

Unit 1 - Different distributions and how to describe them

Populations and Samples

- Parameters are facts about population. σ and M_x
- Statistics are drawn from samples; do tests: \bar{X} and s_x from samples to infer information about population.

Numerical data

- numbers



Discrete - whole #'s

continuous - #'s with fractions

Organize: dot plots, box + whisker plot, stem + leaf, histograms, cumulative.

Describing Distributions

- shape: symmetric, normal,

skew left  skew right 

- Center: median - more robust

mean, IQR: $Q_3 - Q_1$

- Variability: range, standard deviation

- Unusual features: outliers, clusters, gaps.

- Context: title, x and y axis, answer with it.

Categorical Data

- nouns or proportions
ex) 60% brown hair

Open - unlimited options

closed - limited set of options

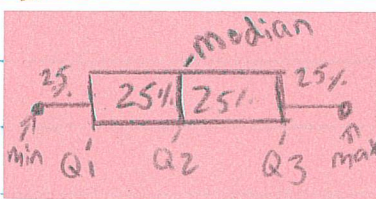
Organize: bar graphs, frequency tables.

Calculator

For mean, S_x , Q_3 , Q_1 , etc:

Stat, edit, fill list, Calc, Vars!

Box Plots



• 25% each section

• Median Center

Outliers:

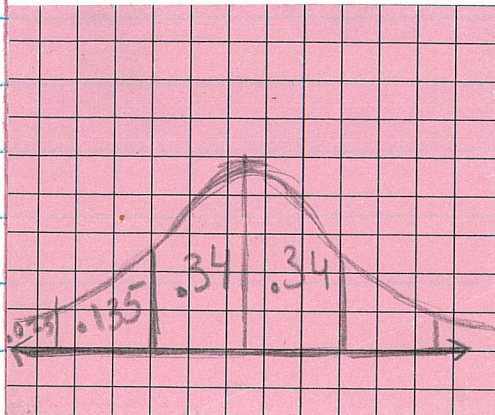
$IQR \cdot 1.5 = \text{fence}$

$Q_3 - Q_1$

$Q_1 - \text{fence}$

and $Q_3 + \text{fence}$

If # is outside that, it's an outlier.



Normal Graphs

- Bell Curve
- Mean in middle
- Symmetric
- unimodal
- Empirical rule: .34, .135, .025

Z-Scores and Percentiles

Z-Score: How many standard deviations a value is away from the mean.

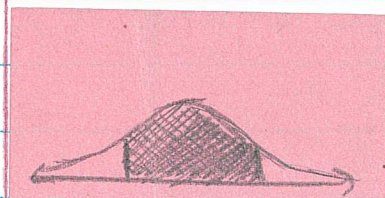
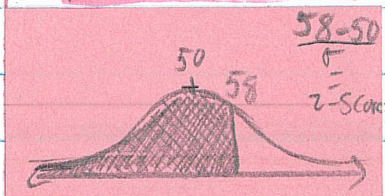
$$z = \frac{\text{obs} - \text{exp}}{\sigma}$$

• Use to find the percent of the data below or above the value. (or in between values)

Convert to % to find how much data is below/above it.

Calc: 2nd vars, normcdf, enter -10000 as lower if trying to find everything below point, 10000 as upper if trying to find all data above.

* Draw Pictures



or... When finding % data between two values do the percent for both through z-scores, then subtract to get what you need.

or... Do 1-combined value percents to find everything less than one value but greater than other value.