Intro to Statistics

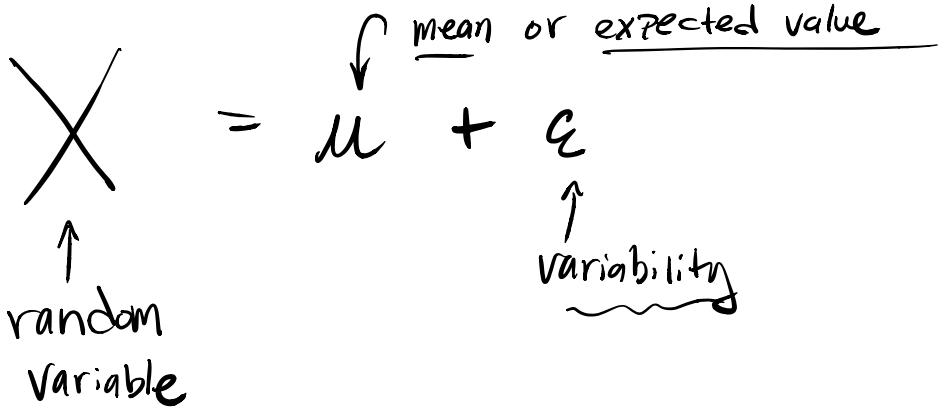
Statistics: collection,
presentation,
analysis, and
use of data

to make decisions,
solve problems, and
design products and processes.

= the science of data

Variability: successive observations of a system do not produce the same results!

random variable: exhibits variability



census : all of a population observed

sample : selected from population

statistical inference : reasoning that sample data applies to population

Types of Studies

a.) retrospective study : based on historical data, no control over the way data was collected

b.) observational study : observe process, record quantities of interest, disturbing as little as possible

*c.) designed experiment : engineer makes deliberate changes to controllable variables; makes inferences about relationships between observed changes and controllable variables; only way to determine cause and effect!

Models

ex: Ohm's Law for ideal resistor:

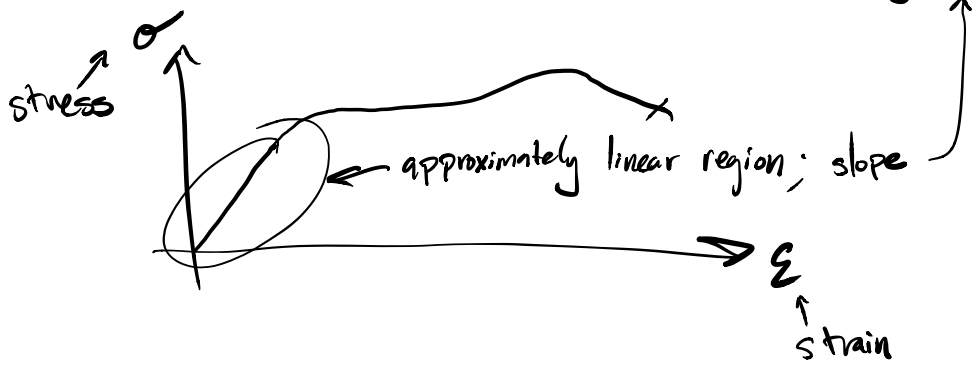
$$I = \frac{E}{R}$$ ← electromotive force ≡ voltage

.. this is a mechanistic model; built from knowledge of physical mechanism of resistance (electric field theory, etc..)

.. actual measurements would exhibit variability!

by contrast: axial stress vs. strain

.. for linearly elastic materials: Young's Modulus



.. this curve comes from data from deformed samples!

∴ empirical model: fits measured data to function

probability: helps to quantify risks in statistical inference

Numerical Data

Sample mean : simple arithmetic average of n observations

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

geometric mean : "multiplicative average"

ex: $\sqrt{x_1 x_2}$

Sample variance : quantifies variability or "scatter" in data

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

← sum of squares of errors

← divide by n if population variance!

Computational formula : sum of squares ↙ , squared sum ↙

$$s^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

Sample standard deviation : $s = +\sqrt{s^2}$

Sample range : $r = \max(x_i) - \min(x_i)$

population mean : μ

population variance : σ^2 } more later!

homework exercise :

ten voltage measurements, all in V_{RMS}

{ 118 121 122 120 115 116 116 119 121 117 }

Compute sample mean, variance, std. dev., range, and population mean and variance

w/ units

answers: $\bar{x} = 118.5\ V_{RMS}$

$s^2 = 6.056\ (V_{RMS})^2$

$s = 2.461\ V_{RMS}$

$r = 7\ V_{RMS}$

Excel

STDDEV.P ← population standard deviation; divides by n

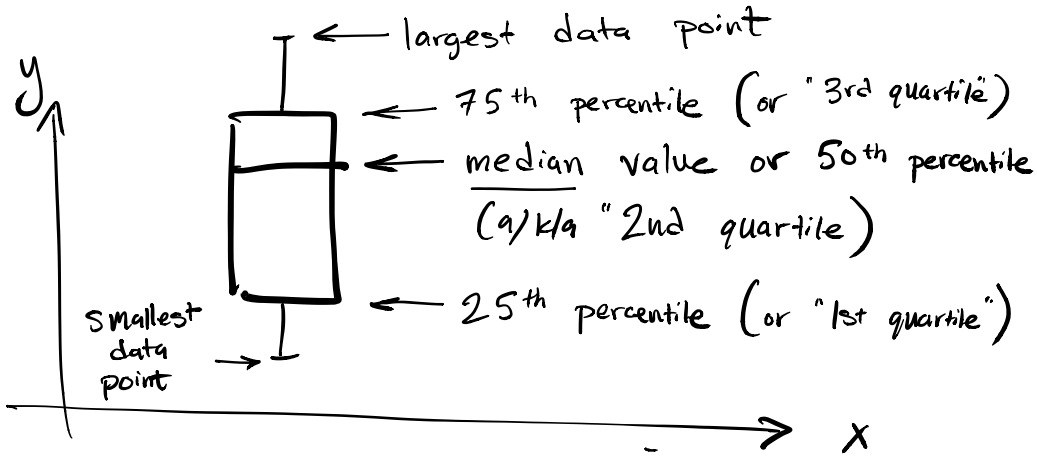
vs. STDDEV.S ← sample; divides by n-1

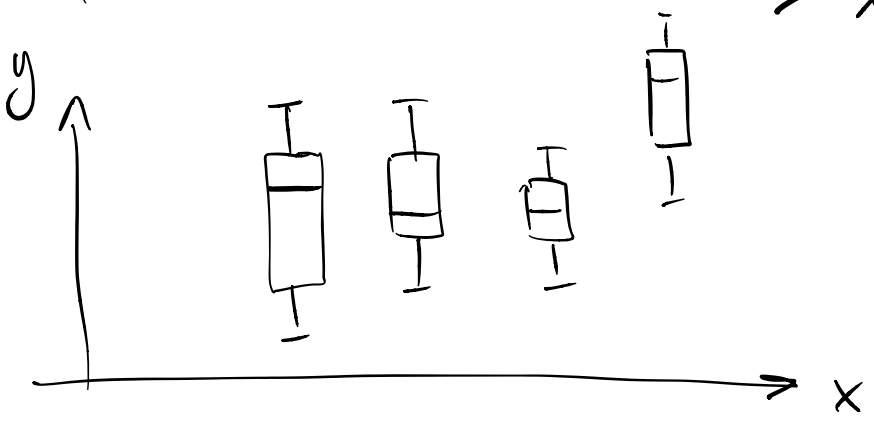
Graphical Displays of Data

scatter diagrams:

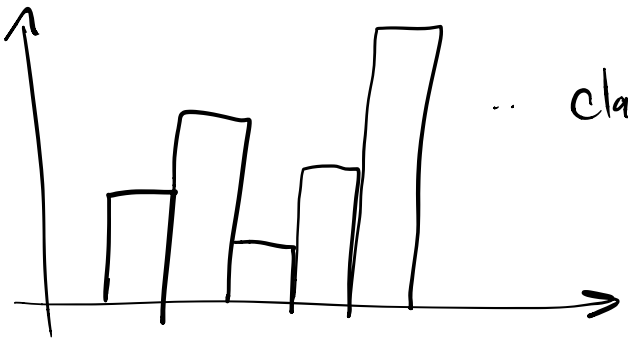
- great for multivariate data (x vs. y, etc..)
- immediately notice trends such as linearity, oscillatory behavior, etc..

Box & Whisker Plots





Histograms a/k/a Frequency Distributions



.. classic bar graph

∴ data is divided into class intervals

a/k/a bins or cells of equal width

rule of thumb: # of bins = $\sqrt{\#\text{ of observations}}$

∴ often bin boundaries are constructed from range of data

ex: ten voltage measurements example

$\sqrt{n} = \sqrt{10} = 3.16 \rightarrow$ use three bins

$x_{i(min)} = 115\ V_{RMS}$

$x_{i(max)} = 122\ V_{RMS}$

∴ $r = 122 - 115 = 7\ V_{rms}$

∴ each bin would have width $\frac{7}{3} = 2\frac{1}{3}\ V_{RMS}$

bin boundaries: $115 + 2\frac{1}{3} = 117\frac{1}{3}$

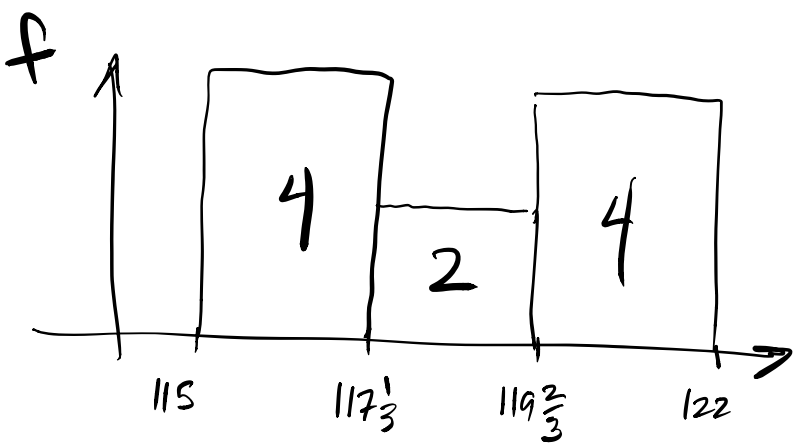
$117\frac{1}{3} + 2\frac{1}{3} = 119\frac{2}{3}$

$119\frac{2}{3} + 2\frac{1}{3} = 122$

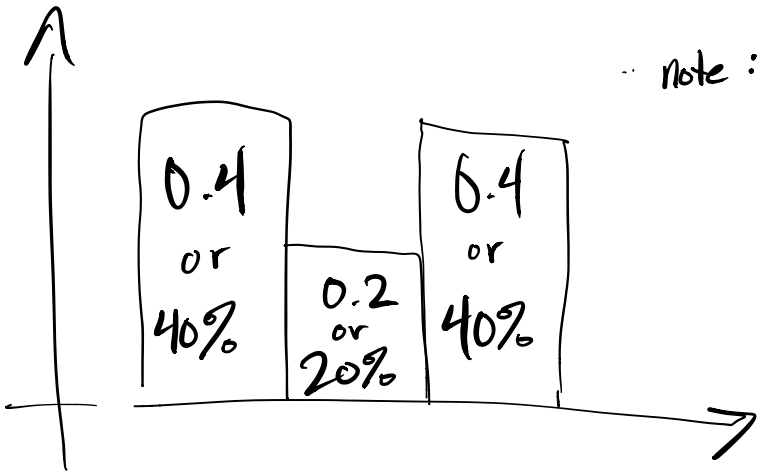
Now we need frequencies associated with each bin

Ordered voltages:

115
116
116
117
$117\frac{1}{3}$
118
119
$119\frac{2}{3}$
120
121
121
122



- relative frequency distribution: divide each frequency by n (n=10)



- note: $\sum = 1$ or 100%

Excel: bin values "up to and including"