Hypothesenprüfung

- Teil der Inferenzstatistik
- Befaßt sich mit der Frage, wie Hypothesen über eine (in der Regel unbekannte) Grundgesamtheit an einer Stichprobe überprüft werden können
- Behandelt werden drei Testverfahren:
  - Mittelwertunterschiede bei bekannter Streuung (z-Test)
  - Mittelwertunterschiede für unabhängige/abhängige Stichproben (t-Test)
  - χ^2-Test auf Unabhängigkeit
- Darüber hinaus existieren zahlreiche andere Testverfahren, die alle auf der gleichen Logik basieren

Alternativhypothese

- Zunächst wird eine (idealerweise) neue, inhaltlich interessante und theoretisch begründete Hypothese über die Verhältnisse in der GG aufgestellt.
- Diese Hypothese kann spezifisch oder unspezifisch, gerichtet oder ungerichtet sein
- Beispiele:
  - „Ostdeutsche verdienen im Durchschnitt weniger als Westdeutsche“
  - Es besteht ein Zusammenhang zwischen autoritären Einstellungen und der Wahl der Republikaner
- Die Alternativhypothese wird mit H_A bezeichnet

Nullhypothese

- Der H_A wird die Nullhypothese H_0 gegenübergestellt, die diesen Zusammenhang negiert:
  - „Ostdeutsche verdienen genauso viel oder mehr als Westdeutsche“
  - Zwischen autoritären Einstellungen und der Wahl der Republikaner besteht kein Zusammenhang
- Idealerweise steht H_0 für den etablierten Forschungsstand
- H_A und H_0 schließen sich wechselseitig aus und beschreiben den Möglichkeitsraum vollständig, d.h. in der GG gilt entweder H_A oder (exklusiv) H_0

Entscheidungsdilemma beim Hypothesentest

		In der *Grundgesamtheit* gilt	
		H_0	H_A
Entscheidung auf Grundlage der *Stichprobe*	H_0	richtig	β-Fehler
	H_A	α-Fehler	richtig

Vorsicht: $\beta \neq 1 - \alpha$!

Irrtumswahrscheinlichkeit

- Das Verfahren ist inhärent konservativ, d.h. es wird ausschließlich der α-Fehler (auch Wahrscheinlichkeit für einen Irrtum 1. Ordnung) betrachtet. Der β-Fehler (auch: Irrtum 2. Ordnung, Teststärke) bleibt zumeist unberücksichtigt
- Diese Irrtumswahrscheinlichkeit wird *vorab* auf 1%, 5%, bei kleinen Stichproben zuweilen auch 10% festgelegt
- α bezeichnet die Wahrscheinlichkeit, daß die H_0 zu unrecht aufgegeben wird, obwohl sie in der GG gilt

Signifikanztest

- Voraussetzung: In der GG gilt tatsächlich die H_0 (kein Zusammenhang)
- Berechnung einer Prüfgröße für die Stärke des Zusammenhangs
- Ermittelung dieser Prüfgröße in einer Zufallsstichprobe ist Realisierung einer Zufallsvariablen, die einer theoretischen Verteilung folgt
- Aus dieser Verteilung läßt sich ein kritischer Wert ablesen, der nur in 5% bzw. 1% aller Fälle überschritten wird, *wenn* die Nullhypothese gilt
- Überschreitet die Prüfgröße diesen Wert, gilt das Ergebnis als *signifikant*
- Dabei muß unterschieden werden, ob es sich um eine gerichtete oder eine ungerichtete H_A handelt

Statistische Signifikanz

- Statistische Signifikanz bedeutet: Das Ergebnis des Tests ist sehr unwahrscheinlich (5% bzw. 1%), wenn die H_0 in der GG gilt
- Deshalb wird die H_0 verworfen
  - Schluß von der Stichprobe auf die Grundgesamtheit.
  - Die Wahrscheinlichkeit, dabei einen Fehler zu begehen, beträgt 5% bzw. 1% (α)
- Von der statistischen Signifikanz ist *die praktische Bedeutsamkeit* der gefundenen Ergebnisse zu unterscheiden

z-Test

- Vergleicht einen Stichprobenmittelwert mit dem bekannten Mittelwert der GG (kommt relativ selten zur Anwendung)
- Frage: Kann der Stichprobenmittelwert tatsächlich aus der GG stammen
- Beispiel (Bortz 1993: 109):
  - In der GG liegt der Mittelwert μ eines schulischen Leistungstests bei 40 Punkten. Die Standardabweichung σ beträgt 8 Punkte
  - Bei einer Stichprobe mit dem Umfang n=100, die nach einem neuen Verfahren unterrichtet wird, stellt man einen Mittelwert von 42 Punkten fest.
  - H_A: Die neue Unterrichtsmethode ist besser als die alte
  - H_0: Die neue Methode ist gleich gut oder schlechter
  - Die Irrtumswahrscheinlichkeit wird auf 5% festgelegt
  - Einseitiger Test, da Hypothese gerichtet

Beispiel: z-Test

- Standardfehler für Stichproben mit dem Umfang n=100
$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{100}} = 0{,}8$$
- Eigentliche Frage: Wie wahrscheinlich ist es, daß in einer Normalverteilung mit einem Mittelwert von 40 und einer Standardabweichung von 0,8 ein Wert von 42 erreicht wird?
- Läßt sich (um die Berechnung zu vereinfachen) umformen: Wie wahrscheinlich ist es, daß in einer Standard-Normalverteilung (z-Verteilung) ein bestimmter z-Wert erreicht wird? Voraussetzung: z-Transformation $z = \frac{x - m_0}{s_{\bar{x}}} = \frac{42 - 40}{0{,}8} = 2{,}5$
- Ein Blick in die Tabelle zeigt, daß dieser Wert eine Fläche von lediglich 0,62% am rechten Rand der z-Verteilung abschneidet. Die Irrtumswahrscheinlichkeit ist also viel kleiner als 5%

t-Test
(für unabhängige und abhängige Stichproben)

- Untersucht die Abweichung der Mittelwerte zweier Stichproben voneinander
- Fragestellung: Wie wahrscheinlich ist die beobachtete Differenz zwischen den Mittelwerten, wenn sich die Mittelwerte in den *Grundgesamtheiten*, aus denen sie gezogen wurden, *nicht* unterscheiden (H_0)
- Beispiele:
  - unterscheidet sich das Durchschnittseinkommen von Ost- und Westdeutschen (unabhängige Stichproben) ?
  - Verbessern sich die Statistik-Kenntnisse durch die Teilnahme an einem Statistik-Kurs (abhängige Stichproben) ?

Statistisches Modell des t-Tests

- Ausprägung des Einkommens bei jedem Befragten ist eine Zufallsvariable
- Einkommensdurchschnitt in beiden Stichproben ist eine Zufallsvariable, die durch Addition von Zufallsvariablen zustandekommt, und folgt einer Normalverteilung
- Differenz zwischen den beiden Einkommensdurchschnitten ist eine weitere Zufallsvariable, die auf zwei normalverteilte Zufallsvariablen zurückgeht
- Diese Variable ist „noch zufälliger“ und folgt deshalb einer t-Verteilung
- Diese Herleitung dient der Veranschaulichung und ist mathematisch nicht ganz korrekt

Eigenschaften der t-Verteilung

- Familie von t-Verteilungen mit unterschiedlicher Streuung und unterschiedlichem Mittelwert
- Mittelwert: Differenz der Mittelwerte der Grundgesamtheiten
- Form: wie bei der Normalverteilung symmetrisch, unimodal und glockenförmig
- Zusätzlich wird die t-Verteilung noch von der Zahl der Freiheitsgrade (df) beeinflußt: Je weniger df, desto schmalgipfliger die Verteilung. Bei einer hohen Zahl von Freiheitsgraden (>1000) nähert sich die t-Verteilung der Normalverteilung an

Freiheitsgrade für den t-Test

- Unter den Freiheitsgraden versteht man allgemein die Anzahl der Parameter, die „frei“ variieren können
- Im Falle des arithmetischen Mittels, das auf der Summe *aller* n Meßwerte beruht, können n-1 Werte frei variieren
- Wenn man das arithmetische Mittel kennt, ist der Wert für den letzten Fall schon festgelegt
- Da beim t-Test zwei Stichproben mit zwei arithmetischen Mitteln involviert sind, beträgt die Zahl der Freiheitsgrade insgesamt $df=(n_1-1)+(n_2-1)=(n_1+n_2)-2$
- Bei großen Stichproben kann die Standardnormalverteilung (=z-Verteilung) verwendet werden.

Festlegung der Hypothesen

- H_0: $\mu_1 = \mu_2$ bzw. $\mu_1 - \mu_2 = 0$
- H_A: $\mu_1 \neq \mu_2$ bzw. $\mu_1 - \mu_2 \neq 0$
- Analog zum z-Test muß nun eine Prüfgröße berechnet und mit dem kritischen Wert aus der t-Verteilung verglichen werden, um zu entscheiden, ob H_A angenommen oder H_0 beibehalten werden soll

Berechnung der Prüfgröße t (unabhängige Stichproben)

- In die Berechnung von t geht ein:
  - die Differenz der beiden Stichprobenmittelwerte
  - die Differenz der Mittelwerte in der GG (unter Gültigkeit von H_0 ist diese gleich null!)
  - die Streuung (=Standardfehler) der Mittelwertdifferenz

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\hat{s}_{(\bar{x}_1 - \bar{x}_2)}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\hat{s}_{(\bar{x}_1 - \bar{x}_2)}}$$

- Der Standardfehler für die Mittelwertdifferenz muß in der Regel geschätzt werden und ist deshalb in der Formel mit einem „Dach“ versehen

Berechnung des Standardfehlers der Mittelwertdifferenz

- Die Formel für den Standardfehler der Mittelwertdifferenz ist mit der Formel für den Standardfehler eines Mittelwertes verwandt:

$$s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}$$

- Da die Varianzen beider Grundgesamtheiten in der Regel nicht bekannt sind, werden für die Berechnung des Standardfehlers die Varianzschätzungen aus den beiden Stichproben benutzt.

- Üblicherweise wird davon ausgegangen, daß die Varianzen gleich sind (Varianzhomogenität). Trifft diese Annahme nicht zu, ist eine Korrektur erforderlich

Signifikanz der Prüfgröße

- t erhält man durch Einsetzen des Standardfehlers in die Formel
- Vergleich des t-Wertes mit der tabellierten t- bzw. z-Verteilung (Freiheitsgrade und Richtung der Hypothese beachten)
- Die Annahme, daß die Varianzen in beiden Grundgesamtheiten gleich sind, kann mit einem weiteren Test überprüft werden (F-Test)
- Unterscheiden sich die Varianzen, wird diese zusätzliche Unsicherheit berücksichtigt, indem die Zahl der Freiheitsgrade nach unten korrigiert wird (breitere Verteilung)

T-Test für abhängige Stichproben (Wiederholungsmessung)

- Typisches Beispiel: Wiederholungsmessungen. Zweite Stichprobe hängt von erster ab (da Befragte identisch)
- Für jeden Teilnehmer wird die Differenz d der Meßwerte x_1 und x_2 berechnet: $d=x_2-x_1$
- Für diese Variable können ebenfalls Mittelwert und Standardabweichung ermittelt werden
- Festlegung der Irrtumswahrscheinlichkeit (α=1%) und der Hypothesen:
  - H_0: $\mu_d \leq 0$
  - H_A: $\mu_d>0$ mit: = μ_d Mittelwert der Differenz in der GG
- Bestimmung der Freiheitsgrade: df=n-1 mit n=Zahl der Meßwertepaare=Zahl der Probanden

Berechnung der Prüfgröße t (abhängige Stichproben)

- Der t-Wert hängt ab von der Differenz in der GG (nach der Nullhypothese=0) und dem Standardfehler der Differenz

$$t = \frac{\bar{x}_d - m_d}{s_{\bar{x}_d}} = \frac{\bar{x}_d}{s_{\bar{x}_d}}$$

- Standardfehler der Differenz hängt (vgl. z-Test und Berechnung der Konfidenzintervalle) ab von Streuung der Differenz und Zahl der Fälle.
- Streuung der Differenz muß wiederum aufgrund der Stichprobe geschätzt werden

$$\hat{s}_{\bar{x}_d} = \frac{\hat{s}_d}{\sqrt{n}} = \sqrt{\frac{SAQ_d/(n-1)}{n}}$$

χ^2-Test auf Unabhängigkeit

- Dient zur Überprüfung der Signifikanz von Zusammenhängen in Kreuztabellen
- Beruht auf der bereits bekannten Maßzahl χ^2
- Das bedeutet: Die Abweichungen zwischen einer empirischen Kreuztabelle von einer Tabelle, in der kein Zusammenhang zwischen den Variablen besteht, also beide voneinander unabhängig sind (Indifferenztabelle) , wird mit einem statistischen Modell (χ^2 -Verteilung) verglichen
- Auch hier stellt man die Frage, wie wahrscheinlich der empirische Befund ist, wenn in der GG H_0 gilt, und vergleicht die Antwort auf diese Frage mit einer vorab festgelegten Irrtumswahrscheinlichkeit

Eigenschaften der χ^2-Verteilung

- Quadriert man Zufallswerte aus einer z-Verteilung, erhält man eine χ^2-Verteilung
- Addiert man zwei,drei,m voneinander unabhängige quadrierte z-Werte, erhält man eine χ^2-Verteilung mit zwei, drei, m df.
- χ^2-Verteilungen sind unimodal, aber asymmetrisch (aller Werte positiv wegen Quadrierung)
- Form und Lage hängen ausschließlich von der Zahl der Freiheitsgrade ab
- Mit steigender Zahl der df geht die χ^2-Verteilung in eine Normalverteilung mit dem Mittelwert df und der Varianz 2*df über

Zusammenhang zwischen χ^2-Wert und χ^2-Verteilung

- Diese Herleitung dient vor allem der Veranschaulichung und ist mathematisch nicht korrekt
- Die Berechnung von χ^2 in einer Kreuztabelle ähnelt der Ziehung von quadrierten z-Werten
  - Subtraktion des erwarteten Wertes und Relativierung am erwarteten Wert erinnert an z-Transformation
  - additives Zusammenwirken der Werte die für jede Zelle berechnet werden ➔ Normalverteilung
  - Quadrieren der Werte

Freiheitsgrade in einer Kreuztabelle

- Gegeben sei die Indifferenztabelle
- In einer empirischen 2x2-Kreuztabelle, die aufgrund einer Stichprobe berechnet wird, kann genau eine Zelle „frei“ (=zufällig) variieren
- Die Besetzung der anderen Zellen liegt fest, weil die Randsummen und die Gesamtsumme gleich bleiben
- Allgemein beträgt die Zahl der Freiheitsgrade in einer k*m-Tabelle (k-1)*(m-1)
- Nach der Zahl der Freiheitsgrade bestimmt sich die χ^2-Verteilung, mit der der empirische Wert verglichen wird

Bestimmung der Signifikanz

- Voraussetzung: In weniger als 20% der Zellen treten erwartete Häufigkeiten < 5 auf
- Festlegung der Hypothesen:
  - H_0 $\chi^2=0$
  - H_A $\chi^2 > 0$
- Wahl einer Irrtumswahrscheinlichkeit
- Berechnung von χ^2
- Vergleich mit dem kritischen Wert aus der Tabelle (bei gegebenem α und df)
- Entscheidung über H_0 und H_A