Math 140
Introductory Statistics

Professor Silvia Fernández

Lecture 3
Based on the book *Statistics in Action*
by A. Watkins, R. Scheaffer, and G. Cobb.

Using the calculator (TI-83)

- For more information go to www.keymath.com/x7065.xml and look for the *Calculator Notes* for Chapters 0, 1, and 2.
- You should know how to
  - Generate a list of *n* random integer numbers between *min* and *max*.

Example: To generate a list of 7 integer numbers between 2 and 10 (inclusive) type

MATH | PRB | 5.randInt(| Enter | 2, 10, 7) | Enter

Using the calculator (TI-83)

- How to generate a list of *n* random numbers between 0 and 1 (exclusive).

Example: Generate 5 random numbers between 0 and 1.

MATH | PRB | 1.randInt(| Enter | 5) | Enter

- How to store a list of numbers.

Example: Store the previous list of 5 random numbers between 0 and 1 on L_1.

2^{nd} | ANS | → | 2^{nd} | L_1

Using the calculator (TI-83)

Example: Store the list 1,2,3,4,5 to L_1.

STAT | 1.Edit | Enter

Move to the first row of column L_1 using the arrows.

Type each of the numbers on the list followed by ENTER.

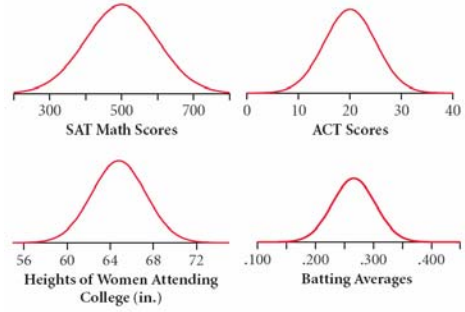
- Compute binomial coefficients.

Example: Compute 10 choose 3.

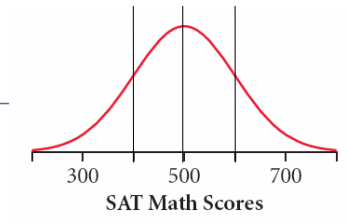
10 | MATH | PRB | nCr | Enter | 3

Practice

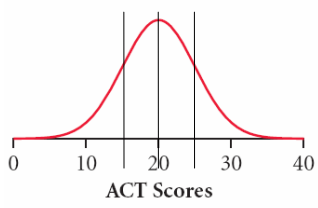
- P3. For each of the normal distributions in below, estimate the mean and standard deviation visually, and use your estimates to write a verbal summary of the form "A typical SAT score is roughly (mean), give or take (*SD*) or so."



Display 2.13 Four distributions that are approximately normal.



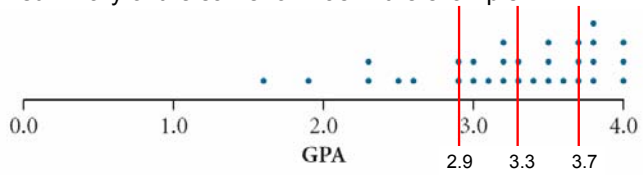
- Mean ~ 500
- SD ~ 100
- A typical SAT score is roughly 500, give or take 100 or so.



- Mean ~ 20
- SD ~ 5
- A typical ACT score is roughly 20, give or take 5 or so.

Practice

- P4. Estimate the median and quartiles for the distribution of GPAs in Display 2.7 on page 34. Then write a verbal summary of the same form as in the example.



Display 2.7 Grade-point averages of 62 statistics students. Each dot represents two points.

Lower quartile ~ 2.9
Median ~ 3.3
Upper quartile ~ 3.7

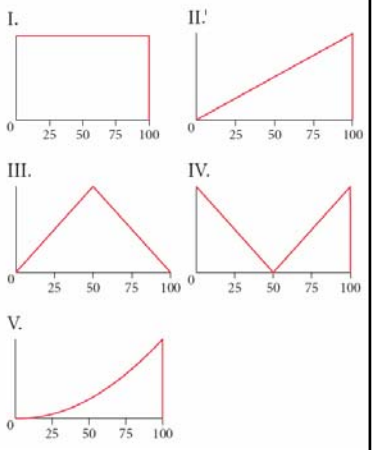
The middle 50% of the GPAs of statistic students were between 2.9 and 3.7, with half above 3.3 and half below.

Practice

P5. Match each plot in Display 2.14 with its median and quartiles (the set of values that divide the area under the curve into fourths).

- a. 15, 50, 85
- b. 50, 71, 87
- c. 63, 79, 91
- d. 35, 50, 65
- e. 25, 50, 75

IV
II
V
III
I



Quantitative vs. Categorical Data

- **Quantitative**: Data about the cases in the form of numbers that can be compared and that can take a large number of values.
- **Categorical**: Data where a case either belongs to a category or not.

Example (D6)

Mammal	Gestation Period (days)	Average Longevity (years)	Maximum Longevity (years)	Speed (mi/h)	Wild (1 = yes; 0 = no)	Predator (1 = yes; 0 = no)
Baboon	187	20	45	*	1	0
Bear, grizzly	225	25	50	30	1	1
Beaver	105	5	50	*	1	0
Bison	285	15	40	*	1	0
Camel	406	12	50	*	1	0
Cat	63	12	28	30	0	1
Cheetah	*	*	14	70	1	1
Chimpanzee	230	20	53	*	1	0
Chipmunk	31	6	8	*	1	0
Cow	284	15	30	*	0	0
Deer	201	8	20	30	1	0
Dog	61	12	20	39	0	1

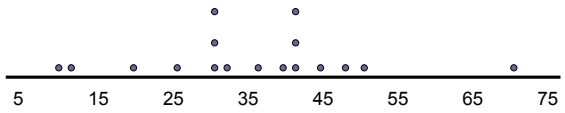
- Quantitative variables: Gestation period, average longevity, maximum longevity, speed.
- Categorical variables: Wild, predator.

Different ways to visualize data

- Quantitative Variables
  - Dot Plots
  - Histograms
  - Stemplots
- Categorical Variables
  - Bar Graphs

Dot Plots

- Each dot represents the value associated to a case.
  - Dots may have different symbols or colors.
  - Dots may represent more than one case.

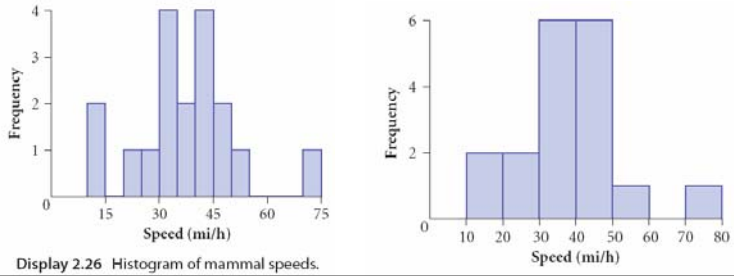


Dot Plots

- Dot Plots work best when
  - Relatively small number of values to plot
  - Want to keep track of individuals
  - Want to see the shape of the distribution
  - Have one group or a small number of groups that we want to compare

Histograms

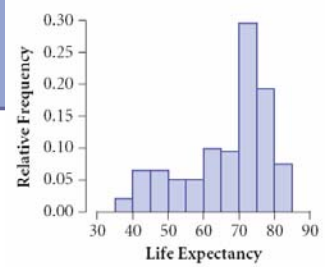
- Groups of cases represented as rectangles or bars
- The vertical axis gives the number of cases (called **frequency** or **count**) for a given group of values.
- By convention borderline values go to the bar on the right.
- There is no prescribed number for the width of the bars.



Display 2.26 Histogram of mammal speeds.

Relative Frequency Histograms

- The height of each bar is the proportion of values in that range. (always a number between 0 and 1)
- The sum of the heights of all the bars equals 1.
- To change a regular histogram to a relative frequency histogram just divide the frequency of each bar by the total number of



This histogram shows the relative frequency distribution of life expectancies for 203 countries around the world.

How many countries have a life expectancy of at least 70 but less than 75 years?
.30 x 203 = 60.9

What proportion of the countries have a life expectancy of 70 years or more?
.30+.19+.07 = .56 = 56 %

Histograms (Relative Frequency)

- Histograms work best when
  - Large number of values to plot
  - Don't need to see individual values
  - Want to see the general shape of the distribution
  - Have one or a small number of distributions we want to compare
  - We can use a calculator or computer to draw the plots