Ao.Prof.DI.Dr Josef Haas
josef.haas@medunigraz.at

Statistik – Methoden

ANOVA (Analysis of Variance) Varianzanalyse

Vergleich von Mittelwerten

Ao.Univ.Prof.DI.Dr. Josef Haas
josef.haas@medunigraz.at

Ausgangssituation

<u>Zwei Variable:</u> 1 Kategorial/Nominal, 1 Quantitativ

<u>Frage:</u> Gibt es einen Unterschied zwischen den Gruppen (=Ausprägung der kategorialen Variablen) bezüglich der quantitativen Variablen, sind also die Mittelwerte unterschiedlich?

Kategoriale Variable mit genau zwei Werten (also 2 Gruppen bzw. 2 Faktorstufen)
→ 2-Stichproben t-Test, Welch-Test

Mehr als zwei Gruppen →ANOVA

Anm. Faktorstufen bzw. Subpopulationen werden hier als Gruppen bezeichnet

ANOVA

	Anzahl der unabhängigen Variable 1	größer 1
Anzahl d. abhängigen Variablen 1	einfaktorielle Varianzanalyse	mehrfaktorielle Varianzanalyse
größer 1	einfaktorielle multivariate Varianzanalyse (MANOVA)	mehrfaktorielle multivariate Varanzanalyse (MANOVA)

Beispiel

Die Anzahl der defekten Teile bei 25 Lieferungen von Montageteilen von 3 Lieferanten wird geprüft:

Lieferanten = Gruppen: A, B, C

Messungen: # defekte Teile pro 1000

Daten	Mittelwerte]:
A: 5,6,6,7,7,8,9,10	[7.25]
B: 7,7,8,9,9,10,10,11	[8.875]
C: 7,9,9,10,10,10,11,12,13	10.11]

Sind die Unterschiede signifikant?

(Gibt es Unterschiede zwischen den Lieferanten bezüglich Produktqualität?)

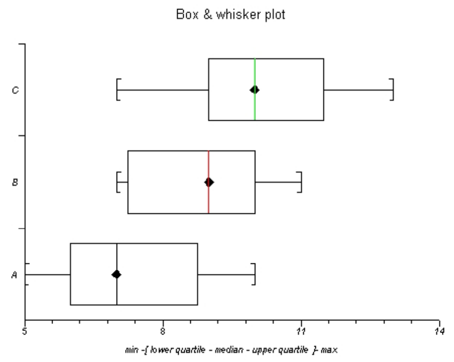
Deskriptive Analyse

Grafische Analyse:
- Boxplots
- multiple Histogramme

Eine mögliche Signifikanz (=Unterschied zwischen den Gruppen) hängt ab von
- den Unterschieden der Mittelwerte
- der Standardabweichung jeder Gruppe
- den Stichprobengrößen

ANOVA berechnet den p-Wert der F-Statistik

Boxplots



Was macht die ANOVA?

Im einfachsten Fall testet die ANOVA folgende Hypothesen:

H_0: Die Mittelwerte aller Gruppen sind gleich.

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

H_a: Nicht alle Mittelwerte sind gleich

H_a: $\mu_i \neq \mu_j$ für irgendein i, j

- Es gibt keine Aussage wo der Unterschied liegt.
- Mit Kontrasten können vorher festgelegte geplante Vergleiche durchgeführt werden
- “Multiple Vergleiche” (Post-Hoc-Tests) können nachgeschaltet werden

Voraussetzungen der ANOVA

- Normalverteilte Werte in jeder Gruppe
  - Überprüfen mit Q-Q-Plots, Histogrammen, Shapiro-Wilks-Test,….
  - robust bei geringen Abweichungen von dieser Annahme, problematisch bei Ausreißern
- Die Standardabweichungen jeder Gruppe sind annähernd gleich (Homoskedastizität)
  - Faustregel: Verhältnis größter zu kleinster Standardabweichung muss kleiner 2:1 sein
- Die Messwerte bzw. Faktorstufen sind voneinander unabhängig

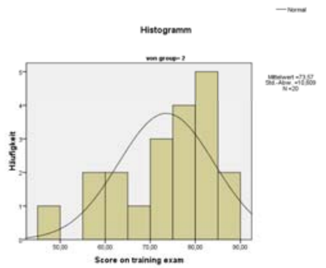
Normality Check

Die Voraussetzung der Normalverteilung können überprüft werden mit

- Annahmen (Vorwissen) über die Population
- Histogramme für jede Faktorstufe
- Quantil-Quantil-Plots für jede Gruppe

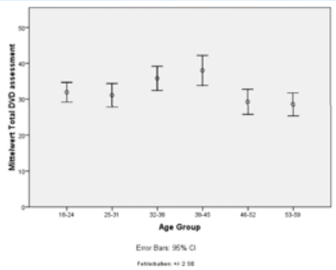
Anm.: Bei kleinen Gruppengrößen gibt es kaum brauchbare Überprüfungen; häufig wird einfach angenommen, dass physikalische Messungen tendenziell normalverteilt sind. Bei Abweichungen von den Voraussetzungen können je nach Sachlage Transformationen hilfreich sein bzw. das Framework der GLM (Generalized Linear Model), poissonverteilte Fehlermodelle,…

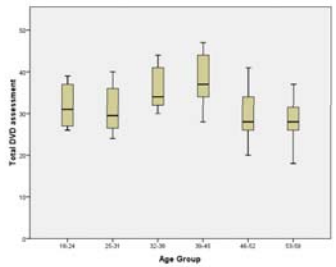
Histogramm (mit Normalverteilungskurve) & Q-Q-Plot



Die hier gezeigte Verteilung ist nicht symmetrisch und weicht von einer Normalverteilung ab. Beim Quantil-Quantil-Plot werden die beobachteten Werte mit Quantilen einer theoretischen Verteilung (hier: Normalverteilung) verglichen. Bei Übereinstimmung liegen die Punkte auf einer Geraden, schiefe Verteilungen und Ausreißer können rasch erkannt werden

Error Bar & Boxplot





Beim Fehlerbalkendiagramm werden Mittelwerte mit einem Vertrauensbereich (hier: +/- 2 Standardabweichungen) angezeigt. In beiden Diagrammen ist ersichtlich, dass die beiden mittleren Altersgruppen höhere Werte aufweisen. Beim Boxplot ist die Variabilität der Daten besser ersichtlich, beim Fehlerbalken die Variabilität der Mittelwerte

Standard Deviation Check

Variable	Lieferant	N	Mean	Median	StDev
Defekte	A	8	7.250	7.000	1.669
	B	8	8.875	9.000	1.458
	P	9	10.111	10.000	1.764

Vergleich der kleinsten und größten Standardabweichung:

- größte: 1.764
- kleinste: 1.458
- 1.458 x 2 = 2.916 > 1.764

Anm: ein StAbw-Verhältnis von 2:1 entspricht einem Verhältnis der Varianzen von 4:1

Notation ANOVA

n = Anzahl der Objekte (Personen, Messungen) – alle Daten
k = Anzahl der Gruppen (Faktorstufen)
$\bar{x}$ = Mittelwert aller Daten

Gruppe i
n_i = Anzahl der Objekte in Gruppe i
x_{ij} = Wert des *j-ten Objekts in Gruppe i*
$\bar{x}_{i.}$ = Mittelwert der Gruppe i
$s_{i.}$ = Standardabweichung der Gruppe i

$i = 1, \cdots k$
$j = 1, \cdots n_i$

Wie ANOVA funktioniert

ANOVA misst zwei Komponenten der Variation in den Daten und vergleicht deren relative Größe. Die Summe der Abweichungsquadrate *(SS=sum of squares)* der Stichproben um das Gesamtmittel lässt sich zerlegen in die

- Variation INNERHALB der Gruppen
  - für jeden Wert betrachtet man den Unterschied zwischen Einzelwert und Gruppenmittel
- Variation ZWISCHEN den Gruppen
  - für jeden Wert betrachtet man den Unterschied zwischen Gruppenmittel und Gesamtmittel
  - Letzte Zeile: beschreibt die Anzahl der Freiheitsgrade (df=degrees of freedom)

$$\begin{array}{ccccc} SS_{total} & = & SS_{within} & + & SS_{between} \\ \sum_{i,j}(x_{ij}-\bar{x})^2 & = & \sum_{i,j}(x_{ij}-\bar{x}_{i.})^2 & + & \sum_i n_i(\bar{x}_{i.}-\bar{x})^2 \\ (n-1) & = & (n-k) & + & (k-1) \end{array}$$

Wie ANOVA funktioniert

Aus den Quadratsummen berechnet man die um die Freiheitsgrade normierten mittleren Quadratsummen *MS*. Stammen die Gruppen aus derselben Grundgesamtheit, dann sollten diese Varianzen etwa gleich groß sein.

$$\begin{aligned} MS_{total} &= SS_{total}/n-1 \\ MS_{between} &= SS_{between}/k-1 \\ MS_{within} &= SS_{within}/n-k \end{aligned}$$

Bei der Hypothesenprüfung wird der Quotient $F_{emp} = MS_{between}/MS_{within}$ mit dem Quantil der F-Verteilung verglichen (mit df k-1 und n-k). Große F-Werte *(dh. mehr Unterschied **zwischen** als **innerhalb** der Gruppen)* führen zur Ablehnung der Nullhypothese

Statsdirect - Output

One way analysis of variance

Variables: A, B, C

Source of Variation	Sum Squares	DF	Mean Square
Between Groups	34,74	2	17,37
Within Groups	59,26	22	2,69
Corrected Total	94	24	

F (variance ratio) = 6,447387 **P = 0,0063**

SPSS - Output

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert Untergrenze	95%-Konfidenzintervall für den Mittelwert Obergrenze	Minimum	Maximum
A	8	7,25	1,669	,590	5,85	8,65	5	10
B	8	8,88	1,458	,515	7,66	10,09	7	11
C	9	10,11	1,764	,588	8,76	11,47	7	13
Gesamt	25	8,80	1,979	,396	7,98	9,62	5	13

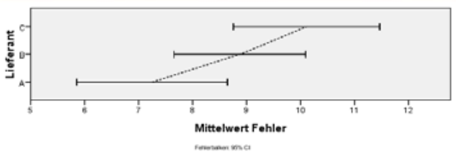
ANOVA

Test auf Homogenität der Varianzen

Levene-Statistik	df1	df2	Signifikanz
,054	2	22	,947

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	34,736	2	17,368	6,447	,0063
Innerhalb der Gruppen	59,264	22	2,694		
Gesamt	94,000	24			

Mittelwerte mit Konfidenzintervallen



Die Konfidenzbereiche von A und B überlappen sich, ebenso von B und C, nicht aber von A und C → A und C sind unterschiedlich

Multiple Vergleiche

Falls ANOVA Unterschiede zwischen den Gruppen aufzeigt, können wir durch (prinzipiell) die Art der Unterschiede durch mehrere 2-Stichproben-t-Tests weiter untersuchen

- Wir müssen dazu den p-Wert (bzw. unsere Signifikanzgrenze anpassen (multiples Testen).
- Je nach Fragestellung gibt es dazu verschieden Methoden.
- Falls wir nur am Unterschied zwischen zwei Gruppen interessiert sind, sollten wir den Untersuchungsaufbau schon vorher darauf ausrichten.

Tuckey's Pairwise Comparisons

```
Tukey's pairwise comparisons

    Family error rate = 0.0500
Individual error rate = 0.0199

Critical value = 3.55

Intervals for (column level mean) - (row level mean)

                 A           B

       B      -3.685
               0.435

       C      -4.863      -3.238
              -0.859       0.766
```

95% Konfidenz

alpha = 0.0199 für jeden Test.

Konf.Int (Sicherheit 98.01%) für jeden paarweisen Vergleich

Nur A vs C ist significant

98% CI für A-C beträgt (-0.86,-4.86)

Weiterführende Methoden

- Zweifaktorielle ANOVA (die Varianz wird zwei Faktoren und Wechselwirkung zugeschrieben)
- Mehrfaktorielle ANOVA
- ANOVA mit Messwiederholungen (Repeated Measurement – mehrere Messungen am gleichen Objekt, häufig ein zeitlicher Verlauf)
- GLM
- Nichtparametrische Verfahren: Kruskal-Wallis, Friedberg, Nemenyi