Single factor analysis of variance (ANOVA)

Ozgur Polat und Kathrin Einzmann

09 01 2020

## Data Record: tips.csv

Var 1 = total\_bill (dependent, invoice is interval scaled)

Var 2 = day (independent variable “day” is categorical)

# 1. Hypothese:

H1: The amount of the invoice changes on the different days of the week.

H0: The amount of the invoice does not change on the different days of the week.

# 2. Requirements for the single factor analysis of variance

* The dependent variable is interval-scaled.
* The independent variables (factors) are categorical (nominal or ordinal scaled)
* The groups formed by the factor are independent
* The dependent variable is normally distributed within each of the groups. From 25 subjects per group, violations of this requirement are unproblematic
* Homogeneity of the variances: The groups come from populations with approximately identical variances of the dependent variables (see Level Test)

# 3. Basic concepts: What is the single factor analysis of variance?

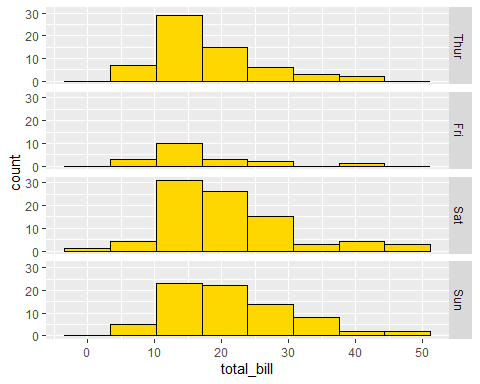
The single factor analysis of variance - also known as “single factor ANOVA”, as “Analysis of Variance” - tests whether the means of several independent groups (or samples) defined by a categorical independent variable differ. This categorical independent variable is called a “factor” in the context of analysis of variance. Accordingly, the expressions of the independent variable are called “factor levels”, and the term “treatments” is also commonly used. An analysis of variance is referred to as “single factor” if it uses only one factor, i.e. a grouping variable (multi-factorial analysis of variance).

The principle of analysis of variance is to decompose the variance of the dependent variable. The total variance is composed of the so-called “variance within groups” and the “variance between groups”. These two parts are compared with each other in an analysis of variance. Single factor ANOVA is a generalisation of the t-test for independent samples for comparison of more than two groups (or samples). The question of single factor analysis of variance is often abbreviated as “Do the means of an independent variable differ between several groups? Which factor levels differ?”

# 4. Descriptive statistics

### Normalverteilung

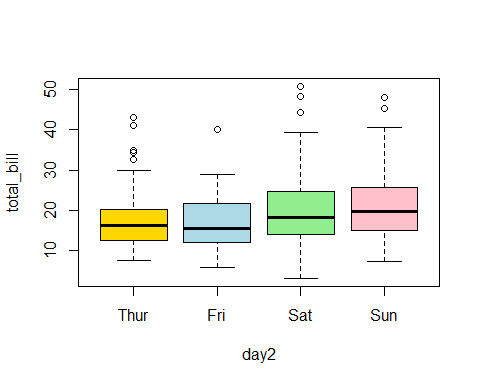
library(reshape2)  
library(ggplot2)  
  
TIPS$day2<-factor(TIPS$day,levels=levels(TIPS$day)[c(4,1,2,3)])  
  
ggplot(TIPS, aes(x = total\_bill)) +  
geom\_histogram(color = "black", fill = "gold", bins = 8) +  
facet\_grid(TIPS$day2)



Apparently there is a normal distribution. The edges are smaller than the center.

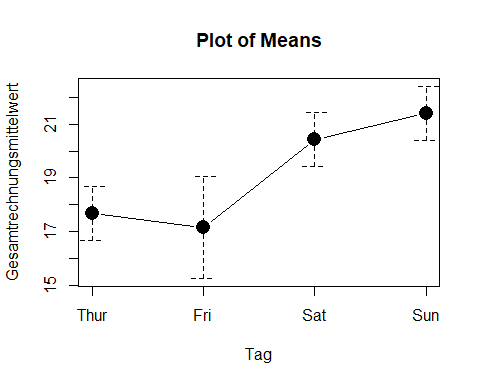
### Boxplots

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

 The box plot shows that both the bill amount and the tip have some outliers upwards.

### Plot for the mean values

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

 As the plot of the mean values shows, there are differences in the mean value of the invoice amount for the four weekdays.

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

On Thursday and Friday, the average invoice amount is lower ($17.6 and $17.1) than on the weekend (average Saturday = $20.4 and Sunday = $21.4). It is noticeable that the number of invoices on Friday (n= 19) is much smaller than on the other days (n zw. 62 and 87).

Conditions fulfilled:

* The dependent variable Invoice is interval-scaled.
* The independent variable Weekday is categorical.
* The groups formed by the factor are independent
* The dependent variable is normally distributed within each of the groups, see figure with histograms for each day of the week.
* Homogeneity of the variances: The groups come from populations with approximately identical variances of the dependent variable (see below result for level test)

# 5. Test of variance homogeneity (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

The Levene test tests the null hypothesis that the variances of the groups do not differ. If the Levene test is not significant, homogeneous variances can be assumed. However, if the Levene test were significant, one of the basic requirements of the analysis of variance would be violated. The analysis of variance is regarded as robust against minor injuries; injuries are not problematic, especially in groups of sufficient size and about the same size. In groups of unequal size, a strong violation of variance homogeneity leads to a distortion of the F-test. Alternatively, the Welch test can then be used. These are adjusted F-tests.

In this example, the Levene test is not significant (F(3,240) = 0.692, p = 0.557, so variance homogeneity can be assumed. This means that no Welch correction has to be performed.

# 6. Results of the single factorial analysis of variance

### 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

It can be seen that the overall model becomes significant F(3,240) = 2.767 , p = .0425).

### Partial das Eta-Quadrat

The partial eta square (partial η2) is a measure of effect size: it relates the variation explained by a factor to the variation not explained by other factors in the model. This means that only the variation that is not explained by the other factors in the model is considered. The partial eta-square shows what proportion of this variation is explained by a factor. In the case of single factor analysis of variance, the partial eta-square is the proportion of the corrected total variation that is explained by the model.

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

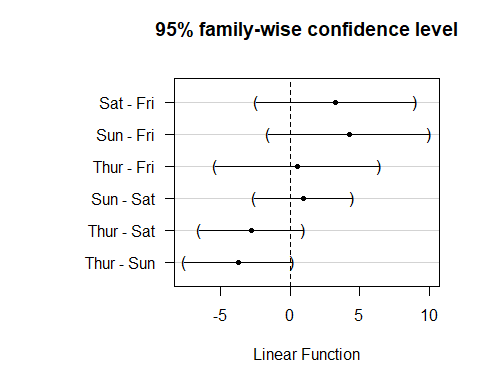
In this example, the partial eta-square is .033, which means that 3.3% of the variation in paid invoice amount is explained by the four weekdays.

# 7. Post-hoc-Tests

When calculating post-hoc tests, in principle a t-test is performed for each combination of two mean values. In the current example with four days of the week, these are 6 possibilities. However, multiple tests are problematic because the alpha error (the false rejection of the null hypothesis) increases with the number of comparisons. If only one t-test with a significance level of .05 is performed, the probability of the alpha error not occurring is 95%. However, if six such pair comparisons are made, the probability of no alpha error is .05/6 = .008, meaning that each test is tested against a level of .008.

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 0.2296   
## Thur - Sun == 0 -3.7273 1.5072 -2.473 0.0636 .  
## ---  
## 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 = 2.5648  
## 95% family-wise confidence level  
##   
##   
## Linear Hypotheses:  
## Estimate lwr upr   
## Sat - Fri == 0 3.2898 -2.4300 9.0096  
## Sun - Fri == 0 4.2584 -1.5351 10.0520  
## Thur - Fri == 0 0.5312 -5.3918 6.4541  
## Sun - Sat == 0 0.9686 -2.5778 4.5151  
## Thur - Sat == 0 -2.7586 -6.5127 0.9955  
## Thur - Sun == 0 -3.7273 -7.5927 0.1382  
##   
## 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

It is apparent that the days of the week do not differ significantly in terms of the amount of the invoice paid. Therefore, no groups of weekdays can be formed.

# 8. Profildiagramm

See figure above.

# 9. Calculation of the effect strength

To assess the significance of a result, effect strengths are calculated. In the example, although some of the mean differences are significant, the question arises as to whether they are large enough to be classified as significant.

There are different ways to measure effect sizes. Among the best known are the effect size of Cohen (d) and the correlation coefficient (r) of Pearson. The correlation coefficient is very suitable because the effect strength is always between 0 (no effect) and 1 (maximum effect). However, if the groups differ greatly in size, it is recommended to choose d by Cohen, since r can be distorted by the differences in size.

Since R is the partial eta-square, this is converted here to the effect size according to Cohen (1992). In this case the effect size is always between 0 and infinity.

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"

In order to assess how great this effect is, one can use the classification of Cohen (1988) as a guide:

f = .10 corresponds to a weak effect f = .25 corresponds to a medium effect f = .40 corresponds to a strong effect

Thus an effect strength of 0.19 corresponds to a weak effect.

# 10. A statement

The choice of training method has a significant influence on endurance (F(3,240) = 2,767 , p = .0425). 3.3% of the variation of the endurance values around the total mean value can be explained by the training methods. The effect strength according to Cohen (1988) is f = 0.19 and corresponds to a weak effect. It can be seen that the days of the week are statistically more significant on the amount of the invoice F(3,240) = 2.767 , p = .0425). On Thursday and Friday the mean values of the invoice amount are lower (M=$17.6 and SD= $7.8 and Friday M=$17.1 and SD=8.3) than on the weekend (mean Saturday = $20.4 and SD= 9.48 and Sunday M = $21.4 and SD= 8.8). It is noticeable that the number of invoices on Friday (n=19) is much smaller than on the other days (n zw. 62 and 87). 3.3% of the spread of the invoice amount around the total average can be explained by the weekdays. The effect strength according to Cohen (1988) is f = 0.19 and corresponds to a weak effect. Post-hoc tests with Tukey show that no four groups of weekdays can be formed (all p > .008). Thus it can be stated that no weekday forms an independent group by itself. Even if the test result (p= .0425) is statistically significant, H0 is assumed, since the post-hoc test does not give significant results and H1 is therefore rejected.