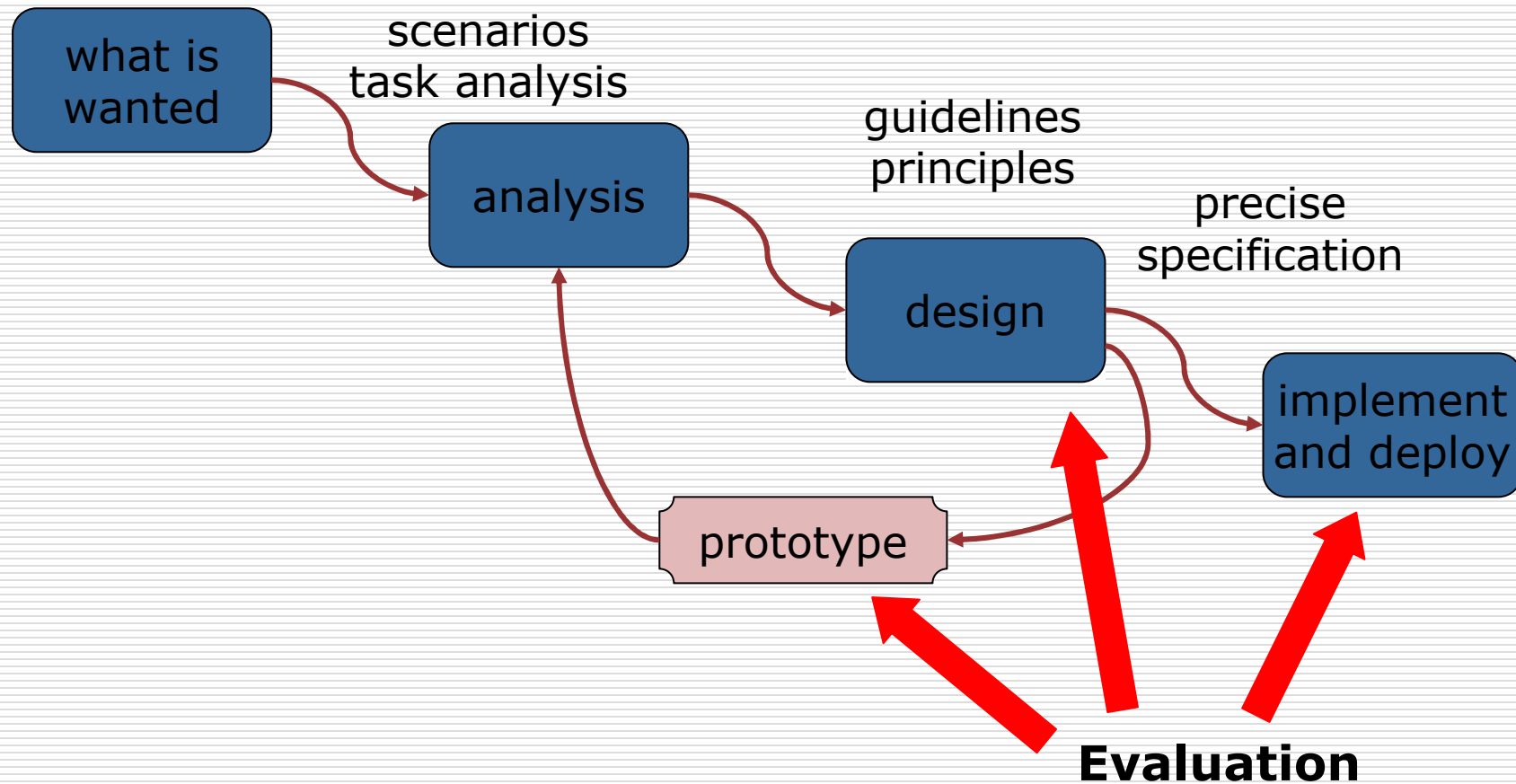


Human-Computer Interaction

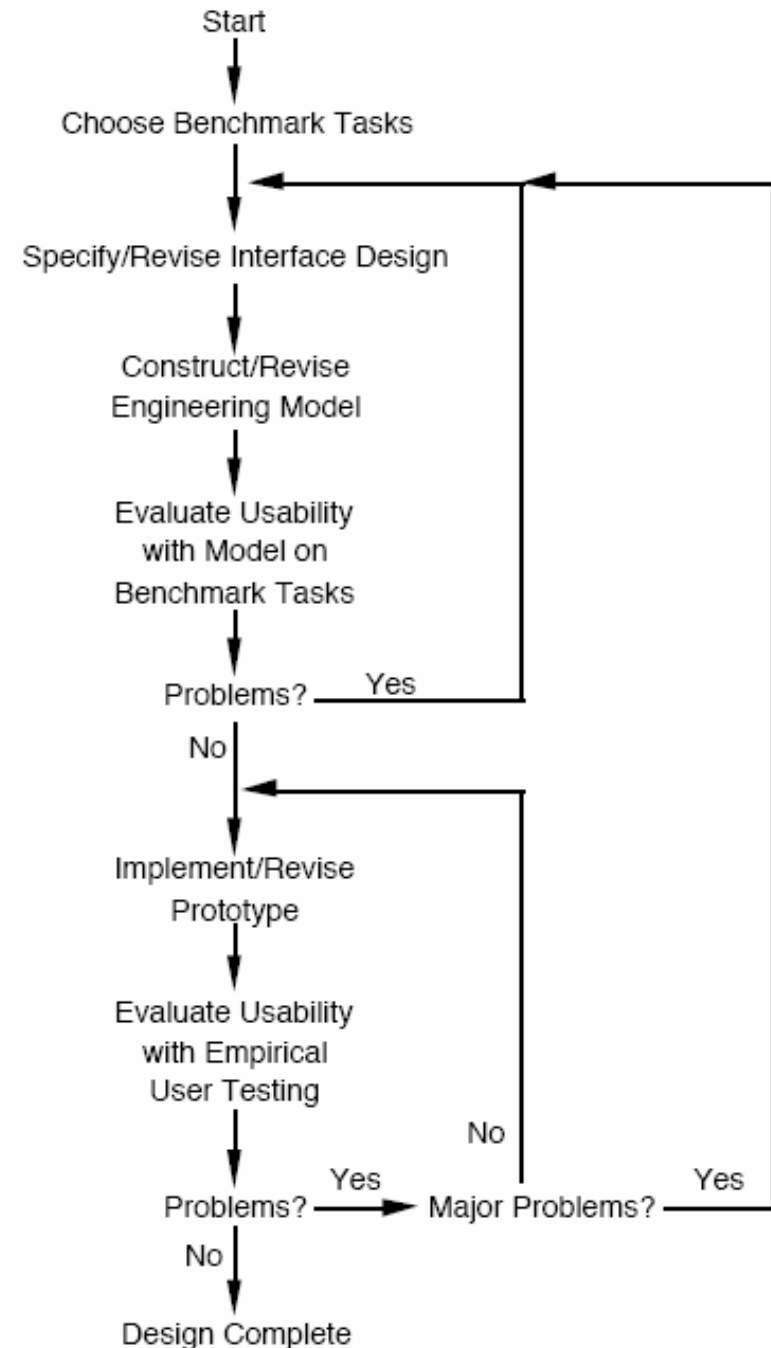
Lecture 8: Evaluating interactive systems II

Designing a usable system



Evaluation

- Analytic models
 - Before prototype implementation
 - Predict execution or learning time, consistency
 - (cf. last course)
- User testing
 - Assess aspects not dealt with by analytic models (user, task, context, system)
 - Protection against errors or oversights in the analysis
 - Sometimes with expert users (cogn. walkthrough, heuristic evaluation)



Evaluation through user participation

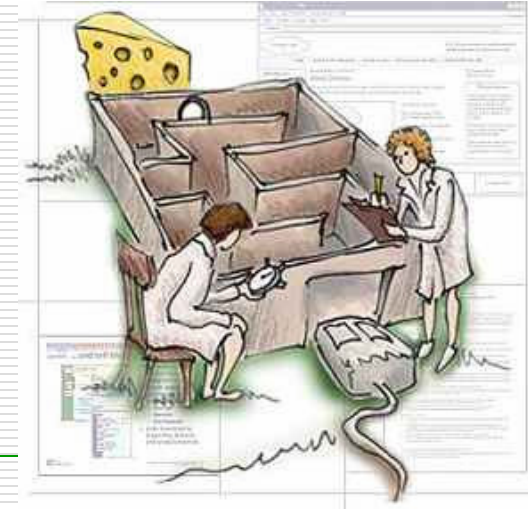
- Needs at least a working prototype
 - Possibly use of *Wizard of Oz* technique, where part of the system functionality is simulated by a wizard.

- Number of different ways to study how the user is doing with the system
 - Query techniques (cf. last course)
 - Observational methods (cf. last course)
 - Empirical/experimental methods
 - Physiological monitoring

- Lab vs field studies



experimental methods

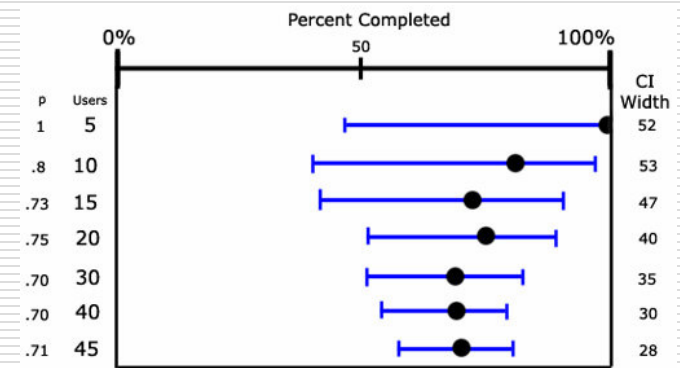
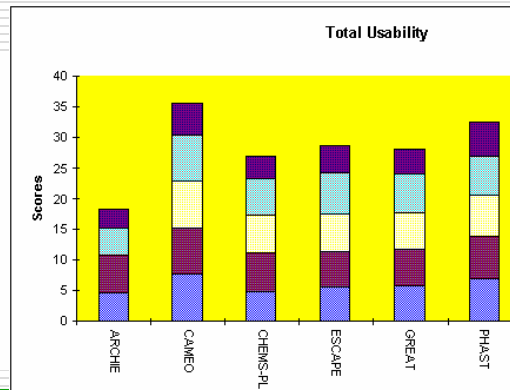
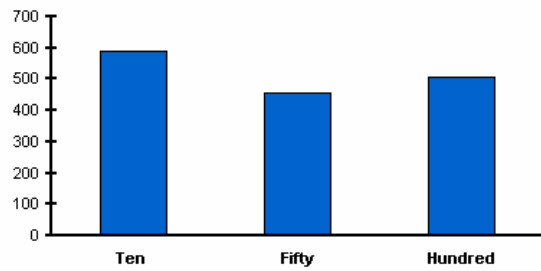
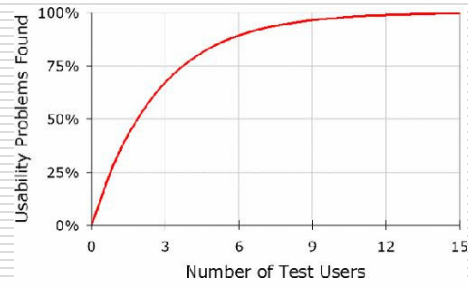


evaluate artefacts: simulation, prototype, full implementation

Readings:

Lorenz Sichelschmidt: "Elementare Statistik", Skript, LiLi-Fakultät.
Oliver Schilling: "Grundkurs: Statistik für Psychologen", München:
Fink Verlag, 1998.

Dix et al., „Human-Computer Interaction“, chapt.9



MMI / SS05

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 *Wertes* einer *Skala* zu einem beobachtbaren *Merkmal* eines Objekts
 - 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 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*
 2. *Ordinalskala*
 3. *Intervallskala*
 4. *Verhältnisskala*
 5. *Absolutskala*



Skalenniveaus: Nominalskala

- erfaßt nur Gleichheit bzw. Unterschiedlichkeit von Objekten bzgl. der Merkmalsausprägung:

Objekte mit gleicher Merkmalsausprägung werden gleichen Werten, Objekte mit verschiedener Merkmalsausprägung verschiedenen Werten zugeordnet.

- erlaubt Aussagen über Kategorienzugehörigkeit ("gleich/ungleich"), aber nicht über Ordnung
- definierte Relationen: =, ≠
- Beispiele: Blutgruppen A, B, AB, Null; Geschlecht m., w.



Skalenniveaus: Ordinalskala

- (auch *Rangskala*) – drückt Rangfolge der Meßobjekte bzgl. der Merkmalsausprägung aus:

Objekte mit größerer Merkmalsausprägung werden größere, Objekte mit kleinerer Merkmalsausprägung kleinere Werte zugeordnet.

- erlaubt Aussagen über Rangordnung ("mehr/weniger"), Abstände nicht interpretierbar
- definierte Relationen: =, ≠, >, <
- Beispiele: Kleidungsgrößen S, M, L, XL, XXL; Urteile --, -, 0, +, ++



Skalenniveaus: Intervallskala

- erfaßt neben der Rangfolge auch den Abstand der Meßobjekte bzgl. der Merkmalsausprägung:

Objekten werden Werte so zugeordnet, daß die Größenordnung von Wertdifferenzen der von Merkmalsunterschieden entspricht.

- erlaubt Aussagen über Größenunterschiede ("wieviel mehr/weniger"), fast alle statistischen Tests sind anwendbar
- definierte Relationen bzw. Operationen: =, ≠, >, <, +, -
- Beispiele: Temperatur in °C; Uhrzeit *hh:mm:ss*
- Bildung von Quotienten ist aber nicht zulässig
 - 10°C in Köln, 20°C in Mailand → doppelt so warm?
 - in Fahrenheit (lineare Transformation!): 59°F vs 68°F



Skalenniveaus: Verhältnisskala

- definiert zusätzlich einen absoluten Nullpunkt:

Objekten werden Werte so zugeordnet, daß die Größenverhältnisse der Werte denen der Merkmalsausprägungen entsprechen.

- erlaubt Aussagen über Größeverhältnisse ("wieviel mal mehr/ weniger") und Bildung von Quotienten
- definierte Relationen bzw. Operationen: =, ≠, >, <, +, -, ·, ÷
- Beispiele: Länge in mm; Zeitintervalle in s; Temperatur in Kelvin



Skalenniveaus: Absolutskala

- definiert eine absolute Zuordnung:

Objekten werden Werte so zugeordnet, daß die absoluten Werte den Merkmalsausprägungen entsprechen.

- Beispiel: Anzahl der Kinder, Stückzahlen
- Metrische Skala mit natürlichen Einheiten und natürlichem Nullpunkt



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



Datenmatrix

- Grundlage der statistischen Auswertung (*Urliste*)
- Zeilen: *Beobachtungseinheiten (Fälle)*; z.B. Personen in der Psychologie, Sätze in der Linguistik, Interaktionen in der MMK
- Spalten: beobachtete Merkmale (*Variablen*)
- Zellen: *Daten*, d.h. beobachtete *Werte (Ausprägungen)* der Variablen
- Fälle bilden i.a. eine *Stichprobe*, d.h. eine Teilmenge der *Grundgesamtheit (Population)*

Fall-Nr.	Geschlecht	Gewicht	Alter
1	w	82	36
2	w	71	29
3	m	75	29
4	w	62	17



descriptive statistics

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?

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 *schiefen* 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

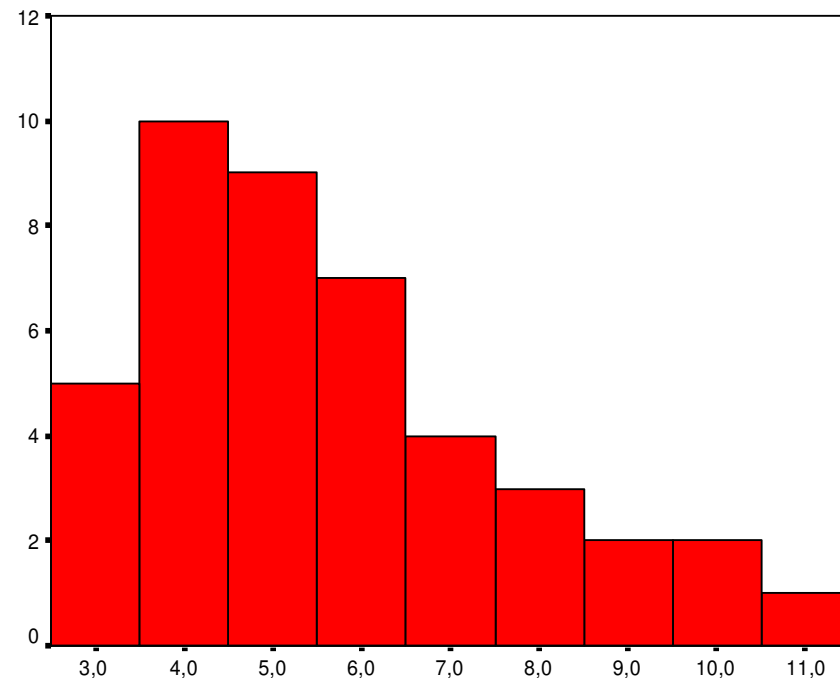
Häufigkeitsverteilung für eine Variable:

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 $x_1 \leq x_2 \leq \dots \leq x_n$
2. Median \tilde{x} bestimmen
3. unteres Quartil $q_{0.25}$ (Median aller $x_i \leq \tilde{x}$) bestimmen
4. oberes Quartil $q_{0.75}$ (Median aller $x_i \geq \tilde{x}$) bestimmen
5. absoluten Quartilsabstand $q = (q_{0.75} - q_{0.25})$ berechnen

- 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 *Dispersion*

Spannweite w

Differenz zwischen dem größten und dem kleinsten Wert der Meßreihe

$$W = x_{\max} - x_{\min}$$

- weniger gebräuchlich, da nur Extremwerte berücksichtigt
- sinnvoll nur dann, wenn in der Domäne keine Ausreißer auftreten



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

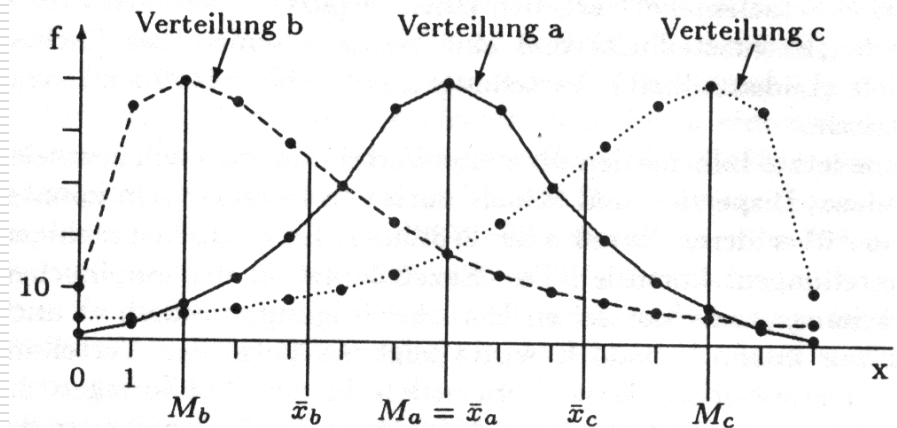
Pearsonsches Schiefemaß

Verhältnis der Differenz zwischen Mittelwert und Modus zur Standardabweichung

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

Bowley-Fishersches Schiefemaß

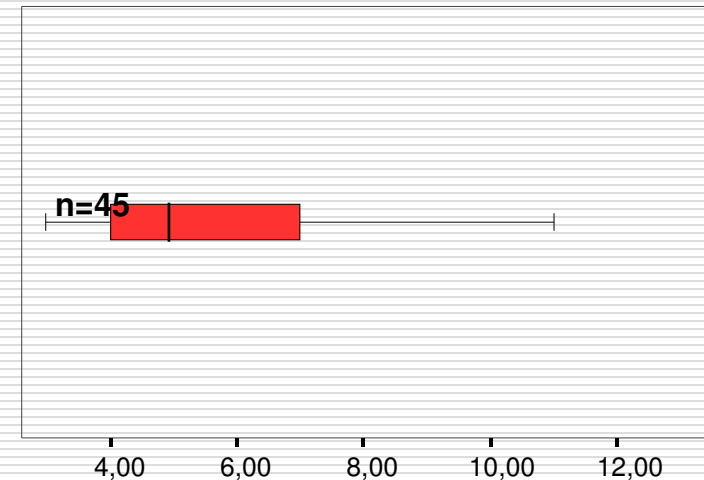
$$S_b = \frac{\sum_i (x_i - \bar{x})^3 / n}{s^3}$$



Graphische Darstellungen

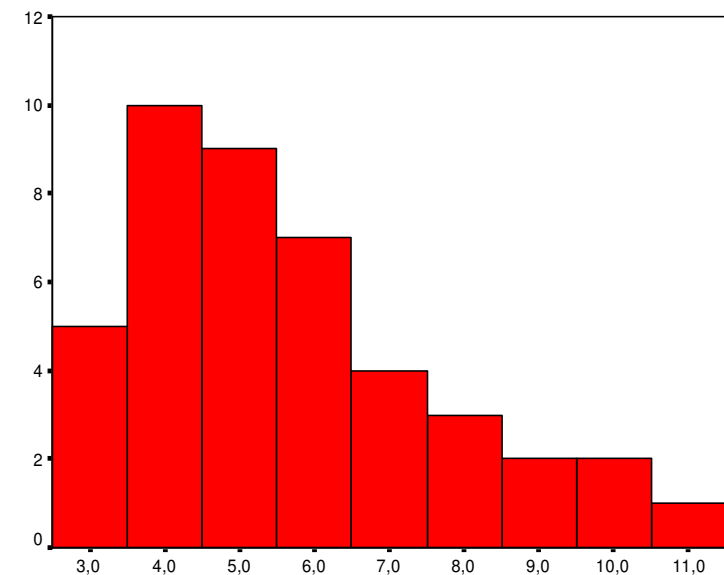
Box-and-Whisker Diagramm

- mind. ordinalskalierte Daten
- senkr. Mittelline: Medianwert
- Box: Quartilsabstand
- waagerechte Linie ("whisker") mit Endmarkern: Spannweite



Histogramm

- Abszisse: intervallskalierte Variable
- Ordinate: Häufigkeit



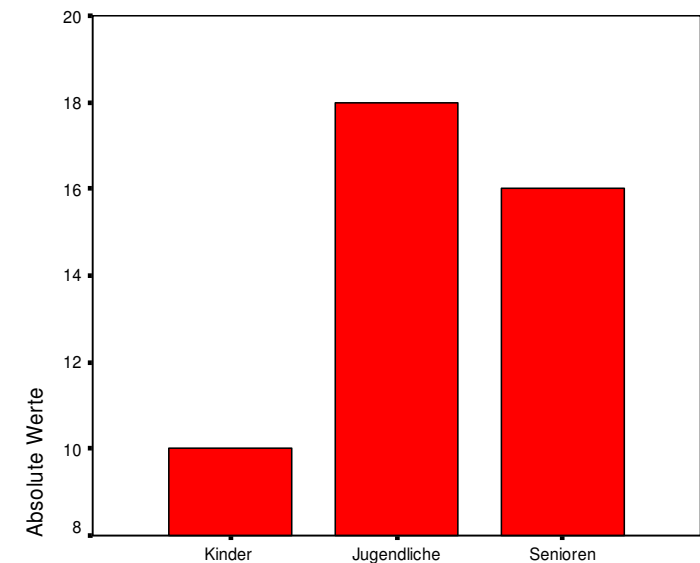
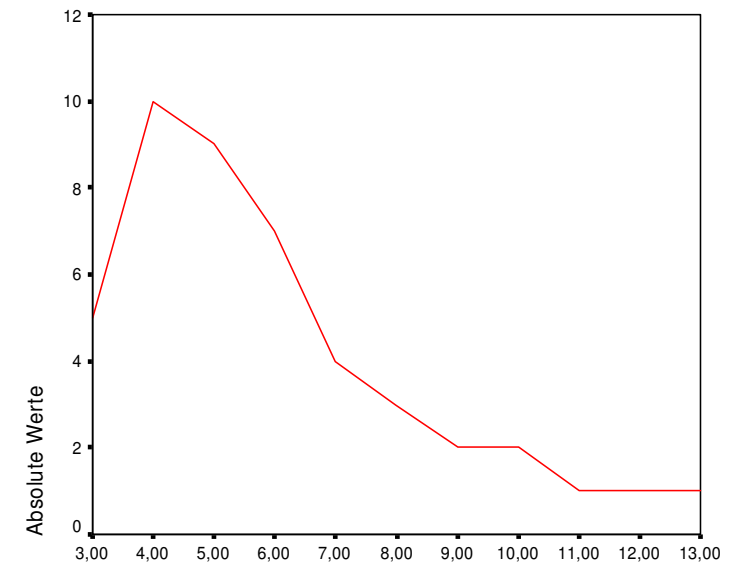
Graphische Darstellungen

Liniendiagramm

- geeignet zur Darstellung von Verläufen und Funktionszusammenhängen
- Abszisse: intervallskalierte Variable
- Ordinate: intervallskalierte Variable

Säulen- bzw. Balkendiagramm

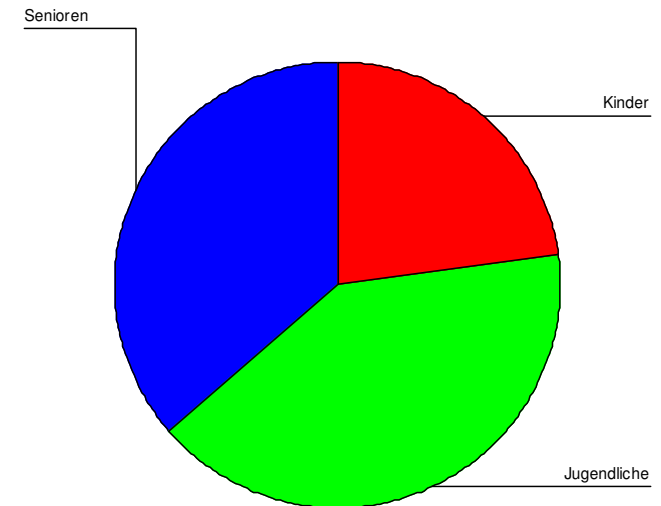
- geeignet zur Darstellung von absoluten Häufigkeiten oder Mittelwerten
- Abszisse: nominale, ordinale oder Intervall-Kategorien
- Ordinate: intervallskalierte Variable



Graphische Darstellungen

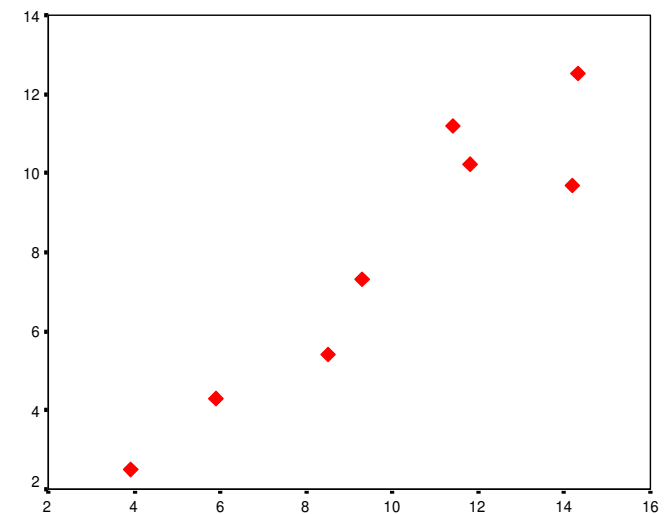
Kreisdiagramm ("Tortendiagramm")

- geeignet zur Darstellung von relativen Häufigkeiten nominalskalierteter Klassen
- Kreis (360°) entspricht Gesamthäufigkeit (100%), der Winkel dem Anteil der betreffenden Klasse



Streudiagramm (Scatterplot)

- geeignet zur Darstellung des Zusammenhangs zwischen intervallskalierten Variablen
- Abszisse: Variable X
- Ordinate: Variable Y
- Punkte repräsentieren Wertepaare



inferential statistics

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
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...

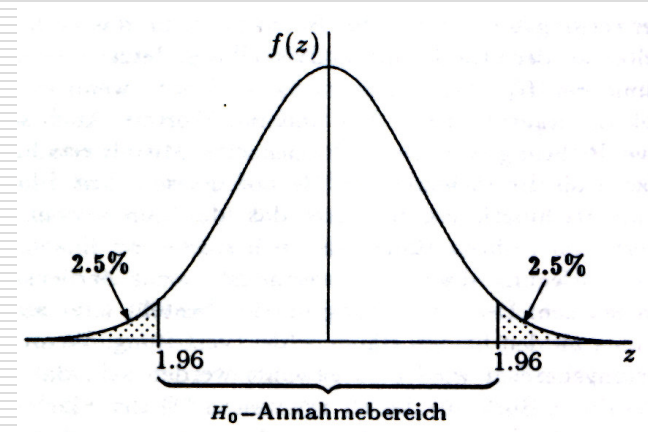
- type of data/variables
 - discrete - can take finite number of values (*levels*)
 - continuous - can take any value
 - ranking scale – interval, nominal, etc.
- type of random experimental variation
 - DVs are subject to random errors
 - do they follow a known probability distribution?
- required information
 - 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?



Analysis - types of test

□ *parametric*

- powerful
- assume normal distribution
- robust (give reasonable results 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 \bar{R}
Retrieved G		$G \cap R$	$G \cap \bar{R}$
Not Retrieved \tilde{G}		$\tilde{G} \cap R$	$\tilde{G} \cap \bar{R}$



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

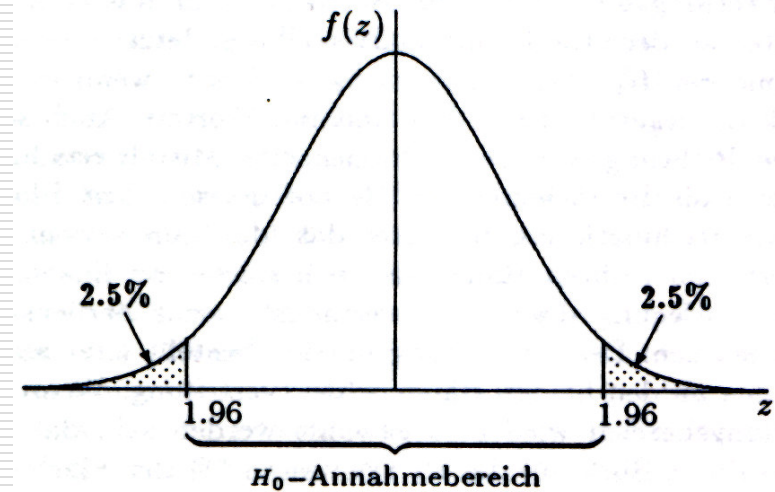


Abbildung 4.3: H_0 -Annahmehbereich und -Verwerfungsbereich bei zweiseitigem Test und $\alpha \leq 0.05$.



Beispiel: Der *t*-Test für Mittelwertsvergleiche

Charakterisierung

Der *t*-Test ist ein Verfahren aus der Gruppe der Unterschiedstests; er beantwortet die Frage:

"Unterscheiden sich die Mittelwerte zweier Meßreihen signifikant?"

Beispiel: Hypothese $\bar{x}(INT.ZEIT_{MAUS}) < \bar{x}(INT.ZEIT_{TASTATUR})$

Jeweils eine Meßreihe für die *Maus*- Bedingung, und eine für die *Tastatur*- Bedingung aufgenommen. Die deskriptiv-statistische Analyse ergibt tatsächlich ein deutlich kleineres arithm. Mittel der Variable Interaktionszeit in der *Maus*-Bedingung. Ist damit die Hypothese bewiesen?

Nein, wir müssen fragen: *"Ist der Unterschied signifikant, d.h. ist die Wahrscheinlichkeit, daß er zufällig zustande gekommen ist, gering?"*



Beispiel: Der t-Test für Mittelwertsvergleiche

Voraussetzungen

Der (einfache) t-Test stellt folgende Anforderungen

- 2 unabhängige Meßreihen (Fallgruppen A und B)
- Gruppe A mit n_a Fällen, Gruppe B mit n_b Fällen
- je 1 Meßwert x_i pro Fall
- mind. intervallskalierte Daten
- Normalverteilung
- homogene Varianzen

Berechnung

- Prüfgröße: "normierte" Differenz der Mittelwerte zweier Stichproben, z.B.: $\bar{x}(INT.ZEIT_{MAUS}) - \bar{x}(INT.ZEIT_{TASTATUR})$
- Prüfverteilung: sog. *t-Verteilung*



Beispiel: Der t-Test für Mittelwertsvergleiche

Berechnungsvorschrift

1. Mittelwerte \bar{x}_A, \bar{x}_B für beide Gruppen berechnen
2. Für beide Gruppen Summen der quadrierten Differenzen zwischen Wert und Mittelwert berechnen

$$Q_A = \sum_i (x_{Ai} - \bar{x}_A)^2 \quad Q_B = \sum_i (x_{Bi} - \bar{x}_B)^2$$

3. Freiheitsgrade berechnen $\nu = n_A + n_B - 2$
4. Mittlere Standardabweichung berechnen

$$s = \sqrt{((Q_A + Q_B) / \nu)}$$

5. Prüfgröße t berechnen

$$t = ((\bar{x}_A - \bar{x}_B) / s) \cdot \sqrt{((n_A n_B) / (n_A + n_B))}$$



Beispiel: Der t-Test für Mittelwertsvergleiche

Entscheidung

Der kritische t -Wert (für gegebene Anzahl von Freiheitsgraden und Alpha-Fehler) kann in einer Tabelle nachgeschlagen werden.

Wenn $t < \text{Tabellen-}t$, dann kein signifikanter Mittelwert-Unterschied

Wenn $t \geq \text{Tabellen-}t$, dann signifikanter Mittelwert-Unterschied

Anmerkung: Die Berechnung des t -Wertes sowie des Tabellen-Wertes erledigt man normalerweise mit einem Statistikprogramm (z.B. SPSS).



Statistische Tests zum Vergleich von zwei Meßreihen

Skala \ Meßreihen- typ	unabhängig	abhängig
Nominaldaten	Chi ² -Test nach Pearson	Chi ² -Test nach McNemar
Ordinaldaten	Rangsummen-Test nach Mann-Whitney	Rangsummen-Test nach Wilcoxon
Intervalldaten	<p>n > 80: z-Test f. unabh. Meßreihen</p> <p>n ≤ 80:</p> <ul style="list-style-type: none"> <input type="checkbox"/> gleiche Varianzen: t-Test nach Student <input type="checkbox"/> ungleiche Varianzen: t-Test nach Student/Welch 	<p>n > 80: z-Test f. abh. Meßreihen</p> <p>n ≤ 80:</p> <ul style="list-style-type: none"> <input type="checkbox"/> t-Test nach Student f. abh. Meßreihen



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

Dix et al., Human-Computer Interaction, p. 334



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



physiological methods

Eye tracking

Physiological measurement

eye tracking

- ❑ head or desk mounted equipment tracks the position of the eye
- ❑ eye movement reflects amount of cognitive processing a display requires
- ❑ measurements include
 - fixations: eye maintains stable position. Number and duration indicate level of difficulty with display
 - saccades: rapid eye movement from one point of interest to another
 - scan paths: moving straight to a target with a short fixation at the target is optimal



physiological measurements

- emotional response linked to physical changes
- may help determine a user's reaction to an interface
- measurements include:
 - heart activity, including blood pressure, volume and pulse.
 - activity of sweat glands: Galvanic Skin Response (GSR)
 - electrical activity in muscle: electromyogram (EMG)
 - electrical activity in brain: electroencephalogram (EEG)
- some difficulty in interpreting these physiological responses - more research needed



Choosing an Evaluation Method

when in process:	design vs. implementation
style of evaluation:	laboratory vs. field
how objective:	subjective vs. objective
type of measures:	qualitative vs. quantitative
level of information:	high level vs. low level
level of interference:	obtrusive vs. unobtrusive
resources available:	time, subjects, equipment, expertise

