Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun
http://blog.ruediger-braun.net

Heinrich-Heine-Universität Düsseldorf

30. Januar 2015

Exakter Test nach Fisher

Mendelsche Erbregeln

- Bei den Mendelschen Erbversuchen tritt das Merkmal *Blütenfarbe* in drei Ausprägungen auf, nämlich weiß, rosa und rot
- weiß und rot haben dieselbe Wahrscheinlichkeit, rosa die doppelte
- 4 Blüten werden beobachtet, alle sind rosa
- Ist diese Beobachtung zum Signifikanzniveau $\alpha = 0.05$ mit den Mendelschen Regeln vereinbar?

Interpretation als Vergleich zweier Verteilungen

- Nullhypothese: Die Mendelschen Regeln gelten für die untersuchte Situation
- Das entspricht der Verteilung

Nummer	Ausprägung	Wahrscheinlichkeit
1	weiß	25%
2	rosa	50%
3	rot	25%

- Zu vergleichen mit der tatsächlichen Verteilung der Blütenfarben in dem Kollektiv
- Der Stichprobenumfang ist 4
- Das ist für praktische Zwecke zu wenig, lässt sich aber gut von Hand rechnen

Mendelsche Erbregeln, Fortsetzung

- Ordne die möglichen Ergebnisse mit aufsteigender Wahrscheinlichkeit an
- Entscheidungsstrategie am Beispiel $\alpha = 0.05$

  Lehne H_0 ab, wenn die Beobachtung zu den 5% unwahrscheinlichsten Ereignissen gehört

Test auf Übereinstimmung zweier Verteilungen

- Unabhängige Zufallsvariable $X_1, \ldots, X_n$, die alle mit Wahrscheinlichkeit p_1 den Wert w_1, mit Wahrscheinlichkeit p_2 den Wert w_2, ..., mit Wahrscheinlichkeit p_s den Wert w_s annehmen
- Vergleichswahrscheinlichkeiten $\pi_1, \pi_2, \ldots, \pi_s$ mit $\pi_1 + \pi_2 + \cdots + \pi_s = 1$
- Nullhypothese und Alternative:

$$\begin{aligned} H_0 &: \quad p_1 = \pi_1, p_2 = \pi_2, \ldots, p_s = \pi_s \\ H_1 &: \quad \text{mindestens ein } p_j \neq \pi_j \end{aligned}$$

- Summenvariable

$$\begin{aligned} Y_1 &= \text{Anzahl aller } X_j \text{ mit } X_j = w_1 \\ Y_2 &= \text{Anzahl aller } X_j \text{ mit } X_j = w_2 \\ &\vdots \\ Y_s &= \text{Anzahl aller } X_j \text{ mit } X_j = w_s \end{aligned}$$

Exakter Test nach Fisher

- Bestimme für jede mögliche Kombination von Werten von $Y_1, \ldots, Y_s$ deren Wahrscheinlichkeit gemäß H_0
- Ordne diese Wahrscheinlichkeiten aufsteigend in einer Liste
- Der kritische Bereich, in dem H_0 abgelehnt wird, besteht aus den obersten Zeilen dieser Liste
- Man nimmt die maximale Anzahl von Zeilen, so dass die erlaubte Fehlerwahrscheinlichkeit erster Art nicht überschritten wird

Beispiel Mendel: Formalisierung

- $s = 3$
- X_1 ist der Zahlencode der Blütenfarbe der ersten Blüte, X_2 dasselbe für die zweite Blüte, ...
- Y_1 bezeichnet die Anzahl der weißen, Y_2 die der rosafarbenen und Y_3 die der roten Blüten
- Dann $Y_1 + Y_2 + Y_3 = 4$
- Im Beispiel $Y_1 = 0$, $Y_2 = 4$, $Y_3 = 0$
- Rechne sämtliche Einzelwahrscheinlichkeiten aus

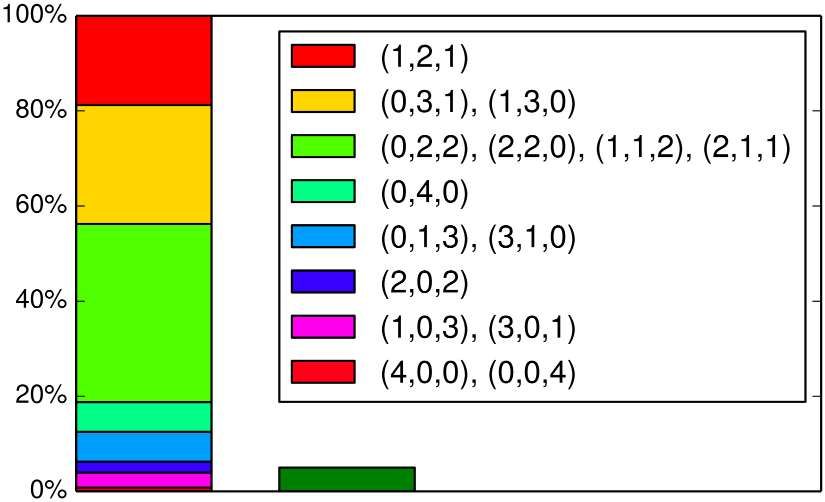
Beispiel Mendel: Wahrscheinlichkeiten der Einzelereignisse

$$
\begin{aligned}
P(Y_1 = k_1, Y_2 = k_2, Y_3 = k_3) \\
&= \binom{4}{k_1} \cdot \binom{4-k_1}{k_2} \cdot \left(\frac{1}{4}\right)^{k_1} \cdot \left(\frac{1}{2}\right)^{k_2} \cdot \left(\frac{1}{4}\right)^{k_3} \\
&= \frac{4! \cdot (4-k_1)!}{k_1! \cdot (4-k_1)! \cdot k_2! \cdot (4-k_1-k_2)!} \cdot \left(\frac{1}{4}\right)^{k_1} \cdot \left(\frac{1}{2}\right)^{k_2} \cdot \left(\frac{1}{4}\right)^{k_3} \\
&= \frac{4!}{k_1! \cdot k_2! \cdot k_3!} \cdot \left(\frac{1}{4}\right)^{k_1} \cdot \left(\frac{1}{2}\right)^{k_2} \cdot \left(\frac{1}{4}\right)^{k_3}
\end{aligned}
$$

Beispiel Mendel: Tabelle der W'keiten der Einzelereignisse

k_1	k_2	k_3	$P(X_1 = k_1, X_2 = k_2, X_3 = k_3)$	kumulierte Summe
0	0	4	0.0039	0.0039
4	0	0	0.0039	0.0078
1	0	3	0.0156	0.0234
3	0	1	0.0156	0.0391
2	0	2	0.0234	0.0625
0	1	3	0.0312	0.0938
3	1	0	0.0312	0.1250
0	4	0	0.0625	0.1875
0	2	2	0.0938	0.2812
1	1	2	0.0938	0.3750
2	1	1	0.0938	0.4688
2	2	0	0.0938	0.5625
0	3	1	0.1250	0.6875
1	3	0	0.1250	0.8125
1	2	1	0.1875	1.0000

Beispiel Mendel: Balkendiagramm



Der linke Balken zeigt die kumulierten Werte aus der Tabelle, der rechte die 5%-Schwelle

Beispiel Mendel: Ergebnis

- In den folgenden Fällen kann die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ abgelehnt werden
  - 4 weiße oder 4 rote Blüten
  - keine rosa, aber 3 weiße oder 3 rote Blüten
- Der p-Wert des beobachteten Ereignisses "4 rosa Blüten" beträgt 18.75%

Der F-Test zum Vergleich zweier Varianzen

Warum vergleicht man zwei Varianzen?

- Um Unterschiedlichkeit zweier Verteilungen nachzuweisen
- Um Voraussetzungen eines anderen Tests zu prüfen
- Um eine ANOVA zu rechnen
- ANOVA="Analysis of Variance": Ein Test, mit welchem man den Einfluss der Gruppenzugehörigkeit auf einen Parameter prüfen kann, indem man Varianzen vergleicht

F-Test zum Vergleich zweier Varianzen

- $X_1, \ldots, X_{n_1}$ und $Y_1, \ldots, Y_{n_2}$ bezeichnen zwei Gruppen von Messwerten
- Verteilungsvoraussetzungen:
  - Die X_j sind verteilt gemäß $N(\mu_1, \sigma_1^2)$, wobei μ_1 und σ_1 unbekannt sind
  - Die Y_j sind verteilt gemäß $N(\mu_2, \sigma_2^2)$, wobei μ_2 und σ_2 unbekannt sind
- Ziel: σ_1 und σ_2 sollen verglichen werden

F-Test, Teststatistik

- x_j und y_j seien Realisierungen.
- Bestimme arithmetische Mittelwerte und Stichprobenstreuungen

$$\overline{x} = \frac{1}{n_1} \sum_{j=1}^{n_1} x_j \qquad s_x = \sqrt{\frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (x_j - \overline{x})^2}$$

$$\overline{y} = \frac{1}{n_2} \sum_{j=1}^{n_2} y_j \qquad s_y = \sqrt{\frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (y_j - \overline{y})^2}$$

- Die Teststatistik ist

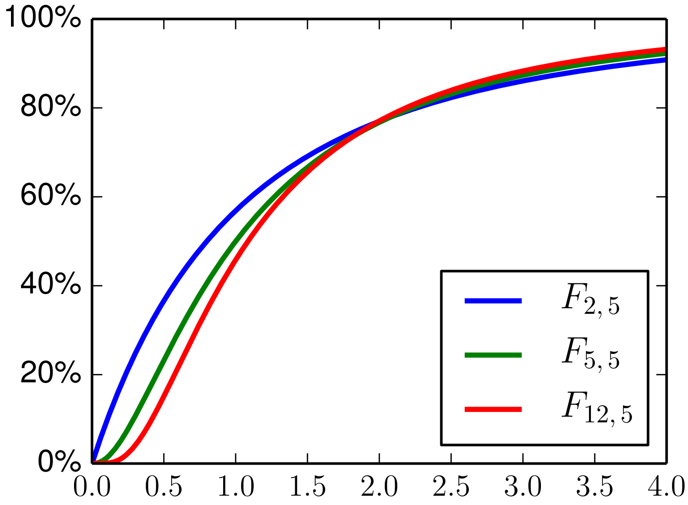
$$t = \frac{s_x^2}{s_y^2}$$

Die F-Verteilung

- Der F-Test verwendet die F-Verteilung
- Für f_1 bzw. f_2 Freiheitsgrade sind die Quantile der F-Verteilung für α nahe 1 tabelliert.
- Für α nahe 0 benutzt man die Formel

$$F_{f_1,\, f_2,\, \alpha} = \frac{1}{F_{f_2,\, f_1,\, 1-\alpha}}$$

Verteilungsfunktionen von F-Verteilungen



Quantile $f_{f_1,\, f_2,\, 0.95}$ der F-Verteilungen

f_2 \ f_1	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	**6.26**	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	**5.19**	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02

Beispiele:

- $f_{4,\, 5,\, 0.95} = 5.19$
- $f_{5,\, 4,\, 0.95} = 6.26$

F-Test, Entscheidungsregel

- Das Signifikanzniveau sei α
- Wir bestimmen die Quantile der F-Verteilung

$F_{n_1-1,\, n_2-1,\, 1-\alpha/2}$	beim zweiseitigen Test
$F_{n_1-1,\, n_2-1,\, 1-\alpha}$	bei einem einseitigen Test

- Entscheidung

$H_0 = \{\sigma_1 = \sigma_2\}$: Die Nullhypothese H_0 wird abgelehnt, wenn $t > F_{n_1-1,\, n_2-1,\, 1-\alpha/2}$ oder $t < \dfrac{1}{F_{n_2-1,\, n_1-1,\, 1-\alpha/2}}$

$H_0 = \{\sigma_1 \leq \sigma_2\}$: Die Nullhypothese H_0 wird abgelehnt, wenn $t > F_{n_1-1,\, n_2-1,\, 1-\alpha}$

$H_0 = \{\sigma_1 \geq \sigma_2\}$: Die Nullhypothese H_0 wird abgelehnt, wenn $t < \dfrac{1}{F_{n_2-1,\, n_1-1,\, 1-\alpha}}$

F-Test, Beispiel

- Beim t-Test für unverbundene Stichproben hatten wir im Beispiel "Bodenbakterium" zwei Datensätze erhalten
- Der erste hatte $n_x = 10$ und $s_x = 7.972$
- Der zweite hatte $n_y = 9$ und $s_y = 6.280$
- Können wir zum Signifikanzniveau $\alpha = 0.10$ ausschließen, dass beide Verteilungen dieselbe Varianz aufweisen?
- Teststatistik
$$t = \frac{s_x^2}{s_y^2} = \frac{63.55}{39.44} = 1.611$$
- Das benötigte Quantil ist $f_{9,8,0.95} = 3.39$
- H_0 kann nicht abgelehnt werden

ANOVA

Beispielhafte Fragestellung: Unterrichtsmethoden

- Jeweils 4 bis 5 Schüler wurden nach einer von 4 Methoden in Statistik unterrichtet. Hat die Wahl der Unterrichtsmethode überhaupt einen Einfluss auf den Lernerfolg?
- Daten: (Der Erfolg wurde auf einer Skala von 0 bis 8 gemessen)

Unterrichtsmethode			
1	2	3	4
2	3	6	5
1	4	8	5
3	3	7	5
1	5	6	3
	0	8	2

Unterrichtsmethoden, Fortsetzung

- Die Nullhypothese ist, dass alle Methoden bis auf zufällige Abweichungen dasselbe Ergebnis liefern
- Wir könnten je zwei Unterrichtsmethoden mit einem t-Test für unverbundene Stichproben testen
- Das sind $\binom{4}{2} = 6$ Paarvergleiche
- Die Bonferroni-Korrektur löst dieses Problem im Prinzip, aber auf Kosten der Power

Quellen der Variabilität: zufällige Effekte

Messfehler: Die Körpergröße einer Person wird fünfmal gemessen. Die Ergebnisse werden voneinander abweichen.
Durch Sorgfalt und geeignete Messmethoden kann der Messfehler beeinflusst werden. Ganz auszuschalten ist er nicht.

Natürliche Variabilität: Innerhalb einer Altersgruppe sind die Probanden unterschiedlich groß.
Diese Variabilität ist unvermeidlich.

Diese beiden Quellen der Variabilität machen die statistische Betrachtung überhaupt erst erforderlich. Man fasst sie häufig als zufällige Effekte zusammen.

Quellen der Variabilität: Gruppenunterschiede

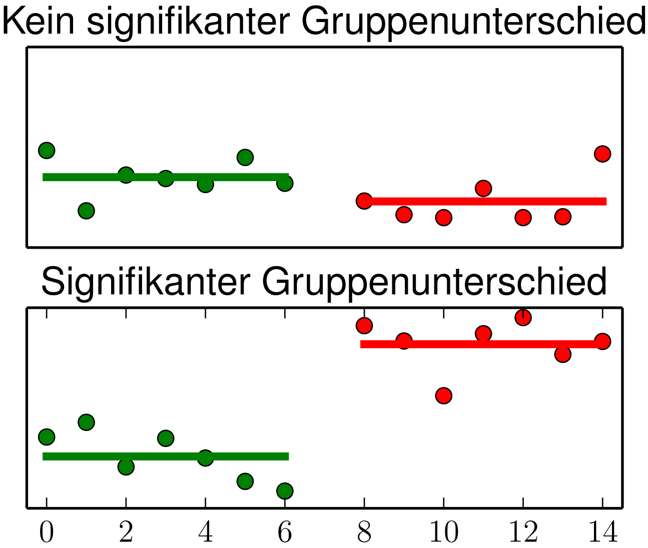
- Ziel ist die Untersuchung des Einflusses eines *Faktors* auf das Messergebnis. Beispiele für Faktoren sind etwa Alter, Unterrichtsmethode, Sonneneinstrahlung.
- Der Einfachheit halber gehen wir von nur einem Faktor aus.
- Der Faktor besitzt endlich viele Faktorstufen, in den Beispielen etwa

  Alter: 10–14, 15–19, 20–24, $\geq$ 25 Jahre
  Unterichtsmethode: Methoden 1 bis 4
  Sonneneinstrahlung: sonnig, halbschattig, schattig
- Zu jeder Faktorstufe werden mehrere Individuen ausgewählt und zu einer *Gruppe* zusammengefasst.
- Falls es Unterschiede zwischen den Gruppen gibt, so tragen sie zur Varianz bei.

Idee der Varianzanalyse

- Es ist rechnerisch möglich, die empirische Varianz aufzuteilen in den von den zufälligen Effekten und den von den Gruppenunterschieden verursachten Anteil.
- Dabei geht man davon aus, dass die zufälligen Effekte in allen Gruppen gleich wirken.
- Falls sich die empirische Varianz der zufälligen Effekte signifikant von dem durch die Gruppenunterschiede verursachten Anteil unterscheidet, dann ist der Einfluss des Faktors auf den Messwert nachgewiesen.
- Zum Vergleich dieser beiden Varianzen wird ein F-Test eingesetzt.

Idee der Varianzanalyse



Gruppenmittelwerte

- Es gibt k Faktorstufen und zu jeder dieser Faktorstufen eine Gruppe.
- Die j-te Gruppe hat n_j Elemente, welche mit

  $$x_{j,1}, x_{j,2}, \ldots, x_{j,n_j}$$

  bezeichnet werden.
- Das arithmetische Mittel der Daten in der j-ten Gruppe ist

  $$\overline{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{j,i}$$

  Es ist der *Gruppenmittelwert.*

Gruppenmittelwerte im Beispiel

Unterrichtsmethode			
1	2	3	4
$x_{1,1} = 2$	$x_{2,1} = 3$	$x_{3,1} = 6$	$x_{4,1} = 5$
$x_{1,2} = 1$	$x_{2,2} = 4$	$x_{3,2} = 8$	$x_{4,2} = 5$
$x_{1,3} = 3$	$x_{2,3} = 3$	$x_{3,3} = 7$	$x_{4,3} = 5$
$x_{1,4} = 1$	$x_{2,4} = 5$	$x_{3,4} = 6$	$x_{4,4} = 3$
	$x_{2,5} = 0$	$x_{3,5} = 8$	$x_{4,5} = 2$
$\overline{x}_1 = 1.75$	$\overline{x}_2 = 3.00$	$\overline{x}_3 = 7.00$	$\overline{x}_4 = 4.00$

Gruppengrößen: $n_1 = 4$, $n_2 = n_3 = n_4 = 5$

Gesamtmittelwert

- Der Gesamtstichprobenumfang ist $N = n_1 + \cdots + n_k$, im Beispiel also $N = 19$
- Der Gesamtmittelwert ist der Mittelwert über alle Daten

$$\overline{\overline{x}} = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{j,i}$$

- Man kann ihn als gewichtetes arithmetisches Mittel der Gruppenmittelwerte berechnen

$$\overline{\overline{x}} = \frac{1}{N} \sum_{j=1}^{k} n_j \cdot \overline{x}_j$$

- Im Beispiel

$$\overline{\overline{x}} = \frac{1}{19} (4 \cdot 1.75 + 5 \cdot 3.00 + 5 \cdot 7.00 + 5 \cdot 4.00) = \frac{77}{19} = 4.053$$

Zerlegung der Varianz

- Für jeden Gruppenmittelwert x_j sei a_j die Differenz zwischen $\overline{x}_j$ und $\overline{\overline{x}}$, also

$$\overline{x}_j = \overline{\overline{x}} + a_j$$

- Für jeden einzelnen Messwert $x_{j,i}$ sei $e_{j,i}$ die Differenz zwischen $x_{j,i}$ und $\overline{x}_j$, also

$$x_{j,i} = \overline{x}_j + e_{j,i} = \overline{\overline{x}} + a_j + e_{j,i}$$

Gruppenmittelwerte im Beispiel

	Unterrichtsmethode			
j	1	2	3	4
$x_{j,1}$	1.75+0.25	3.00+0.00	7.00−1.00	4.00+1.00
$x_{j,2}$	1.75−0.75	3.00+1.00	7.00+1.00	4.00+1.00
$x_{j,3}$	1.75+1.25	3.00+0.00	7.00+0.00	4.00+1.00
$x_{j,4}$	1.75−0.75	3.00+2.00	7.00−1.00	4.00−1.00
$x_{j,5}$		3.00−3.00	7.00+1.00	4.00−2.00
$\overline{x}_j$	4.05−2.30	4.05−1.05	4.05+2.95	4.05−0.05

$$\overline{\overline{x}} = 4.05$$

Berechnung der Quadratsumme

- Für die bei der Berechnung der empirischen Varianz auftauchenden Differenzen bedeutet das

$$x_{j,i} - \overline{\overline{x}} = a_j + e_{j,i} = (\overline{x}_j - \overline{\overline{x}}) + (x_{j,i} - \overline{x}_j)$$

- Zur Berechnung des Quadrats ziehen wir die erste binomische Formel heran

$$(x_{j,i} - \overline{\overline{x}})^2 = (\overline{\overline{x}} - \overline{x}_j)^2 + 2(\overline{\overline{x}} - \overline{x}_j) \cdot (\overline{x}_j - x_{j,i}) + (\overline{x}_j - x_{j,i})^2$$

Zerlegung der Varianz, Fortsetzung

Das muss zuerst über alle i und dann noch über alle j summiert werden. Wir fangen mit der Summe über alle i an. Dabei ist j die Nummer einer festen Gruppe

$$\sum_{i=1}^{n_j}(x_{j,i}-\overline{\overline{x}})^2 = n_j \cdot (\overline{\overline{x}}-\overline{x}_j)^2 + 2(\overline{\overline{x}}-\overline{x}_j) \cdot \underbrace{\sum_{i=1}^{n_j}(\overline{x}_j - x_{j,i})}_{=0} + \sum_{i=1}^{n_j}(\overline{x}_j - x_{j,i})^2$$

$$= n_j \cdot (\overline{\overline{x}}-\overline{x}_j)^2 + \sum_{i=1}^{n_j}(\overline{x}_j - x_{j,i})^2$$

Aufsummiert über alle j

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{j,i}-\overline{\overline{x}})^2 = \sum_{j=1}^{k} n_j \cdot (\overline{\overline{x}}-\overline{x}_j)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\overline{x}_j - x_{j,i})^2$$

Abkürzungen

Quadratsumme, welche die Gesamtvariabilität repräsentiert:

$$\mathrm{SQT} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{j,i} - \overline{\overline{x}})^2$$

Quadratsumme, welche die Variabilität zwischen den Faktor-Stufen repräsentiert:

$$\mathrm{SQZ} = \sum_{j=1}^{k} n_j \cdot (\overline{\overline{x}} - \overline{x}_j)^2$$

Quadratsumme, welche den Anteil der zufälligen Effekte repräsentiert:

$$\mathrm{SQI} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\overline{x}_j - x_{j,i})^2$$

Freiheitsgrade

- Abgekürzt lautet die gefundene Gleichung

$$\mathrm{SQT} = \mathrm{SQZ} + \mathrm{SQI}$$

- Wenn der Faktor keinen Einfluss hat, dann kann man aus jeder dieser drei Größen die Varianz schätzen. Man muss nur die Zahl der Freiheitsgrade kennen.
- Der Gesamtversuch hat $N-1$ Freiheitsgrade,

$$\mathrm{MQT} = \frac{\mathrm{SQT}}{N-1}$$

ist die empirische Varianz.

Freiheitsgrade, Fortsetzung

- In SQI stecken k Schätzer, nämlich die Gruppenmittelwerte. Deswegen ist

$$\text{MQI} = \frac{\text{SQI}}{N-k}$$

ebenfalls ein Schätzer für die Varianz, wenn die Gruppe keinen Einfluss auf die Verteilung hat.

- $k-1$ Freiheitsgrade bleiben übrig

$$\text{MQZ} = \frac{\text{SQZ}}{k-1}$$

ist ebenfalls ein Schätzer für die Varianz.

Entscheidungsregel

- Die Teststatistik ist

$$t = \frac{\text{MQZ}}{\text{MQI}}$$

- Die Nullhypothese, dass der Faktor keinen Einfluss besitzt, wird abgelehnt, wenn

$$t > f_{k-1,\, N-k,\, 1-\alpha}$$

- Hierbei ist $f_{k-1,\, N-k,\, 1-\alpha}$ ein Quantil der F-Verteilung.

Anova, Zusammenfassung

- Gegeben k-Gruppen von Messwerten

$$
\begin{array}{l}
X_{1,1}, X_{1,2}, \ldots, X_{1,n_1} \\
X_{2,1}, X_{2,2}, \ldots, X_{2,n_2} \\
\ldots \\
X_{k,1}, X_{k,2}, \ldots, X_{k,n_k}
\end{array}
$$

- Dann ist $N = n_1 + \cdots + n_k$ der Gesamtstichprobenumfang.
- Verteilungsvoraussetzungen: Die j-te Gruppe von Messwerten ist verteilt gemäß $N(\mu_j, \sigma^2)$ für unbekannte Werte μ_j und σ. Dabei hängt σ nicht von der Gruppe ab.
- Ziel: Die μ_j sollen miteinander verglichen werden.

Anova, Fortsetzung

- $x_{1,1}$ usw. seien die Realisierungen.
- Bestimme arithmetische Mittelwerte und Schätzer für die Varianzen

$$\overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{j,i} \qquad \overline{\overline{x}} = \frac{1}{N}\sum_{j=1}^{k} n_j \cdot \overline{x}_j$$

$$\text{SQZ} = \sum_{j=1}^{k} n_j \cdot (\overline{\overline{x}} - \overline{x}_j)^2 \qquad \text{MQZ} = \frac{\text{SQZ}}{k-1}$$

$$\text{SQI} = \sum_{j=1}^{k}\sum_{i=1}^{n_j} (\overline{x}_j - x_{j,i})^2 \qquad \text{MQI} = \frac{\text{SQI}}{N-k}$$

- Die Teststatistik ist

$$t = \frac{\text{MQZ}}{\text{MQI}}$$

Anova, Fortsetzung

- Die Nullhypothese ist $H_0 = \{\mu_1 = \mu_2 = \cdots = \mu_k\}$.
- Das Signifikanzniveau sei α
- Das folgende Quantil einer F-Verteilung wird benötigt

$$f_{k-1,\, N-k,\, 1-\alpha}$$

- Entscheidung: Die Nullhypothese wird abgelehnt, wenn

$$t > f_{k-1,\, N-k,\, 1-\alpha}$$

Gruppenmittelwerte im Beispiel

	Unterrichtsmethode			
j	1	2	3	4
$x_{j,1}$	1.75+0.25	3.00+0.00	7.00−1.00	4.00+1.00
$x_{j,2}$	1.75−0.75	3.00+1.00	7.00+1.00	4.00+1.00
$x_{j,3}$	1.75+1.25	3.00+0.00	7.00+0.00	4.00+1.00
$x_{j,4}$	1.75−0.75	3.00+2.00	7.00−1.00	4.00−1.00
$x_{j,5}$		3.00−3.00	7.00+1.00	4.00−2.00
$\overline{x}_j$	4.05−2.30	4.05−1.05	4.05+2.95	4.05−0.05

$$\text{SQZ} = \sum_{j=1}^{4} n_j(\overline{\overline{x}} - \overline{x}_j)^2 = 4 \cdot 2.30^2 + 5 \cdot 1.05^2 + 5 \cdot 2.95^2 + 5 \cdot 0.05^2 = 70.20$$

$$\text{SQI} = \sum_{j=1}^{4} \sum_{i=1}^{n_j} (\overline{x}_j - x_{ji})^2$$

$$= 0.25^2 + 0.75^2 + 1.25^2 + 0.75^2 + 0.00^2 + \cdots + 1.00^2 + 2.00^2 = 28.75$$

Beispiel Unterrichtsmethode

- SQZ $= 70.20$, also MQZ $= \frac{\text{SQZ}}{4-1} = 23.4$
- SQI $= 28.75$, also MQI $= \frac{SQI}{19-4} = 1.917$
- Der zufällige Streuungsanteil ist $\sqrt{\text{MQI}} = 1.385$
- SQT $=$ SQZ $+$ SQI $= 98.95$, also MQT $= \frac{\text{SQT}}{19-1} = 5.497$
- Die Streuung über alle Daten ist $\sqrt{\text{MQT}} = 2.282$
- Die Teststatistik ist gleich
$$t = \frac{\text{MQZ}}{\text{MQI}} = 12.21$$
- Das Quantil ist $f_{3,\,15,\,0.95} = 3.287$
- Zum Signifikanzniveau $\alpha = 0.05$ ist nachgewiesen worden, dass die Unterrichtsmethode Einfluss auf den Lernerfolg hat

Quantile $f_{f_1,\,f_2,\,0.95}$ der F-Verteilungen

	f_1					
f_2	1	2	3	4	5	6
1	161.45	199.50	215.71	224.58	230.16	233.99
2	18.51	19.00	19.16	19.25	19.30	19.33
3	10.13	9.55	9.28	9.12	9.01	8.94
4	7.71	6.94	6.59	6.39	**6.26**	6.16
5	6.61	5.79	5.41	**5.19**	5.05	4.95
6	5.99	5.14	4.76	4.53	4.39	4.28
7	5.59	4.74	4.35	4.12	3.97	3.87
8	5.32	4.46	4.07	3.84	3.69	3.58
9	5.12	4.26	3.86	3.63	3.48	3.37
10	4.96	4.10	3.71	3.48	3.33	3.22
11	4.84	3.98	3.59	3.36	3.20	3.09
12	4.75	3.89	3.49	3.26	3.11	3.00
13	4.67	3.81	3.41	3.18	3.03	2.92
14	4.60	3.74	3.34	3.11	2.96	2.85
15	4.54	3.68	3.29	3.06	2.90	2.79