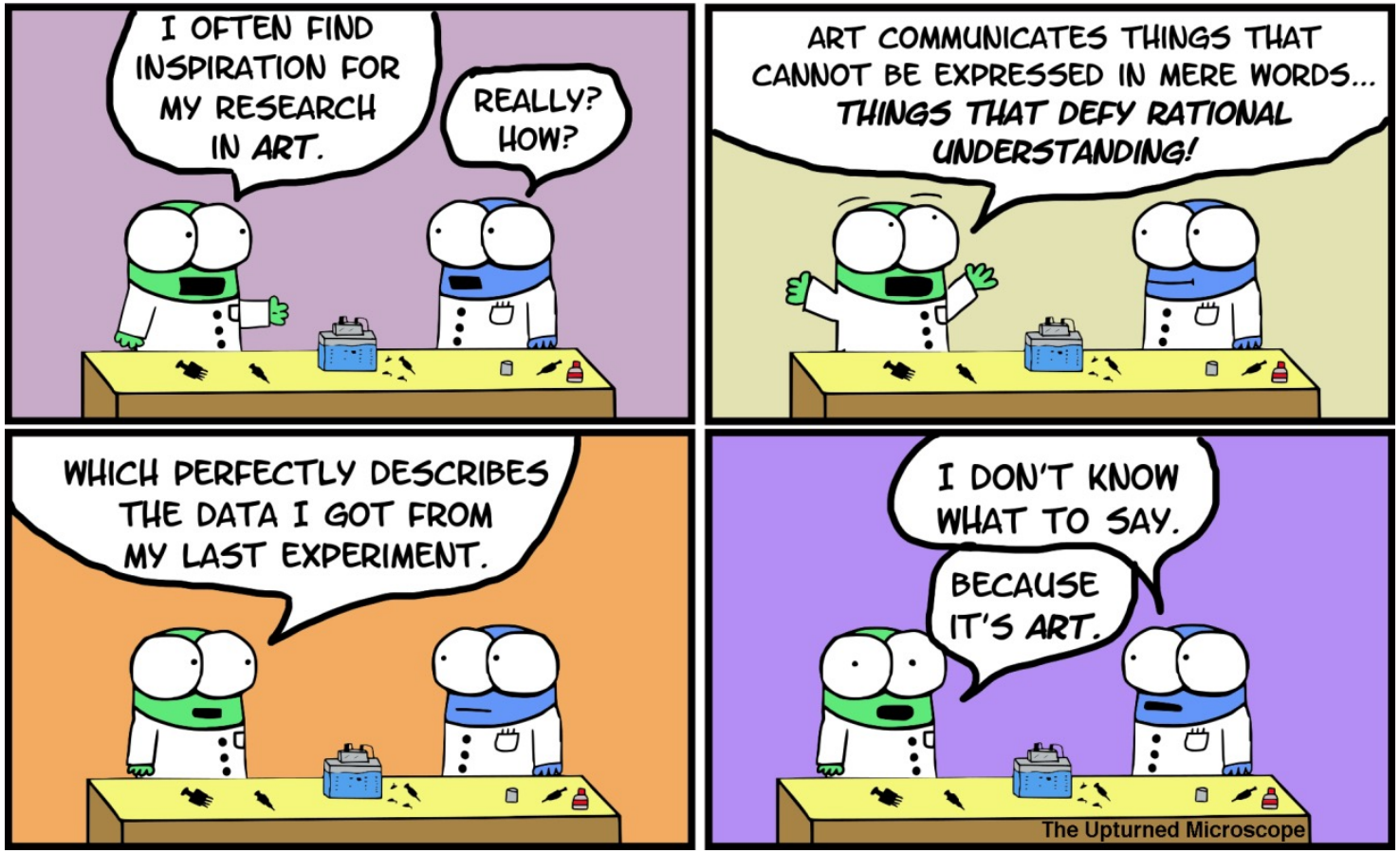
M1D7: Analyze data using statistical methods

1. Quiz
2. Prelab
    1. Statistics
    2. 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, final revision due 10/23

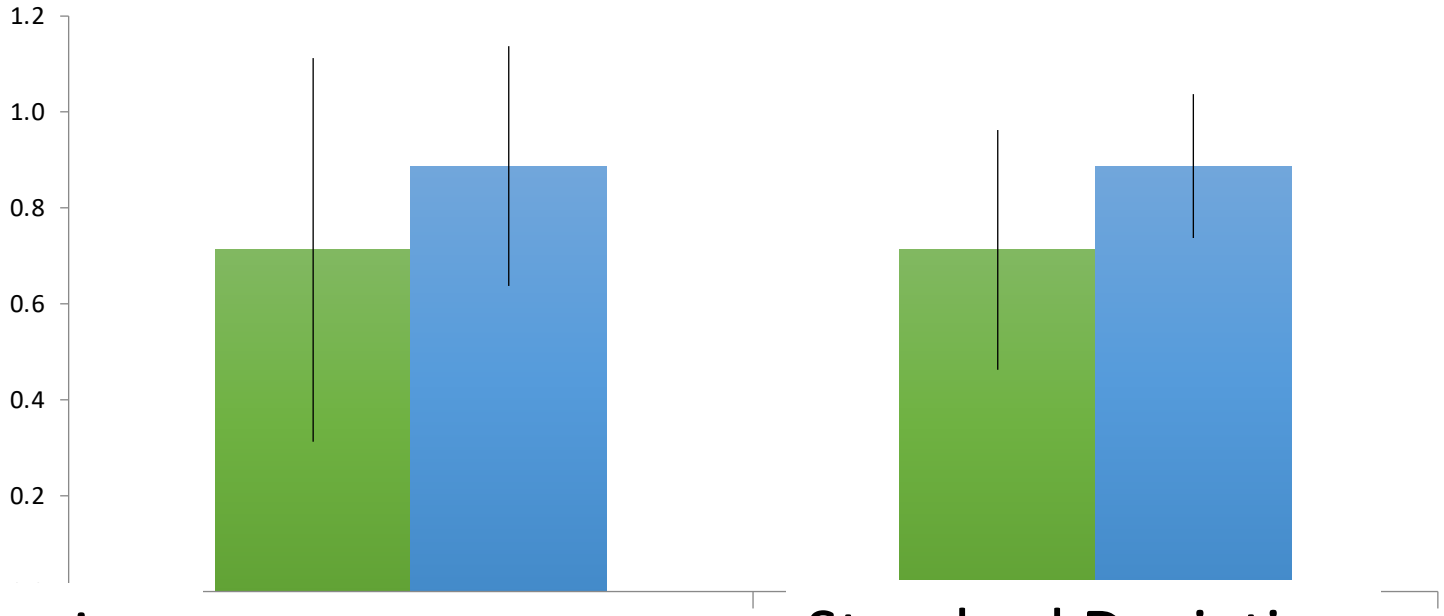
- Research talk (5%)
  - completed individually and submitted **via Gmail**: bioeng20.109@gmail.com
  - due 10/16

- Notebook (collectively 5%) Rubric on Wiki
  - **Submit pdf of M1D6** entry by 10pm Wednesday

- Blog (part of 5% Participation)
  - due 10/18 via Slack

Confidence intervals show the variance in the data set

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

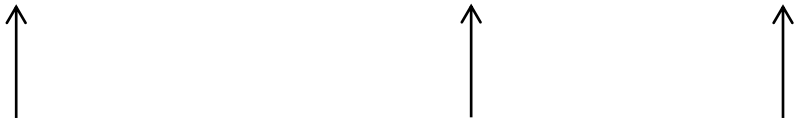


Error bars depict: 95% CI Standard Deviation

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

Calculating Confidence interval in Excel

= CONFIDENCE.T(alpha, standard dev., size)



Can be calculated in Excel using =STDEV(data)

Sample formula =CONFIDENCE(0.05, (STDEV(A3:A12)), 10)

Once you have calculated the confidence interval you will enter this value as your “custom” error bar in excel

CI in Python and Matlab

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

- alpha = significance level
- df = degrees of freedom; (n-1)
- loc = sample mean
- scale = standard deviation of sample

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

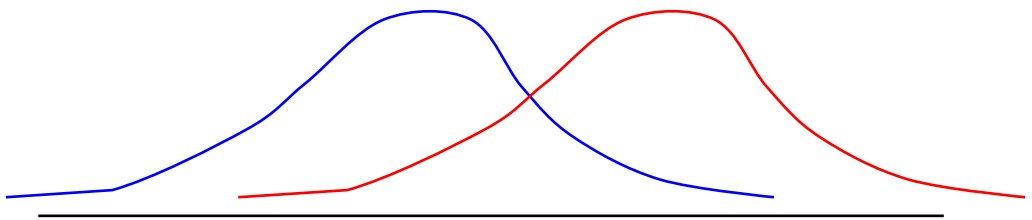
- 'Vartype' = 'equal' or 'unequal' in place of **X**
- 'Alpha' = significance level, # in place of **A**

Student's *t*-test used to determine if populations are significantly different

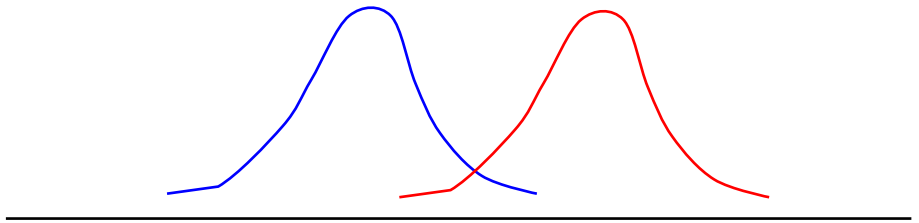
- Assume data follows *t*-distribution
  - Smooth & symmetric distribution (continuous variable)
  - Data results in a normal distribution
  - Two populations being compared have similar variance

- 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

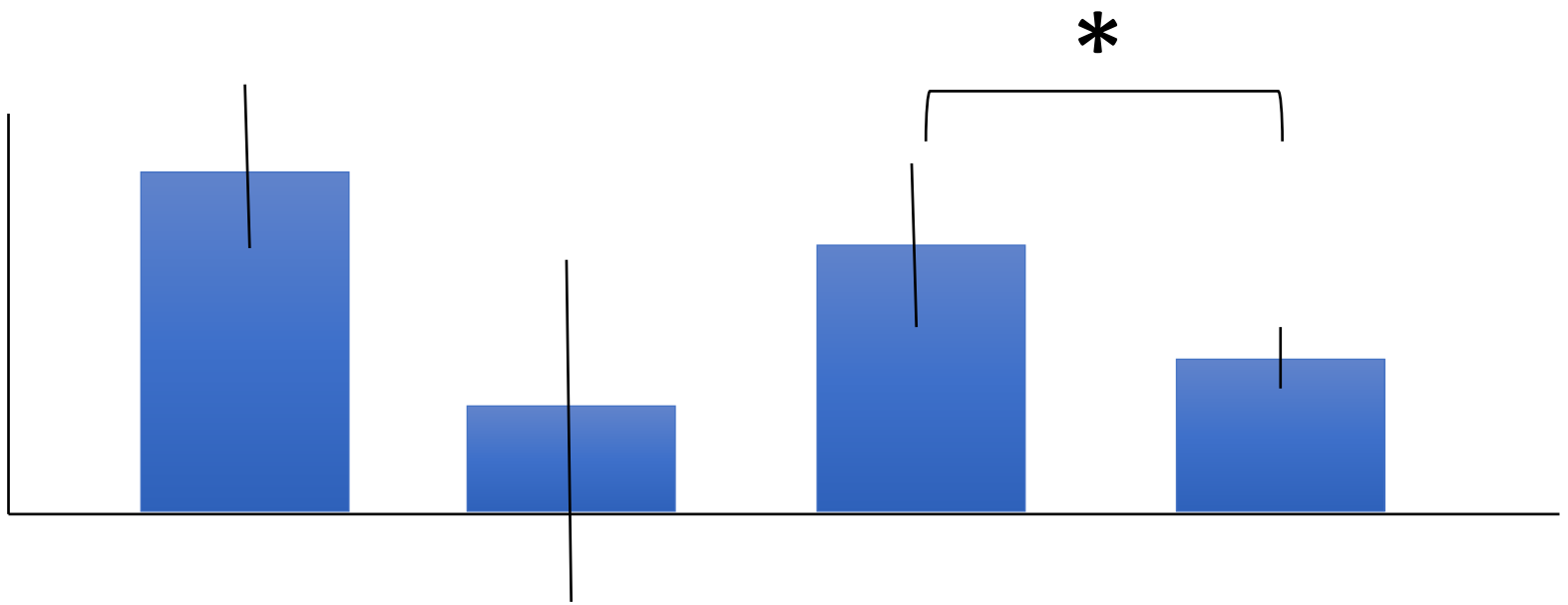
(stat, pvalue) = scipy.stats.ttest_ind(a, b, equal_var)

- 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

For Today

- Complete statistics for **γ**H2AX and CometChip experiments
- Work on Data Summary plans with partner

Discuss specifics of the Data Summary at the end of class

For M2D1

- Read Intro for Mod 2

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 I 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

Review Mod 1 project goals

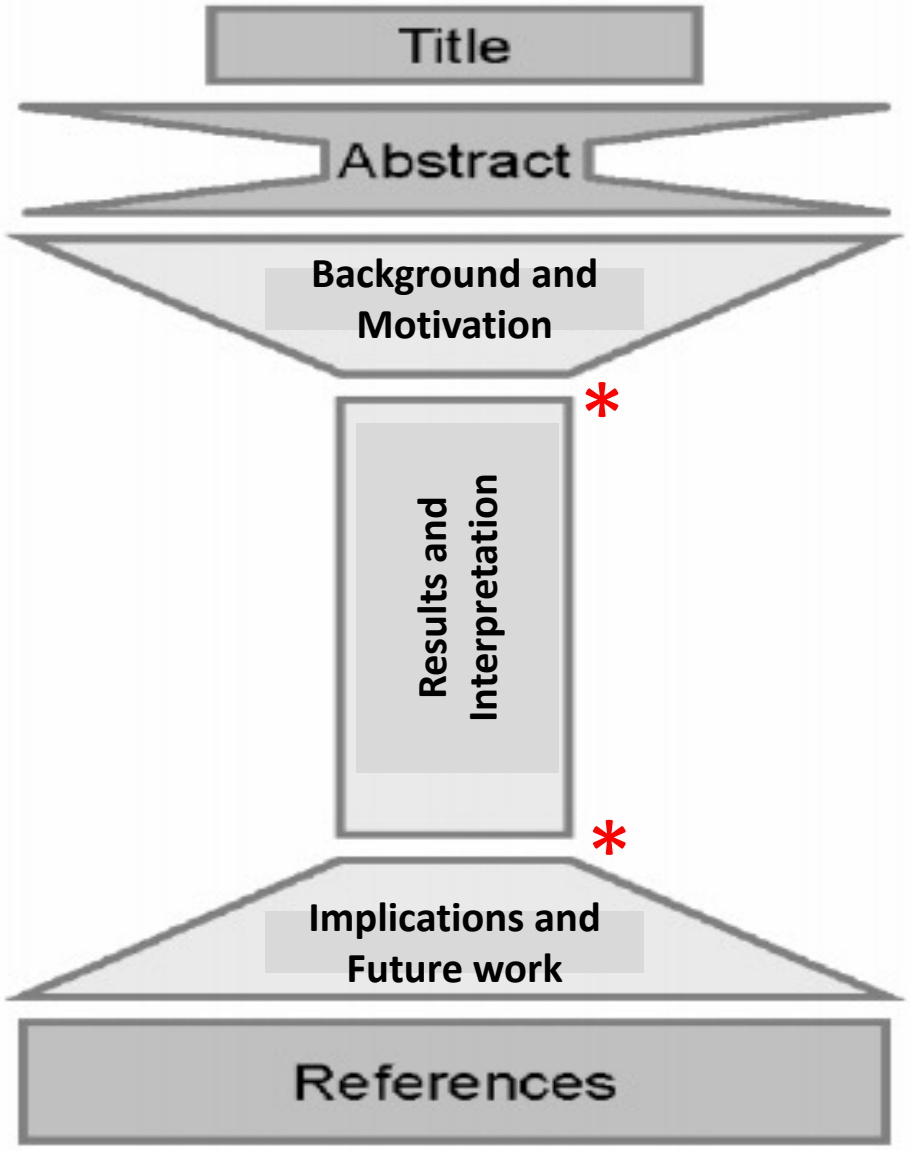
What is our overall goal/question in this project:

What are the conditions we are using to address this:

What experiments are we are using to address this:

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
  - Reference schematic 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

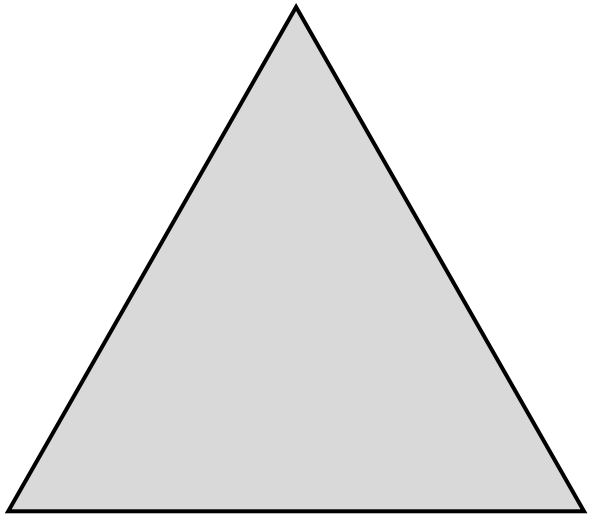
- Figures and captions
  - ***Decide on the figures first***
  - Use figure subpanels (label with letters)
  - Text: limited on figure, explicit in caption
  - reasonable size
  - descriptive title
  - Intro/purpose at beginning of in caption
  - caption descriptive of image, very light on methods

- 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…”
  - What does this indicate: interpret your result, what does it mean?
  - What does this motivate you to do next: **transition** to next experiment

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?