31. Chi-Quadrat-Unabhängigkeitstest

Der Mangel an mathematischer Bildung gibt sich durch nichts so auffallend zu erkennen wie durch maßlose Schärfe im Zahlenrechnen

C.F. Gauß

Der **χ^2-Unabhängigkeitstest** überprüft, ob zwei (beliebig skalierte) Merkmale X und Y voneinander unabhängig sind, was auch als **Kontingenzanalyse** bezeichnet wird. Die jeweiligen Realisierungen liegen dabei als eine (ggf. aus einem Scatter-Plot zu ermittelnde) **Kreuztabelle** mit den absoluten Häufigkeiten h_{ij} vor: h_{ij} gibt die Zahl der Realisierungen an, bei der X in die Klasse i und Y in die Klasse j fällt bzw. X und Y die Ausprägungen x_i und y_j haben z.B.

- Zahl der Prüfungswiederholungen X *vs.* Nebenjob-Stunden pro Woche Y,
- Storchenpopulation *vs.* Geburtenrate,
- Hängt die Art des Schulabschlusses vom Elternhaus ab? ("Wie der Vater, so der Sohn". "Der Apfel fällt nicht weit vom Stamm"). Oder vielleicht vom Geschlecht?
- Klimaerwärmung: Mittlere Temperatur *vs.* Kohlendioxydgehalt.

Der Unabhängigkeitstest prüft, ob eine festgestellte *empirische Unabhängigkeit*, vgl. Kap. 18.2, auch *statistisch signifikant* ist.

Wiederholung der empirischen Unabhängigkeit:

$$f_{ij} = f(x_i)f(y_j) \text{ bzw. } h_{ij} = \frac{h(x_i)h(y_j)}{n}$$

$f_{ij} = \frac{h_{ij}}{n}$ = relative Häufigkeit, n = Stichprobenumfang
$f(x_i)$, $f(y_j)$ = relativen Randhäufigkeiten,
$h(x_i)$, $h(y_j)$ = absolute Randhäufigkeiten.

31.1 Durchführung des Unabhängigkeitstests

Das "4-Schritte-Schema" samt Vorbereitungsschritt ist genau dasselbe wie beim χ^2-Anpassungstest:

0. **Vorbereitung:** Erstellen eines Kreuzdiagramms mit den Klassenzahlen K_x und K_y, genau wie in der deskriptiven Statistik.

1. **Nullhypothese:** X und Y sind unabhängig. Die unter H_0 erwartete absolute Häufigkeit h_{ij}^{e} wird direkt aus dem Unabhängigkeitskriterium der deskriptiven Statistik ermittelt:

$$H_0: \quad h_{ij}^{\mathrm{e}} = \frac{h_i h_j}{n}.$$

Voraussetzung: $h_{ij}^{\mathrm{e}} \geq 5$, ggf. Zusammenfassung von Klassen.

2. **Testvariable:**

$$Q = \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \frac{(h_{ij} - h_{ij}^{\mathrm{e}})^2}{h_{ij}^{\mathrm{e}}} \sim \chi^2(m) \text{ -verteilt}$$

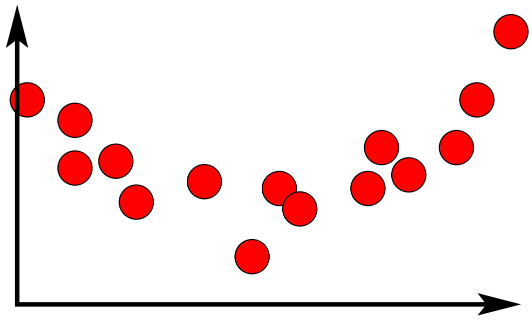
mit $m = (K_x - 1)(K_y - 1)$ Freiheitsgraden.

3. **Realisierung** q von Q aus der Kreuztabelle.

4. **Entscheidung** mit Irrtumswahrscheinlichkeit α: Falls $q > q_{1-\alpha}^{(m)}$, kann die Nullhypothese verworfen werden.

31.2. Bemerkungen zum Unabhängigkeitstest

1. Dieser Unabhängigkeitstest ist weitergehender als ein Korrelationstest; auch signifikante Abhängigkeiten, die eine Korrelation von 0 ergeben (wie in diesem Scatter-Plot), werden detektiert!



2. In der Praxis ist die folgende Umformung der Testfunktion Q vorteilhaft:

$$Q = \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \frac{(h_{ij} - h^{\mathrm{e}}_{ij})^2}{h^{\mathrm{e}}_{ij}} = \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left(\frac{h^2_{ij}}{h^{\mathrm{e}}_{ij}} \right) - n$$

Aufgabe: Zeigen Sie die Gültigkeit der Umformung

3. Die Verallgemeinerung auf mehr als zwei Zufallsvariable und deren paarweise Prüfung ist offensichtlich.

31.3 Beispiel: Ergebnisse zweier Klausuraufgaben

Zugrundegelegt werden die realen Ergebnisse derselben Klausur (OR/Logistik, SS 2001) wie im Beispiel zum χ^2-Anpassungstest.

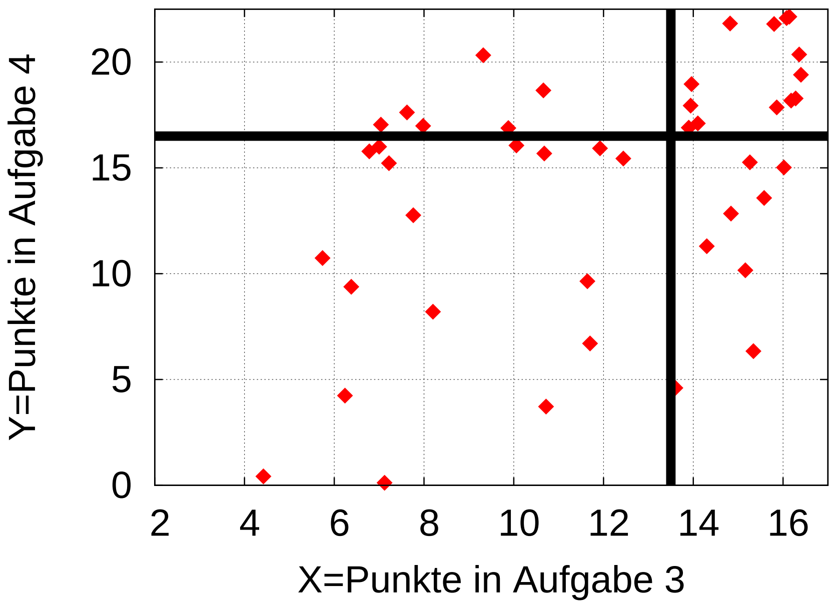
Während dort die Verteilung der $Gesamt$ergebnisse auf die Annahme einer Gaußverteilung getestet wurde, wollen wir nun wissen, ob die erzielten Punktezahlen bei den Aufgaben 3 und 4 dieser Klausur voneinander unabhängig sind, wobei die Art der Verteilung hier $keine$ $Rolle$ spielt.

Anders ausgedrückt: Beeinflusst ein z.B. sehr gutes Abschneiden bei Aufgabe 3 die bedingte Wahrscheinlichkeit, bei Aufgabe 4 gut abzuschneiden?

31.3 (b) Beispiel (Forts.)

Rohdaten: Punktezahlen Aufgabe 3 (Variable X) und Aufgabe 4 (Variable Y):

16-14	14-5	8-18	12-10	11-19	11-16	12-7	11-4
6-11	8-13	7-16	16-22	15-22	15-13	16-18	10-17
14-17	12-16	14-18	14-19	8-17	7-16	16-15	7-17
10-16	16-22	14-17	7-0	16-22	15-10	16-18	8-8
7-15	6-4	15-15	16-18	14-11	9-20	15-6	16-20
6-9	16-19	4-0	12-15				



0. **Vorbereitung: Kreuzdiagramm**
   Wie beim χ^2-Anpassungstest gilt die Bedingung, dass $h_{ij}^{\mathrm{e}} \geq 5$ sein muss. Da $h_{ij}^{\mathrm{e}} \geq 5$ a-priori nicht bekannt, werden als Anhaltspunkt die empirischen absoluten Häufigkeiten genommen.

Aufg. 4 \ Aufg. 3	0-13	14-16	$h(y)$
17-22	6	13	19
0-16	17	8	25
$h(x)$	23	21	$n = 44$

31.3 (c) Beispiel (Forts.)

1. **Nullhypothese:** Ergebnisse sind unabhängig,
$$h^{\mathrm{e}}(x_i, y_j) = \frac{h(x_i)h(y_j)}{n}$$

2. **Testvariable:**
$$Q = \sum_{j=1}^{2} \sum_{k=1}^{2} \frac{(h_{ij} - h_{ij}^{\mathrm{e}})^2}{h_{ij}^{\mathrm{e}}} \sim \chi^2(m) \text{ -verteilt}$$
mit $m = (K_x - 1)(K_y - 1) = 1$ Freiheitsgrad.

3. **Realisierung** aus der Kreuztabelle. Zunächst benötigt man die Summenhäufigkeiten: $h(x_1) = 23$, $h(x_2) = 21$, $h(y_1) = 25$, $h(y_2) = 19$ sowie den Stichprobenumfang $n = \sum h(x_i) = \sum h(y_j) = 44$.

(i, j)	$h(x_i, y_j)$	$h_{ij}^{\mathrm{e}} = \frac{h(x_i)h(y_j)}{n}$	$\frac{h_{ij}^2}{h_{ij}^{\mathrm{e}}}$
(x_1, y_1)	17	23*25/44=13.1	22.1
(x_1, y_2)	6	9.9	3.6
(x_2, y_1)	8	11.9	5.4
(x_2, y_2)	13	9.1	18.6
$\sum$	44	44	49.7

Damit ist die Realisierung q von Q gegeben durch
$$q = 49.7 - 44 = \underline{\underline{5.7}}$$

NB: Das Anwendbarkeitskriterium $h_{ij}^{\mathrm{e}} \geq 5$ ist immer erfüllt.

31.3 (d) Beispiel (Forts.)

4. **Entscheidung und Interpretation:** Falls $q > q_{1-\alpha}^{(m)}$, verwirf die Nullhypothese bei einer Irrtumswahrscheinlichkeit von maximal α.
   Hier ist
   $$q_{1-\alpha}^{(m)} = q_{0.95}^{(1)} = 3.84 < q$$
   Damit ist H_0 verworfen!
   Erlauben wir hingegen nur 1% Irrtumswahrscheinlichkeit für eine fälschliche Ablehnung, wird der Test ungenauer: Es gilt nämlich $q < q_{0.99}^{(1)} = 6.63$, d.h. falls nur 1% Fehlerwahrscheinlichkeit erlaubt wird, ist die Nullhypothse angenommen!

Bemerkungen hirzu:

- Ob bei Ablehnung von H_0 und Vorliegen einer Korrelation die Variablen positiv oder negativ korreliert sind (hier wohl positiv korreliert), sagt uns der Unabhñagigkeitstest *nicht!* Dies ist ggf. mit einem Korrelationstest zu klären, vgl. Formelsammlung.
- Hingegen schlägt er bei *beliebigen* signifikanten Abhängigkeiten an, auch wenn der dazugehörige Korrelationskoefizient =0 sein sollte.
- Der Unabhängigkeitstest und der Korrelationstest ergänzen sich also.

32. Chi-Quadrat-Homogenitätstest

Wenn die Tatsachen nicht mit der Theorie übereinstimmen - um so schlimmer für die Tatsachen.

Hegel, zitiert von Marcuse

Der **χ^2-Homogenitätstest** überprüft, ob *mehrere* unabhängige Stichproben *einer* Variablen X mit Gesamt-Umfang n jeweils zur gleichen (nicht weiter spezifizierten) Verteilung gehören können oder nicht, z.B.

- Einkommensverteilung in Ost- und Westdeutschland,
- Verteilung von Bruchlasten bei Tests von Fahrstuhlseilen vor 1 Monat und jetzt,
- Verteilung der täglichen Veränderungen des DAX-Index im November und Dezember 2001.

Allgemein müssen für diesen Test die Stichproben als Liste von absoluten Häufigkeiten für gewisse Ausprägungen x_i oder Klassen i vorliegen:

Stichprobenindex	x_1	$\cdots$ x_i $\cdots$	x_K	Umfang
1	$h(x_1, 1)$	$\cdots$ $h(x_i, 1)$ $\cdots$	$h(x_K, 1)$	$h(1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	$h(x_1, j)$	$\cdots$ $h(x_i, j)$ $\cdots$	$h(x_K, j)$	$h(j)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$h(x_1)$	$\cdots$ $h(x_i)$ $\cdots$	$h(x_K)$	n

Formal ist der Homogenitätstest nichts anderes als ein Unabhängigkeitstest des zu untersuchenden Merkmals mit dem Stichprobenindex!

32.1 (b) Durchführung des Homogenitätstests

0. **Vorbereitung:** Die Spalten i des Kreuzdiagramms stellen die K_x Ausprägungsklassen von X dar, während der Zeilenindex j die laufende Nummer der M unabhängigen Stichproben bezeichnet.

1. **Nullhypothese:** Die "erwartete" *relative* Häufigkeit $f^{\mathrm{e}}(x_i|j) = h_{ij}/h(j)$ der i-ten Klasse in der Stichprobe j (mit Umfang =Zeilensumme $h(j)$) hängt nicht vom Stichprobenindex j ab: $f^{\mathrm{e}}(x_i|j) = f^{\mathrm{e}}(x_i|j') = f^{\mathrm{e}}(x_i)$ für alle Paare (j, j'). Um den konkreten Wert der relativen Häufigkeiten abzuschätzen, werden alle M Stichproben zu einer "großen" Stichprobe vom Umfang $n = \sum_j h(j)$ (Gesamtzahl der Daten) zusammengefasst: $f^{\mathrm{e}}(x_i) = h(x_i)/n = \sum_{j=1}^{M} h_{ij}/n$ mit der Spaltensumme $h(x_i)$. Die absoluten Häufigkeiten erhält man durch Multiplikation mit dem jeweiligen Stichprobenumfang $h(j)$:

$$H_0 : \quad h_{ij}^{\mathrm{e}} = \frac{h(x_i)h(j)}{n}$$

Dabei sollte, wie beim Unabhängigkeitstest, $h_{ij}^{\mathrm{e}} \geq 5$ gelten.

2-4. Die **Testvariable**, die **Realisierung** sowie die **Entscheidungskriterien** sind identisch zum Unabhängigkeitstest.

32.2 Beispiel: Einkommensverteilung Ost-West

Gegeben sei ein Stichprobenpaar von je 1000 Haushalten in West- und Ostdeutschland, bei denen jeweils die Verteilung des Einkommens ermittelt wurde (Zahlen fiktiv; ähnliches Beispiel vgl. Eckstein):

Stichprobe	0-1500	1500-3000	3000	Umfang $h(j)$
Ost $(j=1)$	400	400	200	1000
West $(j=2)$	200	600	200	1000
$h(x_i)$	600	1000	400	$n=2000$

Zu überprüfen ist die Annahme einer gleichen (aber nicht näher spezifizierten) Verteilung.

1. **Nullhypothese:**
$$h_{ij}^{\mathrm{e}} = \frac{h(x_i)h(j)}{n}.$$
Da die beiden Stichprobenumfänge $h(1) = h(2) = 1000$, gilt für beide Stichproben $j = 1, 2$ jeweils dieselbe erwartete Häufigkeit:
$$\begin{aligned} h_{1j}^{\mathrm{e}} &= 600/2 = 300, \\ h_{2j}^{\mathrm{e}} &= 1000/2 = 500, \\ h_{3j}^{\mathrm{e}} &= 400/2 = 200. \end{aligned}$$

2. **Testvariable:**
$$Q = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{(h_{ij} - h_{ij}^{\mathrm{e}})^2}{h_{ij}^{\mathrm{e}}} \sim \chi^2(m) \text{ -verteilt}$$
mit $m = (K_x - 1)(M - 1) = 2$ Freiheitsgraden.

32.2(b) Beispiel (Forts.)

3. **Realisierung** q von Q aus der Kreuztabelle.

(X_i, j)	h_{ij}	h_{ij}^{e}	$\frac{(h_{ij}-h_{ij}^{\mathrm{e}})^2}{h_{ij}^{\mathrm{e}}}$
$(x_1, 1)$	400	300	33.3
$(x_2, 1)$	400	500	20.0
$(x_3, 1)$	200	200	0.0
$(x_1, 2)$	200	300	33.3
$(x_2, 2)$	600	500	20.0
$(x_3, 2)$	200	200	0.0
$\sum$	2000	2000	$q = 106.7$

4. **Entscheidung und Interpretation:** Falls $q > q_{1-\alpha}^{(m)}$, verwirf die Nullhypothese bei einer Irrtumswahrscheinlichkeit von maximal α.
   Dies ist selbst für $\alpha = 1\%$ *bei weitem* der Fall, da (mit $m = 2$) $q_{0.99}^{(2)} = 9.2 < q \Rightarrow$ die Einkommensverteilung in Ost und West unterscheidet sich signifikant.

Achtung! Der weite Abstand von jedweglicher vernünftigen Fehlergrenze kann trügerisch sein, da beim Test *unabhängige* und *repräsentative* Stichproben vorausgesetzt werden. Eine die Grundgesamtheit repräsentierende Stichprobenauswahl ist aber häufig problematisch zu bekommen und viel schwieriger zu überprüfen als der Test selbst.

Wie würde man da vorgehen?