LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Graduate School of Systemic Neurosciences
LMU Munich

Applied statistics for neuroscientists

Stefan Glasauer

Center for Sensorimotor Research, Clinical Neuroscience
Ludwig-Maximilians-Universität München

What is this course all about?

Hypothesis testing

Parametric tests

T-test

ANOVA

General linear model

Non-parametric tests

And loads of more stuff related to all this …

http://jp.physoc.org/cgi/collection/stats_reporting
http://advan.physiology.org/search?tocsectionid=Staying+Current&submit=Submit

Hypothesis testing

Hypothesis: the next throwing of the die will show a 6.

Correct.

Hypothesis testing

Hypothesis: every throwing of the die will show a 6.

Wrong.

Hypothesis testing

Hypothesis: 6 appears in 1/6 of all cases (fair die)

???

Hypothesis testing

Ok, let's do an experiment

We roll the die 235 times.

The 6 appears 51 times.

If the die is fair, we would expect 6 to come up 235/6 = 39.17 times.

How do we proceed now?

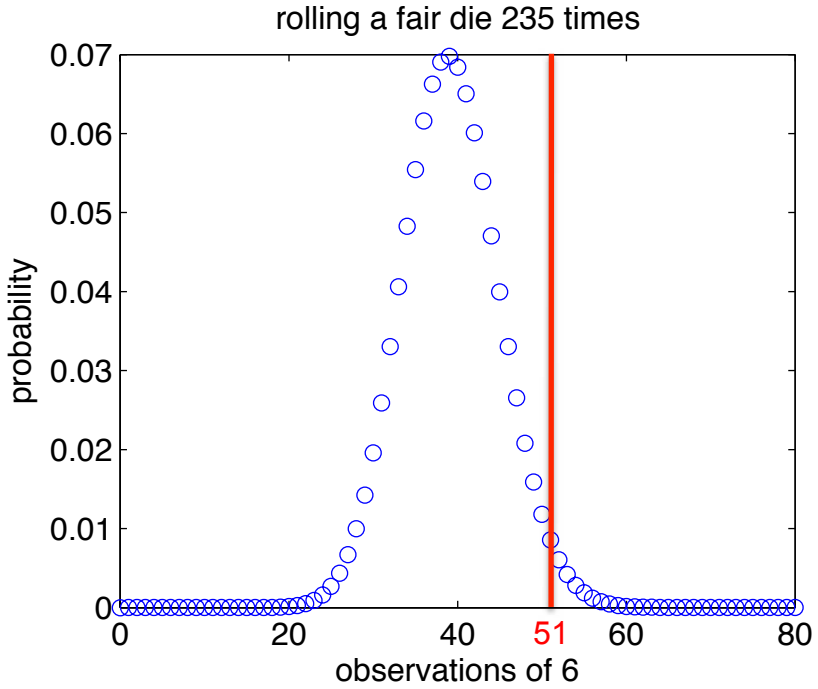
Hypothesis testing

In each trial, the die either shows 6 or not.

There are only two possibilities occurring with probabilities *p* and 1-*p*.

The corresponding probability distribution is the binomial distribution (parameters *n* and *p*) with

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$



Hypothesis testing

We want to know how likely it is to observe 51 or more times the 6 when rolling 235 times.

The binomial distribution gives the probability for exactly *k* 6s for n trials:

$$P(k = 51; n = 235, p = 1/6) = 0.0086$$

The probability we're interested in:

$$P(k \geq 51) = \sum_{k=51}^{n} P(k; n = 235, p = 1/6) = 0.0265$$

Hypothesis testing

That is, with a probability of p=0.026 we will observe 51 or more times the 6 when rolling a fair dice 235 times.

Therefore, for this die, we reject the null hypothesis that it is a fair die.

Hypothesis testing

How do we do this?

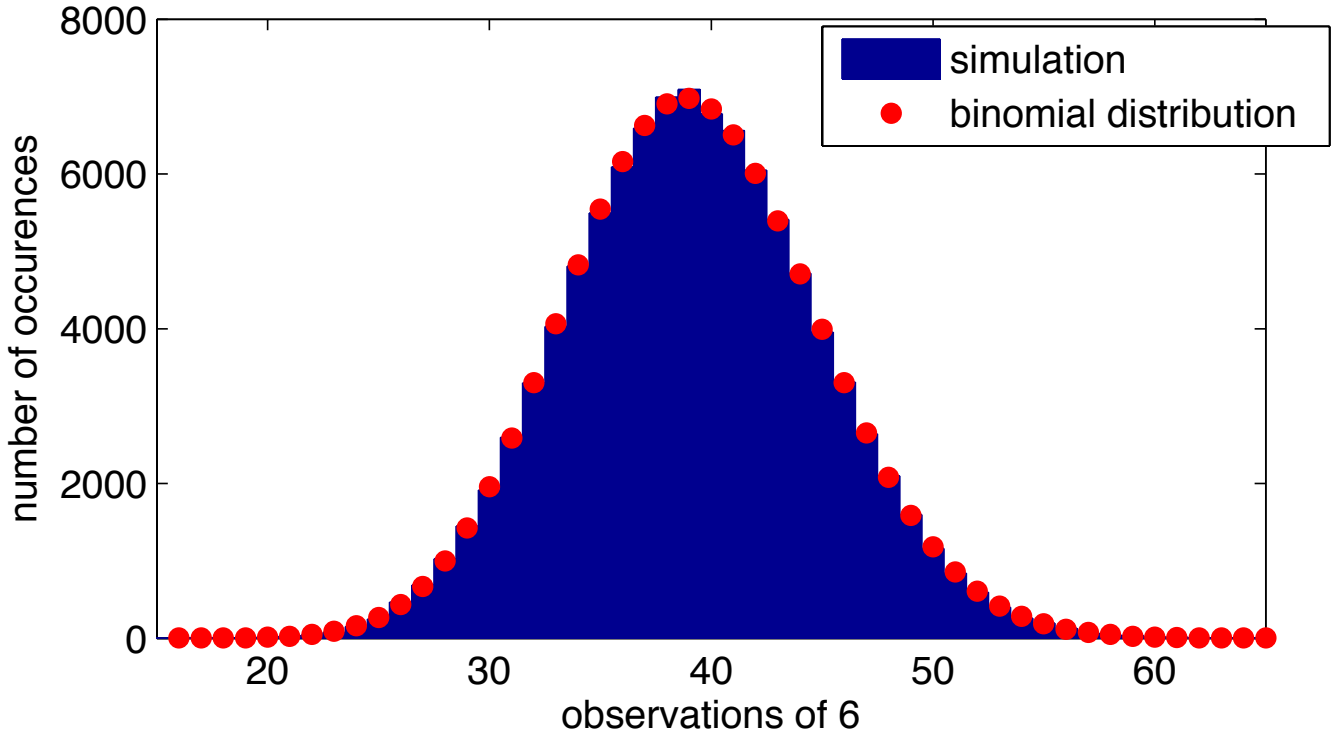
Matlab (using the cumulative binomial distribution):

```
p=1-binocdf(51-1,235,1/6)
```

Alternative (e.g., if we don't know the distribution): numerical simulation!

```
n=100000;
rslt=zeros(1,n);
for i=1:n,
    rslt(i)=sum(randi(6,1,235)==6);
end
p=sum(rslt>=51)/n
```

Hypothesis testing




```
set(gca,'fontsize',12)
hist(rslt,[1:235])
hold on
plot(0:80,binopdf(0:80,235,1/6)*n,'or','markerface','r')
xlim([15 65])
xlabel('observations of 6')
ylabel('number of occurences')
legend('simulation','binomial distribution')
hold off
```

Hypothesis testing

The idea behind null hypothesis testing:

In Null Hypothesis Significance Testing, after collecting data, a researcher computes the value of a summary statistic such as t or F or χ^2 and then determines the probability that so extreme a value could have been obtained by chance alone from a population with no effect if the experiment were repeated many times. If the probability of obtaining the observed value is small (e.g. $p < 0.05$), then the null hypothesis is rejected and the result is deemed significant.

Kruschke JK (2010) What to believe: Bayesian methods for data analysis.
TICS 14:293-299

See also: Drummond GB & Vowler SL (2012) Different tests for a difference: how do we do research?
J Physiol 590: 235–238.

Hypothesis testing

1) Define a clear hypothesis.
2) Define a significance level *before* doing an experiment, e.g., α=0.05 means that the null hypothesis will be rejected, if $p<\alpha$.
3) Decide how many subjects to test *before* starting the experiment.

	H_0 is true	H_1 is true
Do not reject H_0	Right decision	Wrong decision Type II error
Reject H_0	Wrong decision Type I error	Right decision

Hypothesis testing

Defining the *significance level* controls the probability to reject the null hypothesis given that it is true.

We declare that we are willing to reject the true null hypothesis $100\alpha\%$ of the time.

Defining the *power* $1\text{-}\beta$ controls the probability to reject the null hypothesis given that it is false.

Curran-Everett D (2009) Explorations in statistics: hypothesis tests and *P* values.

Adv Physiol Educ 33: 81–86

Some simple examples

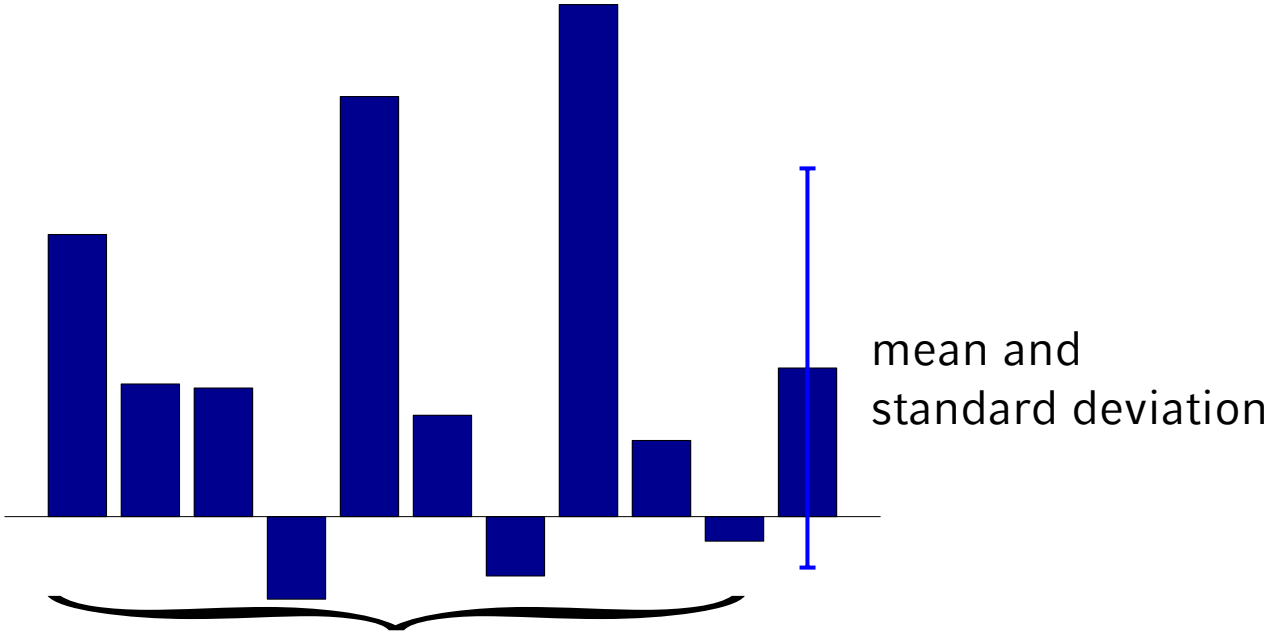
Testing whether

- a measured value is different from zero

- value A is different from value B

- two methods to measure a value yield different results

Measurement



We measure a certain variable ten times.

Is the mean different from zero?

T-test

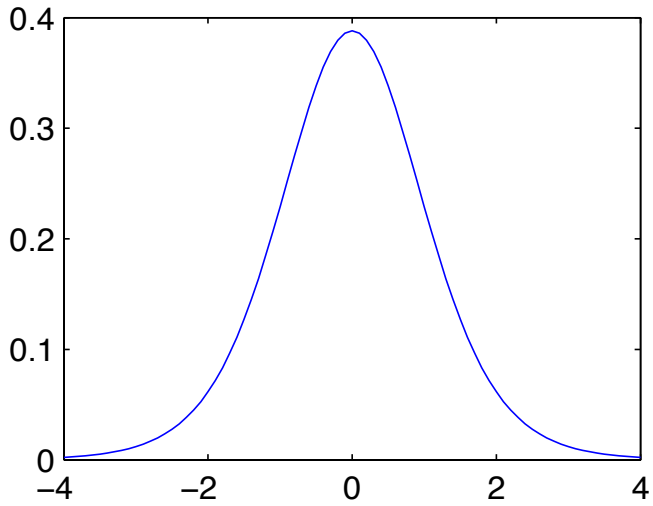
The one-sample t-test is used for to answer this question.

We first compute the t-statistic:

$$t = \frac{mean(x)}{standarderror(x)} = \frac{\mu}{\sigma/\sqrt{n}}$$

with n being the number of measurements, μ the mean and σ the standard deviation of the measurements.

We then calculate the probability that such a t (or a more extreme one) is found by chance. We use Student's t distribution, which has one parameter, the degree-of-freedom df=n-1.



The actual test result

In Matlab, this is done by a dedicated function:

```
[h,p]=ttest(x)
```

In our example, p=0.0429. That is, given a significance level of 0.05, we can reject the null hypothesis that the mean is zero.

Test without the dedicated function:

```
t=sqrt(numel(x))*mean(x)/std(x);
p=2*(1-tcdf(t,numel(x)-1))
```

Bootstrap instead of t-test

We can also approach this problem from a different perspective: given that our sample reflects the population, what's the distribution of the mean?

We do this by a bootstrap method that yields a confidence interval:

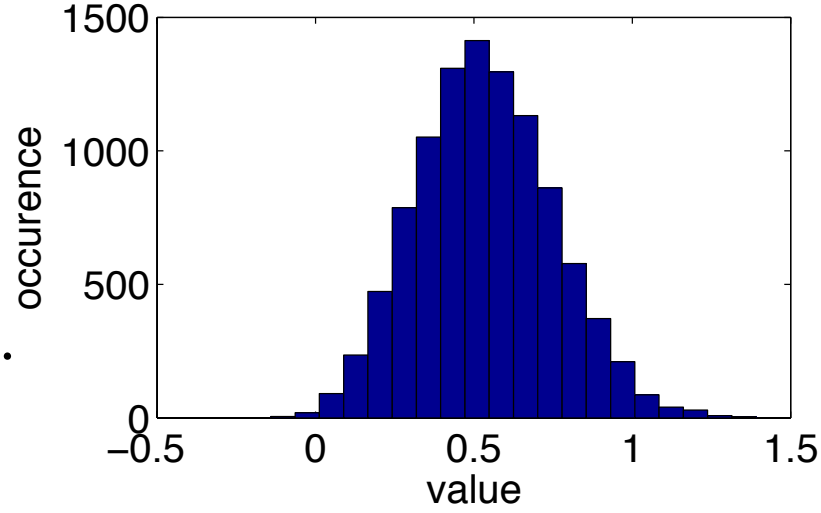
```
bootci(10000,{@mean,r},'type','per‘)
```

The 95% confidence interval is [0.132 0.981] and does not include zero. The p value is p=0.012 (computed without using bootci).

Curran-Everett D (2009) Explorations in statistics: the bootstrap. *Adv Physiol Educ* 33: 286-292

Bootstrap instead of t-test

The basic bootstrap method: randomly draw samples (with replacement) from you data sample and estimate its statistics.



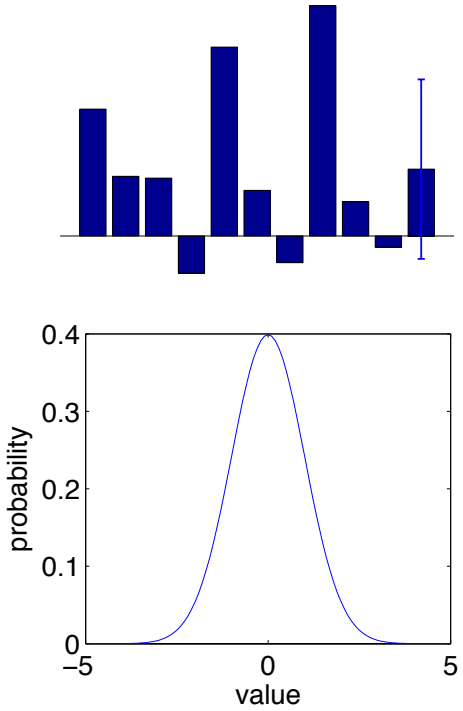

```
n=10000;
m=zeros(1,n);
for i=1:n,
   m(i)=mean(randsample(r,numel(r),true));
end
mest=mean(m);
pest=(sum(m<=0)+sum(m>=2*mest))/n;
ms=sort(m);
ci=[ms(n*5/100) ms(n*95/100)]
```

Curran-Everett D (2009) Explorations in statistics: the bootstrap. *Adv Physiol Educ* 33: 286-292

T-test: prerequisites

The one-sample t-test is applicable, if the sample mean is normally distributed, the sample variance is χ^2 distributed, and sample mean and variance are independent. This is the case, if the measurements come from a normal distribution.

But what if we know that this is not the case?



Alternative: the non-parametric Wilcoxon-Mann-Whitney test

```
p=signrank(x)
```

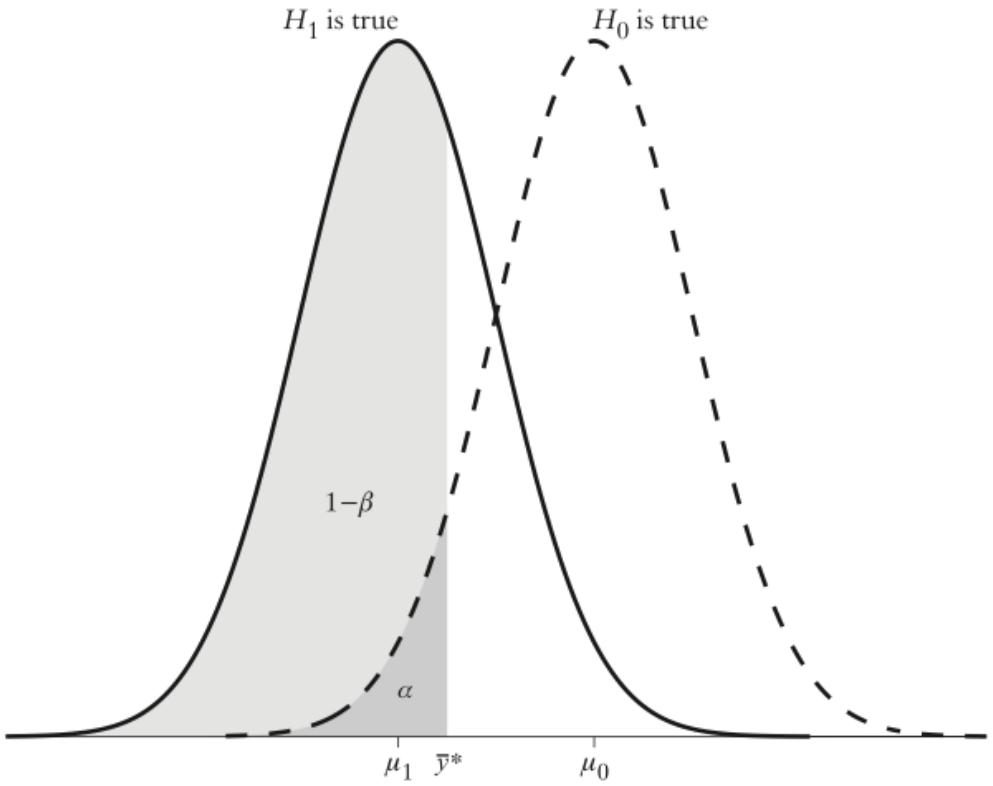
But see also:

Drummond GB & Tom BD (2011) Statistics, probability, significance, likelihood: words mean what we define them to mean. J Physiol 589: 3901–3904.

McElduff F, Cortina-Borja M, Chan SK, Wade A (2010) When t-tests or Wilcoxon-Mann-Whitney tests won't do. Adv Physiol Educ 34: 128–133.

The power

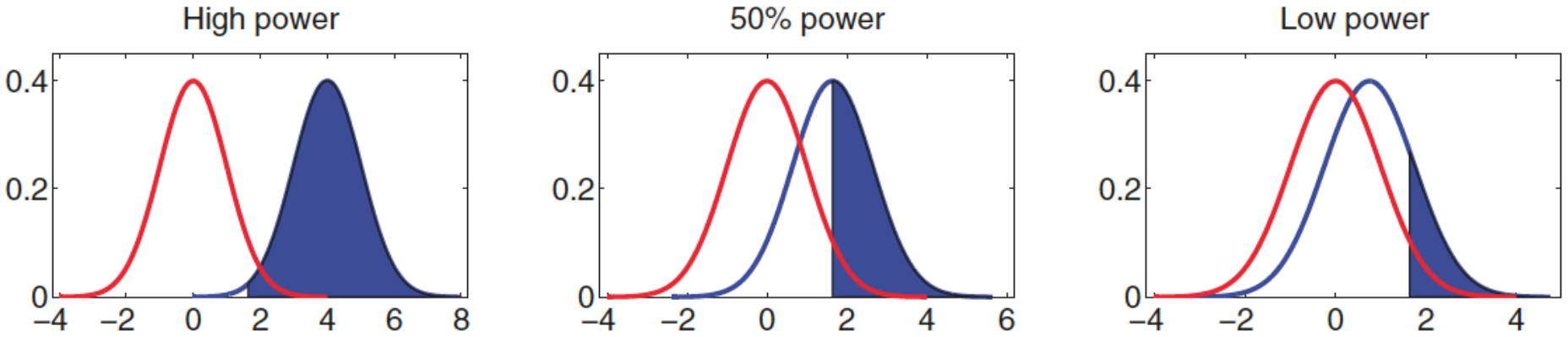
The *power* 1-β controls the probability to reject the null hypothesis given that it is false.



Curran-Everett D (2010) Explorations in statistics: power. *Adv Physiol Educ* 34: 41–43

The power

The *power* 1-β controls the probability to reject the null hypothesis given that it is false.



Tools for power calculation:

G*Power: tool for many statistical tests

http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3

Faul F et al (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. Behavior Research Methods 39:175-191

fMRIPower: tool specifically for fMRI experiments

http://fmripower.org

Mumford JA (2012) A power calculation guide for fMRI studies. SCAN 7:738-742

The power

Example:

Null hypothesis: the mean is $\mu_0=0$

Alternative hypothesis: the mean is $\mu_1=0.5$

The *effect size* is $|\mu_0-\mu_1|/\sigma$.

If we know that $\sigma=1.0$, then we can calculate the power of the test given n observations (significance level 0.05):

```
power=sampsizepwr('t',[0 1],0.5,[],10)
```

Or we can calculate how many observations we need for a power of 90%:

```
n=sampsizepwr('t',[0 1],0.5,0.9)
```

Here the result is n=44, that is, we need 44 (and not 10) observations for a power of 90%.

Curran-Everett D (2010) Explorations in statistics: power. *Adv Physiol Educ* 34: 41–43

The positive predictive value

The lower the power of a study, the lower the probability that an observed effect that passes the required threshold of claiming its discovery actually reflects a true effect.

The positive predictive value (PPV) is the probability that a 'positive' research finding reflects a true effect (that is, the finding is a true positive):

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

R: the pre-study odds (that is, the odds that a probed effect is indeed non-null among the effects being probed).

Remark: odds = p/(1-p)

Button KS et al. (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14: 365-376

What do I report in my paper?

- Analyze your data using the appropriate statistical procedures and identify these procedures in your manuscript: *Guidelines 2–4*.
- Report variability using a standard deviation, not a standard error: *Guideline 5*.
- Report a precise *P* value and a confidence interval when you present the result of an analysis: *Guidelines 6–10*.
- If in doubt, consult a statistician when you design your study, analyze your data, and communicate your findings: *Guideline 1*.

Curran-Everett D & Benos DJ (2004) Guidelines for reporting statistics in journals published by the American Physiological Society. Physiol Genomics 18: 249 –251.

What do I report in my paper?

Guideline 5. Report variability using a standard deviation. Because it reflects the dispersion of individual sample observations about the sample mean, a standard deviation characterizes the variability of those observations.

Curran-Everett D & Benos DJ (2004) Guidelines for reporting statistics in journals published by the American Physiological Society. Physiol Genomics 18: 249 –251.

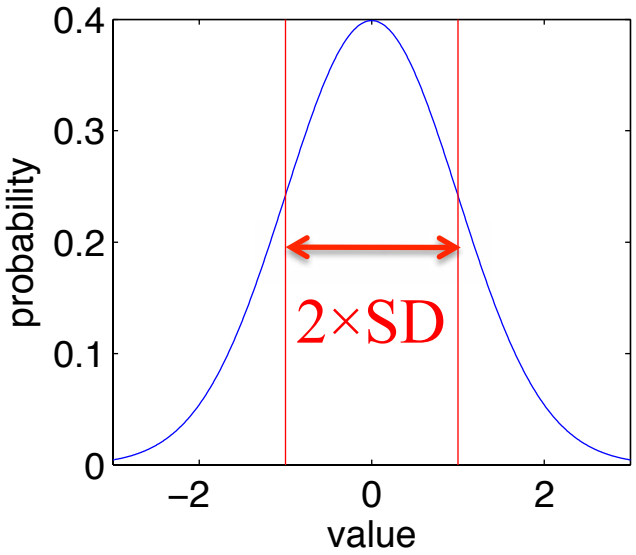
The policy of the BMJ and many other journals is to remove ± signs and request authors to indicate clearly whether the standard deviation or standard error is being quoted. All journals should follow this practice.

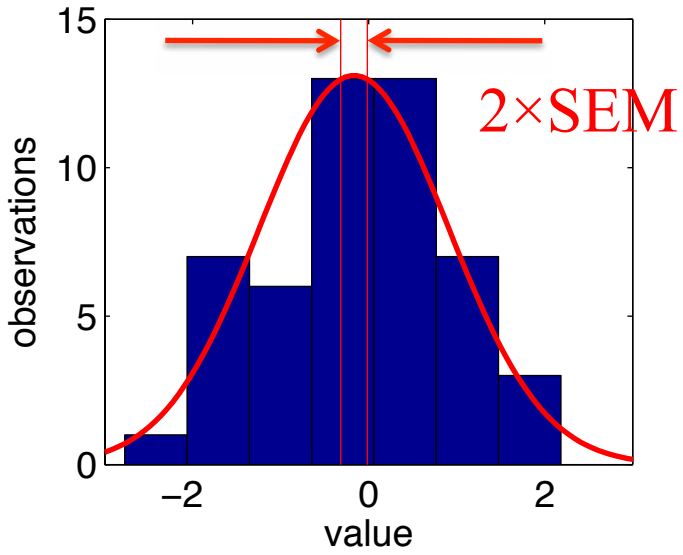
Altman DG & Bland MJ (2005) Standard deviations and standard errors. BMJ 331:903

Standard error and standard deviation

The Standard Deviation (SD) reflects the variability of the underlying probability distribution.

The Standard Error of the Mean (SEM) reflects the variability of the estimated mean of the distribution (decreases with the number of observations).





... a standard deviation estimates the variability among sample observations whereas a standard error of the mean estimates the variability among theoretical sample means.

Curran-Everett D (2008) Explorations in statistics: standard deviations and standard errors. Adv Physiol Educ 32: 203-208.

Confidence interval

Why confidence intervals?

Drug	$\bar{y}$	s	P	Conf Int
A	0.797	0.702	0.005	0.36 to 1.23
B	0.008	0.007	0.005	0.004 to 0.01
C	0.797	2.106	0.14	−0.51 to + 2.10

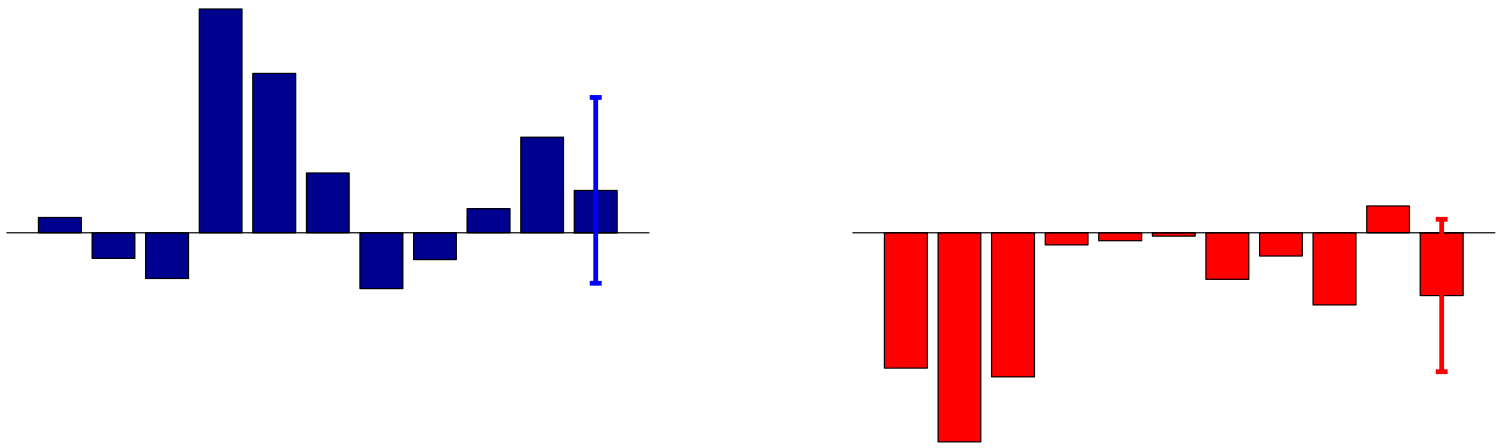
Drug A has an 80% effect (significant with p=0.005), the confidence interval suggests that the true effect is between 36% and 120%, which is scientifically meaningful.

Drug B has an effect of 1%, also significant with p=0.005, with a true effect between 0.4% and 1%, which is scientifically without importance.

Drug C has also an effect of 80%, but not significant (p=0.14), with a true effect between -51% and 210%. The true effect might thus be of scientific importance and the drug bears further study using a larger sample size.

Curran-Everett D (2009) Explorations in statistics: confidence intervals. Adv Physiol Educ 33: 87-90.

Two measurements



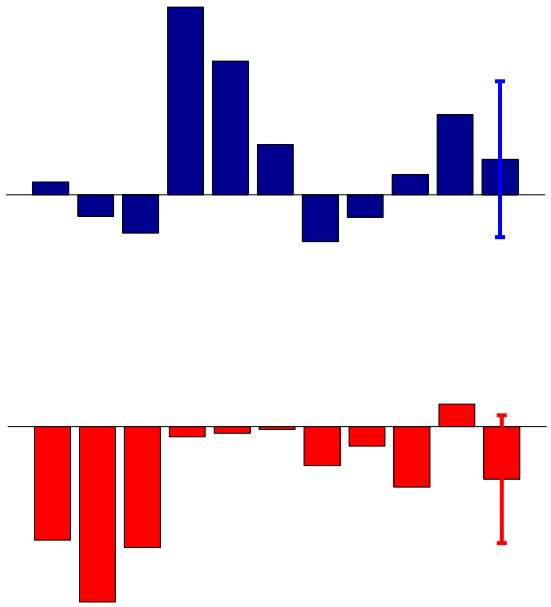
Are the means of these two variables different?

And what's the difference to the previous question?

Paired and unpaired

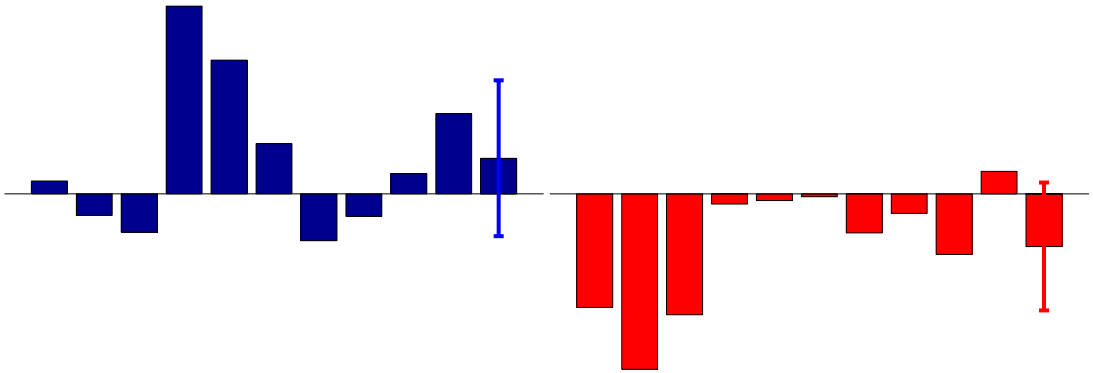
Paired measurements

e.g. two variables measured in ten subjects



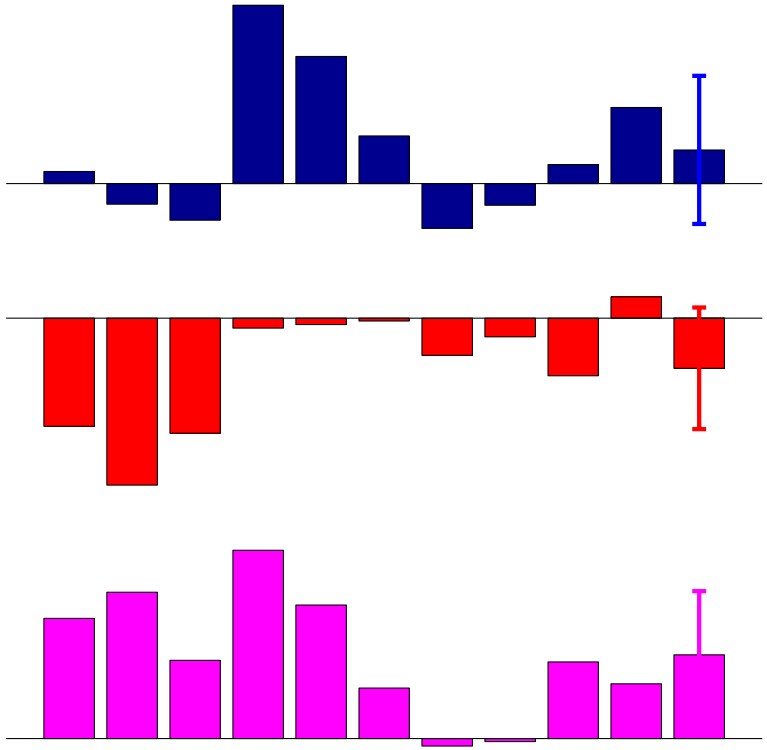
Unpaired measurements

e.g. one variable measured in two groups of ten subjects each



Paired and unpaired data require different treatment

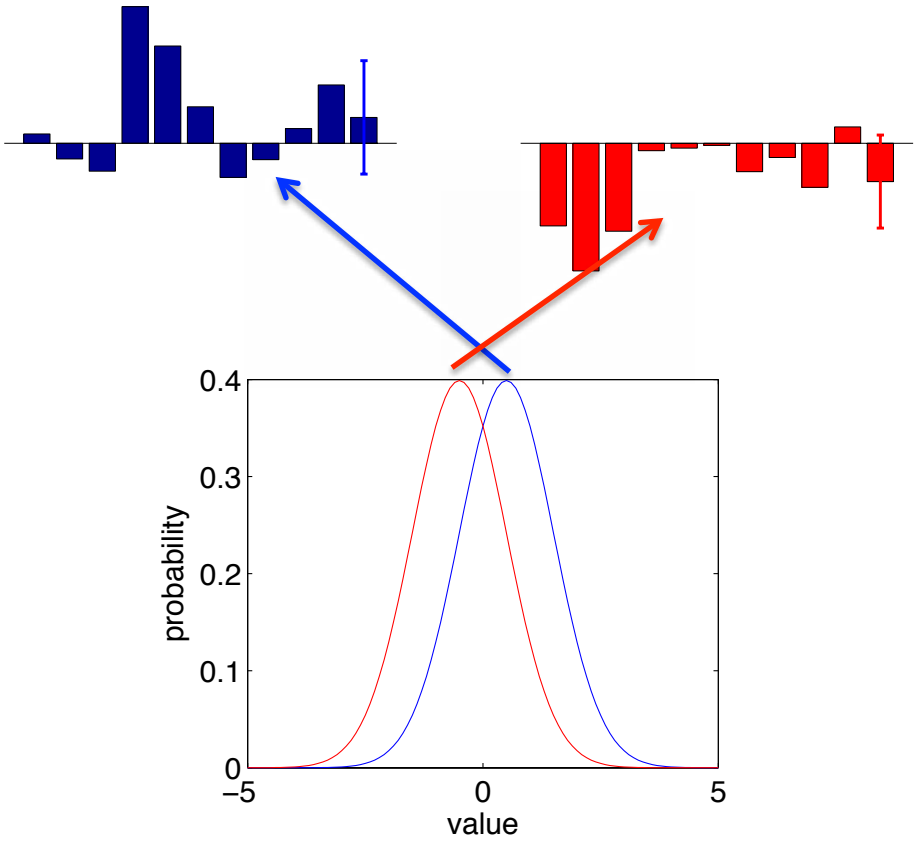
Paired measurements



With paired data, the null hypothesis is that the average *difference* of the measurements is equal to zero.

The t-test is performed on the differences.

Unpaired measurements



With unpaired data, the null hypothesis is that the two population means are equal.

The t-test differs depending on whether sample size or SD are equal for both groups.

How to do the t-test

Unpaired t-test

```
[h,p]=ttest2(x1,x2)
```

Result: p=0.0127

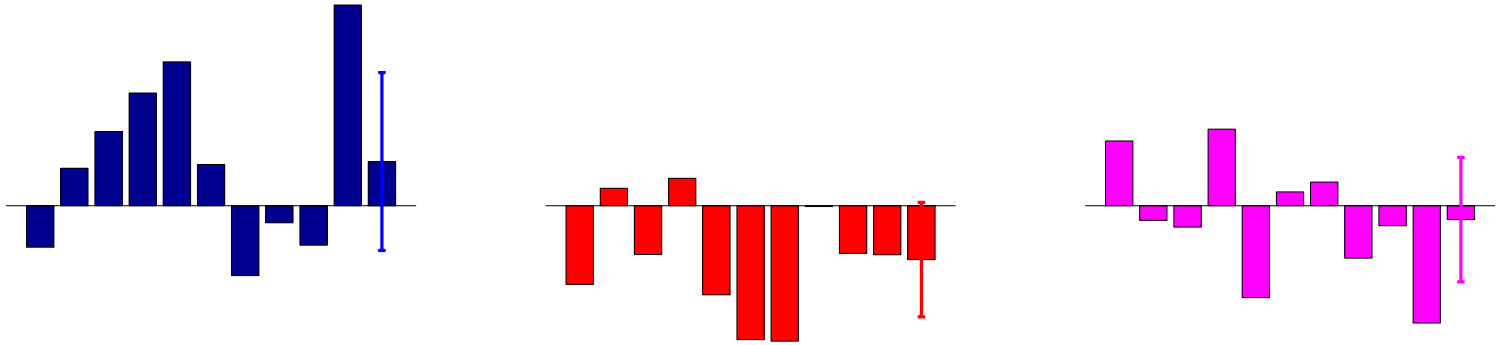
Paired t-test

```
[h,p]=ttest(x1,x2)
```

Result: p=0.0024

The paired t-test often has a better p value than the unpaired test, because the individual differences rather than the difference between population means is assessed.

More than two measurements



With more than two measurements, we would like to know whether any of two means are different. In other words, the null hypothesis states that all means are the same.

Multiple t-tests are not recommended.

Multiple tests increase the chance of falsely rejecting the null hypothesis.

Instead, we use an ANOVA (Analysis of Variance).

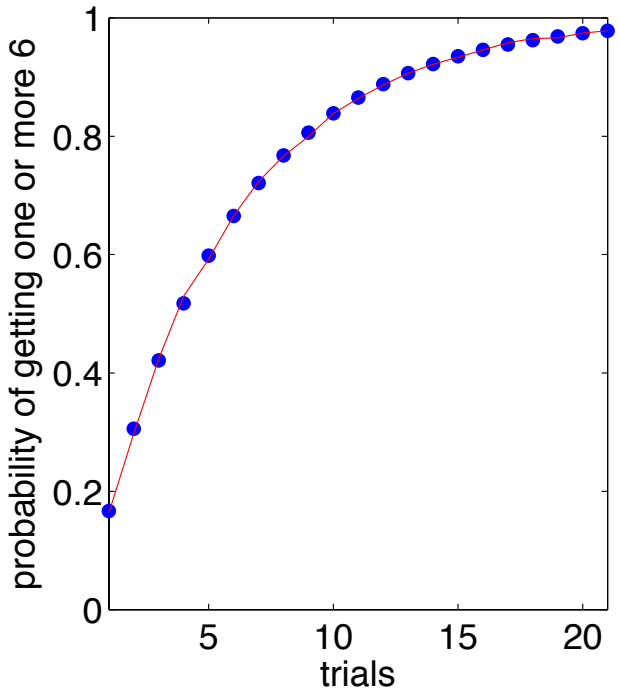
Why is multiple testing a problem?

The problem can be illustrated with a simple example.

Throwing a fair die will give a 6 in 1/6 of all cases.

But if we throw the die k times, the probability of getting one or more times the 6 is much higher than 1/6!

P(getting one or more 6 in k trials)
$=1-(1-1/6)^k$



Why is multiple testing a problem?

The error rate α for a single comparison does not change.

What changes is the *family-wise error rate* α_F.

Consider a single comparison with error rate

$$P(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$$

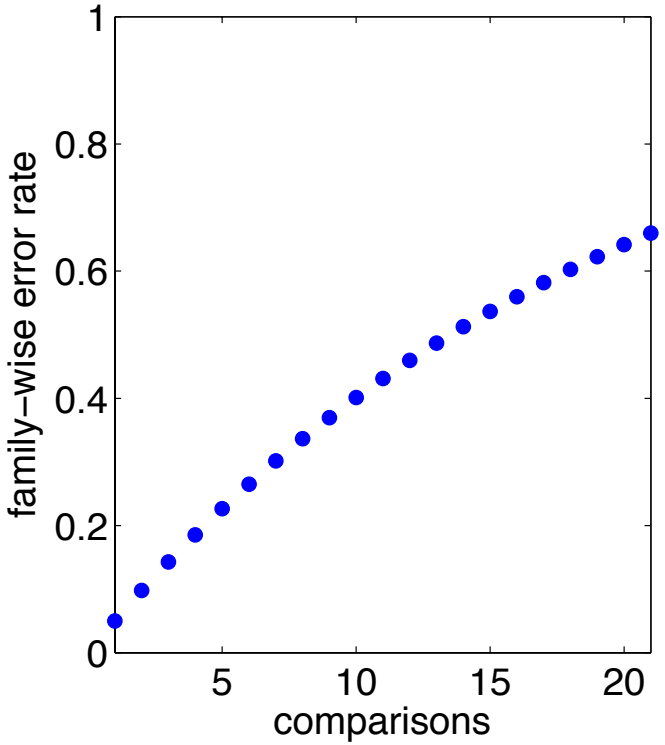
If we perform k such comparisons, what's the probability that we erroneously reject the null hypothesis in one case? With

$$\alpha = P(\text{reject } H_0 | H_0 \text{ is true}) =$$
$$1 - P(\text{fail to reject } H_0 | H_0 \text{ is true}) = 1 - (1-\alpha)$$

it follows that the family-wise error rate is

$$1 - P(\text{fail to reject all } H_0 | \text{all } H_0 \text{ is true}) =$$
$$1 - (1-\alpha)^k = \alpha_F$$

Note: for n groups, we have n!/(2(n-2)!) comparisons!



Curran-Everett D (2000) Multiple comparisons: philosophies and illustrations. Am J Physiol Regul Integr Comp Physiol 279:R1-8.

Drummond GB Vowler SL (2012) Type I: families, planning and errors. J Physiol 590:4971–4974.

Keselman HJ, Miller CW, Holland B (2011) Many tests of significance: new methods for controlling type I errors. Psychol Methods. 16:420-431.

Solutions to multiple testing

1) Don't do multiple testing. Use family-wise 'omnibus' testing procedures such as ANOVA.

2) Control for the *family-wise error rate* α_F.

E.g., Bonferroni correction: From $1-(1-\alpha)^k=\alpha_F$, we can calculate the adjusted significance level α required to keep the family-wise error rate at a constant level: $\alpha_{adj} = 1-(1-\alpha_F)^{1/k} \cong \alpha_F/k$

This helps to avoid false-positive results, but we may now miss cases in which the null hypothesis is not 'true' (more false negatives).

3) Control for the rate at which false positive conclusions are likely.

E.g. False Discovery Rate or similar procedures

Curran-Everett D (2000) Multiple comparisons: philosophies and illustrations. Am J Physiol Regul Integr Comp Physiol 279:R1-8.
Drummond GB Vowler SL (2012) Type I: families, planning and errors. J Physiol 590:4971–4974.
Keselman HJ, Miller CW, Holland B (2011) Many tests of significance: new methods for controlling type I errors. Psychol Methods. 16:420-431.

Pitfalls of fMRI analysis

fMRI is specifically prone to the problem of multiple testing, since in a standard analysis about 45.000 time courses are tested for each participant.

Analysis software such as SPM corrects for this.

BUT there are other problems. Some examples:

1) Low statistical power due to low sample size (Button et al. 2013; Yarkoni 2009)
2) Incorrect analysis of interactions (Nieuwenhuis et al. 2011)
3) Circularity errors (Vul & Pashler 2012)
4) …

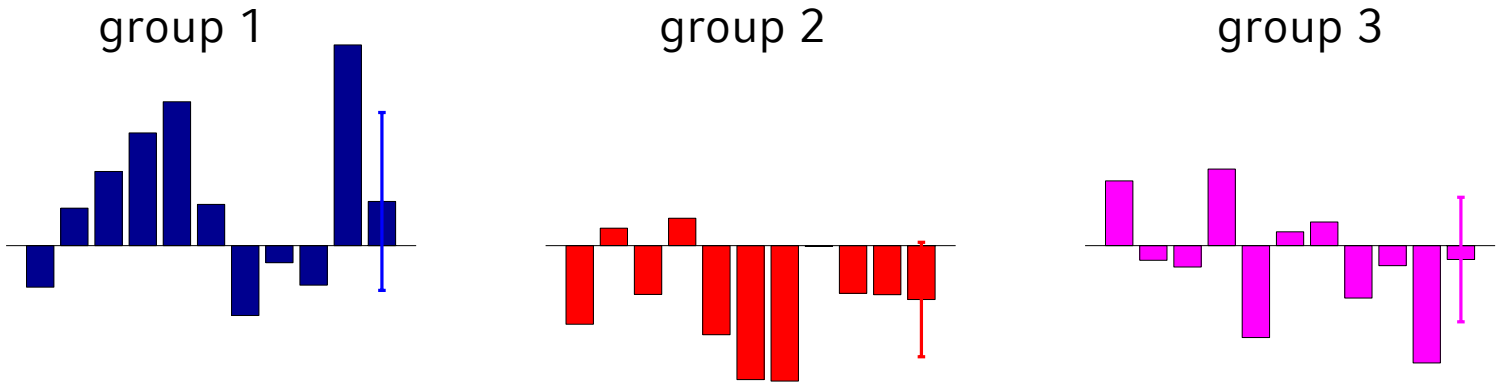
Button KS et al. (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14:365-76.

Nieuwenhuis S et al. (2011) Erroneous analyses of interactions in neuroscience: a problem of significance. Nat Neurosci. 2011;14:1105-7

Vul E, Pashler H. Voodoo and circularity errors. NeuroImage 2012;62:945–948.

Yarkoni T (2009) Big correlations in little studies: Inflated fMRI correlations reflect low statistical power. Perspect Psychol Sci. 2009;4:294-8.

One-way ANOVA



Is there an effect of group?

Null hypothesis: all means are equal.

```
anova1([x1 x2 x3])
```

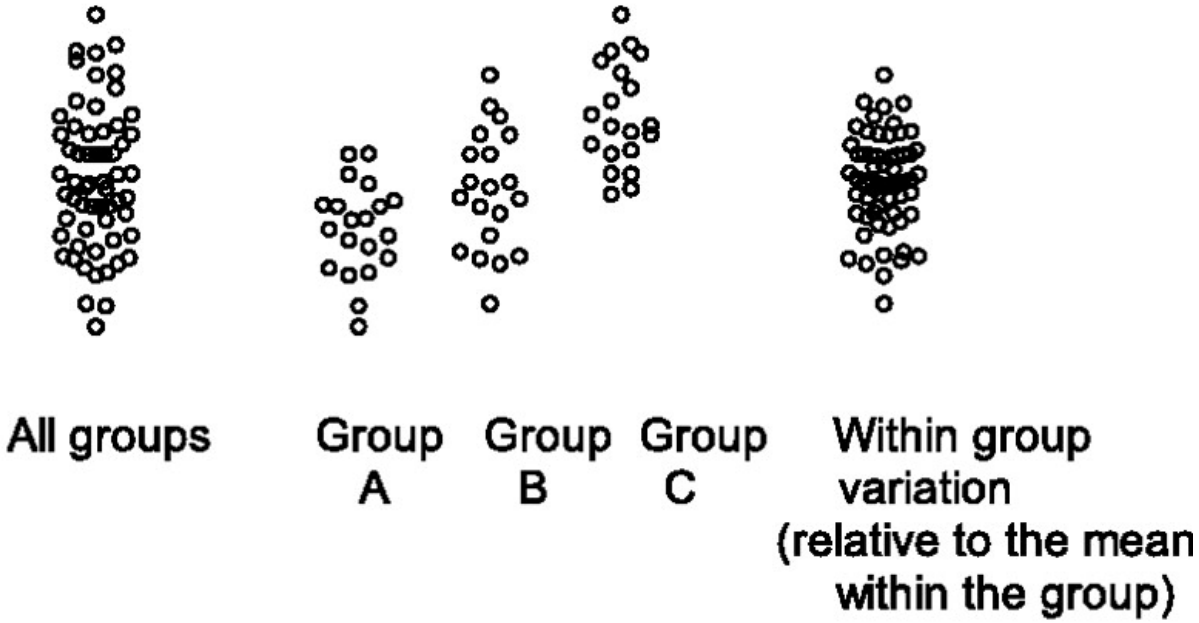
Source	SS	df	MS	F	Prob>F
Columns	9.8098	2	4.90491	4.84	0.0159
Error	27.3431	27	1.01271		
Total	37.1529	29			

What does that mean???

One-way ANOVA

The one-way ANOVA has one *factor* with k different *levels*.

The basic idea: if the overall within-group variation (relative to the mean of each group) is much smaller than the total variation, then the means are not all the same and the factor explains variation in the data.



Drummond GB Vowler SL (2012) Variation: use it or misuse it – replication and its variants. J Physiol 590:2539–2542.

One-way ANOVA

The one-way ANOVA has one *factor* with k different *levels*.

A one way ANOVA with 2 levels is equivalent to an unpaired t-test.

The test statistic of the ANOVA is the F value, which is F-distributed.

The F value for the ANOVA is the quotient of two variance terms (mean squares, MS), which are computed as quotient of the respective sum-of-squares (SS) and the degrees of freedom (df).

$$F = \frac{\text{explained variance}}{\text{unexplained variance}} = \frac{\text{between group variability}}{\text{within group variability}}$$

The F-distribution has 2 parameters, the 2 degrees of freedom of numerator and denominator.

The effect size η^2 is a standardized measure of the variance explained by the model. In our case, $\eta^2=9.8/37.2=0.26$.

One-way ANOVA

The one-way ANOVA implemented in Matlab

```
anova1([x1 x2 x3])
```

allowed only for balanced designs.

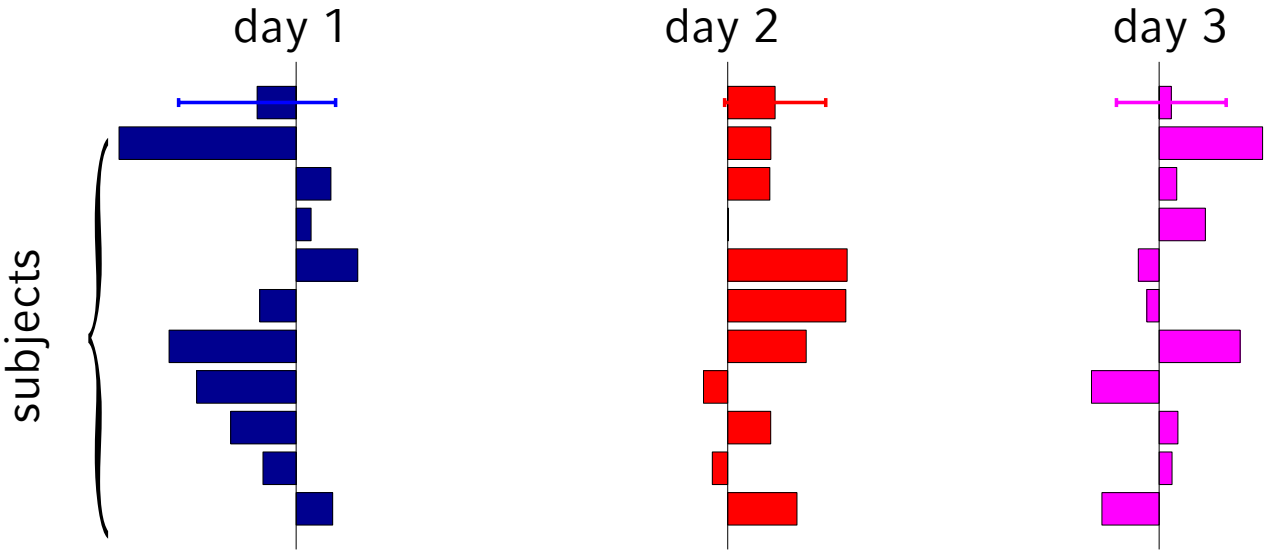
An alternative formulation allows for unbalanced designs (unequal number of samples per group). To perform this, define a group variable:

```
group=[zeros(10,1);ones(10,1);2*ones(10,1)];
anova1([x1; x2; x3],group)
```

Alternatively, we can use the much more versatile function `anovan`, which allows also for more complicated models:

```
anovan([x1; x2; x3],group)
```

Repeated measures ANOVA

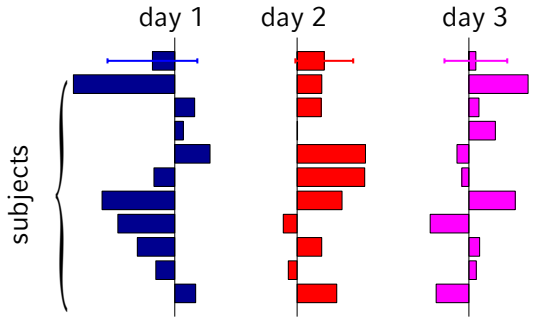


Is there an effect of ‘day’?

Source	SS	df	MS	F	Prob>F
Day	9.8098	2	4.90491	4.34	0.0289
Subject	7.0126	9	0.77918	0.69	0.7094
Error	20.3304	18	1.12947		
Total	37.1529	29			

Why is that different from the one-way ANOVA result?

Repeated measures ANOVA



The repeated measures ANOVA (rANOVA) looks at the *differences* between repeated measurements and accounts for the correlations between measures taken on the same individual.

Repeated measures ANOVA can be used with multiple factors, also between-subjects factors, which then divide subjects into groups.

rANOVA requires an additional prerequisite, the sphericity of the covariance matrix. Corrections or the *multivariate* repeated measures ANOVA can be used in case of sphericity violation.

Alternatives to rANOVA: hierarchical linear modeling, growth-curve modeling, mixed effects models.

Dijkers MP (2012) Chasing change: Repeated measures ANOVA is so yesterday! Arch Phys Med Rehabil. 2012 Nov 20.
Bagiella E, Sloan RP, Heitjan DF (2000) Mixed-effects models in psychophysiology. Psychophysiology, 37:13–20.

Repeated measures ANOVA

Doing repeated measures ANOVA in Matlab is not straight-forward, because there is no dedicated function for it.

However, we can use the `anovan` function for this. To we define two grouping variables, one is 'group', the other 'subject':

```
group=[zeros(10,1) (1:10)';ones(10,1) (1:10)';...
       2*ones(10,1) (1:10)'];
```

The actual call tells Matlab to use the 'subject' group as *random* factor, while 'group' is a *fixed* factor:

```
anovan([x1; x2; x3],group,'random',[2])
```

Bagiella E, Sloan RP, Heitjan DF (2000) Mixed-effects models in psychophysiology. Psychophysiology, 37:13–20.

Nonparametric alternatives

For the one-way ANOVA and the repeated-measures ANOVA, there are non-parametric alternatives.

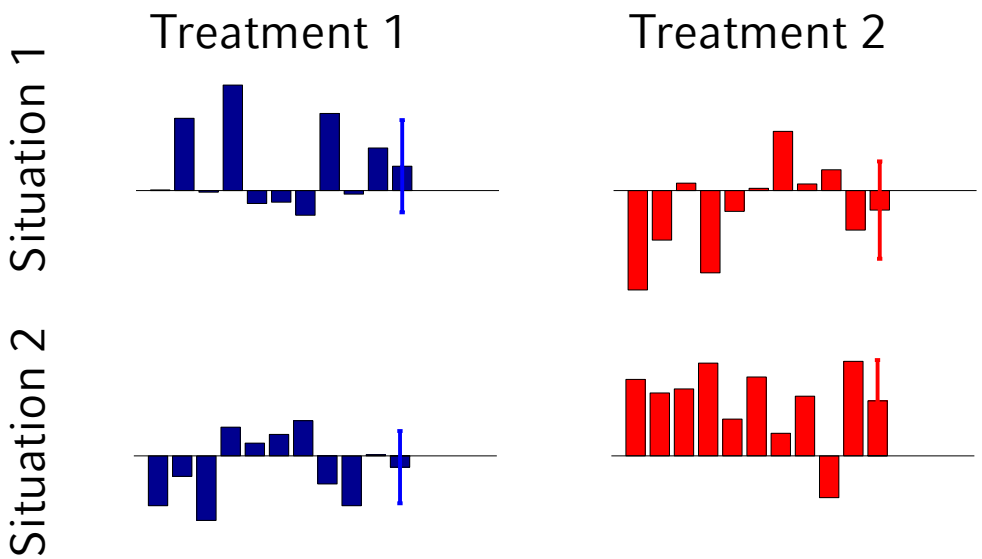
One-way ANOVA – Kruskal-Wallis test

```
kruskalwallis([x1; x2; x3],group)
```

Repeated measures ANOVA – Friedman's test

```
Friedman([x1 x2 x3])
```

Factorial ANOVA

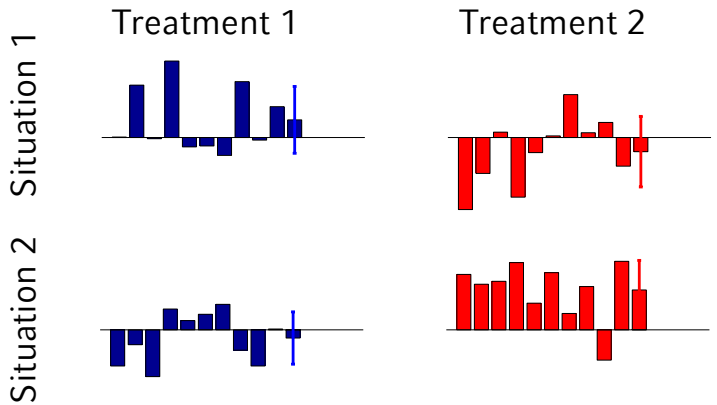


Several independent groups of measurements, which can be separated into factors.

Is there an effect of one of the factors?

Null hypothesis: all means are equal.

Factorial ANOVA

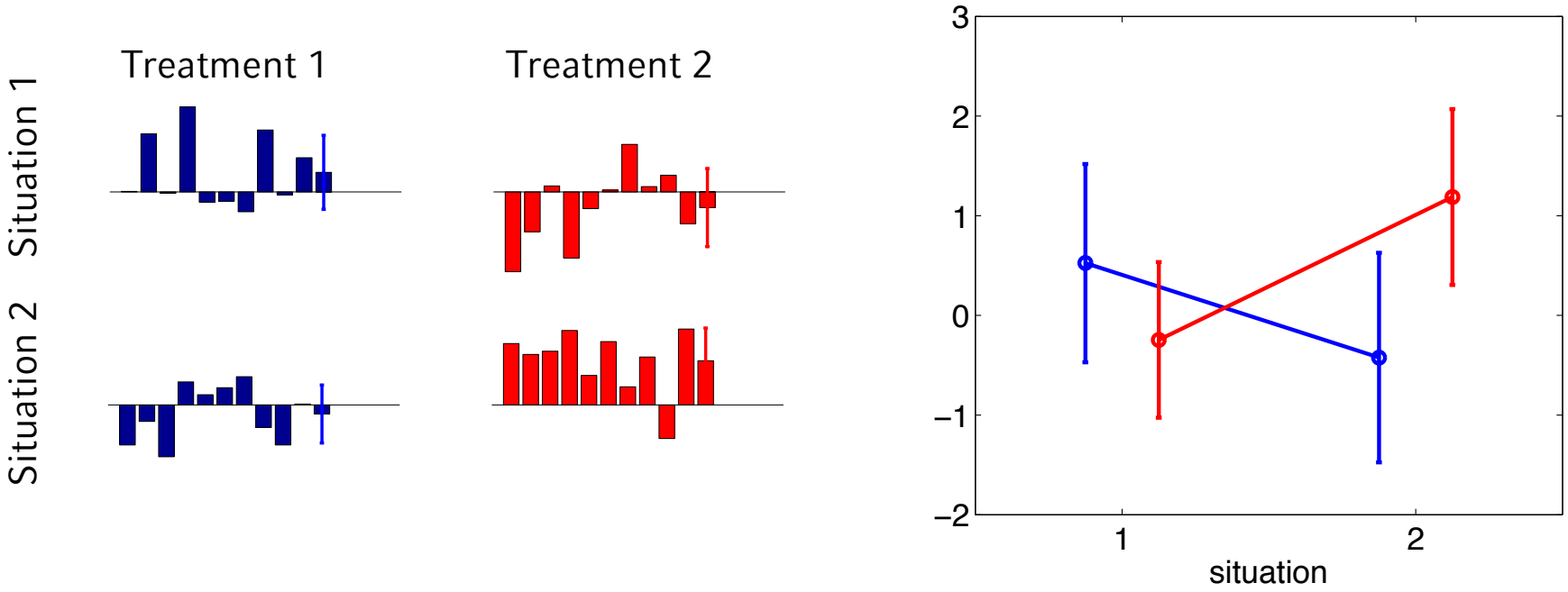


Two factors with 2 levels each.
10 measurements for each condition.
Balanced design.

```
anova2([r1 r2;r3 r4],10)
```

Source	SS	df	MS	F	Prob>F
Columns	0.5889	1	0.5889	0.68	0.4164
Rows	1.7676	1	1.7676	2.03	0.1629
Interaction	14.1726	1	14.1726	16.27	0.0003
Error	31.3613	36	0.8711		
Total	47.8904	39			

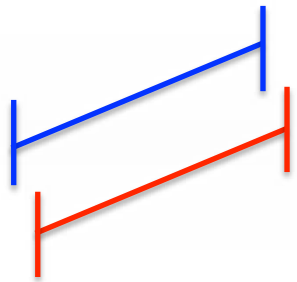
Factorial ANOVA



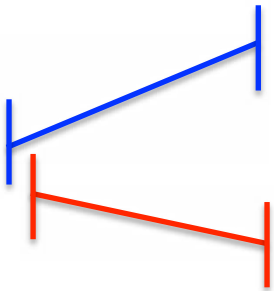
The interaction plot shows that the effect of treatment strongly differs depending on situation. This is an *interaction*.

The main effects are not significant (no difference for treatment when summed over situations and vice versa).

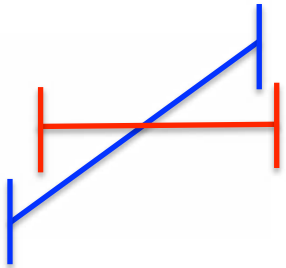
Factorial ANOVA



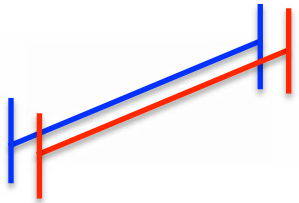
Main effect of treatment and situation.
No interaction.



Main effect of treatment.
Significant interaction.



Main effect of situation.
Significant interaction.

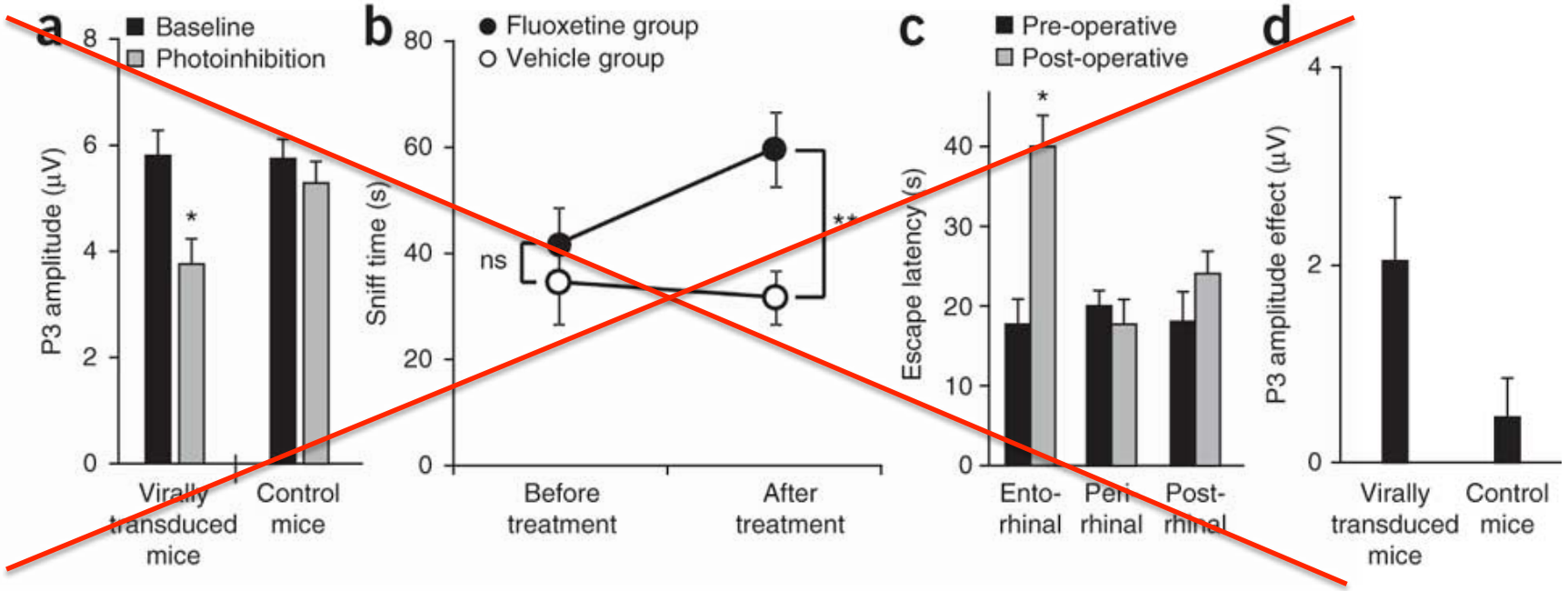


Main effect of situation.
No interaction.

Understanding the concept of main effects and interactions is essential for effectively using ANOVA.

The importance of interaction

"The percentage of neurons showing cue-related activity increased with training in the mutant mice ($P < 0.05$), but not in the control mice ($P > 0.05$)."



Comparing p-values is **not** the right thing to do to draw conclusions! Rather you have to test interactions.

Nieuwenhuis S, Forstmann BU, Wagenmakers EJ (2011) Erroneous analyses of interactions in neuroscience: a problem of significance. Nat Neurosci 14:1105-1107

Prerequisites of univariate ANOVA

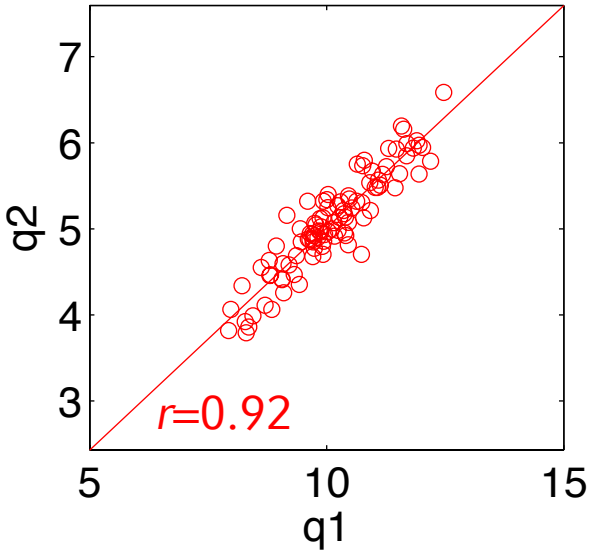
- Independent observations

- Normal distribution
(Shapiro–Wilk test, Lilliefors test, …)

- Homogeneity of variance
(Bartlett's test, Levene's test, ...)

- Sphericity
(Mauchly's test)
apply corrections: Greenhouse-Geisser, Huynh-Feldt, Lower-bound correction
or, if sample size is not small, use multivariate ANOVA (MANOVA)

Correlation



Correlation estimates the magnitude of a straight-line relationship between two variables.

The correlation coefficient ranges from -1 to 1.

Pearson's *r* is defined as:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

The variance explained by the corresponding regression is $R^2=r^2$.

Meaningful only when the two variables are true random variables.

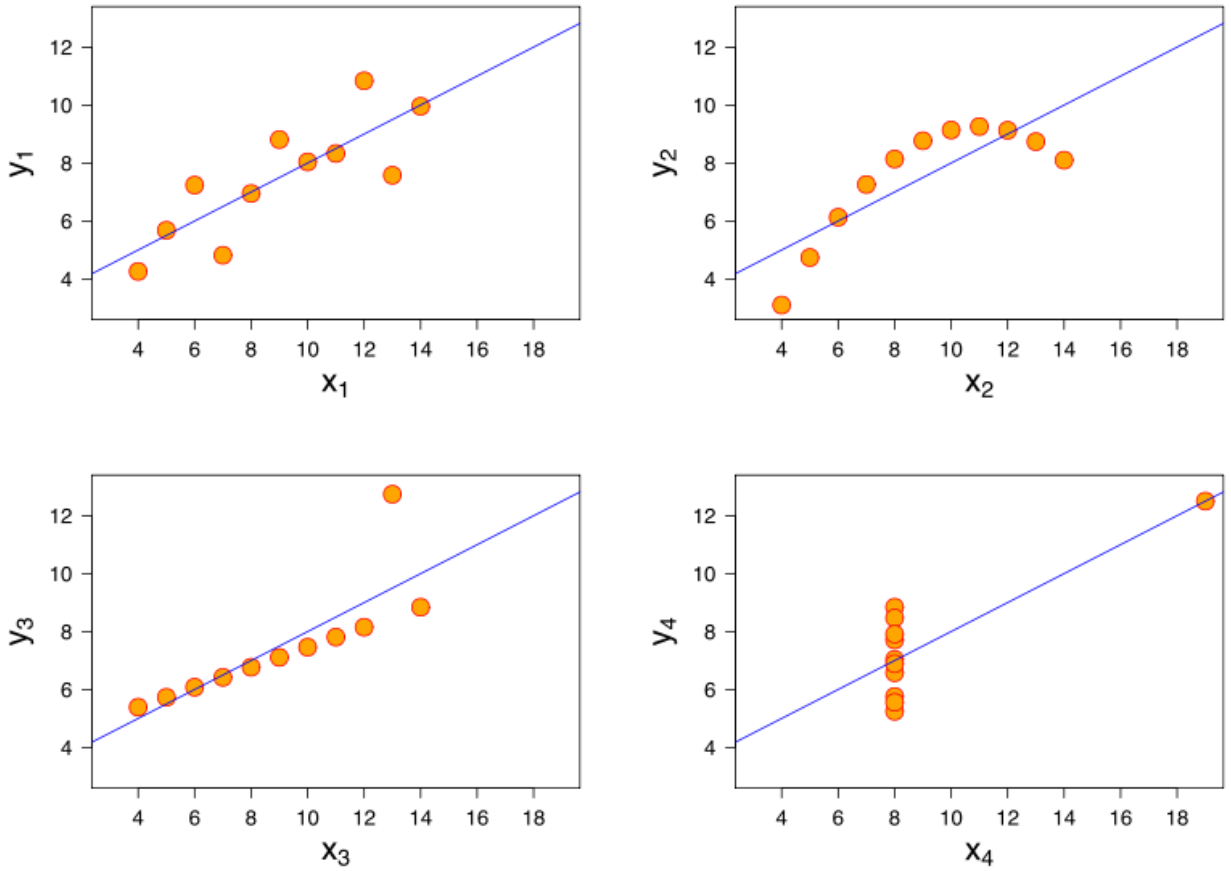
Cannot be used to make inferences about causality.

Does not give an indication of the slope (effect size).

Very sensitive to outliers.

Curran-Everett D (2010) Explorations in statistics: Correlation. Adv Physiol Educ 34:186-191.

Correlation

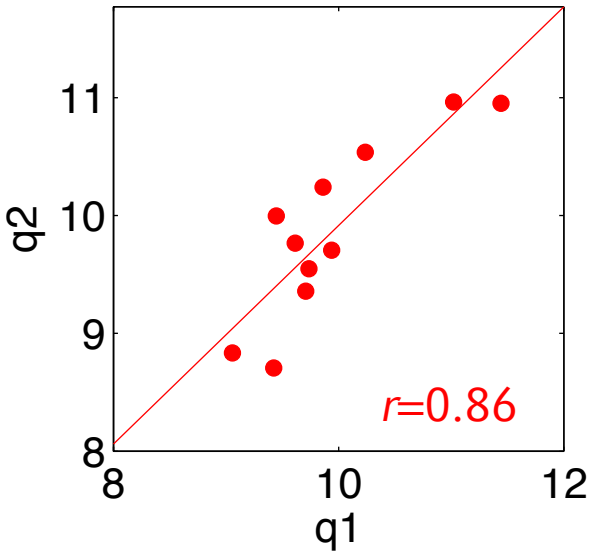


Anscombe's quartet: all four y variables have the same mean, SD, correlation coefficient, and regression line. Only for the upper-left relation the correlation is meaningful.

Always check with scatterplot whether it's meaningful!

Curran-Everett D (2010) Explorations in statistics: Correlation. Adv Physiol Educ 34:186-191.

Correlation: permutation method



Statistical test for significance

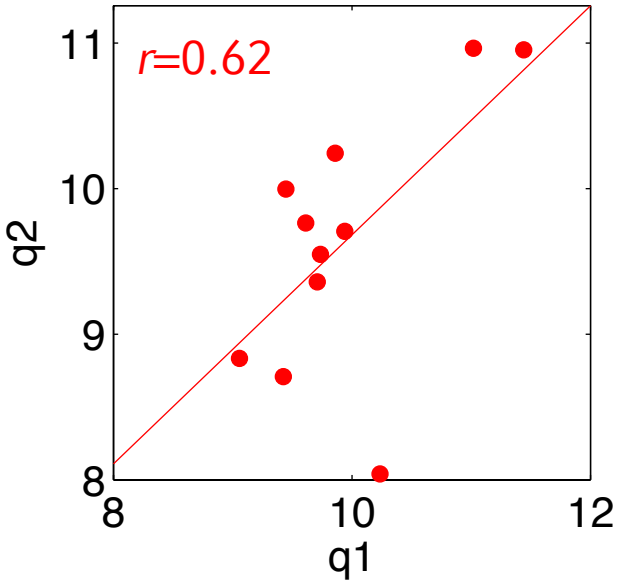
Null hypothesis: no correlation (r=0)

Test either with a t-test statistic (exact when both variables are normally distributed) or by numerical testing (permutation test).

For the permutation test, we take the actual data set and check how often we would find a |correlation coefficient| >|r| for all possible permutations of the y data (with x fixed). Since that's not practicable for large sample sizes (here we already have 10! =3628800 cases), we randomly select about 10000 of them.

Curran-Everett D (2010) Explorations in statistics: Correlation. Adv Physiol Educ 34:186-191.
Curran-Everett D (2012) Explorations in statistics: Permutation methods. Adv Physiol Educ 36:181-187.

Correlation: permutation method




```
[r p]=corr(x,y)
```

With the t-test statistic p=0.043.

The regression explains 38% of the variance.

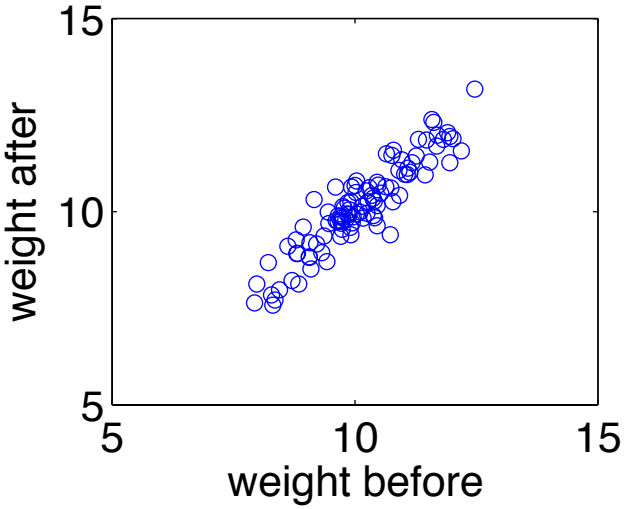
The permutation method:

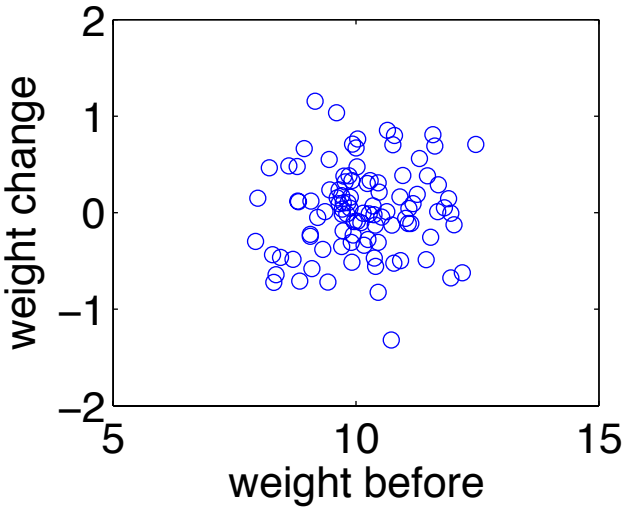
```
permcorr=zeros(10000,1);
for i=1:length(permcorr),
  permcorr(i)=corr(x,randsample(y,length(y),false));
end
p=sum(abs(permcorr)>abs(r))/length(permcorr)
```

We get p=0.041. That's a good match and validates the standard test.

Curran-Everett D (2010) Explorations in statistics: Correlation. Adv Physiol Educ 34:186-191.
Curran-Everett D (2012) Explorations in statistics: Permutation methods. Adv Physiol Educ 36:181-187.

Spurious Correlation





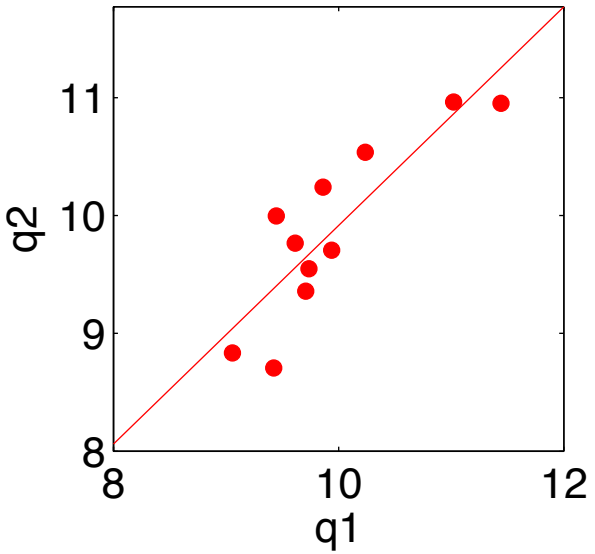
We measure the weight of an animal before and after an intervention and find a significant correlation (r=0.92).

This is due to mathematical coupling ($w_{post}=w_{pre}+\Delta w$) and is one example of spurious correlation.

„A correlation is not meaningful if x and y are related through computation."

Curran-Everett D (2010) Explorations in statistics: Correlation. Adv Physiol Educ 34:186-191.

Regression



In regression, we are interested in the relationship between two variables, not just in the correlation between them.

Regression can also handle more complex relations than correlation.

The idea underlying regression is that the variable Y can be explained by a sum of a scaled and shifted version of X and a normally distributed noise ε:

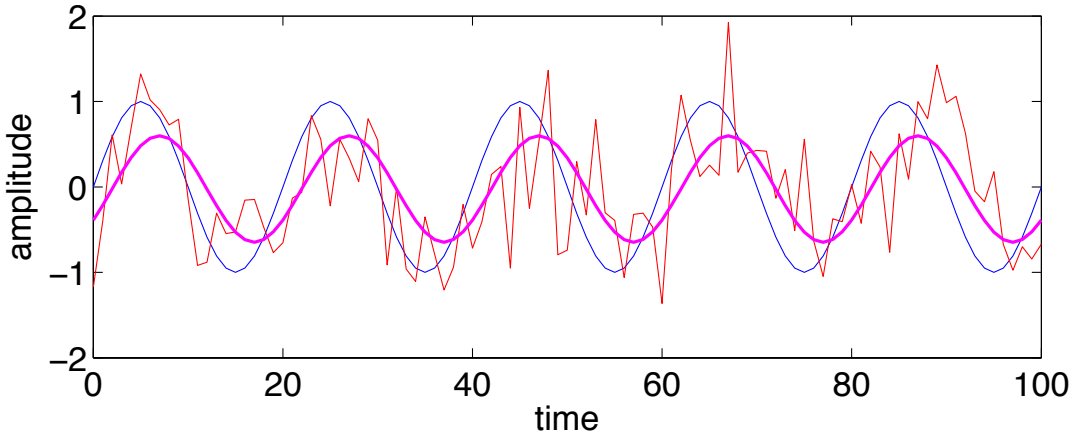
$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Regression estimates the parameters β of this relation.

Curran-Everett D (2011) Explorations in statistics: Regression. Adv Physiol Educ 35:347-352.

Example

We show a sinusoidally moving grating to our subject and see that the result is a sinusoidal eye movement with the same frequency – but shifted, smaller, and noisy. How can we estimate the gain (or amplitude) of this eye movement?



This problem can be solved with linear regression by transforming the variable X appropriately:

$$Y = \beta_0 + \beta_1 \sin(\omega T) + \beta_2 \cos(\omega T) + \varepsilon$$

Regression estimates the parameters β of this relation.

Least-squares estimation

$$\hat{y}_i = b_0 + b_1 \cdot x_i$$

We want to get the parameters that fit 'best'. 'Best' means that we *minimize* the least-squares distance between actual values and estimates.

$$\sum_i \left(\hat{y}_i - y_i\right)^2 = \sum_i \left(y_i - b_0 - b_1 \cdot x_i\right)^2 = \min$$

How is that done?

$$\Rightarrow \left\langle \begin{array}{c} 2 \cdot \sum_i \left(y_i - b_0 - b_1 \cdot x_i\right) = 0 \\ 2 \cdot \sum_i \left(y_i - b_0 - b_1 \cdot x_i\right) \cdot x_i = 0 \end{array} \right. \Rightarrow \left\langle \begin{array}{c} b_0 \cdot n + b_1 \cdot \sum_i x_i = \sum_i y_i \\ b_0 \cdot \sum_i x_i + b_1 \cdot \sum_i x_i^2 = \sum_i x_i \cdot y_i \end{array} \right.$$

$$\Rightarrow \begin{bmatrix} \sum_i 1 & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i \cdot y_i \end{bmatrix} \Rightarrow \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum_i 1 & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_i y_i \\ \sum_i x_i \cdot y_i \end{bmatrix}$$

Least-squares estimation

For the general case (more than one independent variable) we can write the regression as matrix equation:

$$\underline{\hat{y}} = X \cdot \underline{b}$$

Why?

$$\hat{y}_i = b_0 + b_1 \cdot x_{1i} + \cdots + b_m \cdot x_{mi}$$

$$\underline{\hat{y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix} \quad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{m1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{mn} \end{bmatrix}$$

$$\left| \underline{y} - \underline{\hat{y}} \right|^2 = \left(\underline{y} - \underline{\hat{y}} \right)^T \cdot \left(\underline{y} - \underline{\hat{y}} \right) = \min$$

$$\Rightarrow \underline{y}^T \cdot \underline{y} - 2 \cdot \underline{y}^T \cdot X \cdot \underline{b} + \underline{b}^T \cdot X^T \cdot X \cdot \underline{b} = \min$$

$$\Rightarrow -2 \cdot X^T \cdot \underline{y} + 2 \cdot X^T \cdot X \cdot \underline{b} = 0$$

$$\Rightarrow X^T \cdot \underline{\hat{y}} = X^T \cdot X \cdot \underline{b}$$

$$\Rightarrow \underline{b} = \left(X^T \cdot X \right)^{-1} \cdot X^T \cdot \underline{y}$$

Least-squares regression

The short version is: $\hat{\underline{y}} = X \cdot \underline{b} \quad \Rightarrow \underline{b} = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot \underline{y}$

Matlab provides a very convenient way of solving this:

```
b=X\y
```

This automagically uses the pseudo-inverse and computes the minimum-least-squares parameter estimate.

However, this simple calculation doesn't provide us with confidence intervals, test statistics, or other useful stuff.

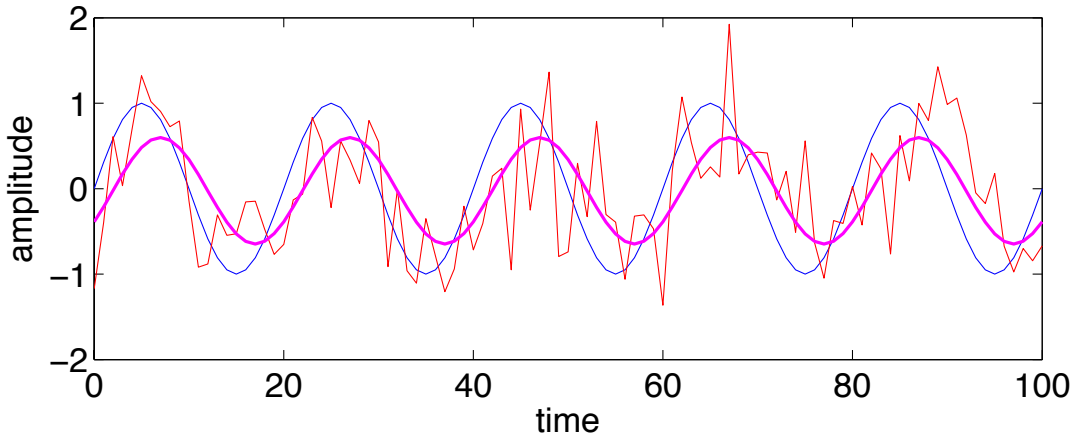
You can use another Matlab command for this:

```
[b,bint,r,rint,stats]=regress(y,[ones(size(x)) x])
```

The output stats contains the R^2 statistic, the F statistic and p value for the full model, and an estimate of the error variance.

Least-squares regression

The R^2 value is the percent-variance explained (corresponding to r for the 1D case), the F statistic and the corresponding p value are computed from the explained variance and the total variance (as in case of the ANOVA).



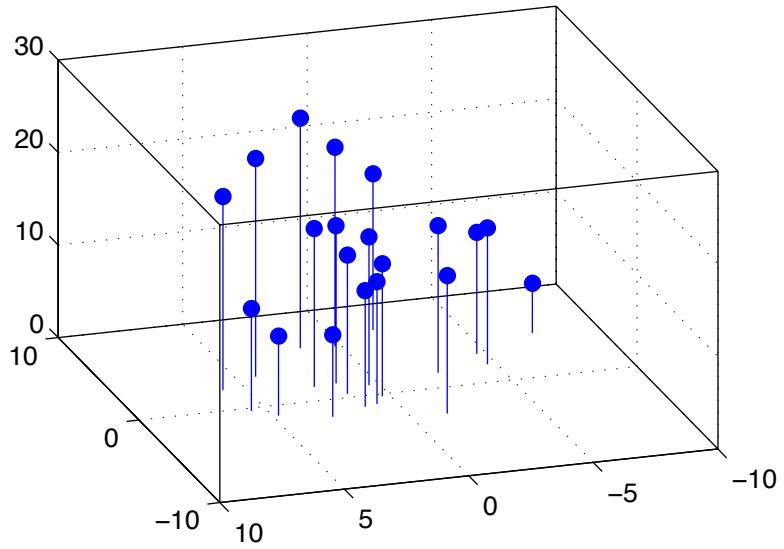
In our example, R^2=0.40 and F=33.1 with p<0.0001.

The parameter estimates are b_0=-0.02, b_1=0.51, b_2=-0.36.

Here, the confidence intervals tell us that the intercept b_0 is not different from 0, but b_1 is different from 1, and b_2 is different from 0.

Multidimensional example

In this example we observe one variable and want to see whether it depends on two other variables.

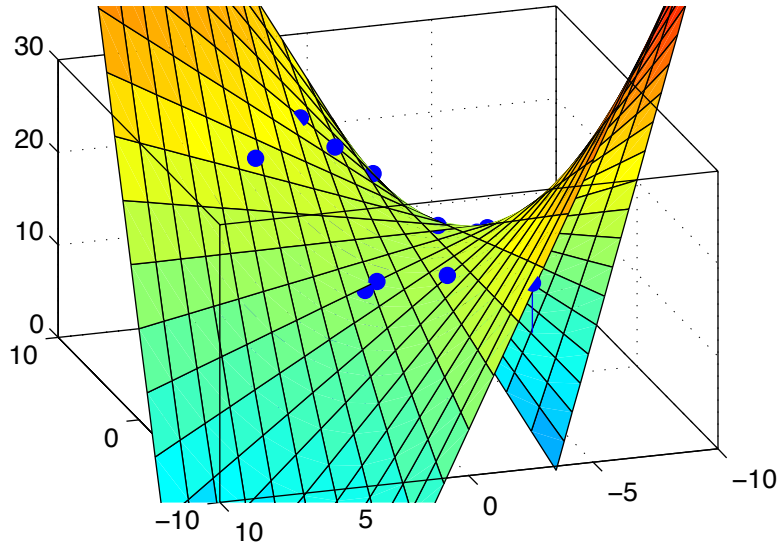


Evidently, this looks pretty random. But lets define a linear model with an interaction term:

```
regress(z,[ones(size(z)) xm ym xm.*ym])
```

Multidimensional example

This definition describes a surface in 3D space. The result yields the following model:



All the data fall very well on this twisted surface. From the parameter confidence intervals we can conclude that we found significant effects of both variables and their interaction.

Stepwise regression

Starting with the initial full model (all independent variables), we can remove variables which do not contribute sufficiently to the fit.

Backward stepwise regression: Remove the variable with the least significant partial r^2.

Forward stepwise regression: enter the variable which will contribute most in increasing R^2.

Best-subset regression: compute the regressions for all possible combinations and take the best model.

Criterion for the "best" model: R^2 cannot be used!!!
Instead, criteria are used which impose a penalty for higher number of parameters:
Smallest residual variance (or max. adjusted R^2), Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (BIC).

General Linear Model

The approach of the General Linear Model to ANOVA and to regression is the same: Define a model, and estimate the solution.

The general formulation includes fixed factors (β) and random factors (γ) in addition to the noise term.

$$\underline{y} = X \cdot \underline{\beta} + Z \cdot \underline{\gamma} + \underline{\varepsilon}$$

Random factors γ and the residual error ε have zero mean, so that $E(y)=X\beta$. If ε and γ have covariance matrices $R=I\sigma^2$ and G, then the covariance matrix of y is thus $V=ZGZ^T+R$. Thus, mixed models allow to define effects on variance or correlation.

Outlook

What I didn't talk about:

Outliers

Multivariate ANOVA and regression

More on the mixed model, such as effects over time

The Bayesian approach to statistics

Techniques such as cluster analysis, principal components analysis, independent component analysis, etc.

And a lot more …