Lehrstuhl fuer Statistik, Oekonometrie und empirische Wirtschaftsforschung, Universitaet Tuebingen

Kreuztabellierung und Kontingenzanalyse

Dr. S. Prohl

11. Mai 2007/ SoSe 2007

Modelle fuer Kreuztabellen

- Test von Homogenitaet:
  - Der Chiquadrat Test.
  - der Fischer-Test.
- Test von Homogenitaet in einer Kontingenztabelle, bei der die Auspraegungen eines Merkmals geordnet sind.
- Test von Homogenitaet in einer Kontingenztabelle, bei der die Auspraegungen beider Merkmale geordnet sind.

Test von Homogenitaet

- Ziel: uebepruefe, ob sich die Verteilung eines Merkmals B mit k-Auspraegungen in c Populationen unterscheidet.
  - Merkmal der i-ten Population ist A mit Auspraegungen A_i.
- Wahrscheinlichkeit des Aufweisens der Auspraegung B_j: $p_{.j} = \sum_{i=1}^{c} p_{ij}$.
- Wahrscheinlichkeit des Aufweisens des Merkmals B in der i-ten Population mit Auspraegung B_j: $p_{j|i} = \frac{p_{ij}}{p_{i.}}$.

Test von Homogenitaet

- Teste die Null-Hypothese H_0:
  $H_0 : p_{j|i} = p_{.j}, i = 1, ..., c.$
- gegen der Alternative-Hypothese H_1:
  $H_1 : p_{j|i} \neq p_{.j}.$

Um die Hypothese zu ueberpruefen, ziehe eine Stichprobe.

Allgemeiner Aufbau einer Kontingenztabelle

Tabelle:

B Population	B_1	B_2	B_k	
1	n_{11}	n_{12}	n_{1k}	$n_{1.}$
2	n_{21}	n_{22}	n_{2k}	$n_{2.}$
.	...	...	...	...
.	...	...	...	...
.	...	...	...	...
c	n_{c1}	n_{c2}	n_{ck}	$n_{c.}$
	$n_{.1}$	$n_{.2}$	$n_{.k}$	n

- wobei: $n_i. = \sum_{j=1}^{k} n_{ij}$ und $n.j = \sum_{i=1}^{c} n_{ij}$.

Welche Tests gibt es zur Ueberpruefung der Nullhypothese?
$H_0 : p_{j|i} = p_{.j}, i = 1, ..., c.$

Chiquadrat-Test von Homogenitaet

- Teststatistik des Chiquadrat-Tests vergleicht die beobachteten mit den erwarteten Haeufigkeiten: $X^2 = \sum_{i=1}^{c} \sum_{j=1}^{k} \frac{n_{ij} - \hat{n}_{ij}}{\hat{n}_{ij}}$.
- Unter H_0-Hypothese (Homogenitaet) ist diese Teststatistik approximativ chiquadrat-verteilt, mit (c-1)(k-1) FG.
- Falls $X^2 \geq \chi^2_{(c-1)(k-1);1-\alpha} \Rightarrow H_0$ wird abgelehnt.
- Dabei ist $\chi^2_{(c-1)(k-1);1-\alpha}$-Quantil χ-verteilt mit (c-1)(k-1) FG.
- Falls H_0 verworfen ist $\Rightarrow$ Paretodiagramm und Mosaikplot.
- Beispiel.

Pruefung der Staerke des Zusammenhangs

- Phi-Koeffizient:
  $\phi = \sqrt{\frac{\chi^2}{n}}$.
  - Beachte: nicht vergleichbar aus verschiedenen Untersuchungen.
- Kontingenzkoeffizient:
  $CC = \sqrt{\frac{\chi^2}{\chi^2+n}}$.
  - Beachte: nimmt Wert in [0,1] an, wobei $CC_{max} = \sqrt{(R-1)/R}$ mit $R = min(I,J)$.
- Cramer's V-Koeffizient:
  $Cramer'sV = \sqrt{\frac{\chi^2}{n(R-1)}}$.
  - Beachte: Test-Statistik nimmt Wert in [0,1] an.
  - Im Falle der binaeren Variablen $\Rightarrow \phi$ = $Cramer'sV$.
- Beispiel.

Der Fischer Test

- Das Testproblem fuer ein binaeres Merkmal in zwei Gruppen $\Rightarrow$ Vergleich zweier Wahrscheinlichkeiten.
  $H_0 : p_{1|1} = p_{1|2}$ gegen $H_1 : p_{1|1} \neq p_{1|2}$
  - Die Fischer-Teststatistik N_{11} hat einen einzigen Freiheitsgrad!
- Die Verteilung ist durch die Wahrscheinlichkeiten von N_{11} unter H_0 gegeben:
  $P(N_{11} = n_{11}) =$
  - Die Teststatistik ist hypergeometrisch verteilt.
  - Fuer die H_1-Hypothese sprechen zu grosse oder zu kleine Werte von N_{11}.
- Beispiel.

Test von Homogenitaet in einer Kontingenztabelle, bei der Auspraegungen eines Merkmals geordnet sind

- Wilcoxon-Test:
  $$W = \frac{R_1 - \frac{n_1(n+1)}{2}}{\sqrt{\frac{n_1 n_2}{12}\left[n+1-\frac{1}{n(n-1)}\sum_{j=1}^{J}(n_{.j}^3 - n_{.j})\right]}}.$$
  - wobei: $n = n_{1.} + n_{2.}$, und R_1 ist die Rangsumme, die ueber das Merkmal A definiert ist.
- H_0 : identische Verteilungen
- Falls $\mid W \mid > z_{1-\alpha/2} \Rightarrow H_0$ wird verworfen.
  - $z_{1-\alpha/2}$ ist $1-\alpha/2$-Quantil der Standardnormalverteilung.
- Beispiel.

Test von Homogenitaet in einer Kontingenztabelle, bei der Auspraegungen beider Merkmale geordnet sind

- JT-Test (Jonckheere und Terpstra) testet:
  $H_0 : \theta_1 = \theta_2 = ... = \theta_c$ gegen $H_1 : \theta_1 < \theta_2 < ... < \theta_c$.
  - wobei: θ_i ist der Lageparameter der X_{ij}.
- JT-Test-Statistik des Vergleichs der ersten mit der zweiten Stichprobe:
  $U_{ij} = \sum_{t=1}^{n_j} \sum_{s=1}^{n_i} D_{st}$
- JT-Test-Statistik des Vergleichs der i-ten mit der j-ten Gruppe:
  $V = \sum_{i<j} U_{ij}$
- Falls $(V - E(V))/\sqrt{Var(V)} \geq z_{1-\alpha} \Rightarrow H_0$ wird abgelehnt.
- Beispiel.

Kontrollfragen

- Definieren Sie die H_0- und H_1-Hypothesen des Chiquadrat-Tests von Homogenitaet.
- Es ist bekannt, dass man fuer kleinere Stichproben den Fischer-Test verwendet. Definieren Sie die H_0- und H_1-Hypothesen.
- Mithilfe von welchen Test ueberpruefen Sie die Homogenitaet in der Kontingenztabelle mit den geordneten Auspraegungen eines Merkmales?
- Mithilfe von welchen Test ueberpruefen Sie die Homogenitaet in der Kontingenztabelle mit den geordneten Auspraegungen beider Merkmale?