8. Vergleich mehrerer Stichproben

Andreas Eberle
Institut für angewandte Mathematik

Dezember 2008

Häufig wollen wir verschiedene Populationen, Verfahren, usw. miteinander vergleichen.

- **Beipiel:**
  - Vergleich der Wohnungsmieten in Köln, Bonn und Frankfurt
  - Gesundheitszustand von Patienten mit und ohne Medikament
- Dies kann man auch als einfachste Version eines **multivariaten Problems** ansehen: Wir wollen wissen, ob und wie die Verteilung einer "**abhängigen Variable**" (Wohnungsmiete, Gesundheitszustand) von den Werten einer "**unabhängigen Variable (Faktor)**" (Ort, Medikament) abhängt.
- Wir können die Grundgesamtheit bzw. die gesamte Stichprobe dann in mehrere Teilpopulationen bzw. Teilstichproben zerlegen, die den verschiedenen Werten des Faktors entsprechen - z.B. Wohnungsmieten in Köln - Wohnungsmieten in Bonn - Wohnungsmieten in Frankfurt.
- Ziel ist der Vergleich der Verteilungen in den Teilpopulationen bzw. das Aufdecken von Abhängigkeiten der Variablen.

Man unterscheidet u.a. folgende Problemstellungen bzw. Verfahren:

- **Zweistichprobenproblem:** Die unabhängige Variable kann nur zwei Werte annehmen (*z.B. Behandlung mit Medikament oder Placebo*). Die Grundgesamtheit zerfällt dann in zwei Teilpopulationen, in denen wir z.B. den Erwartungswert der abhängigen Variable vergleichen wollen.
- **Einfaktorielle Varianzanalyse (ANOVA):** Die unabhängige Variable (der Faktor) kann mehr als zwei Werte annehmen - die Grundgesamtheit zerfällt dementsprechend in mehrere Gruppen. Wir wollen beipielsweise wissen, ob sich der Erwartungswert in irgendwelchen dieser Gruppen unterscheidet (*z.B. Düngung mit fünf verschiedenen Düngersorten: Gibt es überhaupt einen signifikanten Unterschied zwischen den Erträgen ?*)
- **Zweifaktorielle ANOVA:** Analyse der gemeinsamen Wirkungsweise zweier Faktoren auf eine unabhängige Variable (*nicht in dieser Vorlesung*).

- **Test auf Unabhängigkeit:** Gibt es eine Abhängigkeit zwischen zwei verschiedenen Variablen ?
  (*z.B. Abhängigkeit der Schulnote von der für Hausaufgaben aufgewendeten Zeit*)
- **Regressionsanalyse:** Zusammenhang zwischen einer abhängigen und einer oder mehrerer unabhängiger Variablen (alle metrisch skaliert) unter Annahme eines linearen Einflusses der unabhängigen Variablen
  (*nicht in dieser Vorlesung*).

8.1. Zweistichprobenproblem für verbundene Stichproben

- Eine **verbundene Stichprobe** liegt vor, wenn n Elemente zufällig ausgewählt werden, an denen **jeweils** zwei verschiedene Experimente durchgeführt werden.
- **Beispiel:**
  - Blutplättchenaggregation **derselben** Versuchsperson vor und nach dem Rauchen einer Zigarette.
  - Vergleich der Englisch- und der Mathematiknoten **derselben** Schüler.
- Falls es nicht möglich ist, die Experimente an den gleichen Einheiten nacheinander durchzuführen, kann man unter Umständen auch n Paare auswählen. Die beiden Elemente im Paar sollten sich dann aber "sehr ähnlich" sein.
- **Beispiel:** Ist es beim Vergleich zweier Medikamente nicht zumutbar, dem Patienten beide Medikamente nacheinander zu verabreichen, dann kann man auch Paare von ähnlichen Patienten bilden, und beiden jeweils nur ein Medikament verabreichen.

Zweistichprobenproblem für verbundene Stichproben

- Will man die Mittelwerte bzw. Mediane der zugrundeliegenden Verteilungen ausgehend von verbundenen Stichproben

  $$x_1, x_2, \ldots, x_n \qquad \text{unter Versuchsbedingung 1,}$$
  $$y_1, y_2, \ldots, y_n \qquad \text{unter Versuchsbedingung 2,}$$

  eines metrischen Merkmals vergleichen, dann bildet man einfach die Differenzen

  $$u_i = x_i - y_i$$

  der Beobachtungswerte, und wendet auf diese einen t-Test, Wilcoxon-Test, oder Vorzeichentest an.
- Kein Unterschied zwischen den beiden Veruchsbedingungen heißt dann einfach $E\left[U_i\right] = 0$ bzw. $Median\left(U_i\right) = 0$.

Beispiel. (Herzfrequenz von Ratten)

i	1	2	3	4	5	6	7	8	9	10
zusammen x_i	523	494	461	535	476	454	448	408	470	437
allein y_i	463	462	462	456	450	426	418	415	409	402
Differenz u_i	60	32	-1	79	26	28	30	-7	61	35

- $H_0: \quad E[X] = E[Y] \quad \Longleftrightarrow \quad E[U] = 0$
- $H_1: \quad E[X] > E[Y] \quad \Longleftrightarrow \quad E[U] > 0$
- **t-Test:**

$$t = \frac{\overline{u}_n}{s_n / \sqrt{n}} = \frac{34.3}{26.8 / \sqrt{10}} = 4.05$$

- Wegen $t_{9,0.99} = 2.7$ ist die Herzfrequenz von Ratten bei Zusammensein signifikant erhöht.
- **Vorzeichentest:**

$$\begin{aligned} v &= \text{Anzahl positiver Vorzeichen} = 8 \\ p &= P[V \geq 8] = \binom{10}{8} \cdot 2^{-10} + \binom{10}{9} \cdot 2^{-10} + \binom{10}{10} \cdot 2^{-10} = 0.055 \end{aligned}$$

- Keine Ablehnung von H_0 zum Signifikanzniveau $\alpha = 0.05$.

Zweistichprobenproblem für verbundene Stichproben

- **Voraussetzung - Statistisches Modell:**
  - $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ sind unabhängige gepaarte Stichproben.
  - $X_1, X_2, \ldots, X_n$ bzw. $Y_1, Y_2, \ldots, Y_n$ sind jeweils identisch verteilt.
  - Eine Abhängigkeit zwischen X_i und Y_i ist möglich.
  - Die Differenzen $U_i = X_i - Y_i$ sind $N(m, \sigma^2)$ verteilt.
- $\mathbf{H}_0 : m = 0$, $\quad \mathbf{H}_1 : m \neq 0$ (bzw $m > 0$, $m < 0$)
- **Teststatistik:**

$$T = \frac{\overline{U}_n}{S_n / \sqrt{n}} \,, \qquad S_n^2 = \frac{1}{n-1} \sum_i \left(U_i - \overline{U}_n \right)^2$$

- **Verteilung unter Nullhypothese:**

$$T \sim t(n-1) \qquad (\approx N(0,1) \text{ für große } n)$$

Zweistichprobenproblem für verbundene Stichproben

t-Test

- **p-Wert:** $t =$ aus Daten berechneter (realisierter) Wert der T-Statistik

  $$p = P\left[|T| \geq |t|\right] \quad \text{bzw.} \quad p = P\left[T \geq t\right] \quad \text{bzw.} \quad p = P\left[T \leq t\right]$$

- **Testentscheidung via p-Wert**: H_0 wird verworfen, falls $p \leq \alpha$.

- **Testentscheidung über Quantile:** H_0 wird verworfen, falls

  $$|t| \geq t_{n-1,1-\alpha/2} \quad \text{bzw.} \quad t \geq t_{n-1,1-\alpha} \quad \text{bzw.} \quad t \leq t_{n-1,\alpha}$$

- **Zu beachten:** Die Normalverteilungsannahme sollte zum Beispiel mithilfe eines Normal QQ-Plots überprüft werden. Evtl. zusätzlich oder alternativ Wilcoxon- bzw. Vorzeichentest verwenden.

Zweistichprobenproblem für verbundene Stichproben

- **Voraussetzung - Statistisches Modell:**
  - $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ sind unabhängige gepaarte Stichproben.
  - $X_1, X_2, ..., X_n$ bzw. $Y_1, Y_2, ..., Y_n$ sind jeweils identisch verteilt.
  - Die Verteilungen sind stetig mit Median μ_X bzw. μ_Y
  - Eine Abhängigkeit zwischen X_i und Y_i ist möglich.
- $\mathbf{H}_0 : \mu_X = \mu_Y$, $\mathbf{H}_1 : \mu_X \neq \mu_Y$ (bzw $\mu_X > \mu_Y$, $\mu_X < \mu_Y$)
- **Teststatistik:**

$$V = \text{Anzahl der positiven Vorzeichen von } X_i - Y_i$$

- **Verteilung unter Nullhypothese:**

$$V \sim Bin(n, 1/2) \quad (\approx N(n/2, n/4) \text{ für große } n)$$

Zweistichprobenproblem für verbundene Stichproben

Vorzeichentest

- **p-Wert:** v =Beobachtungswert der V-Statistik

$$p = P\left[\left|V - \frac{n}{2}\right| \geq \left|v - \frac{n}{2}\right|\right]$$

$$\text{bzw. } p = P[V \geq v] \qquad \text{bzw. } p = P[V \leq v]$$

- **Testentscheidung via p-Wert**: H_0 wird verworfen, falls $p \leq \alpha$.

- **Testentscheidung über Quantile:** H_0 wird verworfen, falls

$$v \geq b_{1-\alpha/2} \text{ oder } v \leq b_{1-\alpha/2} \,,$$

bzw. $v \geq b$[redacted]$_{1-\alpha}$, bzw. $v \leq b$[redacted]$_{\alpha}$,

wobei b_α Quantil von $Bin(n, 1/2)$.

- **Zu beachten:** Beim Zweistichprobenproblem kann auch oft der mächtigere Wilcoxontest angewendet werden.

Zweistichprobenproblem für verbundene Stichproben

Wilcoxon-Test

- **Voraussetzung - Statistisches Modell:**
  - $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ sind unabhängige, identisch verteilte gepaarte Stichproben.
  - Die Verteilungen der X_i bzw. Y_i sind stetig.
  - Die Verteilung von $U_i = X_i - Y_i$ ist symmetrisch mit Median μ.
  - Eine Abhängigkeit zwischen X_i und Y_i ist möglich.
- $\mathbf{H}_0 : \mu = 0$, $\quad \mathbf{H}_1 : \mu \neq 0$ (bzw $\mu > 0$, $\mu < 0$)
- **Teststatistik:**

$$W = \text{Summe der Rangplätze } R_i \text{ der postiven Paardifferenzen } U_i,$$
$$\text{wenn man alle Differenzen dem Absolutbetrag } |U_i| \text{ nach ordnet.}$$

- **Verteilung unter Nullhypothese:** Für große n näherungsweise normalverteilt, für kleine n benutze Statistikprogramm.

Zweistichprobenproblem für verbundene Stichproben

Wilcoxon-Test

- **p-Wert:** Berechnung via Statistikprogramm.
- **Testentscheidung via p-Wert**: H_0 wird verworfen, falls $p \leq \alpha$.

8.2. Zweistichprobenproblem für unverbundene Stichproben

- Bei einer **unverbundenen (ungepaarten) Stichprobe** wählt man aus der Grundgesamtheit zufällig n Objekte der ersten Teilpopulation (bzw. n Objekte, für die ein Experiment unter der ersten Versuchsbedingung durchgeführt wird), und, unabhängig davon, m Objekte der zweiten Teilpopulation (bzw. m Objekte, für die das Experiment unter der zweiten Versuchsbedingung durchgeführt wird).
- Dieses Vorgehen bei der Versuchsplanung heißt auch **Randomisierung**: Die Reihenfolge der Versuche und die Zuordnung von Versuchseinheit zu Versuchsbedingung ist zufällig.
- **Beispiel:**
  - Zufällige Zuordnung von 100 Testpatienten zu Gruppe der Größe 60 mit Medikamentenbehandlung und zu anderer Gruppe der Größe 40 mit Placebo-Behandlung.
  - Vergleich der Mietpreise bei 280 zufällig ausgewählten Wohnungen in Köln, und 133 zufällig ausgewählten Wohnungen in Bonn.

Unverbundenes Zweistichprobenproblem

- **Daten:** $x_1, x_2, \ldots, x_n$ und $y_1, y_2, \ldots, y_m$
- **Modellierungsannahme:** Die Daten sind Realisierungen von unabhängigen Zufallsvariablen $X_1, X_2, \ldots, X_n$, bzw. $Y_1, Y_2, \ldots, Y_m$. Die Zufallsvariablen X_i und Y_i sind jeweils identisch verteilt.
- Wieder kann man einen Zweistichproben-t-Test (unter Normalverteilungsannahmen) oder einen Wilcoxon-Test (unter der Annahme, daß sich die Verteilungen nur in der Lage unterscheiden) durchführen.
- Wir beschreiben hier nur den Zweistichproben-t-Test unter **Annahme gleicher Varianzen** im Detail.

- **Voraussetzung - Statistisches Modell:**
  - $X_1, X_2, ..., X_n,\ Y_1, Y_2, ..., Y_m$ sind unabhängige Zufallsvariablen
  - $X_1, X_2, ..., X_n \sim N(m_X, \sigma^2)$ und $Y_1, Y_2, ..., Y_m \sim N(m_Y, \sigma^2)$ sind jeweils identisch normalverteilt mit denselben Varianzen.
- $\mathbf{H}_0 : m_X = m_Y$, $\mathbf{H}_1 : m_X \neq m_Y$ (bzw. $m_X > m_Y$, $m_X < m_Y$)
- **Punktschätzer:**

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{Schätzer für } m_X$$

$$\overline{Y}_m = \frac{1}{m}\sum_{i=1}^{m} Y_i \qquad \text{Schätzer für } m_Y$$

$$S_{pool}^2 = \frac{1}{n+m-2}\left(\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 + \sum_{i=1}^{m}\left(Y_i - \overline{Y}_m\right)^2\right)$$

gepoolter Schätzer für σ^2 (**erwartungstreu !**)

Unverbundenes Zweistichprobenproblem

Zweistichproben-t-Test

- **Teststatistik:**

$$T = \frac{\overline{X}_n - \overline{Y}_m}{S_{pool} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

- Warum wählt man den Nenner so ?
  - $Var\left(\overline{X}_n - \overline{Y}_m\right) = Var\left(\overline{X}_n\right) + Var\left(-\overline{Y}_m\right) = \sigma^2 \cdot \left(\frac{1}{n} + \frac{1}{m}\right)$
  - Der Nenner ist also gerade der Schätzwert für die Standardabweichung des Zählers.

- **Verteilung unter Nullhypothese:**

$$T \sim t(n+m-2) \qquad (\approx N(0,1) \text{ für große } n, m)$$

Unverbundenes Zweistichprobenproblem

- **p-Wert:**

  $$p = P\left[|T| \geq t\right] \quad \text{bzw.} \quad p = P\left[T \geq t\right] \quad \text{bzw.} \quad p = P\left[T \leq t\right]$$

- **Testentscheidung via p-Wert**: H_0 wird verworfen, falls $p \leq \alpha$.

- **Testentscheidung über Quantile:** H_0 wird verworfen, falls

  $$|t| \geq t_{n+m-2,1-\alpha/2} \text{ bzw. } t \geq t_{n+m-2,1-\alpha} \text{ bzw. } t \leq t_{n+m-2,\alpha}$$

- **Zu beachten:** Die Normalverteilungsannahme sollte zum Beispiel mithilfe eines Normal QQ-Plots überprüft werden. Evtl. zusätzlich oder alternativ Wilcoxontest verwenden.

Beispiel. (Schmelzwärme von Eis)

Wiederholte Messungen der freigesetzten Wärme beim Übergang von Eis bei -0.72° C zu Wasser bei 0° C ergaben die folgenden Werte (in cal/g):

Methode A	*79.98*	*80.04*	*80.02*	*80.04*	*80.03*	*80.03*	*80.04*	*79.97*
Methode A	*80.05*	*80.03*	*80.02*	*80.00*	*80.02*			
Methode B	*80.02*	*79.94*	*79.98*	*79.97*	*79.97*	*80.03*	*79.95*	*79.97*

Können die Unterschiede zwischen den Methoden als zufällig angesehen werden, oder ist ein systematischer Unterschied plausibler ?

- $\mathbf{H}_0 : m_X = m_Y , \quad \mathbf{H}_1 : m_X \neq m_Y$
- $\overline{x}_{13} = 80.021 \ , \quad \overline{y}_8 = 79.979 \ , \quad s^2_{pool} = 7.2 \cdot 10^{-4}$
- $\Longrightarrow \qquad t = 3.47$
- Da das Quantil $t_{19,0.975} = 2.09$ unterhalb des beobachteten Werts der Teststatistik liegt, verwerfen wir die Nullhypothese zum Signifikanzniveau $\alpha = 5\%$. Methode A liefert also systematisch höhere Werte als Methode B.

Unverbundenes Zweistichprobenproblem

Weitere Tests

- **Zweistichproben-t-Test bei ungleichen Varianzen:**
  Die Verallgemeinerung des t-Tests auf den Fall ungleicher Varianzen der Zufallsvariablen X_i und Y_i ist in der Literatur zu finden, und in vielen Statistikprogrammen implementiert.

Unverbundenes Zweistichprobenproblem

Weitere Tests

- **Zweistichproben-Wilcoxon-Test (Mann-Whitney-Test):**
  Die Voraussetzungen für den Zweistichproben-Wilcoxon-Test sind wie folgt:
  - $X_1, X_2, ..., X_n,\ Y_1, Y_2, ..., Y_m$ sind unabhängige stetige Zufallsvariablen
  - $X_1, X_2, ..., X_n \sim$ beliebige Verteilungsfunktion $F(x)$
  - $Y_1, Y_2, ..., Y_m \sim F(y - \delta)$
- Mit anderen Worten: Die Verteilung der Y_i ist die um einen Wert δ verschobene Verteilung der X_i.
- Die Berechnung des p-Werts erfolgt wieder mit der gängigen Statistiksoftware.
- Der Wilcoxon-Test ist dem t-Test in vielen Fällen vorzuziehen.

Beispiel. (Einüben von Gehreflexen)

Kann das aktive Einüben von angeborenen Gehreflexen bei Neugeborenen dazu führen, daß das Kind eher laufen lernt ? Dazu wurden 12 männliche Babys zufällig auf zwei gleich große Gruppen aufgeteilt. In der ersten Gruppe wurden die angeborenen Gehreflexe aktiv eingeübt, in der zweiten nicht. Das Laufl ernalter war:

Gruppe 1	*9*	*9.5*	*9.75*	*10*	*13*	*9.5*
Gruppe 2	*11.5*	*12*	*9*	*11.5*	*13.25*	*13*

- $\mathbf{H}_0 : m_X = m_Y , \quad \mathbf{H}_1 : m_X < m_Y$
- Zur Illustration führen wir die Tests diesmal mit der Statistik-Software R durch:

```
> x<-c(9,9.5,9.75,10,13,9.5)
> y<-c(11.5,12,9,11.5,13.25,13)
> t.test(x,y,alternative="less",var.equal=T)

 Two Sample t-test
data: x and y
t = -1.8481, df = 10, p-value = 0.04717
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -0.03049963
sample estimates:
mean of x mean of y
 10.12500 11.70833
```

Wegen $p < 0.05$ *wird die Nullhypothese beim t-Test zum Signifikanzniveau 5% verworfen.*

> *wilcox.test(x,y,alternative="less",correct=F)*

Wilcoxon rank sum test

data: x and y

W = 9, p-value = 0.07334

alternative hypothesis: true location shift is less than 0

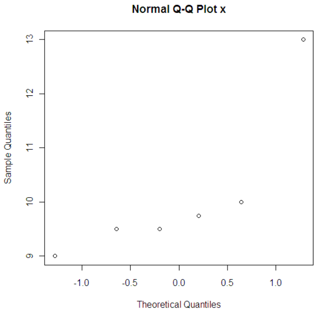
Warning message:

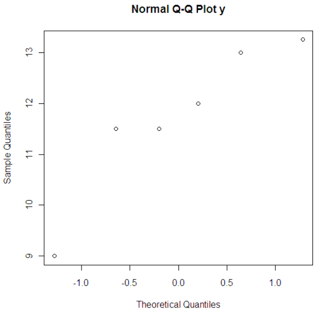
In wilcox.test.default(x, y, alternative = "less", correct = F) :

kann bei Bindungen keinen exakten p-Wert Berechnen

Der p-Wert des Wilcoxon-Tests ist 0.073, sodass die Nullhypothese hier zum Signifikanzniveau 5% nicht verworfen werden kann.

- Um zu sehen, ob wir uns auf den t-Test verlassen können, erstellen wir Normal-QQ-Plots für die beiden Datensätze:
  > *qqnorm(y, main="Normal Q-Q Plot y")*
  > *qqnorm(x, main="Normal Q-Q Plot x")*





- Das sieht nicht nach einer überzeugenden Approximation der Normalverteilung aus.

- *Wir verlassen uns also lieber auf den Wilcoxon-Test, und verwerfen die Nullhypothese nicht !*

8.3. Einfaktorielle Varianzanalyse

ANOVA = Analysis of Variance

Beispiel. (PISA-Studie)

Im Rahmen der PISA-Studie wurde auch der mittlere Zeitaufwand der Schüler für Hausaufgaben in den beteiligten Ländern gemessen. Die folgende Tabelle zeigt die Punktzahlen der Länder im Bereich Mathematische Grundbildung unterteilt nach Ländern mit geringem (Gruppe 1), mittlerem (Gruppe 2), und großem Zeitaufwand für Hausaufgaben:

G1	536	557	514	446	515	510	529	498					
G2	533	520	334	514	490	517	514	533	547	537	499	454	493
G3	447	529	503	457	463	387	470	478	476	488			

Wir wollen herausfinden, ob die Punktzahlen vom Zeitaufwand für Hausaufgaben abhängen, d.h., ob sich die Verteilungen in den drei Gruppen unterscheiden.

- **Daten:** $y_{ij}, \quad 1 \leq i \leq k \ , \quad 1 \leq j \leq n_i \ .$
  Hierbei ist y_{ij} der j-te Beobachtungswert in der i-ten Gruppe. Insgesamt gibt es n_i Beobachtungswerte in der i-ten Gruppe, und k verschiedene Gruppen.
- $n = n_1 + \cdots + n_k$ Gesamtzahl der Beobachtungswerte.
- In unserem Beispiel ist also $k = 3$, $n_1 = 8$, $n_2 = 13$, $n_3 = 10$, und $n = 31$.
- **Modellierungsannahme:** Die Daten sind Realisierungen von unabhängigen normalverteilten Zufallsvariablen Y_{ij} mit **derselben** Varianz σ^2. Die Zufallsvariablen Y_{ij}, $1 \leq j \leq n_i$, aus der i-ten Gruppe sind identisch verteilt mit Mittelwert m_i.

- **Voraussetzung - Statistisches Modell:**
  - Y_{ij} , $1 \leq i \leq k$, $1 \leq j \leq n_i$,
    sind unabhängige Zufallsvariablen
  - $Y_{i1}, Y_{i2}, ..., Y_{in_i} \sim N(m_i, \sigma^2)$
    sind jeweils identisch normalverteilt mit Varianz σ^2.
- $\mathbf{H}_0 : m_1 = m_2 = \cdots = m_k$
- $\mathbf{H}_1 : m_i \neq m_l$ für mindestens ein Paar (i, l)

Grundlegende Statistiken/ Schätzer:

$$\overline{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \qquad \text{Schätzer für } m_i$$

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{k}\sum_{i=1}^{k} n_i \cdot \overline{Y}_i \quad \text{Gesamtmittelwert}$$

$$QS_b = \sum_{i=1}^{k} n_i \cdot \left(\overline{Y}_i - \overline{Y}\right)^2 \qquad \left(= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\overline{Y}_i - \overline{Y}\right)^2 \right)$$

Streuung (Quadratsumme) zwischen (**b**etween) Gruppen

$$QS_w = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_i\right)^2$$

Streuung (Quadratsumme) innerhalb (**w**ithin) der Gruppen

Einfaktorielle Varianzanalyse

Streuungszerlegung

- **Gesamtstreuung:**

$$QS_t = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}\right)^2 = QS_b + QS_w$$

- Die Gesamtstreuung zerfällt in Anteile, die auf die Wirkung des Faktors (Gruppenvariable) zurückzuführen sind, und in Anteile, die durch zufällige Streuungen der Meßwerte innerhalb einer Gruppe entstehen.
- Die Idee ist nun, die Nullhypothese zu verwerfen, wenn der Anteil der Streuung zwischen den Gruppen im Vergleich zur Streuung innerhalb der Gruppen relativ groß ist.

Einfaktorielle Varianzanalyse

Varianzanalyse bei Normalverteilung

- **Teststatistik:**

$$F = \frac{MQS_b}{MQS_w} = \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i \cdot \left(\overline{Y}_i - \overline{Y}\right)^2}{\frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_i\right)^2}$$

- Hierbei stehen im Zähler und Nenner die **mittleren Streuungen**

$$MQS_b = \frac{1}{k-1} \cdot QS_b \quad \text{und} \quad MQS_w = \frac{1}{n-k} \cdot QS_w$$

  zwischen bzw. innerhalb der Gruppen.
- Die Mittelungen wie oben verwendet man, da die Quadratsummen QS_b und QS_w Chiquadrat-Verteilungen mit $k-1$ bzw. $n-k$ Freiheitsgraden haben.

Einfaktorielle Varianzanalyse

Der F-Test

- **Verteilung unter Nullhypothese:**
  Die Statistik F hat unter der Nullhypothese eine **F-Verteilung (Fisher-Verteilung)** mit $k-1$ und $n-k$ Freiheitsgraden:
  $$F \sim F(k-1, n-k)$$
- Die F-Verteilung mit Freiheitsgraden n und m ist die Verteilung des Quotienten $\frac{X/n}{Y/m}$, wobei X und Y unabhängige Chiquadrat verteilte Zufallsvariablen sind. Verteilungsfunktion und Quantile sind tabelliert bzw. mit Statistikprogrammen abrufbar.
- **p-Wert:** Ist f der beobachtete Wert der F-Statistik, dann ist
  $$p = P\left[F \geq f\right] \; = 1 - F_{k-1,n-k}(f) \; ,$$
  wobei $F_{k-1,n-k}$ die Verteilungsfunktion von $F(k-1,n-k)$ ist.
- **Testentscheidung via p-Wert**: H_0 wird verworfen, falls $p \leq \alpha$.

- **Testentscheidung über Quantile:** H_0 wird verworfen, falls

$$f \geq f_{k-1,n-k,1-\alpha} \ ,$$

wobei $f_{k-1,n-k,1-\alpha}$ das $1-\alpha$ Quantil der zugrundeliegenden F-Verteilung ist.

- **Zu beachten:** Die Normalverteilungsannahme sollte zum Beispiel mithilfe eines Normal QQ-Plots überprüft werden. Alternativ Kruskal-Wallis-Test verwenden.

- Ähnlich wie beim Zweistichprobenproblem gibt es auch hier einen Test, der nur voraussetzt, daß die zugrundeliegenden Verteilungen stetig sind, und daß sich die Gruppen nur hinsichtlich der Lage unterscheiden.
- Dieser **Kruskal-Wallis-Test** ist ähnlich zum Wilcoxon-Test aufgebaut.
- Der Test kann wieder mit der gängigen Statistiksoftware durchgeführt werden.

Einfaktorielle Varianzanalyse

Beispiel zum F-Test

- Um den F-Test in übersichtlicher Form durchzuführen, erstellt man eine ANOVA-Tabelle.
- Wir demonstrieren dies im PISA-Beispiel von oben:

Quelle der Variation	QS	Freiheitsgrade	MQS	f
zwischen den Gruppen	9066.03	3-1=2	4533.01	2.238
innerhalb der Gruppen	56720.17	31-3=28	2025.72	
Gesamt	65786.2	31-1=30		

- Das 95%-Quantil der F-Statistik mit 2 und 28 Freiheitsgraden ist 3.34 (aus Tabelle). Also können wir die Nullhypothese zum Signifikanzniveau 0.05 nicht verwerfen.
- Beim Kruskal-Wallis-Test ergibt sich dagegen in diesem Fall eine Ablehnung der Nullhypothese zum Niveau 0.05 !

8.4. Chiquadrattest auf Unabhängigkeit

- **Voraussetzung - Statistisches Modell:**
  - $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ sind unabhängige, identisch verteilte gepaarte Stichproben.
  - Die zugrundeliegenden Merkmale X und Y sind nominalskaliert (z.B. klassierte Daten) mit Merkmalsausprägungen $a_1, \ldots, a_r$ bzw. $b_1, \ldots, b_s$.
- $\mathbf{H}_0$: Die zugrundeliegenden Merkmale X und Y sind unabhängig
- $\mathbf{H}_1$: Die Merkmale X und Y sind nicht unabhängig
- **Teststatistik: Chiquadrat (quadratische Kontingenz)**

$$\chi^2 = \sum_{k,l} \frac{\left(n_{kl} - \widehat{n}_{kl}\right)^2}{\widehat{n}_{kl}} \qquad \text{mit} \qquad \widehat{n}_{kl} = \frac{1}{n} n_k^X n_l^Y \ .$$

- Die Chiquadrat-Statistik erhalten wir aus der Kontingenztabelle der Häufigkeiten der Merkmalsausprägungen, siehe Abschnitt 3 oben.

Chiquadrattest auf Unabhängigkeit

- **Verteilung unter Nullhypothese:** Für große n ist die quadratische Kontingenz näherungsweise Chiquadratverteilt mit $(r-1)\cdot(s-1)$ Freiheitsgraden.

- **Testentscheidung über Quantile:** H_0 wird verworfen, falls

$$\chi^2 \geq \chi^2_{(r-1)(s-1),1-\alpha} \; .$$

Beispiel. (Haar- und Augenfarbe von Statistikstudenten)

n_{kl}	Brown	Blue	Hazel	Green	$\sum$
Black	68	20	15	5	108
Brown	119	84	54	29	286
Red	26	17	14	14	71
Blond	7	94	10	16	127
$\sum$	220	215	93	64	592

$$\chi^2 = 138.3$$

$\widehat{n}_{kl}$	Brown	Blue	Hazel	Green	$\sum$
Black	$\frac{220*108}{592} = 40.1$	$\frac{215*108}{592} = 39.2$	17.0	11.7	108
Brown	$\frac{220*286}{592} = 106.3$	$\frac{215*286}{592} = 103.9$	44.9	30.9	286
Red	$\frac{220*71}{592} = 26.4$	$\frac{215*71}{592} = 25.8$	11.2	7.7	71
Blond	$\frac{220*127}{592} = 47.2$	$\frac{215*127}{592} = 46.1$	20.0	13.7	127
$\sum$	220	215	93	64	592

- In diesem Beispiel ist $r = s = 4$, also

$$(r-1)\cdot(s-1) = 9$$

- Das 99.5% Quantil der Chiquadratverteilung mit 9 Freiheitsgraden ist 23.6.
- Dies liegt weit unterhalb des beobachteten Wertes 138.3 der quadratischen Kontingenz.
- Wir können also erwartungsgemäß die Nullhypothese der Unabhängigkeit von Haar- und Augenfarbe mit ziemlicher Sicherheit verwerfen.