University of Zurich[UZH]

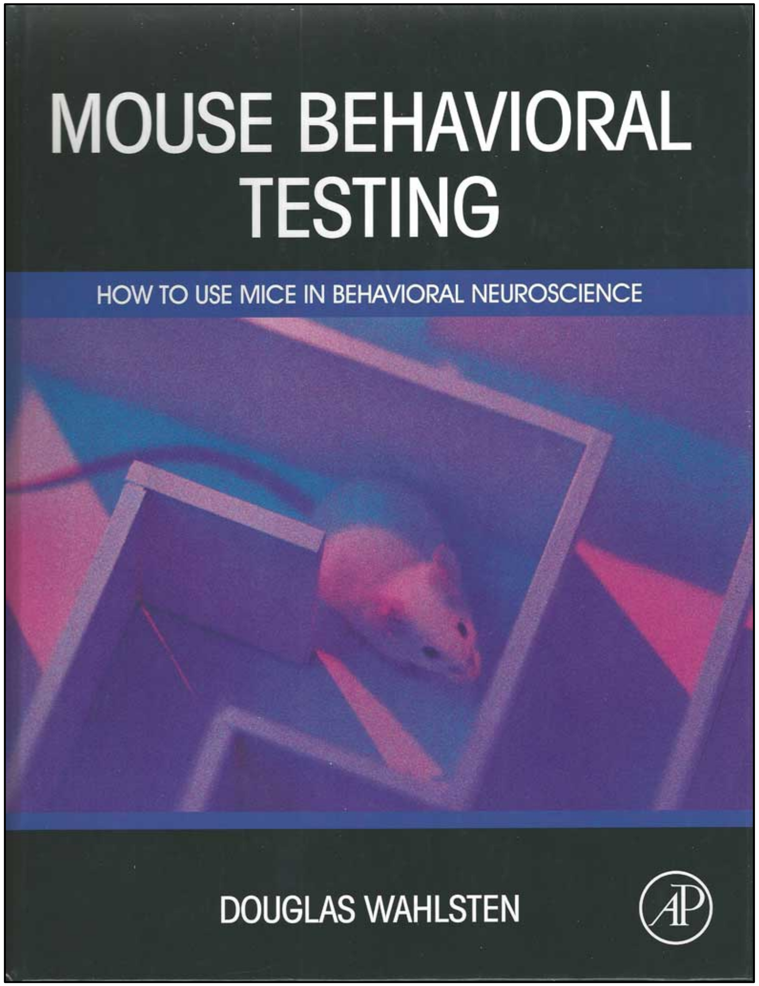
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

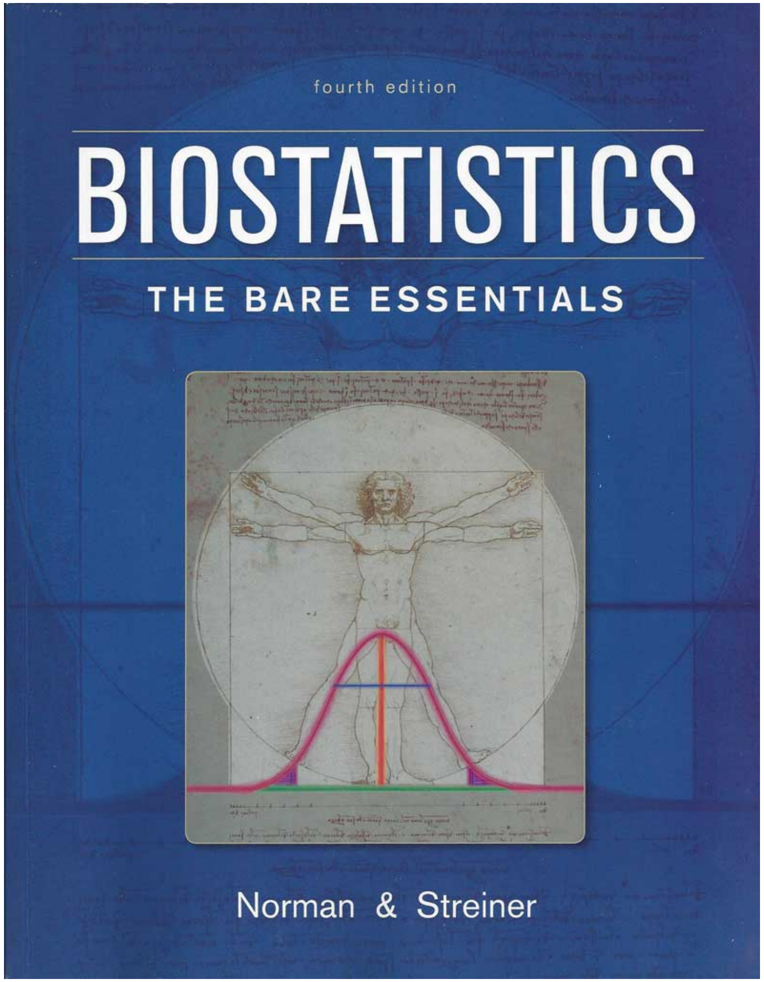
ZNZ MD/PhD Neuroscience Course, Module BIO628
Thu 01.06.2017

Data analysis and presentations: Examples from basic statistics

David P. Wolfer MD

Institute of Anatomy, University of Zurich
Institute for Human Movement Sciences and Sport, ETH Zurich
http://www.dpwolfer.ch
davidp.wolfer@.uzh.ch, dwolfer@ethz.ch

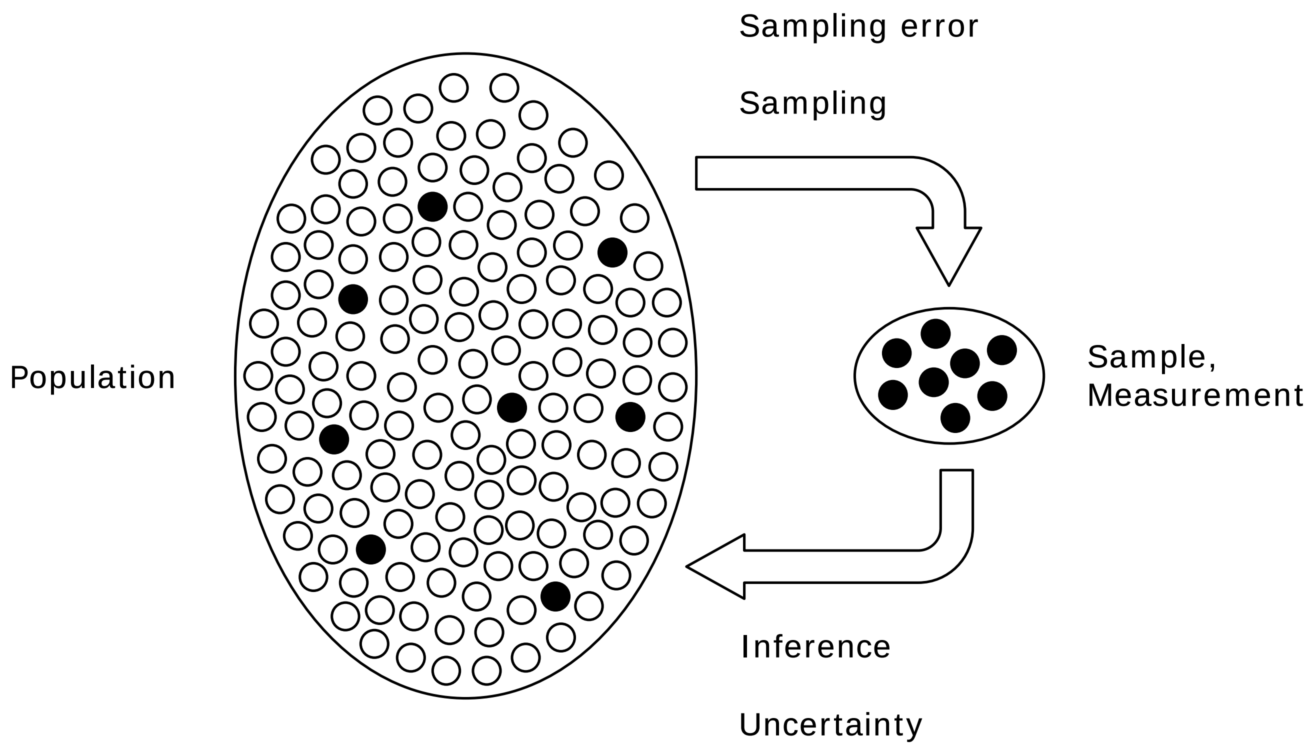




Wahlsten D
Mouse Behavioral Testing
Academic Press, 1. edition, 2011

Norman GR, Streiner DL
Biostatistics, the bare essentials
BC Decker, 4. edition, 2014

Sampling


Sampling error
Sampling
Population
Sample,
Measurement
Inference
Uncertainty

Mean, variance and standard deviation

Sample		Population
$M = \frac{\Sigma (X)}{n}$	mean	$\mu = \frac{\Sigma (X)}{N}$
$S^2 = \frac{\Sigma (X-M)^2}{n-1}$	variance	$\sigma^2 = \frac{\Sigma (X-\mu)^2}{N}$
S	standard deviation	σ
$\frac{S}{\sqrt{n}}$	standard error of mean	

Effect size

	true effect size, population	estimated effect size, sample
t-test	$\delta = \Delta\mu/\sigma$ 0.2 = small 0.5 = medium 0.8 = large	$d = \Delta M/S_{pooled}$
ANOVA	$\omega^2 = \frac{\sigma^2 \text{ between groups}}{\sigma^2 \text{ total}}$ 5% = small 20% = large	$\eta^2 = \frac{S^2 \text{ between groups}}{S^2 \text{ total}}$

Hypothesis testing 1

	Ho true	**Ho false**
reject Ho	**α** P(reject Ho | Ho true) Type I error false positive	1-**β** P(reject Ho | Ho false) power true positive
retain Ho	1-**α** P(retain Ho | Ho true) true negative	**β** P(retain Ho | Ho false) Type II error false negative

Type-I error
- important if we **reject** Ho
- calculated from experimental samples by statistical tests

Type-II error
- important if we **retain** Ho
- depends on:
  - type-I-error threshold
  - effect size, sample size

Sample size for t-test

2 groups: n & Power to detect δ > 0 when Null is δ = 0

Using Wahlsten's (1991) eqn (5) $n = 2C_{\alpha,\beta}/\delta^2 + 2$

δ = 1.000

Note: $\delta = (\mu_1 - \mu_2)/\sigma$

To do a quick calculation of effect size from means:

μ_1 =	15.000
μ_2 =	20.000
σ =	4.500
δ =	-1.111

For one-tailed test of null hypothesis using t test — General result

Type I error (α)	Type II error (β)	Power (%)	Z_α 1-tail	$Z_{1-\beta}$	$2C_{\alpha,\beta}$	n per group	Rounded up
0.05	0.05	95	-1.645	1.645	21.644	23.644	24
0.05	0.1	90	-1.645	1.282	17.128	19.128	20
0.05	0.2	80	-1.645	0.842	12.365	14.365	15
0.01	0.05	95	-2.326	1.645	31.541	33.541	34
0.01	0.1	90	-2.326	1.282	26.034	28.034	29
0.01	0.2	80	-2.326	0.842	20.072	22.072	23
0.001	0.05	95	-3.090	1.645	44.842	46.842	47
0.001	0.1	90	-3.090	1.282	38.225	40.225	41
0.001	0.2	80	-3.090	0.842	30.919	32.919	33

For two-tailed test of null hypothesis using t test — General result

Type I error (α)	Type II error (β)	Power (%)	$Z_{\alpha/2}$ 2-tail	$Z_{1-\beta}$	$2C_{\alpha,\beta}$	n per group	Rounded up
0.05	0.05	95	-1.960	1.645	25.989	27.989	28
0.05	0.1	90	-1.960	1.282	21.015	23.015	24
0.05	0.2	80	-1.960	0.842	15.698	17.698	18
0.01	0.05	95	-2.576	1.645	35.628	37.628	38
0.01	0.1	90	-2.576	1.282	29.759	31.759	32
0.01	0.2	80	-2.576	0.842	23.358	25.358	26
0.001	0.05	95	-3.291	1.645	48.716	50.716	51
0.001	0.1	90	-3.291	1.282	41.808	43.808	44
0.001	0.2	80	-3.291	0.842	34.149	36.149	37

Samplesize 2 2x2 Jgroups P.xls

http://www.elsevierdirect.com/v2/companion.jsp?ISBN=9780123756749

3 bad consequences of low power

1. high probability to miss true effects
2. tendency to overestimate effect size (selection effect)
3. lower positive predictive value

Nat Rev Neurosci 14:365, 2013
Power failure: why small sample size undermines the reliability of neuroscience

Hypothesis testing 2

	Ho true	Ho false
reject Ho	**α** P(reject Ho | Ho true) Type I error false positive	1-**β** P(reject Ho | Ho false) power true positive
retain Ho	1-**α** P(retain Ho | Ho true) true negative	**β** P(retain Ho | Ho false) Type II error false negative

P(Ho false | Ho rejected)
positive predictive value

$$\text{PPV} = \frac{\text{true positives}}{\text{true + false positives}} = \frac{(1-\beta)\mathbf{R}}{(1-\beta)\mathbf{R}+\alpha}$$

R = pre-study odds

4 caveats regarding the interpretation of ɑ

1. ɑ is not a measure of effect size.

2. If ɑ is larger than the rejection threshold, this is no prof that Ho is true.
Under low power conditions, negative test results are inconclusive!
special statistics are needed to demonstrate equivalence.

3. ɑ does not tell us how likely it is that the effect is real if the test tells us to reject Ho. For this we need the positive predictive value (PPV).

4. ɑ is only valid under the assumption that only one test is done to test Ho:
with N tests true type-I error probability = $1-(1-\alpha)^N$
Example: 2 Tests at 5% $1-0.95^2 = 9.8\%$

Nat Rev Neurosci 14:365, 2013
Power failure: why small sample size undermines the reliability of neuroscience

Nature 506:150,2014
Statistical errors