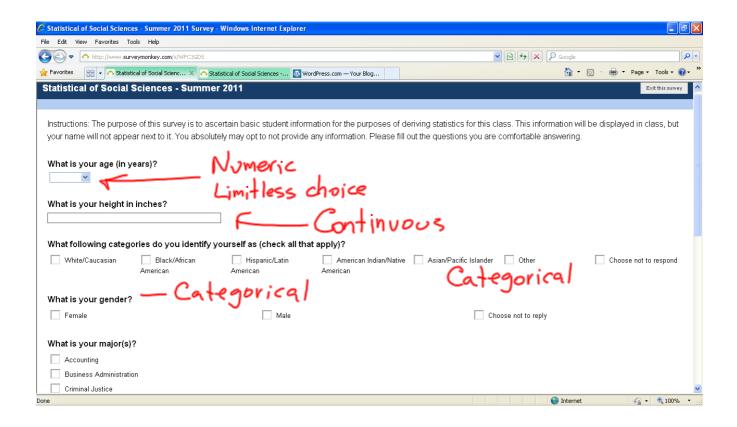## Last Class

- Scientific Method

- Statistics
  - Descriptive
    - Inferential

- Data
  - Observations
  - Variables
    ↓   ↓
  Continuous   Categorical

Measures of Centrality

19	1	→ Minimum
20	2
21	4
22	2
25	2
26	1
27	1
28	1
29	2
30	2
32	2
33	1
35	1
39	1
45	1
47	1
48	2
50	1
70	1

- Mean

$$\bar{x} = 31.55$$

- Median
- Mode

→ Maximum

$$N = 29$$

Mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{N} x_i$$

number $\frac{1}{n}$, $\ldots$ N $\sum_{i=1}^{i-1}$ variable v

$$x_i = \frac{x_1 + x_2 + x_3 + \cdots x_N}{N}$$

i^{th} observation

ex:

$N = 29$

$$19 + 20 + 20 + 21 + 21 + 21 + \cdots + 70$$

$$= 915 = \sum_{i=1}^{N} x_i$$

$$\Rightarrow \frac{915}{29} = \boxed{31.55} = \bar{x}$$

Avg. Age

$\rightarrow$ Mean is the expected value

— Median

└ Middle value

└ Half above, half below.

①	19
2	20
3	20
4	21
5	21
6	21
7	21
8	22
9	22
10	25
11	25
12	26
13	27
14	28
15	29
	29
	30
	30
	32
	32
	33
	35
	39
	45
	47
	48
	48
	50
	70

$$N = 29$$

$$\text{Median} = \frac{(N+1)}{2}$$
$$= i^{th} \text{ observation}$$

ex: $N = 29$
$$\text{Median} = \frac{29+1}{2} = 15$$
$$\Rightarrow 29$$

↳ If odd number of observations, then formula works.

↳ If even number of observations, then average the two middle values.

i.e., suppose that $N = 10$, then
$$\frac{10+1}{2} = 5.5, \text{ so}$$

average the 5th & 6th observation

Mode

⤷ The most frequent observation

ex: 4 students were 21, thus its the mode.

ex:

Mean = 31.55

Median = 29

Mode = 21

Measures of variability

Class 1: A, B, C - variability
$$\bar{x} = B$$
Class 2: B, B, B - no variability
$$\bar{x} = B$$

	1st Yr	2nd	3rd
Stock A:	-$50	$100	-$10
Stock B:	$13	$14	$13

Same average return
$$= \$13.\bar{3}$$

- Range
- Quantile
- Variance
- Standard Deviation

Range

$$= \text{Maximum value} - \text{Minimum value}$$

Stock A: $100 - $-50

$= $150

Stock B: $14 - $13

$= $1

Quantile

↳ Generic term used to denote splitting data into several equally numbered groups.

→ 4 groups ⇒ quartiles
→ 5 groups ⇒ quintiles
→ 10 groups ⇒ deciles

Class: F, F, C, C, C, C, B, B, A, A, A, A

$$\overline{x} = 2.5$$

Quartiles

$\{F\}$ F, $\{C\}$ min = F

$\textcolor{red}{Q_I = C}$

C, C, $\{C\}$

$\textcolor{green}{median = C/B}$

$\{B\}$ B, $\{A\}$

$\textcolor{blue}{Q_3 = A}$

$\{A, A, A\}$ Max = A

ex:

F, C, C, C, C, C, B, B, B, B, B, A

$$\bar{x} = 2.5$$

Quartiles:
{F, C, C}   min = F

{C, C, C} $Q_1 = C$

{B, B, B}   median = C/B

{B, B, A} $Q_3 = B$

max = A

Median = Half above, half below

Quartiles = Groups of 25% of observations

Quintiles (5 groups): Groups of 20% of observations

Deciles (10 groups): Groups of 10% of observations

Finding Quantiles

N = 12, for quartiles, that's a group of 3.

$$\Rightarrow \frac{N}{\# \text{ of groups}}$$

.

Interquartile Range

$$Q_3 - Q_1$$

ex. 1: A - C

$$(4.0) - (2.0) = 2.0$$

ex. 2: B - C

$$(3.0) - (2.0) = 1.0$$

Interquintile Range

$$Q_4 - Q_1$$

ex: 1, 2, 3, 4, 5

Quartiles?

$$\frac{5}{4} = 1.25$$

$$\{1 \to 1.5\}$$

Variance

↳ Variability around the mean.

$$\sum_{i=1}^{N} \frac{(x_i - \overline{x})^2}{N}$$

ex: (ages)

$(19 - 31.55)^2$
$+ (20 - 31.55)^2$
$\vdots$
$+ (70 - 31.55)^2$

Ex: Grades as GPA (e.g. 1.0, 2.0, etc)

1.0, 2.0, 2.0, 3.0, 4.0
(F)

Calculate: Mean, Median, Mode, Range, Variance.

$$\text{Mean} = \frac{1+2+2+3+4}{5}$$

$$= \frac{12}{5} = \boxed{2.4}$$

$$\text{Median} \Rightarrow \frac{5+1}{2} = \frac{6}{2} = 3^{rd} \text{ obs.}$$

$$\Rightarrow \boxed{2.0}$$

$$\text{Mode} = \boxed{2.0} \text{ (appears twice)}$$

$$\text{Range} = 4.0 - 1.0 = \boxed{3.0}$$

Variance $\cdot N = 5, \ \overline{X} = 2.4$

$x_1: \ 1.0 - 2.4 = -1.4^2 = 1.96$
$x_2: \ 2.0 - 2.4 = -0.4^2 = 0.16$
$x_3: \ 2.0 - 2.4 = -0.4^2 = 0.16$
$x_4: \ 3.0 - 2.4 = 0.6^2 = 0.36$
$x_5: \ 4.0 - 2.4 = 1.6^2 = 2.56$

$$= 5.2$$

$$5.2/5 = \boxed{1.04}$$

Average grade is 2.4, but varies 1.04^2

Standard Deviation

$$= \sqrt{Variance}$$

ex: $\sqrt{1.04} = 1.02$

$\bar{x} = 2.4$, but varies 1.02

$$\sigma = \text{Standard deviation} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}$$

$\uparrow$
Sigma

$$= \sqrt{Variance}$$