Hypothesis Testing

Basis of Idea: X Chip brand claims there are 500 grams of chips in the bag?

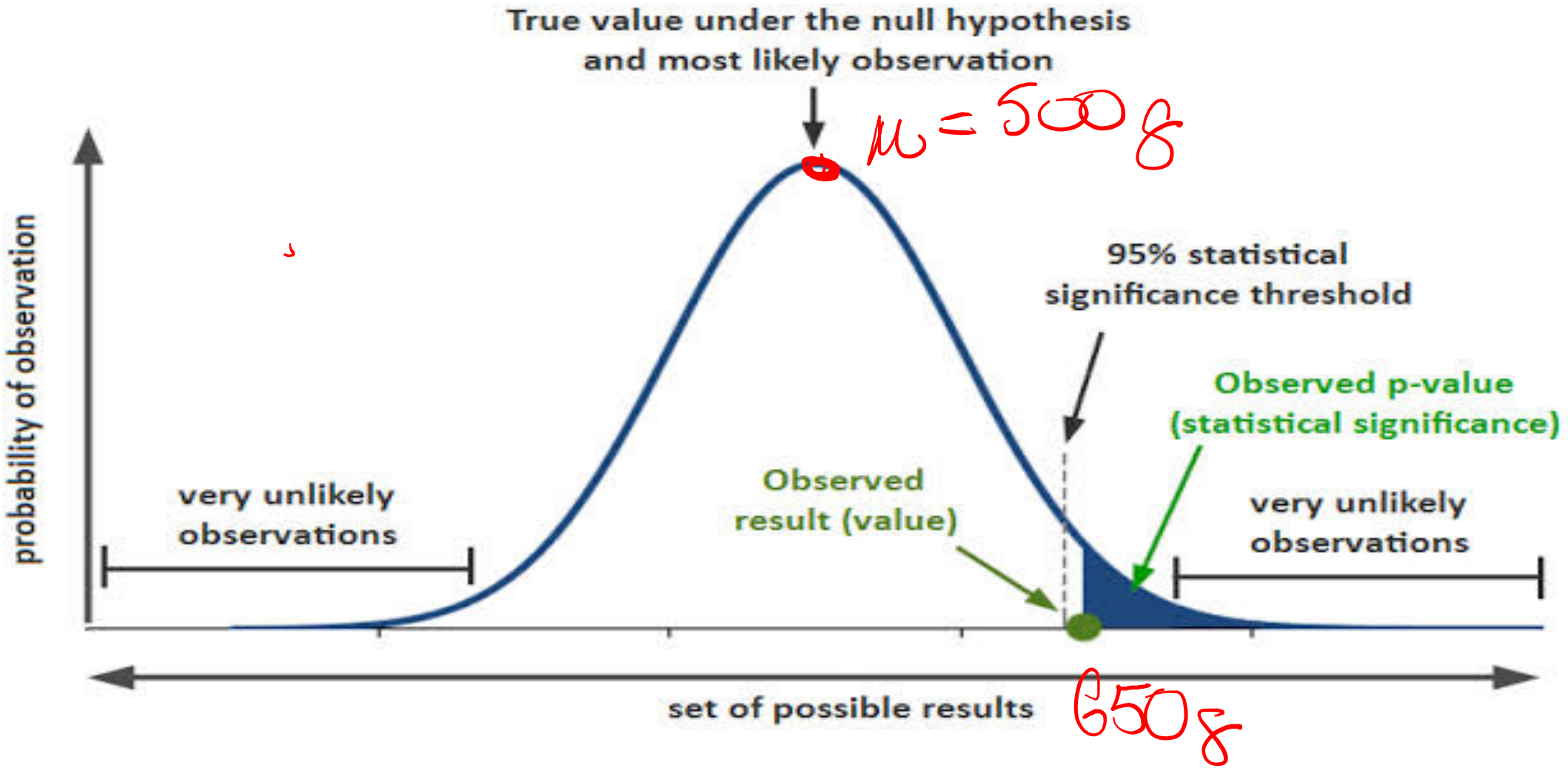
- Null Hypothesis:

500 grams in bag

- Alternative Hypothesis

If I buy bag of chips, there will be less than 500 g.

Probability & Statistical Significance Explained

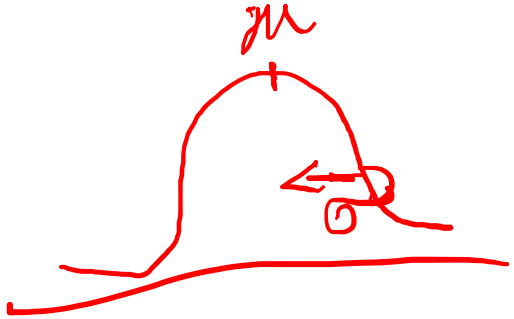


Standard Deviation show the variance in the data set

- Standard deviation describes the variability or scatter within a distribution relative to its mean

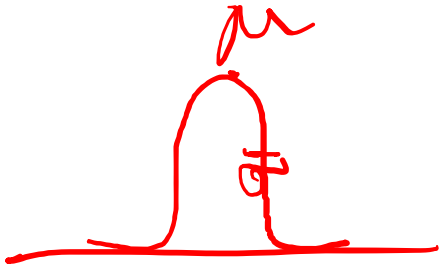
- What does a high standard deviation describe?

Show high variance/variability



- What does a low standard deviation describe?

Show low variability



Calculating Standard Deviation in Excel

VAR= STDEV(array1)

Sample formula =STDEV(A3:A12)

Calculating Standard Deviation in Python and Matlab

Var = stats.stdev(data)

- data = sequence, list, etc.

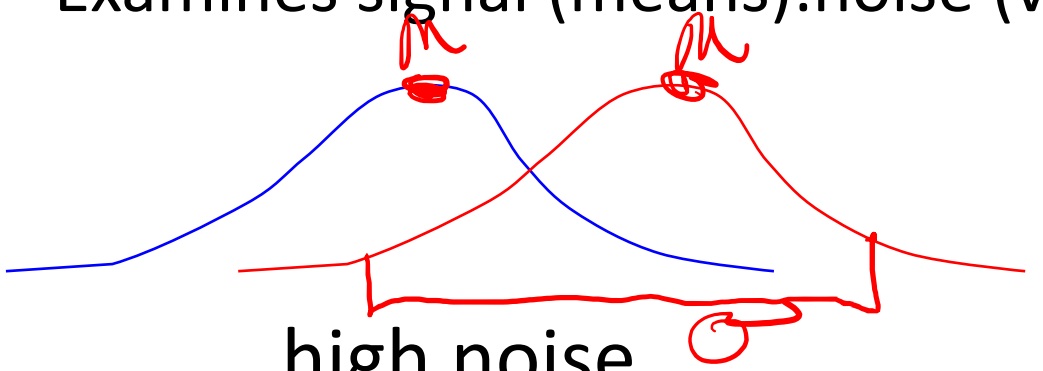
Var = std(A)

- A = array

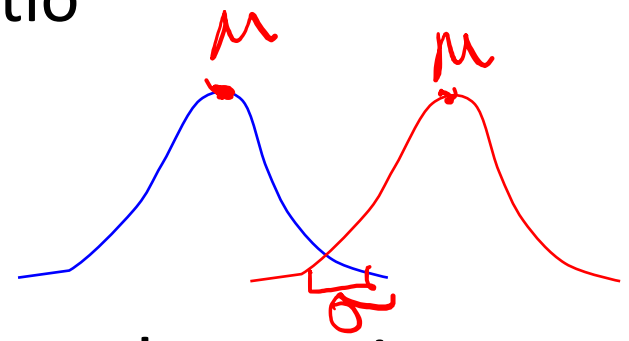
T Test: Determine if populations are significantly different by comparing the mean of two groups

- Assumption:
  - Smooth & symmetric distribution (continuous variable)
  - Data results in a normal distribution
  - Two populations being compared have similar variance

- At $p < 0.05$, there is less than a 5% chance that populations are the same (95% chance that populations are different)
- Examines signal (means):noise (variance) ratio



high noise



low noise

Calculating Student's *t* in Excel

p = T.TEST (array1, array2,2,3)

two-tailed

unequal variance

Sample formula =T.TEST(A2:A10, B2:B10, 2, 3)

Can only compare two data sets at a time

*Make sure it is clear on your plots/writing which conditions are being compared

T-Test in Python & Matlab

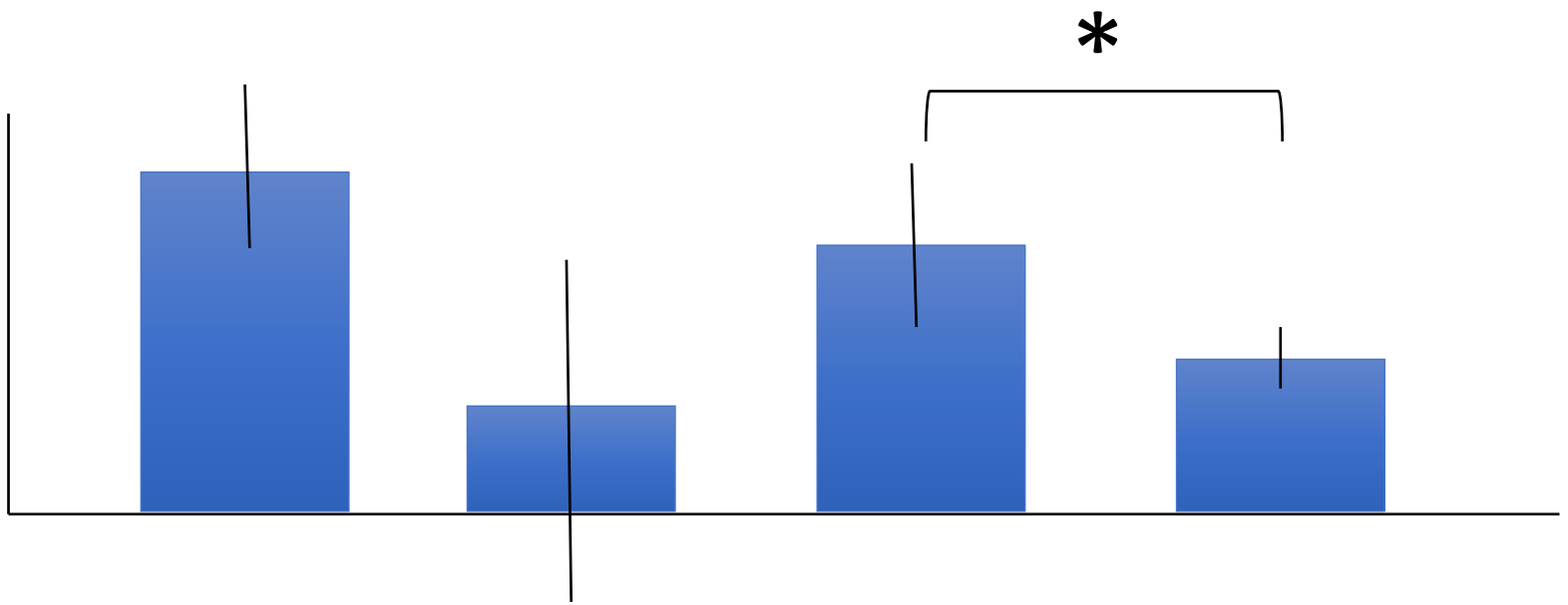
(stat, pvalue) = scipy.stats.ttest_ind(a, b, equal_var)

- a, b = separate lists containing each dataset
- equal_var
  - True assumes equal population variances
  - False assumes unequal → Welch's T-Test

[h, pvalue, ~, stats] = ttest2(data1, data2, 'Vartype', X 'Alpha', A)

- 'Vartype' = 'equal' or 'unequal' in place of **X**
- 'Alpha' = significance level, # in place of **A**

How will you use statistics in your data analysis?



What if the data are not statistically significant?

p = 0.055

THE BLOG

Over 100 Million Now Receiving Federal Welfare

2:40 PM, AUG 8, 2012 - BY DANIEL HALPER

SINGLE PAGE PRINT LARGER TEXT SMALLER TEXT ALERTS

A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."