

Basic Statistics

Stuff you can use!

Percentages

- ◆ Percent = Part/Whole
- ◆ Part = 92 Men, Part = 108 Women
- ◆ Whole = $92+108 = 200$
- ◆ Percent of Men = $92/200 = .46$ or 46%
- ◆ Percent of Women = $108/200 = .54$ or 54%
- ◆ Note to self: If you calculate percentages of each "part" of a group, those parts should add up to 100 percent!

Rates

- ◆ Sometimes we can rely on raw numbers alone. Other times, in order to accurately compare things we need to "level the playing field"... So we compute rates.

State	# of Violent Crimes in 2002
Alabama	19,931
Alaska	3,627
Tennessee	41,562
Texas	126,018
Washington	20,964

Rates, cont.

- ◆ Calculating a rate in this case helps us to compare across different sized populations.

State	# of Violent Crimes (in 2002)	State Population (in millions)	Crime Rate (per 100,000 inhabitants)
Alabama	19,931	4.5	444.2
Alaska	3,627	0.6	563.4
Tennessee	41,562	5.8	716.9
Texas	126,018	21.8	578.6
Washington	20,964	6.1	345.4

Source: FBI's UCR data


$VC/SP \times 100,000 = \#$ of violent crimes per 100,000

Descriptive Statistics

- ◆ We use numbers to describe things.
- ◆ Sounds fun, right?!


Measures of Central Tendency

- ◆ Mean
- ◆ Median
- ◆ Mode




Mean

- ◆ Otherwise known as the “average”
- ◆ Equals the sum of all the Xs, divided by the number of Xs you have.
- ◆ Six people’s ages: 18, 23, 19, 22, 20, 18
- ◆ Mean = $(18+23+19+22+20+18)/6$
 $= 120/6$
 $= 20$




Median

- ◆ The middle number in a list of numbers.
- ◆ Organize the numbers from low to high. If there is no exact middle value, calculate the mean of the middle two numbers.
- ◆ Six people’s ages: 18, 23, 19, 22, 20, 18
- ◆ Reorder: 18, 18, 19, 20, 22, 23
- ◆ No exact middle value – so calculate mean of middle two numbers:
 $(19 + 20)/2 = 19.5$




Mode

- ◆ The most frequently occurring value.
- ◆ If no number appears more than once, then there is no mode!
- ◆ Six people’s ages: 18, 23, 19, 22, 20, 18
- ◆ Mode = 18




Note to self!

- ◆ The mean can be very sensitive to extreme values (outliers).
- ◆ The median is not.
- ◆ Seven people’s ages: 18, 23, 19, 22, 20, 18, 90
 - Mean = 30
 - Median = 20
 - Mode = 18



Excel: Central Tendencies

- ◆ Mean:
 $=\text{AVERAGE}(A2:A6)$
- ◆ Median:
 $=\text{MEDIAN}(A2:A6)$
- ◆ Mode:
 $=\text{MODE}(A2:A6)$



Measures of Variation

- ◆ Range
- ◆ Variance
- ◆ Standard Deviation

Range

- ◆ The difference between the highest and lowest number.
- ◆ Six people's ages: 18, 23, 19, 22, 20, 18
– Range = 23 - 18 = 5
- ◆ Six people's ages: 58, 63, 59, 62, 60, 58
– Range = 63 - 58 = 5

Variance & Standard Deviation

- ◆ You need the values of your variable, the number of cases, and the mean of that variable. That's it!
- ◆ Variance = $\sigma^2 = \frac{\sum(x_i - \mu)^2}{n-1}$
- ◆ Six people's ages (x_i): 18, 23, 19, 22, 20, 18
- ◆ Number of cases (n) = 6
- ◆ Mean (μ) = 20

Variance, cont.

- ◆ Variance = $\sigma^2 = \frac{\sum(x_i - \mu)^2}{n-1}$

Six people's ages (x_i): 18, 23, 19, 22, 20, 18
 Number of cases (n) = 6
 Mean (μ) = 20

(18-20)² = 2² = 4 4 + 9 + 1 + 4 + 0 + 4 = 22
 (23-20)² = -3² = 9
 (19-20)² = -1² = 1 Variance = $\sigma^2 = 22/5 = 4.4$
 (22-20)² = 2² = 4
 (20-20)² = 0² = 0
 (18-20)² = 2² = 4

Standard Deviation

- ◆ It's just the square root of the variance!
- ◆ Standard deviation = $\sigma = \sqrt{\sigma^2}$
- ◆ In this case $\sigma = \sqrt{\sigma^2} = \sqrt{4.4} = 2.098$

Excel: Variance and Standard Deviation

- ◆ Variance:
=VAR(A2:A11)
- ◆ Standard Deviation:
=STDEV(A2:A11)

Need to know...

- ◆ How to calculate by hand:
 - percentages, rates, mean, median, mode, range
- ◆ How to calculate in Excel:
 - mean, median, mode, variance, standard deviation