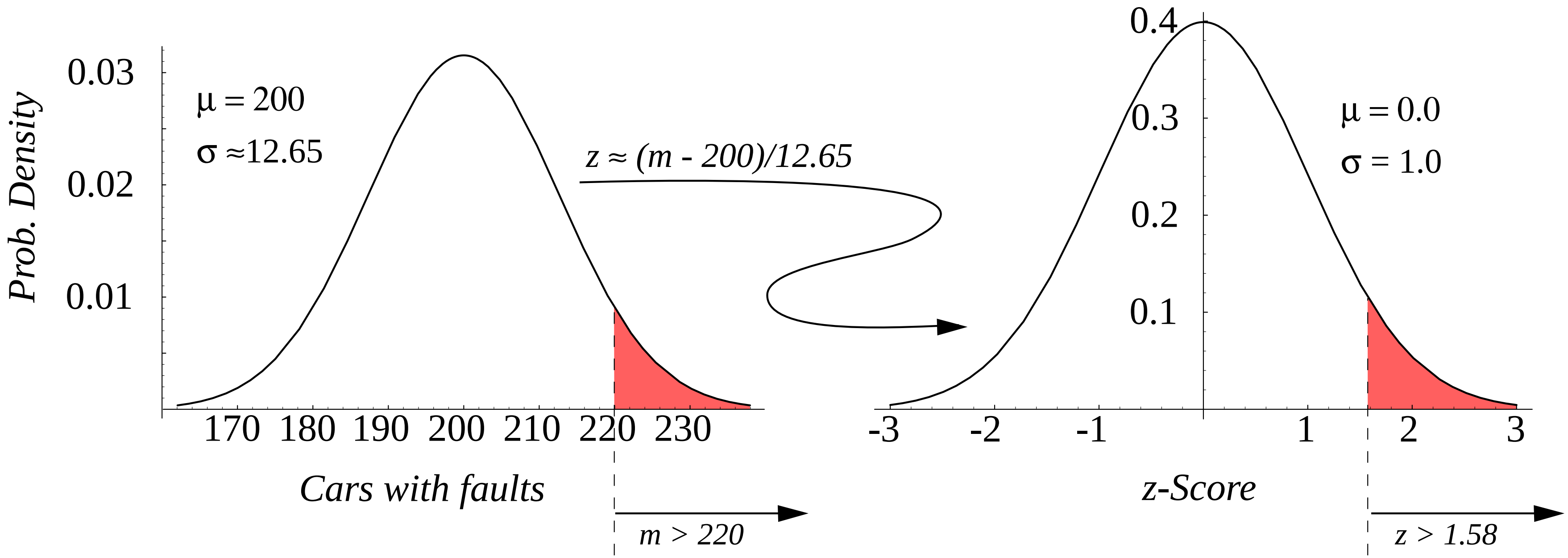
- As $p = 0.2$, the expected number of faults is $\mu = 1000p = 200$.
- Standard deviation is

$$\sigma = \sqrt{1000p(1-p)} = \sqrt{160} \approx 12.65.$$

- So the z-score is:

$$\begin{aligned} z &= \frac{220-\mu}{\sigma} = \frac{220-200}{\sqrt{160}} \\ &\approx (20/12.65) \approx 1.58 \end{aligned}$$

Using the normal approximation



220 examples or more (in a sample of 1000) is a fairly unlikely outcome: $P(z \geq 1.58) \approx 0.0571$.