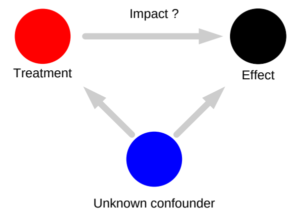
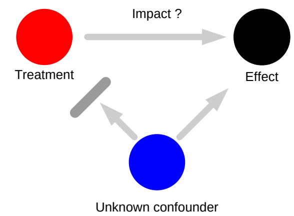
Voluntary and anonymous self assessment

"The goal of empirical research is - or should be - to increase our understanding of the phenomena, rather than displaying our mastery of technique." [1]

If you have no idea feel free to circle "?"

[1] Freedman, David A. (2006). S*tatistical Models Theory and Practice*. Cambridge University Press.





Types of study

Controlling the allocation of treatments through randomization	(Observational / Experimental / ?)
Controlling *** unknown **** confounders	(Observational / Experimental / ?)
Controlling only known confounders	(Observational / Experimental / ?)
Causal conclusions principally possible	(Observational / Experimental / ?)
Cohort study	(Observational / Experimental / ?)
Case-Control study	(Observational / Experimental / ?)
Matched Case-Control study	(Observational / Experimental / ?)
Most frequent type of study in social science	(Observational / Experimental / ?)
Most frequent type of study in natural science	(Observational / Experimental / ?)

Propensity score matching (psm)

Using psm changes the type of study from observational to experimental	(Yes / No / ?)
Because psm matches on (known / unknown / ?) confounders.	(Yes / No / ?)

p-Value (Fisher's hypothesis testing)

The p-value is (informally):

- Probability(H_0|Sample) (true / false / ?)
- Probability(Sample|H_0) (true / false / ?)
- A measure of compatibility between the sample and the hypothesis (true / false / ?)

Remark:
"Sample" should be more correctly read as: T(Sample) ≥ T(current sample)
Meaning "A test statistic (T) at least as extreme as the one that was actually observed "
Examples for test statistics are: mean, relative frequency, …

Testing H_0: Mean income of male = Mean income of female
yields p=0.95
Significance level is 0.05

We can conclude:

- Male and female mean income are equal (true / false / ?)
- Male and female mean income are equal with probability 0.95 (true / false / ?)
- Male and female mean income are equal with probability 0.05 (true / false / ?)
- The difference between male and female mean income is not significant (true / false / ?)
- We can not reject that male and female mean income are equal (true / false / ?)
- Data are compatible with the assumption that male and female mean income are equal (true / false / ?)

"not significant" of a test could have three main reasons

?

With large n even the smallest effect is significant (true / false / ?)

To overcome the problem of interpreting "non significant" test results one might

- avoid tests at all (true / false / ?)
- add effect-estimates (CI) to tests (true / false / ?)
- use equivalence tests in some cases (true / false / ?)
- perform Power-calculation before testing (true / false / ?)
- perform Power-calculation after testing (true / false / ?)

Necessary parameter(s) for a power calculation that are mostly estimated through a pilot study

- effect size (true / false / ?)
- significance level (true / false / ?)
- power (true / false / ?)
- standard deviation (true / false / ?)

Confidence interval (CI)

A 95% confidence interval (CI) for a parameter

- contains the parameter with 95% probability (true / false / ?)
- is a random interval which contains the parameter with 95% probability (true / false / ?)
- is (smaller / larger / ?) when the sample size increases*
- is (smaller / larger / ?) when the standard deviation increases*

* rest being equal

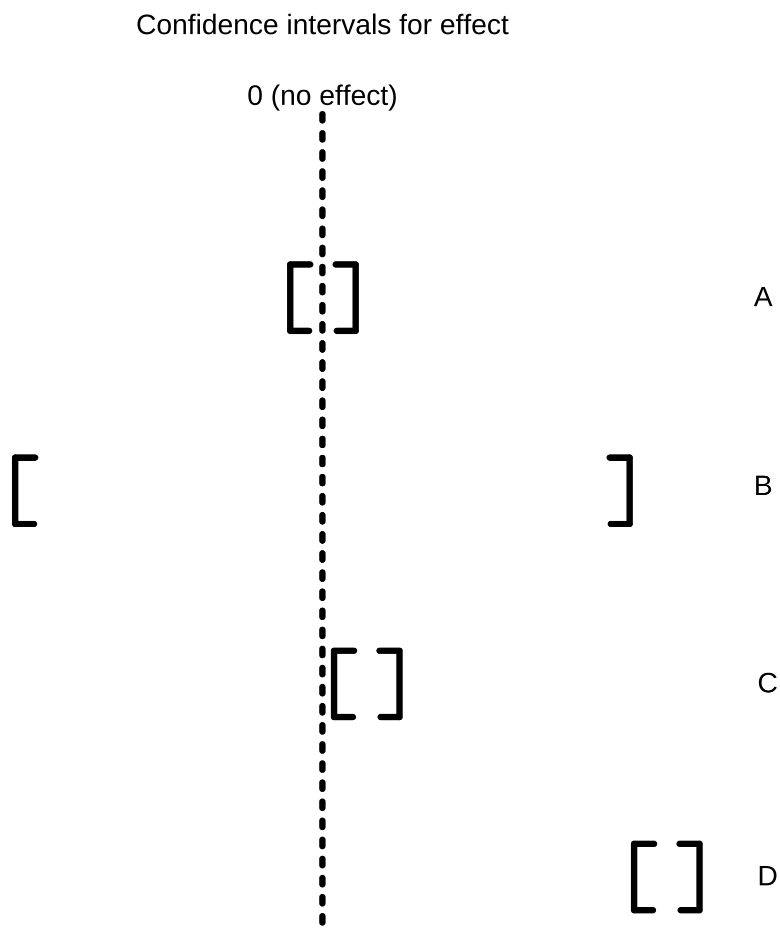
CI and test

- If the $CI_{1-\alpha}$ for an effect contains zero (0) then the respective two-sided test H_0: "effect=0" (is / is not / ?) significant at α

- If intervals $CI_{1-\alpha}$ overlap then the respective two-sided test H_0: "parameters are equal" (is / is not / ?) significant at level α

- If intervals CI1-α do not overlap then the respective two-sided test H0: "parameters are equal" (is / is not / ?) significant at level α

Interpret results from tests and estimations (confidence interval)



Briefly describe the cases and highlight in which case(s) you would recommend further research use (sig.; not sig.; relevant; not relevant; potentially relevant; typical if n small; typical if n big)

A____

B____

C____

D____

Further research with bigger n urgently recommended in _______

Most likely further research will show as irrelevant results as this (these) _______

A "dream result" in _______

Imagine 4 different results from an observational study on monthly income and gender

$$\text{Income} = \mu + \beta_1 * \text{male}$$

The variable "male" is coded 0/1

Result A

Sample size	=	30		
			p	**CI 95%**
$\hat{\mu}$	=	1,500.00 $	0.0305	[1,350; 1650]
$\hat{\beta}_1$	=	0.50 $	0.2406	[-0.10; 1.10]

Result B

Sample size	=	30		
			p	**CI 95%**
$\hat{\mu}$	=	1,500.00 $	0.0305	[1,350; 1650]
$\hat{\beta}_1$	=	50.00 $	0.2406	[-200; +300]

Result C

Sample size	=	10,000		
			p	**CI 95%**
$\hat{\mu}$	=	1,500.00 $	0.0001	[1,450; 1550]
$\hat{\beta}_1$	=	0.50 $	0. 0001	[0.10; 0.90]

Result D

Sample size	=	10,000		
			p	**CI 95%**
$\hat{\mu}$	=	1,500.00 $	0.0001	[1,450; 1550]
$\hat{\beta}_1$	=	50.00 $	0. 0001	[49; 51]

Give your judgment regarding the effect of gender on monthly income:

	Significance (level 0.05)		**Relevance**	
	significant:	+	relevant:	+
	not significant :	0	not relevant :	0
			potentially relevant:	!!!
A				
B				
C				
D				

In case of what result(s) would you

- initiate further research A / B / C / D / ?
- suggest that further research seems not reasonable A / B / C / D / ?
- be convinced that a relevant gender effect exist A / B / C / D / ?
- be convinced that no relevant gender effect exist A / B / C / D / ?

-

ANOVA / ANCOVA

Interactions (product terms)

- Interactions measure the biological (mechanical) interactions between variables (true / false / ?)
- Interactions account for effect-measure modifications rather than confounding (true / false / ?)
- Factors involved in interactions should be included into the model (hierarchy principle) (true / false / ?)

Effect = μ + Therapy + Gender + Therapy * Gender

Fig.1

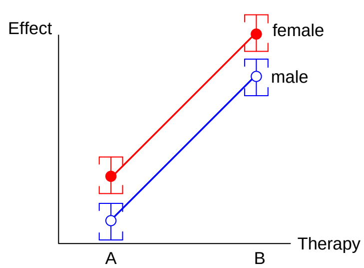


Fig.2

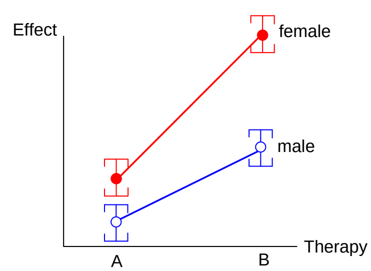
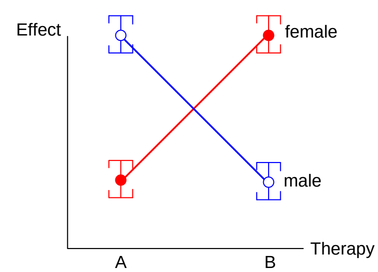


Fig.3



- Which case(s) show(s) significant interaction ? (Fig 1 / Fig 2 / Fig 3 / ?)
- In which case(s) would it make sense to discuss main effects ? (Fig 1 / Fig 2 / Fig 3 / ?)

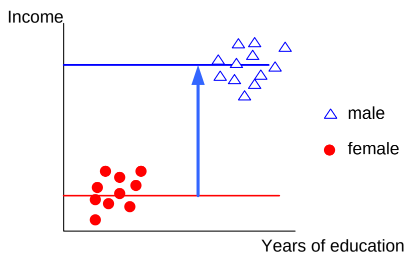
Hypothetical data

Fig. 4 Unadjusted mean income of male is substantially **greater** than mean income of female

Fig. 5 After the adjustment for education mean male income is **smaller** than mean income of female

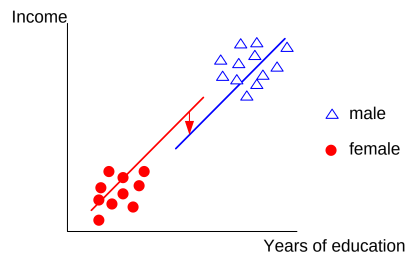
Fig. 6 Using a curvelinear model there is **no gender effect** in mean income

Fig. 4



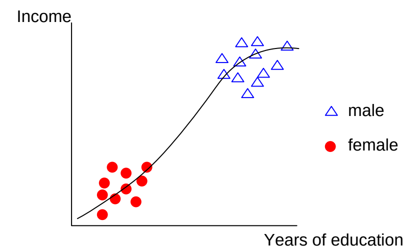
Income = µ + gender
ANOVA

Fig. 5



Income = µ + gender + ß*years of Edu.
ANCOVA or Regression

Fig. 6



Income = sigmoid(gender; years of Edu)
Curvelinear relation

The model related to

- Fig. 4 is appropiate (that's a nice example never to mix categorical and scale variables) (true / false / ?)
- Fig. 5 is appropiate (that's a nice example that one should always adjust for known confounders)(true / false / ?)
- Fig. 6 is appropiate (that's a nice example that simple is not always best) (true / false / ?)
- None (that's a nice example for a model artifact) (true / false / ?)

Modified from Rothman KJ (2012). Epidemiology: An Introduction,3rd edition. Oxford University Press.

Regression

In an observational study
an independent (right-side) variable is convincingly causal if

a) the associated parameter is significant (true / false / ?)
b) if a) and if the model adjusts for known confounders (true / false / ?)
c) if b) and if the semi parametric and consistent GEE-technique is used (true / false / ?)
d) if c) and if Granger's causality test is significant (true / false / ?)
d) it is consistent with background knowledge (true / false / ?)

Model selection

- Stepwise and best-subset regressions are generally accepted methods to detect the correct model (true / false / ?)
- The link-test, a test for omitted variables (Pregibon 1980), can be used to detect unknown confounders (true / false / ?)

- Model sensitivity analysis checks for insensitivity of estimates across acceptable models (true / false / ?)
- Model artifacts can be detected through factoring (true / false / ?)
- A variable should join a model equation if it is a significant (test) confounder (true / false / ?)
- A variable should join a model equation if it is a relevant confounder (true / false / ?)
- A relevant confounder changes the target parameter by at least 20% (rule of thumb) (true / false / ?)

A squared right-side variable like age²

- would change the linear model to a non-linear model (true / false / ?)
- is generally not a good idea (true / false / ?)
- is useful to estimate extrema (true / false / ?)

Collinearity

- The right-hand side variables must not be collinear (e.g. correlated) at all (true / false / ?)
- Result under extreme collinearity: Excellent fit but no variable shows a significant effect (true / false / ?)
- Drop variables with variance inflation factor (VIF) < 1.0 (true / false / ?)
- Counter measures depend on the degree of collinearity (true / false / ?)

Endogeneity * could be

- caused through an omitted confounder (true / false / ?)
- caused through measurement error (true / false / ?)
- caused through simultaneity (true / false / ?)
- "corrected" through instrumental variables (true / false / ?)

* Presence of an explanatory variable that is correlated with the error term.
See e.g. Wooldridge JM (2009). Introductory Econometrics, 4th ed. International Student Edition, South-Western.

An instrumental variable must be correlated

- to the original error term (true / false / ?)
- to the new error term (true / false / ?)
- to the endogenous variable (true / false / ?)

After adding an instrumental variable in the "endogeneous regression" y = ax + b on observational data

- the relationship between y and x becomes causal (true / false / ?)

If the independent (left-side) variable y in a multiple regression is a fraction (percentage) then

- y can be used untransformed (true / false / ?)
- logit(y) is more appropriate (true / false / ?)

log(salary) = 4.822 + 0.257*log(sales)

- 0.257 is the elasticity of salary with respect to sales (true / false / ?)
- 0.257 is the semi-elasticity of salary with respect to sales (true / false / ?)
- 1% increase in sales increases salary by about 0.257% (true / false / ?)
- salary increases by 25.7% for one unit change in sales (true / false / ?)

where log(x) is the natural logarithm of x

Which result(s) of a simple regression y = µ + ßx + u should be checked (n was 300)

- b=2 SE=0.1 p=0.95 (OK / suspect / ?)
- b=2 SE=1 p=0.05 (OK / suspect / ?)
- b=2 SE=10 p=0.05 (OK / suspect / ?)

Logistic regression

logit (Probability of success) = -11.77 + 0.12*math score + 0.98*female

the probability for success
- for female is 2% less than the probability for male (true / false / ?)
- increases by 12% for a one-unit increase in math score (true / false / ?)

the odds for success
- for female are 170% higher than the odds for males (true / false / ?)
- increases by 13% for a one-unit increase in math score (true / false / ?)

Remark: exp(0.12)=1.13 exp(0.98)=2.7
for all statements: holding the rest at mean

Interpretation of logistic regressions using the probability framework * is

- easier to communicate than traditional odds-ratio interpretation (true / false / ?)
- not appropriate when the prevalence is totally unknown (e.g. in some case control studies)(true / false / ?)
- appropriate as long as the prevalence is small (true / false / ?)

* Long JS, Freese J (2014). Regression for Categorical Dependent Variables Using Stata. 3rd ed. Stata Press.

ROC (Receiver operating characteristic)

- ROC is used to compare competing multinomial logistic models (true / false / ?)
- ROC is used to compare competing binary logistic models (true / false / ?)
- A model has a higher classification accuracy if its ROC-area is smaller (true / false / ?)
- A model with its ROC graph close to the diagonal is a good one (true / false / ?)