Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

University of Zurich[UZH]

ZNZ Zentrum für Neurowissenschaften Zürich
Neuroscience Center Zurich

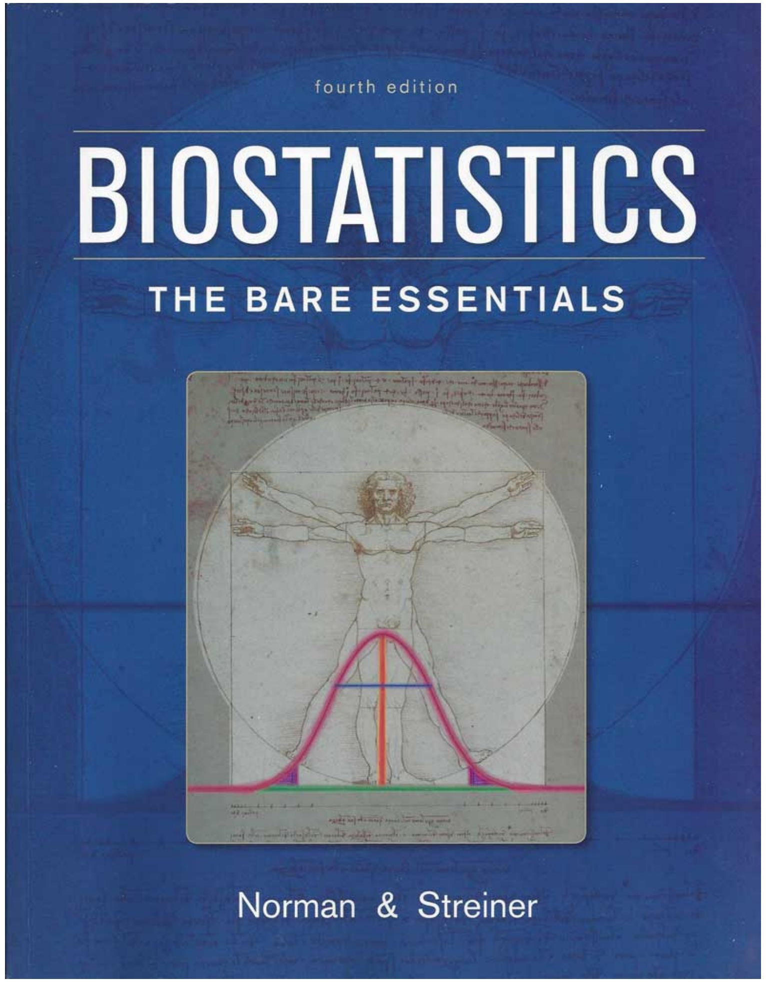
ZIHP

Data analysis and presentations:
Examples from basic statistics

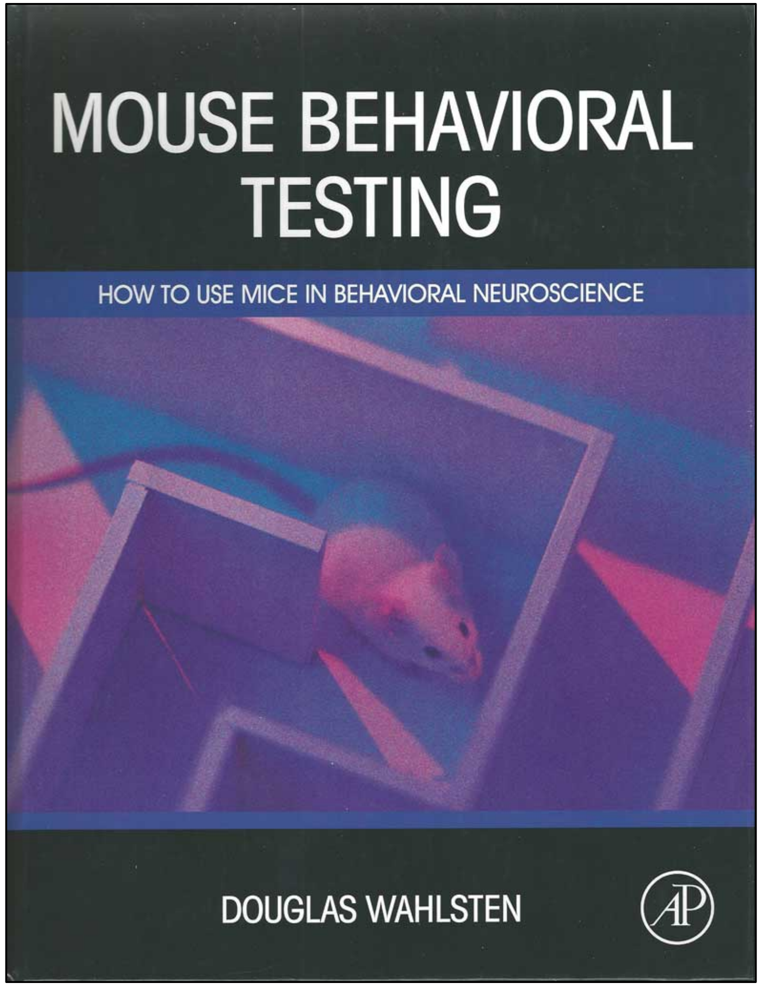
Fri 01.06.2018 13:00-14:30
ZNZ MD/PhD Neuroscience Course, Module BIO628
Room Y35 F47

David P. Wolfer, MD

Institute of Human Movement Sciences and Sport, ETH Zurich
Institute of Anatomy, University of Zurich



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



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

Population, sample and sampling error

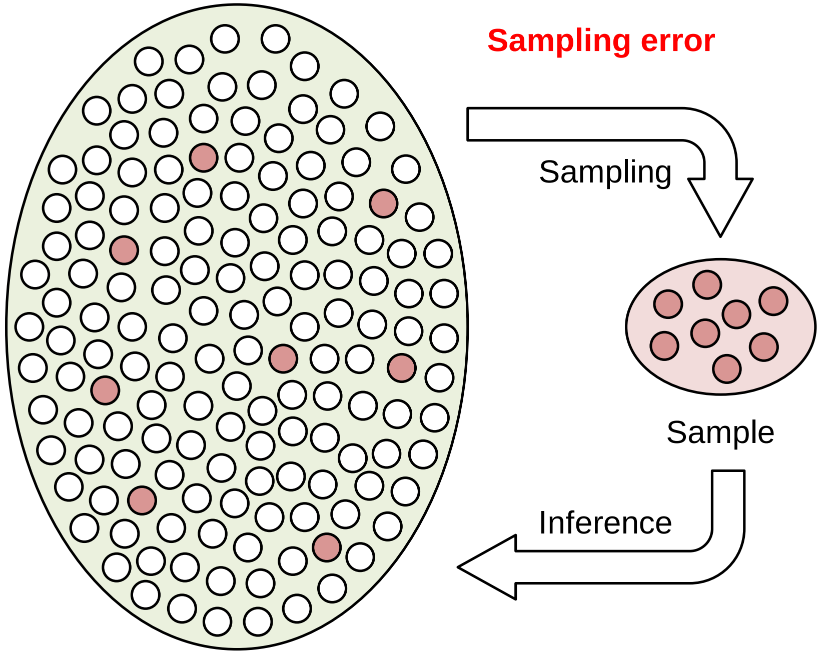
Population

$$\mu = \frac{\Sigma\,(X)}{N}$$

$$\sigma^2 = \frac{\Sigma\,(X-\mu)^2}{N}$$

σ

Sampling error



Population

Uncertainty

Sample

$$M = \frac{\Sigma\,(X)}{n}$$ mean

$$S^2 = \frac{\Sigma\,(X-M)^2}{n-1}$$ variance

S standard deviation

Sampling error

- *Due to sampling error, mean and standard deviation of samples always differ from the true population values, they are only estimates of the mean and standard deviation of the population*

- *If multiple samples are drawn from the same population, due to sampling error their means and standard deviations will always differ from each other*

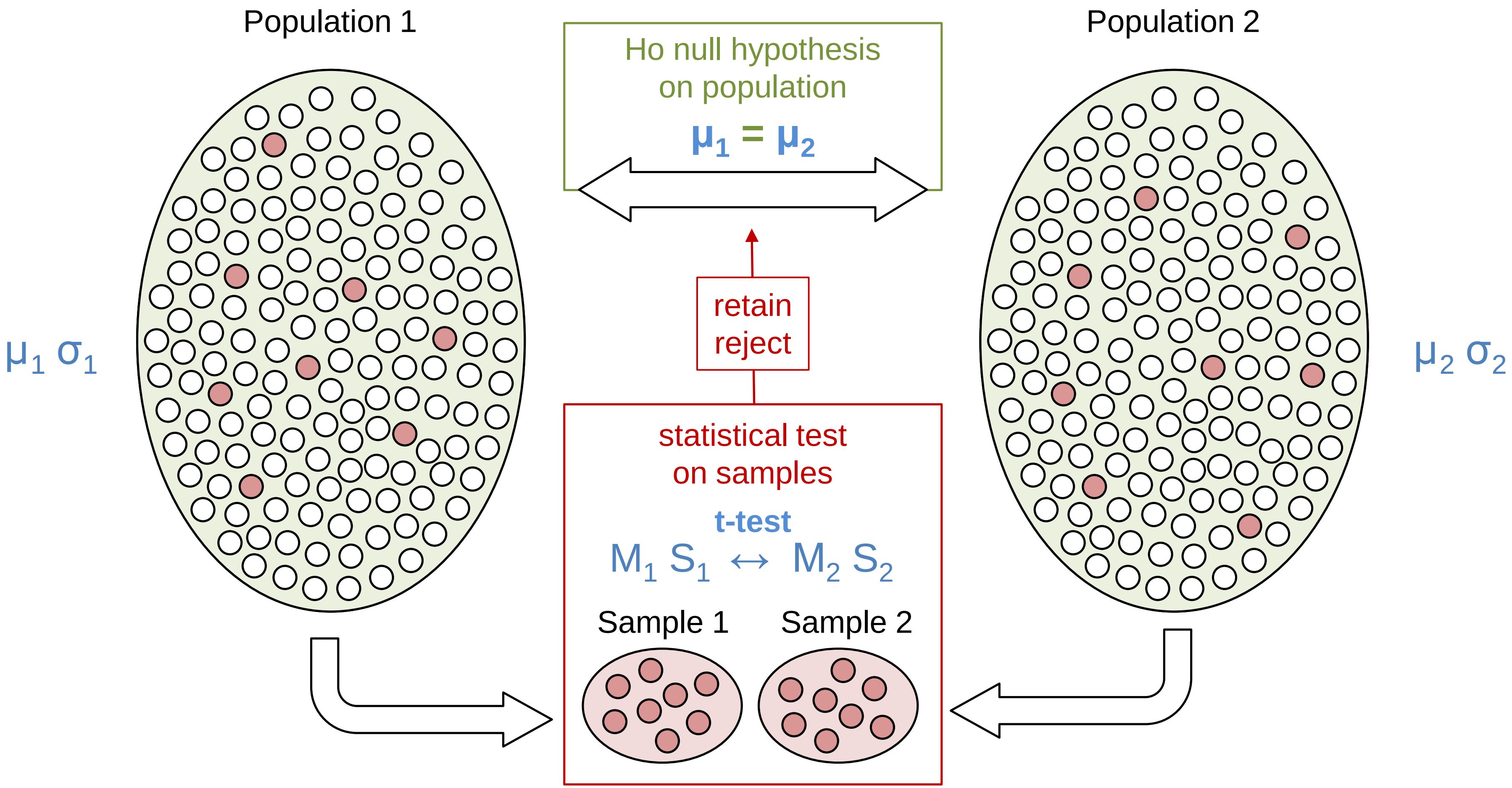
- *Repeated estimates of mean and standard deviations converge on the true population values, provided that*
  - *population data are normally distributed*
  - *sampling is unbiased*

- *Dispersion of estimates of mean and standard deviation decreases with increasing sample size:*

$$SE = \frac{s}{\sqrt{n}}$$

standard error of the mean

Statistical comparison of two populations



Hypothesis testing 1

	Ho false	Ho true
test positive: reject Ho	true positive	false positive
test negative: retain Ho	false negative	true negative

Sampling error:
We may be correct or not when using the results of a statistical test as criterion to reject or retain the null hypothesis

Hypothesis testing 1

<table>
<tr><th></th><th>Ho false</th><th>Ho true</th></tr>
<tr><td>test positive: reject Ho</td><td>true positive</td><td>α
P(reject Ho | Ho true)
Type I error
false positive</td></tr>
<tr><td>test negative: retain Ho</td><td>false negative</td><td>1-α
P(retain Ho | Ho true)
specificity
true negative</td></tr>
</table>

Type-I error

- *reject Ho when it is in fact true = false positive*
- *likelihood **α** estimated from experimental sample data by statistical tests*
- *Ho rejected if estimate ≤ threshold, typically **0.05***

Hypothesis testing 2

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

Type-II error

- *retain Ho when it is in fact false = false negative*
- *likelihood β determined by experimental design:*
  - *- sample size*
  - *- type-I error threshold*
  - *- effect size*
- *Typically accepted $\beta \leq$ **0.2**, same as power 1-$\beta \geq$ **0.8***

Measures of 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	$\omega^2 = \dfrac{\sigma^2 \text{ between groups}}{\sigma^2 \text{ total}}$ 5% = small 20% = large	$\eta^2 = \dfrac{S^2 \text{ between groups}}{S^2 \text{ total}}$

Hypothesis testing 2

<table>
<tr><th></th><th>Ho false</th><th>Ho true</th></tr>
<tr><td>test positive: reject Ho</td><td>1-β
P(reject Ho | Ho false)
power = sensitivity
true positive</td><td>α
P(reject Ho | Ho true)
Type I error
false positive</td></tr>
<tr><td>test negative: retain Ho</td><td>β
P(retain Ho | Ho false)
Type II error
false negative</td><td>1-α
P(retain Ho | Ho true)
specificity
true negative</td></tr>
</table>

Determining sample size:

- *before experiment is done!*
- *depends on*
  - *type-I error threshold, typically* $\alpha=0.05$
  - *expected effect size, eg.* $\delta=1$
  - *desired power, typically* $1-\beta=0.8$

Hypothesis testing 3

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

***FDR** (false discovery ratio)*

P(Ho true | test positive)

FDR = 1 - PPV

$$FDR = \frac{\alpha}{(1-\beta)\cdot \mathbf{R} + \alpha}$$

R = Ho false / Ho true (pre-study odds)

Hypothesis testing 3

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

PPV *(positive predictive value)*

P(Ho false | test positive)

PPV = 1 - FDR

$$PPV = \frac{(1-\beta)\cdot \mathbf{R}}{(1-\beta)\cdot \mathbf{R} + \alpha}$$

R = Ho false / Ho true
(pre-study odds)

Hypothesis testing 4

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

***NPV** (negative predictive value)*

P(Ho true | test negative)

$$NPV = \frac{(1-\alpha)}{(1-\alpha) + \beta \cdot \mathbf{R}}$$

R = Ho false / Ho true
(pre-study odds)

Interpretation of test outcomes

- positive test outcome:
  type-I error p-value is not a measure of false discovery ratio

- false discovery ratio is typically larger than type-I error p-value, especially with
  - underpowered studies
  - low pre-study odds
  - multiple tests without correction

- Typically don't assume pre-study odds >1
  - ~ 1 in phase 3 clinical trials
  - pre-study odds <1 in basic research
  - <<1 for compound screening,
    "fishing expeditions" with poor hypotheses

- negative test outcome:
  absence of evidence is not necessarily evidence of absence of an effect

- negative predictive value is low and result inconclusive with
  - underpowered studies
  - high pre-study odds

- absence of effect should be demonstrated using dedicated tests for equivalence