

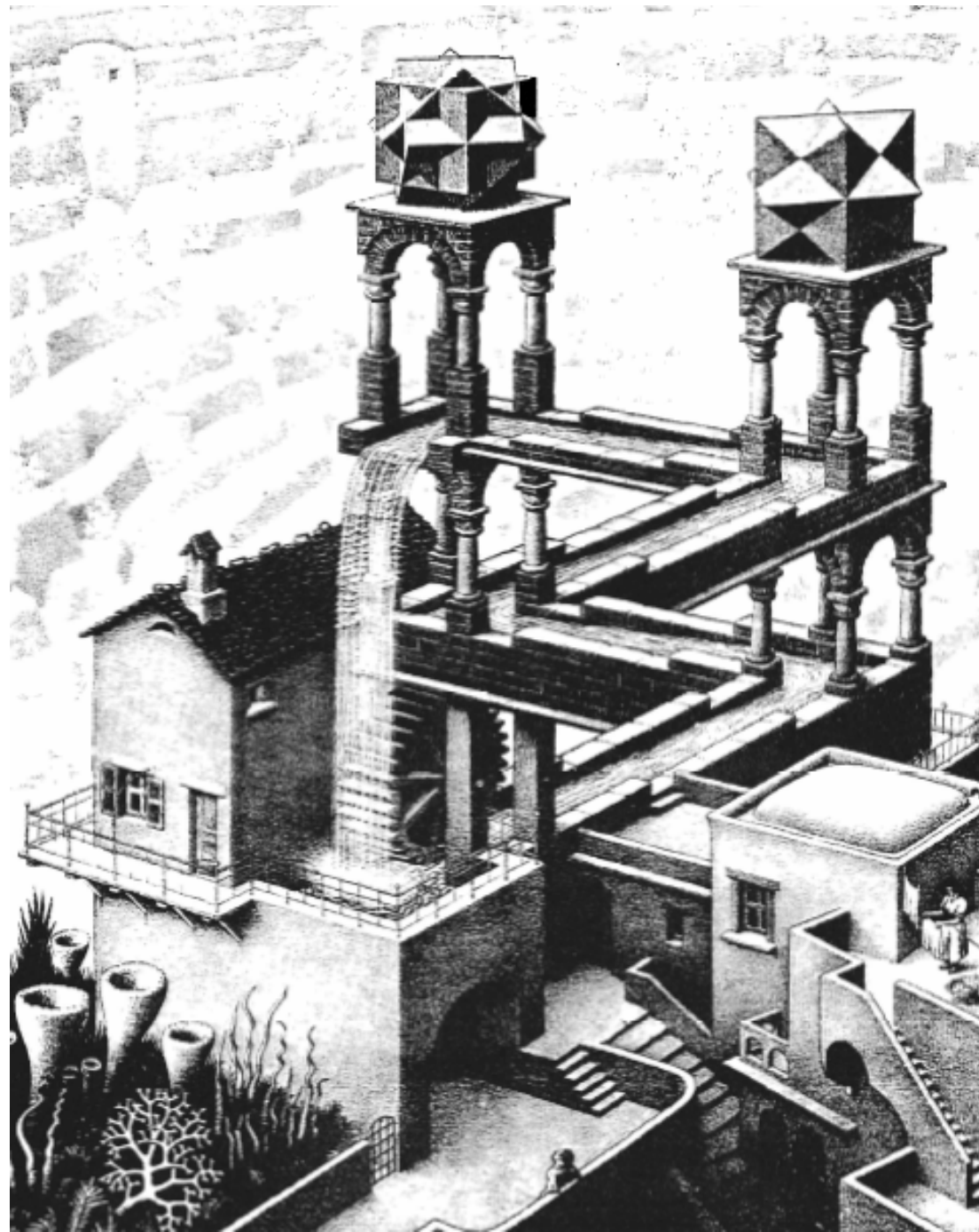


Vorlesung
Wahrscheinlichkeit und Regression
WS 2005/06

Kapitel 1
Einführung und Überblick

Prof. Dr. Rolf Steyer

**Lehrstuhl für Methodenlehre
und Evaluationsforschung**





The Simpson paradox

Table 1. *Evaluation of a treatment*

success	treatment		total
	yes ($X = 1$)	no ($X = 0$)	
yes ($Y = 1$)	500	600	1100
no ($Y = 0$)	500	400	900
	1000	1000	2000

Note. After Novick (1980). The numbers are fictitious.



The Simpson paradox (2)

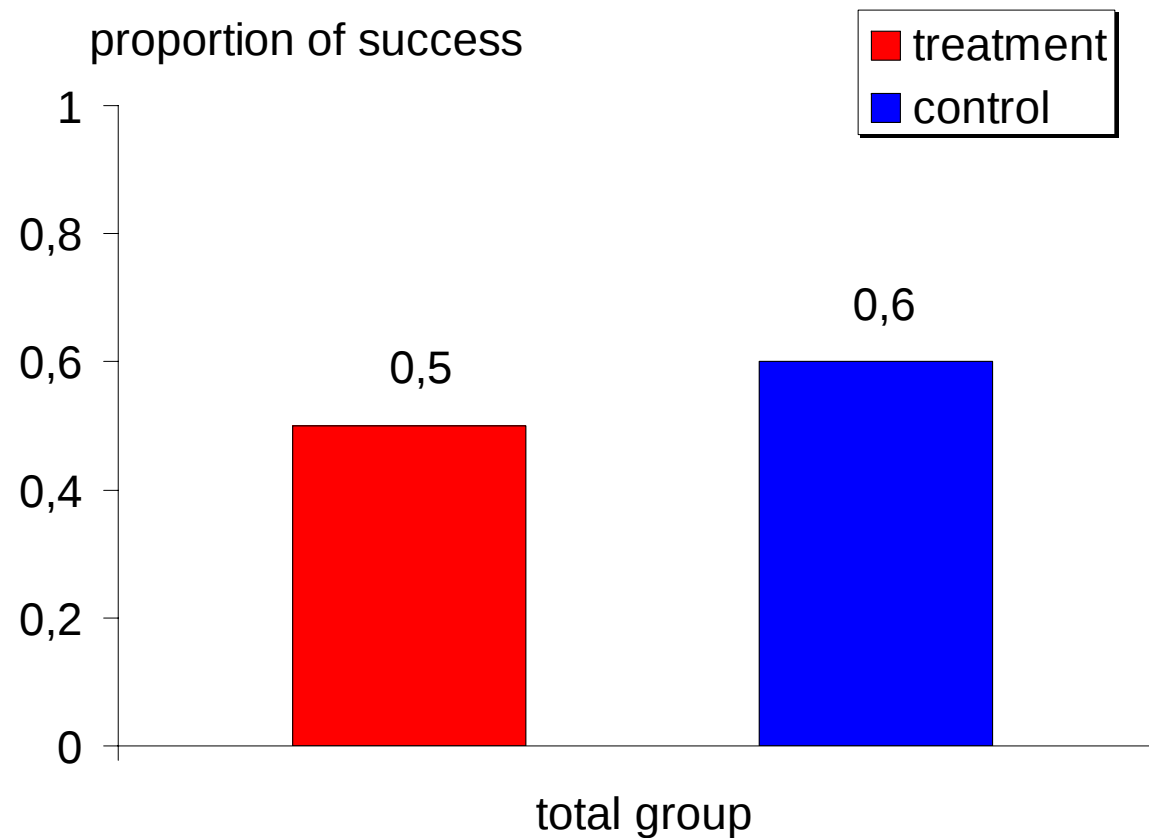


Figure 1. Histogram of the relative frequencies of success



The Simpson paradox (3)

Table 2. *Evaluation of a treatment within gender groups*

A. Males ($W = 1$)

success	treatment		total
	yes ($X = 1$)	no ($X = 0$)	
yes ($Y = 1$)	300	75	375
no ($Y = 0$)	450	175	625
	750	250	1000



The Simpson paradox (4)

B. Females ($W = 0$)

success	yes ($X = 1$)	no ($X = 0$)	total
yes ($Y = 1$)	200	525	725
no ($Y = 0$)	50	225	275
	250	750	1000

Note. After Novick (1980). The numbers are fictitious.



The Simpson paradox (5)

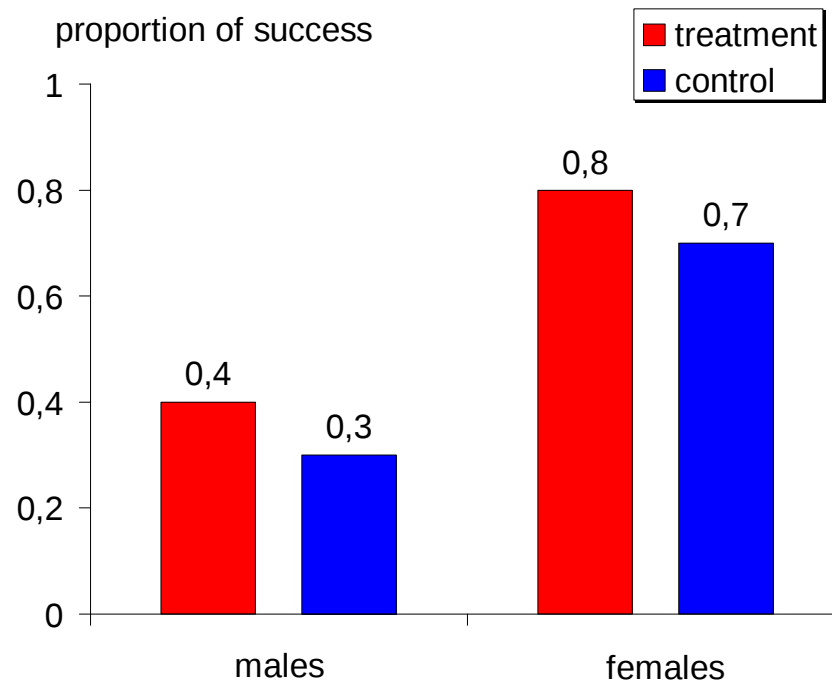


Figure 2. Bar chart of the relative frequencies for success within gender groups



Einführung

Warum sind stochastische Abhängigkeiten wichtig?

- Beispiele

Gründe für Nichtdeterminismus

- Multiple Determiniertheit
- Messfehler
- Situative Effekte
- Zufall



Überblick Kapitel 1

- Arten stochastischer Abhängigkeiten
- Wo kommen regressive Abhängigkeiten vor?
- Hauptaufgaben von Regressionsmodellen
- Wissenschaftstheoretische Bemerkungen
- Zur Geschichte der Regressionstheorie
- Regression als Teil der Theoriesprache
- Überblick über die Kapitel des Buches
- Voraussetzungen zum Verständnis



Arten der stochastischen Abhängigkeit

Abhängigkeiten von Ereignissen

Abhängigkeiten von Zufallsvariablen

- Stochastische Abhängigkeit
- Regressive Abhängigkeit
- Korrelative Abhängigkeit
- Andere

Ziel: Lernen, zwischen den verschiedenen Abhängigkeiten zu unterscheiden und mit diesen Begriffen Inhalte (psychologische und sozialwissenschaftliche Hypothesen) zu formulieren. Dies ist wichtig, da verschiedene Abhängigkeiten verschiedene psychologische und soziale Phänomene bedeuten.



Wo kommen regressive Abhängigkeiten vor?

Eine *Regression* $E(Y | X)$ einer numerischen Zufallsvariablen Y auf den Regressor X ist selbst eine Zufallsvariable, deren Werte die bedingten Erwartungswerte $E(Y | X = x)$ sind. Dabei kann der Regressor X aus mehreren Zufallsvariablen bestehen, d. h. $X = (X_1, \dots, X_m)$

Regressionen kommen zum Beispiel bei Vergleichen von Gruppenmittelwerten (t -Tests, ANOVA), bei der einfachen und multiplen linearen Regression, bei Faktorenanalysen und bei Strukturgleichungsmodellen vor.



Verschiedene Formulierungsarten

Aussagen über Regressionen können in verschiedener Weise formuliert werden, z. B.

- als Regressionskurve in einem kartesischen Koordinatensystem
- als Säulendiagramm, mit dem man bedingte Wahrscheinlichkeiten oder Erwartungswerte angibt,
- als Tabelle, in der man Erwartungswerte in Gruppen angibt
- als Pfaddiagramm oder auch
- als Gleichung

In allen genannten Fällen geht es um Aussagen darüber, wie die bedingten Erwartungswerte $E(Y | X = x)$ einer Zufallsvariablen Y von den Werten x einer Zufallsvariablen X abhängen.



Hauptaufgaben von Regressionsmodellen

Sie explizieren die Verknüpfung zwischen empirischen und theoretischen Begriffen und damit das *Messmodell*.

Sie explizieren die Verknüpfung zwischen den theoretischen (bzw. empirischen) Begriffen und damit die *Abhängigkeitsbegriffe*.

Beispiele



Wissenschaftstheoretische Bemerkungen

Theorien sollten logisch widerspruchsfrei sein

Aus Theorien sollten Aussagen über empirische Beobachtungen
ableitbar sein

Vorteile einer formalen Sprache

Regression als Begriff der Theoriesprache



Historical Remarks

Galton und andere im 19-ten Jahrhundert
Kolmogoroff 1933



Überblick über die Vorlesung

Wahrscheinlichkeit und Regression

Teil I: Wahrscheinlichkeitstheorie

Teil II: Regressionstheorie

Teil III: Kausale Regression

Theorien psychologischer Tests

Klassische Testtheorie

Latent-state-trait-Theorie

Item-response-Theorie

Latent-class-Theorie



Voraussetzungen zum Verständnis

Grundbegriffe der Aussagen- und Prädikatenlogik

Mengenlehre

Relation und Abbildung

Elementare deskriptive und inferenzielle Statistik (nützlich)

(s. Anhänge von Steyer & Eid, 2002, Messen und Testen)



Zusammenfassung I

Abhängigkeiten in den Sozial- und Verhaltenswissenschaften sind meist nichtdeterministisch. Gründe dafür sind *Multiple Determiniertheit*, *Messfehler* und *situationale Spezifität*.

Bei den in empirischen Untersuchungen der Sozial- und Verhaltenswissenschaften betrachteten Abhängig- und Unabhängigkeitsarten handelt es sich nicht um inhaltlich irrelevante Methoden, die beliebig austauschbar und ersetzbar sind. Vielmehr handelt es sich bei ihnen um abstrakte Inhalte insofern, als ihre charakteristischen Eigenschaften vielen inhaltlichen Abhängigkeiten gemeinsam sind.

In den meisten Fällen, in denen in den empirischen Sozial- und Verhaltenswissenschaften statistische Verfahren verwendet werden, geht es um Aussagen darüber, wie die bedingten Erwartungswerte $E(Y | X = x)$ einer Variablen Y (des „Regressanden“) von den Werten einer Variablen X oder auch mehrerer Variablen $X_1, \dots, X_m$ (den „Regressoren“) abhängen, oder um globale Aussagen darüber, wie stark diese regressive Abhängigkeit ist, z. B. durch Angabe des Determinationskoeffizienten.



Zusammenfassung II

Regressionsmodelle haben zwei Hauptaufgaben. Sie explizieren erstens die Verknüpfung zwischen empirischen und theoretischen Begriffen, und damit Messmodelle, und zweitens die Verknüpfung zwischen den theoretischen Begriffen und damit die Abhängigkeitsbegriffe, mit denen die Beziehungen zwischen theoretischen Begriffen beschrieben werden können.

Die Analyse von Daten ist sowohl bei der Entwicklung von Theorien, als auch bei deren Überprüfung von großer Bedeutung. Dennoch gilt der *Theorie* das zentrale Interesse und daher auch der Sprache, in der die theoretischen Aussagen formuliert werden. Regressionstheorie, wie sie in diesem Buch behandelt wird, ist ein wichtiger Teil der Theoriesprache der empirischen Wissenschaften.