Einfaktorielle Varianzanalyse (ANOVA)

Ozgur Polat und Kathrin Einzmann

09 01 2020

## Datensatz:

Var 1 = total\_bill (abhängig, Rechnung ist intervallskaliert)

Var 2 = day (unabhängige Variable „Tag“ ist kategorial)

# 1. Hypothese:

H1: An den unterschiedlichen Wochentagen ändern sich die Höhe der Rechnung.

H0: An den unterschiedlichen Wochentagen ändern sich die Höhe der Rechnung nicht.

# 2. Voraussetzungen für die einfaktoriellen Varianzanalyse

• Die abhängige Variable ist intervallskaliert.

• Die unabhängigen Variable (Faktoren) sind kategorial (nominal- oder ordinalskaliert)

• Die durch den Faktor gebildeten Gruppen sind unabhängig

• Die abhängige Variable ist normalverteilt innerhalb jeder der Gruppen. Ab 25 Probanden pro Gruppe sind Verletzungen dieser Voraussetzung unproblematisch • Homogenität der Varianzen: Die Gruppen stammen aus Grundgesamtheiten mit annähernd identischen Varianzen der abhängigen Variablen (siehe Levene-Test)

# 3. Grundlegende Konzepte: Was ist die einfaktoriellen Varianzanalyse?

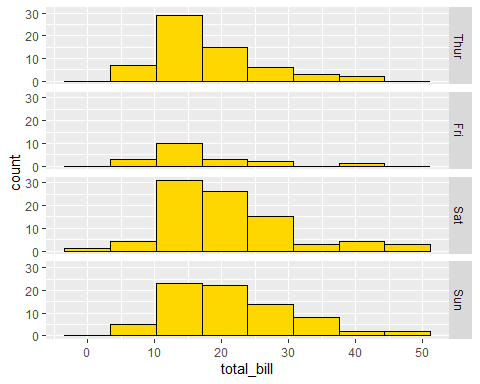
Die einfaktorielle Varianzanalyse – auch “einfaktorielle ANOVA”, da in Englisch “Analysis of Variance” – testet, ob sich die Mittelwerte mehrerer unabhängiger Gruppen (oder Stichproben) unterscheiden, die durch eine kategoriale unabhängige Variable definiert werden. Diese kategoriale unabhängige Variable wird im Kontext der Varianzanalyse als “Faktor” bezeichnet. Entsprechend werden die Ausprägungen der unabhängigen Variable “Faktorstufen” genannt, wobei auch der Begriff der “Treatments” gebräuchlich ist. Als “einfaktoriell” wird eine Varianzanalyse bezeichnet, wenn sie lediglich einen Faktor, also eine Gruppierungsvariable, verwendet mehrfaktorielle Varianzanalyse).

Das Prinzip der Varianzanalyse besteht in der Zerlegung der Varianz der abhängigen Variable. Die Gesamtvarianz setzt sich aus der sogenannten “Varianz innerhalb der Gruppen” und der “Varianz zwischen den Gruppen” zusammen. Diese beiden Anteile werden im Rahmen einer Varianzanalyse miteinander verglichen. Die einfaktorielle ANOVA stellt eine Verallgemeinerung des t-Tests für unabhängige Stichproben für Vergleich von mehr als zwei Gruppen (oder Stichproben) dar. Die Fragestellung der einfaktoriellen Varianzanalyse wird oft so verkürzt: “Unterscheiden sich die Mittelwerte einer unabhängigen Variable zwischen mehreren Gruppen? Welche Faktorstufen unterscheiden sich?”

# 4. Deskriptive Statistiken

### Normalverteilung

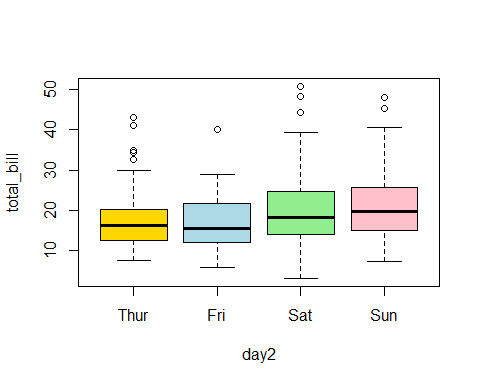
library(reshape2)  
library(ggplot2)  
  
TIPS$day2<-factor(TIPS$day,levels=levels(TIPS$day)[c(4,1,2,3)])  
  
ggplot(TIPS, aes(x = total\_bill)) +  
geom\_histogram(color = "black", fill = "gold", bins = 8) +  
facet\_grid(TIPS$day2)



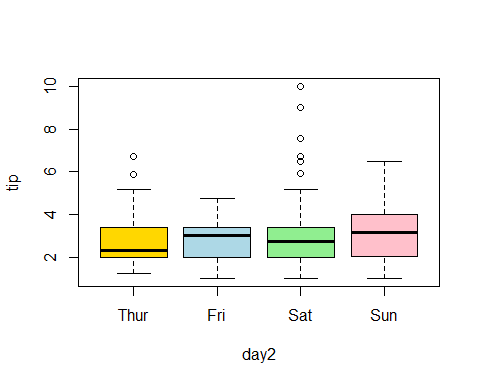
Es liegt augenscheinlich eine Normalverteilung vor. Die Ränder sind kleiner als die Mitte.

### Boxplots

Boxplot(total\_bill~day2, data=TIPS, id=list(method="identify"), col=c("gold", "lightblue", "lightgreen", "pink"))



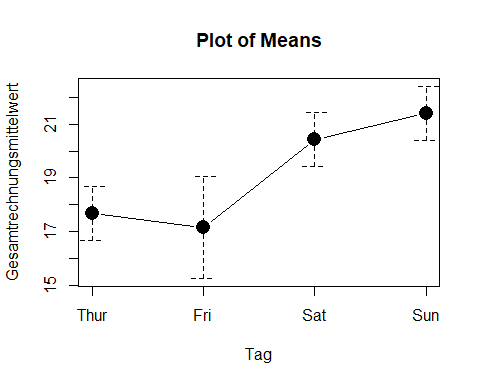
Boxplot(tip~day2, data=TIPS, id=list(method="identify"), col=c("gold", "lightblue", "lightgreen", "pink"))



Der Boxplot zeigt, dass sowohl der Rechnungsbetrag als auch das Trinkgeld einige Ausreißer nach oben haben.

### Plot für die Mittelwerte

with(TIPS, plotMeans(total\_bill, day2, error.bars="se", connect=TRUE, xlab='Tag', ylab="Gesamtrechnungsmittelwert"))



Wie der Plot der Mittelwerte erkennen lässt, bestehen bezüglich der vier Wochentage Unterschiede im Mittelwert der Rechnungshöhe.

library(abind, pos=24)  
library(e1071, pos=25)  
numSummary(TIPS[,"total\_bill", drop=FALSE], groups=TIPS$day2, statistics=c("mean", "sd", "IQR", "quantiles", "skewness",   
 "kurtosis"), quantiles=c(0,.25,.5,.75,1), type="2")

## mean sd IQR skewness kurtosis 0% 25% 50% 75%  
## Thur 17.68274 7.886170 7.7125 1.3607702 1.8199742 7.51 12.4425 16.20 20.1550  
## Fri 17.15158 8.302660 9.6550 1.3167591 1.9944517 5.75 12.0950 15.38 21.7500  
## Sat 20.44138 9.480419 10.8350 1.2621122 1.7156316 3.07 13.9050 18.24 24.7400  
## Sun 21.41000 8.832122 10.6100 0.8230851 0.4222415 7.25 14.9875 19.63 25.5975  
## 100% total\_bill:n  
## Thur 43.11 62  
## Fri 40.17 19  
## Sat 50.81 87  
## Sun 48.17 76

Am Donnerstag und Freitag sind die Mittewerte des Rechnungsbetrages niedriger ($17.6 und $17.1) als am Wochenende (Mittelwert Samstag = $20.4 und Sonntag= $21.4). Auffällig ist, dass die Anzahl von Rechnungen am Freitag (n= 19) wesentlich kleiner ist als an den anderen Tagen (n zw. 62 und 87).

Erfüllte Voraussetzungen:

• Die abhängige Variable Rechnung ist intervallskaliert.

• Die unabhängigen Variable Wochentag ist kategorial.

• Die durch den Faktor gebildeten Gruppen sind unabhängig

• Die abhängige Variable ist normalverteilt innerhalb jeder der Gruppen, siehe Abbildung mit Histogrammen für die einzelnen Wochentage.

• Homogenität der Varianzen: Die Gruppen stammen aus Grundgesamtheiten mit annähernd identischen Varianzen der abhängigen Variablen (siehe unten Ergebnis für Levene-Test)

# 5. Prüfung der Varianzhomogenität (Levene-Test)

with(TIPS, tapply(total\_bill, day2, var, na.rm=TRUE))

## Thur Fri Sat Sun   
## 62.19168 68.93416 89.87834 78.00638

leveneTest(total\_bill ~ day2, data=TIPS, center="mean")

## Levene's Test for Homogeneity of Variance (center = "mean")  
## Df F value Pr(>F)  
## group 3 0.6921 0.5577  
## 240

Der Levene-Test prüft die Nullhypothese, dass die Varianzen der Gruppen sich nicht unterscheiden. Ist der Levene-Test nicht signifikant, so kann von homogenen Varianzen ausgegangen. Wäre der Levene-Test jedoch signifikant, so wäre eine der Grundvoraussetzungen der Varianzanalyse verletzt. Gegen leichte Verletzungen gilt die Varianzanalyse als robust; vor allem bei genügend grossen und etwa gleich grossen Gruppen sind Verletzungen nicht problematisch. Bei ungleich grossen Gruppen führt eine starke Verletzung der Varianzhomogenität zu einer Verzerrung des F-Tests. Alternativ können dann auf den den Welch-Test zurückgegriffen werden. Dabei handelt es sich um adjustierte F-Tests.

Im vorliegenden Beispiel ist der Levene-Test nicht signifikant (F(3,240) = 0.692, p = 0.557, so dass von Varianzhomogenität ausgegangen werden kann. Das heißt, es muss keine Welch-Korrekturdurchgeführt werden.

# 6. Ergebnisse der einfaktoriellen Varianzanalyse

### Deskriptive Statistiken und erste Ergebnisse

AnovaModel.1 <- aov(total\_bill ~ day, data=TIPS)  
summary(AnovaModel.1)

## Df Sum Sq Mean Sq F value Pr(>F)   
## day 3 644 214.65 2.767 0.0425 \*  
## Residuals 240 18615 77.56   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

with(TIPS, numSummary(total\_bill, groups=day, statistics=c("mean", "sd")))

## mean sd data:n  
## Fri 17.15158 8.302660 19  
## Sat 20.44138 9.480419 87  
## Sun 21.41000 8.832122 76  
## Thur 17.68274 7.886170 62

Es ist zu erkennen, dass das Gesamtmodel signifikant wird F(3,240) = 2.767 , p = .0425).

### Das partielle Eta-Quadrat

Das partielle Eta-Quadrat (partielles η2) ist ein Mass für die Effektgrösse: Es setzt die Variation, die durch einen Faktor erklärt wird, in Bezug mit jener Variation, die nicht durch andere Faktoren im Modell erklärt wird. Das heisst, es wird ausschliesslich jene Variation betrachtet, welche nicht durch die anderen Faktoren im Modell erklärt wird. Das partielle Eta-Quadrat zeigt, welchen Anteil davon ein Faktor erklärt. Im Falle der einfaktoriellen Varianzanalyse ist das partielle Eta-Quadrat ist jener Anteil der korrigierten Gesamtvariation, der durch das Modell erklärt wird.

library(sjstats)

## Warning: package 'sjstats' was built under R version 3.6.2

AV <- TIPS$total\_bill  
UV <- TIPS$day  
Anova\_test <- aov(AV~ UV)  
eta <- eta\_sq(Anova\_test)  
eta

## term etasq  
## 1 UV 0.033

AnovaModel.1 <- aov(total\_bill ~ day, data=TIPS)  
.myAnova <- summary(AnovaModel.1)  
.myAnova

## Df Sum Sq Mean Sq F value Pr(>F)   
## day 3 644 214.65 2.767 0.0425 \*  
## Residuals 240 18615 77.56   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# EFFECT SIZE  
.effectSize <- data.frame(format(round(.myAnova[[1]]$`Sum Sq` /   
 sum(.myAnova[[1]]$`Sum Sq`), 4), nsmall = 4), row.names =   
 rownames(.myAnova[[1]]))  
colnames(.effectSize) <- c("part. eta sq.")  
.effectSize

## part. eta sq.  
## day 0.0334  
## Residuals 0.9666

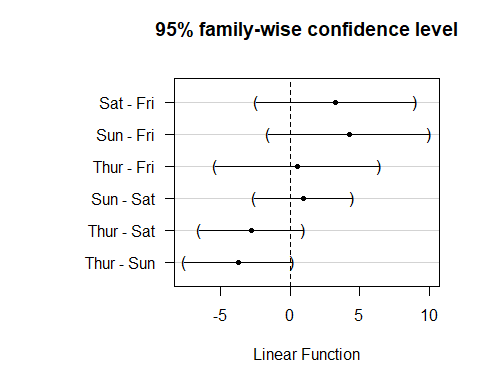
Im vorliegenden Beispiel beträgt das partielle Eta-Quadrat .033. Das heisst, es wird 3.3% der Variation in gezahlten Rechnungsbetrag durch die vier Wochentage erklärt.

# 7. Post-hoc-Tests

Bei der Berechnung von Post-hoc-Tests wird im Prinzip für jede Kombination von zwei Mittelwerten ein t-Test durchgeführt. Im aktuellen Beispiel mit vier Wochentagen sind dies 6 Möglichkeiten. Multiple Tests sind jedoch problematisch, da der Alpha-Fehler (die fälschliche Ablehnung der Nullhypothese) mit der Anzahl der Vergleiche steigt. Wird nur ein t-Test mit einem Signifikanzlevel von .05 durchgeführt, so beträgt die Wahrscheinlichkeit des Nicht-Eintreffens des Alpha-Fehlers 95%. Werden jedoch sechs solcher Paarvergleiche vorgenommen, so beträgt die Nicht-Eintreffens-Wahrscheinlichkeit des Alpha-Fehlers .05/6 = .008. Das heisst, jeder Test wird gegen ein Niveau von .008 geprüft.

local({  
 .Pairs <- glht(AnovaModel.1, linfct = mcp(day = "Tukey"))  
 print(summary(.Pairs)) # pairwise tests  
 print(confint(.Pairs)) # confidence intervals  
 print(cld(.Pairs)) # compact letter display  
 old.oma <- par(oma=c(0,5,0,0))  
 plot(confint(.Pairs))  
 par(old.oma)  
})

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: aov(formula = total\_bill ~ day, data = TIPS)  
##   
## Linear Hypotheses:  
## Estimate Std. Error t value Pr(>|t|)   
## Sat - Fri == 0 3.2898 2.2302 1.475 0.4444   
## Sun - Fri == 0 4.2584 2.2589 1.885 0.2293   
## Thur - Fri == 0 0.5312 2.3094 0.230 0.9955   
## Sun - Sat == 0 0.9686 1.3828 0.700 0.8934   
## Thur - Sat == 0 -2.7586 1.4637 -1.885 0.2295   
## Thur - Sun == 0 -3.7273 1.5072 -2.473 0.0639 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)  
##   
##   
## Simultaneous Confidence Intervals  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: aov(formula = total\_bill ~ day, data = TIPS)  
##   
## Quantile = 2.5662  
## 95% family-wise confidence level  
##   
##   
## Linear Hypotheses:  
## Estimate lwr upr   
## Sat - Fri == 0 3.2898 -2.4331 9.0127  
## Sun - Fri == 0 4.2584 -1.5383 10.0551  
## Thur - Fri == 0 0.5312 -5.3950 6.4573  
## Sun - Sat == 0 0.9686 -2.5798 4.5170  
## Thur - Sat == 0 -2.7586 -6.5148 0.9975  
## Thur - Sun == 0 -3.7273 -7.5948 0.1403  
##   
## Fri Sat Sun Thur   
## "a" "a" "a" "a"



oneway.test(total\_bill ~ day, data=TIPS) # Welch test

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: total\_bill and day  
## F = 3.0113, num df = 3.000, denom df = 75.305, p-value = 0.03531

Es wird ersichtlich, dass sich die Wochentage bezüglich des gezahlten Rechnungsbetrages nicht signifikant unterscheiden. Es können also keine Gruppen von Wochentagen gebildet werden.

# 8. Profildiagramm

Siehe obrige Abbildung.

# 9. Berechnung der Effektstärke

Um die Bedeutsamkeit eines Ergebnisses zu beurteilen, werden Effektstärken berechnet. Im Beispiel sind zwar einige der Mittelwertsunterschiede zwar signifikant, doch es stellt sich die Frage, ob sie gross genug sind, um als bedeutend eingestuft zu werden.

Es gibt verschiedene Arten die Effektstärke zu messen. Zu den bekanntesten zählen die Effektstärke von Cohen (d) und der Korrelationskoeffizient (r) von Pearson. Der Korrelationskoeffizient eignet sich sehr gut, da die Effektstärke dabei immer zwischen 0 (kein Effekt) und 1 (maximaler Effekt) liegt. Wenn sich jedoch die Gruppen hinsichtlich ihrer Grösse stark unterscheiden, wird empfohlen, d von Cohen zu wählen, da r durch die Grössenunterschiede verzerrt werden kann.

Da R das partielle Eta-Quadrat ausgibt, wird dieses hier in die Effektstärke nach Cohen (1992) umgerechnet. In diesem Fall befindet sich die Effektstärke immer zwischen 0 und unendlich.

library(sjstats)  
AV <- TIPS$total\_bill  
UV <- TIPS$day  
Anova\_test <- aov(AV~ UV)  
eta <- eta\_sq(Anova\_test)  
eta\_s <- eta$etasq  
  
print(paste0("Eta-Quadrat liegt bei: ", eta\_s))

## [1] "Eta-Quadrat liegt bei: 0.0334367974221405"

# Effektstaerke  
  
f <- sqrt(eta\_s/(1-eta\_s))  
eff\_staer <- round(f,2)  
print(paste0("Effektstaerke liegt bei: ", eff\_staer))

## [1] "Effektstaerke liegt bei: 0.19"

Um zu beurteilen, wie gross dieser Effekt ist, kann man sich an der Einteilung von Cohen (1988) orientieren:

f = .10 entspricht einem schwachen Effekt f = .25 entspricht einem mittleren Effekt f = .40 entspricht einem starken Effekt

Damit entspricht eine Effektstärke von 0.19 einem schwachen Effekt.

# 10. Eine Aussage

Die Auswahl der Trainingsmethode hat einen signifikanten Einfluss auf die Ausdauer (F(3,240) = 2.767 , p = .0425). 3.3% der Streuung der Ausdauer-Werte um den Gesamtmittelwert kann durch die Trainingsmethoden erklärt werden. Die Effektstärke nach Cohen (1988) liegt bei f = 0.19 und entspricht einem schwachen Effekt. Es ist zu erkennen, dass die Wochentage statistisch signifikanten auf die Höhe des Rechnungsbetrages sind F(3,240) = 2.767 , p = .0425). Am Donnerstag und Freitag sind die Mittewerte des Rechnungsbetrages niedriger (M=$17.6 und SD= $7.8 und Freitag M=$17.1 und SD=8.3) als am Wochenende (Mittelwert Samstag = $20.4 und SD= 9.48 und Sonntag M = $21.4 und SD= 8.8). Auffällig ist, dass die Anzahl von Rechnungen am Freitag (n= 19) wesentlich kleiner ist als an den anderen Tagen (n zw. 62 und 87). 3.3% der Streuung des Rechnungsbetrages um den Gesamtmittelwert kann durch die Wochentage erklärt werden. Die Effektstärke nach Cohen (1988) liegt bei f = 0.19 und entspricht einem schwachen Effekt. Post-hoc-Tests mit Tukey zeigen, dass sich keine vier Gruppen von Wochentagen bilden lassen (alle p > .008). Damit kann festgehalten werden, dass kein Wochentag für sich eine unabhängige Gruppe bildet. Auch wenn das Testergebnis (p= .0425) statistisch signifikant ist, wird H0 angenommen, da der Post-Hoc-Test keine signifikanten Ergebnisse liefert und H1 wird somit verworfen.