Chapter 1

Logan | Nick

1.1 Populations, Samples, and Processes

Statistics provides methods for organizing and summarizing data and for drawing conclusions from that data

Def

Data : a collection of facts

Def

Population : A well defined collection of objects for which we wish to obtain info

Def

Census : When desired info is obtained from every member of the population

- problems : Time, money, practical

Def

Sample : A subset of the population

1.

You want the home price in Edwardsville

- Fewer well trained appraisers gives better results than many poorly trained

2. Tree Age Study

Testing is destructive, so a sample is better

Def

variable : any characteristic whose clue may differ from one subject to another.

- denote with low letters

Note

- Don't say $McDonald's = 10$

- Do say $x =$ the length of the tibia bone in 10 year old boys.

Def

univariate data : result from making observations of 1 variable

- these variable can be qualitative / quantitative

Def

Bivariate data : when observations are made on each of 2 variables for each individual

- (weight.mpg) of cars

Def

Multivariate data : observations made on many variables

- patient data

Ex

Labor force, sample 60,000, find population + sample

- population = labor force, sample size = 60,000 households

Branches of Stats

1. Descriptive Stats : data are collected and you wish to summarize and describe features of the data (graphs, numerical summaries)

2. Inferential stats : data is collected from a sample and used to draw a conclusion about the population

   - confidence intervals, hypothesis test, prediction, etc. . .

Types of sampling

- Simple random sampling : random choice / draw of the hat sampling
- Systematic sampling : selecting every k^{th} member of the population

- Cluster sampling : divide population into groups, then select some of these groups @ random

- Stratified sampling : divide population into groups. Find subgroups of groups (strata) and then draw random sample in strata

- Convenience sampling : sampling in the most convenient way

  - best to avoid , but a good starter

Notate

sample size : n

- For a dataset with n observations on some variable x, the individual observations will be denoted as $x_1, x_2, \ldots, x_n$.

1.2

Stem and leaf plots

Ex

$(54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 46, 41, 34, 22)$
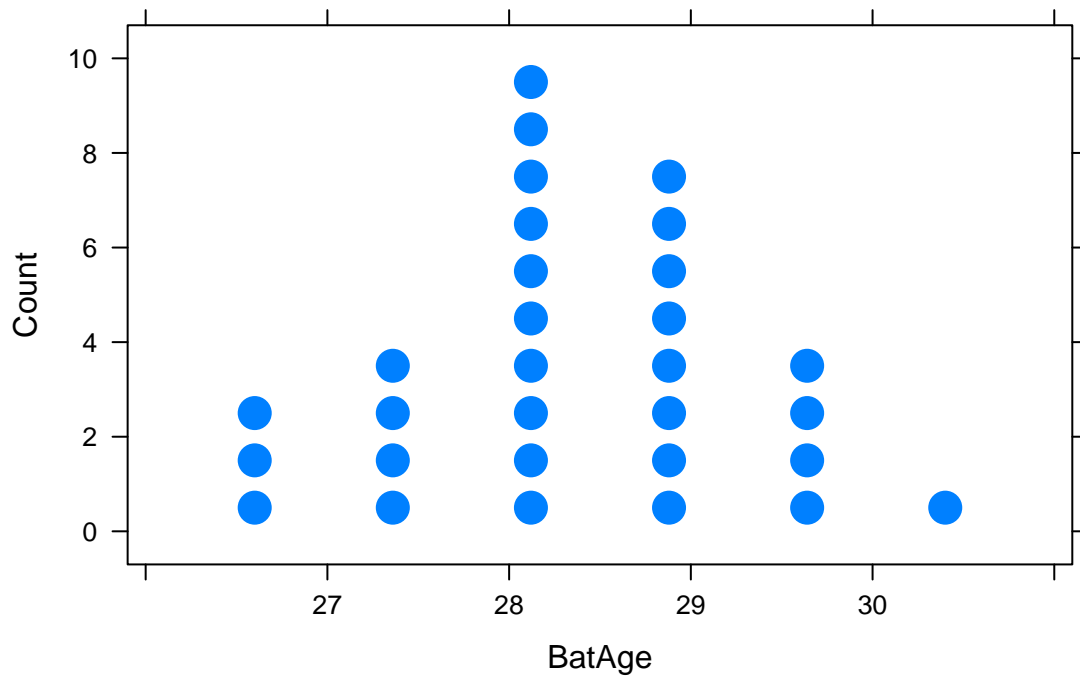
During these problems it helps to first orgranize the numebrs in the list first

$$
\begin{array}{l}
2 \mid 2, 5 \\
3 \mid 4, 5 \\
4 \mid 1, 1, 6, 6, 6, 7, 9 \\
5 \mid 4, 4, 9 \\
6 \mid 0
\end{array}
$$

Dot plots

Histograms

left blank for nick to fill in

Skewed (Right and left)

add a dataset to show?

Bell

add a dataset to show?

Flat uniform

add a dataset to show?

nonsymmetric

add a dataset to show?

bimodal symmetric

add a dataset to show?

1.3

Def

mean : numerical value of average

Notate

Sample mean : $\overline{x}$

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum\limits_{i=1}^{n} x_i}{n}$$

Notate

Population mean : μ

- avg of all values in the entire pop.

Ex

$2, 2, 5, 3, 8, 9, 2, 3, 1$

$$\overline{x} = \frac{\sum\limits_{i=1}^{10} x_i}{10} = 3.6$$

The mean is inappropriate in some cases b/c of outliers.

- this makes the mean a **nonresistant measure**

Def

Median : middle value /avg of 2 middle values when sorted

Notate

Median : $\widetilde{x}$

- if $n = odd$, median is at $\dfrac{n+1}{2}$

- if $n = even$ median are b/n $\dfrac{n}{2}$ & $\dfrac{n+1}{2}$

Notate

Population Mean : $\widetilde{\mu}$

1.4 Measures of Variability

One way to describe a distribution is by using the standard deviation

Quartiles

- Q_1 - lower quartile separates bottom 25%
- Q_2 - median middle 50%
- Q_3 - upper quartile separates upper 25%

Ex

2, 2, 5, 1, 3, 8, 9, 2, 31

SORT

$$1, 1, 2, 2, 2$$
$$3, 3, 5, 8, 9$$

$\widetilde{x} = 2 + 3 = 2.5$

Five number summary

- Find min, Q_1, median, Q_3, max

Note : If median is found in list, use it in both top half and lower half.

Ex : 2 2 5 1 3 8 9 2 3 1 100

$\overline{x} = \frac{36 + 100}{11} \approx 12.36$

Sort to find median. $\widetilde{x} = 3$.

Mean vs. Median

- median is the equal parts point

- mean is the balance point

Notate

Trimmed mean : $\overline{x}_{tr}$

- compromise b/n the mean & median

- to find it, remove top & bottom 10%, then calculate the mean