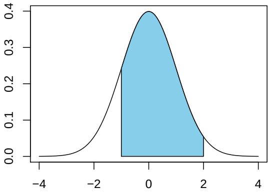
Hypothesentests II

Statistik I

Sommersemester 2009


0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Informationen Modulabschlußklausur

- Betrifft BA (nicht Lehramt)
- Modulabschlußklausur setzt sich aus drei Teilen zusammen („Einführung“, Statistik I, Methoden I)
  - Methodologie-Teil entfällt
  - Modulteilprüfung Methoden I entfällt
- Für BAs im ersten Semester: Modulteilprüfung Methoden I am 16. Juli entfällt, Inhalte werden in Modulabschlußklausur abgeprüft
- BAs im zweiten Semester, die Modulteilprüfung noch nicht geschrieben haben oder durchgefallen sind $\rightarrow$ Modulabschlußklausur
- BAs im zweiten Semester, die Modulteilprüfung erfolgreich abgelegt haben, haben Wahlmöglichkeit

Welche Wahlmöglichkeit besteht, wenn Methoden I bereits bestanden ist?

1. Reduzierte Klausur
   - Nur „Einführung“ + Statistik I, Dauer 70 Minuten
   - Bestehende Methoden I Note geht mit $\frac{3}{14}$ in Modulnote ein
2. Normale Klausur
   - 90 Minuten, alle drei Teile
   - Alte Methoden I Note wird **gelöscht**

- Keine Ex-post-Wahlmöglichkeit zur Notenverbesserung
- Verbindliche Entscheidung am Tag der Prüfung durch Auswahl am Computer

Anmeldung

- Anmeldung zur Prüfungsleistung:
  - Seit Montag, *06.07.2009*, Möglichkeit sich über Jogustine zur Modulabschlussprüfung in der neuen Form anzumelden
  - Bestehende Anmeldungen werden übertragen
  - Anmeldefrist läuft bis einschließlich *27.07.2009*
  - Abmeldungen über Jogustine bis dahin möglich
- Freischaltung für die Klausur
  - Freischaltung in ILIAS zwingend notwendig für Teilnahme an Klausur
  - Kategorie: POWI Einführungsmodul $\longrightarrow$ Kurs Einführungsmodul
  - Bis spätestens *27.07.2009*
  - Paßwort: `Schreibtisch`

Zum Nachlesen

- Agresti/Finlay: Kapitel 6+7
- Gehring/Weins: Kapitel 12
- (Schumann: Kapitel 7.3)

Wie sieht die Logik des Hypothesentests aus?

- Aufteilung Möglichkeitsraums in Nullhypothese vs. Alternativhypothese (gerichtet/ungerichtet)
- Berechnung einer Teststatistik mit bekannter Verteilung
- Vergleich von Teststatistik mit theoretischer Verteilung $\rightarrow$ „signifikantes" Ergebnis?

Worum geht es bei statistischer Signifikanz?

- Wie wahrscheinlich ist das Stichprobenergebnis, wenn in GG Nullhypothese gilt?
- Bzw. ist diese Wahrscheinlichkeit geringer als vorab festgelegte Irrtumswahrscheinlichkeit?
- Irrtumswahrscheinlichkeit = Wahrscheinlichkeit, Nullhypothese zu Unrecht aufzugeben
- Statistische Signifikanz $\neq$ inhaltliche Bedeutsamkeit

Worum geht es beim *t*-Test?

- Test für Mittelwertunterschiede zwischen Stichproben
- Fragestellung:
  - Wie wahrscheinlich ist die beobachtete Differenz zwischen den Mittelwerten
  - Wenn sich die Mittelwerte in den Grundgesamtheiten unterscheiden?
- Unterscheidung zwischen
  1. Unabhängigen Stichproben: zwei separate Gruppen (z. B. Ost- vs. Westdeutsche)
  2. Abhängige Stichproben: dieselbe Gruppe wird zweimal untersucht (Experimente, Teilnahme an Statistik-Kurs)

t-Test und η^2

- *t*-Test und Berechnung η^2 (für zwei Gruppen) verwandt
- η^2
  - Summe SAQ um Gruppenmittelwerte kleiner als
  - SAQ_{gesamt} um Gesamtmittelwert?
  - Stärke des Zusammenhangs
- Wert hängt ab von
  - Streuung innerhalb der Gruppen
  - Unterschied zwischen Mittelwerten der Gruppen ($\rightarrow$ SAQ_{gesamt})
- *t*-Test: Wie sicher sind wir uns über diesen Unterschied?

t-Test für unabhängige Stichproben: Beispiel

Durchschnittseinkommen Männer/Frauen

In sogenannten Mischberufen (40-60% Frauenanteil) lag 1997 das Durchschnittseinkommen von Männern ein Jahr nach Abschluß ihrer Ausbildung bei 3 548 DM bei einer Standardabweichung von 357 DM ($n = 912$). Frauen verdienten im Durchschnitt 3 496 DM bei einer Standardabweichung von 354 DM ($n = 851$). Der Gesamtdurchschnitt beträgt 3523 DM mit einer Standardabweichung von 362 DM.

t-Test für unabhängige Stichproben: Beispiel

Durchschnittseinkommen Männer/Frauen

In sogenannten Mischberufen (40-60% Frauenanteil) lag 1997 das Durchschnittseinkommen von Männern ein Jahr nach Abschluß ihrer Ausbildung bei 3 548 DM bei einer Standardabweichung von 357 DM ($n = 912$). Frauen verdienten im Durchschnitt 3 496 DM bei einer Standardabweichung von 354 DM ($n = 851$). Der Gesamtdurchschnitt beträgt 3523 DM mit einer Standardabweichung von 362 DM.

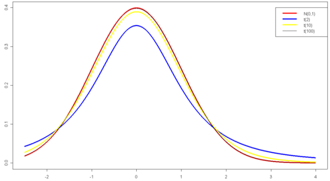
- Kann man aus dem Mittelwertunterschied zwischen beiden Stichproben schließen
- Daß weibliche Berufsanfänger in diesen Berufen in der GG im Mittel schlechter verdienen?

Wie sieht das Modell des t-Tests aus?

- Einkommen ist eine Zufallsvariable (Befragte zufällig auswählen, Einkommen erfragen)
- Mittelwert in einer Stichprobe (Frauen) weitere Zufallsvariable
  - Additiv aus einfachen Zufallsvariablen zusammengesetzt
  - Deshalb (näherungsweise) normalverteilt (große Stichproben)
- Differenz zweier Mittelwerte ist ebenfalls eine Zufallsvariable und t-verteilt

Wie sieht die t-Verteilung aus?

- Ähnlich wie Normalverteilung
- Symmetrisch, unimodal, glockenförmig
- Familie von t-Verteilungen
  1. Mittelwert
  2. Standardabweichung
  3. *Freiheitsgrade* (d.f.)
- Je weniger d. f., desto schmalgipfliger $\rightarrow$ mehr extreme Werte
- Für große Zahl von Freiheitsgraden (> 1000) Übergang in Normalverteilung



Was sind Freiheitsgrade?

- Freiheitsgrade = Anzahl der unabhängigen Informationen die in Berechnung einer Größe einfließen
- Bei Mittelwert d. f. $= n$
- Bei Varianz d. f. $= n - 1$
  - Basiert auf quadratierten Abweichungen vom Mittelwert
  - Einfache Abweichungen müssen sich zu null summieren
  - $n - 1$ zufällige Einflüsse
  - Letzter Einfluß trägt keine unabhängige/neue Information mehr bei da Mittelwert schon bekannt
  - Bzw. ein Freiheitsgrad schon für Mittelwertberechnung „verbraucht“
- Bei *z*-Test Varianz σ in GG bekannt
- Bei Konfidenzintervallen Schätzung $\hat{\sigma} \rightarrow t$-Verteilung mit $n - 1$ Freiheitsgraden verwenden

Wie lauten die Hypothesen?

- *Zwei* wahre Mittelwerte ($\mu_F; \mu_M$) aus zwei GG
- Alternativhypothese
  - $H_A: \quad \mu_M \neq \mu_F \Leftrightarrow \mu_M - \mu_F \neq 0$ (ungerichtet)
  - $H_A: \quad \mu_M > \mu_F \Leftrightarrow \mu_M - \mu_F > 0$ (gerichtet)
- Nullhypothese
  - $H_0: \quad \mu_M = \mu_F \Leftrightarrow \mu_M - \mu_F = 0$ (ungerichtet)
  - $H_0: \quad \mu_M \leqslant \mu_F \Leftrightarrow \mu_M - \mu_F \leqslant 0$ (gerichtet)
- *Wenn* Nullhypothese gilt, ist Prüfgröße (Quotient aus Mittelwertdifferenz und dessen Standardfehler) t-verteilt
- „Kritische Werte", die bei Gültigkeit H_0 nur sehr selten (1/5/10%) überschritten werden

Wie wird Prüfgröße t berechnet?

- t berücksichtigt:
  - Wahre Mittelwertdifferenz $\mu_M - \mu_F$ (bei Gültigkeit von $H_0 = 0$
  - Mittelwertdifferenz zwischen Stichproben: $\bar{x}_M - \bar{x}_F$
  - Standardfehler der Mittelwertdifferenz
- Standardfehler muß aus den Daten geschätzt werden

$$t = \frac{(\bar{x}_M - \bar{x}_F) - (\mu_M - \mu_F)}{\hat{\sigma}_{(\bar{x}_M - \bar{x}_F)}} = \frac{(\bar{x}_M - \bar{x}_F)}{\hat{\sigma}_{(\bar{x}_M - \bar{x}_F)}}$$

- t: Wieviele Standardfehler ist die beobachtete Mittelwertdifferenz von der wahren Mittelwertdifferenz entfernt $\rightarrow$ Vergleich mit bekannter Verteilung

Wie wird Standardfehler für Mittelwertdifferenz berechnet?

- Verwandt mit Berechnung für Standardfehler *eines* Mittelwertes
- Hängt ab von
  - Umfang *beider* Stichproben
  - Varianz/Standardabweichung des Merkmals (Einkommen) in *beiden* Grundgesamtheiten
- Varianzen in der Regel nicht bekannt → Schätzung aus Stichproben:

$$\hat{\sigma}_{(\bar{x}_M - \bar{x}_F)} = \sqrt{\frac{\hat{\sigma}^2_M}{n_M} + \frac{\hat{\sigma}^2_F}{n_F}}$$

- Normalerweise Annahme daß Varianzen in beiden GG identisch (Varianzhomogenität), ansonsten Korrektur notwendig

Einkommensdifferenzen signifikant?

- Signifikanzniveau/Richtung festlegen: einseitiger Test, $\alpha = 0.01$
- $\bar{x}_M - \bar{x}_F = 3548 - 3496 = 52$ – Richtung stimmt
- Standardfehler berechnen: $\hat{\sigma}_{(\bar{x}_M - \bar{x}_F)} = \sqrt{\frac{357^2}{912} + \frac{354^2}{851}} = 16.94$
- Freiheitsgerade bestimmen: d.f. $=$ $n_M - 1 + n_F - 1 = 911 + 850 = 1761$
- $> 1000 \rightarrow$ Normalverteilung gute Approximation
- Welcher Wert trennt links 99% und rechts 1% der Normalverteilung ab?

Einkommensdifferenzen signifikant?

- Signifikanzniveau/Richtung festlegen: einseitiger Test, $\alpha = 0.01$
- $\bar{x}_M - \bar{x}_F = 3548 - 3496 = 52$ – Richtung stimmt
- Standardfehler berechnen: $\hat{\sigma}_{(\bar{x}_M - \bar{x}_F)} = \sqrt{\frac{357^2}{912} + \frac{354^2}{851}} = 16.94$
- Freiheitsgerade bestimmen: d.f. $=$ $n_M - 1 + n_F - 1 = 911 + 850 = 1761$
- $> 1000 \rightarrow$ Normalverteilung gute Approximation
- Welcher Wert trennt links 99% und rechts 1% der Normalverteilung ab? 2.33

Einkommensdifferenzen signifikant?

- Signifikanzniveau/Richtung festlegen: einseitiger Test, $\alpha = 0.01$
- $\bar{x}_M - \bar{x}_F = 3548 - 3496 = 52$ – Richtung stimmt
- Standardfehler berechnen: $\hat{\sigma}_{(\bar{x}_M - \bar{x}_F)} = \sqrt{\frac{357^2}{912} + \frac{354^2}{851}} = 16.94$
- Freiheitsgerade bestimmen: d.f. $=$ $n_M - 1 + n_F - 1 = 911 + 850 = 1761$
- $> 1000 \rightarrow$ Normalverteilung gute Approximation
- Welcher Wert trennt links 99% und rechts 1% der Normalverteilung ab? 2.33
- $t = \frac{52}{23.9} \approx 2.18$

Ist die Differenz relevant?

- Stärke des Zusammenhangs $\rightarrow$ η bzw. η^2
- Wie läßt sich das hier berechnen?

t-Test für unabhängige Stichproben: Beispiel

Durchschnittseinkommen Männer/Frauen

In sogenannten Mischberufen (40-60% Frauenanteil) lag 1997 das Durchschnittseinkommen von Männern ein Jahr nach Abschluß ihrer Ausbildung bei 3 548 DM bei einer Standardabweichung von 357 DM ($n = 912$). Frauen verdienten im Durchschnitt 3 496 DM bei einer Standardabweichung von 354 DM ($n = 851$). Der Gesamtdurchschnitt beträgt 3523 DM mit einer Standardabweichung von 362 DM.

Ist die Differenz relevant?

- Stärke des Zusammenhangs $\rightarrow$ η bzw. η^2
- Wie läßt sich das hier berechnen?
- (geschätzte) Standardabweichung $\rightarrow$ (geschätzte) Varianz $\rightarrow$ SAQ
  - $SAQ_{gesamt} = 362^2 \times (n_{gesamt} - 1) = 131044 \times 1762 = 171\,520\,128$
  - $SAQ_F = 354^2 \times (n_F - 1) = 125316 \times 850 = 106\,518\,600$
  - $SAQ_M = 357^2 \times (n_M - 1) = 127449 \times 911 = 116\,106\,039$

Ist die Differenz relevant?

- Stärke des Zusammenhangs $\rightarrow \eta$ bzw. η^2
- Wie läßt sich das hier berechnen?
- (geschätzte) Standardabweichung $\rightarrow$ (geschätzte) Varianz $\rightarrow$ SAQ
  - $SAQ_{gesamt} = 362^2 \times (n_{gesamt} - 1) = 131044 \times 1762 = 171\,520\,128$
  - $SAQ_F = 354^2 \times (n_F - 1) = 125316 \times 850 = 106\,518\,600$
  - $SAQ_M = 357^2 \times (n_M - 1) = 127449 \times 911 = 116\,106\,039$
- $\eta^2 = \frac{SAQ_{gesamt} - SAQ_F - SAQ_M}{SAQ_{gesamt}} = 0.04$
- Schwacher Zusammenhang zwischen Einkommen/Geschlecht

Ist die Differenz relevant?

- Stärke des Zusammenhangs $\rightarrow$ η bzw. η^2
- Wie läßt sich das hier berechnen?
- (geschätzte) Standardabweichung $\rightarrow$ (geschätzte) Varianz $\rightarrow$ SAQ
  - $SAQ_{gesamt} = 362^2 \times (n_{gesamt} - 1) = 131044 \times 1762 = 171\,520\,128$
  - $SAQ_F = 354^2 \times (n_F - 1) = 125316 \times 850 = 106\,518\,600$
  - $SAQ_M = 357^2 \times (n_M - 1) = 127449 \times 911 = 116\,106\,039$
- $\eta^2 = \frac{SAQ_{gesamt} - SAQ_F - SAQ_M}{SAQ_{gesamt}} = 0.04$
- Schwacher Zusammenhang zwischen Einkommen/Geschlecht
- (Mit dem Standardfehler der Mittelwertdifferenz läßt sich auch 99% Konfidenzintervall berechnen ($52 \pm 2.33 \times 23.9$))

Was bedeutet „abhängige Stichproben“?

- Einkommensunterschiede beziehen sich auf unterschiedliche Personen (Männer vs. Frauen)
- Messungen sind unabhängig voneinander
- Vs. Wiederholungsmessung: Dieselben Personen werden zweimal untersucht
- Messungen „abhängig“, individuelle Meßfehler werden u. U. repliziert
- Häufig bei Panel- und Experimentaldesigns (Vorher-Nachher-Messung)

t-Test für abhängige Stichproben: Beispiel

Statistik-Noten

Eine Gruppe von 49 Studierenden schreibt zu Beginn des Semesters eine Statistik-Vorklausur. Im Mittel werden 33 von 100 möglichen Punkten erreicht. Am Ende des Semesters erzielen die Studierenden in einer vergleichbaren Klausur im Mittel 45 Punkte, haben sich also im Durchschnitt um 12 Punkte verbessert ($d_{\bar{x}_2-\bar{x}_1} = 12$). Die Standardabweichung $\hat{\sigma}_{d_{\bar{x}_2-\bar{x}_1}}$ der Verbesserung beträgt 8 Punkte (hier schon als Schätzung für GG).

t-Test für abhängige Stichproben: Beispiel

Statistik-Noten

Eine Gruppe von 49 Studierenden schreibt zu Beginn des Semesters eine Statistik-Vorklausur. Im Mittel werden 33 von 100 möglichen Punkten erreicht. Am Ende des Semesters erzielen die Studierenden in einer vergleichbaren Klausur im Mittel 45 Punkte, haben sich also im Durchschnitt um 12 Punkte verbessert ($d_{\bar{x}_2 - \bar{x}_1} = 12$). Die Standardabweichung $\hat{\sigma}_{d_{\bar{x}_2 - \bar{x}_1}}$ der Verbesserung beträgt 8 Punkte (hier schon als Schätzung für GG).

- Ist die Verbesserung signifikant bzw. lohnt sich der Aufwand?

Hypothesen und Freiheitsgrade

- H_A: $\mu > 0$, d. h. positiver Lerneffekt in der GG
- H_0: $\mu \leqslant 0$, d. h. kein Lerneffekt in GG oder sogar Verschlechterung
- Freiheitsgrade: d.f. $= n - 1$, d. h. Zahl der Meßwertpaare (=Probanden) - 1
- Irrtumswahrscheinlichkeit wird auf $\alpha = 0.01$ festgesetzt

Standardfehler Differenz/Prüfgröße

- t-Wert hängt ab von Lerneffekt in Stichprobe, Lerneffekt in GG (nach Nullhypothese =0) und Standardfehler der Differenz
- $t = \frac{\bar{x}_d - \mu_d}{\sigma_{\bar{x}_d}} = \frac{\bar{x}_d}{\sigma_{\bar{x}_d}}$
- Standardfehler hängt ab von Streuung der Differenz und Fallzahl
- $\sigma_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}} = \sqrt{\frac{SAQ_d/(n-1)}{n}}$

Signifikanter Lerneffekt?

- Standardfehler berechnen: $\sigma_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}} \frac{8}{7} = 1.14$
- t-Wert bestimmen: $t = \frac{\bar{x}_d}{\sigma_{\bar{x}_d}} = \frac{12}{1.14} = 10.5$
- Freiheitsgrade = n-1 = 48 $\longrightarrow$ kritischen $t - Wert$ für $\alpha = 0.01$, einseitiger Test, d.f.=48 finden

t-Tabelle/Computer

	Fläche (1-α)									
df	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656
2	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

t-Tabelle/Computer

	Fläche (1-α)									
df	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656
2	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

t-Tabelle/Computer

	Fläche (1-α)									
df	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656
2	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
29	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

t-Tabelle/Computer

	Fläche (1-α)									
df	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656
2	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

t-Tabelle/Computer

	Fläche (1-α)									
df	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656
2	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

```
> qt(.99,df=48)
[1] 2.406581
```

Signifikanter Lerneffekt?

- Standardfehler berechnen: $\sigma_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}} \frac{8}{7} = 1.14$
- t-Wert bestimmen: $t = \frac{\bar{x}_d}{\sigma_{\bar{x}_d}} = \frac{12}{1.14} = 10.5$
- Freiheitsgrade = n-1 = 48 $\longrightarrow$ kritischen $t - Wert$ für $\alpha = 0.01$, einseitiger Test, d.f.=48 finden
- Empirischer t-Wert überschreitet kritischen Wert bei weitem

Signifikanter Lerneffekt?

- Standardfehler berechnen: $\sigma_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}} \frac{8}{7} = 1.14$
- t-Wert bestimmen: $t = \frac{\bar{x}_d}{\sigma_{\bar{x}_d}} = \frac{12}{1.14} = 10.5$
- Freiheitsgrade = n-1 = 48 $\longrightarrow$ kritischen $t - Wert$ für $\alpha = 0.01$, einseitiger Test, d.f.=48 finden
- Empirischer t-Wert überschreitet kritischen Wert bei weitem
- Irrtumswahrscheinlichkeit sehr viel kleiner als 1% (einseitig)
- Exakte Irrtumswahrscheinlichkeit: 2.509×10^{-14}

Signifikanter Lerneffekt?

- Standardfehler berechnen: $\sigma_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}} \frac{8}{7} = 1.14$
- t-Wert bestimmen: $t = \frac{\bar{x}_d}{\sigma_{\bar{x}_d}} = \frac{12}{1.14} = 10.5$
- Freiheitsgrade = n-1 = 48 $\longrightarrow$ kritischen $t - Wert$ für $\alpha = 0.01$, einseitiger Test, d.f.=48 finden
- Empirischer t-Wert überschreitet kritischen Wert bei weitem
- Irrtumswahrscheinlichkeit sehr viel kleiner als 1% (einseitig)
- Exakte Irrtumswahrscheinlichkeit: 2.509×10^{-14}
- Statistisch signifikante und inhaltliche relevante (+36%) Verbesserung

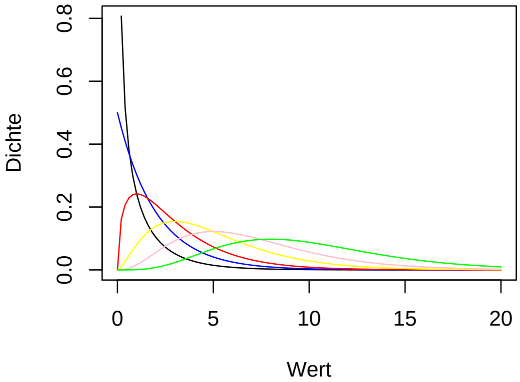
Was ist der χ^2-Test? auf Unabhängigkeit?

- Basiert auf Maßzahl χ^2
- Überprüft statistische Signifikanz in Kreuztabellen (nominale Variablen)
- Abweichungen zwischen
  - Empirischer Kreuztabelle
  - Tabelle, wenn beide Variablen statistisch unabhängig sind (Indifferenztabelle)
- Vergleich mit Verteilungsmodell χ^2-Verteilung
- Wie wahrscheinlich ist beobachtete Abweichung von Unabhängigkeit, wenn in GG Nullhypothese (kein Zusammenhang, Indifferenz) gilt?
- Vergleich mit vorab festgelegter Irrtumswahrscheinlichkeit α

Welche Eigenschaften hat die χ^2 -Verteilung?

- Ziehung von Zufallswerten aus z-Verteilung $\rightarrow$ quadrieren $\rightarrow$ χ^2 -Verteilung
- Summe aus $1, 2, \cdots, m$ unabhängigen quadrierten z-Werten $\rightarrow$ χ^2 -Verteilung mit m Freiheitsgraden
- χ^2 -Verteilungen
  - Unimodal, aber asymmetrisch (nur positive Werte wg. Quadrierung)
  - Form und Lage hängen nur von Zahl der Freiheitsgrade ab
  - Mit steigenden d. f. Übergang zu Normalverteilung mit $\mu =$ d. f. und $\sigma^2 = 2\times$ d. f.

χ^2-Verteilungen mit verschiedenen Freiheitsgraden



d. f.=1 (schwarz), 2 (blau), 3 (rot), 10 (grün)

Wieviele Freiheitsgrade?

- Wieviele unabhängige zufällige Einflüsse wirken in Kreuztabelle zusammen?
- (Abweichung von Indifferenztabelle)
- In 2×2-Tabelle kann eine Zelle „frei“ variieren
  - Rand- und Gesamtsummen liegen fest
  - Sobald eine Zelle zufällig variiert hat, ergibt sich Inhalt anderer Zellen aus diesem Wert plus Randsummen
- Allgemein: d. f. für $k \times m$-Tabelle: $(k-1) \times (m-1)$
- Aus der Zahl der Freiheitsgrade ergibt sich theoretische χ^2-Verteilung, mit der empirischer Wert verglichen wird

Wie werden Hypothesen getestet?

- Voraussetzung: in weniger als 20% der Zellen erwartete Häufigkeit < 5
- Hypothesen festlegen (Richtung spielt keine Rolle – warum?)
  - $H_0 : \chi^2 = 0$
  - $H_A : \chi^2 > 0$
- Wahl α
- χ^2 und Freiheitsgrade bestimmen
- Vergleich mit kritischem Wert aus theoretischer Verteilung; Entscheidung über Hypothesen

Ein Beispiel: Konfession und Wahlabsicht

beobachtete Werte

	Protestantisch	Katholisch	Keine	Σ
SPD	40	20	20	80
CDU	30	60	10	100
Andere	20	10	30	60
Σ	90	90	60	240

$\alpha = 0.05$

Ein Beispiel: Konfession und Wahlabsicht

erwartete Werte

	Protestantisch	Katholisch	Keine	Σ
SPD	30	30	20	80
CDU	37.5	37.5	25	100
Andere	22.5	22.5	15	60
Σ	90	90	60	240

Ein Beispiel: Konfession und Wahlabsicht

erwartete Werte

	Protestantisch	Katholisch	Keine	Σ
SPD	30	30	20	80
CDU	37.5	37.5	25	100
Andere	22.5	22.5	15	60
Σ	90	90	60	240

d. f.$= (k-1) \times (m-1) = 2 \times 2 = 4$

Ein Beispiel: Konfession und Wahlabsicht

erwartete Werte

	Protestantisch	Katholisch	Keine	Σ
SPD	30	30	20	80
CDU	37.5	37.5	25	100
Andere	22.5	22.5	15	60
Σ	90	90	60	240

d. f.$= (k-1) \times (m-1) = 2 \times 2 = 4$

Ein Beispiel: Konfession und Wahlabsicht

erwartete Werte

	Protestantisch	Katholisch	Keine	Σ
SPD	30	30	20	80
CDU	37.5	37.5	25	100
Andere	22.5	22.5	15	60
Σ	90	90	60	240

d. f.$= (k-1) \times (m-1) = 2 \times 2 = 4$

Ein Beispiel: Konfession und Wahlabsicht

(beobachtete Werte - erwartete Werte)

	Protestantisch	Katholisch	Keine
SPD	100	100	0
CDU	56.25	506.25	225
Andere	6.25	156.25	225

Ein Beispiel: Konfession und Wahlabsicht

(beobachtete Werte - erwartete Werte)/erwartete Werte

	Protestantisch	Katholisch	Keine
SPD	3.33	3.33	0
CDU	1.5	13.5	9
Andere	0.28	6.94	15

Ein Beispiel: Konfession und Wahlabsicht

(beobachtete Werte - erwartete Werte)/erwartete Werte

	Protestantisch	Katholisch	Keine
SPD	3.33	3.33	0
CDU	1.5	13.5	9
Andere	0.28	6.94	15

$\chi^2 = 52.89$

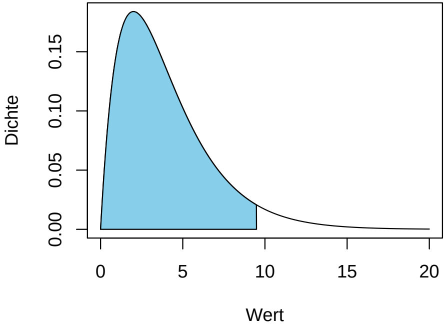
Ein Beispiel: Konfession und Wahlabsicht

(beobachtete Werte - erwartete Werte)/erwartete Werte

	Protestantisch	Katholisch	Keine
SPD	3.33	3.33	0
CDU	1.5	13.5	9
Andere	0.28	6.94	15

$\chi^2 = 52.89 \qquad > \chi^2_{\alpha, df=4}$?

Kritischer Wert



$\chi^2_{\alpha,df=4} = 9.49 \rightarrow$ Zusammenhang statistisch signifikant; $V = \sqrt{\frac{52.89}{240 \times 2}} = 0.33$

Lektüre für nächste Woche (Anwendungen)

- Berühren sich die Extreme? Ein empirischer Vergleich von Personen mit extrem linken und extrem rechten Einstellungen in Europa. In: Uwe Backes und Eckhard Jesse (Hrsg.): Gefährdungen der Freiheit. Extremistische Ideologien im Vergleich. Göttingen, Vandehoeck und Ruprecht 2006: 253-281 `http://www.kai-arzheimer.com/extremismus-im-europaeischen-vergleich/Links-Rechts-Vergleich.pdf`
- Politikverdrossenheit - eine Frage der Persönlichkeit? Der Zusammenhang zwischen Persönlichkeitsfaktoren und Verdrossenheitseinstellungen. (PDF) In: Siegfried Schumann und Harald Schoen (Hrsg.): Persönlichkeit. Eine vergessene Größe der empirischen Sozialforschung. Wiesbaden: VS 2005, S. 193-207 `http://www.kai-arzheimer.com/pollitikverdrossenheit-und-persoenlichkeit/politikverdrossenheit-und-persoenlichkeit.pdf`

Zusammenfassung

- t-Test und χ^2-Test funktionieren im Prinzip wie z-Test
- Logik von Alternativhypothese/Nullhypothese
- Freiheitsgrade: Zahl der unabhängigen Informationen über zufällige Einflüsse $\rightarrow$ adäquate Verteilung
- Modell für Vielzahl weiterer Tests (in Abhängigkeit von Problemstellung)