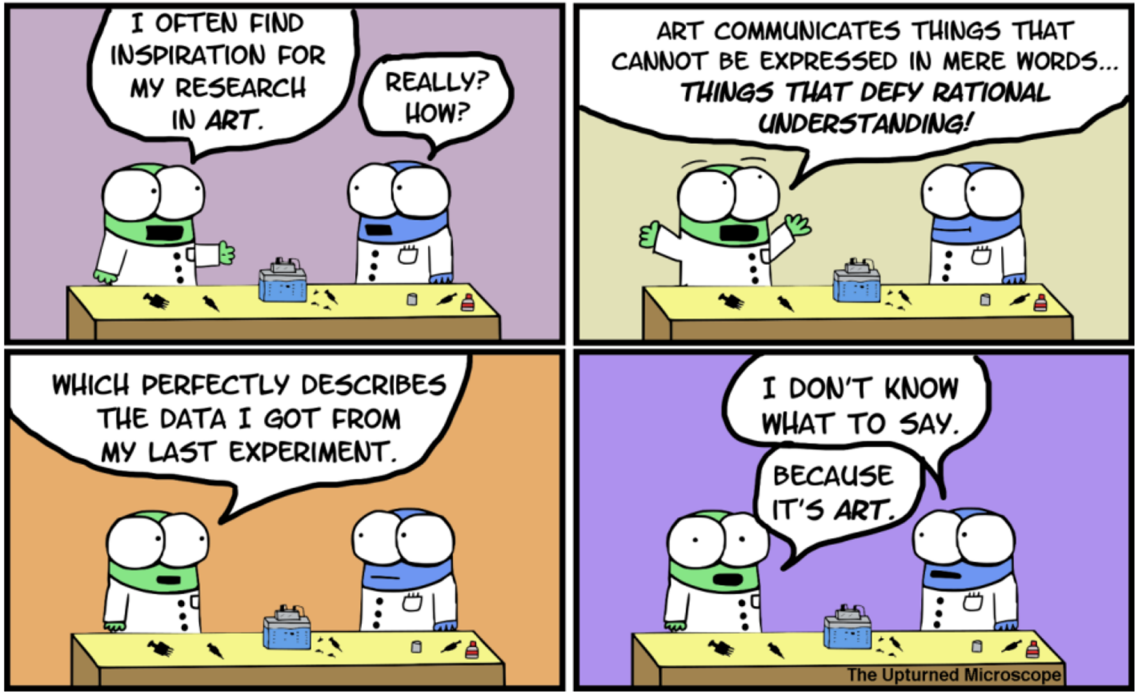
M1D7: Complete data analysis using statistical methods

1. Quiz!
2. Prelab
    a. Statistics
    b. Mod 1 Review
3. Complete stats analysis on data
4. Work on Data Summary



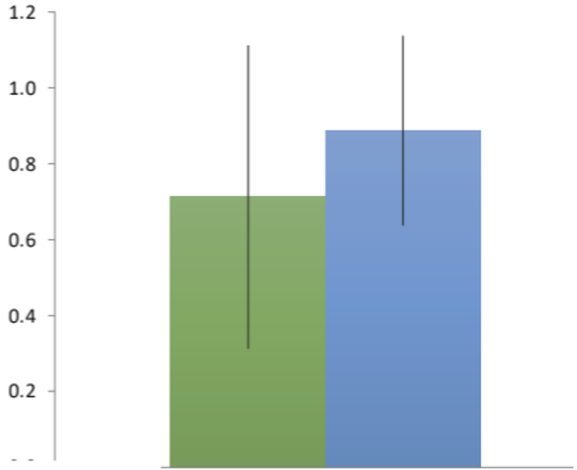
Mod 1 Due Dates

- Data Summary (15%)
  - Completed **in teams** and submitted via **Stellar**
  - Draft due 10/13 (W) at 10pm
  - Revision is due 10/23 (Sat.) at 10pm
- Research Talk (5%)
  - Completed **individually** and submitted by 10pm on 10/16 (Sat.) via **Gmail**: bioeng20.109@gmail.com
- Notebook (collectively 5%) [Rubric on Wiki]
  - Submit a pdf of **M1D6** entry by 10pm on the day after M1D7 (W/R) via **Stellar**

Confidence Intervals show the variance in the data

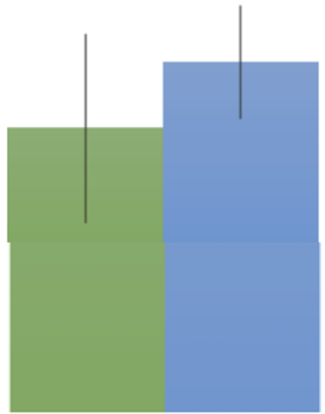
- With a 95% confidence interval (alpha = 0.05), there's a 95% chance the true population mean is within the defined range

$$\bar{x} \pm t\frac{s}{\sqrt{n}}$$



Error bars depict:

95% CI



Standard Deviation

With small sample sizes, 95% CI can be more reflective of sample variance

Confidence Intervals in Excel

=CONFIDENCE.T(alpha, stdev, size)

alpha | significance level

stdev | standard deviation

Can be calculated w/in Excel using =STDEV(data)

size | sample size; (n)

Sample formula =CONFIDENCE(A2, STDEV(A3:A12), COUNT(A3:A12))

- Enter the calculated confidence interval as your “custom” error bar in Excel

Confidence Intervals in Python & Matlab

(low, high) = scipy.stats.t.interval(alpha, df, loc, scale)

alpha | confidence level i.e. 95% → 0.95

df | degrees of freedom; (n-1)

loc | sample mean

scale | sample standard deviation

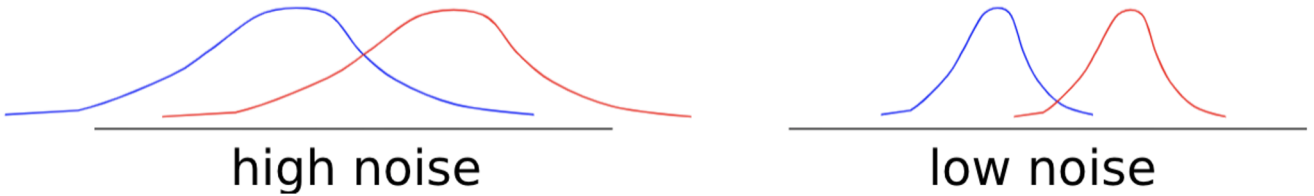
[~, ~, ci, ~] = ttest2(data1, data2, 'Vartype', X, 'Alpha', A)

'Vartype' | 'equal' or 'unequal' # in place of X

'Alpha' | significance level # in place of A

Student's T-Test evaluate how **significant** the differences between two groups

- Assumes a t-distribution
  - Smooth & symmetric distribution (continuous variable)
  - Data results in a normal distribution
  - Two populations being compared have similar variance
- Null hypothesis: there is no relationship between the two samples
  - $p < 0.05$ | there's less than a 5% chance the populations are the same
- Examines ratio of signal (means) : noise (variance)



Student T-Test in Excel

=T.TEST(data1, data2, tails, type)

tails | specifies number of distribution tails

type | (1) paired

(2) two-sample, equal variance

(3) two-sample, unequal variance

Sample formula =T.TEST(A2:A10, B2:B10, 2, 1)

Student 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

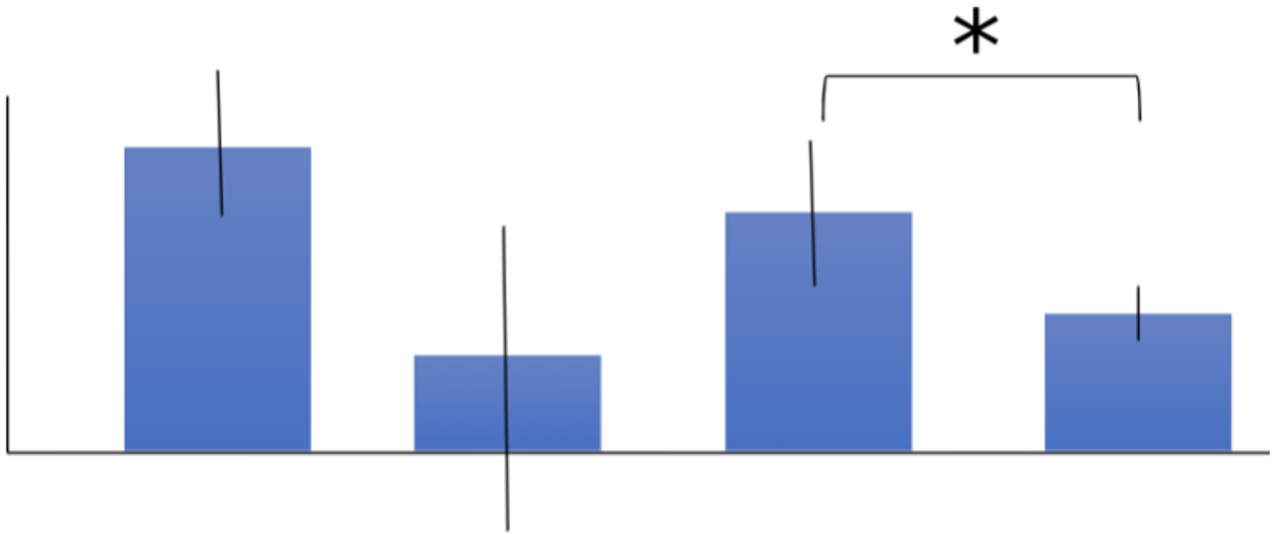
False assumes unequal → Welch's T-Test

[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

Using Statistics in your data analysis



What if the data are not statistically significant?

p = 0.1 **p = 0.55**

Grading rubric for Research Talk

Category	Elements of a strong presentation	Weight
Introduction	• Introduce yourself and the research • Summarize the background information necessary to understand the research • Provide a clear and concise description of the central question / hypothesis	25%
Methods & Data	• Provide ONLY the method information necessary to understand the results • Give complete and concise explanations of the results • Relate the results to the central question	25%
Summary & Conclusions	• Highlight the key finding(s) relevant to the central question / hypothesis	25%
Organization	• Give a logical, easy-to-follow narrative • Include transition statements	15%
Delivery	• Show confidence / enthusiasm and speak clearly • Use appropriate language (technical or informal, as appropriate) • Be mindful of the time limit (3 minutes +/- 15 seconds!)	10%

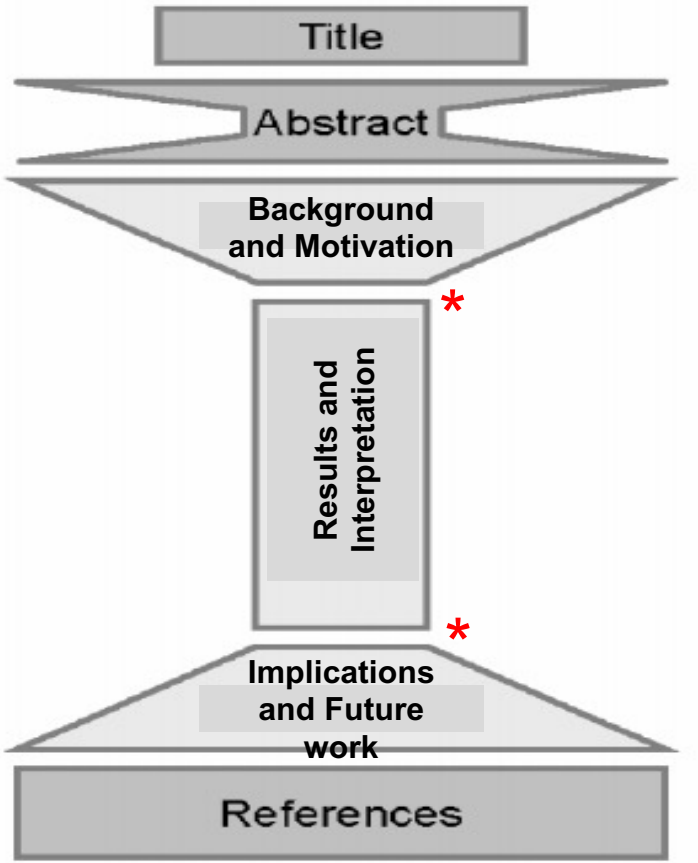
The mini-presentation will be graded by Dr. Noreen Lyell with input from Dr. Leslie McClain, and Dr. Becky Meyer.

Additional guidance for the Data Summary

- Noreen and Becky will hold extra office hours in preparation for this assignment
- Groups can also request meetings to go over questions that come up when working
  - Email both Noreen and Becky and we will set up a meeting with one of us

M1 Data Summary

Format: Portrait 8.5x11” .ppt slides
See wiki for more details



Title: take-away message
Abstract: the only section ***not*** in bullet points
ALL bullet points:
-background and motivation (include references)

-Results and interpretation

Implications and future work (include references)

References (*see wiki for format suggestions*)

Background & Motivation

- Impact statement
  - General background
  - Describe previous work in the field
- Specific background (e.g. BER, H_2O_2, Arsenic, CometChip, H2AX)
  - Introduce topics, pathways and specific technologies necessary to understand the experimental approach
  - Include BER pathway figure
  - Narrow focus to the specific question addressed in your study
- Knowledge gap/statement of problem
  - What is unknown, therefore motivating your study
- Hypothesis
  - What do you propose will be the outcome of your study?
- A brief preview of your findings
  - Here we show…
  - End with broad implications of the study

Results & Interpretation

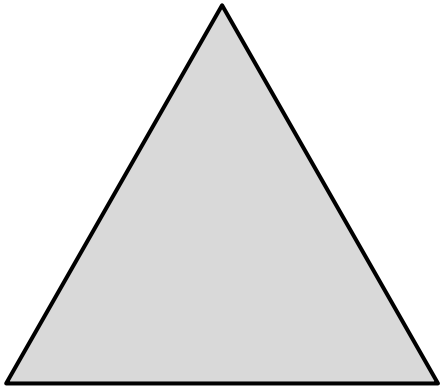
- Figures and captions
  - ***Decide on the figures first***
  - Use figure subpanels (label with letters)
  - Text: limited on figure, explicit in caption
  - reasonable size
  - descriptive title
  - Intro/purpose at beginning of in caption
  - caption descriptive of image, very light on methods

- Results and Interpretation (each page needs subtitle below figure caption)
  - **Goal** / intent / purpose of experiment = intro topic bullet
  - What you did: experiments and expectations, describe controls
  - What you found: quantitatively describe your result, referring to the figure (”Figure 1a shows…”

Notes on Implications & Future works…

- Start with ‘here we showed…’
  - **Restate major results and broad implications**
- Follow same order as in Figures/Results
  - Tie together the conclusions from your data
  - If necessary, describe caveats of experiment and suggest improvements
  - Identify unknowns and speculate (within reason)
  - Don’t make huge generalizations or overreach the results shown
- Propose future experiments, identify new questions that arise
  - Incremental next steps that can be tested / measured
- **Come back to the big picture / impact statement topic introduced in background**

How should you conclude your story?



- What are the main findings / conclusions?
- What are the implications of the results?
- How do the results relate to the research question / hypothesis?
- How do the results advance what is known?

- ***Topic***: What are the main conclusions from key experiments?
- ***Topic***: How do the main conclusions answer the research question?
- ***Topic***: Did your results match your expectations?
  - If no, provide a putative explanation. If yes, how can you further test if your hypothesis is correct?
- ***Topic***: Based on the results, whether they matched your expectations or not, what experiments might you recommend next?
  - Follow-up experiments could distinguish between competing explanations of a given outcome or broaden the sample set for a question you already asked, to give just two examples.
- ***Topic***: What are the limitations of your experimental approach?
- ***Topic***: How might your experimental approach be improved?

Ideas for Future works:

- What are some next steps?

- What are some broader possibilities?