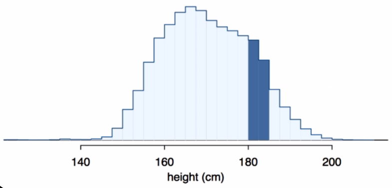
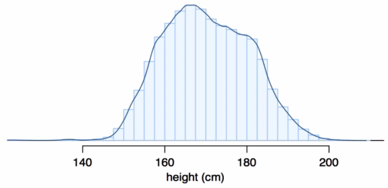
Continuous distributions

Heights of US adults
Shaded: 180-185 cm (about 5'11" - 6'1")



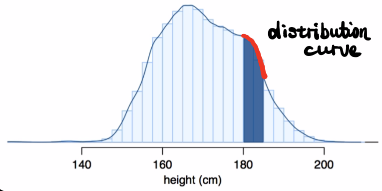
Continuous distributions

Imagine drawing a smooth curve that traces the shape of the histogram



Continuous distributions

P(random US adult is 180-185 cm) = shaded area under the curve.



P(exact value) = 0

The probability of an exact value in a continuous distribution is always 0

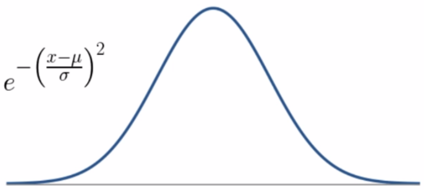
We can think of this as the mathematical representation of the fact that observations have limited precision

If we measure to 1 decimal place, 180.4 cm really represents 180.35-180.45

Normal Distribution

- Unimodal, symmetric, bell-shaped
- Many variables are nearly normal
- $\boxed{N(\mu, \sigma)}$ → Normal distribution with mean μ and standard deviation σ

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

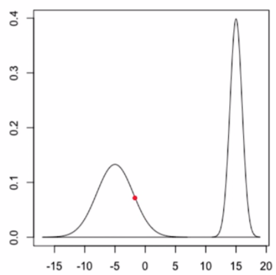


Example

The two normal distributions graphed below have:

μ=-5, σ=3

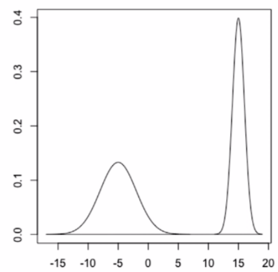
μ=15, σ=1



Shifting and scaling

All normal distributions have the same shape, modified by a shift and a scale

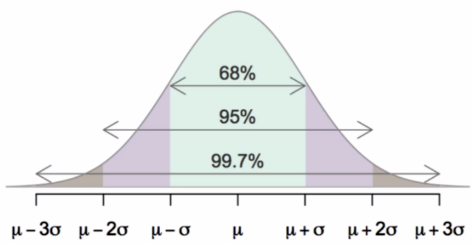
We can go from one to another by moving the center point and squishing/stretching

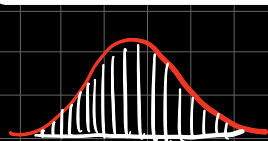


68-95-99.7 Rule

For normally distributed data:

- about 68% falls within 1 SD of the mean
- about 95% falls within 2 SD of the mean
- about 99.7% falls within 3 SD of the mean

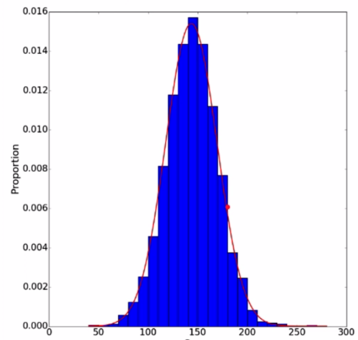




area = probability

Example: Basketball scores

Histogram of 1706 college basketball game scores



N (0, 1) Standard ND

overlay curve follows the histogram ⇒ normal distribution

Z-score (or standardized score)

Number of standard deviations an observation is above or below the mean

$$Z = \frac{x - \mu}{\sigma} = \frac{\text{observation} - \text{mean}}{\text{standard deviation}}$$

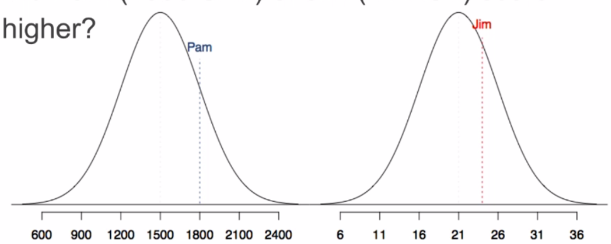
Can use Z-scores to compare observations from different distributions

Z scores typically used for normal distributions but can be used for any distribution

Comparing Across Distributions

SAT scores ~ N(1500, 300)
ACT scores ~ N(21, 5)

Did Pam (1800 SAT) or Jim (24 ACT) score higher?

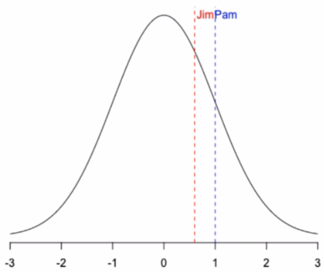


Comparing Across Distributions

SAT scores ~ N(1500, 300)
ACT scores ~ N(21, 5)

Pam (1800 SAT)
Z = (1800 - 1500) / 300
1 SD above the mean

Jim (24 ACT)
Z = (24 - 21) / 5
0.6 SD above the mean



Example

- The distribution of basketball scores is approximately N(143.5, 25.9). Find the Z-scores of games with scores of 120, 180, and 205 points.

  Answer: -0.91, 1.41, 2.37

- Another quantity we could look at is the difference between the winning and losing scores. About 1.3% of games have a Z-score over 3.0 for this quantity. What does this tell you about the distribution of the score difference? 0.15%

  Answer: the distribution is not normal

5%

mean

$$76 - 1.65 * 4 = 70$$