UNDERSTANDING STATISTICS & EXPERIMENTAL DESIGN

Content

About this course

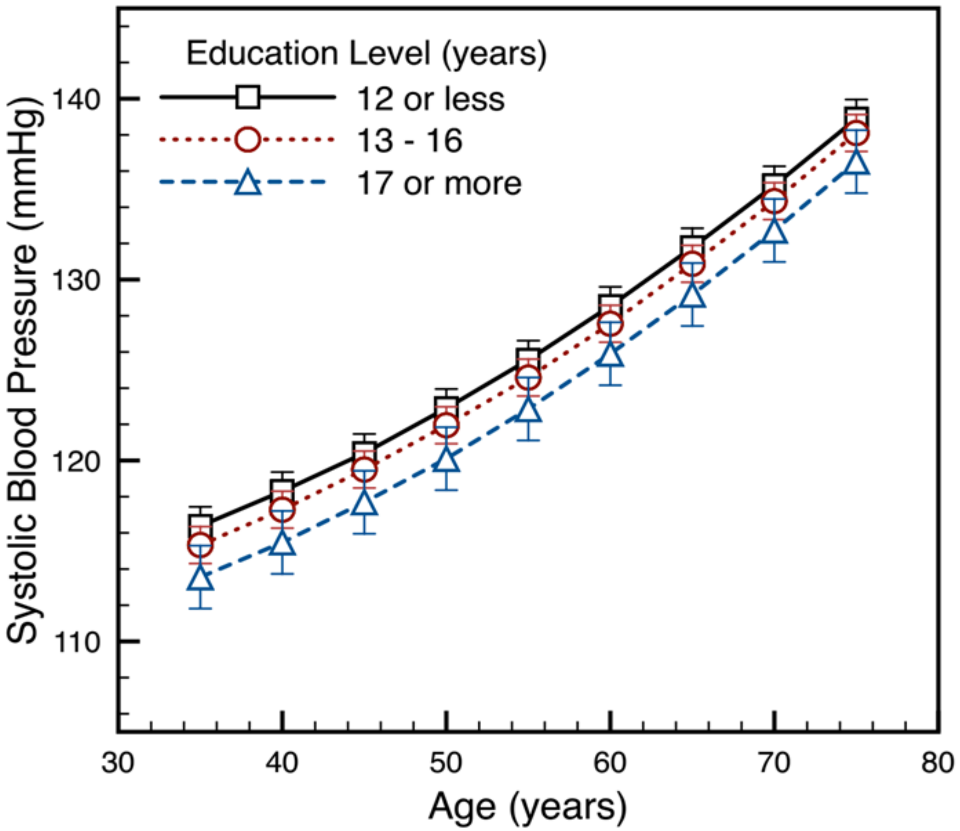
EPFL

Understanding statistics does NOT require computing statistics!

3. Statistics via SDT: t-test

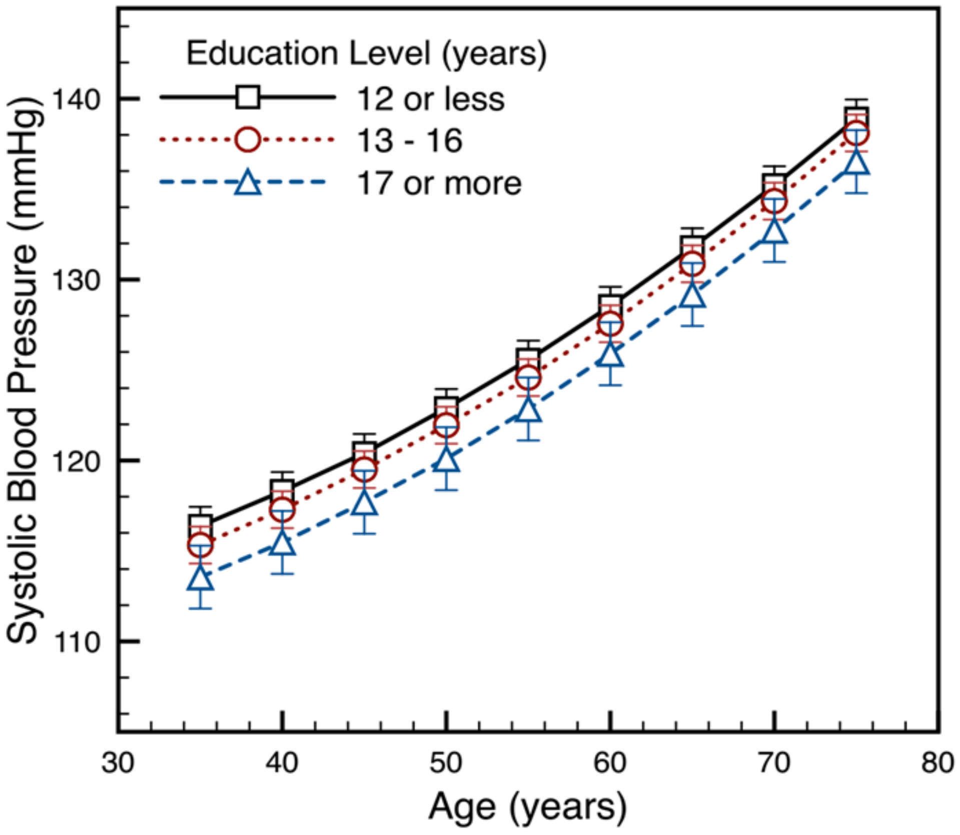
Framingham Studies

Framingham Studies



Associations of education with 30 year life course blood pressure trajectories: Framingham Offspring Study
Eric B Loucks, Michal Abrahamowicz, Yongling Xiao & John W Lynch
BMC Public Health **volume 11**, Article number: 139 (2011)

Framingham Studies

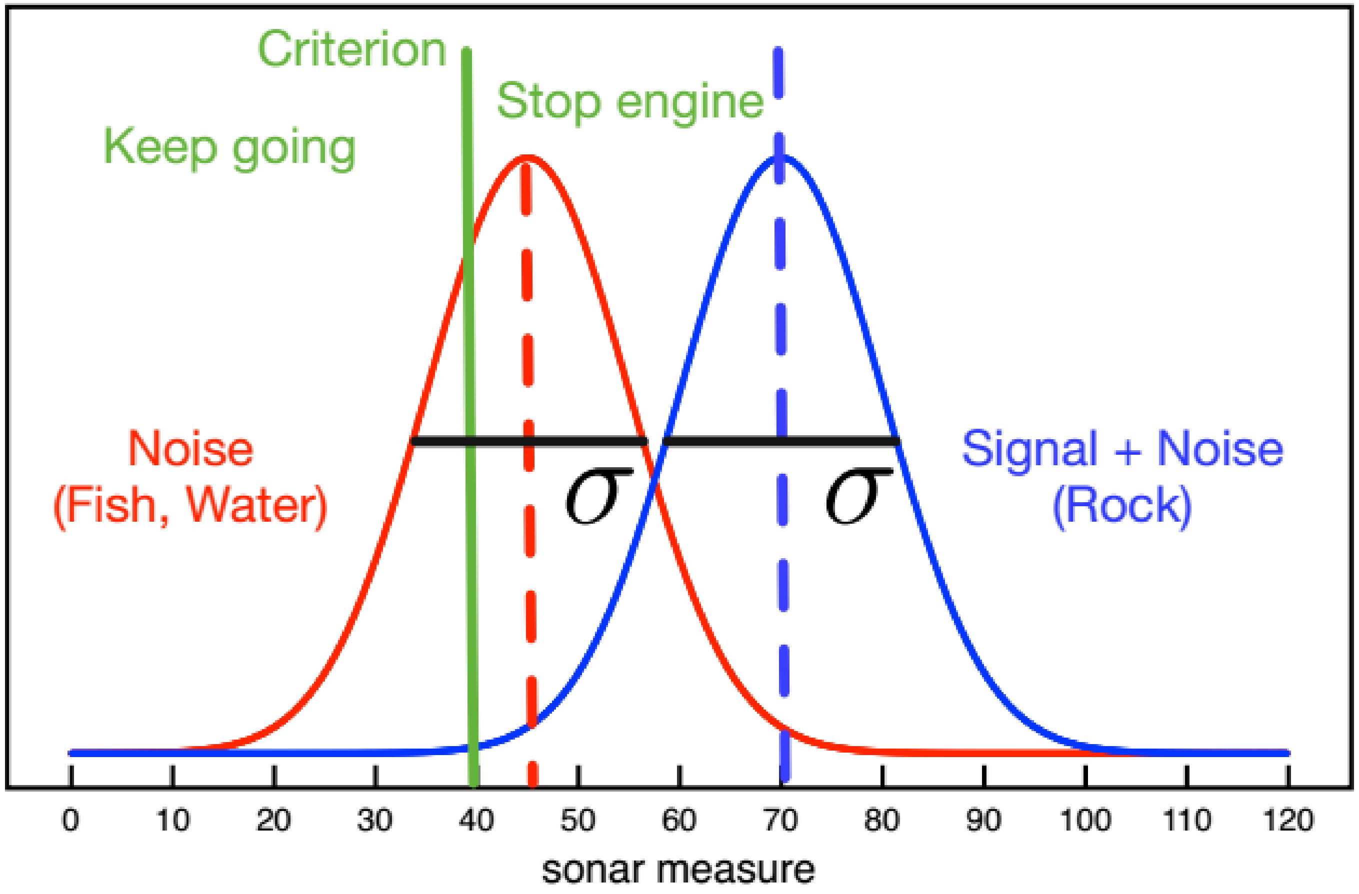


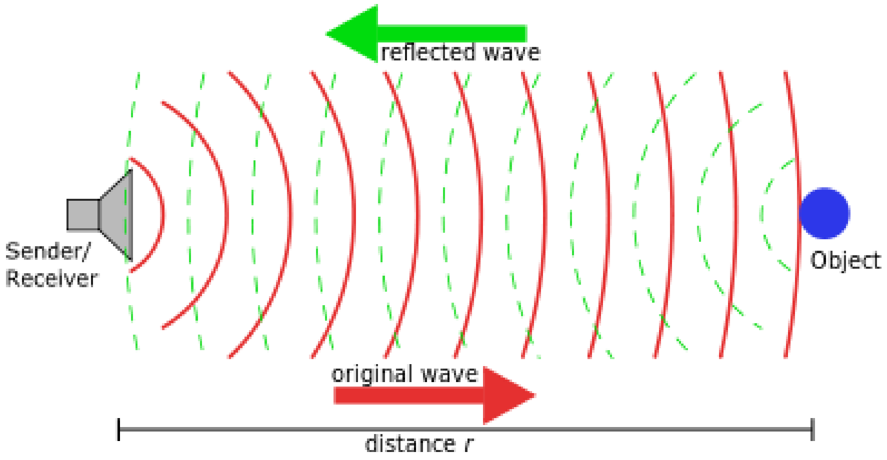
What matters is effect size d'

Small effects are not significant, isn't it?

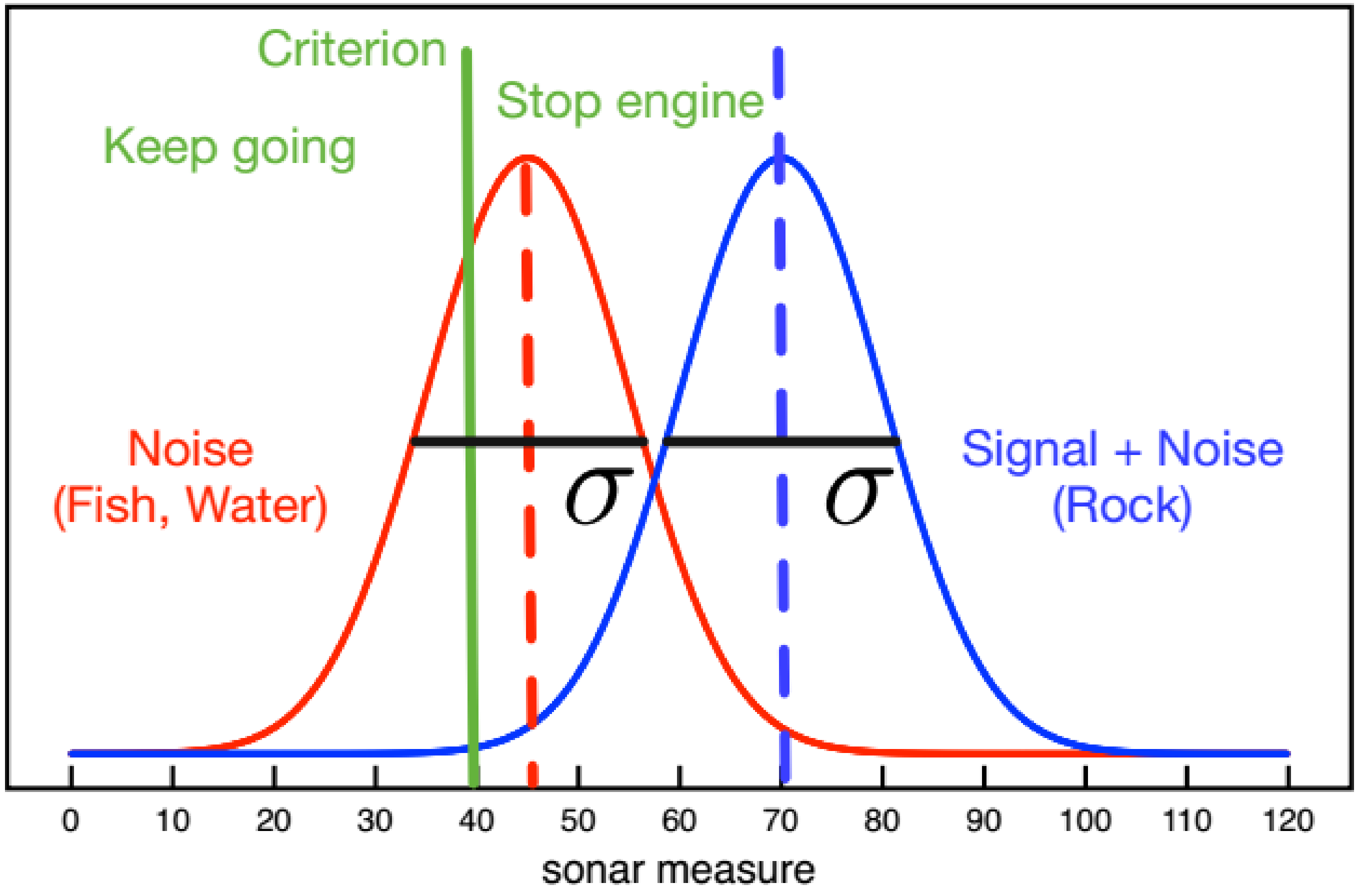
The t-test

Conditional Probability: Noise (fish, water)





Conditional Probability: Noise (fish, water)



Statistics:

1. Is the sonar ok, ie d‘=0?
2. Average out noise
3. Estimate mean and variance

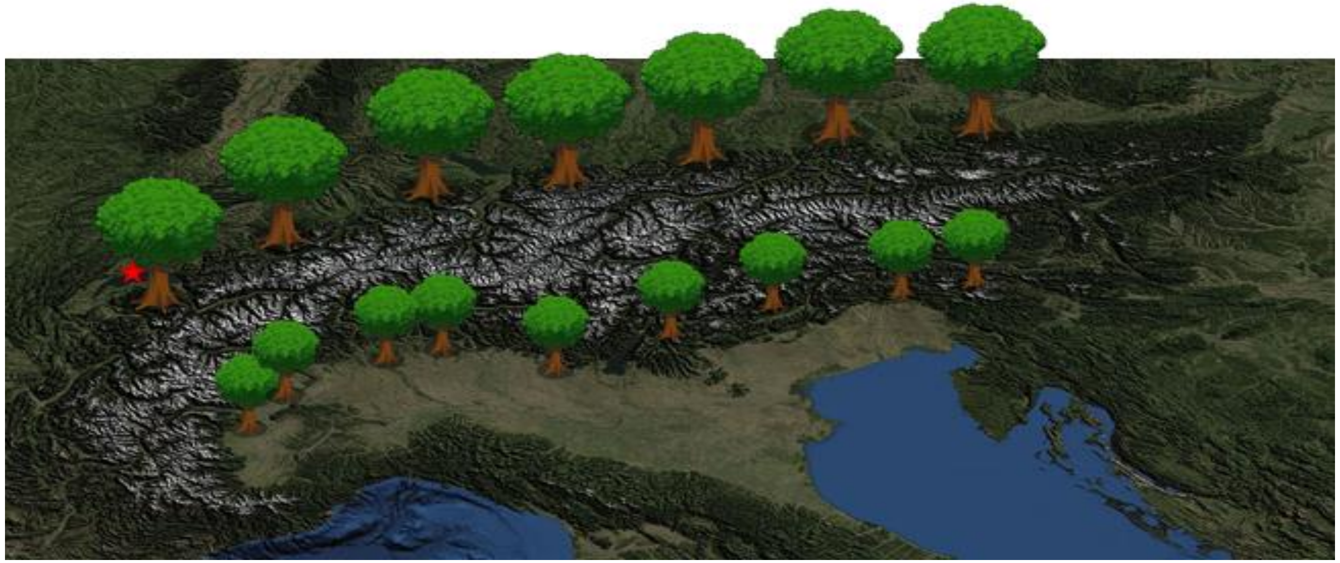
The Problem: Undersampling

Example

Bergmann's rule states that "populations and species of **larger size** are found **in colder environments**, and species of **smaller size** are found in **warmer regions**" (Wikipedia.org).

Is this true for plants too?

Example



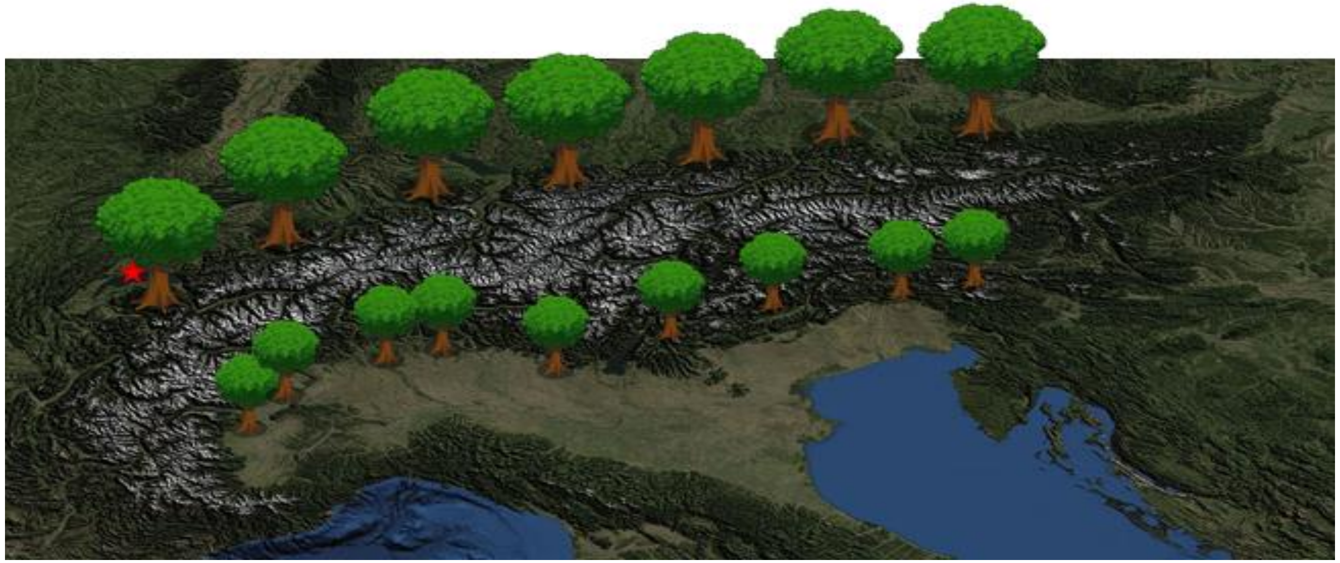
Bergmann's rule states that "populations and species of **larger size** are found **in colder environments**, and species of **smaller size** are found in **warmer regions**" (Wikipedia.org).

Compare mean height of oaks in the North and in the South

μ(North) > μ(South)

Example

$$\bar{x}(North) = \frac{1}{n}\sum_{i=1}^{n} x_i$$



$$\bar{x}(South) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Bergmann's rule states that "populations and species of **larger size** are found **in colder environments**, and species of **smaller size** are found in **warmer regions**" (Wikipedia.org).

Compare mean height of oaks in the North and in the South

μ(North) > μ(South)

Measure the height of all these oaks?

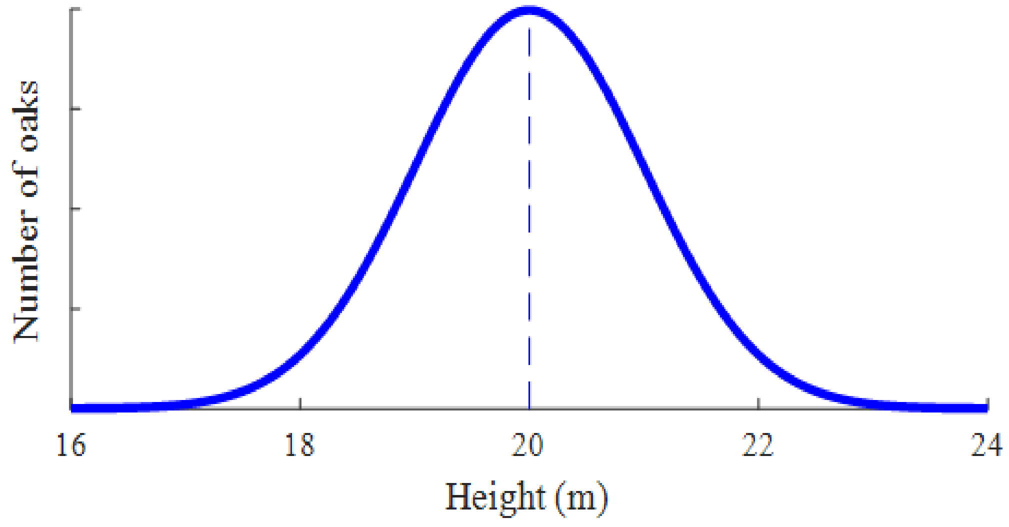
Measure the height of all these oaks?

Undersampling

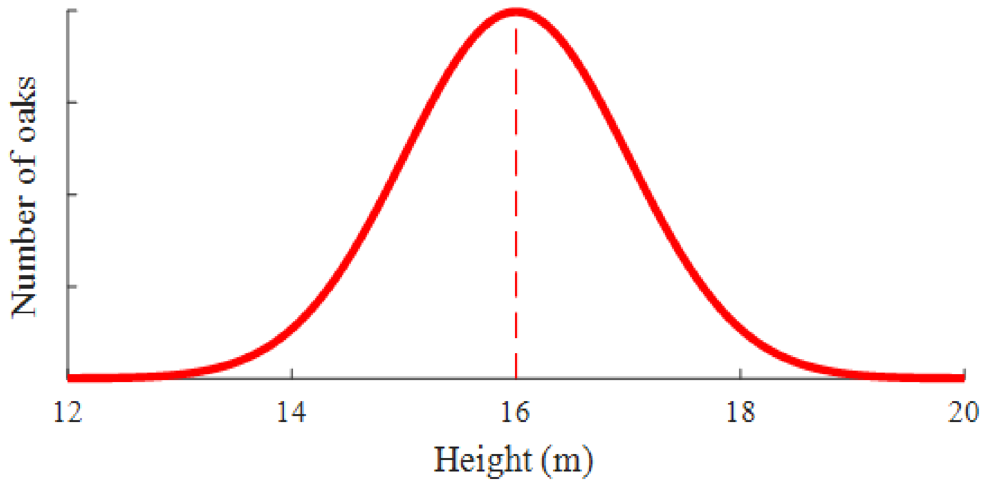
Estimate μ(North) and μ(South) by sub-samples **n <<< N:**

$$\bar{x}(North) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x}\ (South) = \frac{1}{n}\sum_{i=1}^{n} x_i$$



North



South

Undersampling causes Errors

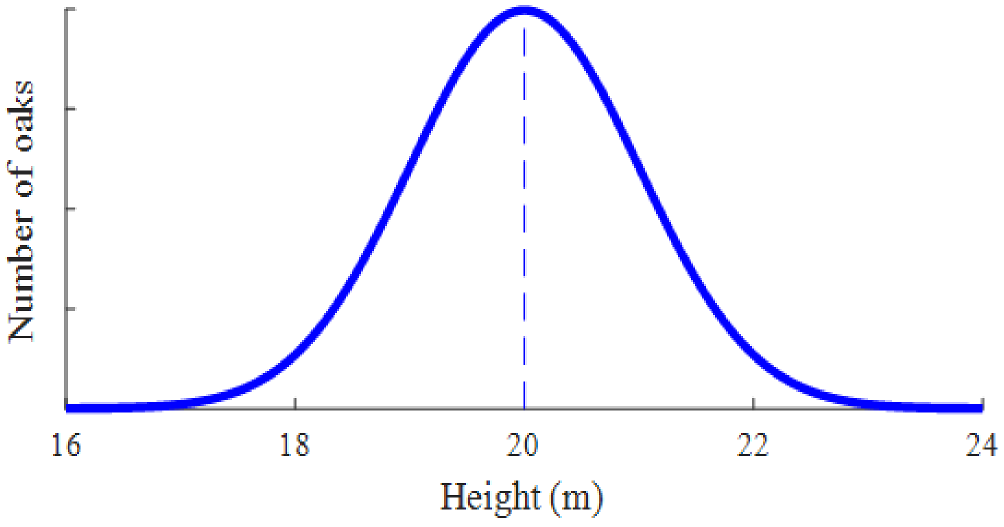
	H_0 is false	H_0 is true
Decide there is a significant difference	Hit	False Alarm (Type I error)
Do not decide there is a significant difference	Miss (Type II error)	Correct Rejection

Undersampling

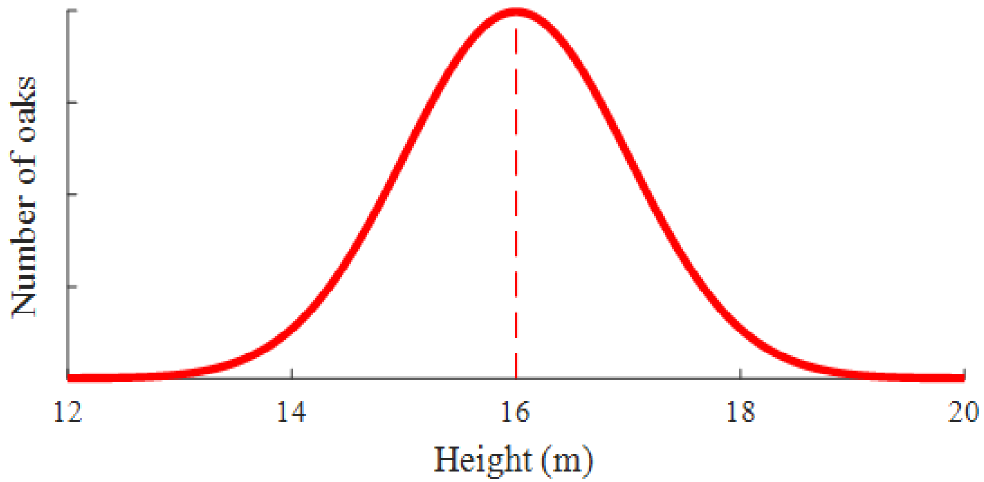
Estimate µ(North) and µ(South) by sub-samples **n <<< N:**

$$\bar{x}(North) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x}\,(South) = \frac{1}{n}\sum_{i=1}^{n} x_i$$



North



South

Most important equation

$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

Significant: $d * \sqrt{n/2} > 1.96\ (depending\ on\ n)$

Step 1: Sampling Error

Notations

$$\delta = d' = \frac{\mu_{SN} - \mu_N}{\sigma}$$

$$d = \frac{\bar{x}_{SN} - \bar{x}_N}{s}$$

Notations

EPFL

Some definitions

We collected a sample of n scores x_i

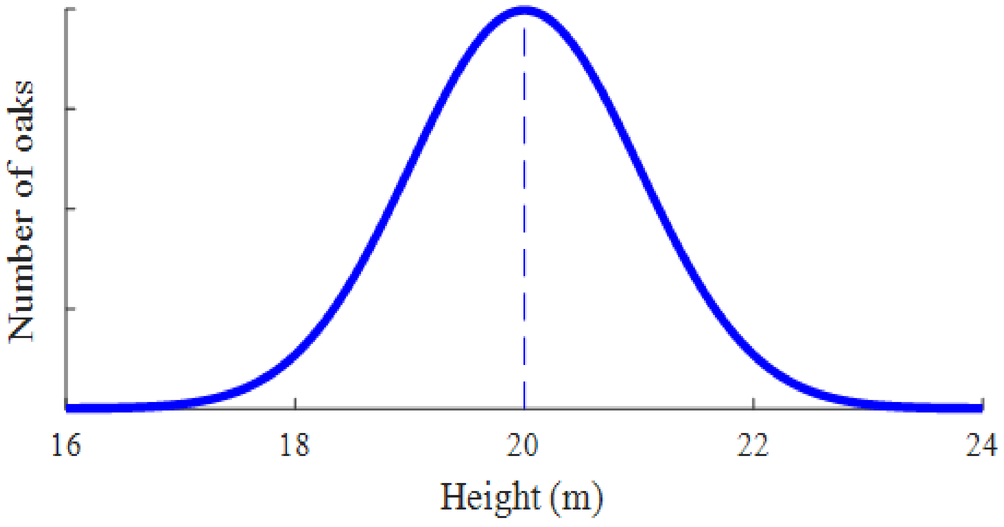
- Sample mean
  $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
  where the symbol $\sum$ means "add up all following terms"
- Sample variance
  $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$
- Sample standard deviation
  $s = \sqrt{s^2}$
- Standard error
  $s_{\overline{x}} = s/\sqrt{n}$

Undersampling

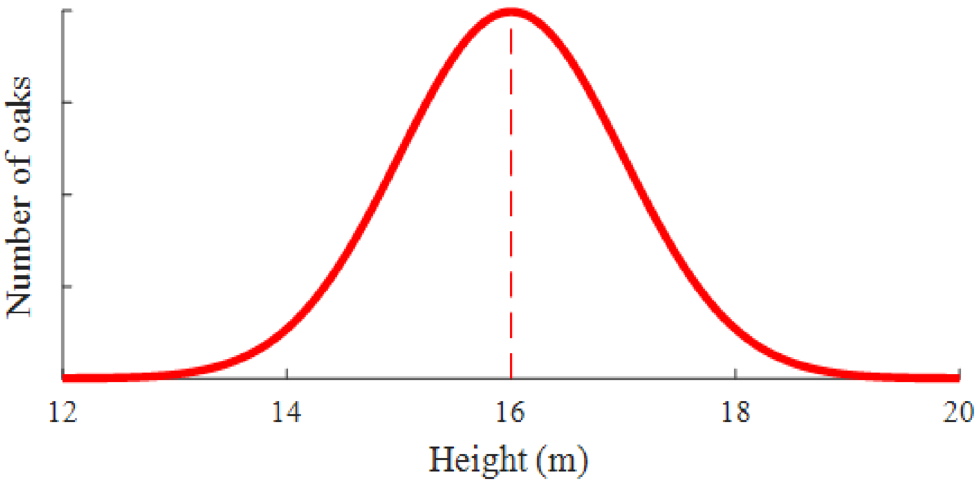
Estimate μ(North) and μ(South) by sub-samples **n <<< N:**

$$\bar{x}(North) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x}\,(South) = \frac{1}{n}\sum_{i=1}^{n} x_i$$



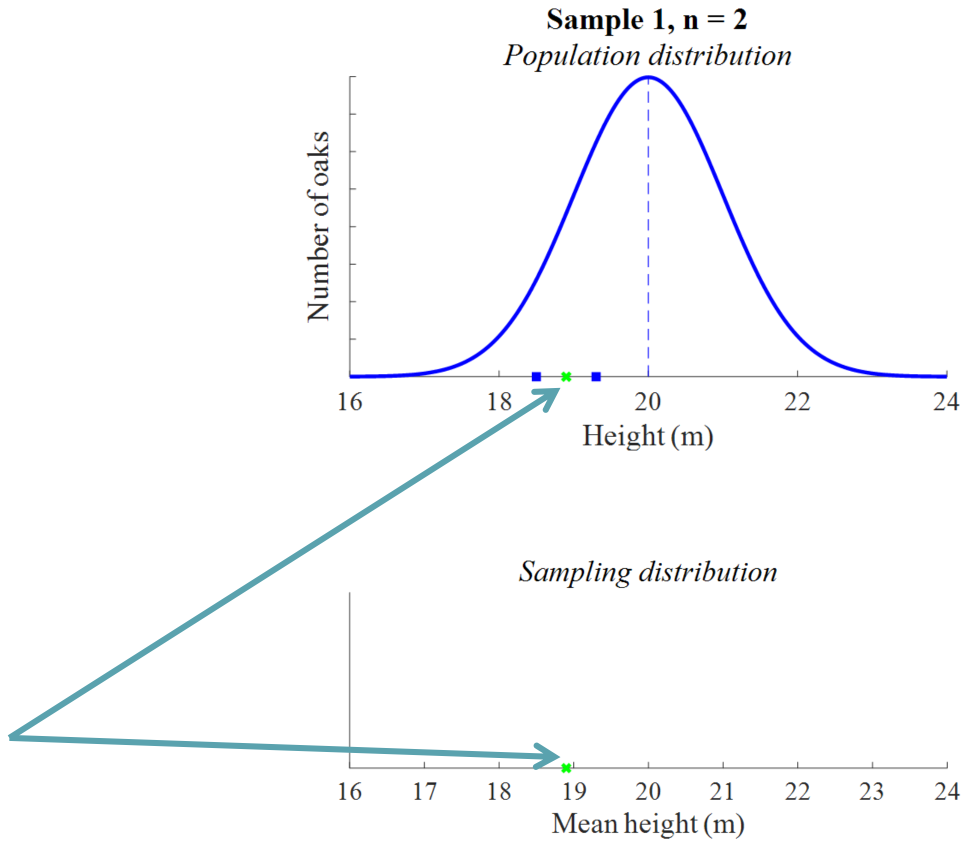
North



South

Step 1: Sampling & Sampling Error: Theory

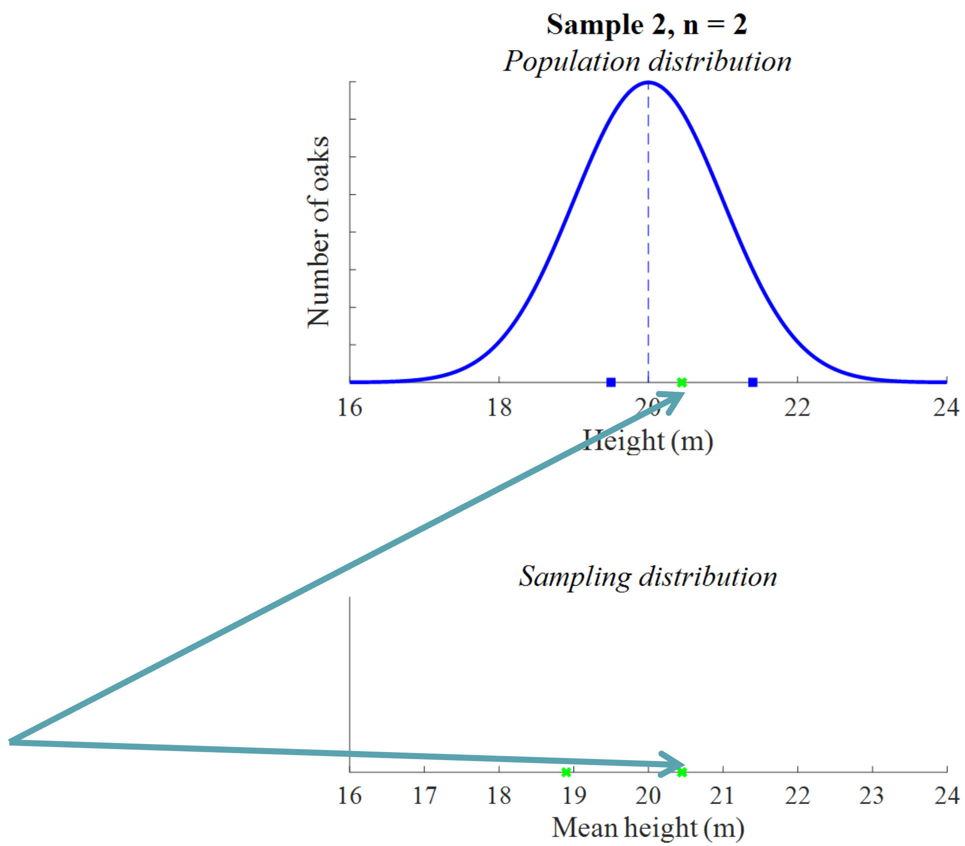
Because of sub-sampling, the sample mean may differ from the true mean



Mean of this random sample of n = 2 trees

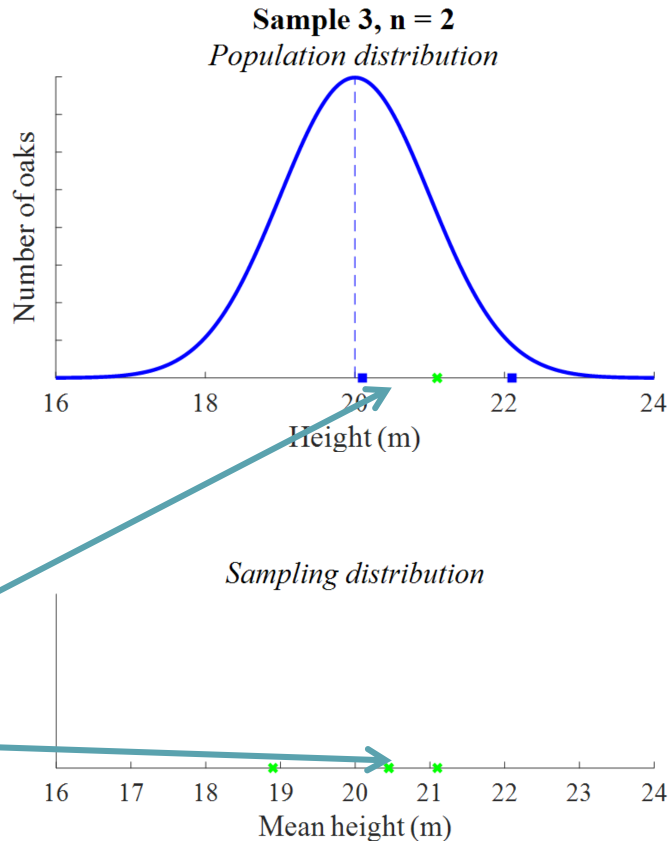
$$\bar{x}(North) = \frac{1}{2}\sum_{i=1}^{2} x_i$$

Step 1: Sampling & Sampling Error



Mean of this random sample of n = 2 trees

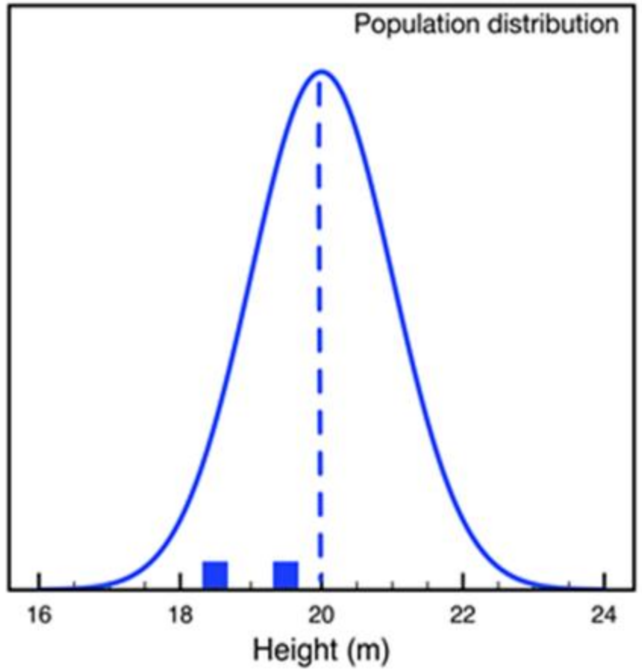
Step 1: Sampling & Sampling Error

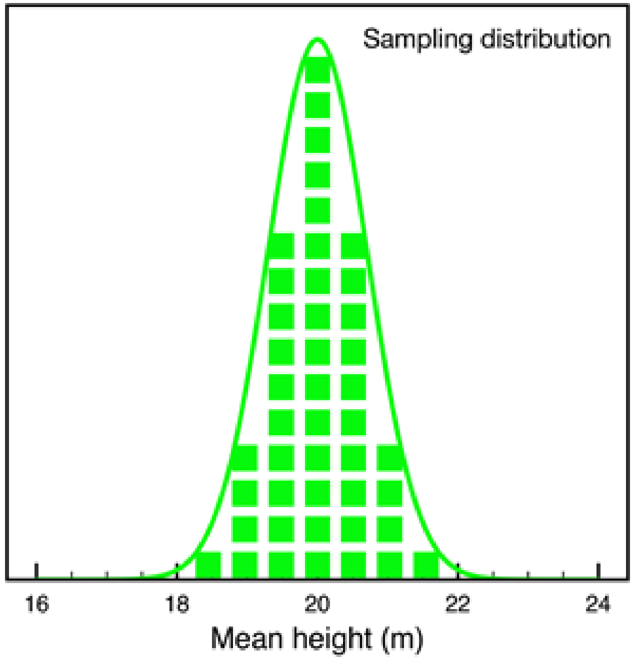


Mean of this random sample of n = 2 trees

Many Samples Later…

Distribution of sample means with sample size n=2

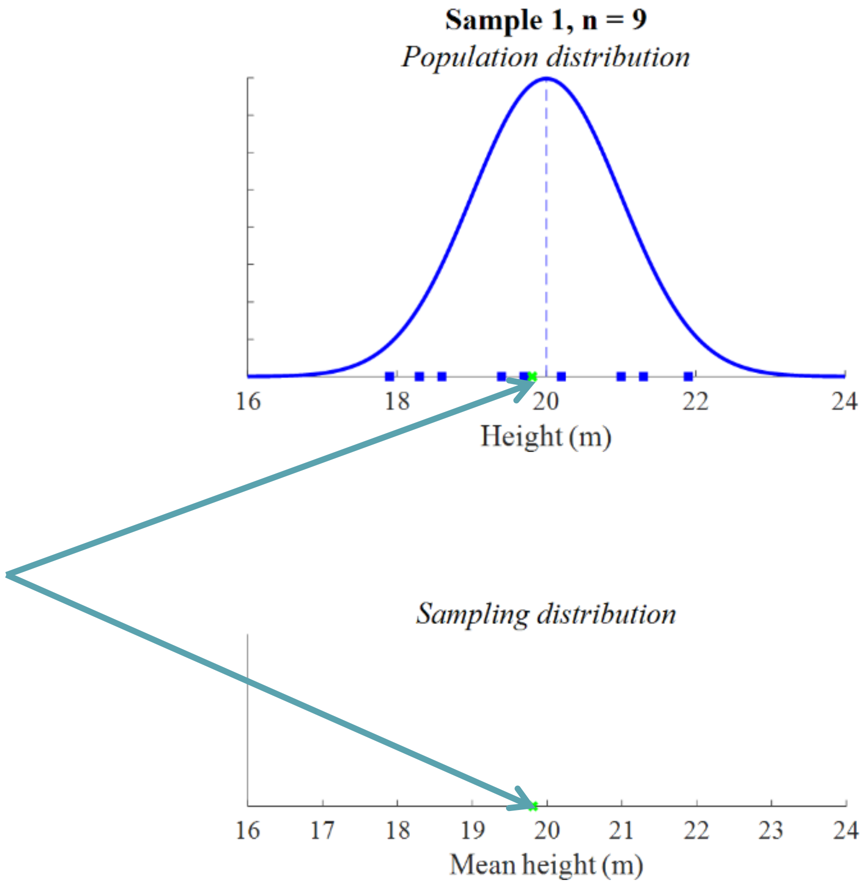




Increase sample size (n) …

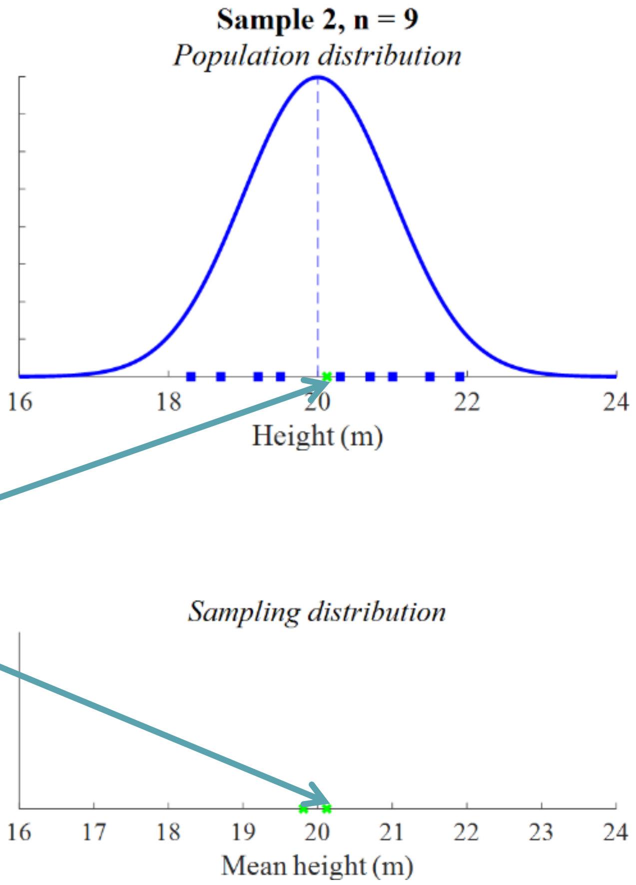
Because of sub-sampling, we make errors in the estimates of the means

Mean of this random sample of n = 9 trees



$$\bar{x}(North) = \frac{1}{9}\sum_{i=1}^{9} x_i$$

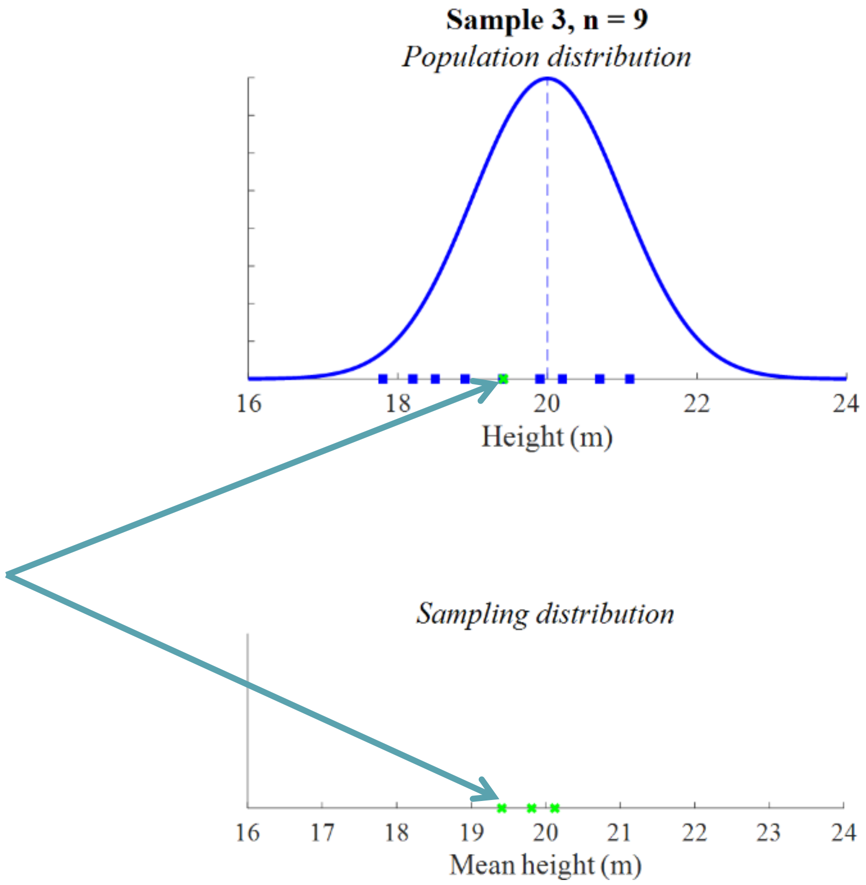
Increase sample size (n) …



Mean of this random sample of n = 9 trees

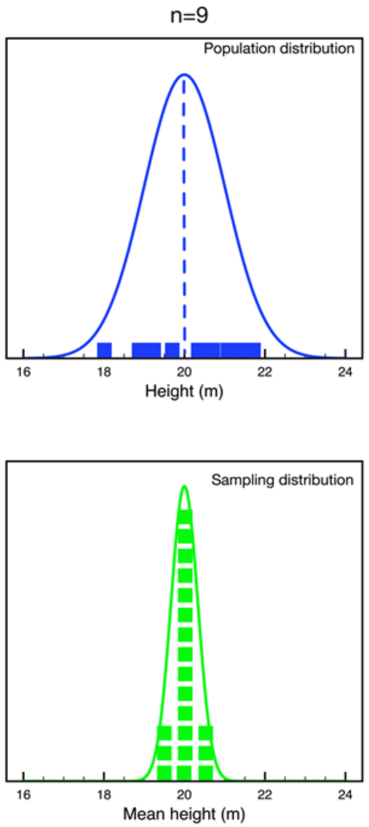
Increase sample size (n) …

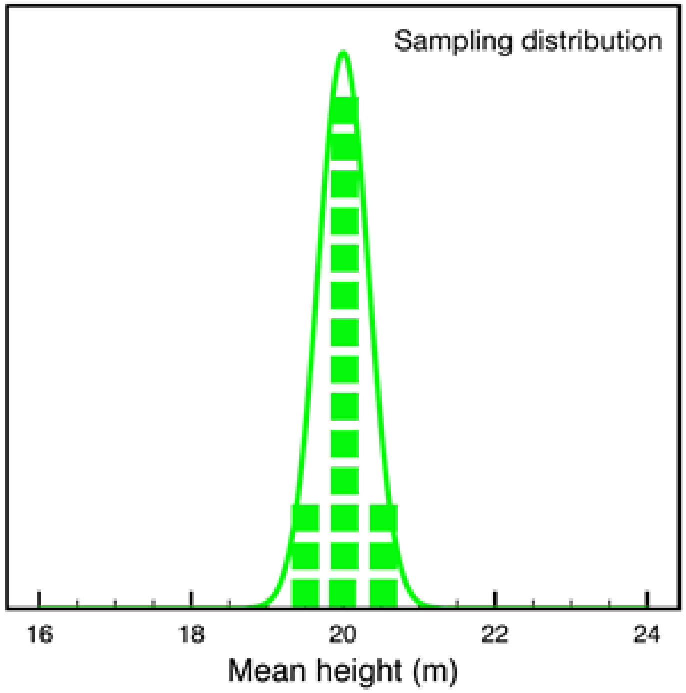
Mean of this random sample of n = 9 trees



Many Samples Later…

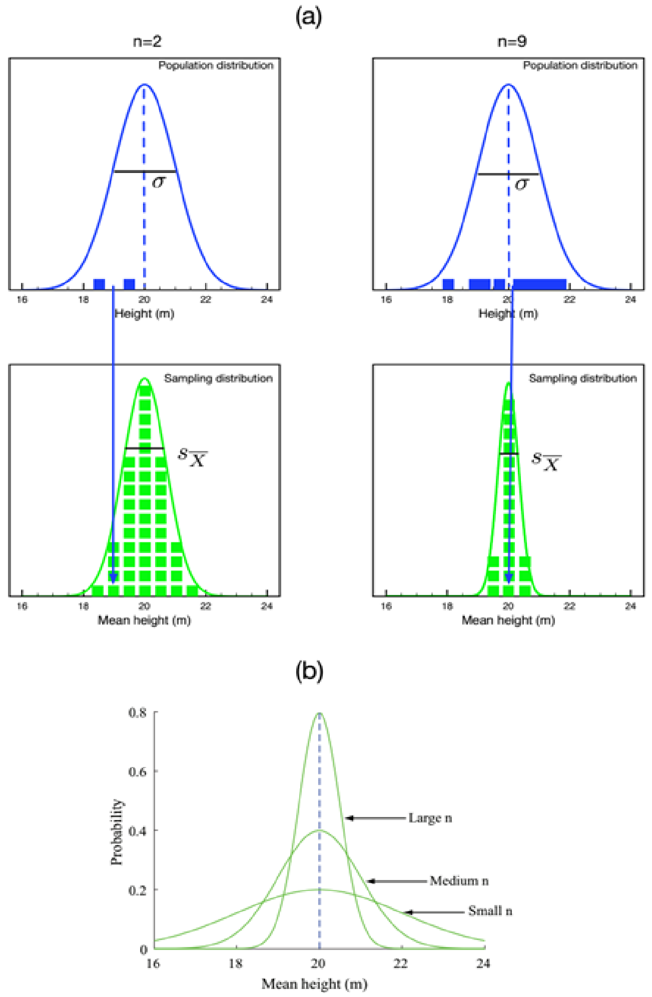
Distribution of sample means with sample size n=9



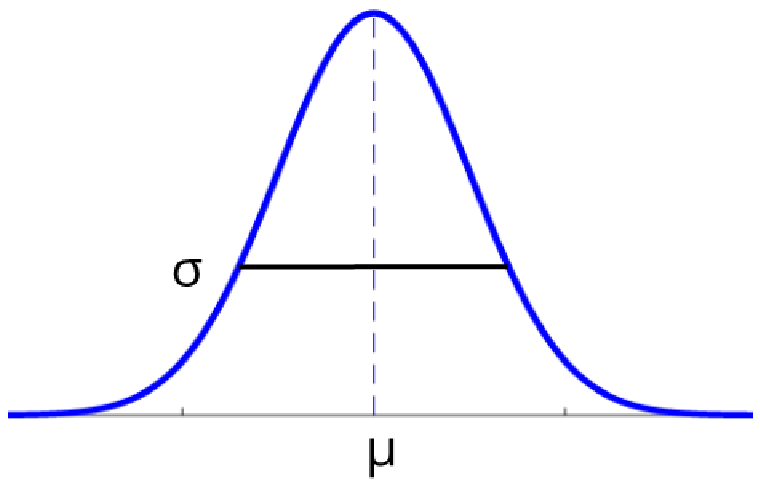


Many Samples Later…

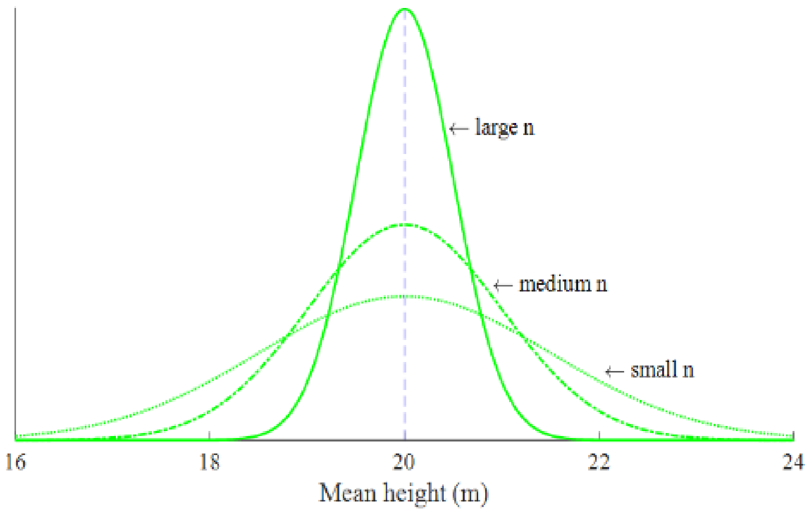
The sampling error becomes smaller when n increases, i.e, the variances of the sampling distributions become smaller



Many Samples Later…



Population distribution



The variances s_x of the sampling distributions become smaller, when n increases

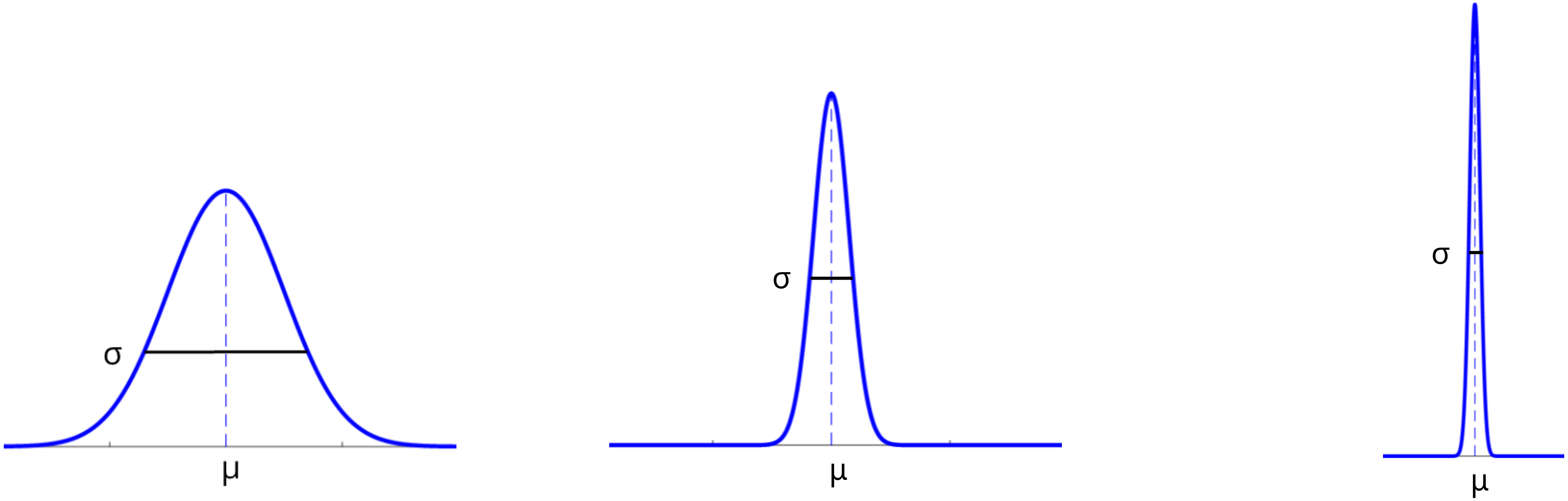
$s_{\bar{x}}$ is a good measure for the sampling error

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

standard error of the mean

Variance

Standard deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$; Standard error: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

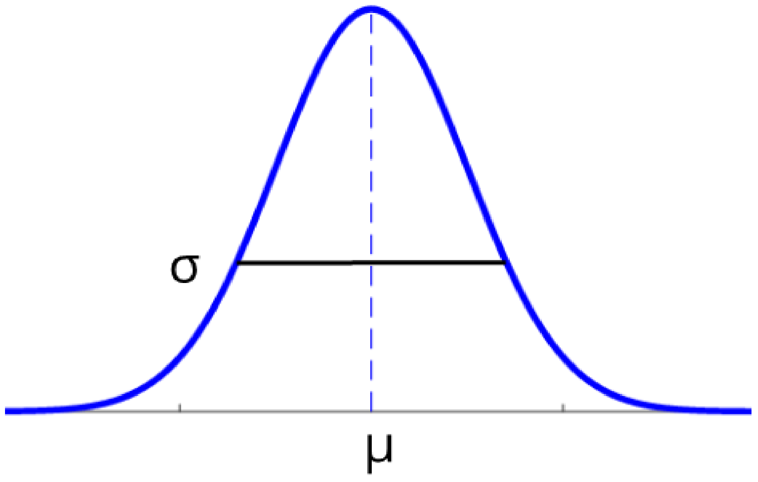


Central Limit Theorem (CLT)

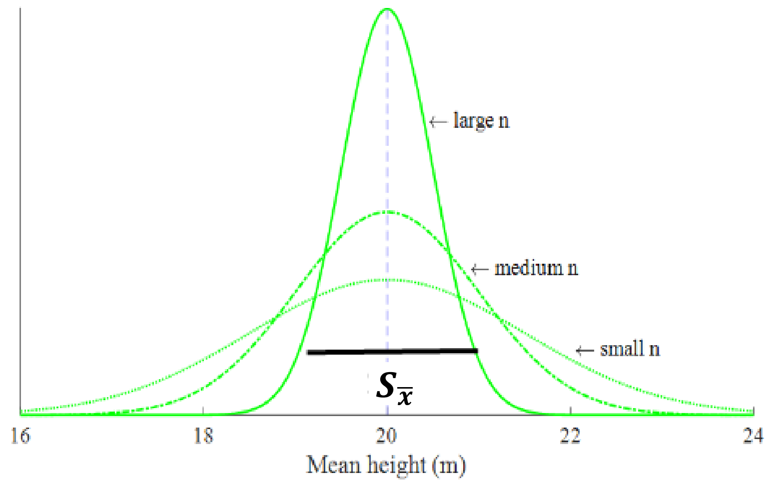
For n → inf:

- $\bar{x} \to \mu$, with $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- $s_{\bar{x}} \to 0$, with $\frac{s}{\sqrt{n}}$ (std error of the mean: SEM)
- Sampling distribution: Gaussian: N(μ,SEM)

Many samples later…



Population distribution



The variances s_x of the sampling distributions become smaller, when n increases

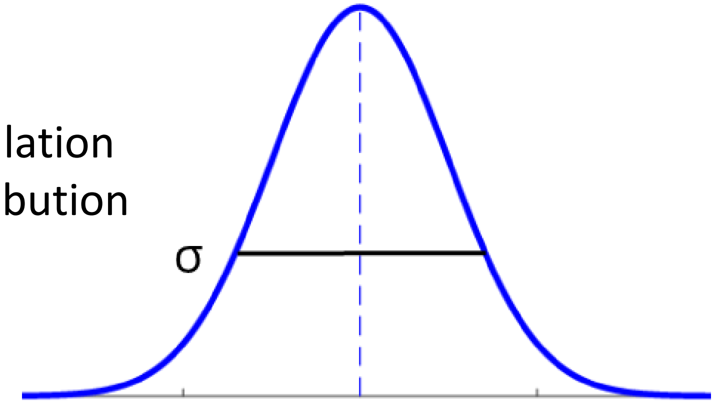
$s_{\bar{x}}$ is a good measure for the sampling error

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

standard error of the mean

Estimates

So much about theory. In practice, we do not know the true means and variance- and we have only one sample. For this reason, we use estimates



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

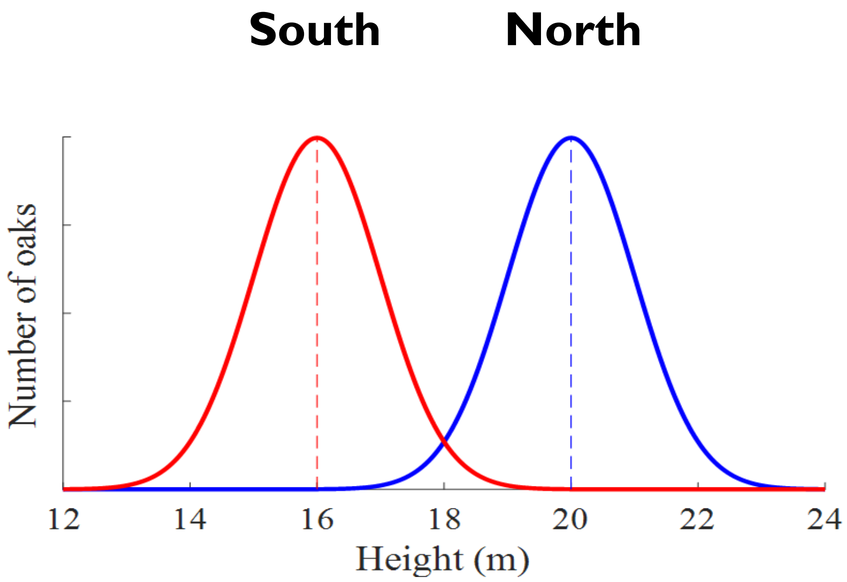
$$\text{Standard deviation: } s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

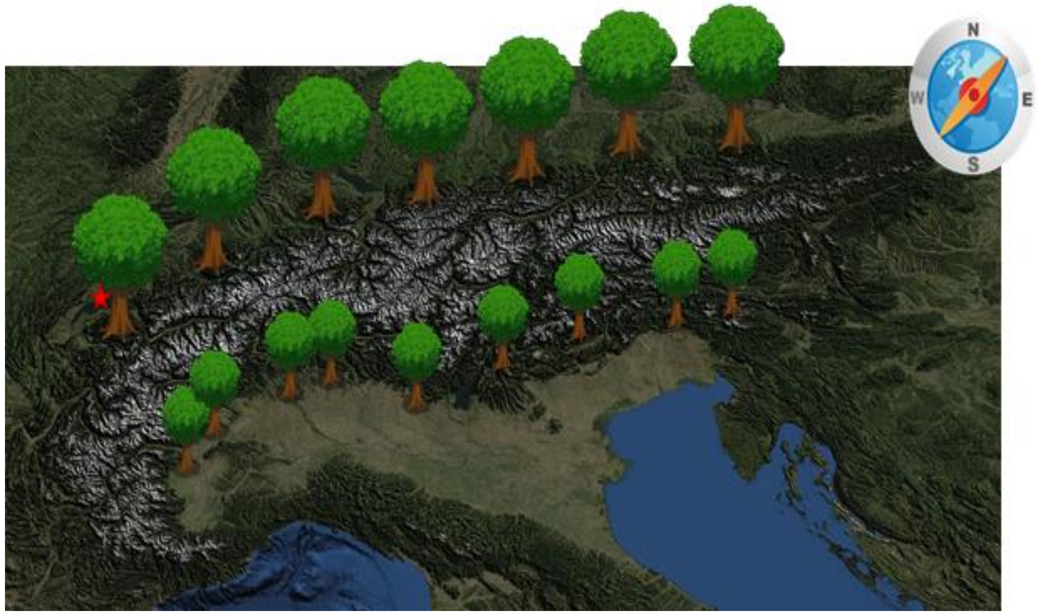
$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$s_{\bar{x}}$ tells us about the "size" of the error for one sample

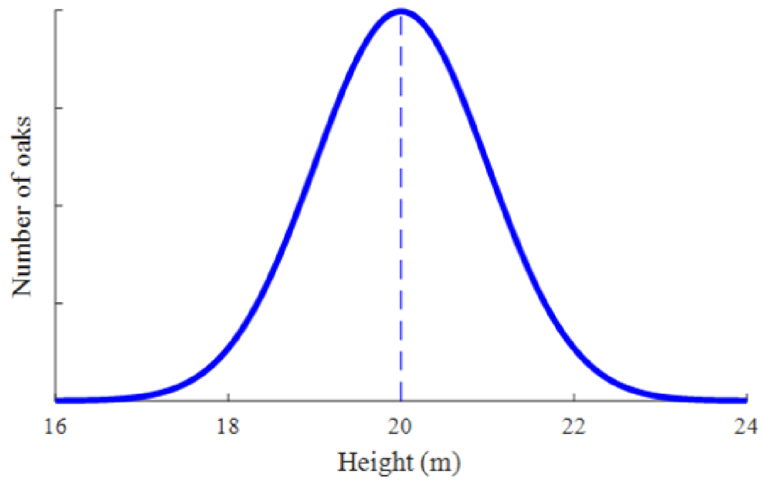
Step 2: Comparing Means

Step 2: Overlap

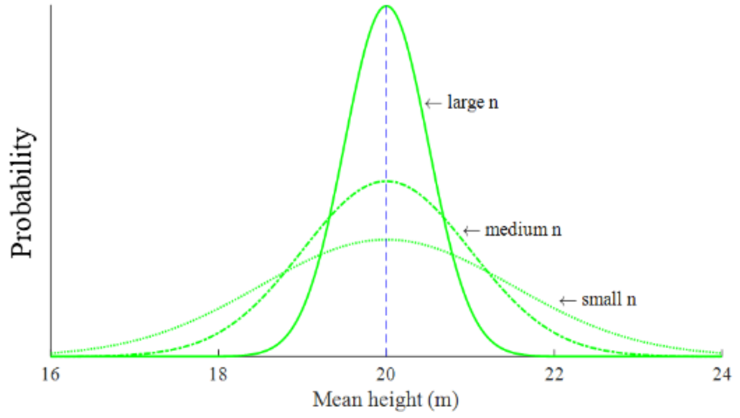




Step 2: Overlap



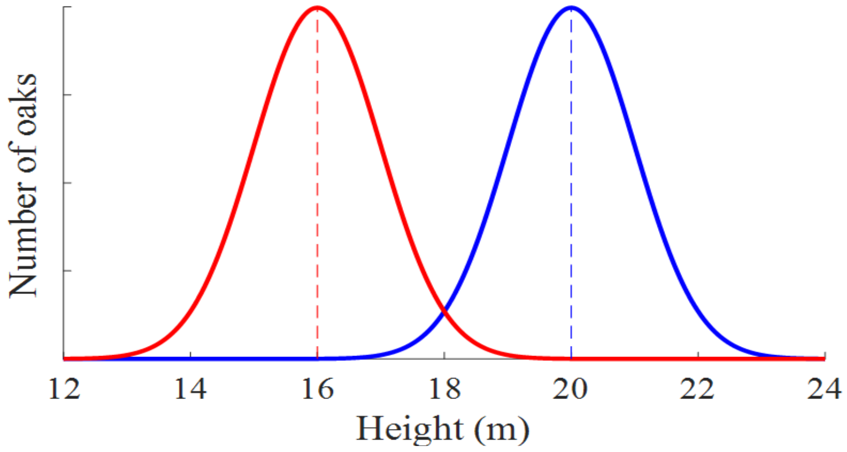
Population distribution



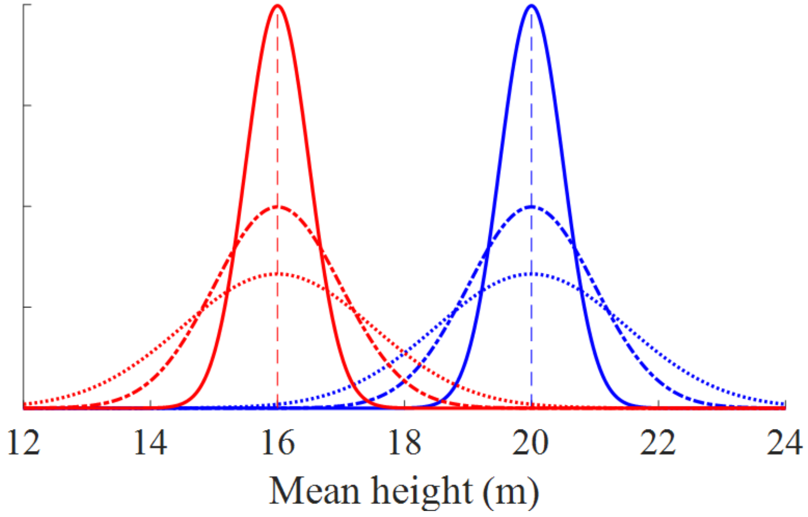
Sampling distribution

Step 2: Overlap

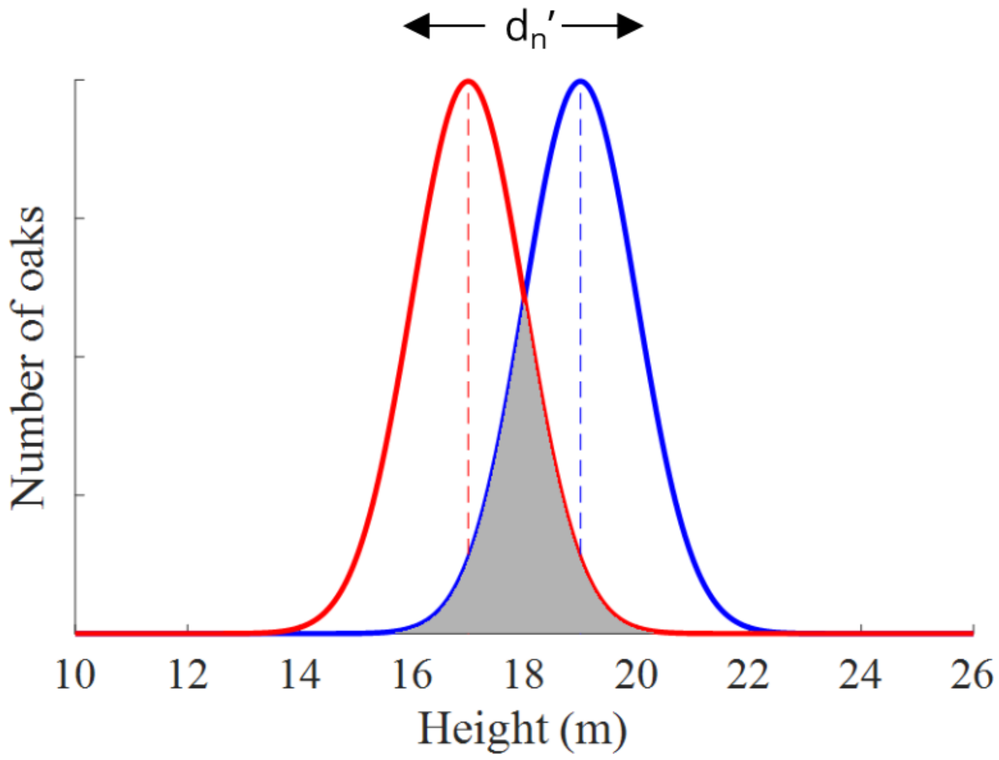
Population distributions



Sampling distributions Families (n)

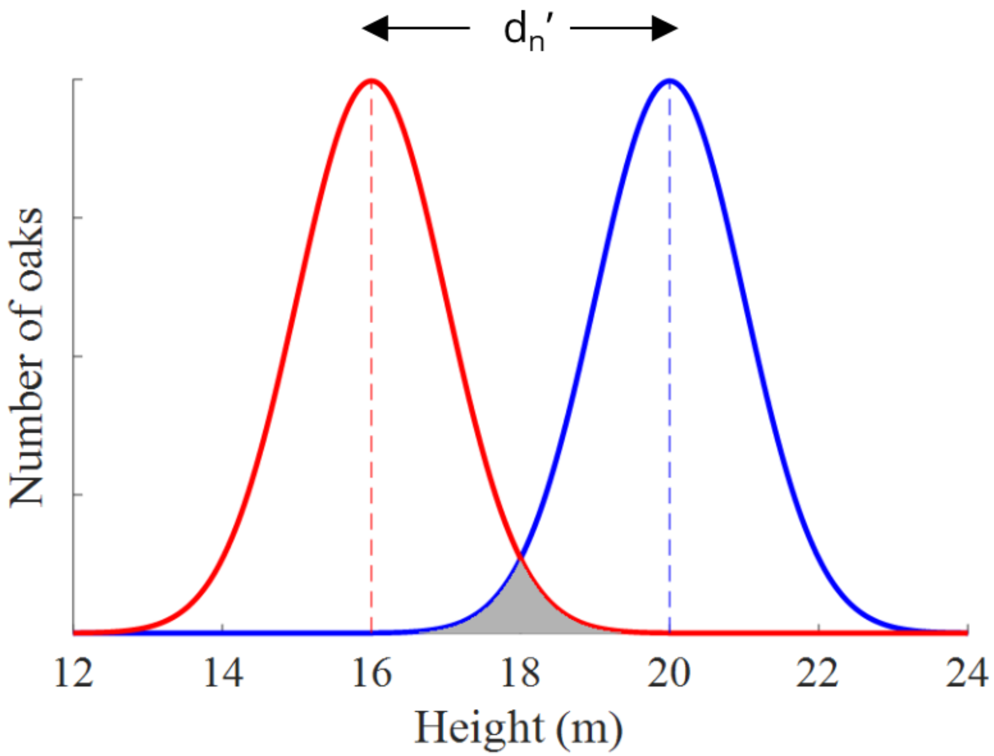


STD for means



Recall our hypothesis:
are the means the same
or different?

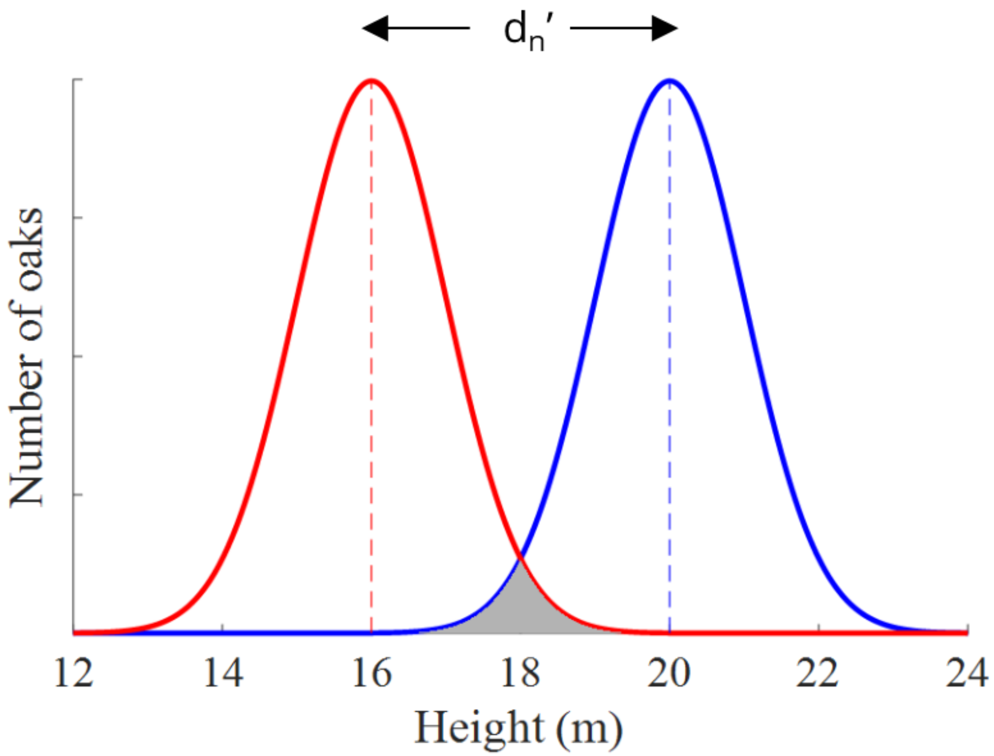
STD for means



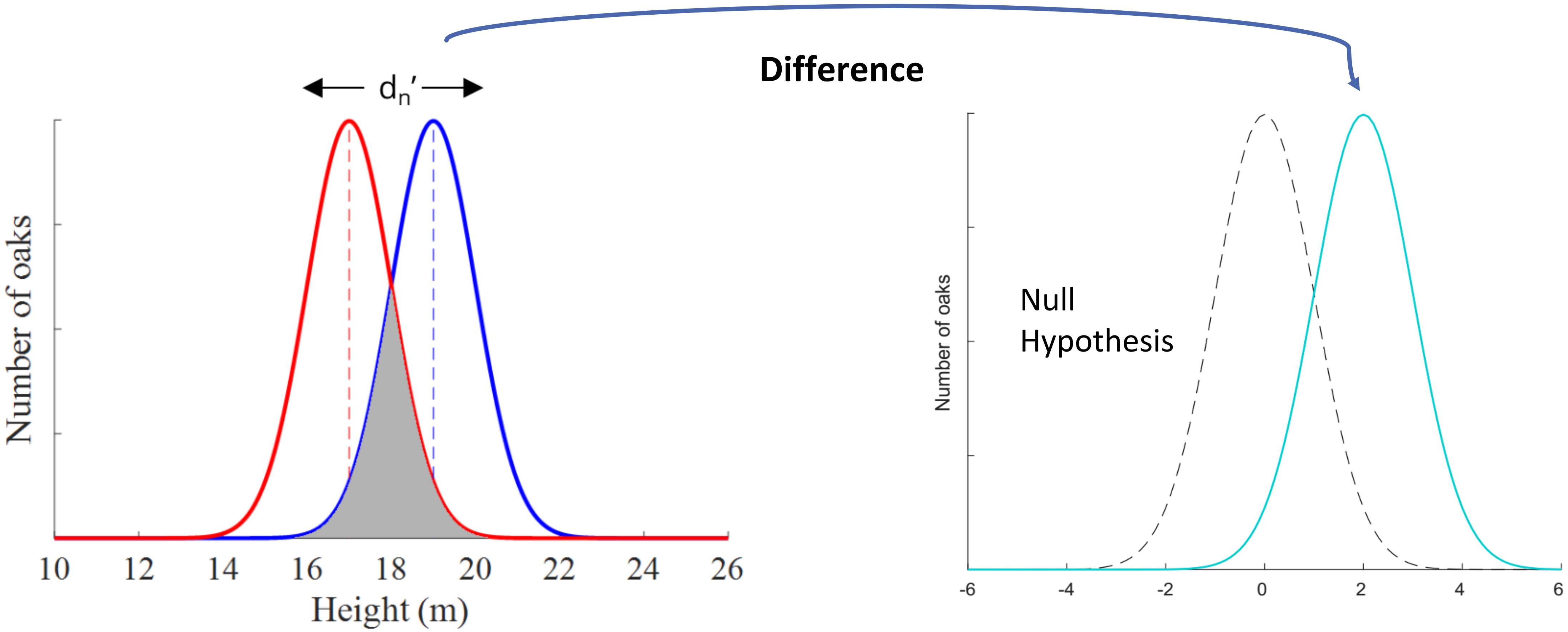
Recall our hypothesis:
are the means the
same or different?

STD for means

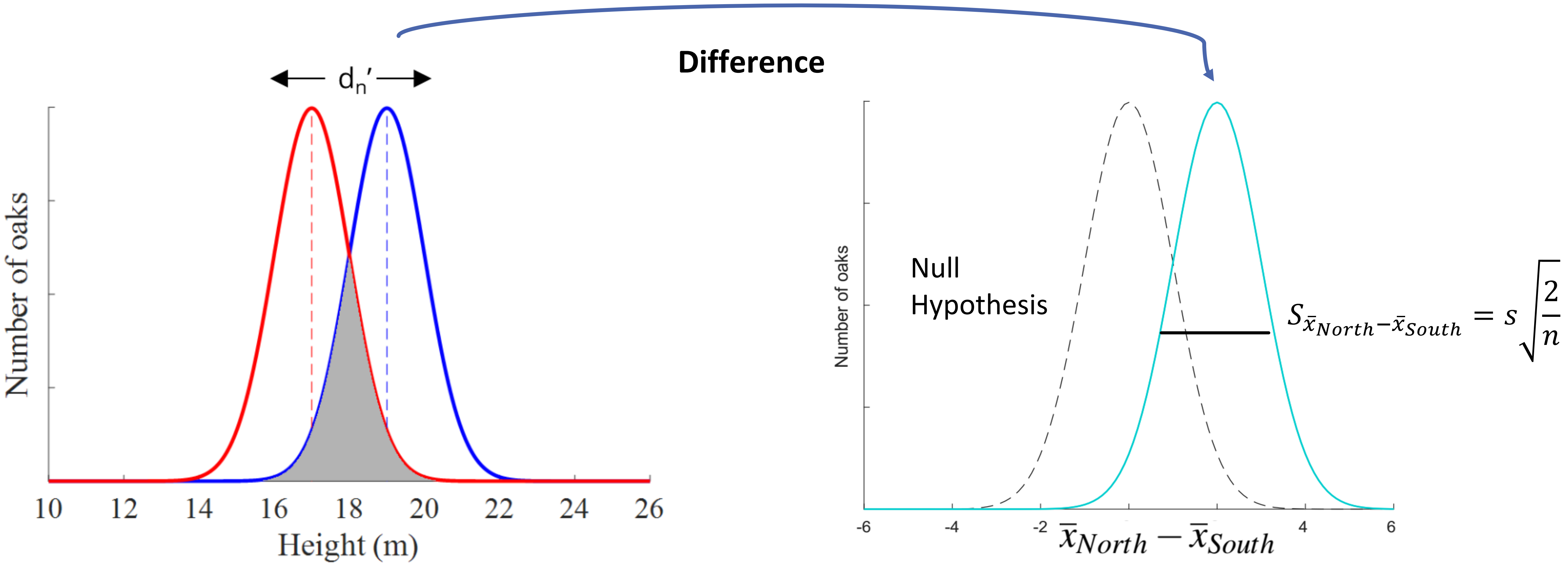
For a given sample of size n, t_n reflects the overlap d'_n



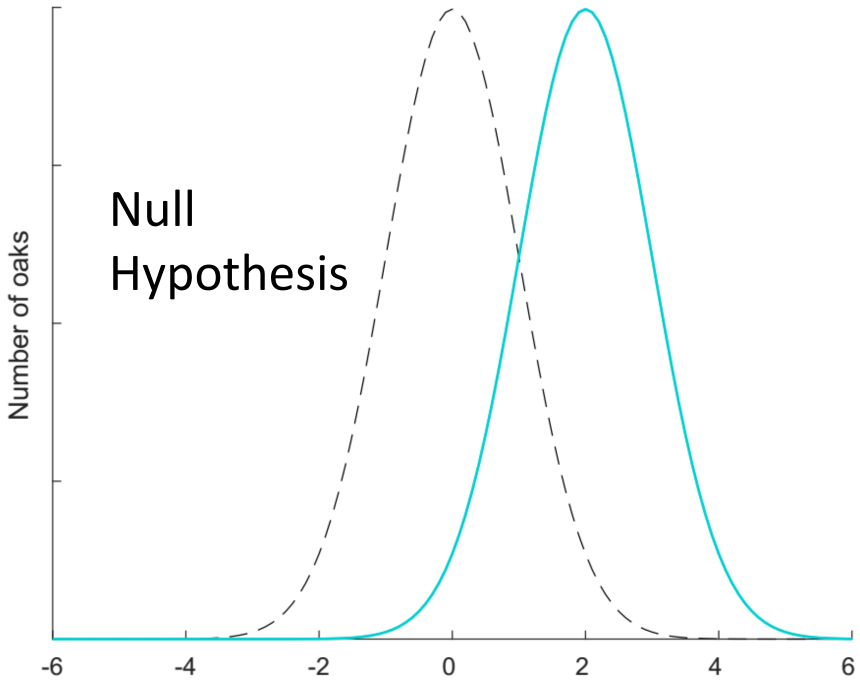
STD for means



STD for means



T-test



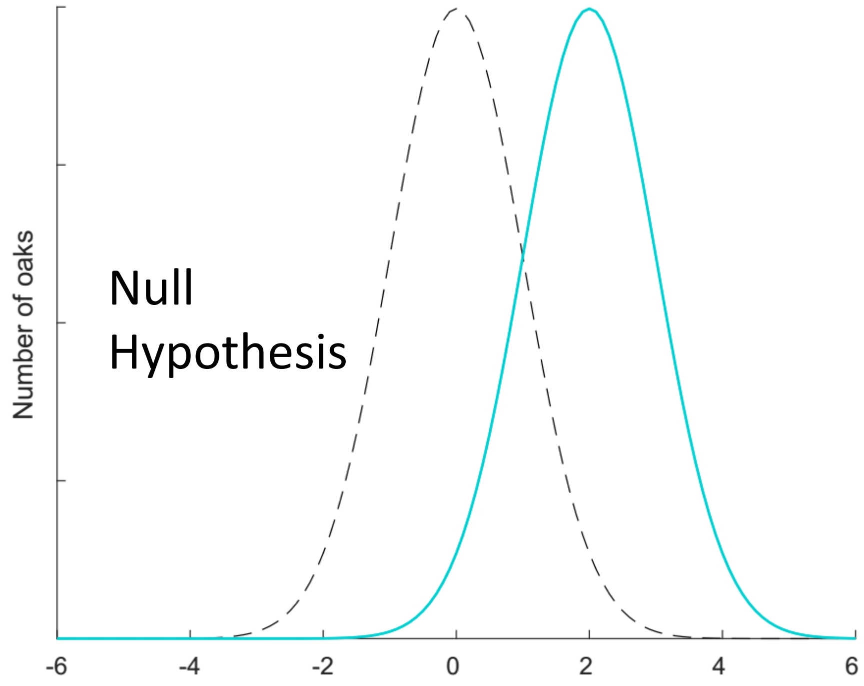
$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{s_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{s\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{s}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

Step 3: The Criterion

Undersampling Causes Errors

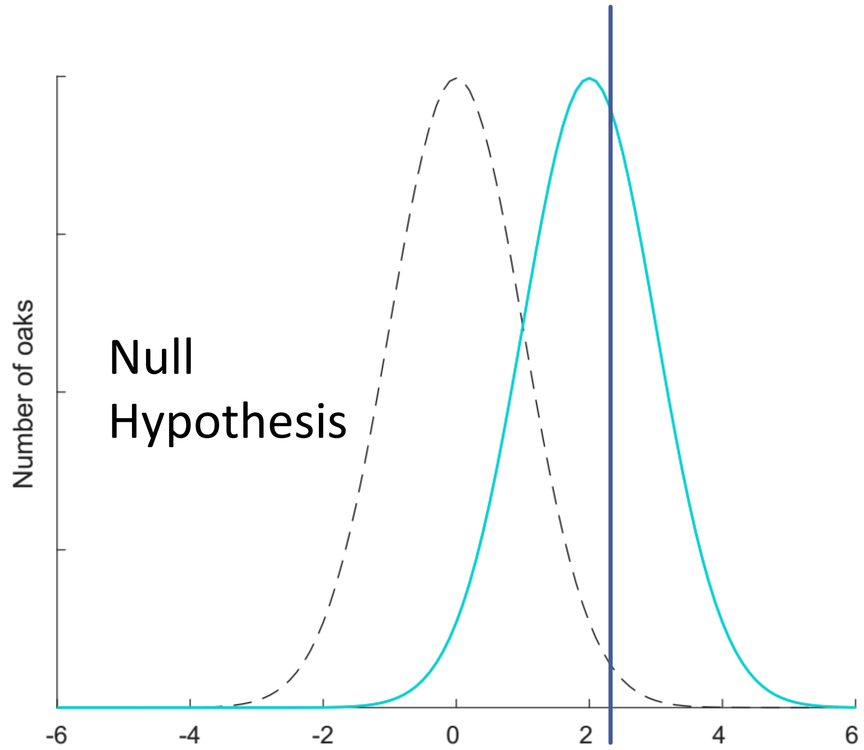
	H_0 is false	H_0 is true
Decide there is a significant difference	Hit	False Alarm (Type I error)
Do not decide there is a significant difference	Miss (Type II error)	Correct Rejection

t-test



$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

t-test

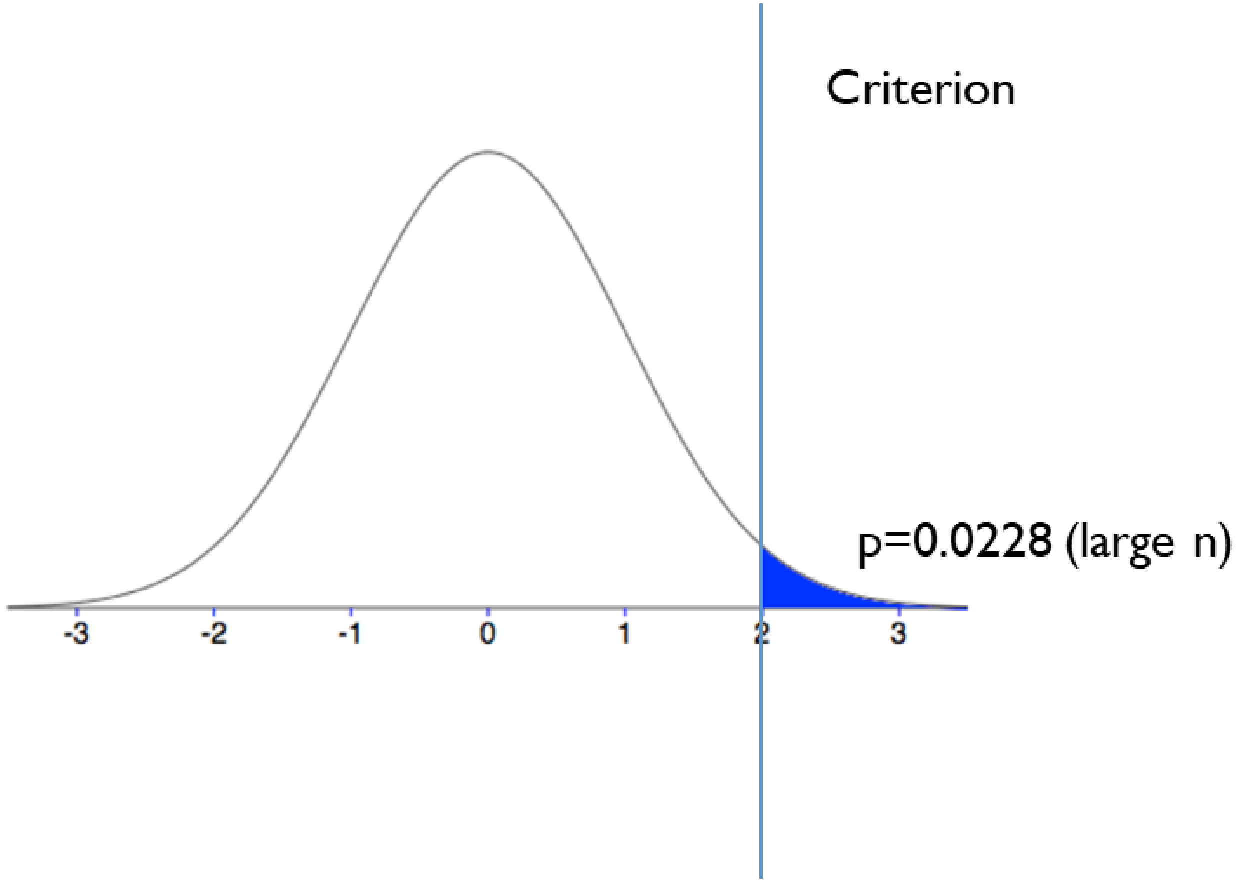


$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{s_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{s\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{s}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

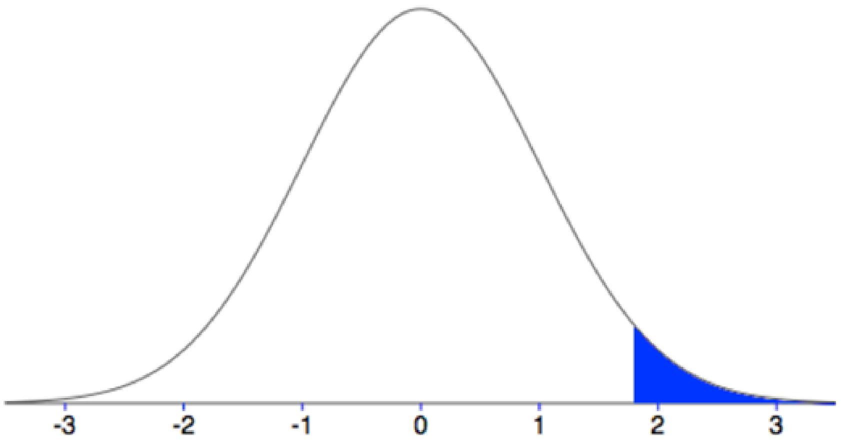
Undersampling Causes Errors

	H_0 is false	H_0 is true
Decide there is a significant difference	Hit	False Alarm (Type I error)
Do not decide there is a significant difference	Miss (Type II error)	Correct Rejection

Type I Error, one tailed t-test



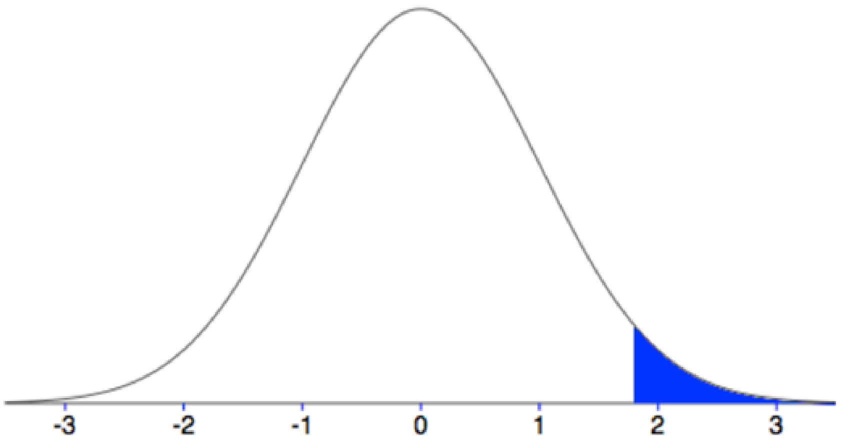
Type I Error, one tailed t-test



Type I Error, one tailed t-test

$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}}$$

$$t = 1.8$$

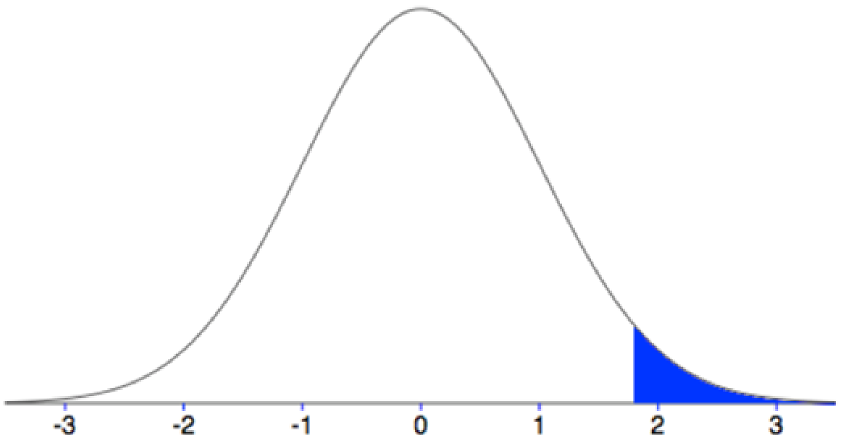


p=0.0359
(dependent n)

Type I Error, one tailed t-test

$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}}$$

$$t = 1.8$$



p=0.0359
(dependent on n)

This is the p-value!

Type I Error, one tailed t-test

Choose your Type I Error Rate, e.g. 5%

Two sample t-test

$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

Significant: $t > 1.96\ (depending\ on\ n)$

Two sample t-test

$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

Significant: d $* \sqrt{n/2} > 1.96$ $(depending\ on\ n)$

Lesson #1

$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

The t-test confounds effect and sample size:

Partial information!

Question 24 Which of the following statements is (are) correct about the p-value?

- The p-value is partial information
- It gives you the probability that the result observed is due to chance
- 1-p is power
- None of them
- A non-significant p-value tells you that there is evidence for H0

Exam

Question 24 Which of the following statements is (are) correct about the p-value?

- [x] The p-value is partial information
- [ ] It gives you the probability that the result observed is due to chance
- [ ] 1-p is power
- [ ] None of them
- [ ] A non-significant p-value tells you that there is evidence for H0

Question 7 Which of the equations below is the equation for the standard error of the mean with standard deviation σ and sample size n?

- $\frac{\sigma}{\sqrt{n-1}}$
- $\sigma\sqrt{\frac{n-1}{n}}$
- $\sigma\sqrt{\frac{2}{n}}$
- None of them
- $\frac{\sigma}{\sqrt{n}}$
- $\sigma\sqrt{\frac{1}{n(n-1)}}$

Question 7 Which of the equations below is the equation for the standard error of the mean with standard deviation σ and sample size n?

- [ ] $\frac{\sigma}{\sqrt{n-1}}$
- [ ] $\sigma\sqrt{\frac{n-1}{n}}$
- [ ] $\sigma\sqrt{\frac{2}{n}}$
- [ ] None of them
- [x] $\frac{\sigma}{\sqrt{n}}$
- [ ] $\sigma\sqrt{\frac{1}{n(n-1)}}$

Take home messages

EPFL

Take Home Messages

- Since the p-value is determined by the t-value, it confounds effect size (d) and sample size (n). The original idea behind the t-test was to provide tools to understand to what extent a significant result is a matter of random sampling, given a certain effect size d. Nowadays, the p-value is often mistaken as a measure of effect size, which was never intended and is simply wrong!
- Partial information: proper conclusions can only be based on both the estimated population effect size, d, and the sample size, n. Hence, it is important to report both values, to take both values into account for conclusions, and to understand whether a significant result is driven by the estimated effect size d, the sample size, or both.

END Class 3a

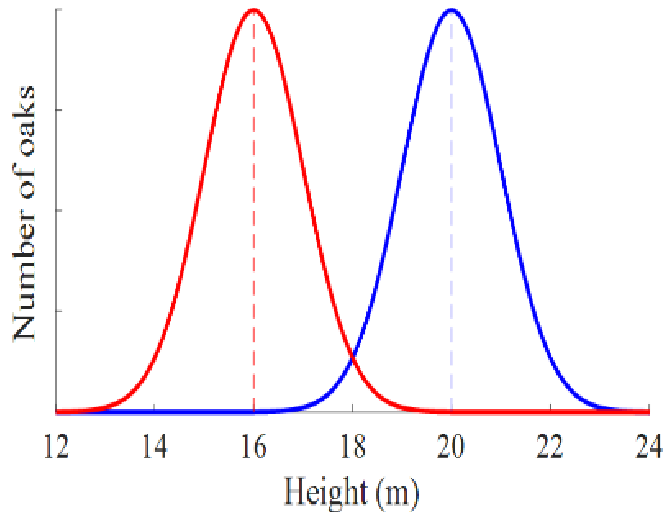
Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I and II
4. Statistics in a nutshell
5. Multiple Testing
6. ANOVA
7. Experimental Design & Statistics
8. Correlations & PCA
9. Meta-Statistics: Basics
10. Meta-Statistics: Too good to be true
11. Meta-Statistics: How big a problem is publication bias?
12. Meta-Statistics: What do we do now?

Summary of the transformations

Summary of the transformations

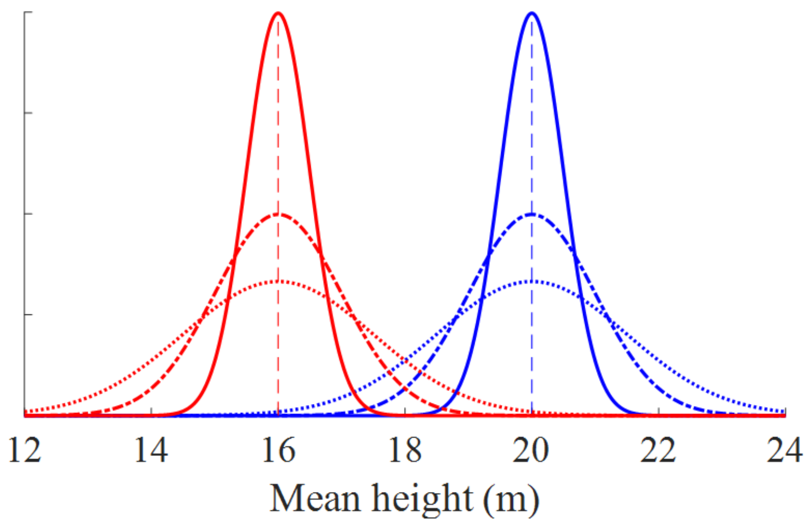
Population distributions



$$d = \frac{\bar{x}_{North} - \bar{x}_{South}}{s}$$

Sampling distributions

$$s_{\bar{x}} = \frac{s}{\sqrt{n/2}}$$

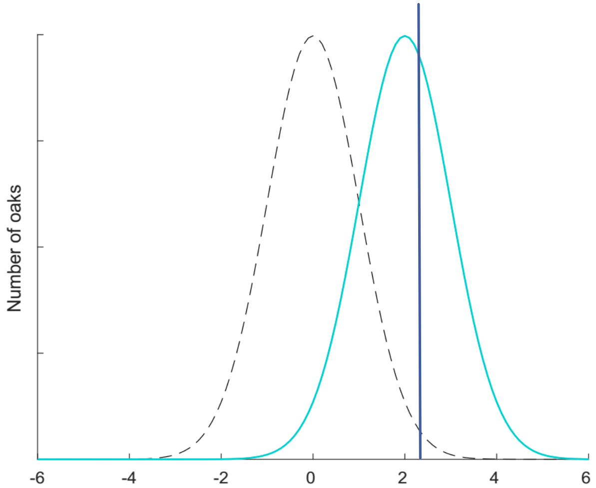


$$t = \frac{(\bar{x}_{North} - \bar{x}_{South})}{\frac{s}{\sqrt{n/2}}}$$

$$t = d * \sqrt{n/2}$$

Null + Difference distribution

$$s_{\bar{x}_{North} - \bar{x}_{South}} = s\sqrt{\frac{2}{n}}$$



$$p = (data|H_0)$$

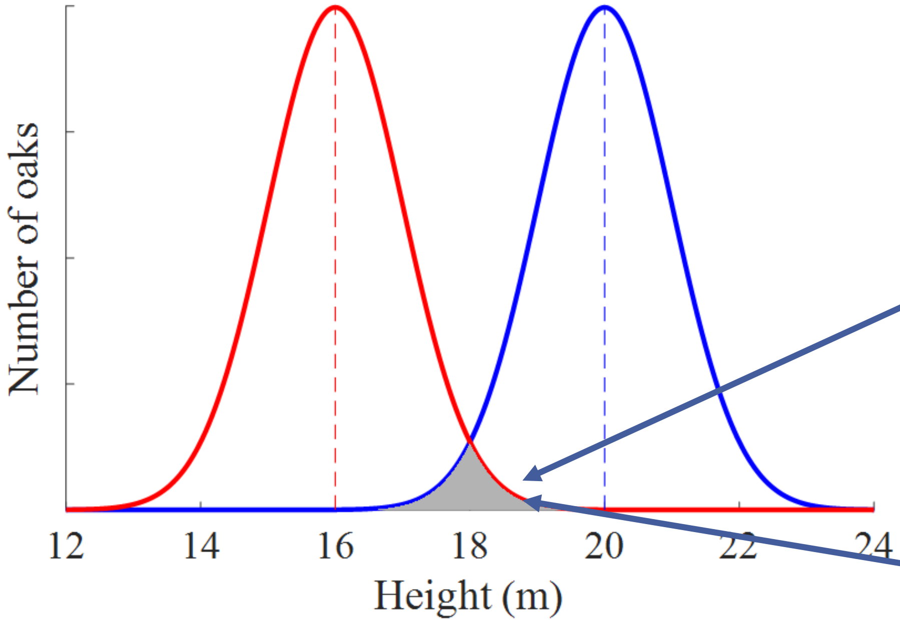
Type II Error

	H_0 is false	H_0 is true
Decide there is a significant difference	Hit	False Alarm (Type I error)
Do not decide there is a significant difference	Miss (Type II error)	Correct Rejection

$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

The t-test confuses effect and sample size: Partial information!

T-Distribution Critical Values



df = n - 1

TABLE B: *t*-DISTRIBUTION CRITICAL VALUES

	Tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level *C*											

Example

Example

North	$(x_{i,North} - \overline{x}_{North})^2$
20.80	$(20.80 - 18.00)^2 = 7.840$
17.81	0.036
17.92	0.006
18.30	0.090
15.17	8.009
$\overline{x}_{North} = \sum_{i=1}^{n} \frac{x_{i,North}}{n} = 18.00$	$s^2_{North} = \frac{\sum_{i=1}^{n} (x_{i,North} - \overline{x}_{North})^2}{n-1} = 3.995$

South	$(x_{i,South} - \overline{x}_{South})^2$
16.91	3.648
15.28	0.078
13.70	1.690
16.81	3.276
12.30	7.290
$\overline{x}_{South} = \sum_{i=1}^{n} \frac{x_{i,South}}{n} = 15.00$	$s^2_{South} = \frac{\sum_{i=1}^{n} (x_{i,South} - \overline{x}_{South})^2}{n-1} = 3.996$

$$df = (5-1) + (5-1) = 4 + 4 = 8$$

$$s_p^2 = \frac{s^2_{North}(n-1) + s^2_{South}(n-1)}{(n-1) + (n-1)}$$

$$= \frac{3.995(4) + 3.996(4)}{4+4} = 3.996$$

$$s_{\overline{x}_{North} - \overline{x}_{South}} = \sqrt{s_p^2\left(\frac{1}{n} + \frac{1}{n}\right)}$$

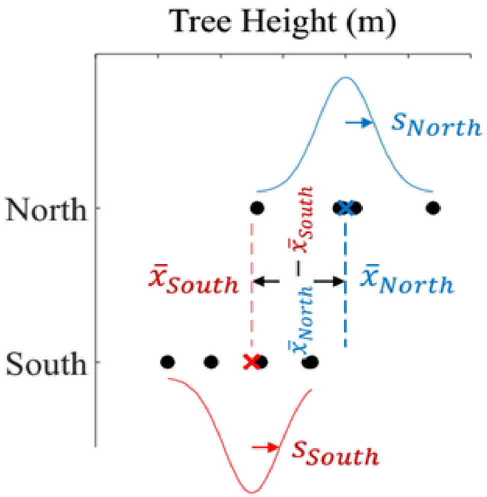
$$= \sqrt{3.996\left(\frac{1}{5} + \frac{1}{5}\right)} = 1.264$$

$$t = \frac{\overline{x}_{North} - \overline{x}_{South}}{s_{\overline{x}_{North} - \overline{x}_{South}}} = \frac{18.00 - 15.00}{1.264} = 2.373$$

p-value=0.045 < 0.05

Example

North	$(x_{i,North} - \bar{x}_{North})^2$	South	$(x_{i,South} - \bar{x}_{South})^2$
20.80	$(20.80\text{-}18.00)^2 = 7.840$	16.91	3.648
17.81	0.036	15.28	0.078
17.92	0.006	13.70	1.690
18.30	0.090	16.81	3.276
15.17	8.009	12.30	7.290
$\bar{x}_{North} = \sum_{i=1}^{n_{North}} \frac{x_{i,North}}{n_{North}} = 18.00$	$s^2_{North} = \frac{\sum_{i=1}^{n_{North}}(x_{i,North} - \bar{x}_{North})^2}{n_{North} - 1} = 3.995$	$\bar{x}_{South} = \sum_{i=1}^{n_{South}} \frac{x_{i,South}}{n_{South}} = 15.00$	$s^2_{North} = \frac{\sum_{i=1}^{n_{South}}(x_{i,South} - \bar{x}_{South})^2}{n_{South} - 1} = 3.996$



$$df = (5-1) + (5-1) = 4 + 4 = 8$$

$$s_p^2 = \frac{s^2_{North}(n_{North} - 1) + s^2_{South}(n_{South} - 1)}{(n_{North} - 1) + (n_{South} - 1)} = \frac{3.995(5-1) + 3.996(5-1)}{(5-1)+(5-1)} = 3.996$$

$$s_{\bar{x}_{North} - \bar{x}_{South}} = \sqrt{\frac{s_p^2}{n_{North}} + \frac{s_p^2}{n_{South}}} = \sqrt{\frac{3.996}{5} + \frac{3.996}{5}} = 1.264$$

$$t = \frac{\bar{x}_{North} - \bar{x}_{South}}{s_{\bar{x}_{South} - \bar{x}_{North}}} = \frac{18.00 - 15.00}{1.264} = 2.373$$

p-value=0.045 < 0.05

Example

	t	df	Sig. (2-tailed)	Cohen's d
Tree Height	2.373	8	0.045	1.5 (large effect)

How to present and read statistical values?

$t(2000)=2.21, p=0.047$ (t-test)

$F(1,5)=28.89, p=0.01$ (ANOVA)

$\chi^2 (1, N=226)=6.90, p<0.01$ (χ^2 Test)

Better:

$t(2000)=2.21, p=0.047, d=0.2$

How to present and read statistical values?

EPFL

Am J Med Genet B Neuropsychiatr Genet. 2016 Dec;171(8):1099-1104. doi: 10.1002/ajmg.b.32489. Epub 2016 Aug 17.

GAD1 gene polymorphisms are associated with hyperactivity in Attention-Deficit/Hyperactivity Disorder.

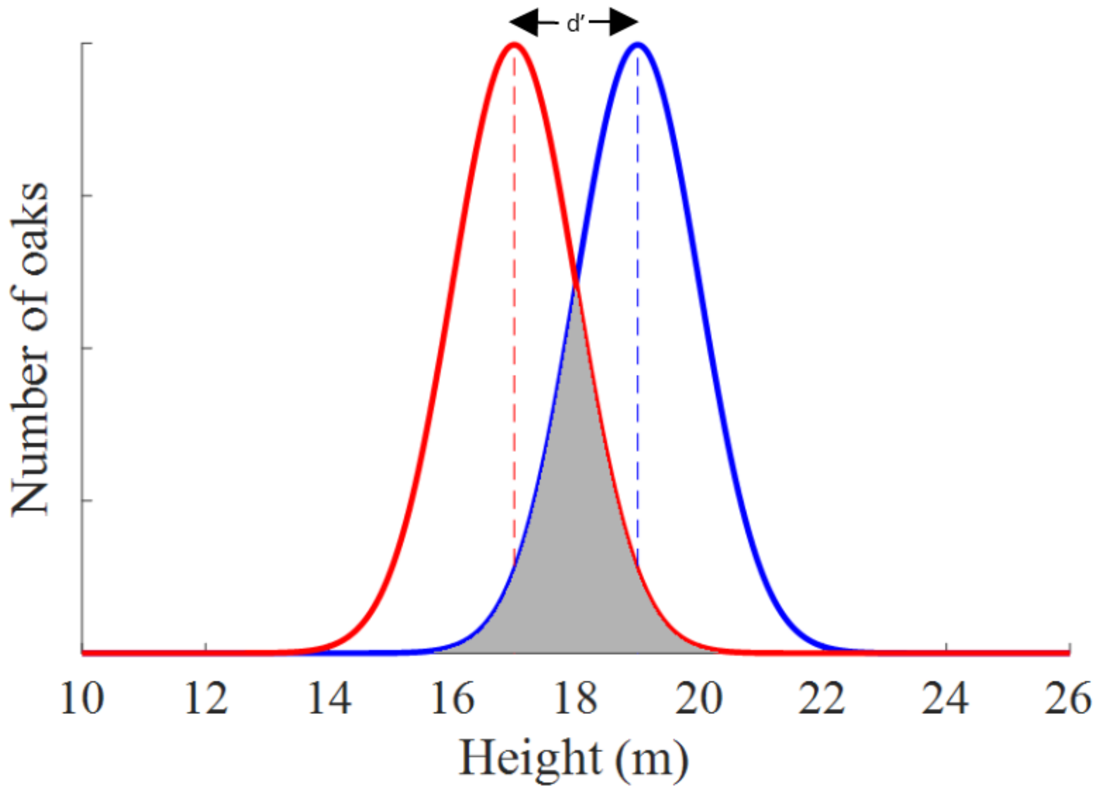
Bruxel EM[1], Akutagava-Martins GC[1], Salatino-Oliveira A[1], Genro JP[1], Zeni CP[2], Polanczyk GV[3], Chazan R[2], Schmitz M[2], Rohde LA[2,3], Hutz MH[1].

⊕ Author information

Abstract

Attention-Deficit/Hyperactivity Disorder (ADHD) is one of the most common neurodevelopmental disorders of childhood. Recent studies suggest a role for γ-aminobutyric acid (GABA) on ADHD hyperactive/impulsive symptoms due to behavioral disinhibition resulting from inappropriate modulation of both glutamatergic and GABAergic signaling. The glutamic acid decarboxylase (GAD1) gene encodes a key enzyme of GABA biosynthesis. The aim of the present study was to investigate the possible influence of GAD1 SNPs rs3749034 and rs11542313 on ADHD susceptibility. The clinical sample consisted of 547 families with ADHD probands recruited at the ADHD Outpatient Clinics from Hospital de Clínicas de Porto Alegre. Hyperactive/impulsive symptoms were evaluated based on parent reports from the Swanson, Nolan, and Pelham Scale-version IV (SNAP-IV). The C allele of rs11542313 was significantly overtransmitted from parents to ADHD probands ($P = 0.02$). Hyperactive/impulsive score was higher in rs3749034G allele ($P = 0.005$, Cohen's $D = 0.19$) and rs11542313C allele ($P = 0.03$; Cohen's $D = 0.16$) carriers. GAD1 haplotypes were also associated with higher hyperactive/impulsive scores in ADHD youths (global P-value = 0.01). In the specific haplotype test, the GC haplotype was the one with the highest hyperactive/impulsive scores ($P = 0.03$). Our results suggest that the GAD1 gene is associated with ADHD susceptibility, contributing particularly to the hyperactive/impulsive symptom domain.

Cohen's d



d'= 0.2 small effect

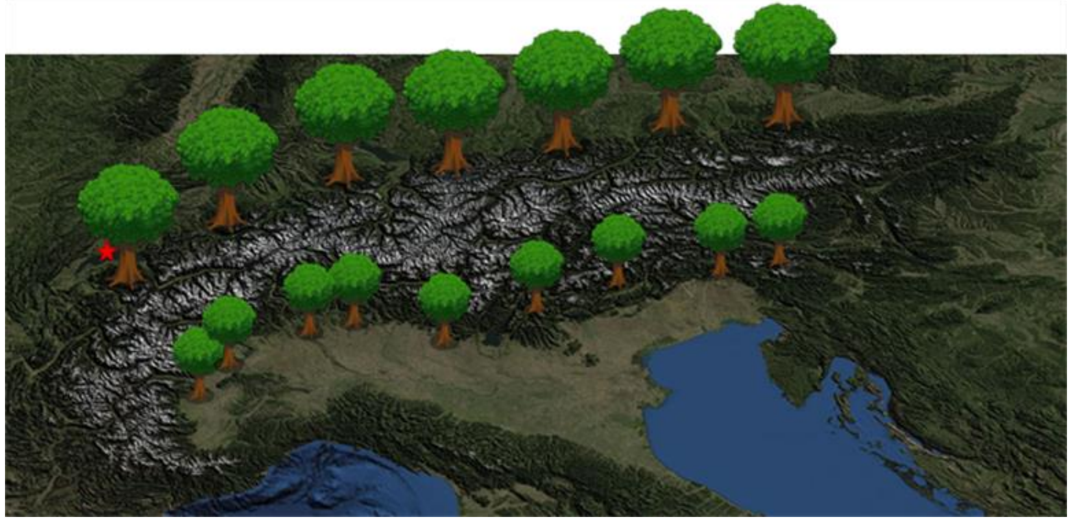
d'= 0.5 medium effect

d'= 0.8 large effect

https://de.wikipedia.org/wiki/Effektst%C3%A4rke

Comments & Extensions

Example

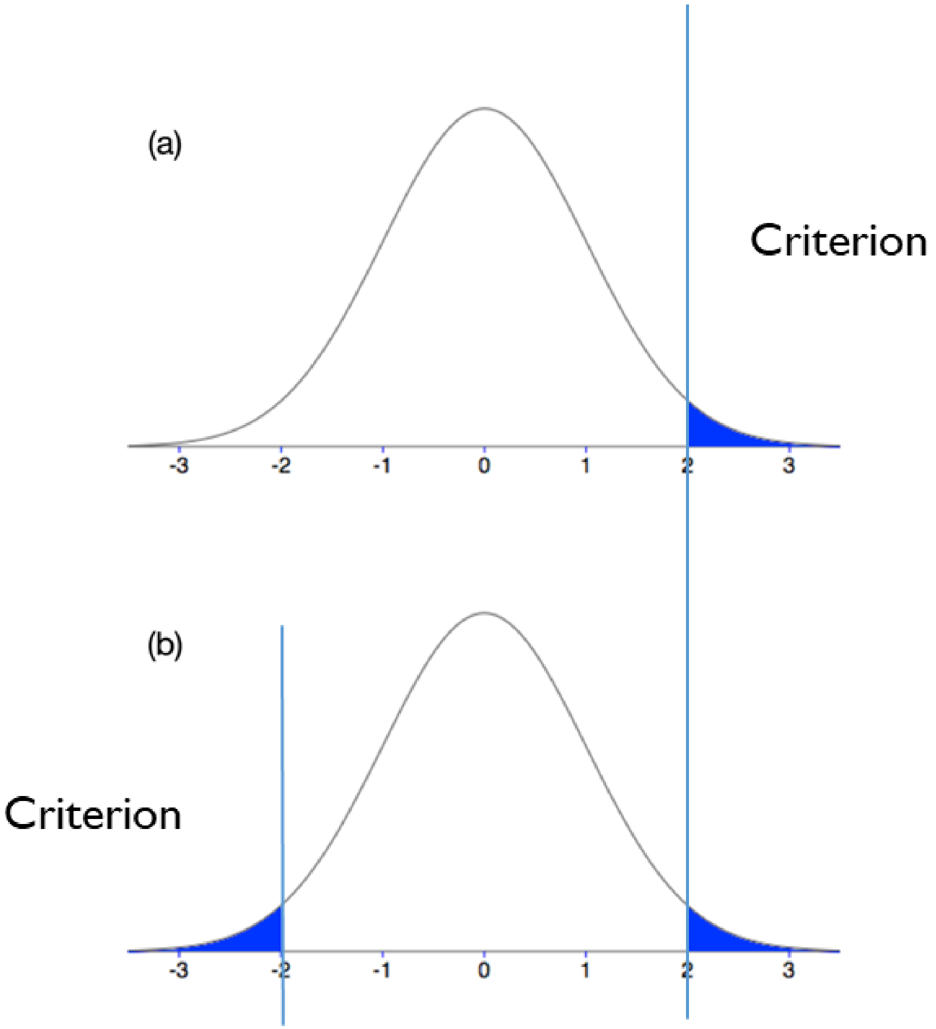


Bergmann's rule states that "populations and species of **larger size** are found in **colder environments**, and species of **smaller size** are found in **warmer regions**" (Wikipedia.org).

Compare mean height of oaks in the North and in the South

μ(North) > μ(South)

Type I Error, two tailed t-test



D' is all over the place

EPFL

Statistic	Description
Cohen's d or Hedges' g	Estimated standardized effect size
t	Test statistic
p	p-value for a two-tailed t test
d_{95}(lower) or g_{95}(lower)	Lower limit of a 95% confidence interval for d or g
d_{95}(upper) or g_{95}(upper)	Upper limit of a 95% confidence interval for d or g
Post hoc power from d or g	Estimated power for experiments with the same sample size
Post hoc v	Proportion of times OLS is more accurate than RLS
Λ	Log likelihood ratio for null and alternative models
ΔAIC, $\Delta AICc$	Difference in AIC for null and alternative models
ΔBIC	Difference in BIC for null and alternative models
JZS BF	Bayes Factor based on the Jeffreys-Zellner-Siow prior

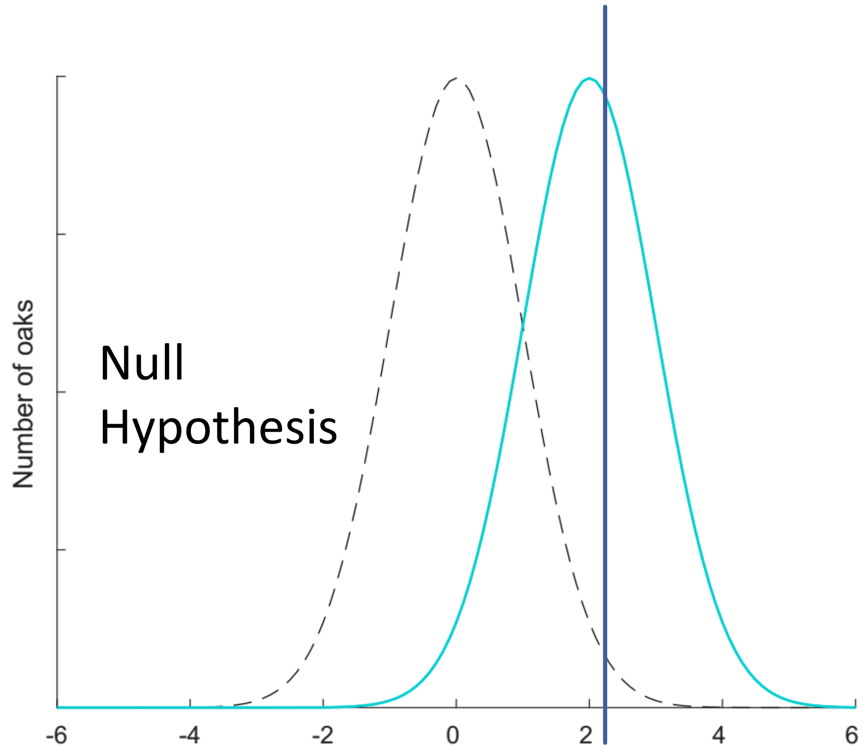
Table 1: For known sample sizes of a independent two-sample t-test, each these terms is an equivalent sufficient statistic for the standardized effect size of the population. For given sample sizes, it is possible to transform any value to all the others. OLS – ordinary least square; RLS – randomized least squares.

Step 4: Type II Error

Type II Error

	H_0 is false	H_0 is true
Decide there is a significant difference	Hit	False Alarm (Type I error)
Do not decide there is a significant difference	Miss (Type II error)	Correct Rejection

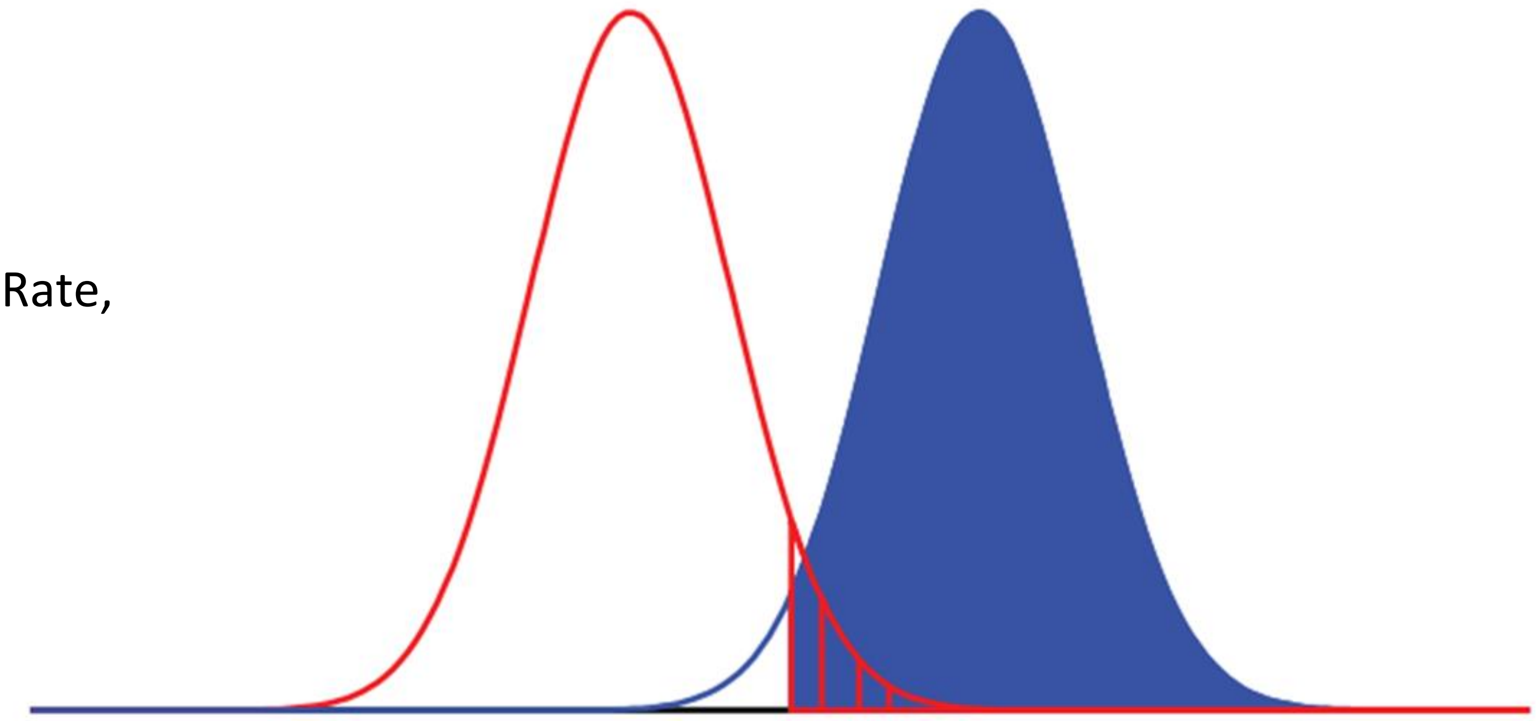
t-test



$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

Type II Error

How to compute the Miss Rate,
we will see in lesson 4



Highlights

Highlights

- Since the p-value is determined by the t-value, it confounds effect size (d) and sample size (n). The original idea behind the t-test was to provide tools to understand to what extent a significant result is a matter of random sampling, given a certain effect size d. Nowadays, the p-value is often mistaken as a measure of effect size, which was never intended and is simply wrong!

- Partial information: proper conclusions can only be based on both the estimated population effect size, d, and the sample size, n. Hence, it is important to report both values, to take both values into account for conclusions, and to understand whether a significant result is driven by the estimated effect size d, the sample size, or both.

Implications

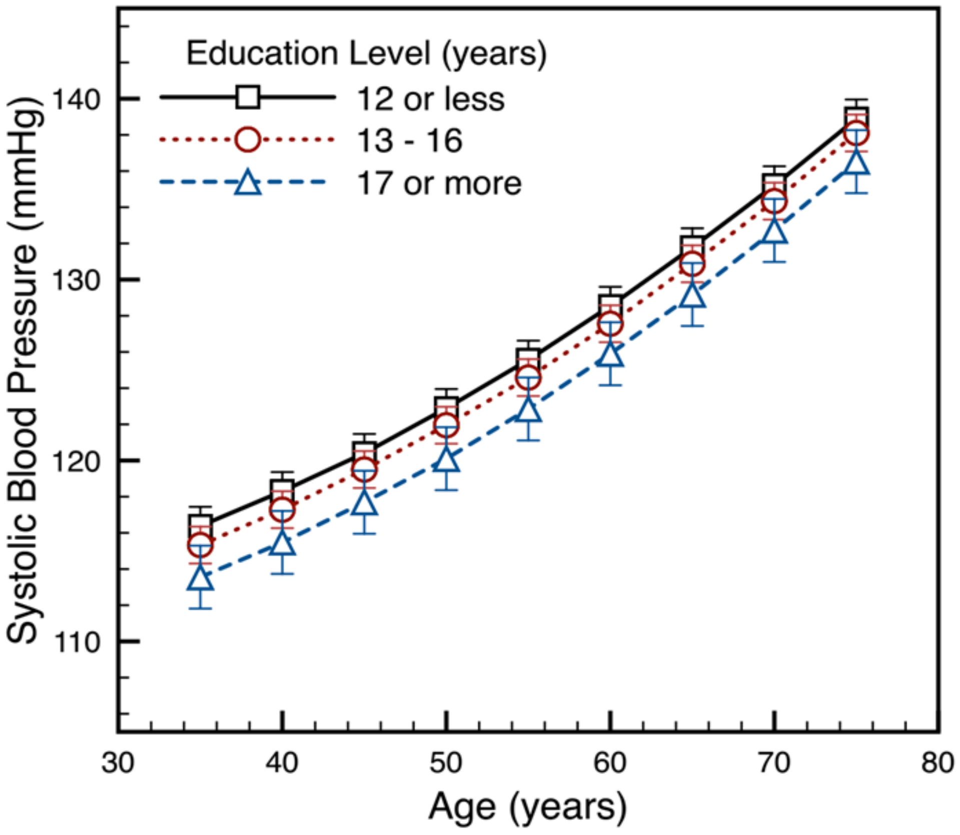
Implications I: Sample size

Implication Ia: for each $d \neq 0$, the t-test is significant if *n* is sufficiently large

$$t = d * \sqrt{n/2}$$

Significance: $d * \sqrt{n/2} > 1.96 \ (depending\ on\ n)$

Framingham Studies



Associations of education with 30 year life course blood pressure trajectories: Framingham Offspring Study
Eric B Loucks, Michal Abrahamowicz, Yongling Xiao & John W Lynch
BMC Public Health **volume 11**, Article number: 139 (2011)

Quiz

Does alcohol make you smart?

Japanese researchers found that moderate consumption of alcohol can positively influence human mental capabilities. Men, who consume less than half liter of sake wine per day, had IQs 3.3 points higher compared to men who never drink alcohol. For women a value of 2.5 points was found.

The study tested 2000 humans in the age from 40 to 79 years. The researchers propose that alcohol protects against arterial sclerosis and thus provides a better blood supply to the brain.

Adapted from Spiegel Online, 7.12.2000

Implications I

Implication Ib: For each $d \neq 0$, the t-test is significant for a sufficiently large n_1 and non-significant for a smaller n_2

Is the effect significant? Sounds like a paradox but it is not........

Implications I

Implication Ib: For each $d \neq 0$, the t-test is significant for a sufficiently large n_1 and non-significant for a smaller n_2

Is the effect significant? Sounds like a paradox but it is not........

	H_0 is false	H_0 is true
Decide there is a significant difference	Hit	False Alarm (Type I error)
Do not decide there is a significant difference	Miss (Type II error)	Correct Rejection

Implications I

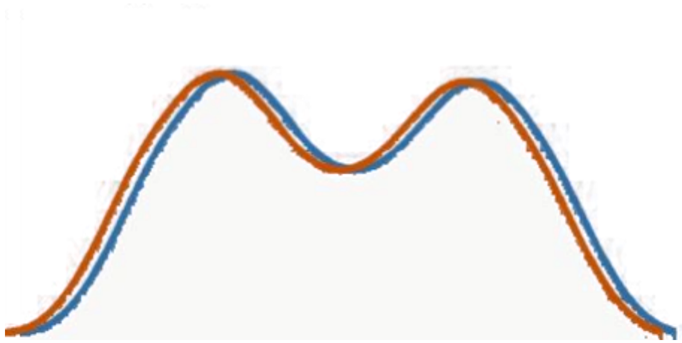
Implication Ic: Aren‘t two experimental conditions always be different, hence, $d' \neq 0$ in general?

We do not need to do experiments! We just need to *assume* we did the experiment with a large n*!*

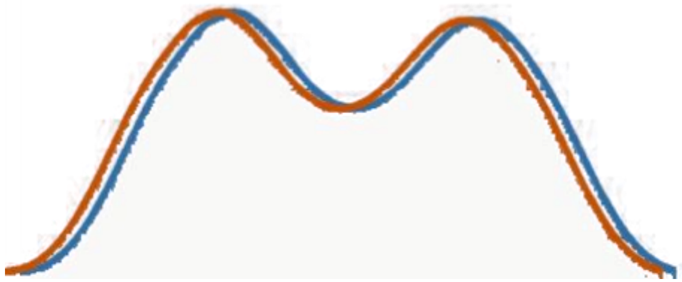
$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

Two Sample t-test

Question: when is $d' = 0$?



Question: when is $d' = 0$?



Part III: optional stopping

Implications II: Effect size

Implication IIa: Significance tells nothing about effect size: Partial Information!

Example: If fish oil were prolonging life by 2min, so what?

Implications II

Implication IIb: The p-value does not tell about effect size. A low p-value does not tell you that there is a strong effect. The p-value depends on n.

The p-value was never intended to be a measure of the effect size. Statistics was developed to understand how many samples are needed for proper conclusions, given a fixed effect size.

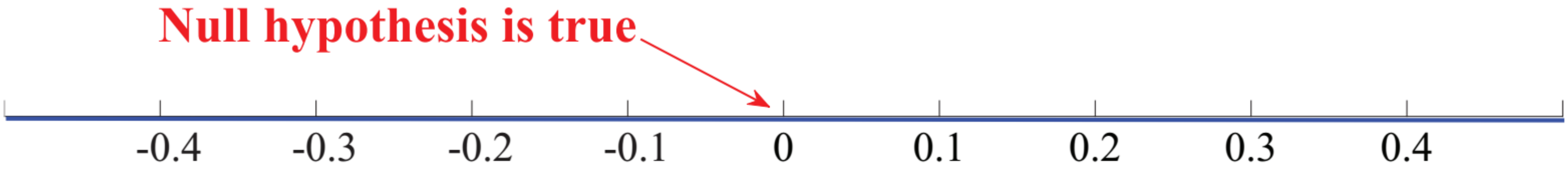
Comparing p-values across studies is non-sense (unless n is the same).

Implications II

Implication IIc: When a study with a small n becomes significant: the effect size must be large (or there is something wrong); the opposite is not true.

Implications II

Implication IId: The Null and the Alternative hypotheses



$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}} * \sqrt{n}$$

Implications III: Null Results

Implication IIIa: Null result: either the Hypothesis is wrong or power is too low. One can never prove the Null Hypothesis: Absence of Proof is not Proof of Absence.

	H_0 is false	H_0 is true
Decide there is a significant difference	Hit	False Alarm (Type I error)
Do not decide there is a significant difference	Miss (Type II error)	Correct Rejection

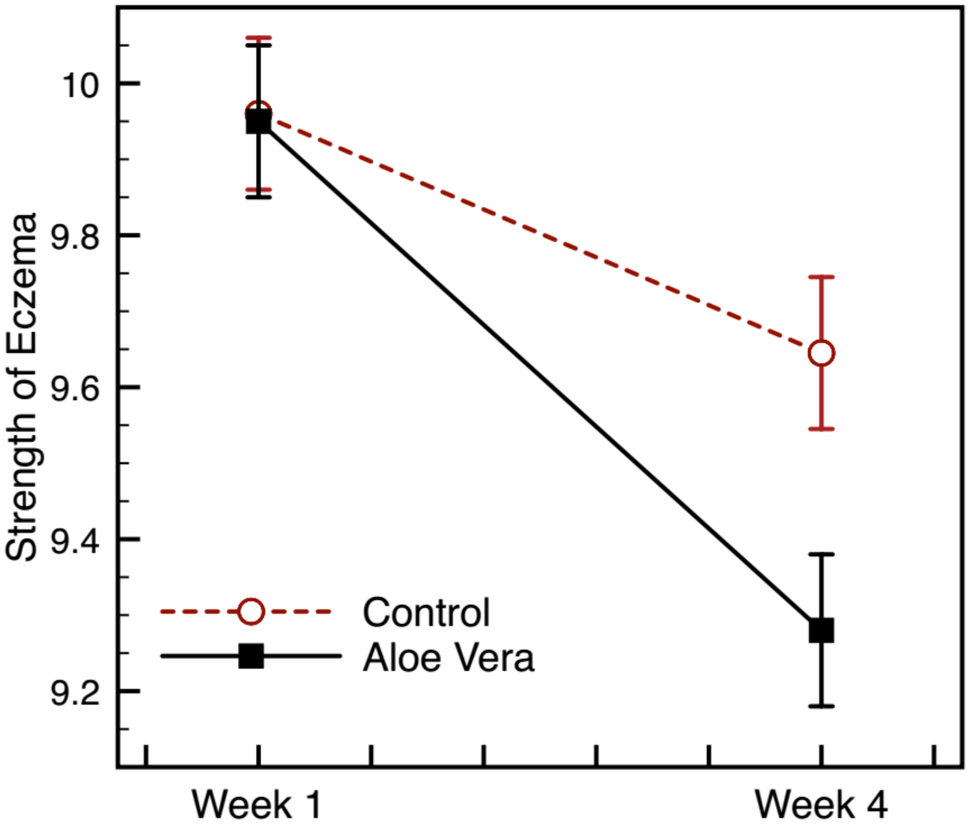
Implications III: Null Results

Implication IIIa: Null result: either the Hypothesis is wrong or power is too low. One can never prove the Null Hypothesis: Absence of Proof is not Proof of Absence.

How to create a Null result? n=1.

$$t = \frac{(\bar{x}_{North} - \bar{x}_{South}) - (0)}{S_{\bar{x}_{North} - \bar{x}_{South}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S\sqrt{\frac{2}{n}}} = \frac{(\bar{x}_{North} - \bar{x}_{South})}{S}\sqrt{\frac{n}{2}} = d\sqrt{\frac{n}{2}}$$

Implications III: Null Results



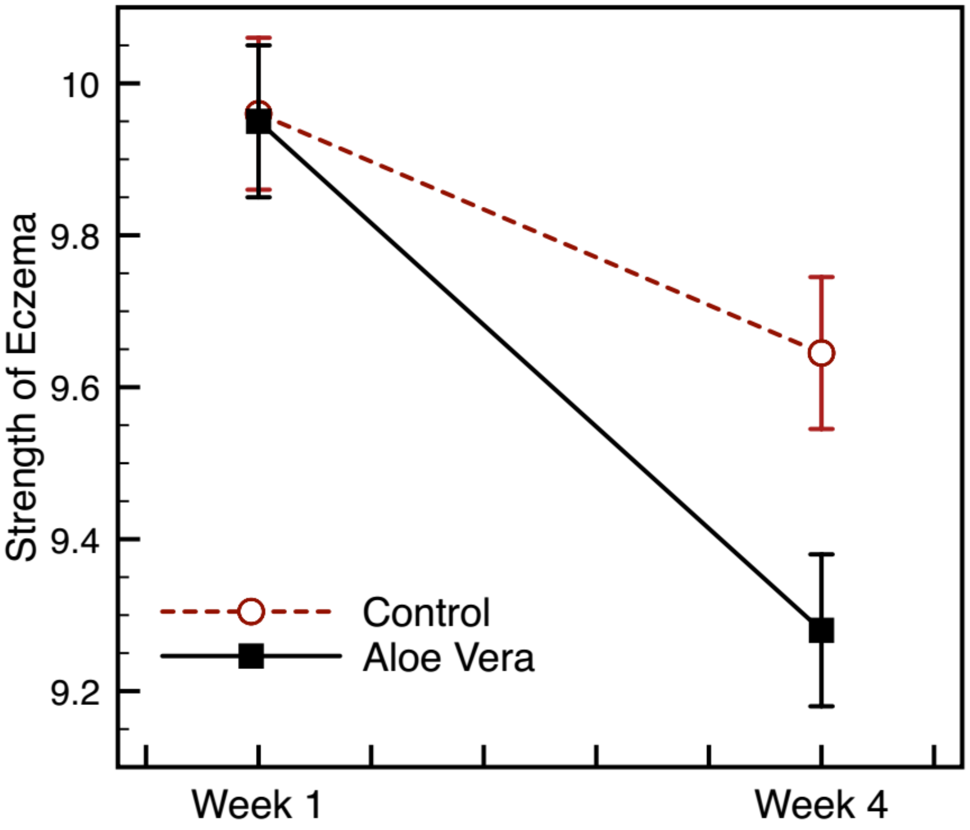
In a placebo-controlled randomized trial of DHA oil for eczema, researchers found a statistically significant improvement in the DHA group but not the placebo group.

The abstract reports: “DHA, but not the control treatment, resulted in a significant clinical improvement of atopic eczema.”

Koch C, Dölle S, Metzger M, et al. Docosahexaenoic acid (DHA) supplementation in atopic eczema: a randomized, double-blind, controlled trial. Br J Dermatol 2008;158:786-792.

Implications III: Null Results

Implication IIIb: A difference of significance is not a significant difference!



The improvement in the DHA group (18%) is not significantly greater than the improvement in the control group (11%).

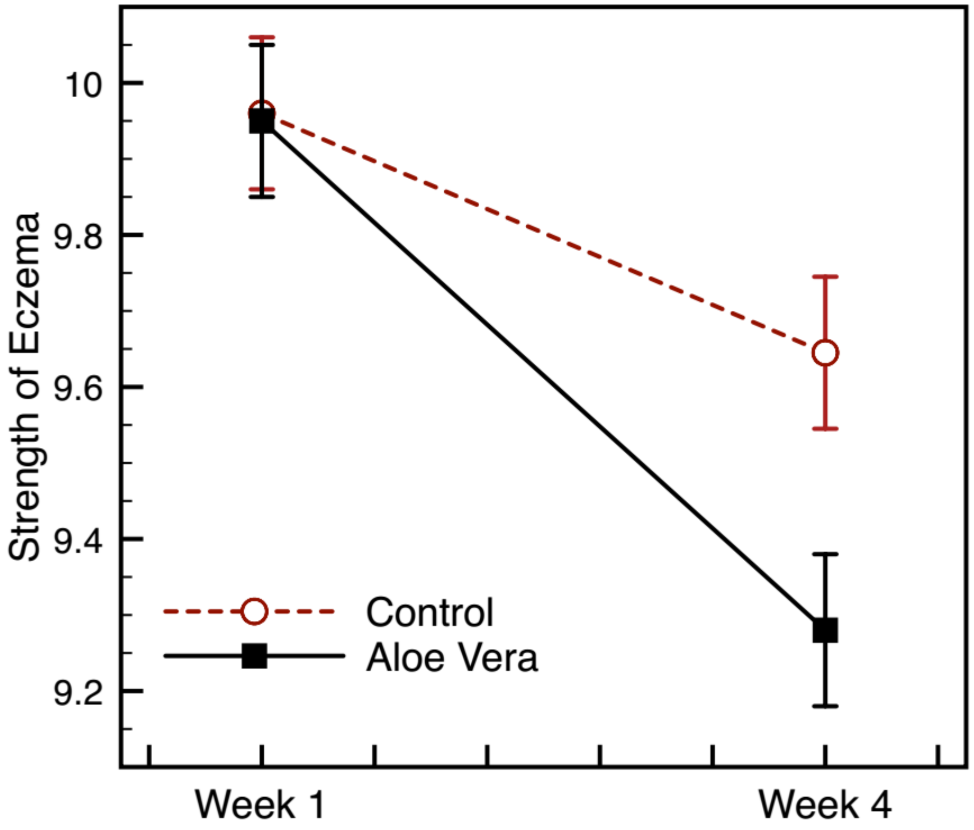
Koch C, Dölle S, Metzger M, et al. Docosahexaenoic acid (DHA) supplementation in atopic eczema: a randomized, double-blind, controlled trial. Br J Dermatol 2008;158:786-792.

Implications III: Null Results

Implication IIIb: A difference of significance is not a significant difference!

This is the same situation whenever a control group is used. It makes usually very little sense to state: “there was an effect in condition A but not B”.

Never ever compare a significant result with a Null result…….



Koch C, Dölle S, Metzger M, et al. Docosahexaenoic acid (DHA) supplementation in atopic eczema: a randomized, double-blind, controlled trial. Br J Dermatol 2008;158:786-792.

Implications I

Implication Ib: For each $d' \neq 0$, the t-test is significant for a sufficiently large n_1 and non-significant for a smaller n_2

Is the effect significant? Sounds like a paradox but it is not......

$$t = d * \sqrt{n/2}$$

Implications III: Null Results

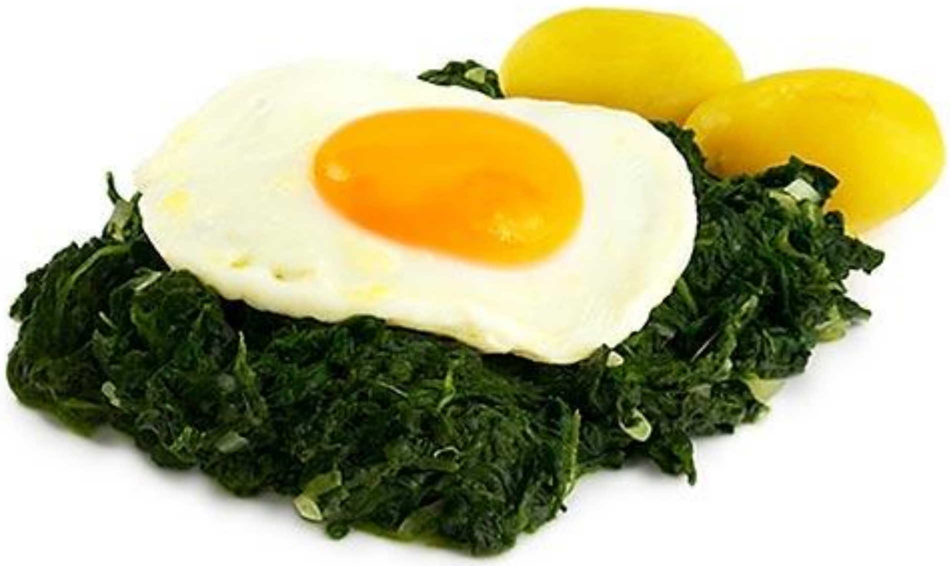
Comment: Null results are hard to publish

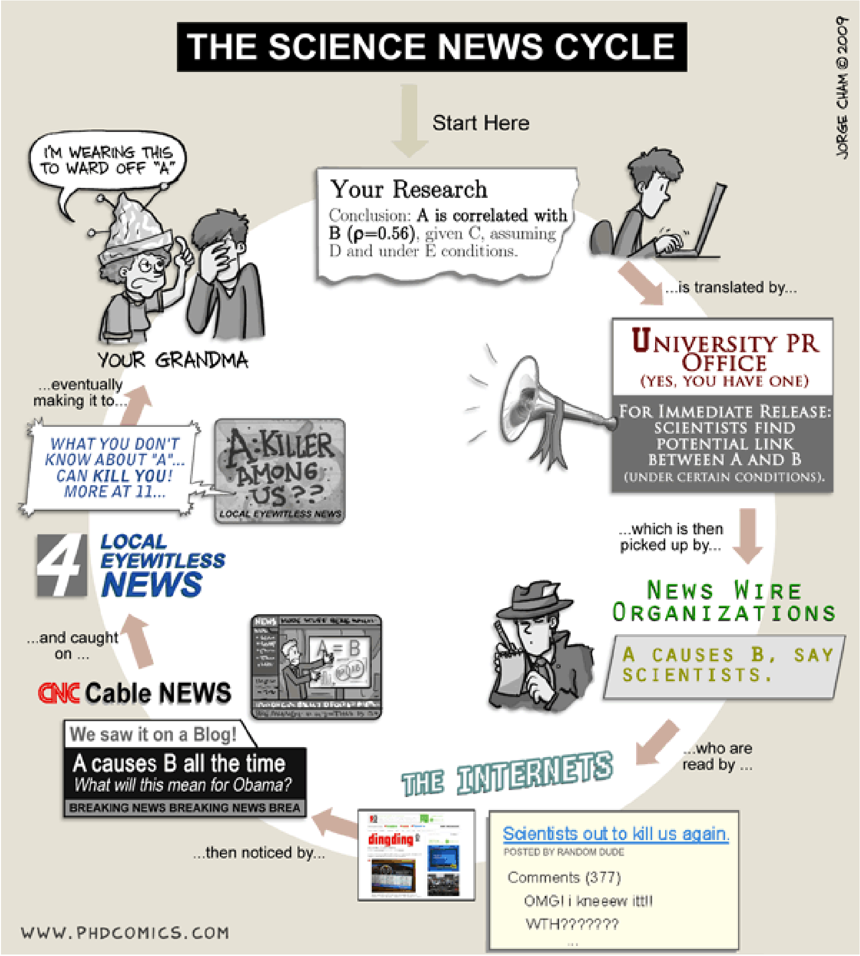
This leads to precedence effects & False Positives survival

Null results are the Archilles Heel of the Frequentist approach

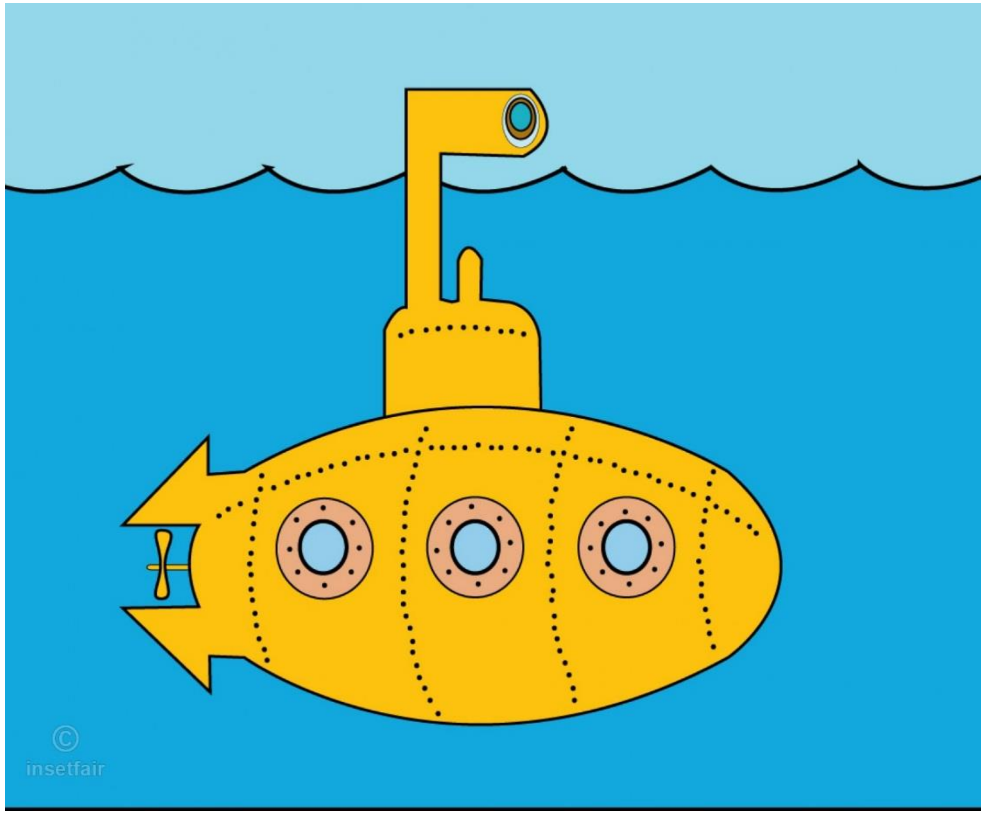
Implications III

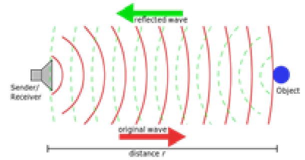
1. Did medieval people believe the earth is a disk?
2. Does spinach contain more iron than other veggies?
3. Are carots good for the eyes?
4. Does egg consumption increase blood cholesterol levels?
5. Does music by Mozart make cows give more milk than rock music (Mozart effect)?

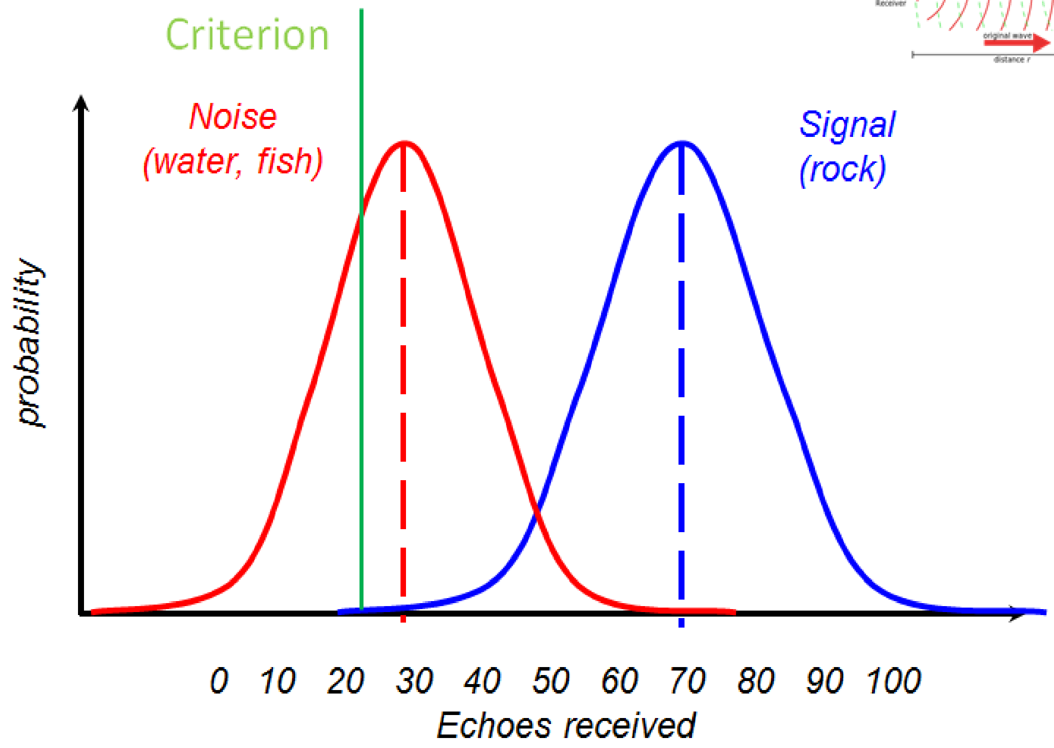



THE SCIENCE NEWS CYCLE
Start Here
JORGE CHAM © 2009
Your Research
Conclusion: A is correlated with B (ρ=0.56), given C, assuming D and under E conditions.
...is translated by...
UNIVERSITY PR OFFICE
(YES, YOU HAVE ONE)
FOR IMMEDIATE RELEASE: SCIENTISTS FIND POTENTIAL LINK BETWEEN A AND B (UNDER CERTAIN CONDITIONS).
...which is then picked up by...
NEWS WIRE ORGANIZATIONS
A CAUSES B, SAY SCIENTISTS.
...who are read by ...
THE INTERNETS
Scientists out to kill us again.
POSTED BY RANDOM DUDE
Comments (377)
OMG! i kneeew it!!
WTH???????
dingding
...then noticed by...
We saw it on a Blog!
A causes B all the time
What will this mean for Obama?
BREAKING NEWS BREAKING NEWS BREA
CNC Cable NEWS
...and caught on ...
4 LOCAL EYEWITLESS NEWS
WHAT YOU DON'T KNOW ABOUT "A"... CAN KILL YOU! MORE AT 11...
A: KILLER AMONG US??
LOCAL EYEWITLESS NEWS
...eventually making it to...
YOUR GRANDMA
I'M WEARING THIS TO WARD OFF "A"
WWW.PHDCOMICS.COM

Implications IV:Truth, Noise & Variability

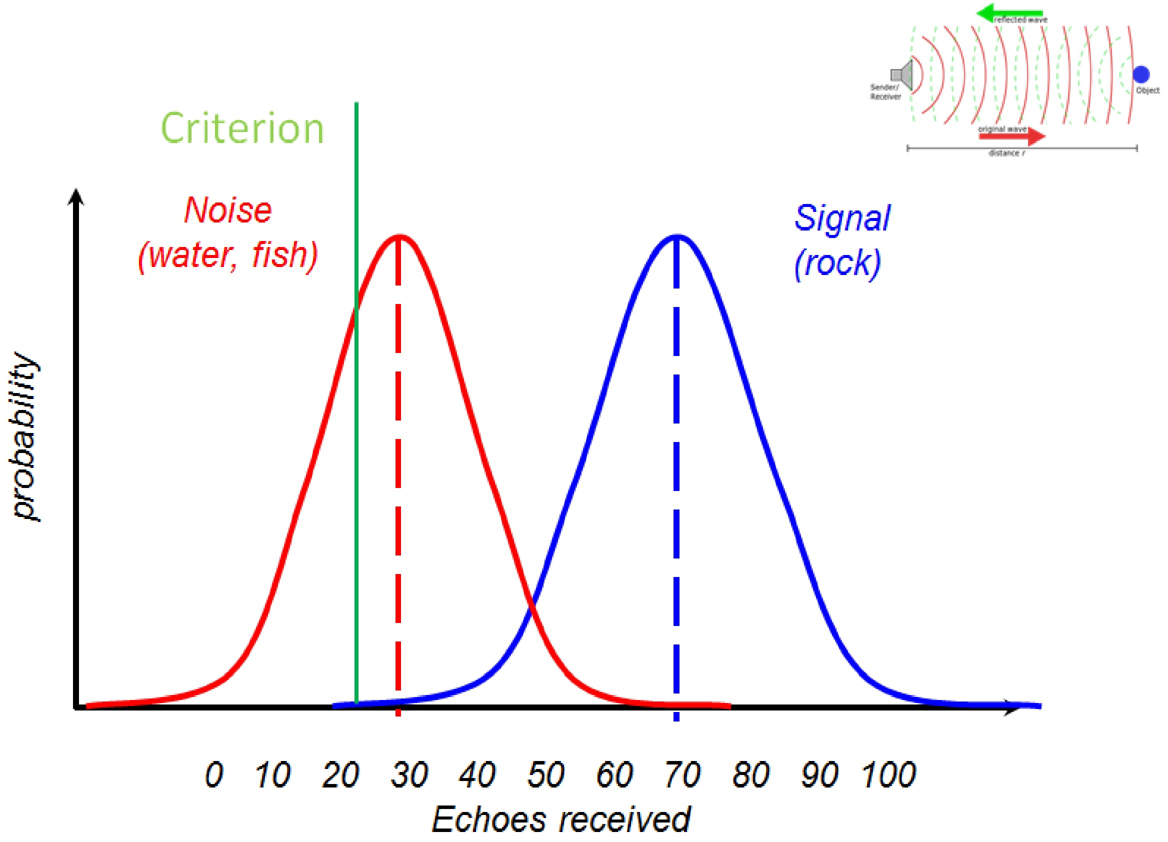






Implications IV:Truth, Noise & Variability

$$x_i = \mu + e_i$$



Where does the noise come from?

Why Synthetic Vitamins Are Harmful

http://www.goodlivingwarehouse.com/why-synthetic-vitamins-are-harmful-supplement-science/

Science has clearly demonstrated a need for humans to supplement with a few key nutrients if we want to live long, healthy lives. Unfortunately, science has also shown that synthetic forms of these nutrients are not only of little value, but can also be harmful to our well-being.

1. In a study published in the New England Journal of Medicine, 22,000 pregnant women were given synthetic Vitamin A. The study was halted because birth defects increased by 400%. N. Eng. J. Med. 1995; 333: 1369-1373

2. Another study involving 29,000 male smokers who were given synthetic beta-carotene and synthetic Vitamin E was also stopped when rates of lung cancer, heart attacks, and death increased. N. Eng. J. Med. 1994: 330; 1029-1035

Why Synthetic Vitamins Are Harmful

http://www.goodlivingwarehouse.com/why-synthetic-vitamins-are-harmful-supplement-science/

Science has clearly demonstrated a need for humans to supplement with a few key nutrients if we want to live long, healthy lives. Unfortunately, science has also shown that synthetic forms of these nutrients are not only of little value, but can also be harmful to our well-being.

1. In a study published in the New England Journal of Medicine, 22,000 pregnant women were given synthetic Vitamin A. The study was halted because birth defects increased by 400%. N. Eng. J. Med. 1995; 333: 1369-1373

2. Another study involving 29,000 male smokers who were given synthetic beta-carotene and synthetic Vitamin E was also stopped when rates of lung cancer, heart attacks, and death increased. N. Eng. J. Med. 1994: 330; 1029-1035

Where does the noise come from?

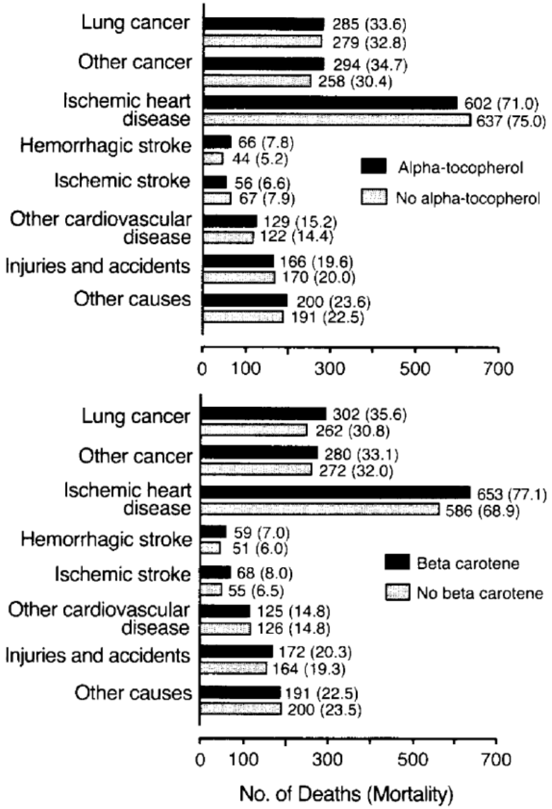


Figure 3. Deaths and Mortality Rates (per 10,000 Person-Years), According to Cause of Death, among Participants Who Received Alpha-Tocopherol Supplements and Those Who Did Not (Upper Panel) and among Participants Who Received Beta Carotene Supplements and Those Who Did Not (Lower Panel).
The cause of death was unknown for four participants.

Why Synthetic Vitamins Are Harmful

Only Beta Carotene

Where does the noise come from?

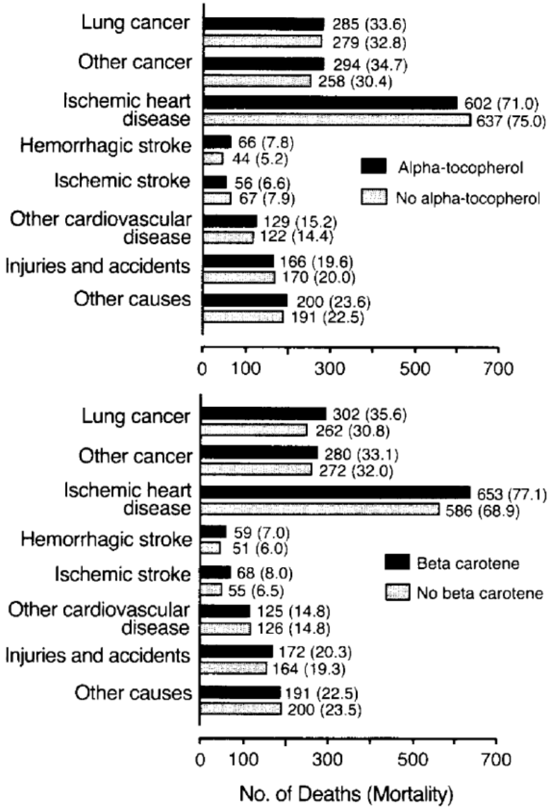


Figure 3. Deaths and Mortality Rates (per 10,000 Person-Years), According to Cause of Death, among Participants Who Received Alpha-Tocopherol Supplements and Those Who Did Not (Upper Panel) and among Participants Who Received Beta Carotene Supplements and Those Who Did Not (Lower Panel). The cause of death was unknown for four participants.

Some participants had no lung cancer even thus they were taken suppl. vitamins and the opposite was true too.

Is this noise? Maybe.

Maybe it is variability of the population. For some people vitamins help whereas for other not.

A small excessive subpopulation may cause the effect.

In this case, it is NOT TRUE that suppl. vitamins harm. Only, on average, there is evidence for the hypothesis

Where does the noise come from?

Where does the noise comes from?

- Measurement noise
- Variability in the population

There are two terms:

$$x_{ij} = \mu + v_i + e_{ij}$$

μ: grand mean of population
v_i: deviation from μ for subject i
e_{ij}: noise for a particular trial for subject i for trial number j

Where does the noise come from?

Some participants had no lung cancer even thus they were taken suppl. vitamins and the opposite was true too.
Is this noise? Maybe.
Maybe it is variability of the population. For some vitamins help whereas for other not.

A small excessive subpopulation may cause the effect.

In this case, it is NOT TRUE that suppl. vitamins harm. Only, on average, there is evidence for the hyopthesis.

The effect may be dose dependent. High doeses hurt, lower one help- for each invidual individually.

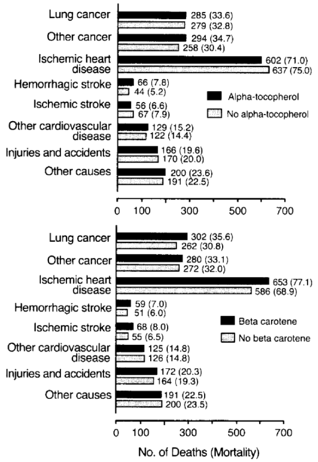
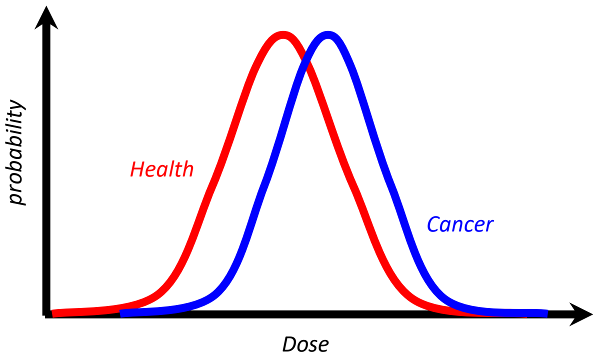


Figure 3. Deaths and Mortality Rates (per 10,000 Person-Years), According to Cause of Death, among Participants Who Received Alpha-Tocopherol Supplements and Those Who Did Not (Upper Panel) and among Participants Who Received Beta Carotene Supplements and Those Who Did Not (Lower Panel).
The cause of death was unknown for four participants.



Where does the noise come from?

EPFL

Why Synthetic Vitamins Are Harmful

http://www.goodlivingwarehouse.com/why-synthetic-vitamins-are-harmful-supplement-science/

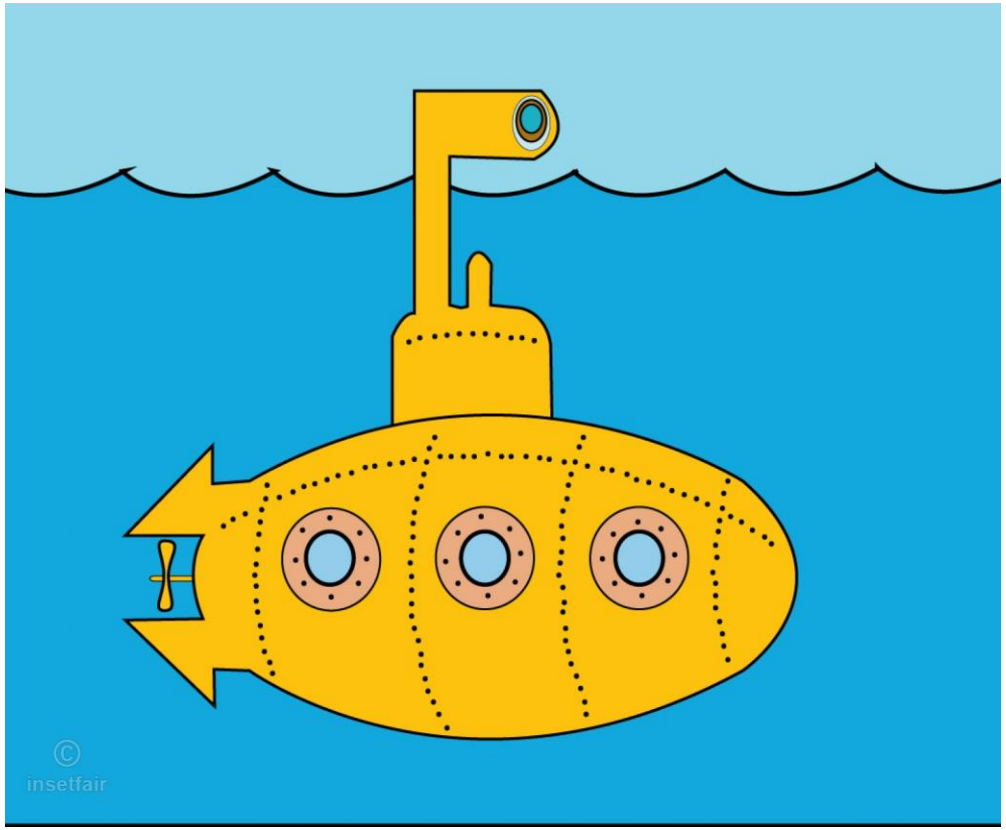
Dose dependent Variability in the Population

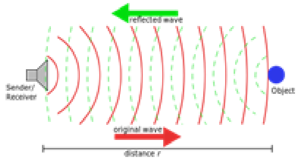
Science has clearly demonstrated a need for humans to supplement with a few key nutrients if we want to live long, healthy lives. Unfortunately, science has also shown that synthetic forms of these nutrients are not only of little value, but can also be harmful to our well-being.

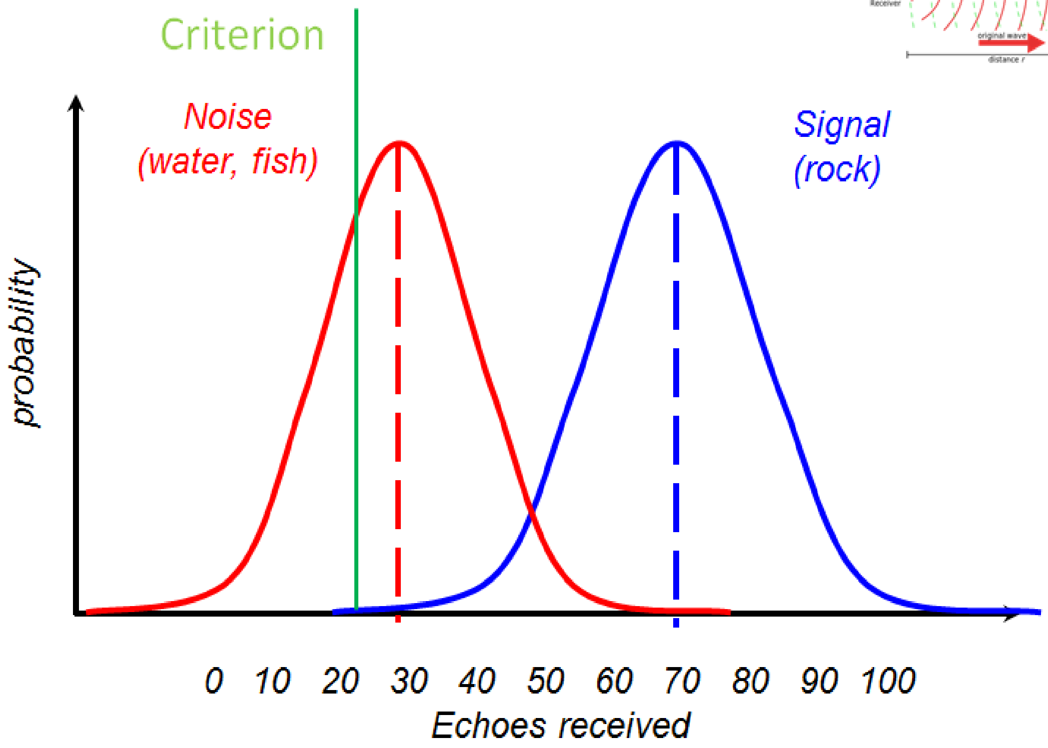
1. In a study published in the New England Journal of Medicine, 22,000 pregnant women were given synthetic Vitamin A. The study was halted because birth defects increased by 400%. N. Eng. J. Med. 1995; 333: 1369-1373

2. Another study involving 29,000 male smokers who were given synthetic beta-carotene and synthetic Vitamin E was also stopped when rates of lung cancer, heart attacks, and death increased. N. Eng. J. Med. 1994: 330; 1029-1035

Implications IV: Truth, Noise & Variability







Implications IV

Implication IVa: A conclusion holds only true for all samples when variability v_i is zero: $x_i = \mu + e_i$

Implications IV

Implication IVb: If the variability v_i not zero or you do not know whether it is zero, you cannot prove facts! There is no truth!

Stastistics is only about means in this case.

i) All humans are mortal	i) Humans are mortal (p=0.0001)
ii) Socrates is a human	ii) Socrates is a human
Hence, Socrates is mortal	Hence, no conclusion

Lesson #2

Often the experiment and statistics are correct but the interpretation is not

Implications IV

Implication IVc: Be aware of heterogeneous and “excess” sub-samples.

Example: A study reported that shorter students are more ill than taller ones because it was found that shorter students see their doctor more often.

Implications IV

Implication IVc: Be aware of heterogeneous and “excess” sub-samples.

Example: A study reported that shorter students are more ill than taller ones because it was found that shorter students see their doctor more often.

1. Female students are on average shorter than male students

2. They visit the gynecologist more frequently than male students the urologist- for preventive check ups

3. Students visit doctors very rarely in general

Implications IV

Implication IVd: How to stratify a sample is a non-trivial task. There is an intricate trade off between generality and variability!

Implications IV

	Cardiovascular Disease	Cancer	Accident	Other
Reference Groups				
US Population Ages 55–64, (*n* = 338, 127)	27%	34%	5%	35%
Non-Flight Astronauts, (*n* = 35)	9%*	29%	53%*	9%*
Astronaut Groups				
All Flight Astronauts, (*n* = 42)	17%	31%	43%*	10%*
Low Earth Orbit Astronauts, (*n* = 35)	11%*	31%	49%*	9%*
Apollo Lunar Astronauts, (*n* = 7)	43%†‡	29%	14%^	14%

From: Apollo Lunar Astronauts Show Higher Cardiovascular Disease Mortality: Possible Deep Space Radiation Effects on the Vascular Endothelium

Implication IV

Implication Ve: What is the Null model?

Implication IV

Implication Ve: What is the Null model?

Crud factor: In the social sciences and arguably in the biological sciences, "everything correlates to some extent with everything else."

Psychological Reports, 1990, 66, 195–244
Monograph Supplement 1–V66

#144

WHY SUMMARIES OF RESEARCH ON PSYCHOLOGICAL THEORIES ARE OFTEN UNINTERPRETABLE[1, 2]

PAUL E. MEEHL

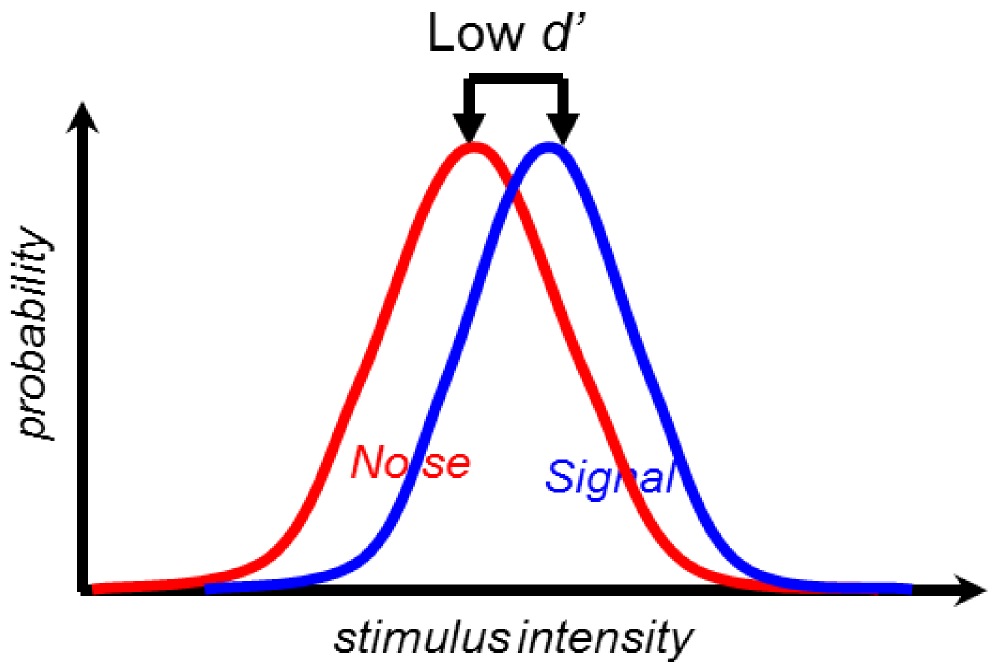
University of Minnesota

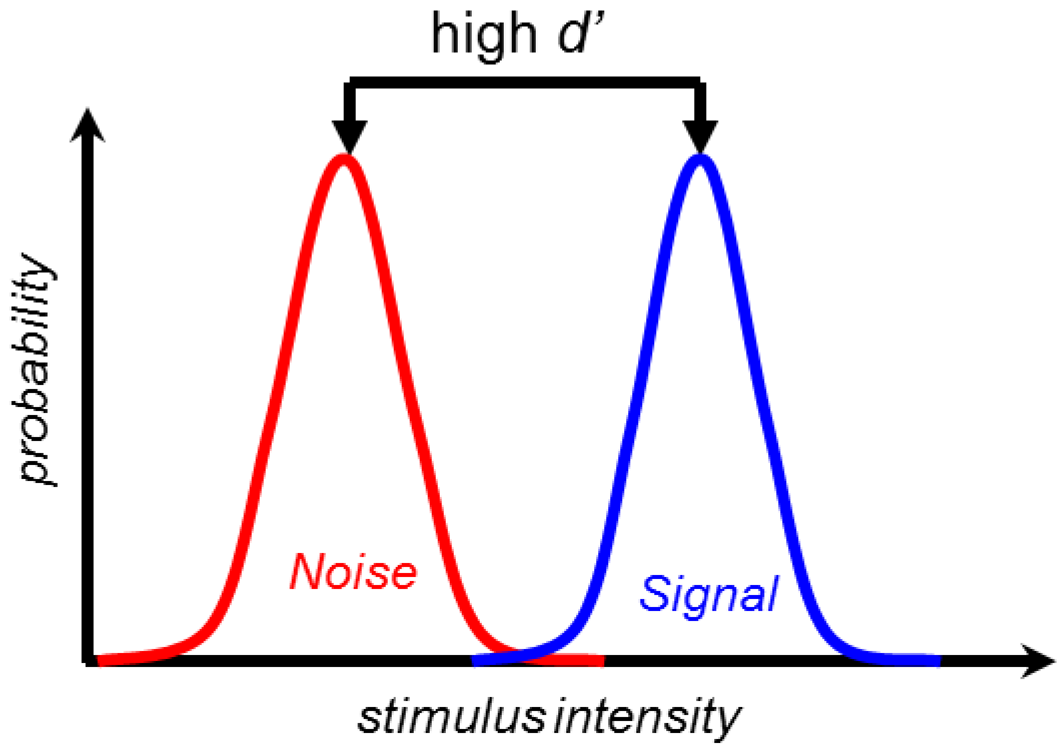
Implications IV

Implication IVf: When effects are dosage dependent, mean value comparisons may be very dangerous: noise, variability, and dosage are hard to disentangle and interactions may be present (see ANOVA).

Implication V: The statistics paradox

Implication Va: Why doesn‘t physics need statistics? Studies with large d‘ do not need statistics!
Studies with small d‘ need statistics- and are often irrelevant
Statistics makes only sense for a range of d‘s, eg. 0.2-0.8





Implication V: The statistics paradox

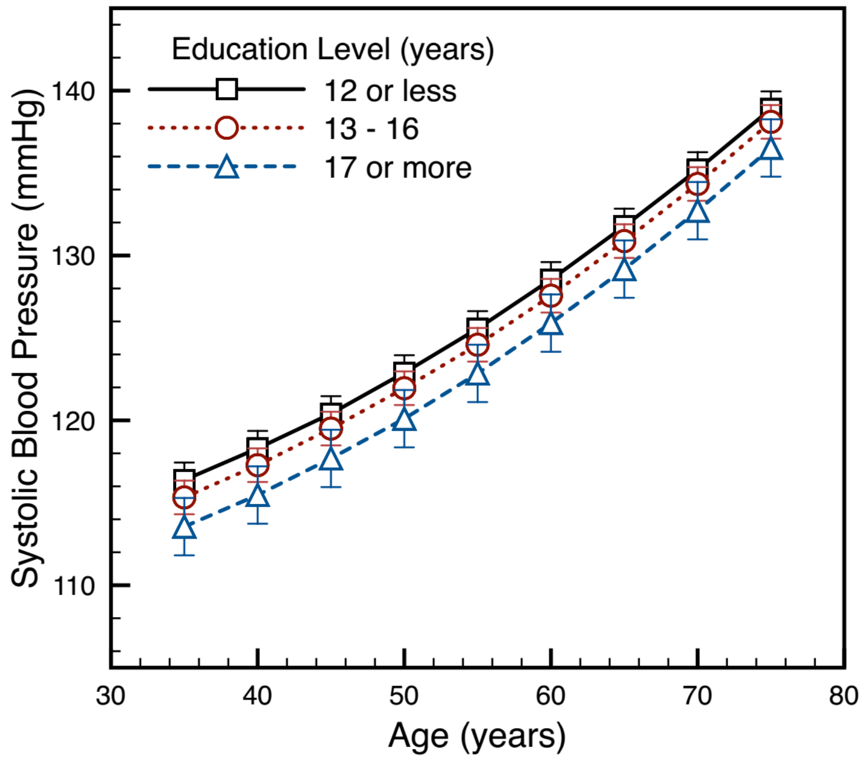
Implication Vb: The problem of small effect sizes: the danger with cohort studies

Cohorts: feature defined, e.g. patients vs. controls

Experimental design: Randomized

Implication V: The statistics paradox

Implication Vb: The problem of small effect sizes: the danger with cohort studies



Cohort Studies, with small effect sizes

- Crud Factor: everything is correlated

- Uncontrolable but systematic differences
in the design:
  - Data collection (multi-center studies)
  - Correlated features (particular, cohort studies)
  - Inhomogenous samples (Subpopulations)

- Collapsed data (dose effects)

- Violations of assumptions of statistical tests

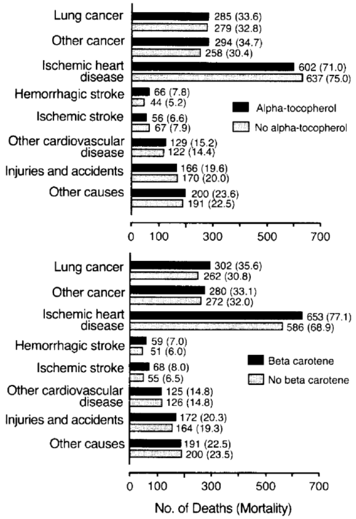
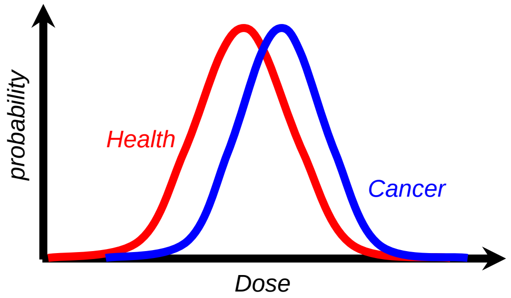


Figure 3. Deaths and Mortality Rates (per 10,000 Person-Years), According to Cause of Death, among Participants Who Received Alpha-Tocopherol Supplements and Those Who Did Not (Upper Panel) and among Participants Who Received Beta Carotene Supplements and Those Who Did Not (Lower Panel).
The cause of death was unknown for four participants.



Implication IV: The statistics paradox

Implication Vc: Small effects may be important too…….

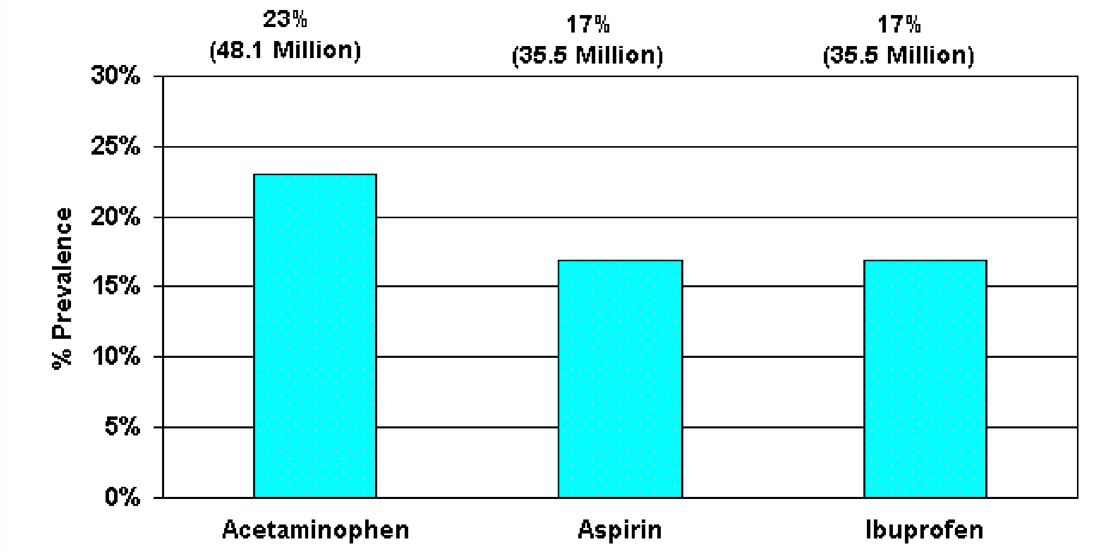
Implication IV: The statistics paradox

Implication Vc: Small effects may be important too…….

DNA, Theory of relativity, First computer....

Implication IV: The statistics paradox

Implication Vc: Small effects may be important too…….



Americans consume 15 tons of aspirin a day, 19 billion tablets per year. Although thought to be harmless, a single aspirin is…

… responsible for 1500-2000 deaths

Implication IV: The statistics paradox

Implication Vc: Small effects may be important too.

The problem is that we often stay with small effects and treat them as true and large effects.

It depends on the context.

Motivation -Examples

EPFL

1. INTRODUCTION

Statistical studies are prominently featured in most major newspapers on a daily or weekly basis, yet most citizens, and even many reporters, do not have the knowledge required to read them critically. When statistics courses were first introduced, they were taken primarily by students who intended to pursue their own research, or were in disciplines that required them to analyze data as part of their training. The focus of those courses was on computation, and little emphasis was placed on how to integrate information from study design to final conclusions in a meaningful way. Much has changed since then, in three ways: the audience, the tools available to students, and the world around us.

Yet many instructors have not made any changes in how they teach introductory statistics.

What Educated Citizens Should Know about Statistics and Probability
Author(s): Jessica Utts
Source: *The American Statistician*, Vol. 57, No. 2 (May, 2003), pp. 74-79
Published by: American Statistical Association
Stable URL: http://www.jstor.org/stable/30037236
Accessed: 25/04/2013 06:59

Summary

- Confusion of Significance and Effect Size
- Confusion of Evidence with Truth
- Inability to handle Null results and Precedence Effects + Piloting and file drawer problem
- Small effect sizes
- and publication pressure
- ...have alltogether led to the problematic situation we are facing at the moment....

Solutions

Solutions

The normativity of statistics:

Significance: $d' * \sqrt{n/2} > \alpha$, $e.g. 1.96$ $(depending\ on\ n)$ for **α of 0.05**

d’ = 0.2 small effect

d’ = 0.5 medium effect

d’ = 0.8 large effect

Fix effect size too! Equivalent fix n

Bayes approach: Part III

Solutions

Keep your eyes open…

Survey for statistics class

Homework for next week, 31st of Oct

- Ask 10 friends/colleagues (not from the course) to fill out this survey
- If possible, ask people from EPFL (at least, they need to know EPFL buildings)
- It consists of basic questions about yourself and your preferences.
- It only takes a few minutes.
- https://forms.gle/5yHGZ9n6zGsKv43TA
- All information (including the link) are on Moodle.

Survey about your preferences for a statistics class

The aim of this survey is to show students from our "Understanding statistics and experimental design" class that statistics can be tricky if you don't know what you're doing. Basically, you can find spurious correlations that are non-sense, but still significant.

It will only take you a few minutes to answer this questionnaire. Thank you in advance!

*Obligatoire

What is your gender? *

- ○ Female
- ○ Male

What is your zodiac sign (French translation in brackets)? *

- ○ Aquarius (Verseau)
- ○ Pisces (Poissons)
- ○ Aries (Bélier)
- ○ Taurus (Taureau)
- ○ Gemini (Gémeaux)
- ○ Cancer (Cancer)
- ○ Leo (Lion)

How tall are you? Please indicate your height in centimeters! *

Votre réponse

Which school/university are you from? If you are neither from EPFL, nor from UNIL, please tick "Other". *

- ○ EPFL
- ○ UNIL
- ○ Other

If you ticked "EPFL", please indicate which faculty you are from:

- ○ ENAC Faculté Environnement Naturel, Architectural et Construit
- ○ SB Faculté Sciences de Base
- ○ STI Faculté Sciences et techniques de l'ingénieur
- ○ IC Faculté Informatique et Communications
- ○ SV Faculté Sciences de la Vie
- ○ CDM Collège du Management de la Technologie
- ○ CdH Collège des Humanités

What is your academic status? *

- ○ Bachelor (BA)
- ○ Master (MA)
- ○ PhD
- ○ Post-Doc
- ○ Professor
- ○ Other

Question 6 You are comparing two treatments against eye itches. You take two samples of 18 people each and measure the strength of eye itch pain. Calculate the t-value for an independent samples t-test with an estimated difference of means of 0.2 and an estimated standard deviation of 0.10.

- 6
- 8.49
- 2.4
- 2
- 4.47
- None of them

Question 6 You are comparing two treatments against eye itches. You take two samples of 18 people each and measure the strength of eye itch pain. Calculate the t-value for an independent samples t-test with an estimated difference of means of 0.2 and an estimated standard deviation of 0.10.

- ■ 6
- ☐ 8.49
- ☐ 2.4
- ☐ 2
- ☐ 4.47
- ☐ None of them

Question 8 An experiment delivered a significant result with a p-value of 0.049. What can you tell about the effect size in terms of $Cohen's\ d$ given Cohen's classification of low, medium, and high effects sizes?

It is high because the result is significant

It is low

It is medium high because the result is barely significant

None of them

Question 8 An experiment delivered a significant result with a p-value of 0.049. What can you tell about the effect size in terms of $Cohen's\ d$ given Cohen's classification of low, medium, and high effects sizes?

- [ ] It is high because the result is significant
- [ ] It is low
- [ ] It is medium high because the result is barely significant
- [x] None of them

Take home messages

Take Home Messages

1. Even small effect sizes lead to significant results when the sample size is sufficiently large.
2. Do not compare the p-value of two experiments if n is not identical: a smaller p does not imply more significance.
3. Statistical significance is not practical significance.
4. Absence of proof is not proof of absence: avoid conclusions from a Null result.
5. Do not pit a significant experiment against a non-significant control experiment.
6. Cohort studies with small effects are usually useless.
7. A statement like "X is true" can only be true for sure if inter-subject variability is zero.

END Class 3b

The Crud Factor

Crud factor: In the social sciences and arguably in the biological sciences, “everything correlates to some extent with everything else.”

In 1966, the University of Minnesota Student Counseling Bureau’s Statewide Testing Program administered a questionnaire to 57,000 high school seniors, the items dealing with family facts, attitudes toward school, vocational and educational

plans, leisure time activities,school organizations, etc. Wecross-tabulated a total of 15 (and then 45) variables including the following (the number of categories for each variable given in parentheses): father’soccupation (7), father’s education (9), mother’s education (9), number of siblings (10), birth order (only, oldest, youngest, neither), educational plans after high school (3), family attitudes towards college (3), do you like school (3), sex (2), college choice (7), occupational plan in ten years (20), and religious preference (20). In addition, there were 22 “leisure time activities” such as “acting,” “model building,” “cooking,” etc., which could be

treated either as a single 22-category variable or as 22 dichotomous variables. There were also 10 “high school organizations” such as “school subject clubs,” “farm youth groups,” “political clubs,” etc., which also could be treated either as a single ten-category variable or as ten dichotomous variables.

Considering the latter two variables as multichotomies gives a total of 15 variables producing 105 different cross-tabulations. All values of $\chi 2$ for these 105 cross-tabulations were statistically significant, and 101 (96%) of them were significant with a probability of less than 0.00001.

These relationships are not, I repeat, Type I errors. They are facts about the world, and with N= 57,000 they are pretty stable.

Motivation - Examples

Ioannides: 12 hazelnut

We can use the overlap for all probability functions (Cohen's U)