Chapter 6 Lecture Notes

The Standard Deviation as a Ruler and the Normal Model

Comparing quantitative variables

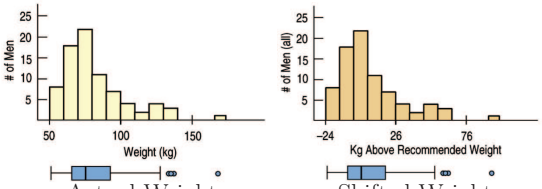
A student gets a 67/75 on exam 1 and a 64/75 on exam 2, but is told she did better on the second exam relative to the class. How can this be?

- Both sets of exam scores **exhibit variation**.

Shifting Data

We Shift Data by...

A shift from men's actual weights in kilograms (LEFT) to kilograms above recommended weight (RIGHT)

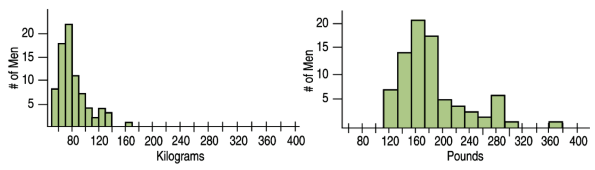


(a) Actual Weight (b) Shifted Weight

Rescaling Data

We Re-scale Data by...

Rescaling from weight in kilograms (LEFT) to weight in pounds (RIGHT).



How do we standardize observations

We can compare individual data values to their mean, relative to their standard deviation by

$$z = \frac{y - \bar{y}}{s}$$

EXAMPLE

A student gets a 67/75 on exam 1 and a 64/75 on exam 2, but is told she did better on the second exam relative to the class. Below is a summary of the class performance on each exam.

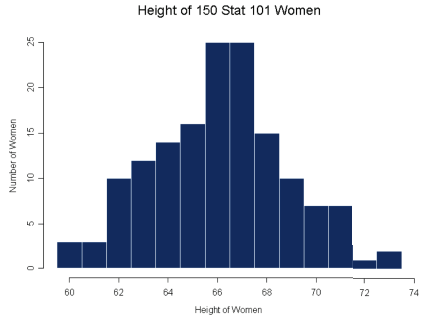
- Exam 1
  - Mean: $\bar{y} = 59.5$
  - Standard Deviation: $s = 8.61$
- Exam 2
  - Mean: $\bar{y} = 50.1$
  - Standard Deviation: $s = 11.86$

How did the student perform with respect to the class on each exam?

When is a z-score BIG?

- There is no universal standard for z-scores, but we can use the **Normal Model** to provide a measure on how extreme a z-score is.

The Normal Distribution is characterized by the following two parameters:



- Notation: to denote the normal distribution we use

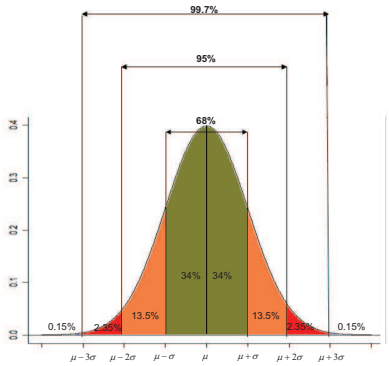
EXAMPLE:

_____ denotes a normal distribution with mean ___ and standard deviation ___, while _____ denotes a normal distribution with mean ___ and standard deviation ___.

- To denote that a variable (e.g. heights, SAT scores, etc.) follows a normal distribution we write

The 68-95-99.7 Rule

For a variable that follows a _____, we have that



Example: The length of human pregnancies follows a normal distribution with mean $\mu = 266$ days and a standard deviation of $\sigma = 16$ days.

1. How long do the middle 95% of all pregnancies last?

2. How long do the shortest 16% of all pregnancies last (at most)?

3. How long do the longest 0.15% of all pregnancies last (at least)?

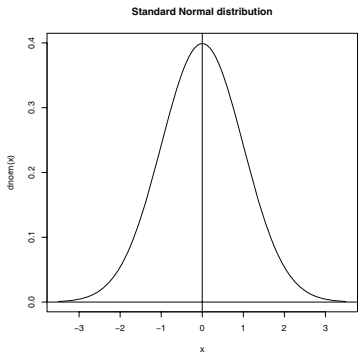
Example: The Standford-Binet Intelligence Quotient (IQ) Test, is a test used to measure a person's general intelligence. The IQ test is designed to so participants' scores follow a normal distribution with mean $\mu = 100$ and a standard deviation of $\sigma = 16$.

1. Identify the range of IQ scores for the central 68% of the population.

2. What percentage of people have an IQ below 52?

3. What percentage of people have an IQ above 116?

Standard Normal Distribution



For the standard normal distribution, the **proportion of observations** falling into a specified range is tabulated.

Knowing the mean and the standard deviation of a normal distribution allows us to determine

We can standardize any given normal distribution to a standard normal distribution using

$$z = \frac{y - \mu}{\sigma}$$

Finding z-scores and corresponding proportions/areas under the normal curve

Why are z-scores helpful?

- IQ's follow a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 16$
- heights of males follow approx. a normal distribution with mean $\mu = 70$ inches and $\sigma = 3$

Who is more unusual? — A man being 73 inches tall or a man having an IQ of 124?

Normal Values table

Once we know the corresponding z-score of an observation we can look up the overall **proportion** (percentage) of men in that population having a height of 73 inches or more.

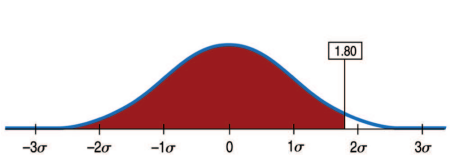
$\Rightarrow$ need to know how to read Table Z (Table of the Standard Normal Distribution)

$\Rightarrow$ Table Z in Appendix D

Note, in the following the terms *proportion*, *probability*, *percentage*, and *area* are all interchangeable, i.e.

proportion = probability = percentage = area

We can find the area to the left of $z = 1.80$ using Table Z.

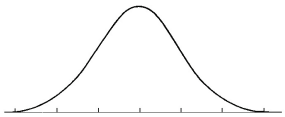


z	.00	.01
1.7	.9554	.9564
1.8	.9641	.9649
1.9	.9713	.9719

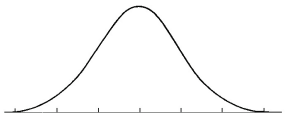
USING TABLE A TO FIND PROPORTIONS UNDER THE NORMAL CURVE

consider the following situations:

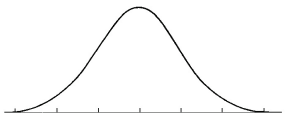
1. What proportion of observations is below $z = -1.67$, i.e. what is the probability of observing a z-score of -1.67 or less?



2. What proportion is below $z = 1.67$?



3. What proportion of observations is greater than $z = 1.67$?

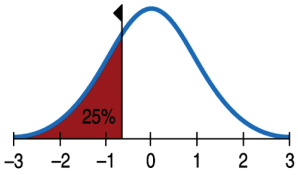


4. What proportion is less than $z = -2.00$ and greater than $z = 2.00$?

Backwards Calculations

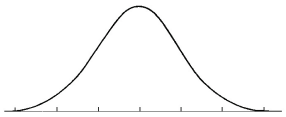
We can also work backwards — given a certain percentile (or proportion), what is the corresponding value of y?

Example: What z-score represents the first quartile in the Normal model?

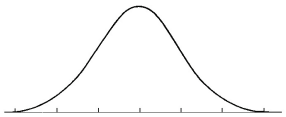


z		.07	.08
−0.6		.2514	.2483

1. What z-score does the 30^{th} percentile correspond to?



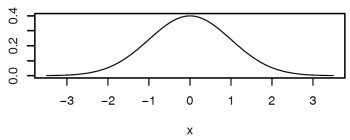
2. What z-scores bound the middle 60%?



Applications of the Normal Distribution

1. State the problem, i.e. state the mean μ, the standard deviation σ and the value of the observation y

2. draw picture, i.e. shade area of interest under curve.

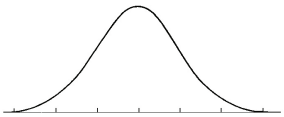


3. To calculate the proportion of observations falling in a certain region
   (a) standardize y, i.e. find the corresponding z-score
   (b) use Table A to find the shaded area

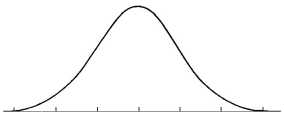
4. To calculate a data value corresponding to a given percentile
   (a) Find the z-score corresponding to given percentile in Table A
   (b) Convert the z-score to a data value with $y = z*\sigma+\mu$

Examples — Male Heights $\sim$ N(70, 3).

What percent of men are shorter than 66 inches?



What value corresponds to the median?

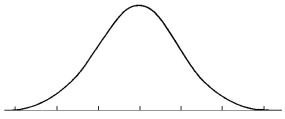


What percent of men are taller than 74 inches?

What percent of men are between 68 and 71 inches tall?

What value corresponds to Q_1?

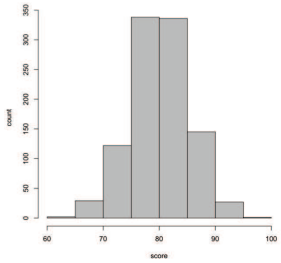
What value corresponds to Q_3?



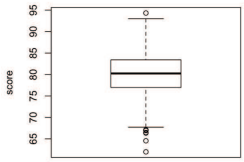
Checking Normality of Data

- In general it is quite risky to assume normality without looking at the data and verifying normality

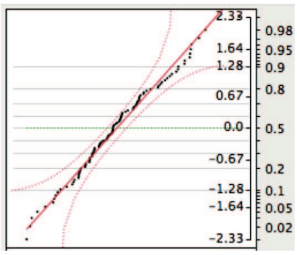
• Histograms



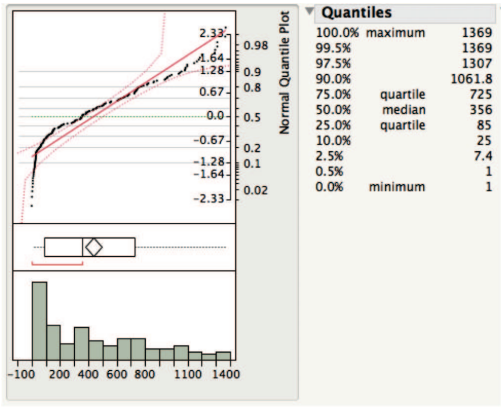
• Boxplots/5 Number Summaries



• Normal Probability Plot



Cannot use the Normal Model.



Should use the Normal Model.