

Human-Computer Interaction

Termin 8:

Statistical analysis

Model-based usability evaluation

Evaluation

- **Warum?** Bewerte *Usability* und Effekte auf Nutzer, finde Probleme, gib Hinweise für Verbesserungen
- **Was?** Usability-Kriterien
- **Wo?** Labor oder Umgebung/Feldstudie
- **Wer?** Experte (mit/ohne Nutzerbeteiligung) oder echte Nutzer
- **Wann?** In *allen* Stadien des Entwicklungszyklus (Ideen, Prototypen, System)



Wie? Evaluationsmethoden

- Inspektionsmethoden (Experten-Review)
- Benutzerstudien
- Model-basierte Evaluation
 - Usability-Test
 - Thinking-Aloud
 - Field studies
 - Interviews & questionnaires



Usability Test - Durchführung

- Suche repräsentative Benutzer
- Kriterien der Auswertung auswählen:
 - Zeit für Aufgabenerfüllung
 - Zeit für Aufgabe nach Ablenkung/neuem Input
 - Anzahl und Art von Fehlern pro Aufgabe oder pro Zeiteinheit
 - Anzahl Zuhilfenahme Onlinehilfe oder Manual
 - ...
- Entwickle relevante Testszenarien
- Ethische Fragen klären
- Pilottests
- Durchführung eigentlicher Test - Daten sammeln
- Auswertung - Was kann man aus den Daten lernen?



Statistische Evaluation

- **Deskriptive Statistik** (beschreibend) für quantitative Erhebungen, etwa
 - Daten von Benutzerbefragungen
 - Daten von Messungen (z.B. Zeit zur Bearbeitung von Aufgaben)
 - Veranschaulichung von Zahlenmaterial

- **Inferenzstatistik** (Test von Hypothesen), etwa
 - konkurrierende Schnittstellen-Designs
(Bsp.: "GUIs mit Zeichnungen als Icons sind einfacher zu bedienen als GUIs mit Fotos")
 - verschiedene Benutzergruppen
(Bsp.: "ältere Benutzer haben mehr Probleme mit der Lesbarkeit als jüngere")
 - verschiedene Eingabemodalitäten
(Bsp: „Multimodale Eingabe mit Sprache und Gestik ist effizienter als direkte Manipulation per Spacemouse“)



Statistik & Daten

- Durchführung eines Experimentes und Beobachtung von empirischen Sachverhalten

- *Beobachtungen bzw. Messungen* ergeben *Daten*
 - im Hinblick auf das jeweilige Erkenntnisinteresse informativ
 - durch den Bezug auf eine *Skala* mit spezifischen mathematischen Eigenschaften statistisch auswertbar



Messen

- *Messen* ist das Zuordnen eines beobachtbaren *Merkmals* eines Objekts zu einem *Wert* einer *Skala*
 - Beispiel: Zuordnung des Werts 182 auf der Zentimeterskala zum Merkmal Größe einer Person

- in der Praxis häufig: "Messung" von Merkmalen, die auf theoretischen Konstrukten basieren, statt auf physikalischen Eigenschaften
 - Beispiel: Unix-Kenntnisse als Merkmal eines Benutzers

- Problem der *Operationalisierung*:
Festlegung, auf welche beobachtbaren (*messbaren*) Merkmale sich ein theoretischer Begriff bezieht



Skalen

- *Skalen* für Messwerte in Typenhierarchie einteilbar
- Typ einer Skala wird auch als *Skalenniveau* bezeichnet
- je nach Skalenniveau sind unterschiedliche mathematische Operationen auf Messwerten definiert
- man unterscheidet die Niveaus
 1. *Nominalskala: gleiche Merkmalsausprägung, gleiche Werte:*
Blutgruppen A, B, AB, Null; Geschlecht m., w.
 2. *Ordinalskala: größerer Merkmalsausprägung, größere Werte:*
Kleidungsgrößen S, M, L, XL, XXL; Urteile --, -, 0, +, ++
 3. *Intervallskala: Größenordnung von Wertdifferenzen entspricht Merkmalsunterschieden:* Temperatur in °C oder Fahrenheit
 4. *Verhältnisskala: Größenverhältnisse der Werte entspricht Merkmalsausprägungen:* Länge in mm
 5. *Absolutskala: absolute Werte entsprechen Merkmalsausprägungen:*
Anzahl der Kinder



Skalenhierarchie

	<i>Skalenniveau</i>	<i>Mögl. Aussagen über Messwerte</i>	<i>Zulässige Transformation</i>
metrisch (Wertabstände sinnvoll)	Absolutskala	Absoluter Messwert	Identität $T(x) = x$
	Verhältnisskala	Gleichheit von Verhältnissen	Ähnlichkeit $T(x) = ux$
	Intervallskala	Gleichheit von Abständen	Positiv linear $T(x) = ux + v$
	Ordinalskala	Größer-kleiner-Relationen	Monot. steigend $x > y \Leftrightarrow T(x) > T(y)$
nichtmetrisch	Nominalskala	Gleichheit, Verschiedenheit	Eineindeutigkeit $x = y \Leftrightarrow T(x) = T(y)$

↑ Messwerte aussagekräftiger



Deskriptive Statistik

Angenommen, zu Evaluationszwecken wurde reichlich Datenmaterial gesammelt bzw. erhoben.

Zentrale Frage der deskriptiven Statistik:

Wie können die empirischen Daten aufbereitet, dargestellt, zusammengefasst und strukturiert werden, so dass zentrale Merkmale sichtbar werden?

Hier Fokus auf

- Darstellung von Häufigkeitsverteilungen durch Kennwerte
- Darstellung durch graphische Visualisierung



Empirische Häufigkeitsverteilungen

- einfache Art der Zusammenfassung von Datenmatrizen
- enthält beobachtete *Häufigkeiten* der Werte einer Variablen
 - *absolute* Häufigkeit f : Gesamtzahl der Auftretens eines Wertes
 - *relative* Häufigkeit f_r : abs. Häufigkeit/ n , mit n =Anzahl Werte
 - *kumulierte* Häufigkeit f_k : sukzessives Aufsummieren der Häufigkeiten; in wie vielen Fällen größerer bzw. kleinerer Wert gemessen (für mind. ordinalskalierte Daten)

Alter	f	f_r	f_k	f_{rk}
17	1	0.25	1	0.25
29	2	0.50	3	0.75
36	1	0.25	4	1.00



Kennwerte der *zentralen Tendenz*

Gesucht ist ein Wert, der die ganze Messreihe, also die Gesamtheit aller Ausprägungen einer Variablen in einer Stichprobe, möglichst gut repräsentiert.

Arithmetisches Mittel (Durchschnitt) \bar{x}

Summe der Messwerte x_i einer Messreihe, dividiert durch Anzahl n

$$\bar{x} = \frac{\sum_i x_i}{n}$$

- geeignet bei Intervalldaten (oder "höheren" Skalen), großen Stichproben und symmetrischer Verteilung
- sensibel gegenüber Extremwerten („*outliers*“), daher bei *schiefer* Verteilung nicht angebracht!



Kennwerte der *zentralen Tendenz*

Median (Zentralwert) \tilde{x}

der Wert, der die nach Größe geordnete Messreihe halbiert (daher auch *0.50-Quantil* genannt)

1. Meßwerte der Größe nach ordnen $x_1 \leq x_2 \leq \dots \leq x_n$
 2. Halbierungsstelle H identifizieren
 - bei geradem n : $H = n/2$
 - bei ungeradem n : $H = (n+1)/2$
 3. Median berechnen
 - bei geradem n : $\tilde{x} = (x_H + x_{H+1})/2$
 - bei ungeradem n : $\tilde{x} = x_H$
- geeignet bei Ordinaldaten oder "höheren" Skalen, kleinen Stichproben oder asymmetrischer Verteilung
 - unsensibel gegenüber Extremwerten



Kennwerte der *zentralen Tendenz*

Modus (Modalwert) \ddot{x}
der am häufigsten vorkommende Wert

1. Häufigkeit $f(x)$ für alle vorkommenden Kategorien bestimmen
2. Modus = Wert der meistbelegten Kategorie

bei mehreren benachbarten Spitzenwerten, Modus des arithmetischen Mittels

bei mehreren nicht-benachbarten Spitzenwerten, alle angeben

- geeignet bei Nominaldaten und bei mehrgipfliger Verteilung
- insgesamt weniger gebräuchlich



Zentrale Tendenz am Beispiel

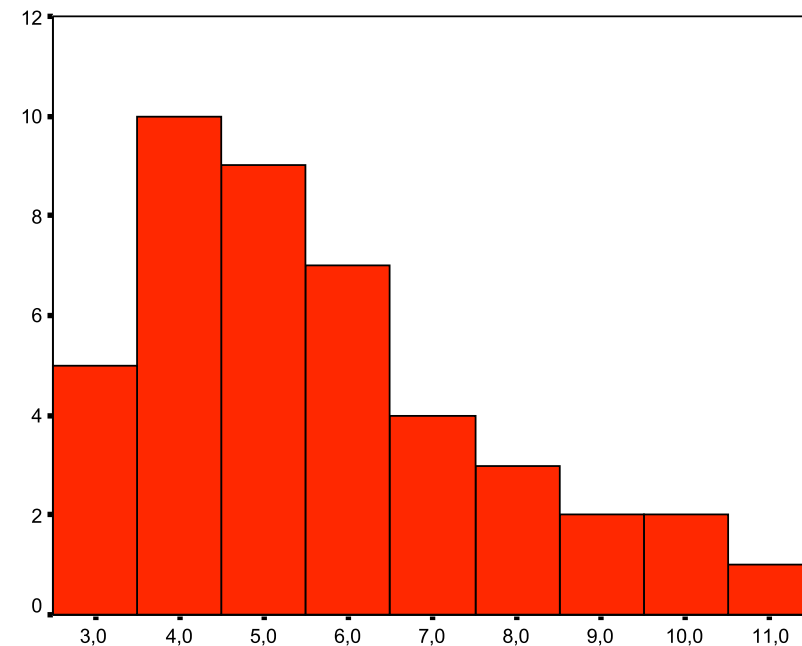
Häufigkeitsverteilung für eine Variable:

Wert	3	4	5	6	7	8	9	10	11
Häufigkeit	5	10	9	7	4	3	2	2	1

arithm. Mittel 5,65

Median 5

Modus 4



Kennwerte der *Dispersion (Streuung)*

Dispersion bezeichnet das Ausmaß, in dem sich die Verteilungen der Werte unterscheiden (wie "weit" oder "eng" sie sich um die zentrale Tendenz gruppieren).

Varianz und Standardabweichung s^2 s

Summe der quadrierten Abweichungen vom arithm. Mittel bezogen auf ihre Anzahl (Standardabweichung: Wurzel davon)

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1} \quad s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}$$

- geeignet bei Intervalldaten oder "höheren" Skalen, großen Stichproben und symmetrischer Verteilung



Kennwerte der *Dispersion*

Quartilsabstand (Interquartilbereich) q

absolute Differenz der Quartile, d.h. der Mediane der am Median geteilten Messreihe (0.25-Quantil bis 0.75-Quantil)

1. Meßwerte der Größe nach ordnen

2. Median bestimmen

$$x_1 \leq x_2 \leq \dots \leq x_n$$

3. unteres Quartil (Median aller $x_i \leq \tilde{x}$) bestimmen

4. oberes Quartil $q_{0.25}$ (Median aller $x_i \geq \tilde{x}$) bestimmen

5. absoluten Quartilsabstand berechnen

$$q = (|q_{0.75} - q_{0.25}|)$$

□ geeignet bei Ordinaldaten oder "höheren" Skalen, kleinen Stichproben oder asymmetrischer Verteilung

□ ebenso werden manchmal Abstände zwischen anderen Quantilen betrachtet, z.B. der *Interdezilbereich* $d = (|q_{0.9} - q_{0.1}|)$



Kennwerte der *Schiefe*

Schiefe bezeichnet anschaulich, wie stark rechts- oder linkslastig eine Verteilung ist:

Schiefe < 0: Verteilung ist linksschief (rechtssteil)

Schiefe > 0: Verteilung ist rechtsschief (linkssteil)

Schiefe = 0: Verteilung ist symmetrisch

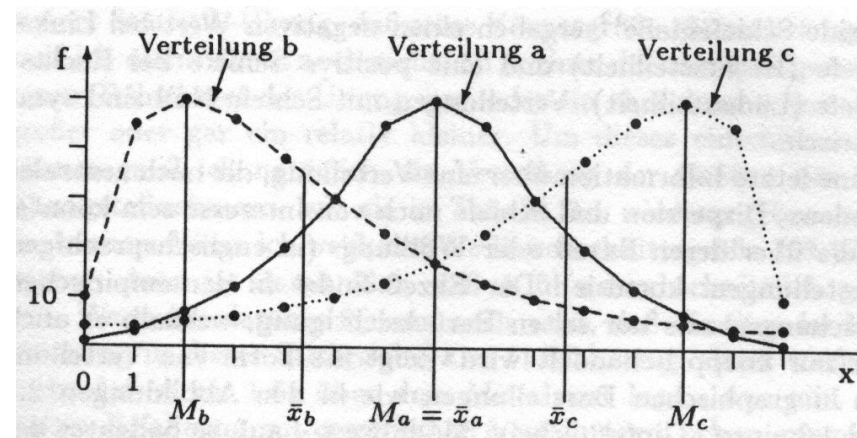
Pearsonsches Schiefemaß

Verhältnis der Differenz zwischen Mittelwert und Modus zur Standardabweichung

$$S_p = \frac{\bar{x} - M}{s}$$

Bowley-Fishersches Schiefemaß

$$S_b = \frac{\sum_i (x_i - \bar{x})^3 / n}{s^3}$$



Inferenzstatistik

Zentrale Frage der Inferenzstatistik:

Wie können auf Basis empirischer Daten
(Wahrscheinlichkeits-)Aussagen über die
Allgemeinheit getroffen werden?

Hier Fokus auf

- Prinzip inferenzstatistischer Analysen
- ein einfaches Beispiel
- Überblick über statistischer Testverfahren



Experimental design - factors

□ Subjects

- who – representative, sufficient sample size

□ Variables

- things to modify and measure

□ Conditions

- experimental conditions, differ only in the value(s) of some controlled variable(s)

□ Hypothesis

- what you'd like to show
- derived from literature or some sort of `theory` (not from data!)



Variables

□ *independent* variable (IV)

- characteristics changed to produce different conditions
e.g. interface style, number of menu items
- also called *controlled* variables

□ *dependent* variable (DV)

- characteristics measured in the experiment
e.g. time taken, number of errors



Hypotheses

- formulate as if-then or the-the („je..desto“) statement
- formulate in three steps
 1. in terms of the underlying theory
"Alle Computerbenutzer können das Programm xy schneller mit der Maus als mit der Tastatur bedienen."
 2. in terms of the variables
"Wenn bei Programm xy Interaktionsgerät = Maus, dann Interaktionszeit < als wenn Interaktionsgerät = Tastatur."
 3. in terms of statistical measures („Kennwerte“)

$$\bar{x}(INTERAKTIONSZEIT_{MAUS}) < \bar{x}(INTERAKTIONSZEIT_{TASTATUR})$$

- Again, need to frame theoretical concepts in statistical terms (*Operationalization*)



Hypotheses

- Statistical formulation calls for comparison of test series under different conditions
- Formulate and test possible explanations

- *Working hypothesis or alternative hypothesis H_1*
 - differences in test series are systematic and due to changes in controlled variables (IVs)
 - H_1 states expected outcome (how IVs influence DVs) $\bar{x}_A \neq \bar{x}_B$

- *Null hypothesis H_0 :*
 - there is no difference between conditions other than random variation
 - contraposition to working hypothesis
 - aim is to disprove this
e.g. null hypothesis = "no change with font size" $\bar{x}_A = \bar{x}_B$



Principle of statistical tests

Disprove the *null hypothesis*, i.e. prove that differences between the conditions did *not* happen by chance.

Note:

Statistical conclusions are always generalizations from a sample to an overall population, where the sample will *always* be affected by random variation. There are thus no absolute decisions against the null hypothesis, but only probabilities of their (in)validity!

Do not reject the null hypothesis before the results disprove it with a sufficient probability (significance).



Signifikanz & Fehler

Signifikanz

Ergebnis der Analyse ist Wahrscheinlichkeitsaussage über Bedeutsamkeit (Signifikanz) des Unterschieds bzw. des Zusammenhangs

Fehler

zwei Arten von Fehlentscheidungen sind möglich:

- α – Fehler: Entscheidung für H_1 obwohl H_0 zutrifft
- β – Fehler: Entscheidung für H_0 obwohl H_1 zutrifft

In der Praxis ist man v.a. daran interessiert, die Wahrscheinlichkeit eines α – *Fehlers* zu minimieren.



Signifikanzniveaus

Die *Irrtumswahrscheinlichkeit*, d.h. die WK einer Fehlentscheidung unter der Annahme, die H_0 treffe zu, ist berechenbar.

Als vertretbar gilt: α – *Fehlerrisiko* von *max. 5 %*

Schreibweise: $p(\alpha) \leq 0.05$

Lesart: *Meßreihen-Unterschied* bzw. *-Zusammenhang* ist mit einer *Wahrscheinlichkeit* von 0.95 oder mehr *systematisch bedingt*

- Aussage auf 95%-Niveau $p(\alpha) \leq 0.05$ "*signifikant*"
- Aussage auf 99%-Niveau $p(\alpha) \leq 0.01$ "*sehr signifikant*"
- Aussage auf 99.9%-Niveau $p(\alpha) \leq 0.001$ "*hochsignifikant*"



Experimental design

Goal: controlled evaluation of aspects of interactive behavior

1. define appropriate task (must encourage cooperation)
2. define variables (IV, DV)
3. formulate hypothesis to be tested in terms of variables
4. choose conditions to test; changes in measure are attributed to different conditions; *control* condition without variable manipulation
5. choose how to gather data
6. choose *statistical technique* to test the hypotheses
7. Before you start to do any statistics
 - look at data, check for *outliers*
 - save original data



Choice of statistical test – depends on...

- type of data/variables
 - discrete - can take finite number of values (*levels*)
 - continuous - can take any value
 - ranking scale – interval, nominal, etc.
- type of random experimental variation
 - DVs are subject to random errors
 - do they follow a known probability distribution?
- required information
 - is there a difference between...
 - distributions? Anpassungstests
 - frequencies? Häufigkeitstests
 - means? Unterschiedstests
 - dispersions? Homogenitätstest
 - correlation of test series? Zusammenhangstests
 - influential factors? Varianzanalyse
 - how accurate is the estimate?



Grundmuster statistischer Tests

- Ausgangspunkt: zwei statistische Testhypothesen H_0 und H_1
- Bestimmung einer *Prüfgröße* aus den Stichprobendaten
- Bestimmung einer *Prüfverteilung*, die bei Gültigkeit der H_0 gegeben wäre
- Signifikanzniveau (akzeptables α -Risiko) festlegen
- durch das Signifikanzniveau ist der H_0 -Verwerfungsbereich der Prüfverteilung festgelegt
- fällt die Prüfgröße in den H_0 -Verwerfungsbereich, dann kann die H_1 angenommen werden

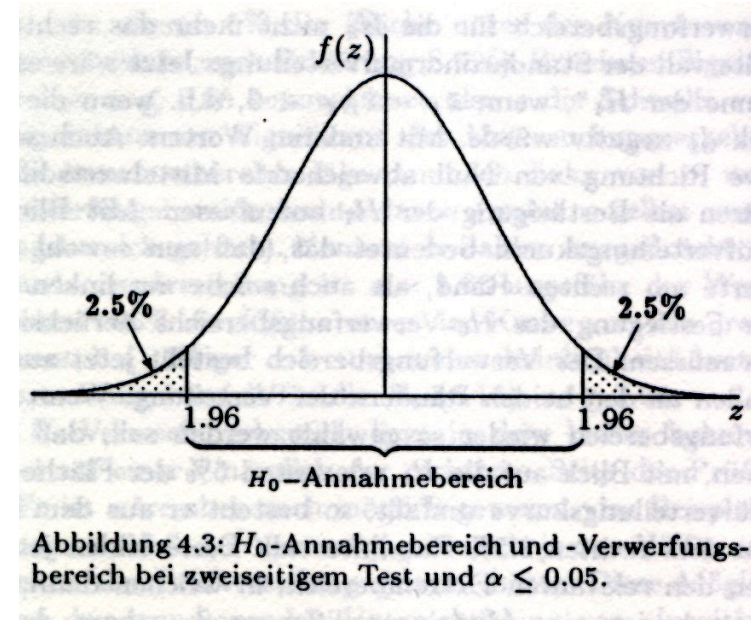
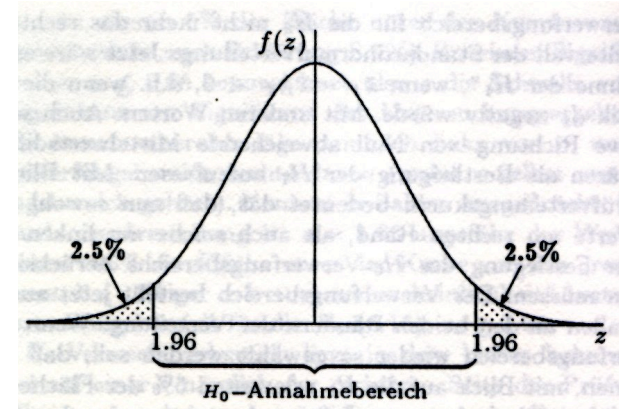


Abbildung 4.3: H_0 -Annahmebereich und -Verwerfungsbereich bei zweiseitigem Test und $\alpha \leq 0.05$.



Analysis - types of test

- *parametric*
 - powerful
 - assume normal distribution of DV
 - robust (give reasonable results also when data not exactly normal)
 - Example: completion time of complex task depends on *independent* subtasks



- *non-parametric*
 - less powerful, more reliable
 - do not assume normal distribution
 - Example: subjective usability rating

- *contingency table*
 - classify data by discrete attributes
 - count number of data items in each group

		Relevant R	Not Relevant \tilde{R}
Retrieved G		$G \cap R$	$G \cap \tilde{R}$
Not Retrieved \tilde{G}		$\tilde{G} \cap R$	$\tilde{G} \cap \tilde{R}$



Statistical test by form of IV and DV

	<i>IV</i>	<i>DV</i>	<i>Test</i>
<i>Parametric</i>	Two-valued	Normal	Student's t-test on difference of means
	Discrete	Normal	ANOVA (ANalysis Of VAriance)
	Continuous	Normal	(Non-)linear regression factor analysis
<i>Non-parametric</i>	Two-valued	Cont.	Wilcoxon/Mann-Whitney rank-sum test
	Discrete	Cont.	Rank-sum versions of ANOVA
	Continuous	Cont.s	Spearman's rank correlation
<i>Contingency test</i>	Two-valued	Discrete	No special test, see next entry
	Discrete	Discrete	Contingency table and Chi-squared test
	Continuous	Discrete	Group indep. Variable and then as above



Summary

- Use statistics to describe experimental data and to test hypotheses on them.
- Statistics can be (roughly) divided in:
descriptive statistics and inferential statistics
- Methods are standardized – in science, everybody knows what you want to say
- Methods, especially of inferential statistics, are not quite easily applied; some experience needed, read text books!
- Make sure the statistical test you are using is applicable, check its requirements!
- Use software for analyzing the data



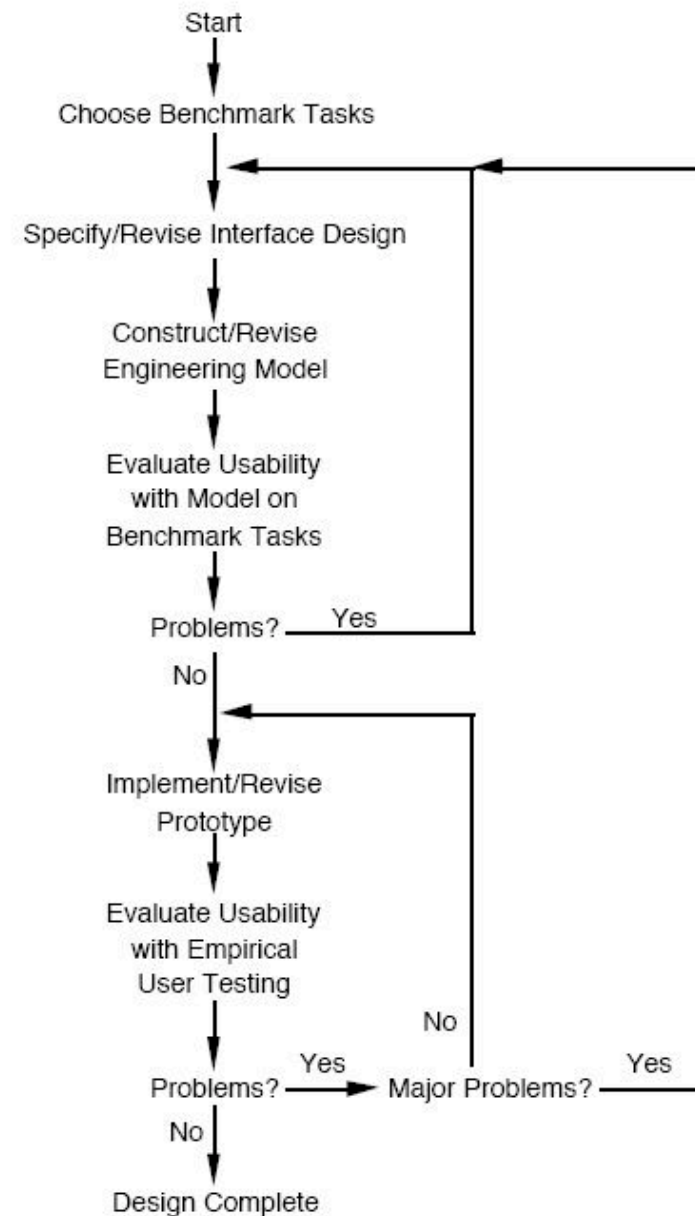
Model-based evaluation

Model-based evaluation

Four steps:

1. Describe interface design in detail
2. Build model of user doing a task
3. Use the model to predict execution or learning time
4. Revise or choose design depending on prediction

- Usability results *before* implementing prototype or user testing
- Engineering model allows more design iterations



Model-based approach

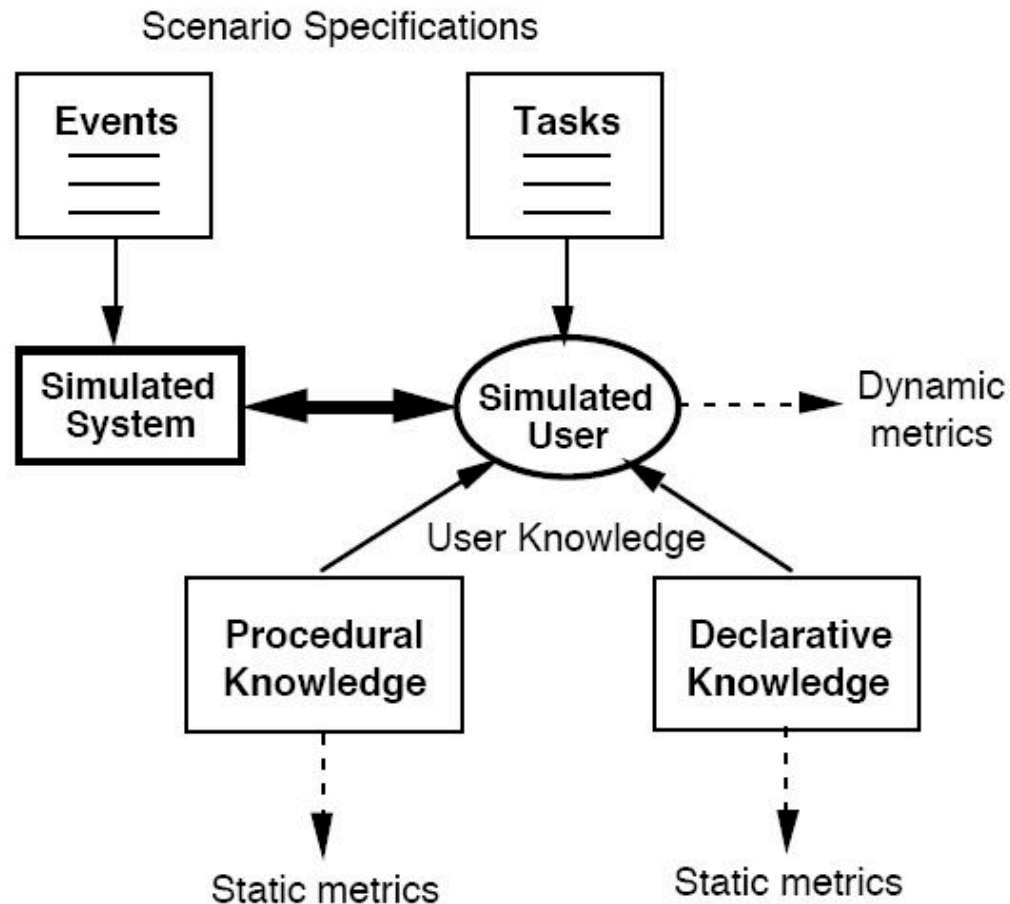
- Model summarizes the interface design from the user's point of view:
 - Represents how the user gets things done with the system.
 - Components of model can be reused to represent design of related interfaces.

- *But*, current models can only predict a few aspects:
 - Time required to execute specific tasks.
 - Ease of learning of procedures, consistency effects

- User testing still required!



Overview



Models = simulations of human-computer interaction

Procedural knowledge: how-to procedures → executable

Declarative knowledge: facts, beliefs → reportable



Psychological constraints

- Evaluation of a proposed design must be a *routine* activity, not a scientific research project.
- Need to be able to build models without inventing psychological theory.

- Modeling system must provide human psychological constraints *automatically*
 - Constrain what the model can do, so modeler can focus on design questions, not psychological basics
 - If model can be programmed to do any task at any speed or accuracy, something's wrong!



Cognitive vs perceptual-motor constraints

□ What dominates a task?

- Heavily cognitive tasks: Human “thinks” most of the time, e.g. stock trading system



□ Many HCI tasks dominated by *perceptual-motor activity*

- A steady flow of physical interaction between human and computer („doing rather than thinking“)
- Time required depends on human characteristics and computer’s behavior (determined by the design)

□ Implication

- Modeling perceptual-motor aspects is often practical, useful, and relatively easy.
- Modeling purely cognitive aspects of complex tasks is often difficult, open-ended, and requires research resources.



Modeling approaches

Three current approaches:

1. Task network models – before detailed design
2. Cognitive Architecture Models – packaged constraints
3. GOMS models – relatively simple & effective

Differ in constraints, detail, when to use.

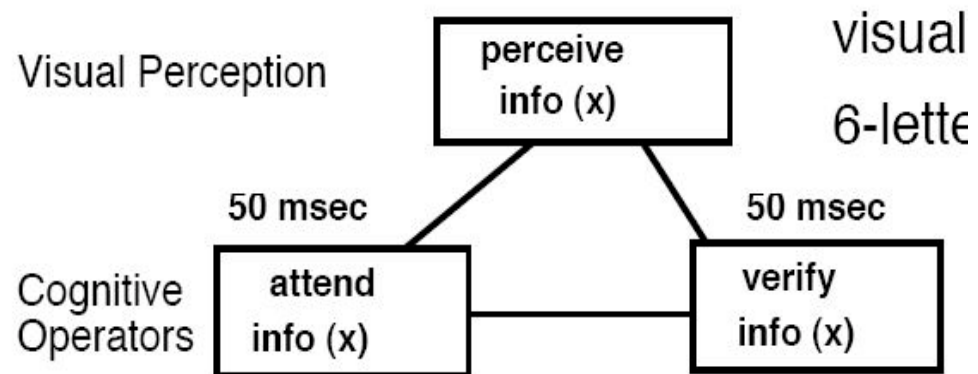


Task Network Models

- Connected network of tasks:
 - Connection: one task is a prerequisite of the other
 - Both serial and parallel execution of tasks
 - Final completion time computed from chain of serial and parallel tasks
 - *Critical path* = chain with largest execution time
 - PERT charts (*Program Evaluation & Review Techn.*), (E) TAGs
- Tasks = mixture of human and machine tasks
- Each task characterized by a distribution of completion times, and arbitrary dependencies and effects



Task network - example



Cognitive architectures

Represent basic human abilities and limitations.

“Programmed” with a strategy to perform specific tasks.

- provides constraints on the form and content of the strategy.
- Architecture + specific strategy = a model of a specific task.

To model a specific task:

- Do a task analysis to arrive at human’s strategy for the task.
- “Program” the architecture with representation of strategy.
- Run the model using task scenarios.

Result: predicted behavior and time course for that scenario and task strategy.

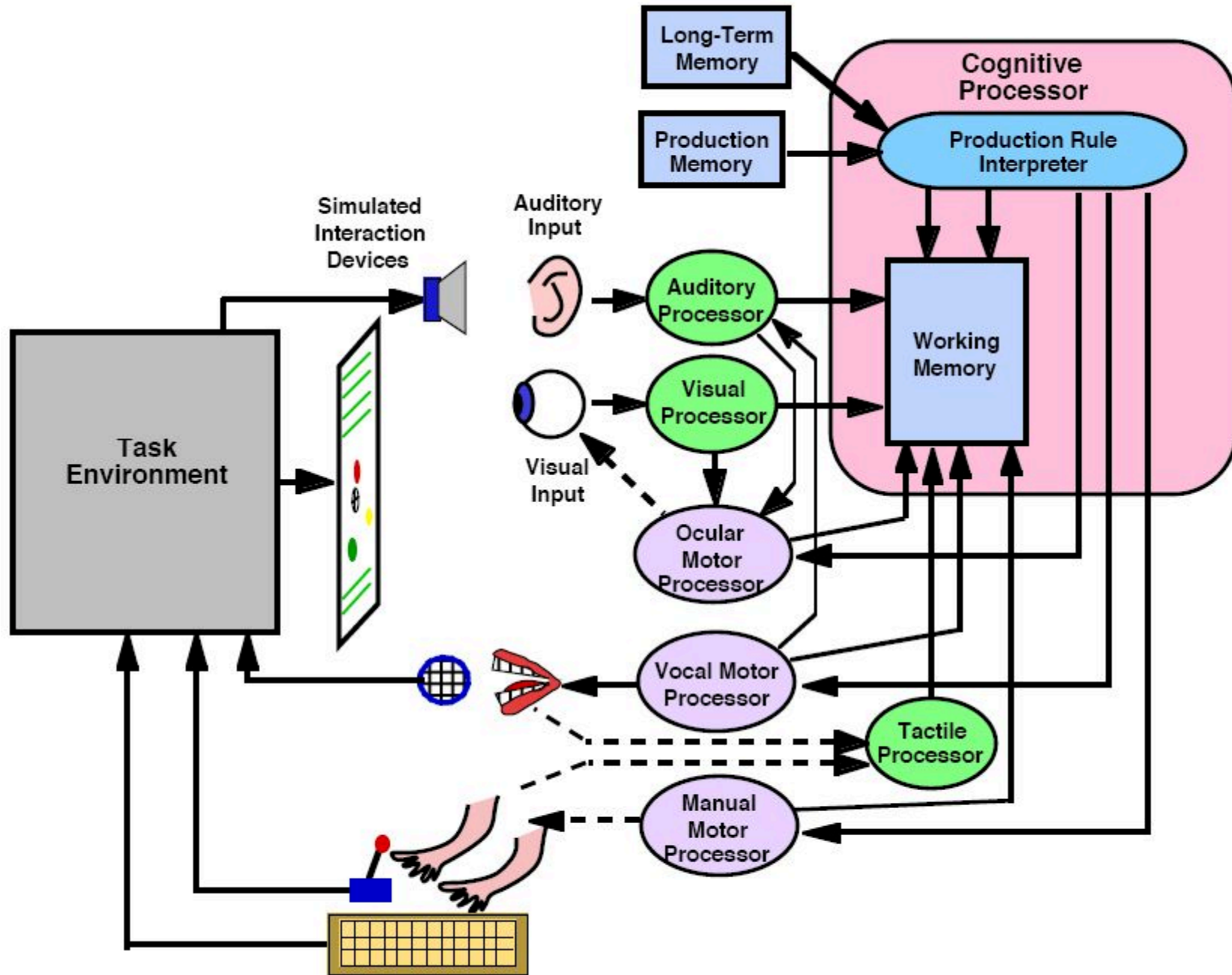
Needs comprehensive psychological theory, so these are quite complex;
used mostly in a research settings



Example: EPIC Architecture

- Developed to represent executive processes that control other processes during *multiple task* performance.
- **Executive-Process Interactive Control**
(Kieras & Meyer, mid-1990s)
- Basic assumptions
 - Production-rule cognitive processor.
 - Parallel perceptual and motor processors.
- Fixed architectural properties
 - Components, pathways, and most time parameters
- Task-dependent properties
 - Cognitive processor production rules (strategy).
 - Perceptual recoding.
 - Response requirements and styles.





GOMS (Card, Moran, & Newell, 1983)

- A key model-based methodology based on simplified cognitive architectures.
- An approach to describing the knowledge of *procedures* that a user must have in order to operate a system
 - **G**oals - what goals can be accomplished with the system
 - **O**perators - what basic actions can be performed
 - **M**ethods - what sequences of operators can be used to accomplish each goal
 - **S**election Rules - which method should be used to accomplish a goal
- Well worked out, quite practical, but limited due to simplifications
- Often in the "sweet spot" - lots of value for modest modeling effort



Keystroke-level model

1. Choose one or more representative task scenarios.
2. Have design specified to the point that keystroke-level actions can be listed.
3. List the keystroke-level actions (operators) involved in doing the task.
4. Insert mental operators for when user has to stop and think.
5. Look up the standard execution time to each operator.
6. Add up the execution times for the operators.
7. The total is the estimated time to complete the task (sum of times for tasks t_i multiplied by frequency n_i)

$$T_{execute} = \sum_i t_i * n_i$$



KLM – operators and times

K - Keystroke (.12 - 1.2 sec; use .28 sec for ordinary user).

- Pressing a key or button on the keyboard.
- Different experience levels have different times.
- Pressing SHIFT or CONTROL key is a separate keystroke.
- Use type operator T(n) for a series of n Ks done as a unit.

P - Point with mouse to a target on the display.

- Follows Fitts' law - use if possible: $0.1 * \log_2 (D/S + 0.5)$
- Typically ranges from .8 to 1.5 sec, average (text editing) is 1.1 sec.

B - Press/release mouse button (.1 sec; click is .2).

- Highly practiced, simple reaction.



KLM – operators and times

H - Home hands to keyboard or mouse (.4 sec).

W - Wait for system response.

- Only when user is idle because can not continue
- Have to estimate from system behavior
- Often essentially zero in modern systems

M - Mental act of thinking.

- Represents pauses for routine activity (not problem-solving).
- New users often pause to remember or verify every step.
- Experienced users pause and think only when logically necessary.
- Estimates ranges from .6 to 1.35 sec; 1.2 sec is good single value.



Example: file deletion in MacOS, original design, experienced user

General procedure

- Find the file icon to be deleted and drag it to the trash can.

Assumptions:

- user thinks of selecting+dragging icon as a single operation.
- Finding to-be-deleted icon is still required
- Moving icons to the trash can is highly practiced:
 - The trash can does not have to be located, so finding the trash can is overlapped with pointing to it.
 - Verifying that the trash can has been hit is overlapped with pointing to it.
 - Final result (bulging can) is not checked since it is redundant with verifying that the can has been hit.

Operator sequence:

initiate the deletion **M**, find the file icon **M**, point to file icon **P**,
press and hold mouse button **B**, drag file icon to trash can icon **P**,
release mouse button **B**, point to original window **P**

- **Total time = 3P + 2B + 2M = 5.9 sec**



Example: command key file deletion, experienced user

General procedure

- Select the file icon to be deleted and hit a command key.

Assumptions

- User operates both mouse + key with right hand.
- Right hand starts and ends on the mouse.

Operator sequence: initiate the deletion **M**, find the icon for the to-be-deleted file **M**, point to file icon **P**, click mouse button **BB**, move hand to keyboard **H**, hit command key **KK**, move hand back to mouse **H**

- **Total time = P + 2B + 2H + 2K + 2M = 5.06 sec**

Only slightly faster, due to need to move the hand!



Other models in GOMS family

- Critical-Path Method GOMS (CPM-GOMS)
 - Express activities in terms of Model Human Processor → task network → analyze for critical path
- Natural GOMS Language (NGOMSL)/ Cognitive Complexity Theory (CCL)
 - basic GOMS concept as simple production system
 - hierarchical actions as sequential/hierarchical rules, eventually keystroke level operators
- Executable GOMS Language (GOMSL)/GLEAN
 - Formalized and executable version of NGOMSL.
 - *GLEAN* - a simplified version of the EPIC simulation system (**G**OMS **L**anguage **E**valuation and **A**nalysis)



Model-based vs. inspection evaluation

	<i>Cognitive walkthrough</i>	<i>Heuristic evaluation</i>	<i>Model-based</i>
<i>Stage</i>	Throughout	Throughout	Design
<i>Style</i>	Lab	Lab	Lab
<i>Objective?</i>	No	No	Somewhat
<i>Measure</i>	Qualitative	Qualitative	Qual. & Quan.
<i>Information</i>	Low level	High level	Low level
<i>Immediacy</i>	N/A	N/A	N/A
<i>Intrusive?</i>	No	No	No
<i>Time</i>	Medium	Low	Medium
<i>Equipment</i>	Low	Low	Low
<i>Expertise</i>	High	Medium	High

