



# **Klinische versus statistische Urteilsbildung**

Probenvorlesung  
10.11.2005

Dr. Carmen Hagemeister

Fakultät Mathematik und Naturwissenschaften  
Fachrichtung Psychologie  
Diagnostik und Intervention



# Überblick

---

- Was heißt "statistisch", was "klinisch"?
- Was ist besser und warum?
- Wird das Bessere auch gemacht?
- Warum wird nicht immer das Bessere gemacht?
- Fazit



## Was heißt "statistisch", was "klinisch"?

- Urteilsbildung
  - Informationen kombinieren
  - Ziel: Diagnose, oft auch Prognose
  - psychologisch: auf menschliches Verhalten bezogen



## Was heißt "statistisch", was "klinisch"?

- "statistisch"
  - = nach festen, expliziten Regeln
    - ◆ "mechanisch"
    - ◆ "buchhalterisch"
  
- Die klinische Psychologin, die ein standardisiertes Verfahren zur Diagnose nutzt, geht in diesem Sinne "statistisch" vor.



## Was heißt "statistisch", was "klinisch"?

- "klinisch"
  - = nicht nach festen, expliziten Regeln
    - ◆ kann "intuitiv" sein, muss aber nicht
    - ◆ kann auch gut begründet sein
  
- Der Personalchef, der ohne eine feste Regel entscheidet, wen er einstellt, geht in diesem Sinne "klinisch" vor.



## Was ist besser und warum?

- Datenkombination Meehl (1954):
  - statistische Datenkombination im Mittel besser als klinische Datenkombination
  
- Problem der Untersuchung von Meehl:
  - nur Vergleich von Datenkombinationen: bestimmte Informationen ausgeschlossen, z. B. Daten klinischer Interviews
  - Kliniker kann Daten erheben, die anders nicht zugänglich sind



## Was ist besser und warum?

---

- Sawyer (1966): Untersuchung von
  - Datensammlung
  - Datenkombination



## Was ist besser und warum?

- Meta-Analyse:
  - 47 Studien
  - Vergleich von mindestens zwei Vorhersagemethoden
  - "Kliniker" waren
    - ◆ Psychologen oder Psychiater (über 50% der Studien)
    - ◆ Ärzte
    - ◆ Offiziere oder andere Soldaten
    - ◆ Studierende
    - ◆ teilweise Personen ohne Training in Diagnostik



# Was ist besser und warum?

## - Klassifikation von Vorhersagemethoden

<b>Daten- sammlung</b>	<b>Datenkombination</b>	
	klinisch	statistisch
klinisch	rein klinisch	Eigenschaftsurteile
statistisch	Profilinterpretation	rein statistisch
klinisch und statistisch	klinische Zusammensetzung	mechanische Zusammensetzung
Synthese	klinische Synthese	mechanische Synthese



# Was ist besser und warum?

## - Klassifikation von Vorhersagemethoden

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	rein klinisch	Eigenschaftsurteile
statistisch	Profilinterpretation	rein statistisch
klinisch und statistisch	klinische Zusammensetzung	mechanische Zusammensetzung
Synthese	klinische Synthese	mechanische Synthese

Diagramm zur Klassifikation von Vorhersagemethoden. Die Tabelle zeigt die Kombination von Datenarten (klinisch, statistisch) und die daraus resultierenden Methoden (Eigenschaftsurteile, Profilinterpretation, mechanische Zusammensetzung, klinische Synthese, mechanische Synthese). Pfeile weisen von den Methoden in der rechten Spalte auf die entsprechenden Methoden in der linken Spalte der letzten Zeile (Synthese) hin.



# Was ist besser und warum?

## - Klassifikation von Vorhersagemethoden

Daten-sammlung	Datenkombination	
	klinisch	statistisch
klinisch	rein klinisch	Eigenschaftsurteile
statistisch	Profilinterpretation	rein statistisch
klinisch und statistisch	klinische Zusammensetzung	mechanische Zusammensetzung
Synthese	klinische Synthese	mechanische Synthese



## Was ist besser und warum?

- relative Güte

<b>Daten- sammlung</b>	<b>Datenkombination</b>	
	klinisch	statistisch
klinisch	20%	43%
statistisch	38%	63%
klinisch und statistisch	26%	75%
Synthese	50%	75%



## Was ist besser und warum? - Datenkombination

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	<b>43%</b>
statistisch	38%	<b>63%</b>
klinisch und statistisch	26%	<b>75%</b>
Synthese	50%	<b>75%</b>



## Was ist besser und warum? - Datenkombination

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	<b>43%</b>
statistisch	38%	<b>63%</b>
klinisch und statistisch	26%	<b>75%</b>
Synthese	50%	<b>75%</b>

statistische Datenkombination ist der klinischen überlegen



# Was ist besser und warum?

## - Datensammlung

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	43%
statistisch	<b>38%</b>	<b>63%</b>



## Was ist besser und warum? - Datensammlung

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	43%
statistisch	<b>38%</b>	<b>63%</b>

nur statistische Daten liefern bessere Vorhersagen als nur klinische



## Was ist besser und warum? - Datensammlung

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	43%
statistisch	<b>38%</b>	63%
klinisch und statistisch	26%	<b>75%</b>



## Was ist besser und warum? - Datensammlung

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	43%
statistisch	<b>38%</b>	63%
klinisch und statistisch	26%	<b>75%</b>

klinische und statistische Daten

- Gewinn bei statistischer Kombination
- kein Gewinn bei klinischer Kombination



# Was ist besser und warum?

## - Synthese

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch		43%
statistisch		<b>63%</b>
klinisch und statistisch		<b>75%</b>
Synthese	50%	

Arrows from the percentages 43%, 63%, and 75% in the 'statistisch' column point to the 'Synthese' row.



# Was ist besser und warum?

## - Synthese

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch		43%
statistisch		<b>63%</b>
klinisch und statistisch		<b>75%</b>
Synthese	50%	

### klinische Synthese

- besser als statistische Kombination klinischer Daten
- schlechter als statistische Kombination statistischer oder klinischer und statistischer Daten



# Was ist besser und warum?

## - Synthese

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	
statistisch	38%	
klinisch und statistisch	26%	
Synthese		<b>75%</b>

Arrows from the 20%, 38%, and 26% cells point to the 75% cell, indicating that the total percentage of data combinations that include synthesis is 75%.



# Was ist besser und warum?

## - Synthese

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	
statistisch	38%	
klinisch und statistisch	26%	
Synthese		<b>75%</b>

### statistische Synthese

- o besser als klinische Datenkombination der verschiedenen Datengrundlagen



## Was ist besser und warum? - Datenkombination

- Das Modell eines Experten ist besser als der Experte, weil
  - es reliabler ist
  - die Urteilsstrategie konsistenter verwendet als der Experte
  - Basisraten berücksichtigt
- Bootstrapping
  - z.B. HYPAG (Wottawa, 1987): isomorphe Abbildung der Urteilsstrategie eines Experten
- Konsequenz:
  - Implementation auf dem Computer



## Was ist besser und warum? - mehr Daten

- klinische Kombination von klinischen und statistischen Daten ist schlechter als klinische Kombination von nur statistischen Daten
  - mehr Information  $\Rightarrow$  schlechteres Urteil
  - falsche Gewichtung
  - hohe Gewichtung der klinischen Daten
  - falls keine Gewichte berechnet
    - ◆ Einheitsgewichtung (= alles gleich gewichtet) am günstigsten
    - ◆ schlechter als optimale Gewichtung
    - ◆ besser als Gewichtung in der falschen Richtung



## Was ist besser und warum? - Schlussfolgerungen

---

- Statistische Datenkombination hat bessere Ergebnisse als klinische Datenkombination.
- Wenn man Daten statistisch kombiniert, lohnt es sich, sowohl klinische als auch statistische Daten zu verwenden; nicht jedoch, wenn man sie klinisch kombiniert.
- Wenn man Daten, die entweder nur statistisch oder statistisch und klinisch sind, statistisch kombiniert hat, sollte man die Entscheidung nicht noch einmal klinisch revidieren.



# Was ist besser und warum?

## - neuere Meta-Analyse

- Grove, Zald, Lebow, Snitz und Nelson (2000):  
Was ist besser, klinisches oder statistisches Vorgehen?
- Meta-Analyse:
  - 136 Studien
  - Medizin und Psychologie
  - Klinische und mechanische Vorhersage



# Was ist besser und warum?

## - neuere Meta-Analyse

Grove et al. (2000)

bessere Datenkombination	Zahl der Untersuchungen
klinisch	8
gleich	65
statistisch	63



## Was ist besser und warum? - neuere Meta-Analyse

Grove et al. (2000)

bessere Datenkombination	Zahl der Untersuchungen
klinisch	8
gleich	65
statistisch	63

feste Regel im Mittel um 9-10% besser als Kliniker



## Was ist besser und warum? - Gibt es Ausnahmen?

- Vorteil der mechanischen Kombination wird nicht beeinflusst von:
  - Publikationsjahr
  - Stichprobengröße
  - Zeitschrift vs. anders publiziert
  - Feld der Vorhersage: Bildung, Finanzen, Forensik, Medizin, Klinisch/Persönlichkeit, andere
  - Urteiler: medizinische vs. psychologische Ausbildung
  - Training oder Erfahrung des Urteilers
  - mehr Daten beim Kliniker als in der mechanischen Vorhersage



## Was ist besser und warum? - Gibt es Ausnahmen?

---

Vorteil der mechanischen Vorhersage wird beeinflusst von

- Einbezug von Interviewdaten

Vorteil der mechanischen Vorhersage größer, als wenn keine Interviewdaten einbezogen werden!!!



## Was ist besser und warum? - Gibt es Ausnahmen?

Vorteil der mechanischen Vorhersage wird beeinflusst von

- Einbezug von Interviewdaten

Vorteil der mechanischen Vorhersage größer, als wenn keine Interviewdaten einbezogen werden!!!

- Interviews

- meist klinische Daten
- führen zu schlechter Vorhersage, wenn keine feste Regel benutzt wird
- wie bei Sawyer: klinische Kombination schlechter bei klinischen und statistischen Daten als bei nur statistischen Daten



## Was ist besser und warum? - wo weniger Daten besser sind

Daten- sammlung	Datenkombination	
	klinisch	statistisch
klinisch	20%	
statistisch	<b>38%</b>	
klinisch und statistisch	26%	

kein Gewinn zusätzlicher klinischer Daten bei klinischer Kombination



## Wird das Bessere auch gemacht? - Klinisch-klassifikatorische Diagnostik

- Diagnostiksysteme bei Praktikern nicht hinreichend bekannt (Wynn, Myklebust & Stensland, 2005)
- Akzeptanz abhängig von der theoretischen Ausrichtung des Praktikers (Wynn et al., 2005)
- Dominanz der weißen europäisch-nordamerikanischen Kultur erlebt (Mezzich, 2002)
- Training wird als unzureichend erlebt (Mezzich, 2002)



## Wird das Bessere auch gemacht? - Eignungsdiagnostik

- regelgeleitetes Vorgehen eher bei häufigen Entscheidungen als bei seltenen
- nicht regelgeleitetes Vorgehen oft, wenn
  - wenige Personen eine derartige Stelle haben
  - Leistung schwierig zu definieren
  - keine Erfolgskriterien verfügbar
- "Trick": 50% Regel, 50% Intuition des Managers bei Vorhersagen (Blattberg & Hoch, 1990).
  - bessere Entscheidungen als rein klinisch
  - höhere Akzeptanz als nur statistisch
- DIN 33430: Anforderungsprofil



## Warum nicht immer das Bessere? - Gefühl

- Menschen, die eine Entscheidung mit einer Regel statt mit ihrem Wissen fällen sollen, haben oft ein schlechtes Gefühl dabei.
  - Argument: Entscheidungen, die explizit und systematisch sind, sind entmenschlichende Entscheidungen.
  - Für wen?
    - Sie sind "entmenschlichende" Entscheidungen, aber nur für den Entscheider.
    - Für die Personen, über die entschieden wird, sind schlechte Entscheidungen entmenschlichend.
- (Dawes, 1988)



# Warum nicht immer das Bessere?

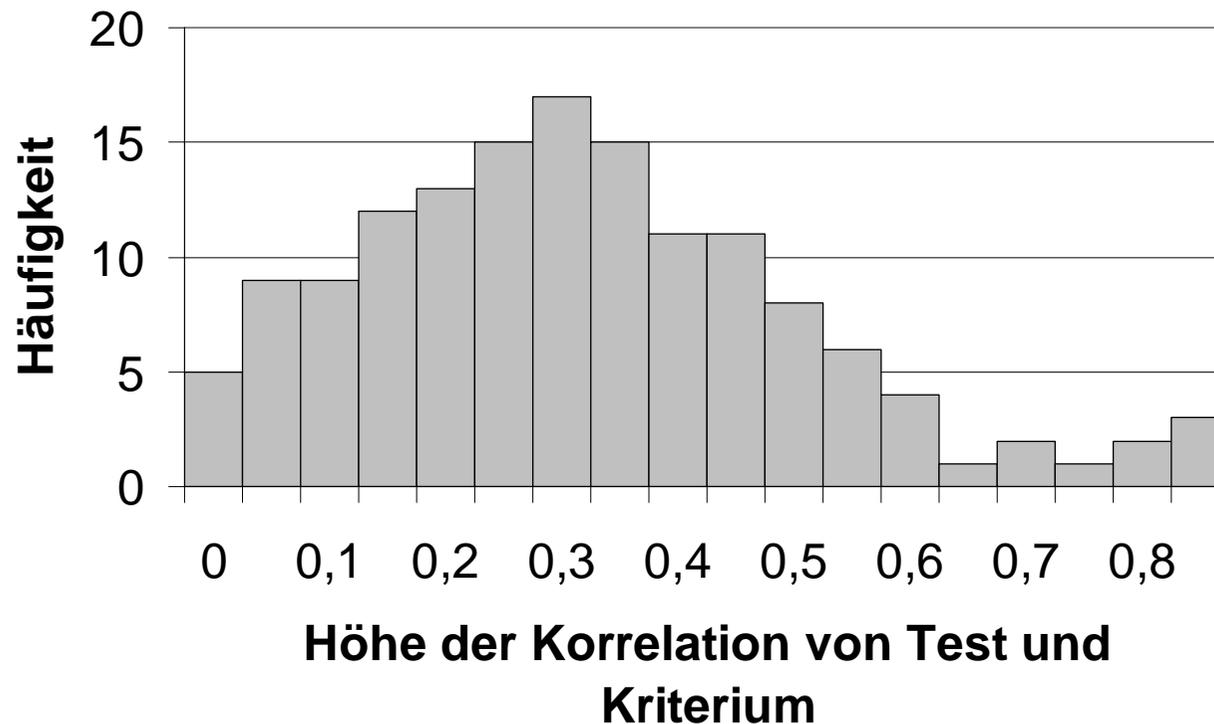
## - Ethik

---

- Es erscheint den Entscheidern unethisch, die Entscheidung an eine Regel zu delegieren.
- auch hier:
  - Für die Personen, über die entschieden wird, sind schlechte Entscheidungen unethisch.



## Warum nicht immer das Bessere? - Wie gut ist besser?



Meyer et al. (2001): Kriteriumsvalidität psychologischer und medizinischer Tests für unterschiedliche Kriterien  
Mittelwert  $r = 0,32$  (SD  $r = 0,19$ )



# Warum nicht immer das Bessere?

## - Wie gut ist besser?

- realistisch

- geringe Validität
- viele Fehlentscheidungen

⇒ Eindruck, man selbst könnte besser sein als die a-priori-Strategie / feste Regel



## Warum nicht immer das Bessere? - Ich kann es besser!

- Experiment mit 4 Instruktionen (Arkes, Dawes & Christensen, 1986)
  - "Mit der Regel schafft man 70% ...
    - ... So gut können Menschen werden." (Kontrollgruppe)
    - ... Vermutlich kann man besser werden. Versuchen Sie es." (Innovation)
    - ... Nicht einmal ein Experte ist besser. Personen, die versucht haben, es besser zu machen, waren in Wirklichkeit schlechter." (Warnung)
    - Gruppe ohne Feedback mit Kontrollgruppeninstruktion



# Warum nicht immer das Bessere?

## - Ich kann es besser!

---

- Belohnung
  - pro richtiger Entscheidung
  - für den besten Entscheider
  - keine Belohnung



## Warum nicht immer das Bessere? - Ich kann es besser!

- beste Leistung (70%): Gruppe ohne Feedback ohne Belohnung
  
- Instruktion:
  - gute Leistung
    - ◆ Gruppe ohne Feedback
    - ◆ Gruppe mit Warnung
  - Leistung gleich schlecht bei
    - ◆ Innovationsgruppe
    - ◆ Kontrollgruppe



# Warum nicht immer das Bessere?

## - Ich kann es besser!

---

- Belohnung:
  - Gruppe ohne Belohnung besser als Gruppen mit Belohnung



# Warum nicht immer das Bessere?

## - Ich kann es besser!

- Feedback plus Belohnung:
  - schneller Strategiewechsel nach Misserfolg
  - hohe Motivation: schädlich für Leistung
  
- ebenfalls ungünstig
  - Expertise
    - ◆ Personen, die wenig Wissen in einem Gebiet haben und sich auch so einschätzen, sind eher bereit, eine Regel anzuwenden



# Warum nicht immer das Bessere?

## - Situationen

---

- Nicht immer feste Regeln verfügbar
  - seltene Entscheidungen
  - bei sehr seltenen Fällen keine Evaluation der Entscheidungsregel möglich



## Warum nicht immer das Bessere? - Empfehlungen (Kleinmuntz, 1990)

- herausfinden, was Menschen schlecht können: Gewichte von Hinweisreizen bestimmen, Formeln durchrechnen – Computer nutzen
- Strategie dokumentieren und ihre Ergebnisse
- Entscheidung nachher evaluieren
- Information suchen, die der eigenen Hypothese widerspricht
- Anwendungsfelder, Validität, Nutzen und Grenzen von festen Regeln erforschen



## Fazit

- Wenn man Entscheidungen zu treffen hat, die vielfach vorkommen, werden diese besser, wenn man sie nach einer festen Regel trifft.
- Wer eine vorliegende feste Regel nicht nutzt, riskiert, dass sie oder er schlechter ist als die Regel. (Das ist die Warnung, die im Experiment geholfen hat. Möge sie auch hier helfen!)
- Man kann eine Entscheidungsstrategie nur verbessern, wenn man sie kennt. Das bedeutet im ersten Schritt Dokumentation. Der nächste Schritt ist die Evaluation.



## Fazit

- Was bleibt den Psychologinnen und Psychologen?
  - Theoriebildung:
    - ◆ Was könnten gute Prädiktoren sein? Entscheiden, was noch erhoben und geprüft werden soll.
    - ◆ Erklärungen: gute Prädiktoren sagen etwas vorher, erklären es aber nicht
  - Therapie statt Diskussionen über Therapiewahl und -prognose (Meehl, 1954)



## Fazit

---

Auch mit der besten Regel kann man nicht nur richtige Entscheidungen treffen.

Wir leben in einer probabilistischen Welt. Das heißt, dass wir akzeptieren müssen, dass nicht alles vorhersagbar ist.

Auch andere konnten sich das nicht vorstellen, zum Beispiel Einstein, der sagte "Gott würfelt nicht".



Falls Sie auf dieser Seite ein Bild vermissen, finden Sie es in Hawking, S. (1998). Raum und Zeit. Reinbek: Rowohlt.



## Literatureempfehlungen

für die Diagnostik-Diplomprüfung:

- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin*, 66, 178-200.

darüber hinaus:

- Arkes, H. R., Dawes, R. M. & Christensen, C. (1986). Factors Influencing the Use of a Decision Rule in a Probabilistic Task. *Organizational Behavior and Human Decision Processes*, 37, 93-110.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E. & Nelson, C. (2000). Clinical versus Mechanical Prediction: A Meta-Analysis. *Psychological Assessment*, 12, 19-30.
- Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative Approach. *Psychological Bulletin*, 107, 296-310.



## Literatureempfehlungen

weitere zitierte Literatur:

- Blattberg, R. C. & Hoch, S. J. (1990). Database models and managerial intuition: 50% model + 50% manager. *Management Science*, 36, 887-899.
- Dawes, R. M. (1988). You can't systematize human judgement: Dyslexia. In J. Dowie & A. S. Elstein (Eds.), *Professional judgement: A reader in clinical decision making* (pp. 150-162). New York: Cambridge University Press.
- Meyer, G. J., Finn, S. E., Eide, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eismann, E. J., Kubiszyn, T. W. & Reed, G. M. (2001). Psychological testing and psychological assessment. A review of evidence and issues. *American Psychologist*, 56, 128-165.
- Mezzich, J. E. (2002). International surveys on the use of ICD-10 and related diagnostic systems. *Psychopathology*, 35, 72-75.
- Wottawa, H. (1987). Hypotheses Agglutination (HYPAG). A method for configuration-based analysis of multivariate data. *Methodika*, 1, 68-92.
- Wynn, R., Myklebust, L.-H. & Stensland, G. O. (2005). Attitudes to the use of diagnostic instruments among clinicians in northern Norway. *Australian and New Zealand Journal of Psychiatry*, 39, 313.