PSY 201: Statistics in Psychology

Lecture 08
Normal distribution
Business decisions.

Greg Francis

Purdue University

Fall 2019

NORMAL DISTRIBUTIONS

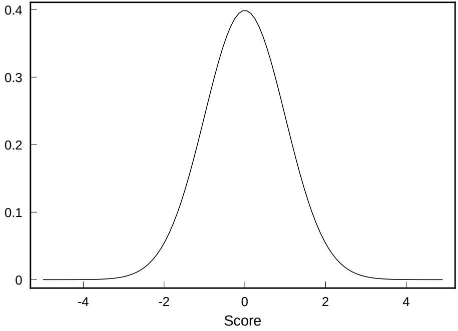
- when the distribution is a normal distribution, we can describe the distribution by just specifying
  - Mean: $\overline{X}$
  - Standard deviation: s
  - Noting it is a normal distribution
- that's all we need!
- That's part of our goal: describe distributions

USE

- same as all other distributions
  - identify key aspects of the data
  - percentiles
  - percentile rank
  - proportion of scores within a range
  - ...
- make it easier to interpret data significance!

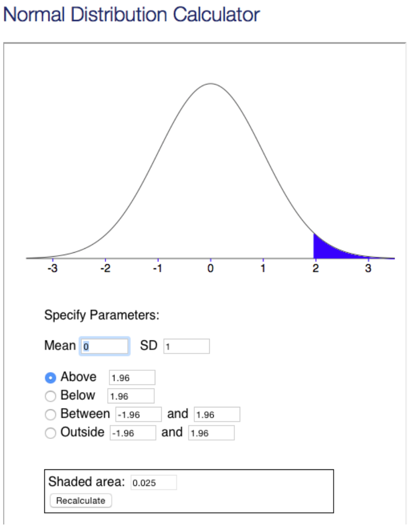
AREA UNDER CURVE

- proportional to the frequency of scores within the designated endpoints
- suppose you want to know the proportion of scores between the mean and another score (z-score)



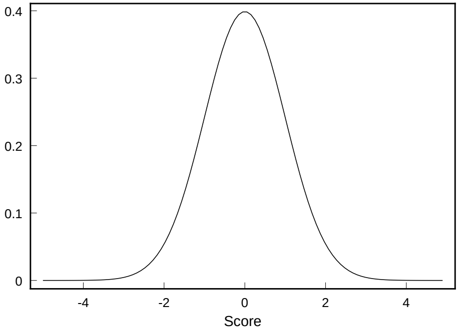
AREA UNDER CURVE

- solving for the area requires calculus and numerical analysis (ack!)
- fortunately, we can also use computers
- our text provides



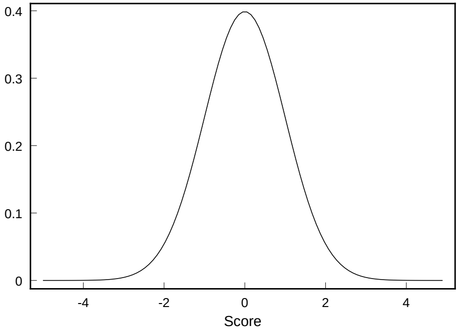
CALCULATOR

- how would you find the area between $z = -0.3$ and $z = 2.4$?



CALCULATOR

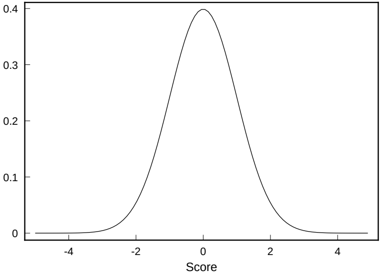
- how would you find the area below $z = 1.4$?



PROPORTIONS

- suppose you have 250 scores from a test that are normally distributed
- you want to know how many scores are **between** 1.0 standard deviations below the mean and 1.5 standard deviations above the mean
- two steps
  1. calculate the area under the standard normal between $z = -1.0$ and $z = 1.5$.
  2. convert the area under the curve to number of scores

PROPORTIONS



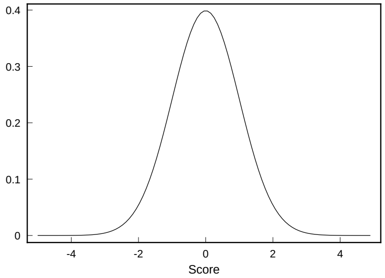
- We find that 77.45% of the scores lie between one standard deviation below the mean and 1.5 standard deviations above the mean
- so how many scores are in that range?
- multiply the total number of scores (250) with the percent in the range (decimal form)

$$(0.7745) \times (250) = 193.625 \approx 194$$

PROPORTIONS

- suppose you have 250 scores from a test that are normally distributed
- you want to know how many scores are **below** 0.5 standard deviations above the mean, and how many scores are **beyond** 2.5 standard deviations above the mean.
- two steps
  1. calculate the area under the standard normal below $z = 0.5$ and above $z = 2.5$.
  2. convert the area under the curve to number of scores

PROPORTIONS

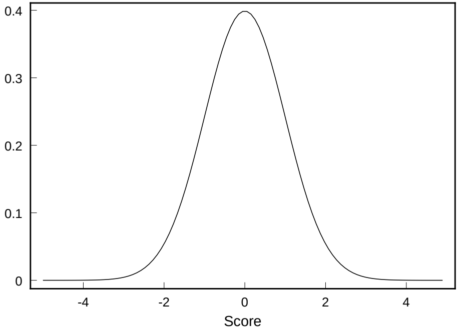


- We find that 69.77% of the scores lie below 0.5 standard deviation above the mean or beyond 2.5 standard deviations above the mean
- so how many scores are in that range?
- multiply the total number of scores (250) with the percent in the range (decimal form)

$$(0.6997) \times (250) = 174.925 \approx 175$$

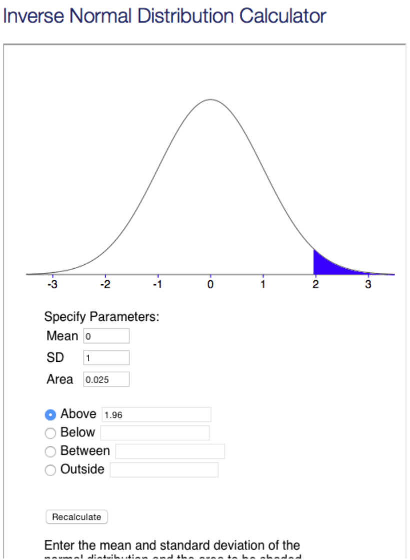
PERCENTILES

- Xth percentile is score for which X percent of scores fall at or below
- 50th percentile is the median (and the mean!)



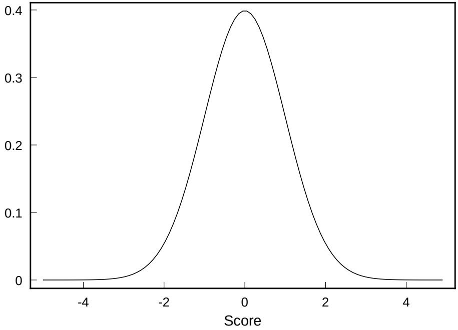
PERCENTILES

- The Inverse Normal Calculator gives the *z*-score that corresponds to different areas
- Click "Below" to make it fill in from the left side



EXAMPLE

- to find P_{75} for a standard normal curve, enter Area$= 0.75$
- and find that the corresponding z-score is 0.674



- what about P_{25}?

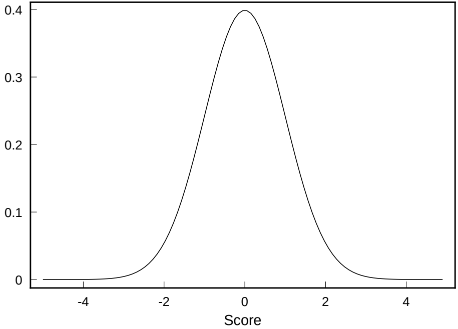
EXAMPLE

- Symmetry!

$$P_{25} = -P_{75}$$

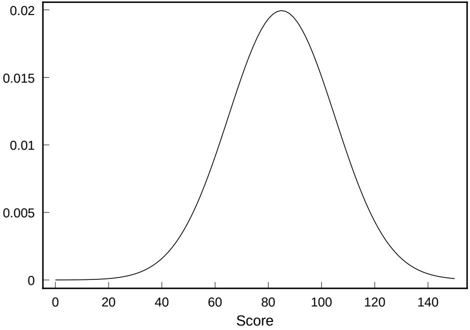
- in general for $X < 50$,

$$P_X = -P_{100-X}$$



CONVERSION

- suppose you have a **normal** distribution with a mean of 85 and a standard deviation of 20
- how would you find the 70th percentile?



z-scores

- Indirect way:
  1. Calculate percentile of z-score distribution.
  2. Convert z-score back to a raw score.
- from z-score we can calculate

$$X = (s)(z) + \overline{X}$$

- the online-app shows that for a standard normal, $P_{70} = 0.5244$, so

$$X = (20)(0.5244) + 85 = 95.49$$

- Or, just change the mean and the standard deviation of the normal distribution in the on-line app

BUSINESS DECISION

- suppose you are part of a company manufacturing what you think will be the "next big thing" in men's pants

BUSINESS DECISION

- You want to produce pants that will fit the center of the distribution of men's waist sizes
- There is no need to make pants for men with really small or really large waists because there are so few of such people
- According to the National Health and Nutrition Examination Survey the distribution of waist circumference is approximately normal with (in centimeters)

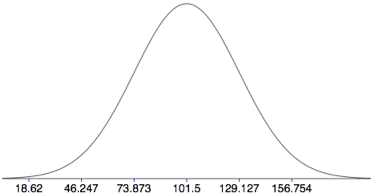
$$\mu = 101.5$$

- (around 40 inches)

$$\sigma = 27.6$$

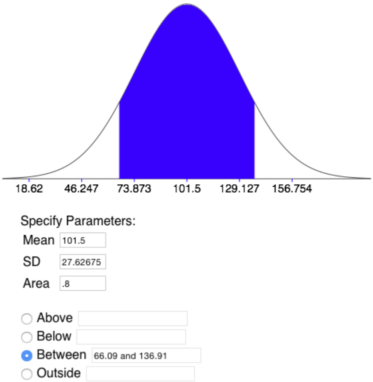
BUSINESS DECISION

- What size waists do you manufacture to cover the middle 80% of the distribution of waist sizes?



BUSINESS DECISION

- What size waists do you manufacture to cover the middle 80% of the distribution of waist sizes?



- (Obviously, there are more things to consider: costs, how many sizes, customer preferences,...)

BUSINESS DECISION

- You plan to set up a canoe business on the Wabash River. You want to purchase canoes that will be able to carry 90% of 3-person families. Canoes that carry more weight cost more, so you want canoes that hold the lower 90% of people (mother, father, child)
- Statistics (pounds)
  - Adult women:

$$\mu = 168.5, \ \sigma = 67.7$$

  - Adult men:

$$\mu = 195.7, \ \sigma = 68.0$$

  - Children (18 year old):

$$\mu = 179.4, \ \sigma = 89.7$$

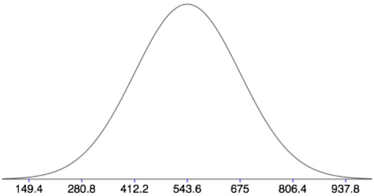
BUSINESS DECISION

- For a *family* we add the means and the variances
- Family:

$$\mu = 168.5 + 195.7 + 179.4 = 543.6$$

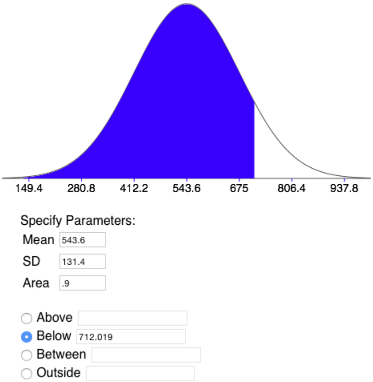
$$\sigma^2 = (67.7)^2 + (68.0)^2 + (89.7)^2 = 17261$$

$$\sigma = 131.4$$



BUSINESS DECISION

- To be able to hold 90% of families, you need a canoe that holds weight of the 90th percentile



CONCLUSIONS

- normal distribution
- area under curve
- proportions
- percentiles

NEXT TIME

- percentile ranks
- examples

A statistical approach to assigning grades.