

Part I: Vocabulary List

Please define each of the following terms from the information on the statrek website. When asked to provide an example of the word, provide a unique example of the word NOT given on the website.

1. **Categorical Variables** (qualitative variables): use words, described by names/labels
Example: Types of flowers ~ peony, rose, tulip, carnation
2. **Quantitative Variables**: use numbers, can be measured
 ↳ continuous: any value from min to max
 discrete: gaps between min & max
Example: Temperature in °C (from -273°C to 100°C)
3. **Univariate Data**: only one variable is being examined in the study
4. **Bivariate Data**: the relationship/connection between two variables is examined in the study
5. **Median**: the middle number of an odd string of values arranged from smallest → largest; the average of 2 middle numbers of an even string of values
6. **Mean**: average of a set of values; sum of values divided by number of values

$$\mu = \frac{\sum x}{N} \quad \bar{x} = \frac{\sum x}{n}$$
7. **Population**: involves all elements of the group in question
Example: Seniors at KSJC
8. **Sample**: specific part(s) of the population
Example: Seniors at KSJC who are school athletes
9. **Center**: point on a graph where half of the data are on either side (right/left); can be located at the median of data set

10. Spread: data set's variability; clustered - less spread

wider range = more spread

11. Symmetry: when both halves of the graph are mirrored when divided at the center

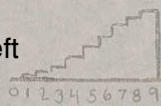
12. Unimodal and Bimodal: unimodal - distributions with one distinct peak

bimodal - distributions with two distinct peaks

13. Skewness: when the graphic display has more data on one side compared to the other

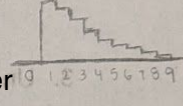
Sketch Skewed left:

When the "tail" is on the left side of the data (fewer observations on the left)



Sketch Skewed right:

When the "tail" is on the right side of the data (fewer observations on the right)



14. Uniform: observations are equally spread across the graphic display; no distinct peaks

15. Gaps: an area of the distribution where there aren't any observations

16. Outliers: extreme values that differ significantly from the other data & are distinct

17. Dot plots: used to compare frequency counts; a graph composed of dots where they are stacked in columns & their heights depict the

frequency of observations

18. Difference between a bar chart and histogram

Bar Chart: columns represent a categorical variable; spaced

Histogram: columns represent a continuous, quantitative variable; can describe

19. Stem plots: show quantitative data usually from small sets to show how individual values are distributed; stem - left entries, leaves - right entries

20. Boxplots: display quantitative data & measures of variability and shape; composed of quartiles, a "box" as the body, median, whiskers which show the smallest & largest non-outliers, & outliers

21. Quartiles: divide a rank-ordered data set into 4 equal parts of a boxplot; Q_1 is the left side of the box, Q_2 the median, & Q_3 the right side of the box

22. **Range**: measure of variation in a data set; difference between the largest and smallest variable

23. **Interquartile Range (IQR)**: measure of variation based on the quartiles of a boxplot; the range of the box ($Q_3 - Q_1$)

24. **Parallel boxplots**: data from 2 distributions are shown on the same chart with the same measurement scale

Maybe it goes without saying, but the data sets are displayed as boxplots.

25. **Difference between a frequency table and relative frequency table**

Frequency Table: shows whole numerical frequency counts for a specific category of a qualitative variable

Relative Frequency Table: shows relative frequencies, such as percentages or proportions, for categories of a qualitative variable

26. **Parameter**:

measurable characteristic of a population

↳ mean, standard deviation

27. **Statistic**: measurable characteristic of a sample

28. **Marginal Distribution** (marginal frequencies): the entries in the "total" row & column of a two-way frequency table
↳ entries in body: joint frequencies

29. **Conditional Distribution** (conditional frequencies): table entries of a two-way relative frequency table

30. **Segmented Bar Chart**: chart that has distinct bars for each level of a categorical variable & is divided into segments which correspond to a proportion/percentage

31. **What are the W's of data (might need to do your own digging for this!)**

Who: cases/individuals the information is about

What: variables, what has been measured or is being looked at

When: context, how recent & up to date the information is

Where: context, information on what things involved are like

Why: shows purpose of the information

How: how information is collected

AP Statistics Summer Homework Part 2: Practice Problems

Categorical or Quantitative?

- | | | |
|-----------------|------------------|------------------|
| 1) quantitative | 6) categorical | 11) categorical |
| 2) quantitative | 7) categorical | 12) categorical |
| 3) categorical | 8) quantitative | 13) categorical |
| 4) quantitative | 9) quantitative | 14) categorical |
| 5) quantitative | 10) quantitative | 15) quantitative |

Statistic - what is that?

Mean: 36.438

Minimum: 3

Maximum: 70

Median: 36

Q1: 25.5

Q3: 50.5

Range: 67

IQR: 25

3, 9, 9, 22, 29, 32, 32, 33, 39, 39, 42, 49, 52, 58, 65, 70

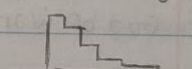
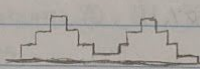
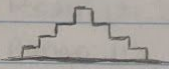
- 16) A bar chart is usually spaced & its columns represent a categorical variable. A histogram doesn't have spaces between columns and they represent a continuous, quantitative variable.

unimodal

bimodal

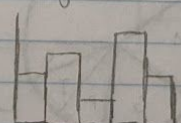
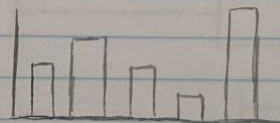
skewed left

skewed right



bar chart

histogram



Standard Deviation: a number that represents the variation in a group; usually from the mean

$$s = \sqrt{\sum (x_i - \bar{x})^2 / (n-1)}$$

$$\sigma = \sqrt{\sum (x_i - \bar{x})^2 / N} \rightarrow \# \text{ of elements in pop}$$

\swarrow \searrow
*i*th element \swarrow \searrow
 from pop \swarrow \searrow
 pop mean

Variance: a numerical value to show how widely individuals in a group vary

$$s^2 = \sum (x_i - \bar{x})^2 / N$$

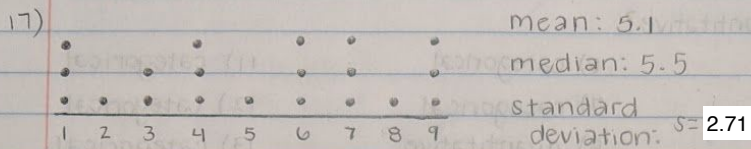
$$\sigma^2 = \sum (x_i - \bar{x})^2 / (n-1)$$

the expectation of the squared deviation of a random variable from its mean.

Standard Score (z-score): the number of standard deviations something is from the mean

$$z = (x - \mu) / \sigma \rightarrow \text{standard deviation}$$

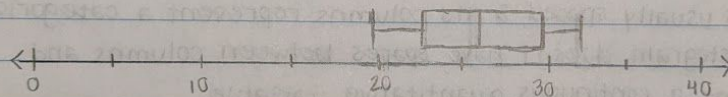
\swarrow \searrow
 value of \swarrow \searrow
 element \swarrow \searrow
 population mean



Standard Deviation is a frustrating calculation to do by hand. We will usually use technology to aid us.

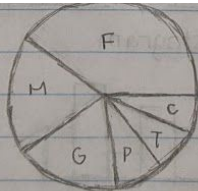
18) 19, 22, 23, 23, 23, 26, 26, 27, 28, 29, 29, 31, 32
 mean: 26 IQR: $29 - 23 = 6$
 Five-number summary: minimum - 19, maximum - 32, median - 26,
 $Q_1 - 23, Q_3 - 29$
 standard deviation: $s = 3.83$

$$s = \sqrt{\frac{(19-26)^2 + (22-26)^2 + 3(23-26)^2 + 2(26-26)^2 + (27-26)^2 + (28-26)^2 + 2(29-26)^2 + (31-26)^2 + (32-26)^2}{12}}$$



19) Fortnite: 38%, PUBG: 9%, The Show: 8%, Madden 2017: 20%,
 God of War: 19%, Call of Duty: 6%

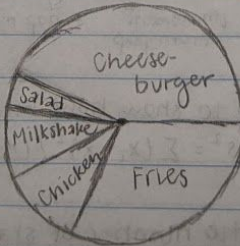
Diego's Video Game Usage



Titles are important!

20) Cheeseburger: 40%, Milkshake: 10%, chicken: 13%, Fries: 33%, Salad: 5%

Food Item's Sold



Titles are important!

Shopping Spree!

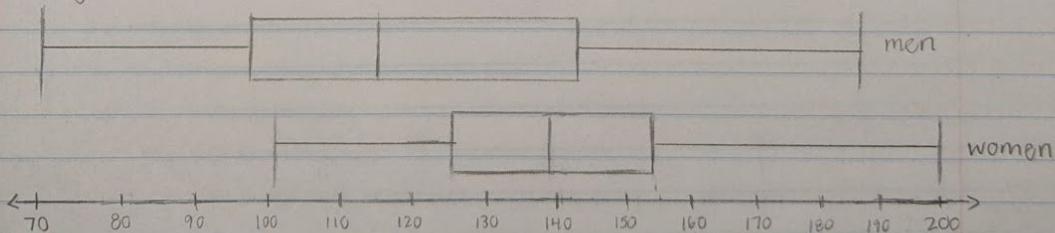
Money (\$) spent in store

(tens of dollars)	Stems	Leaves	(dollars)
9	3		93
8	3 6 6		83, 86, 86
7	0		70
6	1		61
5	0 3 5 9		50, 53, 55, 59
4	1 3 4 5 5 7 9		41, 43, 44, 45, 45, 47, 49
3	2 5 6 9 9		32, 35, 36, 39, 39
2	0 0 0 1 2 3 4 5 5 6 6 8 8 8 8		20, 20, 20, 21, 22, 23, 24, 25, 25, 26, 26, 28, 28, 28, 28
1	1 3 4 5 6 7 7 8 8 9		11, 13, 14, 15, 16, 17, 17, 18, 18, 19
0	3 9 9		03, 09, 09

SSHA Scores

a) Women: 101, 103, 109, 115, 126, 126, 129, 137, 137, 140, 140, 152, 154, 154, 165, 165, 178, 200
 Mean: 141.056, Minimum: 101, Q1: 126, Median: 138.5, Q3: 154, Maximum: 200
 Range: 99, IQR: 28

Men: 70, 75, 88, 91, 92, 104, 108, 109, 113, 114, 115, 115, 126, 132, 140, 146, 151, 169, 180, 187
 Mean: 121.25, Minimum: 70, Q1: 98, Median: 114.5, Q3: 143, Maximum: 187
 Range: 117, IQR: 45

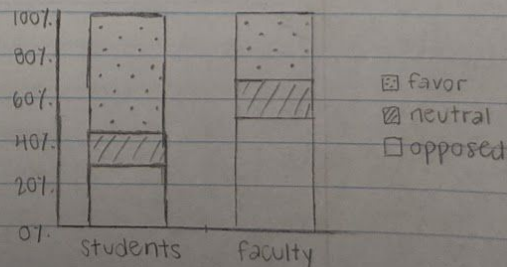


I know it can be confusing, but the MAIN variable of interest here is categorical: people's opinions.

New Grading Policy

- a) students, faculty, categorical
- b) 54%, 57%, 97%
- c) in favor: 56%, neutral: 12%, opposed: 32%
- d) in favor: 57%, neutral: 12%, opposed: 31%
- e) in favor: 32%, neutral: 15%, opposed: 53%

Student and Faculty Opinions on Policy Change



Algebra Section

$$1) z = \frac{x - \bar{x}}{s} \quad 2.5 = \frac{102 - 100}{s} \quad 2.5 = \frac{2}{s} \quad 2.5s = 2 \quad s = 0.8$$

$$2) z = \frac{x - \bar{x}}{s} \quad -3.35 = \frac{60 - \bar{x}}{4} \quad -13.4 = 60 - \bar{x} \quad \bar{x} = 60 + 13.4 = 73.4$$

$$3) 0.05 = 1.96 \sqrt{\frac{0.5^2}{n}} \quad 0.0255 = \sqrt{\frac{0.5^2}{n}} \quad 0.00065 = \frac{0.5^2}{n}$$
$$0.00065n = 0.5^2 \quad n = \frac{0.5^2}{0.00065} \quad n = 384.16$$

$$4) -1.64 = \frac{60 - \mu}{\sigma} \quad 1.96 = \frac{95 - \mu}{\sigma}$$

$$-1.64\sigma = 60 - \mu \quad 1.96\sigma = 95 - \mu$$

$$\mu = 60 + 1.64\sigma \quad 1.96\sigma = 95 - (60 + 1.64\sigma)$$

$$\mu = 60 + 1.64(9.72) \quad 1.96\sigma + 1.64\sigma = 35 \rightarrow 3.6\sigma = 35$$

$$\mu = 75.94 \quad \sigma = 9.72$$