

2013 - Das Jahr der Statistik

Marloes Maathuis
Seminar für Statistik
ETH Zürich

300 Jahr nach "Ars Conjectandi" von Jakob Bernoulli



Basel: 1654 - 1705



1713

Was ist das Jahr der Statistik?

- Ziele:
 - “Increase public awareness of the power and impact of statistics on all aspects of society”
 - “Nurture statistics as a profession, especially among young people”
 - “Promote creativity and development in the sciences of probability and statistics”
- Website:
 - <http://www.statistics2013.org/>
(Posters, Videos, etc)

2013 INTERNATIONAL YEAR OF STATISTICS

RECOGNIZING THE CONTRIBUTIONS OF STATISTICS TO SOCIETY WORLDWIDE

Gesundheit

Wirtschaft und Politik

Biologie

Landwirtschaft und Umwelt

Sport

Naturkatastrophen



Es gibt immer mehr Daten: wir brauchen mehr Statistik!

- Hal Varian (Chief Economist at Google), Jan 2009:
 - “I keep saying the sexy job in the next ten years will be statisticians.”
 - “The ability to take data – to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it – is going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.”

Internationale Konferenz Ars Conjectandi

- Internationale Konferenz Ars Conjectandi 1713-2013
 - 15.- 16. Oktober 2013 in Basel
 - <http://www.statoo.ch/bernoulli13/>
- **Einladung zum öffentlichen Vortrag:**
 - Sprecher: **Gerd Bosbach**, Hochschule Koblenz
Ko-autor vom Buch „Lügen mit Zahlen. Wie wir mit Statistiken manipuliert werden“ (mit Jens Jürgen Korff)
 - Titel: **“Vorsicht Statistik”**
 - Zeit: **Mittwoch 16. Oktober**, 16.40 - 17.25
 - Ort: Congress Center Basel, Messeplatz 21, Basel

Inhaltsübersicht des Vortrags

- 2013 - Das Jahr der Statistik
- **Permutationstests / Randomisierungstests**
- Kausalität

Warum Randomisierungstests?

- Intuitiv, verständlich für Schüler
- Man kann alle wichtigen Konzepte eines statistischen Tests anhand eines Permutationstests erklären
- Wird häufig in der Praxis benutzt

Beispiel: Anzahl Stunden bis zum Abklingen einer Erkältung

- **Frage:**
Zeigen die Daten, dass ein Multi-Vitamin Präparat eine Erkältung verkürzt?

Können die Unterschiede zwischen den Gruppen als zufällig angesehen werden, oder ist ein systematischer Unterschied plausibler?

- **Applet**

Placebo	Multi-Vitamin
43.0	34.4
48.7	39.0
55.2	41.8
57.2	49.6
65.8	50.4
67.1	59.7
71.2	61.5
89.0	63.6
90.4	69.5
101.0	76.9
	78.3

(hypothetischer Datensatz)

Was haben wir gemacht?

- **Gedanken-Experiment: wir nehmen mal an, dass die Behandlung keine Auswirkung hat (*)**
 - Dann würden die Zahlen (Stunden bis zum Abklingen) sich nicht ändern, wenn wir die Behandlung wechseln
 - Wir schauen alle $\binom{21}{10} \approx 352716$ möglichen Gruppen-Zuteilungen an und rechnen jedes mal den Unterschied in Mittelwerten. Das ergibt eine Referenzverteilung.
 - Wir vergleichen den beobachteten Wert 12.06g zu dieser Referenzverteilung:
 - "passt gut" → unsere Annahme (*) könnte gut stimmen
 - "sehr extrem" → unsere Annahme (*) stimmt eher nicht

Formell: ein statistischer Test

- Daten:
 - Annahme: Gruppen gemacht durch Randomisierung
- Wir formulieren die Null- und Alternativ-Hypothesen:
 - H_0 : Die Behandlung hat keine Auswirkung (Null-Hypothese)
 - H_1 : Die Behandlung verkürzt die Erkältung (Alternativ-Hypothese; einseitig)
- Wir formulieren eine Teststatistik; soll extreme Werte annehmen, wenn die Alternativ-Hypothese gilt:
 - Unterschied in Mittelwerten (12.06g)
- Ziel:
 - Entscheiden zwischen H_0 und H_1

Fehler der 1. und 2. Art

- Wir haben nur eine zufällige Gruppenzuteilung und die dazugehörige Teststatistik.
Deshalb sind wir nie sicher, ob H_0 oder H_1 stimmt.
- Wir können zwei Arten von Fehlern machen:

Entscheidung Wirklichkeit	H_0	H_1
H_0	OK	Fehler der 1. Art (false positive)
H_1	Fehler der 2. Art (false negative)	OK

- Wir möchten Kontrolle über den Fehler der 1. Art:
 - $P(\text{Fehler der 1. Art}) \leq \alpha$ **(Signifikanzniveau / Niveau)**
 - Ein Test ist **exakt**, falls diese Kontrolle genau stimmt

Ein statistischer Test (Fortsetzung)

- Wir definieren H_0 , H_1 , Teststatistik und Niveau α
- Wir konstruieren die Verteilung der Teststatistik unter H_0 (Referenzverteilung)
- Wir vergleichen die beobachtete Teststatistik zur Referenzverteilung und treffen eine Entscheidung zum Niveau α
- Zwei Ansätze:
 - Verwerfungsbereich
 - P-Wert

Entscheidung anhand des Verwerfungsbereiches

- Verwerfungsbereich = Gebiet A der extremen Werte (bzgl. H_1), so dass $P(A|H_0) \leq \alpha$
 - Beispiel: $A \approx [12.65, \infty)$
- Entscheidung: Verwerfe H_0 , falls die beobachtete Teststatistik im Verwerfungsbereich ist
 - Beispiel: 12.069 ist nicht im Verwerfungsbereich
→ wir verwerfen H_0 nicht
- Schlussfolgerung:
Die Daten widersprechen H_0 nicht zum Niveau $\alpha = 0.05$.
(das heisst aber nicht, dass H_0 bewiesen ist!)

Entscheidung anhand eines P-Wertes

- P-Wert = die Wahrscheinlichkeit eine Teststatistik die mindestens so extrem ist wie unsere zu beobachten, falls H_0 stimmt (extrem bzgl. H_1)
 - Beispiel: P-Wert ≈ 0.06
- Entscheidung: Verwerfe H_0 , falls P-Wert $< \alpha$
 - Beispiel: $0.06 > 0.05 \rightarrow$ wir verwerfen H_0 nicht
- Schlussfolgerung:
Die Daten widersprechen H_0 nicht zum Niveau $\alpha = 0.05$.
(das heisst aber nicht, dass H_0 bewiesen ist!)

Randomisierungstests: Geschichte

- Zuerst vorgeschlagen von Fischer (1935) and Pitman (1937,1938)
- Fisher (1936):
“the statistician does not carry out this very simple and very tedious process, but his conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method.”
- Heutzutage ist Computing keine Einschränkung mehr. Man benutzt Monte-Carlo Simulationen, wenn die Anzahl der möglichen Permutationen zu gross ist.
- Einige übliche statistische Tests sind Randomisierungstests:
 - Wilcoxon Rank Sum Test
 - Wilcoxon Signed Rank Test

Randomisierungstests: Eigenschaften

- Exakt und nicht-parametrisch:
 - Die Tests halten das Signifikanzniveau α exakt ein ohne Voraussetzungen an der Verteilung
- Formell betrachtet man die Daten als gegeben. Der Zufall kommt nur via der Gruppenzuteilung rein.
(Man kann die Daten aber auch als zufällig betrachten; dann konditioneller Test.)
- Man kann beliebig komplexe (auch robustere) Teststatistiken benutzen
- Man kann auch Vertrauensintervalle konstruieren
- Nicht limitiert zum Vergleich zweier Stichproben mit Randomisierungsschritt. Voraussetzungen: die Beobachtungen sind unter H_0 gleich verteilt und unabhängig.

Inhaltsübersicht des Vortrags

- 2013 - Das Jahr der Statistik
- Permutationstests / Randomisierungstests
- **Kausalität**

Warum Kausalität?

- Wichtig für "statistical literacy"
- Meine Forschung ist in diesem Gebiet

Beispiele aus der Zeitung / dem Internet

- “Eat breakfast if you want to reduce your risk of coronary heart disease”
(The Guardian)
- “Anesthesia may harm children's brains. Study: anesthesia before age 3 linked to later mental problems”
(WebMD)
- “Breastfeeding may reduce Alzheimer’s risk”
(Cambridge University)
-

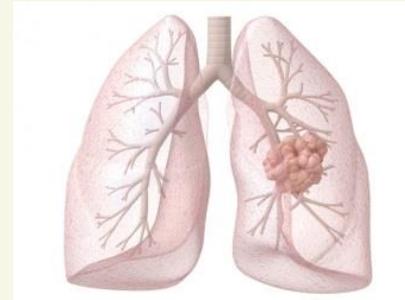
- Oft interpretiert man solche Aussagen als Warnungen oder Ratschläge

Beispiel

- Personen mit gelben Fingern haben ein höheres Risiko an Lungenkrebs zu erkranken



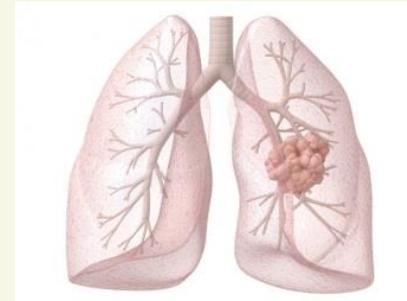
Korrelation



- Ratschlag: Man muss die Hände besser waschen (?!)

Beispiel

- Personen mit gelben Fingern haben ein höheres Risiko an Lungenkrebs zu erkranken



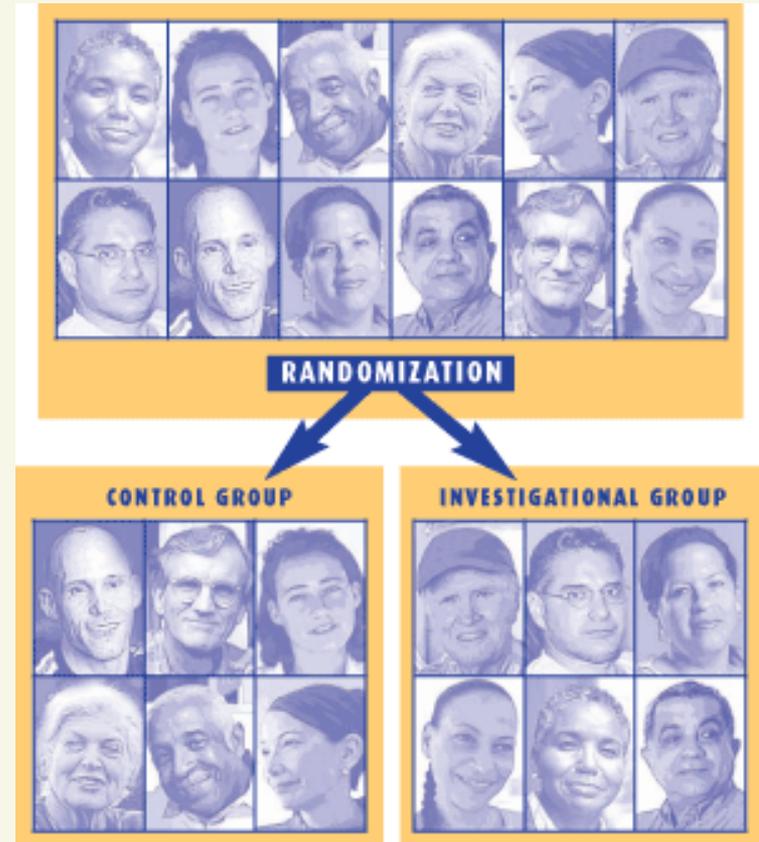
Nein, Rauchen ist die wahre Ursache!

- Ratschlag: Man muss die Hände besser waschen (?!)

Für sinnvolle Ratschläge muss man Kausalität verstehen

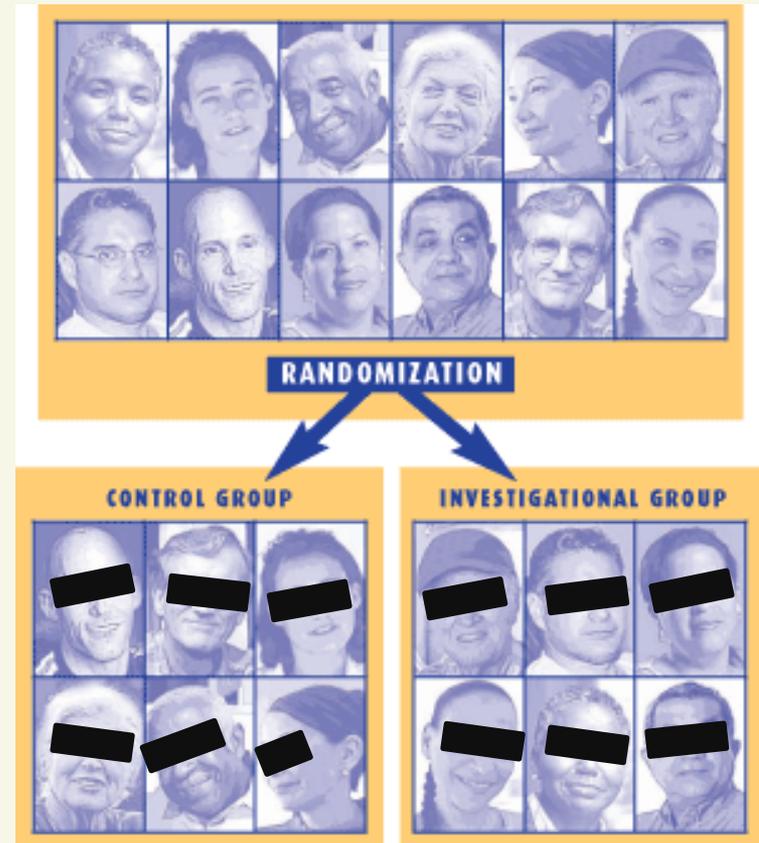
- Erste Frage:
 - Wie wurden die Daten gesammelt?
- Es gibt zwei wichtige Arten von Studien:
 - Randomisierte kontrollierte Studien
 - Beobachtungsstudien

Randomisierte kontrollierte Studien



- Es gibt eine **Kontrollgruppe**
- Die Behandlung wird **randomisiert**
- Am besten auch noch **doppelblind**

Randomisierte kontrollierte Studien

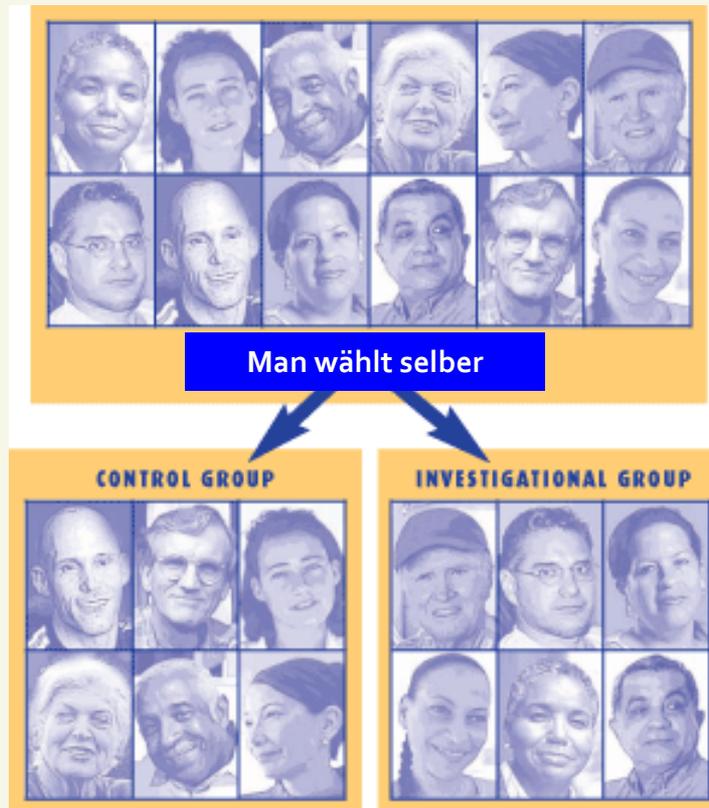


- Es gibt eine **Kontrollgruppe**
- Die Behandlung wird **randomisiert**
- Am besten auch noch **doppelblind**

Grundidee: vergleichbare Gruppen

- Durch die Randomisierung sind die Gruppen gleich in allen möglichen Aspekten, ausser der Behandlung
- Wenn es einen grossen Gesundheitsunterschied zwischen den Gruppen gibt, dann **muss das wegen der Behandlung sein**

Beobachtungsstudien



- Die Versuchspersonen bestimmen selber was sie machen; die Forscher beobachten nur.

Beobachtungsstudien

- Beispiel: Frühstück oder nur Kaffee

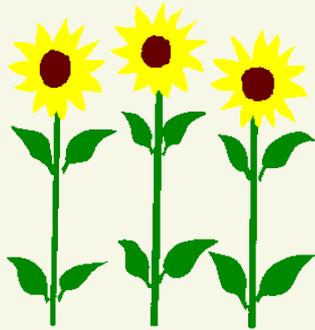
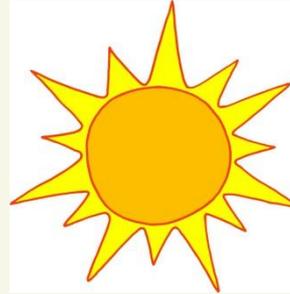


Weniger Stress?
Mehr Sport?
Gesündere Snacks?

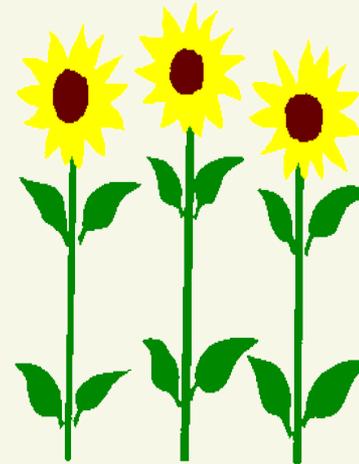
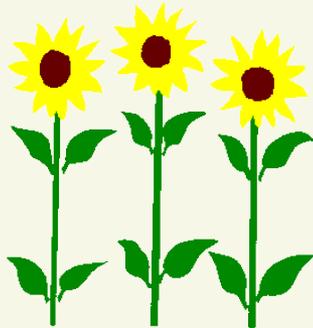
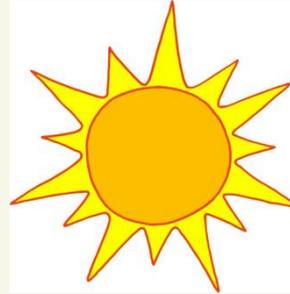
....

- Was ist die Ursache der geringeren Anzahl Herzkrankheiten?
- Wir wissen es nicht...

Vergleich: Experimente in den Naturwissenschaften



Vergleich: Experimente in den Naturwissenschaften



- Warum wachsen die Blumen schneller? Wegen der Sonne oder des Wassers?
- Wir wissen es nicht, weil wir auf einmal mehrere Sachen geändert haben

Wie findet man kausale Zusammenhänge?

- Randomisierte kontrollierte Studien:
 - Die Behandlung wird randomisiert
 - Die Gruppen sind gut vergleichbar
 - **Kausalität ist relativ einfach**
- Beobachtungsstudien:
 - Die Versuchspersonen wählen ihre Behandlung selber
 - Die Gruppen sind schlecht vergleichbar
 - **Kausalität ist schwierig**

Randomisierte kontrollierte Studien nicht immer möglich

- Eine randomisierte kontrollierte Studie ist aber nicht immer möglich oder praktisch.
- Beispiele:
 - Rauchen
 - Gen-Knockout Experimente
- Wie findet man kausale Zusammenhänge aus Beobachtungsstudien?

Ein Ansatz: konstruiere vergleichbare Subgruppen

- Beispiel: Frühstück oder nur Kaffee



Wenig Stress
Viel Sport
Gesunde Snacks

....



Wenig Stress
Viel Sport
Gesunde Snacks

....

- Statistisch: Man kontrolliert/adjustiert für diese Faktoren
- Schwierigkeit:
 - Für welche Faktoren muss man das machen / nicht machen?
 - Die Antwort hängt von der kausalen Struktur ab.
Man muss Annahmen über diese Struktur treffen.

Beispiel: Rauchen

- Eine randomisierte kontrollierte Studie war nicht möglich
- Jetzt ist man wirklich überzeugt, dass Rauchen Lungenkrebs verursacht
- Man hat kontrolliert für:
 - Stress
 - Geschlecht
 - Ernährung
 - Alkoholkonsum
 -
- Es hat auch sehr geholfen, dass man den biologischen Mechanismus verstanden hat

Take home message

- Erste Frage: Wie wurden die Daten gesammelt?
 - Randomisierte kontrollierte Studie:
 - Kausalität ist relativ einfach
 - Beobachtungsstudie:
 - Kausalität ist viel schwieriger
 - Für welche Faktoren wurde kontrolliert?
 - Gibt es alternative Erklärungen für die Resultate?
- Bei Beobachtungsstudien braucht es zusätzliche Annahmen. Die kann man oft nicht testen, weshalb man nie auf das Niveau einer randomisierten kontrollierten Studie kommt. Aber man kann mit modernen statistischen Methoden das meiste aus den Daten herausholen.

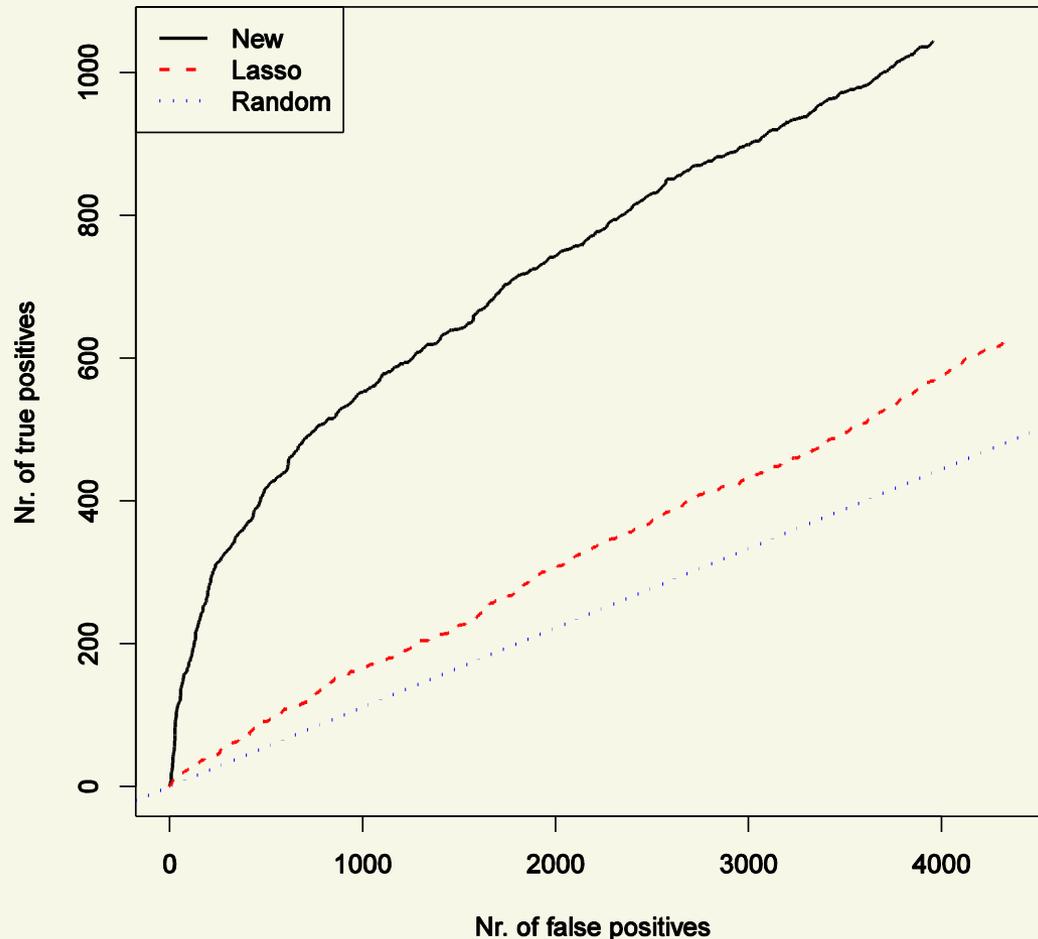
Einige Fragestellungen in meiner Forschung

- Wie findet man kausale Zusammenhänge aus Beobachtungsstudien?
 - Oft gibt es **wichtige Faktoren, die nicht gemessen wurden**. **Wie kann man damit umgehen?**
 - Die kausalen Strukturen kann man mittels Graphen darstellen. **Wie kann man solche Graphen schätzen? Was sind die Eigenschaften der verschiedenen Schätzmethoden?**
 - In manchen Situationen gibt es **sehr viele Variablen und wenig Beobachtungen** (Beispiel: Gen-Expressionsdaten). **Wie kann man in solchen Fällen vorgehen?**

Beispiel: Genregulationsnetzwerk von Hefe

- Daten:
 - Beobachtungsstudie mit Gen-Expressionsdaten:
 - 5000 Gene
 - 63 Hefe-Organismen
- Methode:
 - Wir treffen gewisse Annahmen
 - Wir schätzen die kausale Struktur / den Graphen
 - Wir schätzen die kausalen Effekte
- Validierung:
 - Wir vergleichen unsere Resultate mit Gen-Knockout Experimenten

Wir können die grossen kausalen Effekte besser vorhersagen



(Nature Methods, 2010)

- Mögliche Anwendung: [Versuchsplanung](#)

Referenzen

- Paper on randomization tests:
M.D. Ernst (2004). Permutation methods: a basis for exact inference. *Statistical Science*, volume **19**, pages 676-685.
- Applet randomization test:
<http://www.rossmanchance.com/applets/randomization20/Randomization.html>
- My favorite book on Introductory Statistics: Freedman, Pisani and Purves (2007). *Statistics*. 4th edition. Norton, New York
- Rice Virtual Lab of Statistics (simulations, demonstrations):
<http://onlinestatbook.com/rvls.html>
- Data and story library: <http://lib.stat.cmu.edu/DASL/>
- The year of statistics: <http://www.statistics2013.org/>
- Conference Ars Conjectandi (with public lecture):
<http://www.statoo.ch/bernoulli13/>

Danke!