

MSc. Research Methods - Statistikteil Lösungen 2018

Gian-Andrea Egeler

November 2018

Musterlösung Aufgabe 2.2: einfaktorielle ANOVA

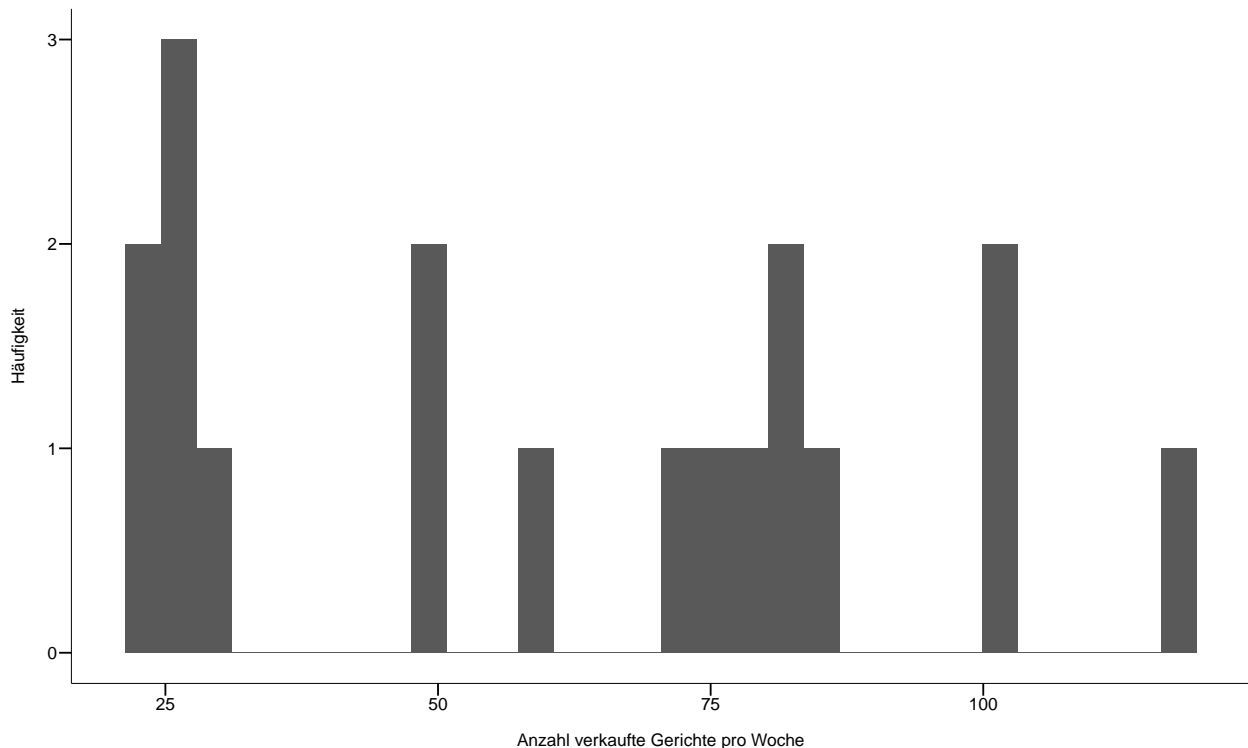
```
df <- nova # kclone den originaler Datensatz

# fasst die vier Inhalte der Gerichte zu drei Inhalten zusammen.

df$label_content[grepl("Pflanzlich+", df$label_content)] <- "Vegetarisch"

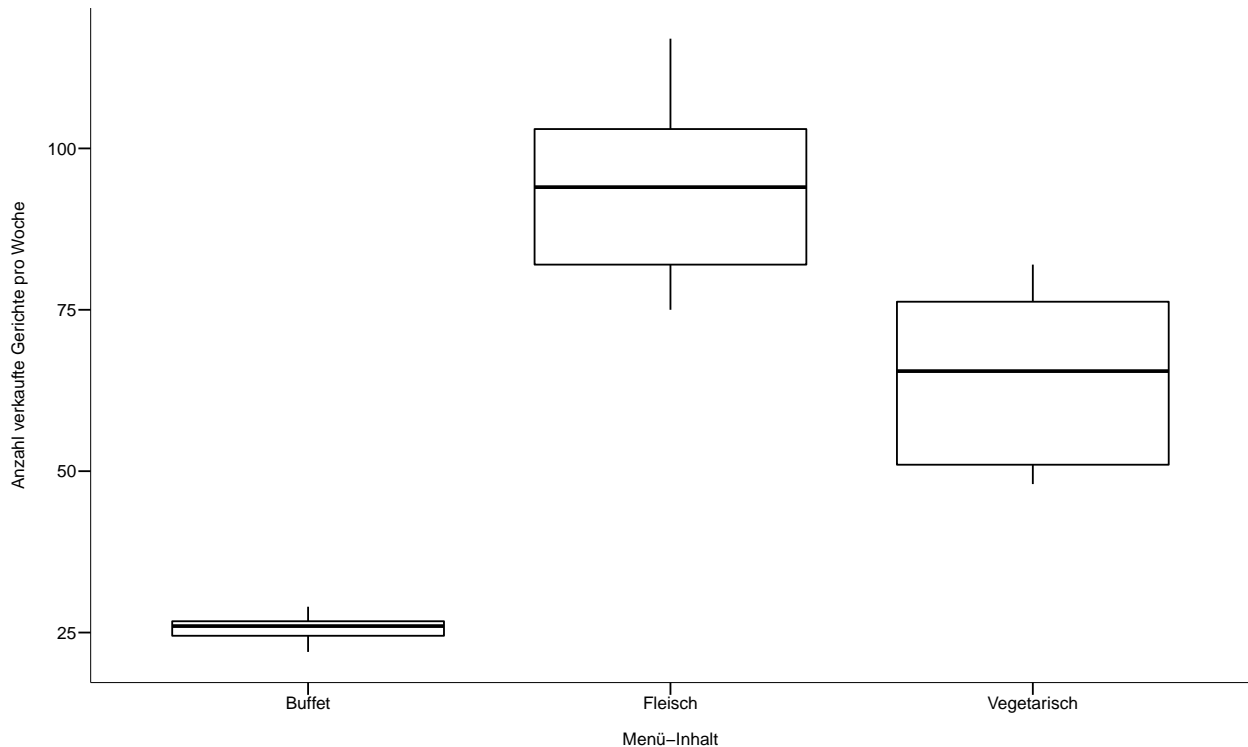
# gruppieren Daten nach Menü-Inhalt und Woche
df_ <- df %>%
  group_by(label_content, week) %>%
  summarise(tot_sold = n()) %>%
  drop_na() # lässt die unbekannten Menü-Inhalte weg

# überprüft die Voraussetzungen für eine ANOVA
# Histogramm, sagt aber nicht viel aus
ggplot(df_, aes(x = tot_sold, y = ..count..)) +
  geom_histogram() +
  labs(x = "\nAnzahl verkaufte Gerichte pro Woche", y = "Häufigkeit\n") +
  mytheme
```



```
# Boxplot
ggplot(df_, aes(x = label_content, y = tot_sold)) +
```

```
geom_boxplot(fill="white", color = "black", size = 1) +
labs(x = "\nMenü-Inhalt", y = "Anzahl verkaufte Gerichte pro Woche\n") +
mytheme # klare Varianzheterogenität
```

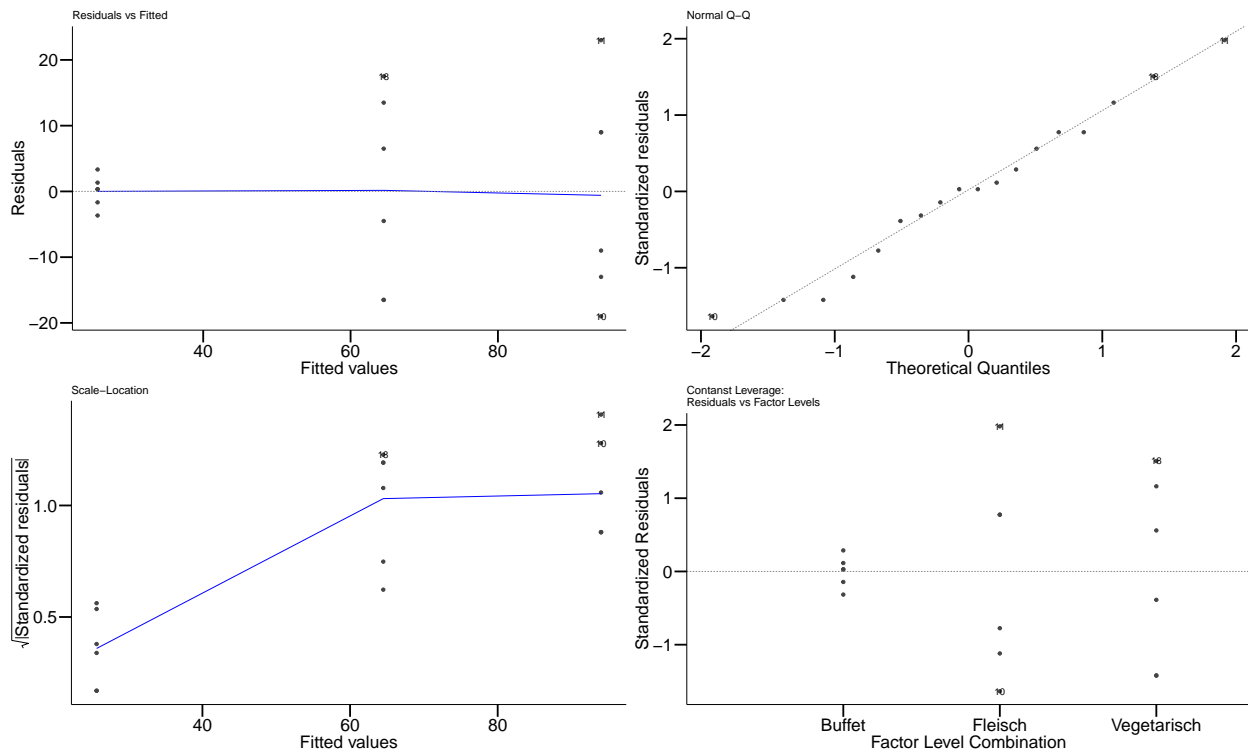


```
# definiert das Modell
model <- aov(tot_sold ~ label_content, data = df_)
```

```
summary.lm(model)
```

```
##
## Call:
## aov(formula = tot_sold ~ label_content, data = df_)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.0000  -7.8750   0.3333   8.3750  23.0000
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      25.667     5.193   4.943 0.000177 ***
## label_contentFleisch    68.333     7.344   9.305 1.28e-07 ***
## label_contentVegetarisch  38.833     7.344   5.288 9.11e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.72 on 15 degrees of freedom
## Multiple R-squared:  0.8531, Adjusted R-squared:  0.8335
## F-statistic: 43.56 on 2 and 15 DF, p-value: 5.653e-07
```

```
autoplot(model) + mytheme
```



Fazit: Inspektion der Modellvoraussetzung zeigt klare Verletzungen der Homoskedastizität. Nächster Schritt Welch-Test

```
# überprüft die Voraussetzungen des Welch-Test.
# Gibt es eine hohe Varianzheterogenität und
# ist die relative Verteilung der Residuen gegeben?
# In diesem Fall wäre ein Welch-Test passend
w_test <- oneway.test(data=df_, tot_sold ~ label_content, var.equal=F)
w_test

##
## One-way analysis of means (not assuming equal variances)
##
## data: tot_sold and label_content
## F = 64.997, num df = 2.0000, denom df = 6.9832, p-value =
## 3.067e-05
```

Methoden

Ziel war es die Unterschiede in den Verkaufszahlen pro Menü-Inhalt aufzuzeigen. Da die Kriteriumvariable (Verkaufszahlen) metrisch und die Prädiktorvariable kategorial sind, wurde eine einfaktorielle ANOVA gerechnet. Die visuelle Inspektion der Voraussetzungen zeigte insbesondere schwere Verletzungen der Homoskedastizität. Der Boxplot bestätigt diesen Befund. Daher wurde in einem weiteren Schritt der Welch-Test für ungleiche Varianzen gerechnet.

Ergebnisse

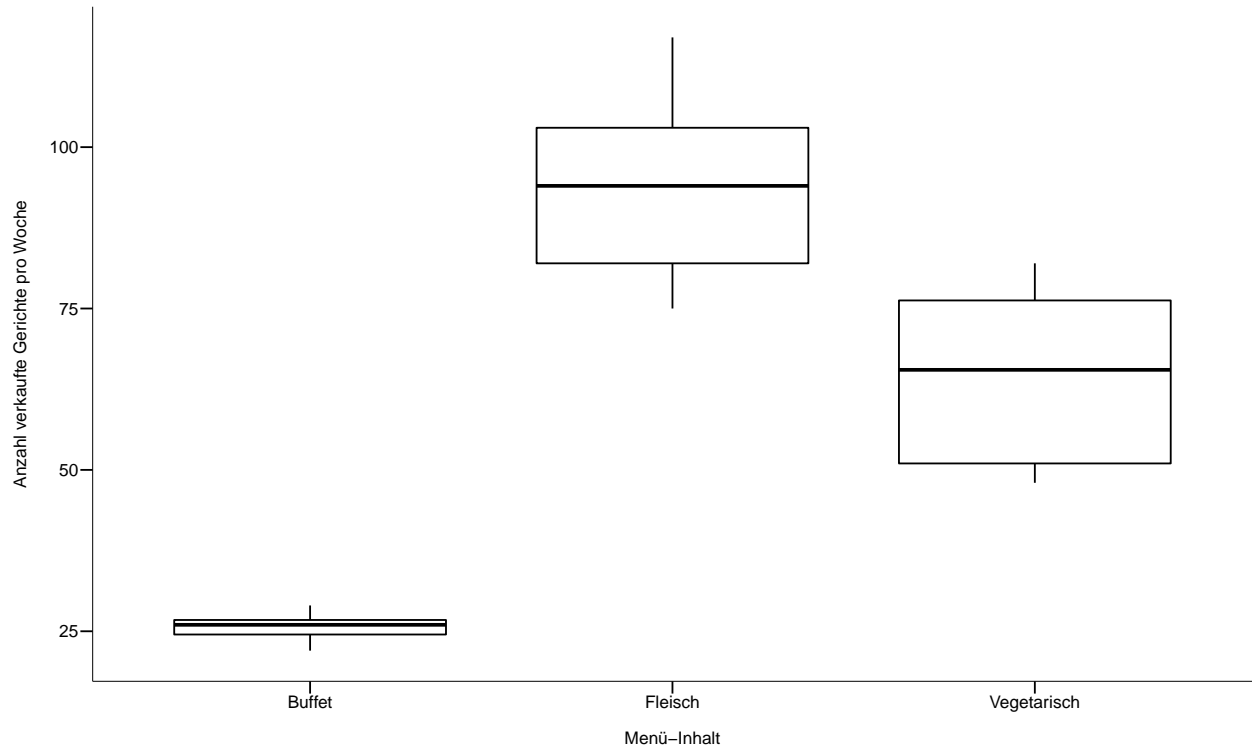


Figure 1: Die wöchentlichen Verkaufszahlen unterscheiden je nach Menü-Inhalt stark.

Die Menü-Inhalte (Fleisch, Vegetarisch und Buffet) unterscheiden sich in den Verkaufszahlen signifikant ($F(2,7) = 65, p < .001$). Die Figure 1 zeigt die Verkaufszahlen pro Menü-Inhalt.