Statistik II für Betriebswirte
Vorlesung 4

Dr. Andreas Wünsche

TU Bergakademie Freiberg
Institut für Stochastik

4. November 2019

5. Varianzanalyse

- "ANOVA" – "**An**alysis **o**f **Va**riance".
- Die Varianzanalyse wurde ursprünglich von Sir R.A. Fisher (1890-1962) für die landwirtschaftliche Versuchstechnik entwickelt und findet heute Anwendung in ganz verschiedenen Gebieten.
- Sie gestattet es, den Einfluss eines qualitativen Merkmales (hier Faktor genannt), auf ein quantitatives oder messbares Merkmal zu untersuchen (Verallgemeinerung des doppelten $t-$Tests zum Vergleich von Mittelwerten zweier unabhängiger Stichproben aus normalverteilten Grundgesamtheiten).
- Unterteilungen:
  - Modell mit festen Effekten, Modell I;
    Modell mit zufälligen Effekten, Modell II;
    Modell mit festen und zufälligen Effekten, Modell III;
  - einfache Klassifikation, einfaktorielle Varianzanalyse;
    zweifache Klassifikation, zweifaktorielle Varianzanalyse; . . . ;
  - eindimensionale oder univariate Varianzanalyse (ANOVA);
    mehrdimensionale oder multivariate Varianzanalyse (MANOVA).

5.1 Einfache Varianzanalyse

- **Frage:** Wie stellt man für ein zufälliges Merkmal X anhand einer Stichprobe fest, ob dieses Merkmal von **einem** Faktor A abhängt, der in mehreren Stufen (Ausprägungen) auftritt ?
- **Antwort:** Man untersucht die Variabilität des Merkmals: Überwiegt die Variabilität zwischen den Gruppen (die durch jeweils eine Stufe des Faktors A erzeugt werden) im Vergleich zu der Variabilität innerhalb der Gruppen, dann ist die Entscheidung für die Ungleichheit der Erwartungswerte und damit für einen Einfluss der Faktorstufe begründet.
- **Beispiele:**
  - X Produktion eines Gutes, A verschiedene Maschinen;
  - X Umsatz einer Firma, A verschiedene Regionen;
  - X Kenngröße für die Wirkung eines Medikaments (einer Behandlung), A verschiedene Medikamente (Behandlungen).

Datenschema

- **Datenschema:**

$j \backslash i$	Gruppen (Stufen)		
	1	$\dots$	p
1	x_{11}	$\dots$	x_{p1}
2	x_{12}	$\dots$	x_{p2}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
	x_{1n_1}	$\dots$	$\vdots$
			x_{pn_p}
n_i	n_1	$\dots$	n_p
$x_{i\bullet}$	$x_{1\bullet}$	$\dots$	$x_{p\bullet}$
$\overline{x}_{i\bullet}$	$\overline{x}_{1\bullet}$	$\dots$	$\overline{x}_{p\bullet}$

- n_i Gruppenumfang, $x_{i\bullet}$ Gruppensumme, $\overline{x}_{i\bullet}$ Gruppenmittel .
- Gilt $n_1 = \dots = n_p$, dann heißt der Versuchsplan balanciert oder orthogonal, ansonsten unbalanciert oder nichtorthogonal .

Beispielaufgabe

- Es soll die Abhängigkeit der Ernteerträge einer bestimmten Getreidesorte von unterschiedlichen Düngemitteln $D_1, \ldots, D_k$ untersucht werden.
- Jedes Düngemittel D_i wird auf n_i gleich großen Feldern angewendet.
- X_{ij} bezeichne den Ertrag (in kg) vom $j-$ten Feld, welches mit dem $i-$ten Düngemittel gedüngt wurde.
- Es soll festgestellt werden, ob ein signifikanter Einfluss des verwendeten Düngemittels auf den Ernteertrag besteht.
- Merkmal X: Ernteertrag.
- Faktor A: Düngemittel mit den Faktorstufen „ohne", D_1 , ..., D_k .

Beispiel 5.1: Düngemittel

	Düngemittel				
	ohne	D_1	D_2	D_3	D_4
x_{ij}	66	60	64	97	90
	68	35	79	99	79
	42	51	72	64	87
	56	69	82	91	71
n_i	4	4	4	4	4
$x_{i\bullet}$	232	215	297	351	327
$\overline{x}_{i\bullet}$	58.00	53.75	74.25	87.75	81.75

(Quelle: J. LEHN, H. WEGMANN: Einführung in die Statistik, B. G. Teubner Verlag, 2006, Beispiel 3.61.)

Allgemeines Modell

$$X_{ij} = \mu_i + \varepsilon_{ij}, \qquad i = 1, \ldots, p, \; j = 1, \ldots, n_i, \quad \text{mit}$$

X_{ij} $j-$te Merkmalszufallsgröße für $i-$te Stufe;

μ_i Erwartungswert des Merkmals auf Stufe i;

ε_{ij} zufälliger Fehler;

Hypothesen:

- H_0 : $\mu_1 = \ldots = \mu_p$.
- H_A : $\mu_i \neq \mu_k$ für mindestens ein Paar $i \neq k$.

Allgemeines Modell (mit Effekten)

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \ldots, p,\ j = 1, \ldots, n_i, \quad \text{mit}$$

μ	allgemeiner Erwartungswert ;
$\alpha_i = \mu_i - \mu$	fester Effekt (systematische Komponente) der Stufe i ;
mit $\sum_{i=1}^{p} n_i \alpha_i = 0$	der Reparametrisierungsbedingung .

Hypothesen:

- H_0 : $\alpha_1 = \ldots = \alpha_p = 0$.
- H_A : $\alpha_i \neq 0$ für mindestens ein i .

Hypothesen und Voraussetzungen für den $F-$Test

- **Hypothese** $H_0: \quad \mu_1 = \ldots = \mu_p \quad (\text{bzw. } \alpha_1 = \ldots = \alpha_p = 0)$.
- **Hypothese** $H_A: \quad \mu_i \neq \mu_k$ für mindestens ein Paar $i \neq k$
  (bzw. $\alpha_i \neq 0$ für mindestens ein i).
- **Voraussetzungen für den $F-$Test:**
  - die Merkmalszufallsgrößen X_{ij} sind unabhängig und normalverteilt mit Erwartungswert μ_i für die $i-$te Stufe jeweils (und damit sind die zufälligen Fehler ε_{ij} unabhängig und normalverteilt mit Erwartungswert 0);
  - die Varianzen der Merkmalszufallsgrößen X_{ij} (und damit der zufälligen Fehler ε_{ij}) sind alle gleich groß,
    $\mathbf{Var} X_{ij} = \mathbf{Var} \varepsilon_{ij} = \sigma^2, \; i = 1, \ldots, p, j = 1, \ldots, n_i$;
  - die Varianz σ^2 der einzelnen Merkmalszufallsgrößen muss nicht bekannt sein.

Hilfsgrößen für den Test

- $\overline{X}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ Mittelwert der $i-$ten Stufe $(i = 1, \dots, p)$.
- $\overline{X}_{\bullet\bullet} = \frac{1}{N} \sum_{i=1}^{p} \sum_{j=1}^{n_i} X_{ij}$ totaler Mittelwert.
- $\mathrm{SSA} = \sum_{i=1}^{p} n_i \left(\overline{X}_{i\bullet} - \overline{X}_{\bullet\bullet}\right)^2$ Summe der Abweichungsquadrate zwischen den Gruppen, charakterisiert die Variabilität zwischen den Stufen (Gruppen), ("**S**um of **S**quares for Factor **A**"), manchmal auch "SST" ("**S**um of **S**quares for **T**reatments") genannt.
- $\mathrm{SSR} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_{i\bullet}\right)^2$ Summe der Abweichungsquadrate innerhalb der Gruppen, charakterisiert die Variabilität innerhalb der Stufen (Gruppen), ("**S**um of **S**quares for **R**esiduals"), auch "SSE" ("**S**um of **S**quares for **E**rrors") genannt.

Testgröße und kritischer Bereich

- $\mathrm{MSA} = \dfrac{\mathrm{SSA}}{p-1}$ ("**M**ean **S**quare for Factor **A**"), auch "MST" ("**M**ean **S**quare for **T**reatments") genannt.
- $\mathrm{MSR} = \dfrac{\mathrm{SSR}}{N-p}$ ("**M**ean **S**quare for **R**esidual"), auch "MSE", ("**M**ean **S**quare for **E**rrors") genannt.
- **Testgröße:** $T = \dfrac{\mathrm{MSA}}{\mathrm{MSR}}$.
- **Kritischer Bereich:** $K = \{t \in \mathbb{R} : t > F_{p-1;N-p;1-\alpha}\}$ mit dem Quantil der $F-$Verteilung mit $(p-1; N-p)$ Freiheitsgraden (FG).
- **Bemerkung:** Es gilt für die Totalvariabilität SST ("**S**um of **S**quares **T**otal") die sogenannte „Streuungszerlegung“:

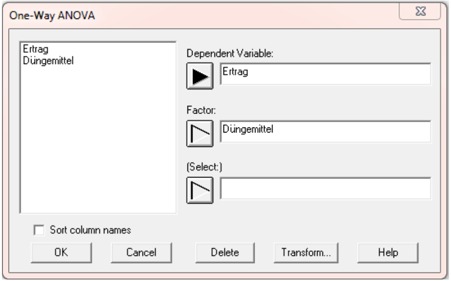
$$\mathrm{SST} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(X_{ij} - \overline{X}_{\bullet\bullet}\right)^2 = \mathrm{SSA} + \mathrm{SSR}\,.$$

ANOVA-Tabelle (ANOVA-Tafel)

Quelle der Variation	Summe der Quadrate	Freiheitsgrade	Mittlere Quadrate	Testgröße
Streuung zwischen der Stufen (Faktor A)	SSA	$p-1$	$\mathrm{MSA} = \frac{\mathrm{SSA}}{p-1}$	$T = \frac{\mathrm{MSA}}{\mathrm{MSR}}$
Streuung innerhalb der Stufen (Rest)	SSR	$N-p$	$\mathrm{MSR} = \frac{\mathrm{SSR}}{N-p}$	
Gesamtstreuung	SST	$N-1$		

ANOVA in Statgraphics für Beispiel 5.1 Ernteerträge

Compare → Analysis of Variance → One-Way ANOVA...



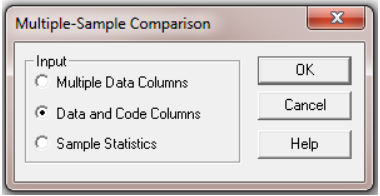
ANOVA Table for Ertrag by Düngemittel

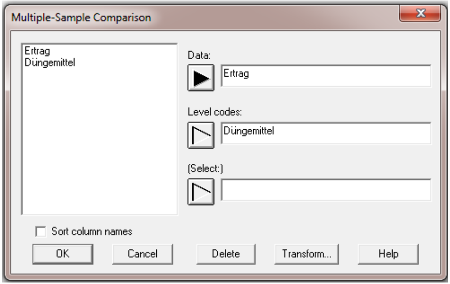
Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Between groups	3492,8	4	873,2	5,81	0,0050
Within groups	2253,0	15	150,2		
Total (Corr.)	5745,8	19			

ANOVA in Statgraphics (Zweite Möglichkeit)

Compare → Multiple-Samples → Multiple-Sample Comparsion...

Wenn sich die Daten für jede Gruppe in einer separaten Spalte befinden, gelangt man nur über diesen Weg zur ANOVA-Tabelle. Im Beispiel 5.1 liegen die Daten aber in einer Spalte (Ertrag) vor. Daneben gibt es die Spalte Düngemittel. In dieser ist angegeben, zu welchem Düngemittel (Level Codes) der Ertrag gehört.





Der Kruskal-Wallis-Test

- Sind die Merkmalszufallsgrößen nicht normalverteilt, dann kann man mit dem Kruskal-Wallis-Test (auch $H-$Test) die Gleichheit der Mediane (bzw. die Gleichheit der Verteilungsfunktionen) der Merkmalszufallsgrößen zu den einzelnen Stufen des Faktors A überprüfen.
- Der Kruskal-Wallis-Test ist eine Verallgemeinerung des Wilcoxon-Rangsummentests auf den Fall von mehr als 2 unabhängigen Stichproben.
- **Voraussetzung:** die Merkmalszufallsgrößen haben eine stetige Verteilung.
- **Bezeichnungen:**
  - p Anzahl der Stufen (Gruppen);
  - n_i Anzahl der Beobachtungswerte in der Stufe i, $i = 1, \ldots, p$;
  - N Gesamtanzahl der Beobachtungswerte, $N = \sum_{i=1}^{p} n_i$.

Vorgehen beim Kruskal-Wallis-Test

- In der gemeinsamen Stichprobe (alle p Gruppen, Stichproben) werden die Ränge bestimmt.

  r_{ij} sei die Rangzahl der $j-$ten Beobachtung in der $i-$ten Stufe, $i = 1, \ldots, p\,,\; j = 1, \ldots, n_i\,.$

- Bei der Bestimmung der Rangzahlen achtet man auf möglicherweise auftretende Bindungen (mehrfach auftretende Werte in der gemeinsamen Stichprobe).
  Es bezeichne
  g die Anzahl der auftretenden Bindungen;
  t_h die Anzahl der übereinstimmenden Beobachtungswerte in der $h-$ten Bindung, $h = 1, \ldots, g\,.$

- Man berechnet die Summe der Ränge in der Stufe i für alle $i = 1, \ldots, p\,,$ diese wird mit $r_{i\bullet}$ bezeichnet.

Testgröße und kritischer Bereich beim Kruskal-Wallis-Test

- **Testgröße:**

$$T = \frac{1}{B}\left[\frac{12}{N(N+1)}\sum_{i=1}^{p}\frac{1}{n_i}r_{i\bullet}^2 - 3(N+1)\right]$$

mit $$B = 1 - \frac{1}{N^3 - N}\sum_{h=1}^{g}(t_h^3 - t_h)\,,$$

  dabei ist $\frac{1}{B}$ ein Korrekturfaktor, falls Bindungen vorkommen, kommen keine Bindungen vor, setzt man $B = 1$.

- **Kritischer Bereich:** $K = \{t \in \mathbb{R} : \; t > \chi^2_{p-1;1-\alpha}\}$.
- Der angegebene kritische Bereich beruht wieder auf einer asymptotischen Verteilung, er gilt nur näherungsweise.

  Als Faustregel kann man mit ihm rechnen, falls alle $n_i > 5$ sind (für $p = 3$ sollte allerdings mindestens ein $n_i > 8$ sein).

Beispiel 5.2: Zugfestigkeit von 3 Drahtsorten

	Drahtsorte					
	1		2		3	
$j \backslash i$	x_{1j}	$\mathbf{r}_{1j}$	x_{2j}	$\mathbf{r}_{2j}$	x_{3j}	$\mathbf{r}_{3j}$
1	9.0	**7**	7.3	**4.5**	18.0	**19**
2	15.4	**16**	15.6	**17**	9.6	**8**
3	8.2	**6**	14.2	**13**	11.5	**11**
4	3.9	**2**	13.0	**12**	19.4	**21**
5	7.3	**4.5**	6.8	**3**	17.1	**18**
6	10.8	**10**	9.7	**9**	14.4	**15**
7	3.8	**1**			19.4	**21**
8					19.4	**21**
9					14.3	**14**
$\mathbf{r}_{i\bullet}$		**46.5**		**58.5**		**148.0**

Quelle: nach J. Hartung, Statistik: Oldenbourg Verlag, 2009, Kap. XI, Abschnitt 1.1.A .

Kruskal-Wallis-Test für Beispiel 5.2

- $B = 1 - \frac{1}{22^3 - 22}((2^3 - 2) + (3^3 - 3)) = 0.9971767\,.$
- **Wert der Testgröße:**

$$t = \frac{1}{0.9971767}\left[\frac{12}{22 \cdot 23} \cdot \left(\frac{46.5^2}{7} + \frac{58.5^2}{6} + \frac{148^2}{9}\right) - 3 \cdot 23\right] = 9.597\,.$$

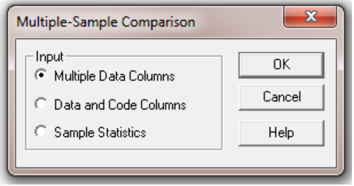
- **Kritischer Bereich** $(\alpha = 0.05)$:

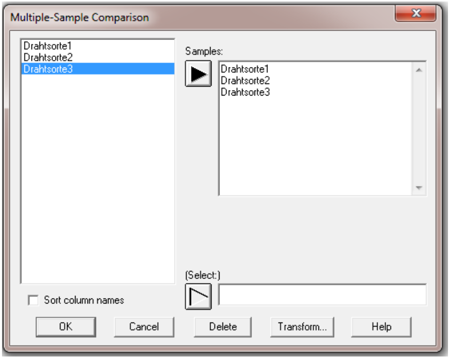
$$K = \left(\chi^2_{2;0.95}; +\infty\right) = (5.99; +\infty)\ .$$

- **Testentscheidung:** $t \in K$, die Nullhypothese über die Gleichheit der Erwartungswerte wird abgelehnt.
- **Testergebnis:** Die drei Drahtsorten unterscheiden sich beim Signifikanzniveau von 5% hinsichtlich der (erwarteten) Zugfestigkeit signifikant voneinander.

Statgraphics für Beispiel 5.2

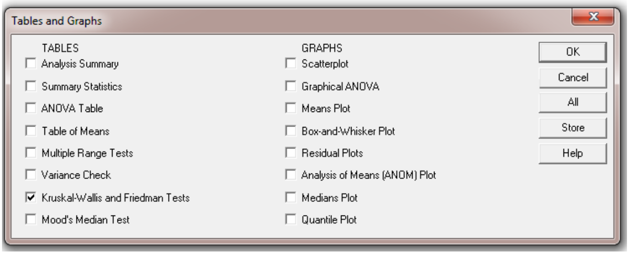
Compare → Multiple-Samples → Multiple-Sample Comparsion...





Hier befinden sich die Daten in der Statgraphics-Datendatei für jede Gruppe (Drahtsorte1,. . . ,Drahtsorte3) in einer separaten Spalte.

Statgraphics-Ausgabe für Beispiel 5.2



Statgraphics-Ausgabe:

Kruskal-Wallis Test

	Sample Size	*Average Rank*
Drahtsorte1	7	6,64286
Drahtsorte2	6	9,75
Drahtsorte3	9	16,4444

Test statistic = 9,59735   P-Value = 0,00824067

Paarweise Tests

- Sind die Mittelwerte von p unabhängigen Stichproben signifikant unterschiedlich (global) und möchte man zusätzlich herausfinden, welche Mittelwerte paarweise verschieden sind, bedient man sich der paarweisen Vergleiche.
- Die bekanntesten Verfahren für normalverteilte Stichproben mit übereinstimmender Varianz sind der Scheffé-Test und der Tukey-Test.
- Der Scheffé-Test ist flexibler, der Tukey-Test besitzt eine höhere Güte.
- **Statgraphics:** Compare $\rightarrow$ Analysis of Variance $\rightarrow$ One-Way ANOVA... , dann im Auswahlfenster für Tables and Graphs unter TABLES Multiple Range Tests aktivieren;

  durch Rechtsklick im Ergebnisfenster für die Multiple Range Tests $\rightarrow$ Pane-Options... kann man das Testverfahren auswählen.

Signifikanzniveau bei paarweisen Vergleichen

- Für die paarweisen Tests gibt es 2 mögliche Bedeutungen für das Signifikanzniveau α:
  - **globales Niveau:** die Wahrscheinlichkeit, mindestens eine wahre Hypothese $H_0^{kl} : \mu_k = \mu_l$ unter der Voraussetzung abzulehnen, dass die Globalnullhypothese richtig ist (d.h. alle Einzelnullhypothesen wahr sind), ist $\leq \alpha$;
  - **multiples Niveau:** die Wahrscheinlichkeit, mindestens eine wahre Hypothese H_0^{kl} unabhängig davon abzulehnen, wieviele und welche der Einzelnullhypothesen wahr sind, ist $\leq \alpha$.
- Sowohl der Scheffé-Test als auch der Tukey-Test halten multiples Niveau, allerdings ist der Scheffé-Test konservativer, d.h. er erkennt weniger Unterschiede als signifikant; deshalb ist der Tukey-Test vorzuziehen.

Weitere Bemerkungen

- Zur Überprüfung der Annahme über die Gleichheit der p Varianzen kann unter anderem der Bartlett-Test verwendet werden (bei ungleichen Stichprobenumfängen n_i und normalverteilten Grundgesamtheiten; siehe Literatur, z.B. STORM, Abschnitt 14.2.5 oder HARTUNG, Kap. XI, Abschnitt 1.3.).
- Ein weiterer Test für dieses Problem ist z.B. der Levene-Test.
- **Statgraphics:** Compare → Analysis of Variance → One-Way ANOVA... , dann im Auswahlfenster für Tables and Graphs unter TABLES Variance Check aktivieren;

  durch Rechtsklick im Ergebnisfenster für die Variance Check → Pane-Options.... kann man das Testverfahren auswählen.