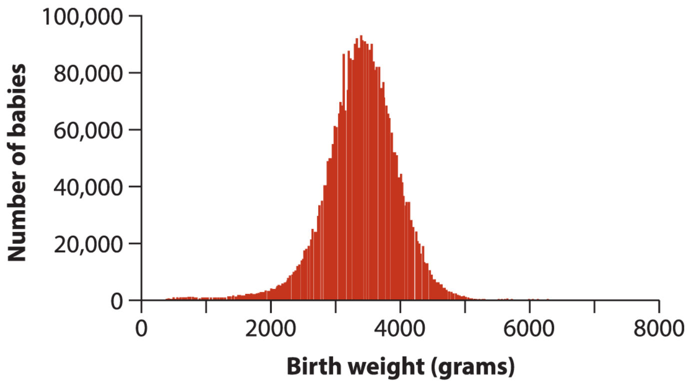
Chapter 10:
The normal distribution

1

Normal distribution

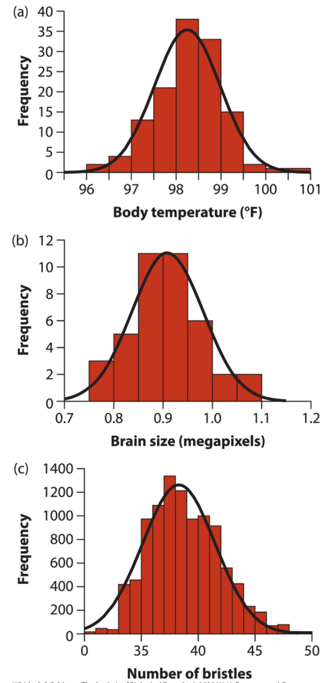
- The **normal distribution** is a continuous probability distribution describing a bell-shaped curve. It is a good approximation to the frequency distributions of many biological variables
- Two parameters: **mean** and **standard deviation**
- Can be used to approximate the sampling distribution of estimates, especially of sample means
  - Many statistical techniques depend on the normal sampling distribution

2



Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

Fig 10.1-1



Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

Fig 10.1-3

3

Probability density of normal distribution

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

- Y can be any real number from positive to negative infinity
- μ is the mean of the distribution, σ is the standard deviation
- Infinite number of distributions, each with its own mean and standard deviation

4

Properties of normal distribution

- Continuous distribution, so probability is measured as area under the curve
- It is symmetrical around its mean (bell-shaped)
- Single mode
- Probability density is highest exactly at the mean

5

Normal distribution symmetry

- For a variable with a normal distribution, about **about two thirds of samples are within one standard deviation**, and **about 95% of samples are within two standard deviations**

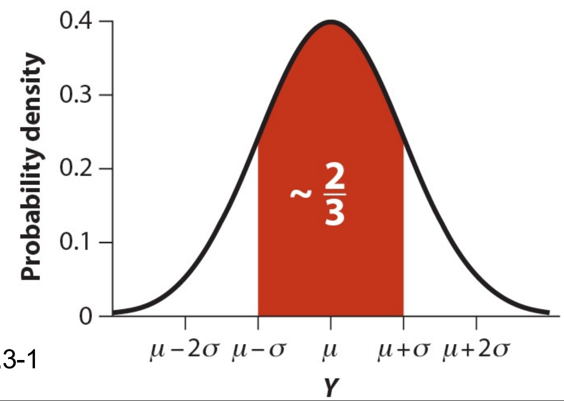


Fig 10.3-1

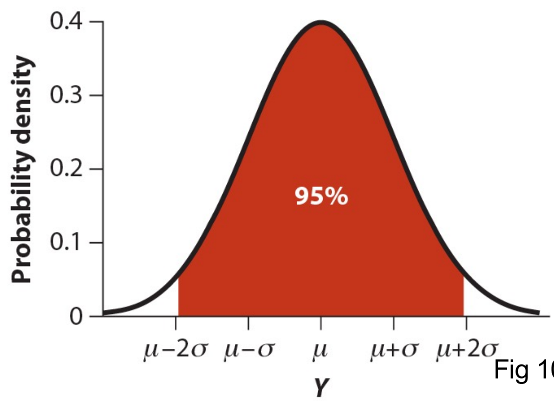


Fig 10.3-2

6

Standard normal distribution

- The **standard normal distribution** is a normal distribution with a mean of zero and a standard deviation of one
- Indicated by symbol Z
- Probabilities for any value of Z can be obtained by a computer function or a statistical table

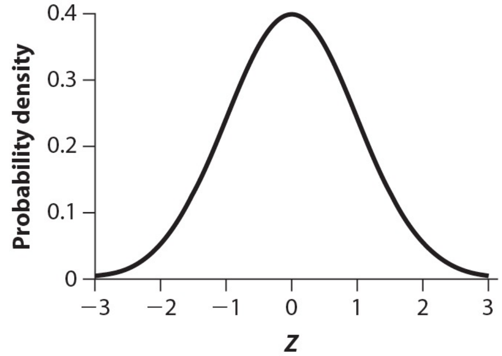


Fig 10.3-2

7

Standard normal table

TABLE 10.4-1 Probabilities of $Z > a.bc$ under the standard normal curve. The digit before and immediately after the decimal (i.e., $a.b$) are given down the first column, and the second digit after the decimal (i.e., c) is given across the top row. The answer highlighted in red shows the probability that $Z > 1.96$. Excerpted from Statistical Table B.

First two digits of *a.bc*	**Second digit after decimal (*c*)**									
	0	1	2	3	4	5	6	7	8	9
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	**0.0250**	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110

Probability that a random draw from the standard normal distribution is above a given Z

8

Standard normal Z probabilities

- Since distribution is symmetrical around mean, the probabilities of negative Z values are the same as the positive Z values
- Pr[Z < -number] = Pr[Z > number]
- Pr[Z < -1.96] = Pr[Z > 1.96] = 0.025

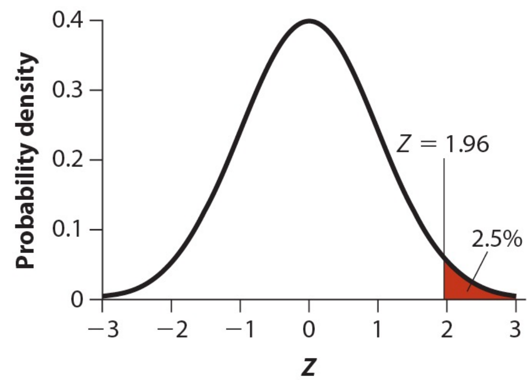
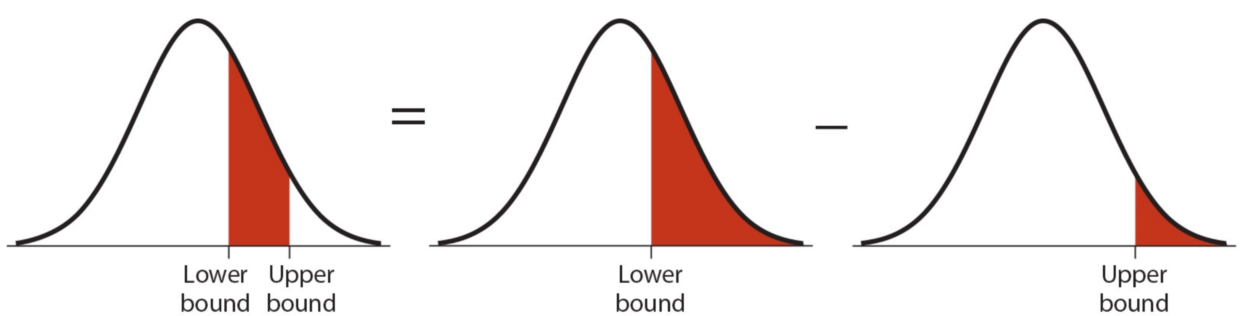


Fig 10.4-2

9

Standard normal Z probabilities

- The probability that Z lies between an upper and lower bound:



Pr[lower bound < Z < upper bound] = Pr[Z > lower bound] – Pr[Z > upper bound

Fig 10.4-3

10

Why standard normal distribution

- There are an infinite number of normal distributions, but they are all similar in shape (i.e., bell curve)
- Able to transform for probability calculation under a standard normal distribution
- A **standard normal deviate**, or Z, tells us how many standard deviations a particular value is from the mean

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

11

Ex 10.4: One small step for man?

- NASA has height restrictions for astronaut pilots
- 157.5 cm < height < 190.5 cm
- What fraction of the American population is excluded?

12

Ex 10.4: One small step for man?

- Height for males 20-29 years old
  - $\mu = 177.6\ cm$
  - $\sigma = 9.7\ cm$
- Those excluded:
  - Pr[height < 157.5] + Pr[height > 190.5]

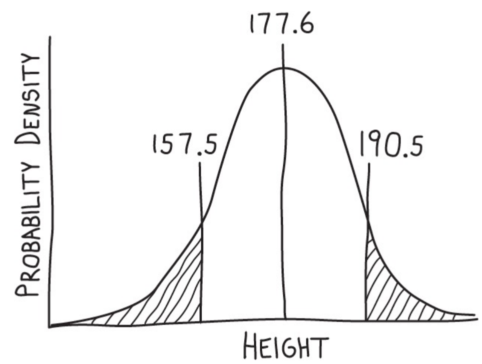


Fig 10.4-4

13

Males too tall?

- Standard normal deviate

$$Z = \frac{Y - \mu}{\sigma} = \frac{190.5 - 177.6}{9.7} = 1.33$$

- 190.5 is 1.33 standard deviations above the mean male height
- What fraction is above this point?
- Now you can use standard normal distribution and Z
  - Pr[Z > 1.33] = 0.09176
  - ~9.2% of males in this age range are too tall

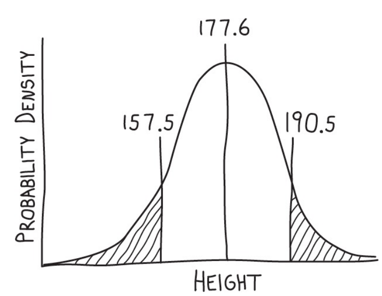


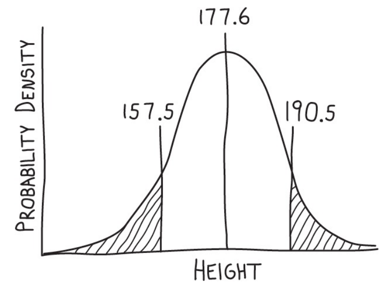
Fig 10.4-4

14

Males too short?

- Standard normal deviate

$$Z = \frac{Y - \mu}{\sigma} = \frac{157.5 - 177.6}{9.7} = -2.07$$



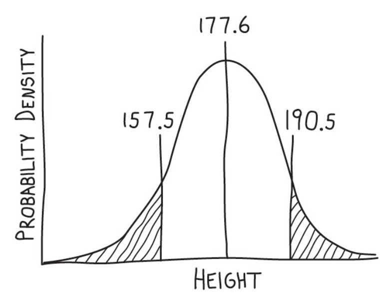
- 127.5 is 2.07 standard deviations below the mean male height
- What fraction is below this point?
- Now you can use standard normal distribution and Z
  - Pr[Z < -2.07] = Pr[Z > 2.07 = 0.01923
  - ~1.9% of males in this age range are too short

Fig 10.4-4

15

Excluded males

- Either too short or too tall (addition rule)
  - Pr[height < 157.5] = 0.01923
  - Pr[height > 190.5] = 0.09176
  - 0.01923 + 0.09176 = 0.11099



- For women using same method?
  - ~28.8% too short, ~0.3% too tall

Fig 10.4-4

16

17

Normal distribution of sample means

- Normal distribution can be used to describe the sampling distribution of many estimates, including the mean
- The **sampling distribution** is the probability of all values for an estimate that we might obtain when we sample a population (from chapter 4)

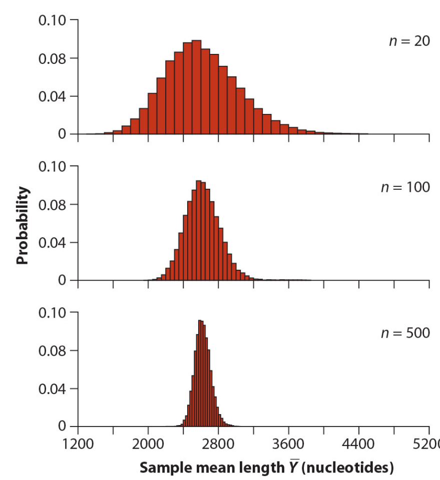
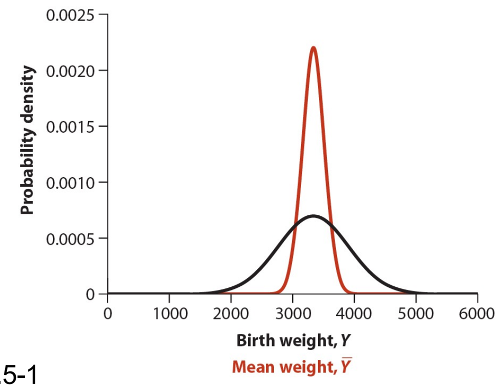


Fig 4.1-3

18

Normal distribution of sample means

- **If a variable Y has a normal distribution in the population, then the distribution of the sample means $\overline{Y}$ is also normal**



Black = normal distribution of population

Red = normal distribution of sample means

$\overline{Y}$ = sample mean

Standard error of the mean

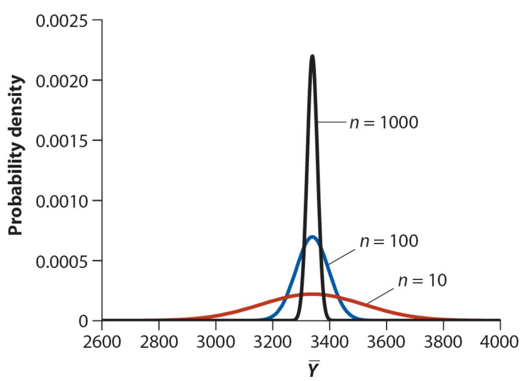
$$\sigma_{\overline{y}} = \frac{\sigma}{\sqrt{n}}$$

Fig 10.5-1

19

Normal distribution of sample means

- **If a variable Y has a normal distribution in the population, then the distribution of the sample means $\overline{Y}$ is also normal**



Increasing the sample size reduces the sample error and increases the precision of the distribution of sample means

Fig 10.5-2

20

Normal distribution of sample means

- **If a variable Y has a normal distribution in the population, then the distribution of the sample means $\overline{Y}$ is also normal**
- This fact will be used in statistical tests in upcoming chapters
- https://www.zoology.ubc.ca/~whitlock/Kingfisher/SamplingNormal.htm
  – Web visualization of sampling from a normal distribution

21

Calculating probabilities of sample means

- Birth weight dataset:
  – μ = 3339 g
  – σ = 573 g
- Suppose random sample of 80 babies produces a mean of 3370
  – What is the probability getting mean of 3370 or greater?

$$Z = \frac{\overline{Y} - \mu}{\sigma_{\overline{Y}}} \qquad \sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{573}{\sqrt{80}} = 64.1$$

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}} \qquad Z = \frac{\overline{Y} - \mu}{\sigma_{\overline{Y}}} = \frac{3370 - 3339}{64.1} = 0.48 \qquad \Pr[Z > 0.48] = 0.316$$

22

23

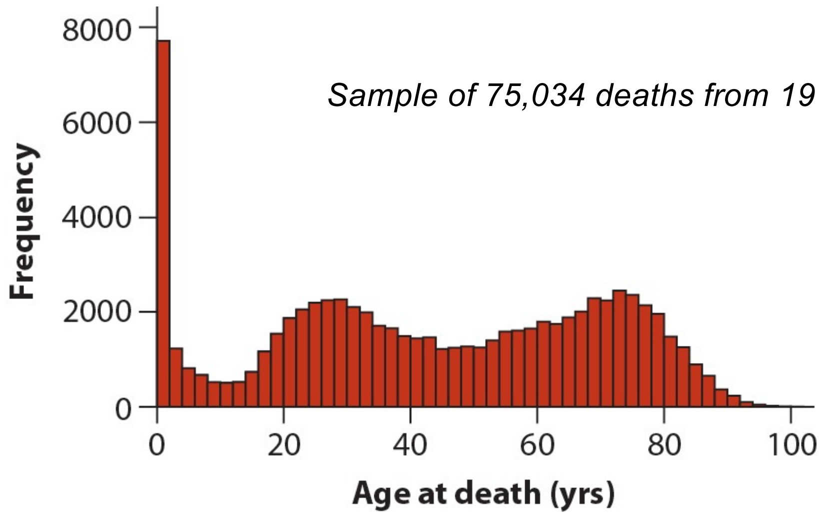
Central limit theorem

- According to the **central limit theorem**, the sum or mean of a large number of measurements randomly sampled from a non-normal distribution is approximately normal
- Many statistical tests that assumes the distribution of means is normal
- So central limit theorem allows use of these tests *even if the distribution of the sampled population parameter is not normal*
- How many samples needed depends on shape of the distribution of observations

24

Ex 10.6: Young adults and the Spanish flu

- Spanish flu (1918-1920)



Sample of 75,034 deaths from 1918 Switzerland

Fig 10.6-1

25

Ex 10.6: Young adults and the Spanish flu

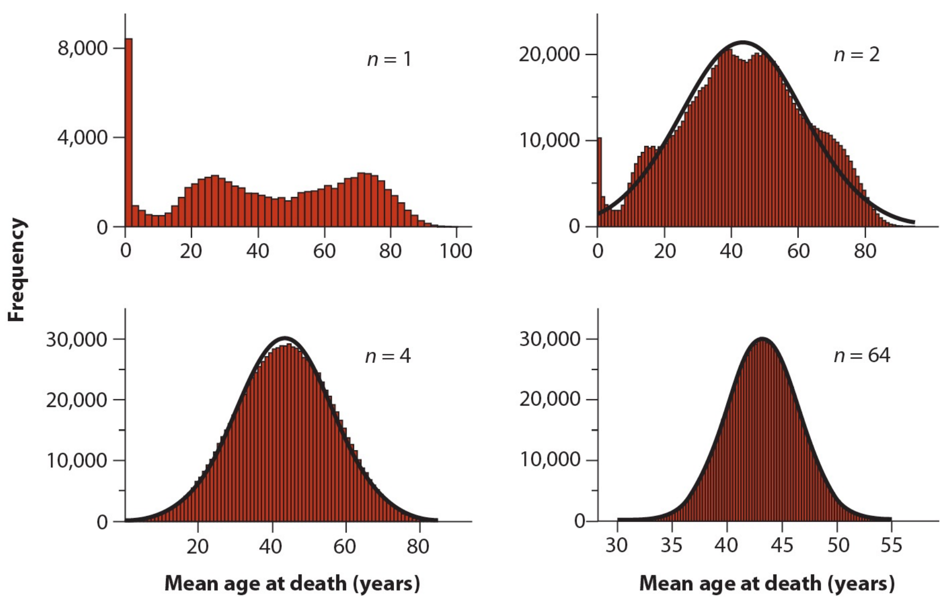


Fig 10.6-2

26

Ex 10.6: Young adults and the Spanish flu

- Spanish flu (1918-1920)
- https://www.zoology.ubc.ca/~whitlock/Kingfisher/CLT.htm
  - Web visualization of central limit theorem

Fig 10.6-1

27

Notes

- Skipping section 10.7

28