Questions:

Was the Pulse Experiment **sufficiently controlled** so that we can interpret our treatment effect estimate as a valid estimate of **the effect of standing**?

There are 3 ways to reduce the standard error in this experiment.

$$\sigma_r(\hat{\delta}) = \sqrt{\frac{\text{Variance of population + Variance of measurements}}{\text{Sample size}}}$$

Should we try to reduce all 3?

Questions:

Was the Pulse Experiment **sufficiently controlled** so that we can interpret our treatment effect estimate as a valid estimate of **the effect of standing**?

Sources of Confusion How could we be misled by our experiment?

TABLE 1. Potential sources of confusion in an experiment and means for minimizing their effect.

Source of confusion	Features of an experimental design that reduce or eliminate confusion
1. Temporal change	Control treatments
2. Procedure effects	Control treatments
3. Experimenter bias	Randomized assignment of experimental units to treatments Randomization in conduct of other procedures "Blind" procedures*
4. Experimenter-generated variability (random error)	Replication of treatments
5. Initial or inherent variability among experimental units	Replication of treatments Interspersion of treatments Concomitant observations
6. Nondemonic intrusion†	Replication of treatments Interspersion of treatments
7. Demonic intrusion	Eternal vigilance, exorcism, human sacrifices, etc.

* Usually employed only where measurement involves a large subjective element.

† Nondemonic intrusion is defined as the impingement of chance events on an experiment in progress.

Control treatments

Positive / Negative controls

Is my procedure good?

Would the response change with no intervention?

Replication

Reduce sampling error

Variation among EU

Before, during, or after experiment

Measurement error

There are 3 ways to reduce the standard error in this experiment.

$$\sigma_r(\hat{\delta}) = \sqrt{\frac{\text{Variance of population + Variance of measurements}}{\text{Sample size}}}$$

Should we try to reduce all 3?

Decrease: Variance of measurements	Problems?
Measure for longer	
Standardize measurements	More effort
Measure multiple times	
Increase: Sample size	More effort
Decrease: Variance of population	
Select more similar people	
Standardize behavior before/during	Reduce **scope**
Standardize behavior before/during	

Our results will be more precise, but for a more limited question

Tradeoff between confidence in our results and the generality of our results

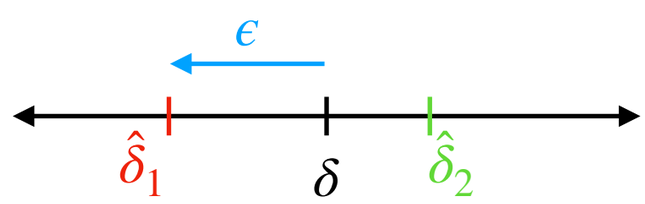
How good is your estimate?

Error, bias, precision

Error: difference between estimate and true value

$$\epsilon = \hat{\delta} - \delta \qquad \epsilon = \hat{\mu}_A - \mu_A$$

We can never know this value



Bias: Average ϵ across repeats of the experiment

With randomization, valid treatments, blinding, bias = 0

Precision: Consistency of estimates across repeats of the experiment

What is the average $|\epsilon|$?

Statement about the **Precision** of this estimate

1) Analytical: Standard Error: $\sigma_r(\hat{\delta})$
2) Estimated: Estimated Standard Error (SED)
3) Confidence Interval

Outcomes of an analysis

Estimate of a treatment effect: $\hat{\delta}_{B-A}$

Statement about the **Precision** of this estimate

1) Analytical: Standard Error: $\sigma_r(\hat{\delta})$

2) Estimated: Estimated Standard Error (SED)

3) Confidence Interval

Analytical: Standard Error: $\sigma_r(\hat{\delta})$

~ **Average error** $|\hat{\delta} - \delta|$ if you were to repeat the experiment many times

$$\hat{\delta}_{B-A} = \frac{1}{n}\sum \hat{\delta}_j$$ Direct estimate = average of *n* observations

$$\sigma_r(\hat{\delta}) = \sqrt{\frac{\text{Variance of population + Variance of measurements}}{\text{Sample size}}}$$

$$= \sqrt{\frac{\sigma_\delta^2 + \sigma_\epsilon^2}{\text{n}}} \longleftarrow \sigma_r^2(\hat{\delta}_j)$$ (Standard Error of measurements)²

Problem: What are σ_δ^2 and σ_ϵ^2 ?

We can't report the value of $\sigma_r(\hat{\delta})$, only the equation

Solution: **estimate it** from our data

Estimated Standard Error

Outcomes of an analysis

Estimate of a treatment effect: $\hat{\delta}_{B-A}$

Statement about the **Precision** of this estimate

1) Analytical: Standard Error: $\sigma_r(\hat{\delta})$

2) Estimated: Estimated Standard Error (SED)

3) Confidence Interval

Estimated Standard Error (SED)

$$\hat{\delta}_{B-A} = \frac{1}{n}\sum \hat{\delta}_j$$ Direct estimate = average of *n* observations

$$\sigma_r(\hat{\delta}) = \sqrt{\frac{\text{Variance of population + Variance of measurements}}{\text{Sample size}}}$$

Each estimate $\hat{\delta}_i$ = TRUE value + measurement error

Sample Variance of $\hat{\delta}_i \approx$ Variance of TRUE values + Variance of errors

$$s^2_{\hat{\delta}} = \frac{\sum (\hat{\delta}_j - \hat{\delta})^2}{n-1}$$ **Observed** variance of estimates around their mean

$$\textbf{SED} = \sqrt{\frac{s^2_{\hat{\delta}}}{n}}$$ This is an estimate of $\sigma_r(\hat{\delta})$

Outcomes of an analysis

Estimate of a treatment effect: $\hat{\delta}_{B-A}$

Statement about the **Precision** of this estimate

1) Analytical: Standard Error: $\sigma_r(\hat{\delta})$

2) Estimated: Estimated Standard Error (SED)

3) Confidence Interval

Estimated: Estimated Standard Error (SED)

Sample Variance of $\hat{\delta}_i \approx$ Variance of TRUE values + Variance of errors

$$s^2_{\hat{\delta}} = \frac{\sum (\hat{\delta}_j - \hat{\delta})^2}{n-1}$$

$$\textbf{SED} = \sqrt{\frac{s^2_{\hat{\delta}}}{n}}$$ This is an estimate of $\sigma_r(\hat{\delta})$

This is a number we can calculate and report

It's interpreted as approximately the average expected size of an error from this experiment, if it were repeated many times

normal units

Note: n is used **two times!** **These are different!**

(n-1) in the denominator of s^2: **Degrees of freedom**

independent deviation used to estimate s^2

(n) in the denominator of SED: **# EU**

independent samples used to estimate $\hat{\delta}$

Outcomes of an analysis

Estimate of a treatment effect: $\hat{\delta}_{B-A}$

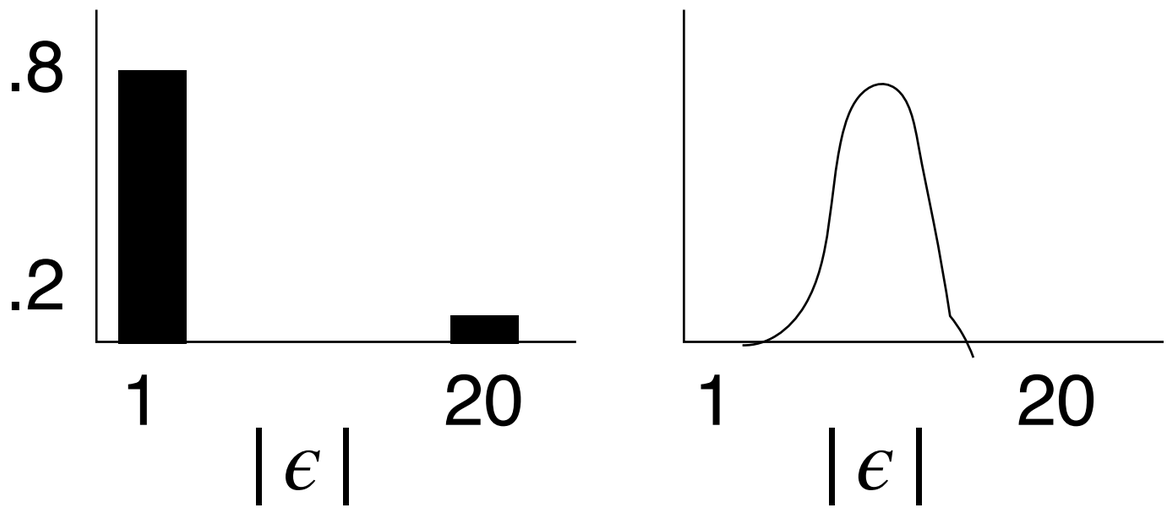
Statement about the **Precision** of this estimate

1) Analytical: Standard Error ($\sigma_{\hat{\delta}}$)

2) Estimated: Estimated Standard Error (SED)

3) Confidence Interval:

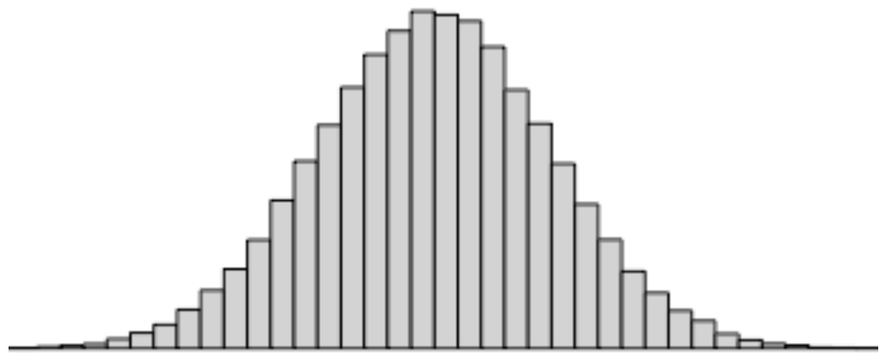
SED is the **average** error. But averages can be misleading



Can we put bounds on errors?

What is the biggest possible error?

If $\hat{\delta}_i$ follows a Normal distribution

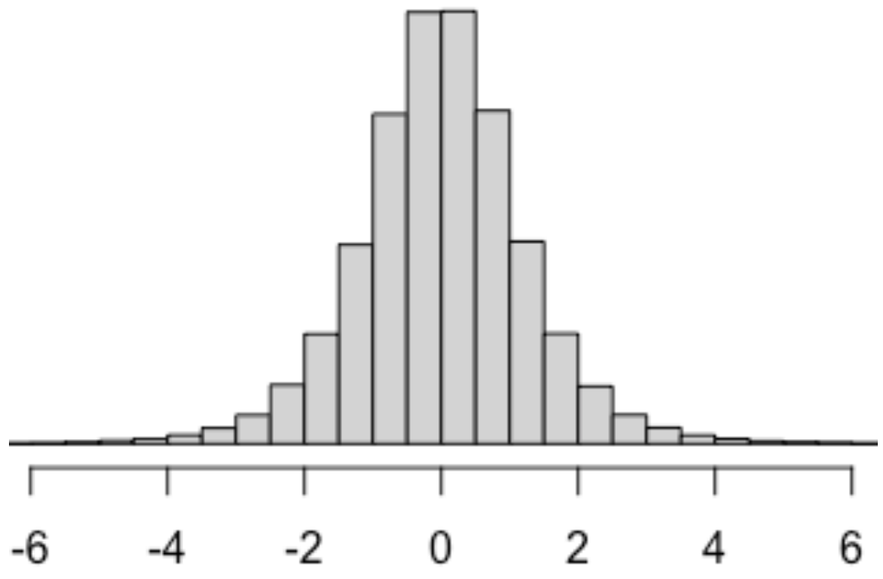


Then the distribution of errors ($\hat{\delta} - \delta$)

follow a *T-distribution*, scaled by SED, with (n-1) Degrees of Freedom

$$\epsilon \sim t_{df} \times SED$$

$$\frac{\hat{\delta} - \delta}{SED} \sim t_{df}$$



Outcomes of an analysis

Estimate of a treatment effect: $\hat{\delta}_{B-A}$

Statement about the **Precision** of this estimate

1) Analytical: Standard Error ($\sigma_{\hat{\delta}}$)

2) Estimated: Estimated Standard Error (SED)

3) Confidence Interval:

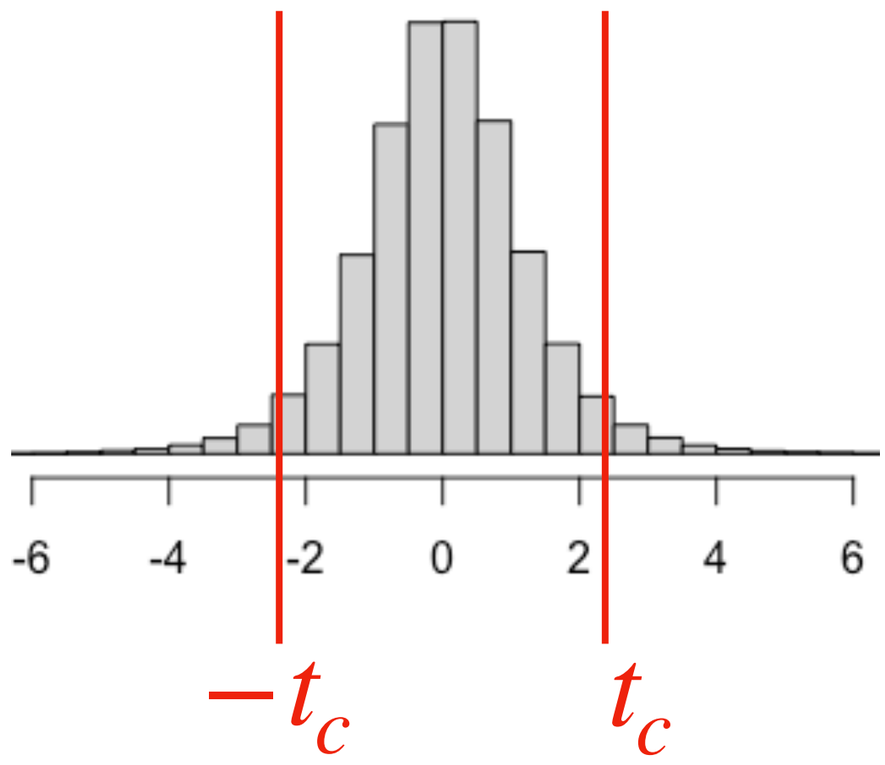
T-distribution: How much bigger than the **SED** is the actual error likely to be?

$$t = \frac{\hat{\delta} - \delta}{\text{SED}}$$

Actual error

Estimated average error

$(1-\alpha) \times 100\,\%$ **of relative errors**



Confidence Interval: Interval drawn around an estimate that should include a defined % of errors from experiments like this

$\hat{\delta} \pm t_c \times SED$ estimate ± X_c * SE

t_c is called the **critical value**

We choose a **level** (e.g. 95%), and set $\alpha = 1 - \text{level}/100$

We use the **qt()** function in R to calculate t_c

t_c = qt(α/2,df,lower.tail=F)

depends on α and df

Outcomes of an analysis

Estimate of a treatment effect: $\hat{\delta}_{B-A}$

Statement about the **Precision** of this estimate

1) Analytical: Standard Error ($\sigma_{\hat{\delta}}$)

2) Estimated: Estimated Standard Error (SED)

3) Confidence Interval:

Confidence Interval:

$$\hat{\delta} \pm t_c \times SED$$

t_c = qt(α/2,df,lower.tail=F)

Size is proportional to $2 \times t_c$

Calculate t_c for the following scenarios:

n	**level=0.95**		**level=0.99**	
2	12.7		63.7	
		4x		10x
4	3.2		5.8	
		1.3x		1.7x
8	2.4		3.5	
		1.1x		1.2x
16	2.1		2.9	

t_c shrinks with increasing sample size

But with decreasing gains

SED also shrinks with increasing sample size for a separate reason

WHY?

$$\textbf{SED} = \sqrt{\frac{s^2_{\hat{\delta}}}{n}}$$

Prepare for Tuesday:

Read Hurlbert 1984 up to “Randomization vs. interspersion

Questions:

What interspersion was needed for **our pulse experiment?**

What interspersion would be needed for **the direct estimate version?**

Our pulse experiment used an indirect design:

Each person was only measured for **one treatment level**

1. How can we estimate the Average Treatment Effect?

2. Will we get the same answer as for a direct version?

3. Can we estimate σ_δ^2 to calculate SED and a CI?