Statistik – Methoden zum Mittelwertvergleich von mehreren Gruppen

Mehrfachvergleichsverfahren

Mehrfachvergleichsverfahren

- Bei der Betrachtung der einfaktoriellen Varianzanalyse haben wir festgestellt, das diese zwei mögliche Ausgänge hat: Verbleib bei der Nullhypothese bzw. Wechsel zur Alternativhypothese

$\boldsymbol{H_0}$ Die Mittelwerte der einzelnen Gruppen unterscheiden sich nicht

$\boldsymbol{H_1}$ Es gibt mindestens einen Unterschied im Mittelwert bei einer Gruppe

- Wechseln wir zur Alternativhypothese, wissen wir das es signifikante Unterschiede gibt

Mehrfachvergleichsverfahren

- Wir sind mit den Ergebnissen der Varianzanalyse aber nicht in der Lage, die Unterschiede genau zu lokalisieren oder sie in ihrer Größe zu beschreiben

- Wir erkennen nicht einmal, ob nur einen oder mehrere Unterschiede gibt

- Wir benötigen Verfahren, mit denen wir nach einer Varianzanalyse mit H_1-Ausgang, die uns dabei unterstützen

- Diese Verfahren werden im Allgemeinen als Post hoc-Verfahren bezeichnet

Mehrfachvergleichsverfahren

- Nachfolgend zur Varianzanalyse können nun alle Stichproben paarweise mit einem 2t-Test verglichen werden
- Vergleichbares Testen ist auch als nicht-parametrisches Verfahren möglich

Mehrfachvergleichsverfahren

Multiples Testen

- Durchführung mehrerer statistischer Tests auf dieselben oder ähnliche Daten
  - Beispiel: Paarweise Mittelwertvergleiche nach einer ANOVA oder einem Kruskal-Wallis-Test, gleichzeitige Überprüfung der Normalverteilungsannahme bei mehreren Datensätzen

- Überprüfung von Hypothesen mit verschiedenen Verfahren
  - Beispiel: Überprüfung der Normalverteilungsannahme mit verschiedenen Verfahren

Mehrfachvergleichsverfahren

Problem des Multiplen Testens

Nehmen wir an, unser Signifikanzniveau beträgt $\alpha = 5\%$

Führen wir nun multiple Tests durch, so werden sich für jeden einzelnen Test Sicherheit bzw. Unsicherheit nicht verändern

Bei einer großen Anzahl von Tests erhöht sich nun aber die Gefahr, dass ein Test falsch interpretiert wird

Mehrfachvergleichsverfahren

Problem des Multiplen Testens

Für zwei aufeinanderfolgende Tests sinkt zum Beispiel die Sicherheit von $1 - \alpha = 95\%$ auf $(1 - \alpha)^2 = 90{,}25\%$

Im folgenden steigt die Wahrscheinlichkeit einer Fehlentscheidung

Mehrfachvergleichsverfahren

Abhilfe: Einsatz von Korrekturverfahren, die vielfach auf Anpassung des Signifikanzniveaus basieren
Beispiele: Bonferroni, Holm, Tukey

Des weiteren werden auch grafische Vergleiche angestellt, die aber keine Anpassung des Signifikanzniveaus beinhalten, die Anpassung muss also manuell erfolgen

Grafischer Vergleich

- Der einfachste Weg eines Mehrfachvergleiches ist die gemeinsame grafische Darstellung der Konfidenzintervalle der einzelnen Gruppenmittelwerte

- Gibt es Gruppen, deren Konfidenzintervalle nicht mit anderen überlappen, hat man den Unterschied schon gefunden, er ist dann auch signifikant

Grafischer Vergleich

Beispiel_Mehrfach.xlsx

```
Rcmdr>  summary(AnovaModel.4)
            Df Sum Sq Mean Sq F value  Pr(>F)
Index        2   9.96   4.979   6.217 0.00299 **
Residuals   87  69.67   0.801
```

- Ausgabe der einfachen Varianzanalyse
- p-Wert < α = 5%: Alternativhypothese

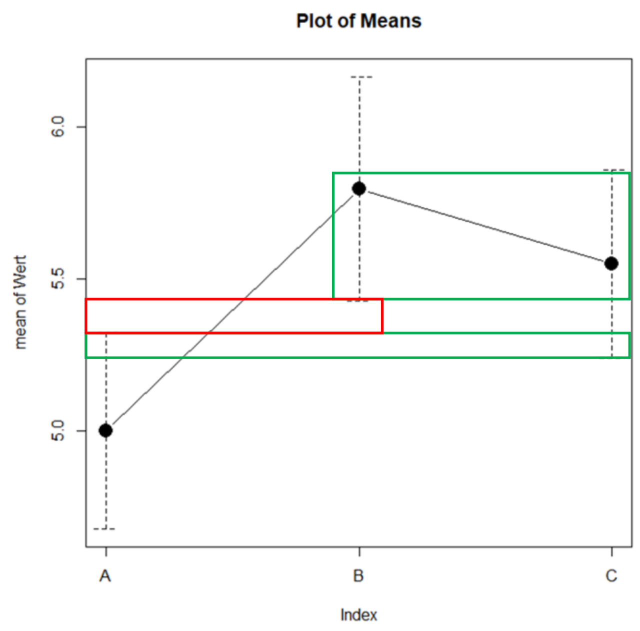
Grafischer Vergleich

- Keine Überschneidung des Konfidenzintervalls A mit den anderen Intervallen

- Der Mittelwert A unterscheidet sich von den anderen Mittelwerten

Überschneidungen

Keine Überschneidung



Anzahl der Vergleiche

Anzahl der Vergleiche am Beispiel der ANOVA

- Ist p die Anzahl der zu vergleichenden Stichproben, bestimmt sich die Anzahl der Vergleiche zu

$$m = \frac{p(p-1)}{2}$$

- Für 3 Stichproben ergeben sich somit 3 Vergleiche, für 5 Stichproben sind dies schon 10 Vergleiche

Bonferroni

- Multiple 2t-Tests mit Bonferroni-Korrektur der Irrtumswahrscheinlichkeit

- Würden wir Mehrfachvergleiche nur paarweise über den t-Test durchführen, würden immer wieder die selben Gruppen gegeneinander getestet

- Das würde zu einem Anstieg des α-Fehlers führen

- Die Bonferroni-Korrektur passt deshalb den α-Wert an

Bonferroni

Korrektur des α-Wertes

- $\propto_{korr} = \frac{1}{m} \propto = \frac{\propto}{Anzahl\ der\ Vergleiche}$
- Bei 3 Stichproben / Vergleichen und einem nicht-korrigierten $\alpha = 0{,}05$ kann mit $\alpha_{korr} = \frac{\alpha}{3} = 0{,}0167$ gerechnet werden, für 5 Stichproben sinkt der Wert auf nur noch $\alpha_{korr} = 0{,}005$

Bonferroni

Alternativ lässt sich auch der p-Wert anpassen

$$p_{korr} = m * p$$

und mit dem ursprünglichen α vergleichen

Bonferroni

- Mit der korrigierten Irrtumswahrscheinlichkeit können jetzt paarweise normale 2t-Tests durchgeführt werden

- Die Umsetzung der Bonferroni-Korrektur ist sehr einfach umzusetzen

- Die Korrektur nach Bonferroni gilt als sehr konservativ, die einzelnen Tests verbleiben sehr lange in der Nullhypothese

- Testung mit korrigiertem α_{korr} führt zu einer Verringerung der Teststärke

Bonferroni

Beispiel Bonferroni-Korrektur für Beispiel_Mehrfach.xlsx

- $\propto_{korr} = \frac{\alpha}{m} = \frac{0{,}05}{3} = 0{,}0167$ ($m = 3$, d.h. 3 Vergleiche)

```
data:  AB by IAB
t = -3.3322, df = 58, p-value = 0.001504 < 0,0167

data:  AC by IAC
t = -2.5132, df = 58, p-value = 0.01476 < 0,0167

data:  BC by IBC
t = 1.0497, df = 58, p-value = 0.2982 > 0,0167
```

Die Vergleiche AB und AC wechseln zur Alternativhypothese

Holm (auch Bonferroni-Holm)

- Weniger konservative Anpassung der Bonferroni-Korrektur
- Höhere Teststärke als der reine Bonferroni
- Sukzessive Anpassung der α-Korrektur

Holm (auch Bonferroni-Holm)

Vorgehen:

- Ordnen der p-Werte von klein nach groß
- Für $j = 1,2, ...$ gilt $\alpha_{korr} = \frac{\alpha}{m-j+1}$
- Abbruch, sobald der erste nicht-signifikante p-Wert erreicht wird, alle weiteren Vergleiche verbleiben in der Nullhypothese
- Nur der erste Vergleich nutzt die ursprüngliche Korrektur nach Bonferroni

Tukey Contrast

- Findet vielfach Anwendung in der post-hoc-Analyse von ANOVA-Rechnungen
- Vergleich aller möglichen Gruppenpaare und Bestimmung von Kontrastwerten zur Feststellung von statistisch signifikanten Unterschieden zwischen den Gruppen
- Berücksichtigung des Problems von multiplem Testen
- Robust gegenüber Verletzungen der Normalverteilungsannahme, solang die Stichproben groß sind

Korrektur in R

- Paarweise Vergleiche inklusive dazugehöriger Korrektur können unmittelbar im RCommander innerhalb der einfaktoriellen ANOVA angewählt werden.

- Andere Korrekturen müssen über RStudio durchgeführt werden

Korrektur in R

Beispiel

```
pairwise.t.test    (x, g, p.adjust.method = p.adjust.methods,
                    pool.sd = !paired, paired = FALSE,
                    alternative = c("two.sided", "less",
                    "greater"), ...)
```

- Unter p.adjust.methods stehen verschiedene Korrekturmethoden zur Verfügung (z.B. Holm, Bonferroni, Benjamini & Hochberg, u.a.).
- pairwise.wilcox.test als nicht-parametrische Variante

Korrektur in R

Eingabe in RStudio:

Beispiel: Die Daten aus Beispiel_Mehrfach.xlsx als Dataset nach R geladen und gerechnet

```
> with(Dataset,pairwise.t.test(Wert,Index,p.adjust.method =
"bonferroni",paired=FALSE))

         Pairwise comparisons using t tests with pooled SD
data:  Wert and Index
  A      B
B 0.0026 -
C 0.0592 0.8634
P value adjustment method: bonferroni
```

Der Vergleich AB mündet in der Alternativhypothese (bei α=5%), AC und CB verbleiben in der Nullhypothese