

# MSc. Research Methods - Statistikteil Lösungen 2018

Gian-Andrea Egeler

November 2018

## Übung 1.2: $\chi^2$ -Test

$H_0$ : Es gibt keine Unterschiede zwischen der Population und der Stichprobe bezüglich Geschlecht und Hochschulzugehörigkeit.

$H_1$ : Es gibt Unterschiede zwischen der Population und der Stichprobe bezüglich Geschlecht und Hochschulzugehörigkeit.

```
# gruppiert die Variablen und fasst sie
# nach Geschlecht und Hochschulzugehörigkeit zusammen

canteen <- group_by(nova, gender, member) %>%
  summarise(tot = n()) %>%
  ungroup() %>%
  mutate(canteen_member = c("Mitarbeiterinnen", "Studentinnen", "Mitarbeiter", "Studenten"))

# definiert einen Vektor mit den erwarteten Wahrscheinlichkeiten,
# beachtet dabei die Reihenfolge

population_exp <- c(.15,.32,.16,.37) # Achtung: die Summe muss 1 ergeben

# berechnet den Chi-Quadrat-Test

chi_sq <- chisq.test(canteen$tot, p = population_exp)
chi_sq

##
## Chi-squared test for given probabilities
##
## data: canteen$tot
## X-squared = 103.81, df = 3, p-value < 2.2e-16

# führt einen Fishers exakten Test durch
# erstellt dafür einen neuen Datensatz;
# tibble() ist eine ähnliche Funktion wie data.frame()

fisher_t <- tibble(member = c("Mitarbeiterinnen", "Studentinnen", "Mitarbeiter", "Studenten"),
  population = round(population_exp * 2138,0),
  stichprobe = canteen$tot,
  population_pct = population_exp,
  canteen_pct = canteen$tot / sum(canteen$tot))

# simulate.p.value simuliert einen p-Wert mit 100000 Replikaten

fish <- fisher.test(fisher_t[,2:3], simulate.p.value=TRUE, B = 100000)
fish

##
## Fisher's Exact Test for Count Data with simulated p-value (based
```

```
## on 1e+05 replicates)
##
## data: fisher_t[, 2:3]
## p-value = 1e-05
## alternative hypothesis: two.sided
```

## Methoden

Ziel war es die novanimal Stichprobe in der Gesamtpopulation bezüglich Geschlecht und Hochschulzugehörigkeit besser einzuordnen. Die Gesamtpopulation definiert sich durch alle aktiven CampusCards. Da die untersuchten Variablen Kategorien sind, wurde einen  $\chi^2$  resp. einen Fishers exakten Test gerechnet.

## Ergebnisse

Der  $\chi^2$ -Test sagt uns, dass die NOVANIMAL-Stichprobe von der Population signifikant unterscheidet ( $\chi^2(3) = 103.808$ ,  $p > .001$ ). Auch der Fishers exakte Test zeigt einen ähnlichen p-Wert ( $p > .001$ ). Demnach ist unsere Stichprobe bezüglich den Variablen Geschlecht und Hochschulzugehörigkeit nicht repräsentativ für die Grundgesamtheit. Es scheint, dass die Studentinnen unter- und die Studenten übervertreten sind (siehe Table 1).

Table 1: Anzahl und Anteil Beobachtungen in der Population und in der Stichprobe

Hochschulzugehörigkeit	Anzahl Population	Anteil Population (%)	Anzahl Stichprobe	Anteil Stichprobe (%)
Mitarbeiterinnen	321	15	209	19
Studentinnen	684	32	212	19
Mitarbeiter	342	16	255	23
Studenten	791	37	440	39

## Übung 1.3: t-Test

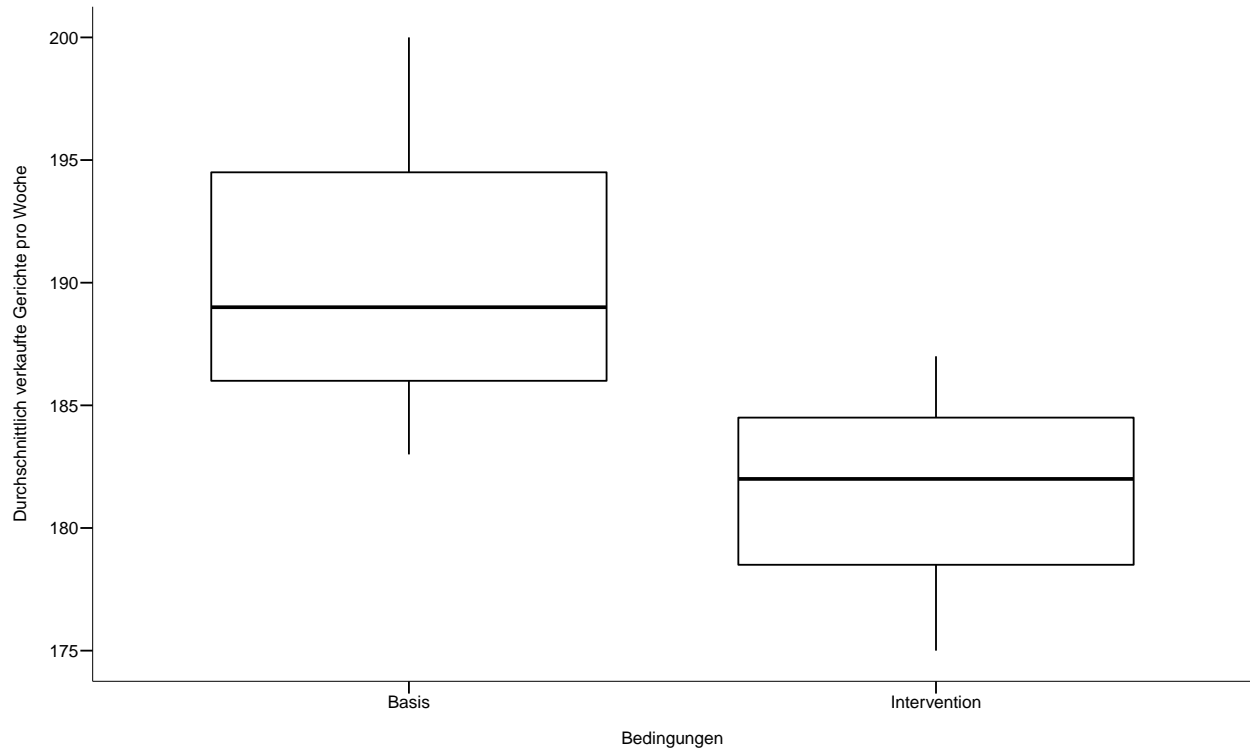
$H_0$ : Es gibt keine Unterschiede in den Verkaufszahlen zwischen Basis- und Interventionswochen.

$H_1$ : Es gibt Unterschiede in den Verkaufszahlen zwischen Basis- und Interventionswochen.

*# Daten müssen zuerst nach "week" und "condition" zusammengefasst werden*

```
df <- nova %>%
  group_by(week, condit) %>%
  summarise(tot_sold = n())

# überprüft die Voraussetzungen für einen t-Test
ggplot(df, aes(x = condit, y= tot_sold)) +
  geom_boxplot(fill = "white", color = "black", size = 1) +
  labs(x="\nBedingungen", y="Durchschnittlich verkaufte Gerichte pro Woche\n") +
  mytheme
```



```
# führt einen t-Tests durch;
# es wird angenommen, dass die Verkaufszahlen zwischen den Bedingungen unabhängig sind

t_test <- t.test(df[df$condit == "Basis", ]$tot_sold,
                 df[df$condit == "Intervention", ]$tot_sold)

t_test

##
## Welch Two Sample t-test
##
## data: df[df$condit == "Basis", ]$tot_sold and df[df$condit == "Intervention", ]$tot_sold
## t = 1.5367, df = 3.5781, p-value = 0.2074
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.343572 27.010238
## sample estimates:
## mean of x mean of y
## 190.6667 181.3333
```

## Methoden

Ziel war es die wöchentlichen Verkaufszahlen zwischen den Interventions- und Basiswochen zu vergleichen. Die Annahme war, dass die wöchentlichen Verkaufszahlen unabhängig sind. Daher können die mittleren Verkaufszahlen pro Woche zwischen den beiden Bedingungen mittels t-Test geprüft werden. Obwohl die visuelle Inspektion keine schwerwiegenden Verletzungen der Modelvoraussetzung zeigte, wurde einen Welch t-Test gerechnet.

## Ergebnisse

In den Basiswochen werden mehr Gerichte pro Woche verkauft als in den Interventionswochen (siehe Figure

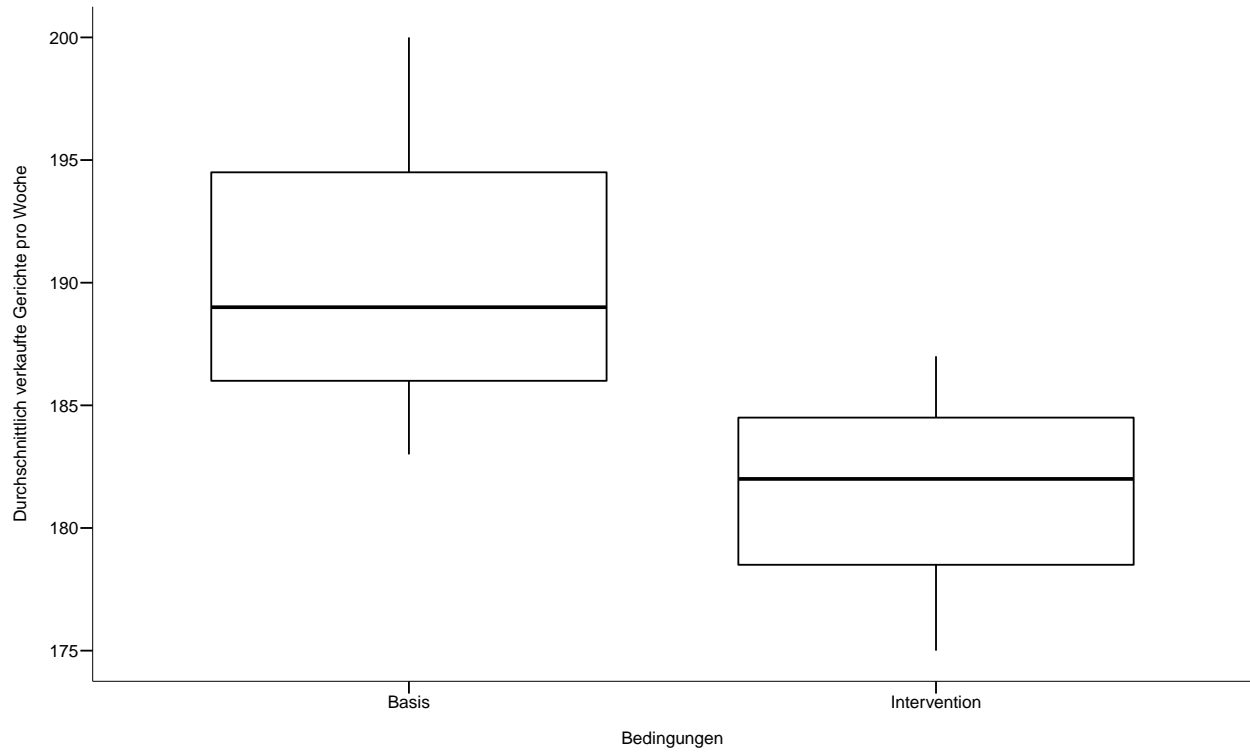


Figure 1: Die wöchentlichen Verkaufszahlen für die Interventions- und Basiswochen unterscheiden sich nicht signifikant.

1). Die wöchentlichen Verkaufszahlen zwischen den Bedingungen (Basis oder Intervention) unterscheiden sich gemäss Welch t-Test jedoch nicht signifikant ( $t(4) = 1.537$ ,  $p = 0.207$ ).