

Lies, Damned Lies, and Statistics

M2D4 20.109 Lecture, F14

We aim to tell the best truth we can.

- Re-visit graphing data
 - Error bars (and the art of “pretty” data)
 - Look at your raw data
- Multiple hypothesis testing
 - Go beyond the two-tailed t-test
- Avoid the lies
 - Plan your analysis *before* you plan your experiments
 - TX sharpshooter fallacy

These x-y pairs have the same mean, variance, and correlation

| I | | II | | III | | IV | |
|------|-------|------|------|------|-------|------|-------|
| x | y | x | y | x | y | x | y |
| 0.0 | 8.04 | 10.0 | 9.14 | 10.0 | 7.46 | 8.0 | 6.58 |
| 8.0 | 6.95 | 8.0 | 8.14 | 8.0 | 6.77 | 8.0 | 5.76 |
| 13.0 | 7.58 | 13.0 | 8.74 | 13.0 | 12.74 | 8.0 | 7.71 |
| 9.0 | 8.81 | 9.0 | 8.77 | 9.0 | 7.11 | 8.0 | 8.84 |
| 11.0 | 8.33 | 11.0 | 9.26 | 11.0 | 7.81 | 8.0 | 8.47 |
| 14.0 | 9.96 | 14.0 | 8.10 | 14.0 | 8.84 | 8.0 | 7.04 |
| 6.0 | 7.24 | 6.0 | 6.13 | 6.0 | 6.08 | 8.0 | 5.25 |
| 4.0 | 4.26 | 4.0 | 3.10 | 4.0 | 5.39 | 19.0 | 12.50 |
| 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15 | 8.0 | 5.56 |
| 7.0 | 4.82 | 7.0 | 7.26 | 7.0 | 6.42 | 8.0 | 7.91 |
| 5.0 | 5.68 | 5.0 | 4.74 | 5.0 | 5.73 | 8.0 | 6.89 |

X mean = 9.00

X stdev = 3.32

Y mean = 7.50

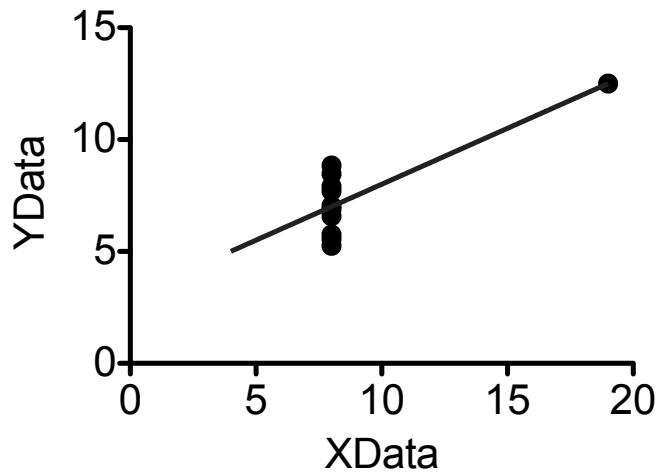
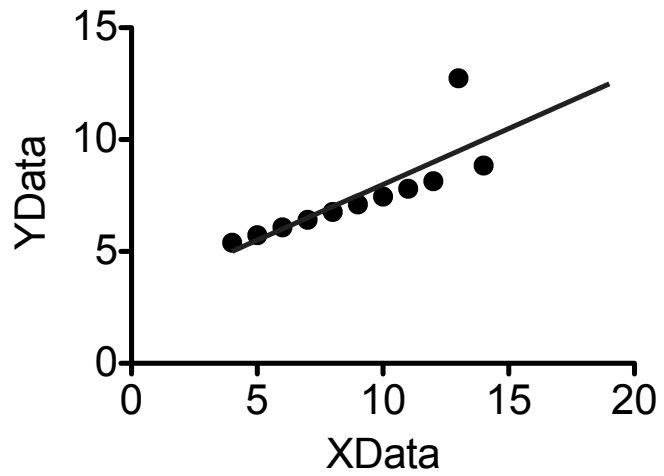
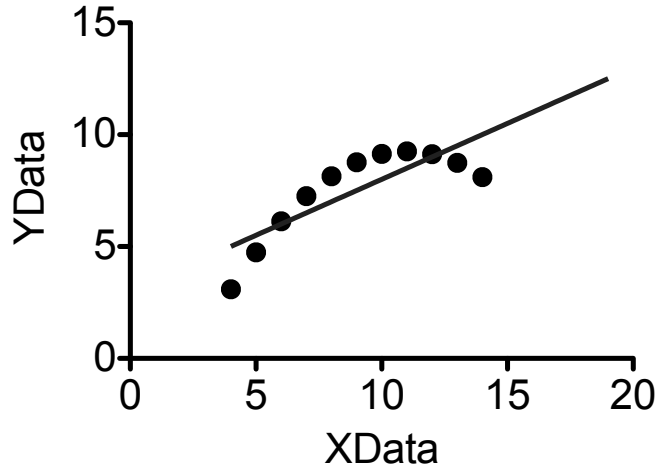
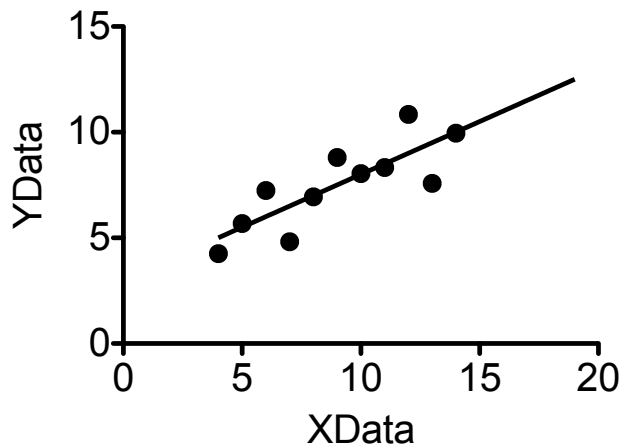
Y stdev = 2.03

R = 0.82

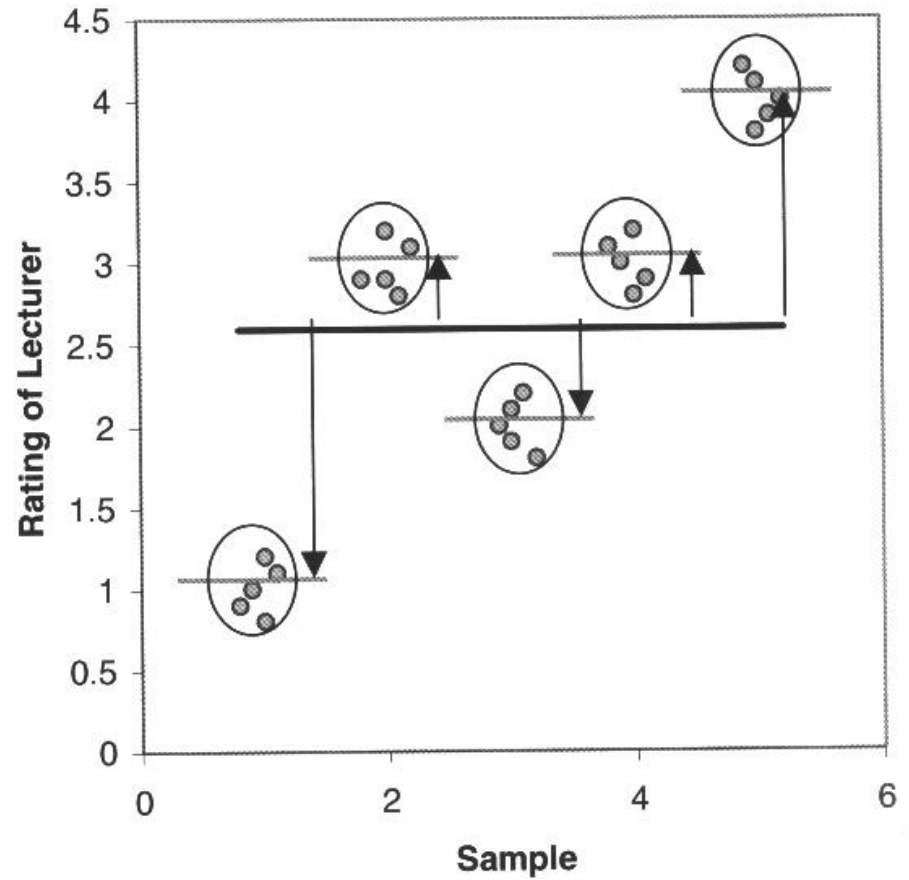
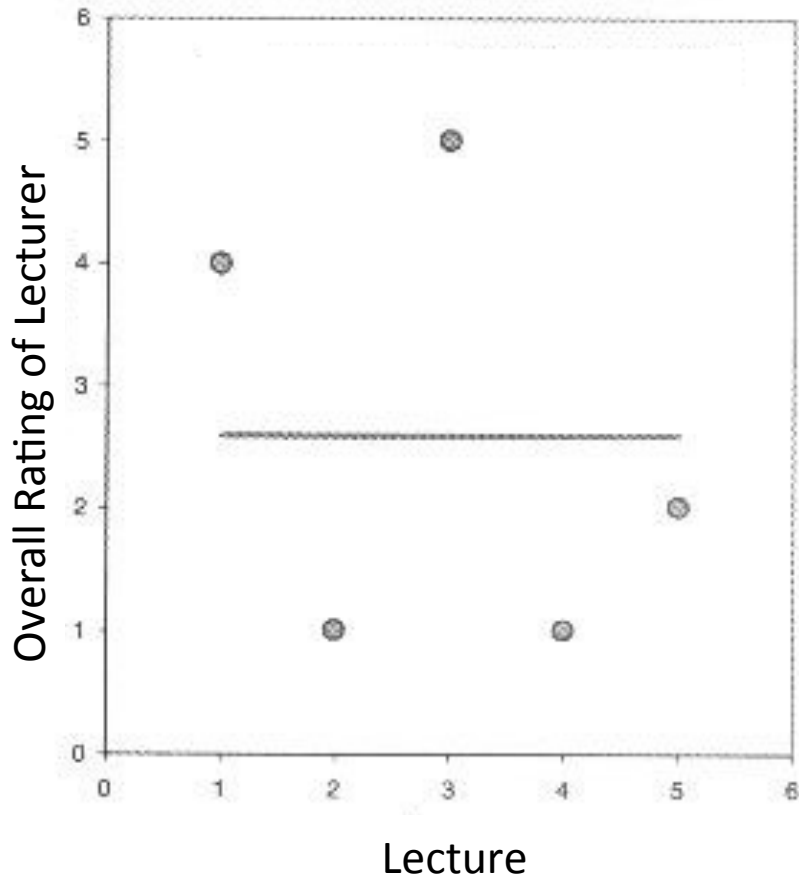
(Pearson correlation)

Anscombe's quartet

These x-y pairs have the same mean, variance, and correlation



There is a difference between SD and SEM!



The SD quantifies the scatter of the data.

The SEM quantifies how far the sample mean is from the true population mean.

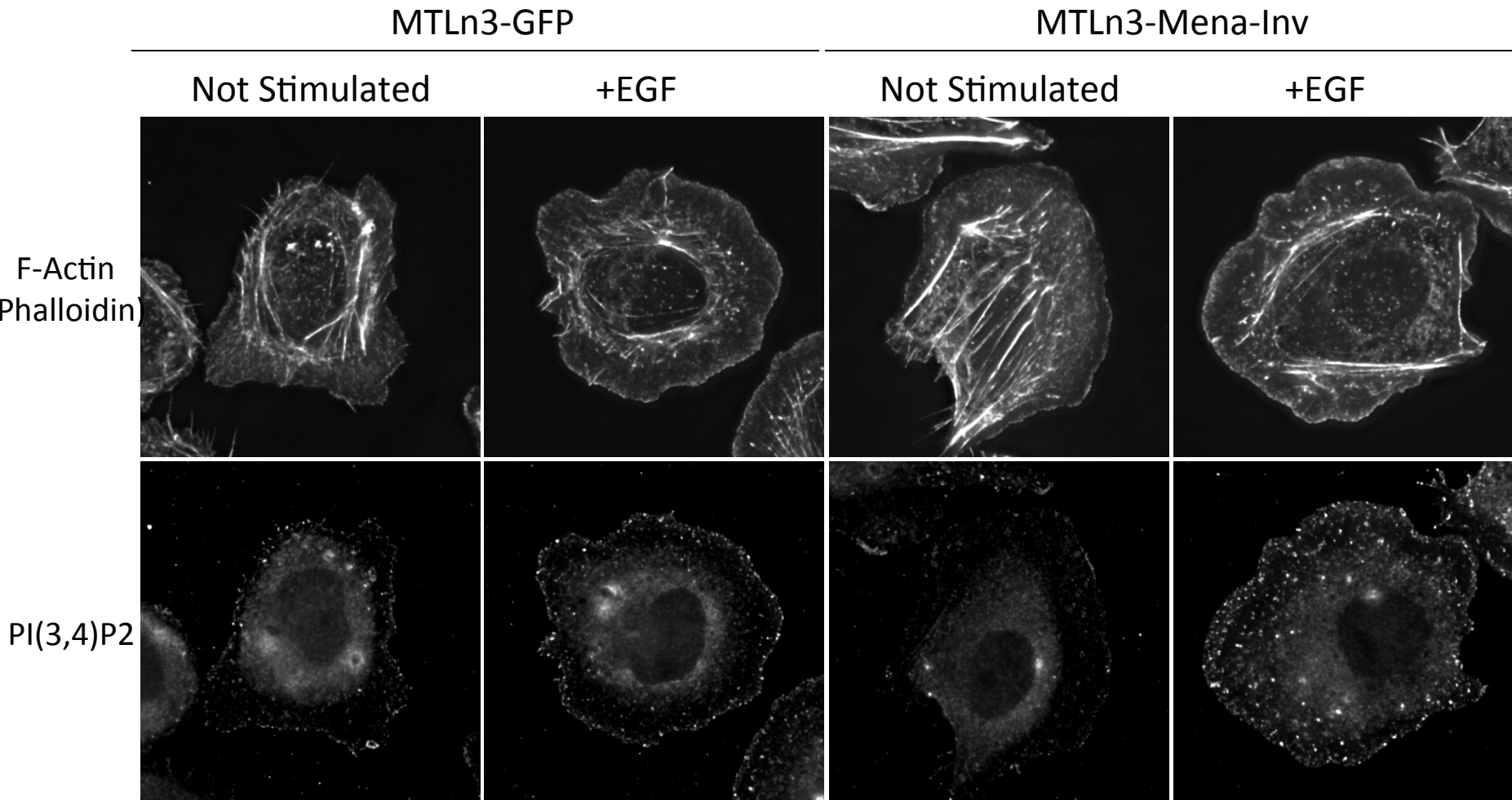
SD or SEM ?

- If the scatter is caused by **biological variability**, it is important to show the variation.
 - **Report the SD rather than the SEM.**
 - Better, show a graph of all data points, or perhaps report the largest and smallest value there is no reason to only report the mean and SD.
- If you are using an in vitro system with no biological variability, the scatter can only result from **experimental imprecision** (no biological meaning).
 - **Report the SEM** since the SD is less useful here.
 - Instead, report the SEM to give your readers a sense of how well you have determined the mean.

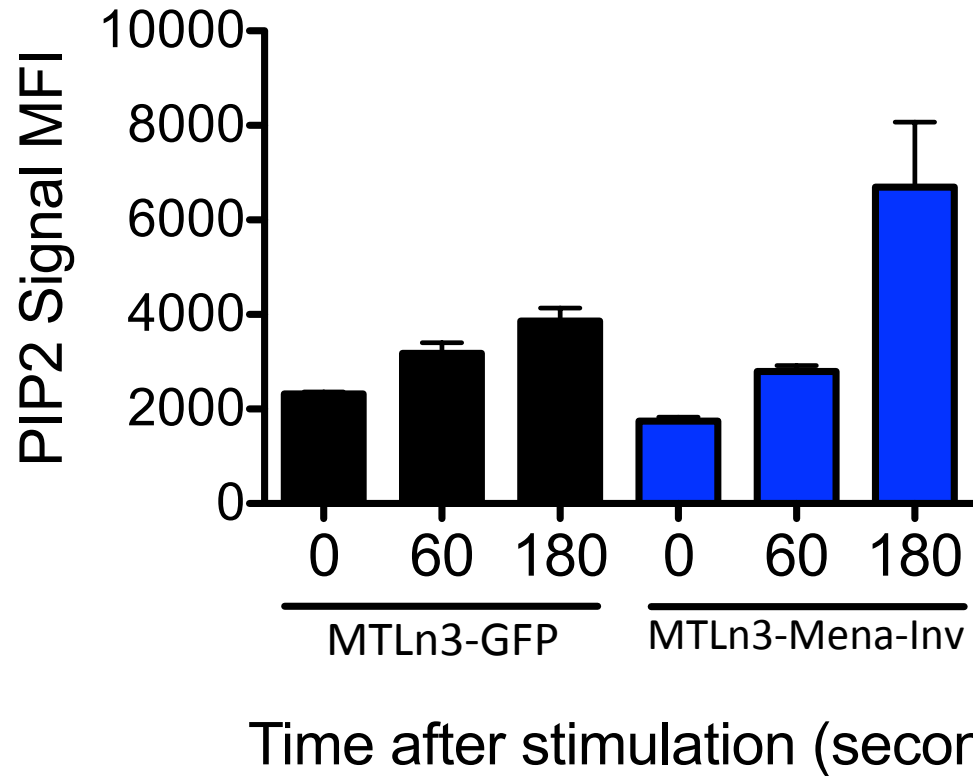
That 'rule of thumb' isn't so handy...

| Error Bar Type | Overlapping | Not overlapping |
|-------------------------|--|---|
| Standard Error of Mean | Not significantly different ($p < 0.05$) | No conclusion |
| Standard Deviation | No conclusion | No conclusion |
| 95% Confidence Interval | No conclusion | $p < 0.05$ *If no multiple comparisons |

It is *really* important to look at your data.



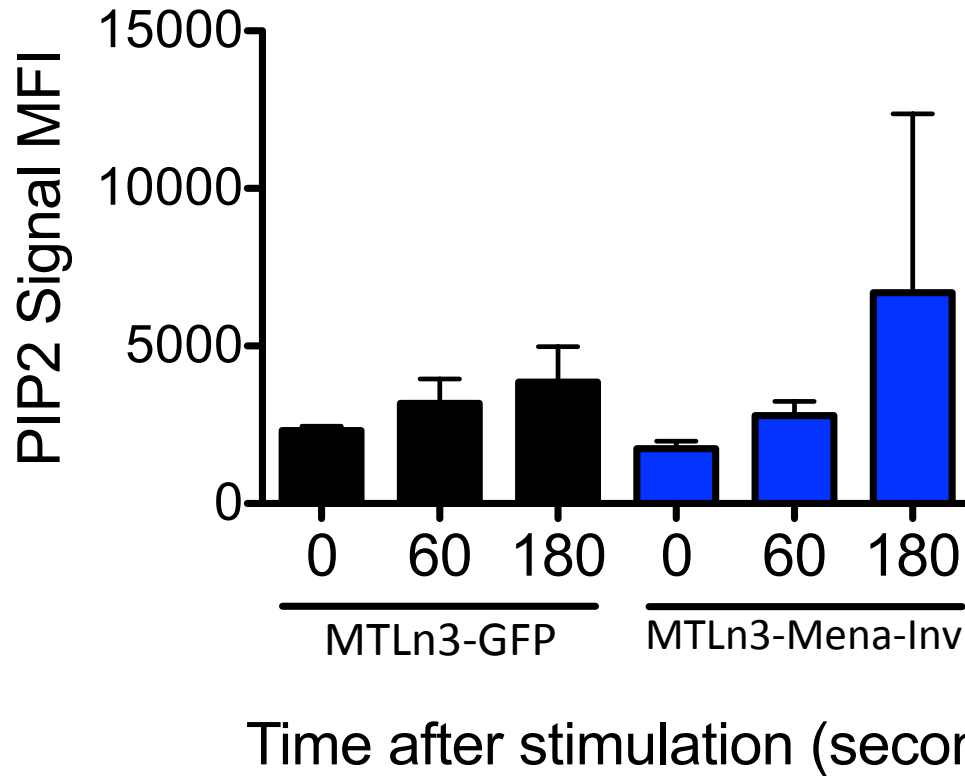
It is *really* important to look at your data.



m = 4 experiments
n = 5 cells / experiment

Data shown as mean +/- SEM

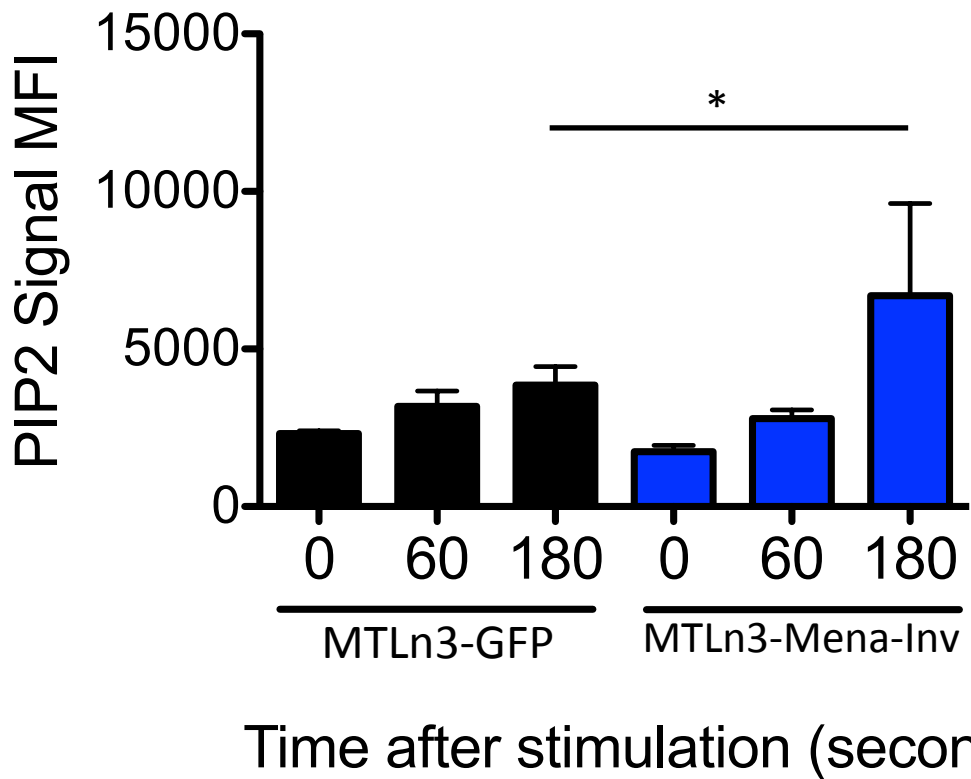
It is *really* important to look at your data.



m = 4 experiments
n = 5 cells / experiment

Data shown as mean +/- SD

It is *really* important to look at your data.



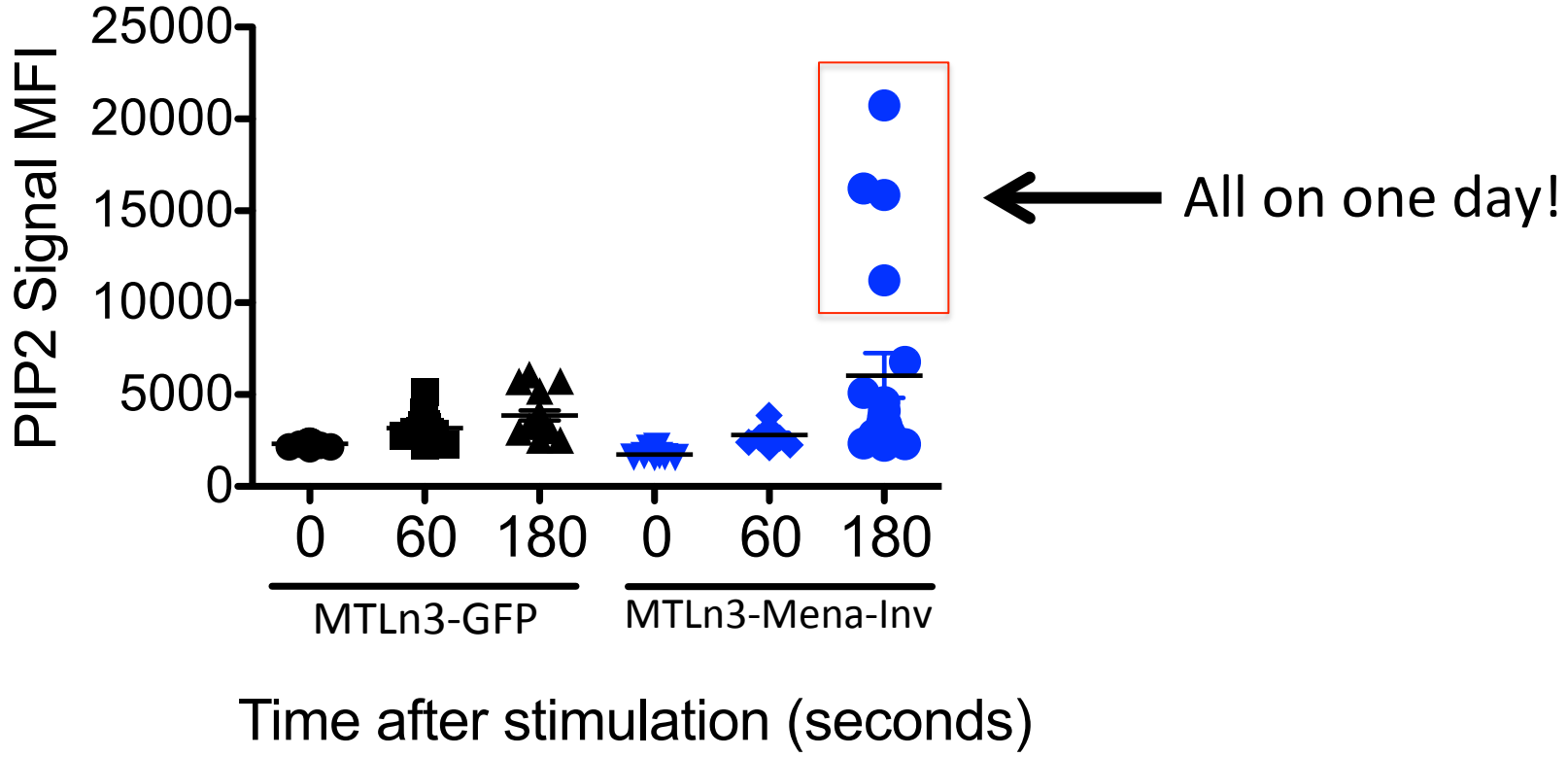
m = 4 experiments

n = 5 cells / experiment

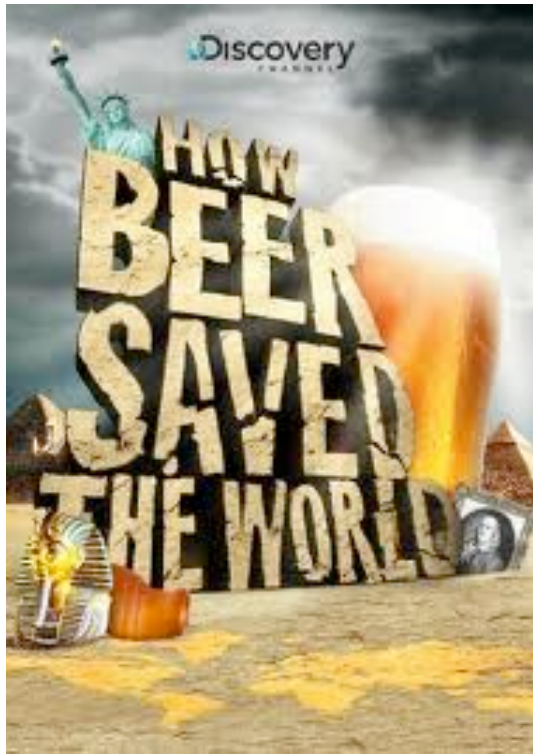
Data shown as mean +/- 95% CI

* $p < 0.05$, two-tailed t-test

It is *really* important to look at your data.



Multiple Comparisons



William Sealy Gosset

aka “Student”

Guinness Brewery,
Dublin, Ireland

“The probable error
of a mean”
Biometrika, 1908



“The circumstances of brewing work, with its variable materials and susceptibility to temperature change and necessarily short series of experiments, are all such as to show up most rapidly the limitations of large sample theory and emphasize the necessity for a correct method of treating small samples. It was thus no accident, but the circumstances of his work, that directed ‘Student’s’ attention to this problem” (McMullen 1939, pp. 205–206).

Calculating significance by t-test

$$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{S_{pooled}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

DOF $n_1 + n_2 - 2$

t_{table} listed by DOF
vs. confidence level

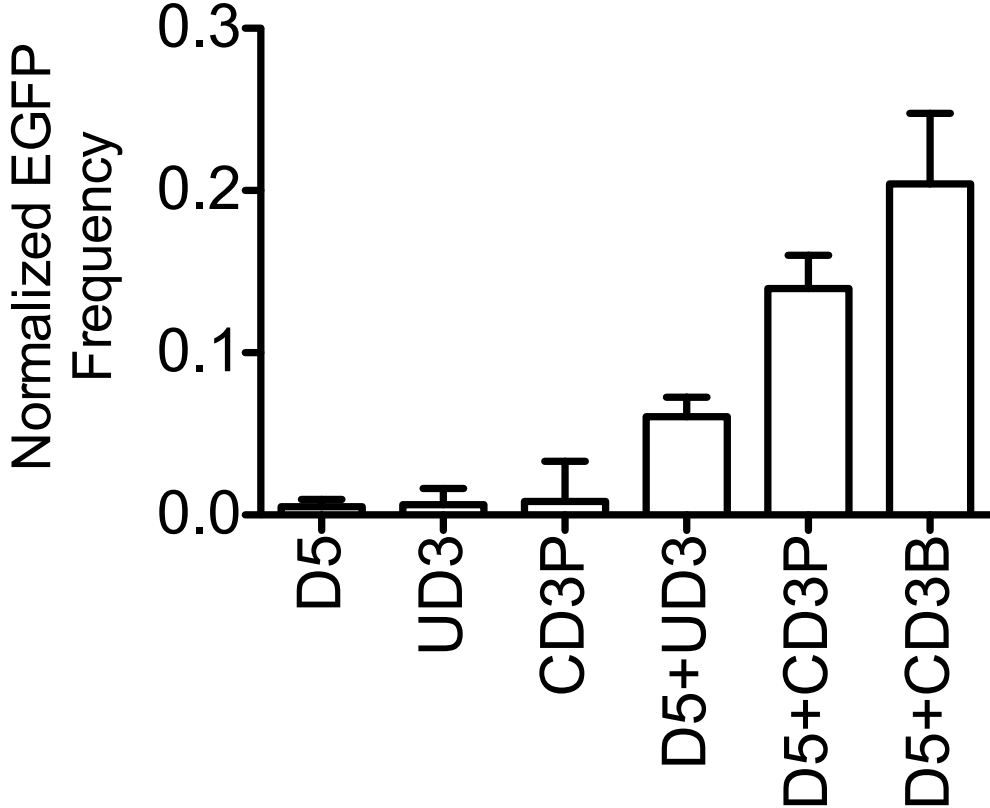
- If $t_{calc} > t_{table}$ difference is significant at that CL
signal:noise *difference in mean vs. spread*
- In Excel, use *TTEST* function
 - returns p -value \rightarrow confidence level (CL)
if $p = 0.01$, C.L. = 99%

Let's revisit some of the t-test assumptions:

Context for statistical comparisons

- Every statistical test
 - Asks a specific question
 - Has *assumptions*
 - Requires *human interpretation (G/GO)*
- Some t-test assumptions
 - normal distribution (cf. Mann-Whitney test)
 - equal variances (type 2 in Excel; type 3 = unequal)
 - not appropriate for multiple comparisons (cf. ANOVA) *(ignore in 105)*
- Posing a question *are means equal and sample*

How many comparisons did you make in Mod1?



How do we make these conclusions more reliable?

| | UD3 | CD3P | D5+UD3 | D5+CD3P | D5+CD3B |
|----------------|------------|-------------|---------------|----------------|----------------|
| D5 | 0.85 | 0.71 | 2.14E-10 | 1.75E-13 | 6.38E-09 |
| UD3 | | 0.80 | 1.73E-08 | 8.95E-14 | 5.09E-09 |
| CD3P | | | 0.0009 | 1.62E-08 | 8.20E-09 |
| D5+UD3 | | | | 3.78E-08 | 9.01E-07 |
| D5+CD3P | | | | | 0.009 |

A conservative approach: the Bonferroni correction

$$\text{Number of pairwise comparisons} = m = \frac{N*(N-1)}{2} = (6*5)/2 = 15$$

$$\alpha_{\text{corrected}} = \alpha/m$$