

# Inferenzstatistik

# Die Induktive Statistik

Im letzten Semester hatten wir den Unterschied zwischen der deskriptiven und der induktiven Statistik kennen gelernt (und uns im letzten Semester dann auch auf die deskriptive Statistik beschränkt).

Dieses Semester wollen wir uns der induktiven Statistik zuwenden.

# Inferenz vs. Vorhersage

Die induktive Statistik lässt sich in zwei Hauptblöcke unterteilen:

## Modellierung zur Erklärung (Inferenz)

Hier versuchen Sie explizit den Zusammenhang zwischen ihrer zu untersuchenden Variablen  $y$  und einer (oder mehreren) erklärenden Variablen  $x$  zu beschreiben und zu quantifizieren. Oberstes Ziel wäre es einen **kausalen** Zusammenhang zwischen den Variablen aufzudecken (falls möglich).

## Modellierung zur Vorhersage

Hier versuchen Sie eine zu untersuchende Variable  $y$  durch mehrere erklärende Variablen  $x$  **vorherzusagen**. Es interessiert Sie hierbei nicht so sehr wie die Variablen zusammenhängen, nur ob Sie mit den erklärenden Variablen ihre Vorhersage verbessern können.

# Inferenz vs. Vorhersage

Die induktive Statistik lässt sich in zwei Hauptblöcke unterteilen:

## Modellierung zur Erklärung (Inferenz)

Hier versuchen Sie explizit den Zusammenhang zwischen ihrer zu untersuchenden Variablen  $y$  und einer (oder mehreren) erklärenden Variablen  $x$  zu beschreiben und zu quantifizieren. Oberstes Ziel wäre es einen **kausalen** Zusammenhang zwischen den Variablen aufzudecken (falls möglich).

## Modellierung zur Vorhersage

Hier versuchen Sie eine zu untersuchende Variable  $y$  durch mehrere erklärende Variablen  $x$  **vorherzusagen**. Es interessiert Sie hierbei nicht so sehr wie die Variablen zusammenhängen, nur ob Sie mit den erklärenden Variablen ihre Vorhersage verbessern können.

Wir konzentrieren uns in diesem Kurs auf die **Inferenz** und versuchen *kausale* Zusammenhänge aufzudecken.

Falls Sie sich für die *Vorhersage* interessieren, so könnten Sie bspw. den Kurs "Economics with R" (im Bachelor, von Herr Prof. Kranz), oder "Data Analysis" (im Master, von Herrn Prof. Gebhardt) anhören.

# Induktive Statistik

Wann sind tatsächlich *kausale* Interpretationen ihrer Ergebnisse zulässig?

Dieser Frage wenden wir uns nun im zweiten Teil des Projektkurses!

# Induktive Statistik

Wann sind tatsächlich *kausale* Interpretationen ihrer Ergebnisse zulässig?

Dieser Frage wenden wir uns nun im zweiten Teil des Projektkurses!

Allgemein gibt es hier zwei Arten von Daten, die wir für die Analyse heranziehen können:

## Experimentelle Daten

Experimente können vom Forscher bewusst gestaltet und beeinflusst werden. Viele Problematiken, wie z.B. die Selektionsproblematik oder ausgelassene Variablen können durch das Design des Experiments und die Randomisierung eliminiert werden.

## Beobachtungsdaten (nicht-experimentelle Daten)

Beobachtungsdaten werden meist retrospektiv erhoben und der Forscher ist nicht am Prozess der Datengenerierung beteiligt. Für die kausale Interpretation der Ergebnisse aus diesen Daten müssen externe Gegebenheiten herangezogen werden, die eine gewisse Art von Zufall im Datengenerierungsprozess erzeugen. Meist steht hier nur eine Stichprobe der Daten zur Verfügung.

# Randomisiertes Experiment

Wie können wir *kausale* Aussagen aus unseren vorliegenden Daten tätigen?

- + Optimalzustand ist ein tatsächliches randomisiertes Experiment
- + Der zu untersuchende Effekt kann *unverzerrt* geschätzt werden
  - + D.h. anderen Einflussgrößen, welche dazu führen könnten, dass ihr Effekt systematisch falsch geschätzt wird können ausgeschlossen werden

**Problem:** In der ökonomischen Forschung ist es sehr schwer oder teuer randomisierte Experimente durchzuführen, deshalb arbeiten wir sehr oft mit Beobachtungsdaten.

# Randomisierung

Warum ist es wichtig, dass die Teilnehmer eines Experiments **zufällig** auf Kontroll und Treatmentgruppe zugeteilt werden?

Durch die zufällige Zuteilung werden alle sonstigen Störgößen eliminiert. Dadurch können Sie mit Hilfe der Wahrscheinlichkeitstheorie Aussagen darüber treffen, ob ihr gemessenes Ergebnis durch reinen Zufall zustande kam oder ob es sich um ein signifikantes Ergebnis handelt.

D.h. bei einem *randomisierten* Experiment können Sie einen Hypothesentest durchführen und Aussagen zur statistischen Signifikanz ihres Ergebnisses tätigen.

Wie genau, das wollen wir uns nun an einem Beispiel anschauen!



# Randomisiertes Experiment

Wir wollen uns nun ein Beispiel für ein randomisiertes Experiment anschauen und wie Sie Hypothesen testen können.

# Randomisiertes Experiment

Wir wollen uns nun ein Beispiel für ein randomisiertes Experiment anschauen und wie Sie Hypothesen testen können.

Wie rational und konsistent ist die Verhaltensweise des typischen amerikanischen College Studenten?

# Randomisiertes Experiment

Wir wollen uns nun ein Beispiel für ein randomisiertes Experiment anschauen und wie Sie Hypothesen testen können.

Wie rational und konsistent ist die Verhaltensweise des typischen amerikanischen College Studenten?

Hat es einen Effekt, wenn Sie einen College Studenten daran erinnern, dass er Geld für später sparen kann?

# Randomisiertes Experiment

Wir wollen uns nun ein Beispiel für ein randomisiertes Experiment anschauen und wie Sie Hypothesen testen können.

Wie rational und konsistent ist die Verhaltensweise des typischen amerikanischen College Studenten?

Hat es einen Effekt, wenn Sie einen College Studenten daran erinnern, dass er Geld für später sparen kann?

**Studie durchführt von:**

Frederick S, Novemsky N, Wang J, Dhar R, Nowlis S. 2009. Opportunity Cost Neglect. Journal of Consumer Research 36: 553-561.

# Teilnehmer und Durchführung

- + 150 Teilnehmer
- + Zufällige Zuteilung in Treatment und Kontroll-Gruppe

# Teilnehmer und Durchführung

- + 150 Teilnehmer
- + Zufällige Zuteilung in Treatment und Kontroll-Gruppe

Alle Teilnehmer bekamen folgenden Text vorgelegt:

Imagine that you have been saving some extra money on the side to make some purchases, and on your most recent visit to the video store you come across a special sale on a new video. This video is one with your favorite actor or actress, and your favorite type of movie (such as a comedy, drama, thriller, etc.). This particular video that you are considering is one you have been thinking about buying for a long time. It is available for a special sale price of \$14.99.

What would you do in this situation? Please circle one of the options below.

# Entscheidungsmöglichkeiten

Entscheidungsmöglichkeiten der Kontrollgruppe:

- + Buy this entertaining video.
- + Not buy this entertaining video.

# Entscheidungsmöglichkeiten

Entscheidungsmöglichkeiten der Kontrollgruppe:

- + Buy this entertaining video.
- + Not buy this entertaining video.

Entscheidungsmöglichkeiten der Treatment-Gruppe:

- + Buy this entertaining video
- + Not buy this entertaining video. Keep the \$14.99 for other purchases.



# Ergebnis

Hat die Nennung des Offensichtlichen, d.h. das die \$14.99 für andere Einkäufe berücksichtigt werden können, eine Konsequenz für die Studenten?

# Ergebnis

Hat die Nennung des Offensichtlichen, d.h. das die \$14.99 für andere Einkäufe berücksichtigt werden können, eine Konsequenz für die Studenten?

**In absoluten Zahlen:**

<b>gruppe</b>	<b>kaufen</b>	<b>nicht kaufen</b>
kontrolle	56	19
treatment	41	34

# Ergebnis

Hat die Nennung des Offensichtlichen, d.h. das die \$14.99 für andere Einkäufe berücksichtigt werden können, eine Konsequenz für die Studenten?

**In absoluten Zahlen:**

<b>gruppe</b>	<b>kaufen</b>	<b>nicht kaufen</b>
kontrolle	56	19
treatment	41	34

**Als Prozentwerte:**

	<b>kaufen</b>	<b>nicht.kaufen</b>
<b>Kontrolle</b>	0.747	0.253
<b>Treatment</b>	0.547	0.453

# Ergebnis

Definition von **Erfolg** in diesem Experiment: Studenten, welche die DVD *nicht* gekauft haben.

# Ergebnis

Definition von **Erfolg** in diesem Experiment: Studenten, welche die DVD *nicht* gekauft haben.

Der Unterschied zwischen Kontroll- und Treatment Gruppe ergibt sich dann als:

$$\hat{p}_{treatment} - \hat{p}_{kontrolle} = 0.453 - 0.253 = 0.200$$

# Ergebnis

Definition von **Erfolg** in diesem Experiment: Studenten, welche die DVD *nicht* gekauft haben.

Der Unterschied zwischen Kontroll- und Treatment Gruppe ergibt sich dann als:

$$\hat{p}_{treatment} - \hat{p}_{kontrolle} = 0.453 - 0.253 = 0.200$$

**Ergebnis: Bestätigt!**

Ist der Unterschied zwischen der Kontrollgruppe und Treatment Gruppe signifikant?

# Hypothesentest

Bei dem Ergebnis der Studenten ist Skepsis angebracht.

Bedenken Sie, dass Sie die Studenten *zufällig zu einer Gruppe zugeordnet haben*, daher unterliegt dieses Ergebnis auch einem *Zufallsprozess*.



# Hypothesentest

Bei dem Ergebnis der Studenten ist Skepsis angebracht.

Bedenken Sie, dass Sie die Studenten *zufällig zu einer Gruppe zugeordnet haben*, daher unterliegt dieses Ergebnis auch einem *Zufallsprozess*.

$$H_0 : \hat{p}_{treatment} = \hat{p}_{kontrolle}$$

**wahr ist**, dann besteht kein Zusammenhang zwischen dem Erinnerungstext und den Handlungen der Studenten.

**Achtung:** Sie könnten aus Zufall mehr Studenten in ihrer Treatment Gruppe haben, welche die DVD sowieso nicht kaufen wollten. Diesen Sachverhalt nennen wir die *Nullhypothese*.

# Hypothesentest

Hypothesen werden meist explizit benannt um sie testen zu können. Hierbei beschreibt  $H_0$  die Nullhypothese und  $H_1$  die Alternativhypothese.

In unserem Fall ergibt sich:

# Hypothesentest

Hypothesen werden meist explizit benannt um sie testen zu können. Hierbei beschreibt  $H_0$  die Nullhypothese und  $H_1$  die Alternativhypothese.

In unserem Fall ergibt sich:

**Null Hypothese (  $H_0$  ):**

Studenten daran zu erinnern, dass Sie Geld für späteren Konsum sparen können hat keinen Einfluss auf ihre Ausgabenentscheidung.

**Alternativhypothese (  $H_1$  ):**

Studenten daran zu erinnern, dass Sie Geld für späteren Konsum sparen können reduziert die Wahrscheinlichkeit das diese mit einem Einkauf fortfahren.

# Hypothesentest

Sie testen folglich:

$$H_0 : p_{treatment} - p_{control} = 0$$

$$H_1 : p_{treatment} - p_{control} > 0$$

# Hypothesentest

Sie testen folglich:

$$H_0 : p_{treatment} - p_{control} = 0$$

$$H_1 : p_{treatment} - p_{control} > 0$$

Nullhypothesen werden als Gegenteil der Alternativhypothese formuliert (häufig in dem Sinn, dass *kein* Effekt vorliegt). Teststatistiken sind so konzipiert, dass die Nullhypothese abgelehnt werden kann und die Alternativhypothese als Erklärung übrig bleibt.

**Achtung:** Wenn die Nullhypothese nicht abgelehnt werden kann, so bedeutet dies *nicht* das sie zutrifft!

# Hypothesentest

Stellen Sie sich ein Paralleluniversum vor, in dem die Erinnerung für die Studenten bzgl. der \$14.99 keinen Effekt auf deren Kaufentscheidung hat. Wie müsste hier die Verteilung innerhalb der Gruppen aussehen?

Wenn die Gruppenzugehörigkeit nicht relevant wäre, dann könnten Sie die Kaufentscheidung zufällig den jeweiligen Gruppen zuteilen, und diese neue Zuteilung hätte keine Konsequenz für ihre Ergebnisse!

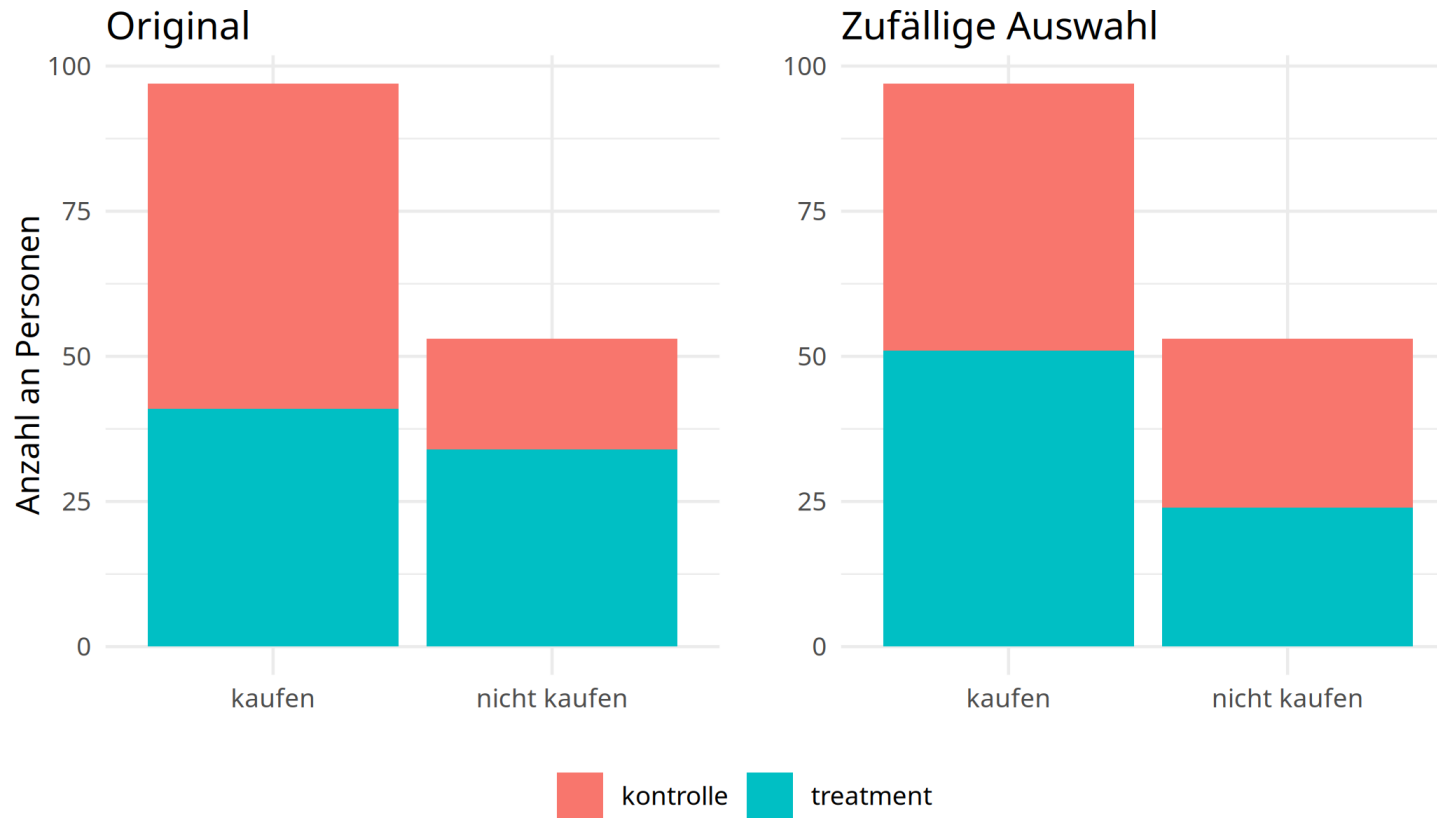
# Hypothesentest

Stellen Sie sich ein Paralleluniversum vor, in dem die Erinnerung für die Studenten bzgl. der \$14.99 keinen Effekt auf deren Kaufentscheidung hat. Wie müsste hier die Verteilung innerhalb der Gruppen aussehen?

Wenn die Gruppenzugehörigkeit nicht relevant wäre, dann könnten Sie die Kaufentscheidung zufällig den jeweiligen Gruppen zuteilen, und diese neue Zuteilung hätte keine Konsequenz für ihre Ergebnisse!

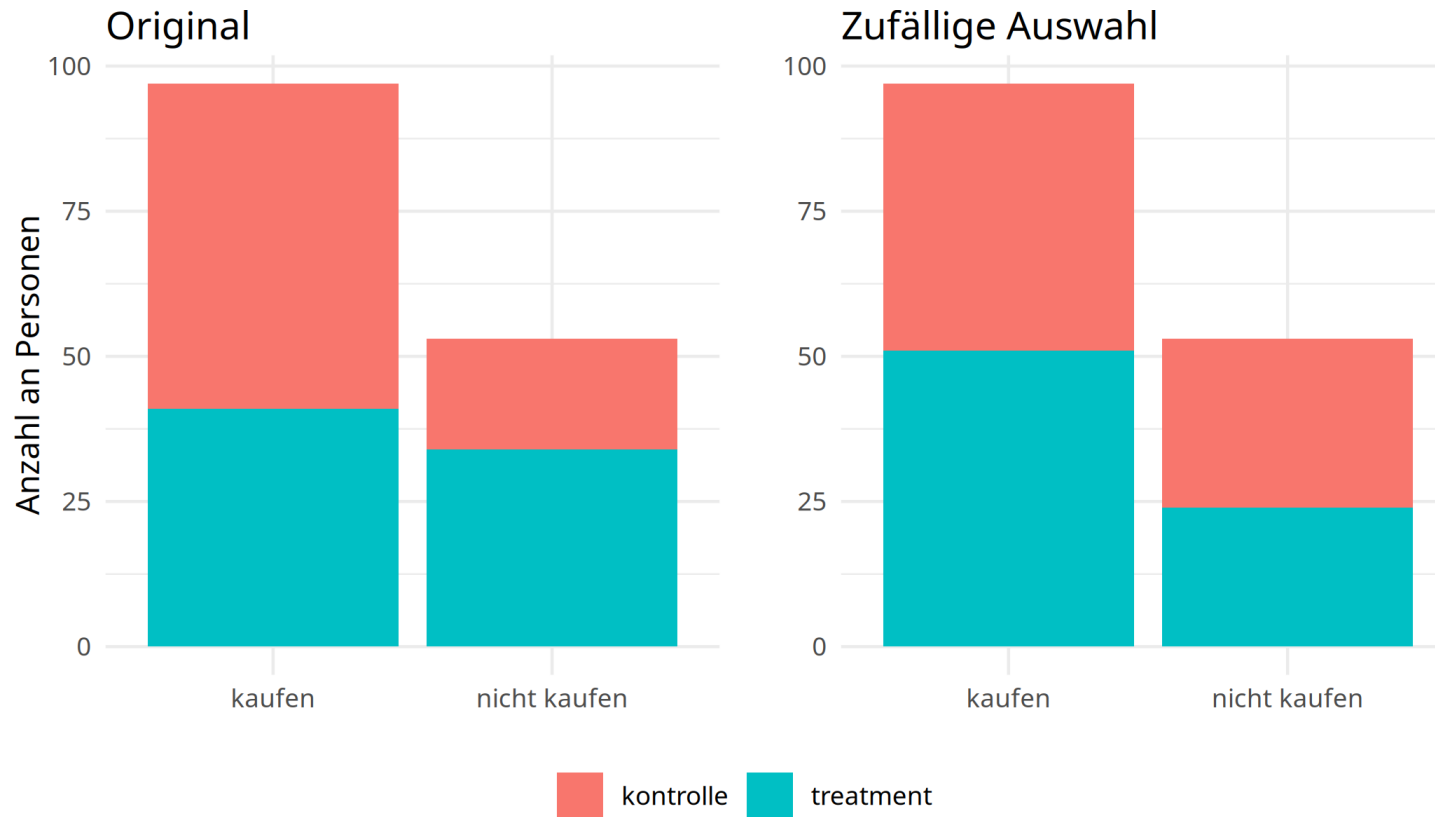
Lassen Sie uns dies einmal machen und die Auswirkungen betrachten

# Hypothesentest



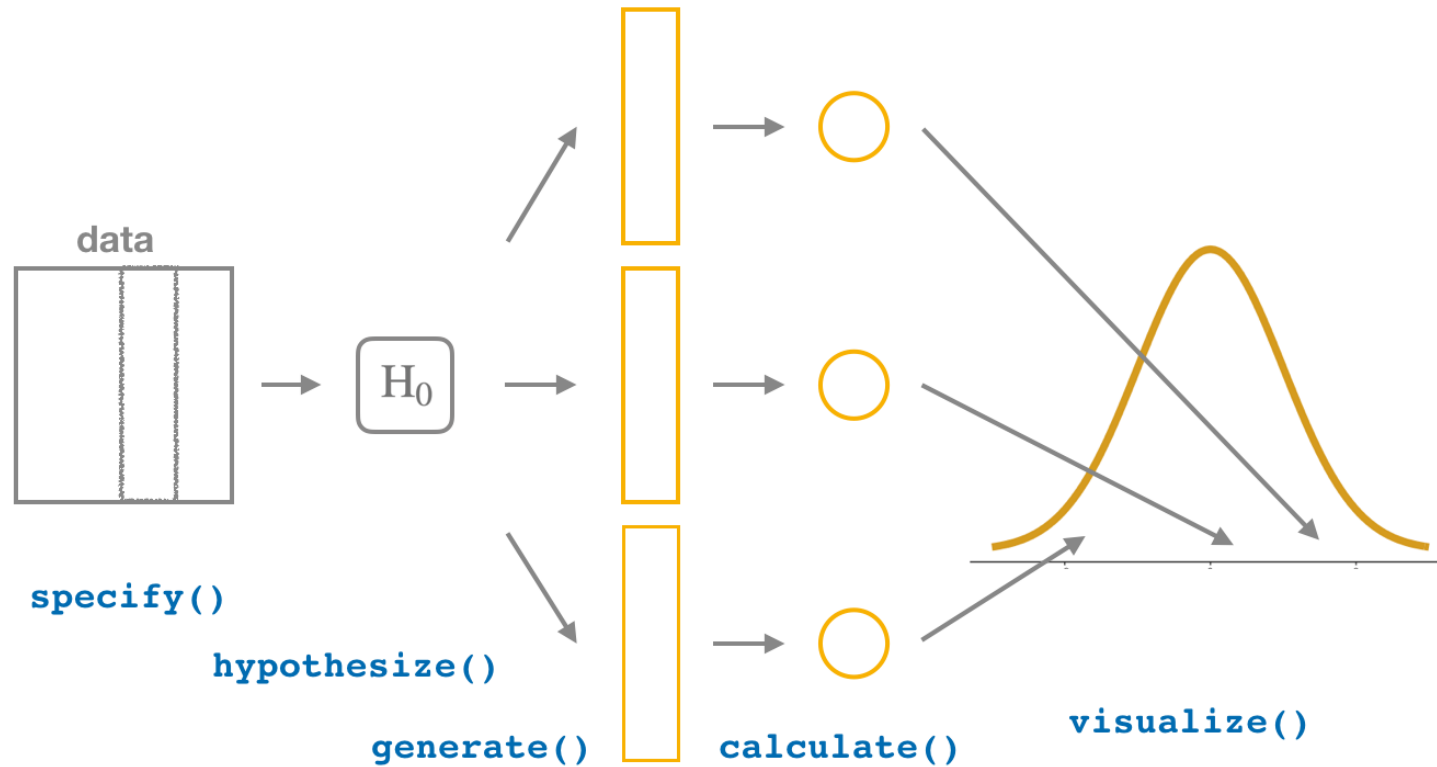


# Hypothesentest



Die zufällige Zuteilung sieht deutlich homogener aus, als die, welche durch das Experiment gemessen wurde.

# HYPOTHESENTEST MIT DEM `INFER` PAKET



Quelle: Bild von <https://moderndive.com/index.html>

# Hypothesentest

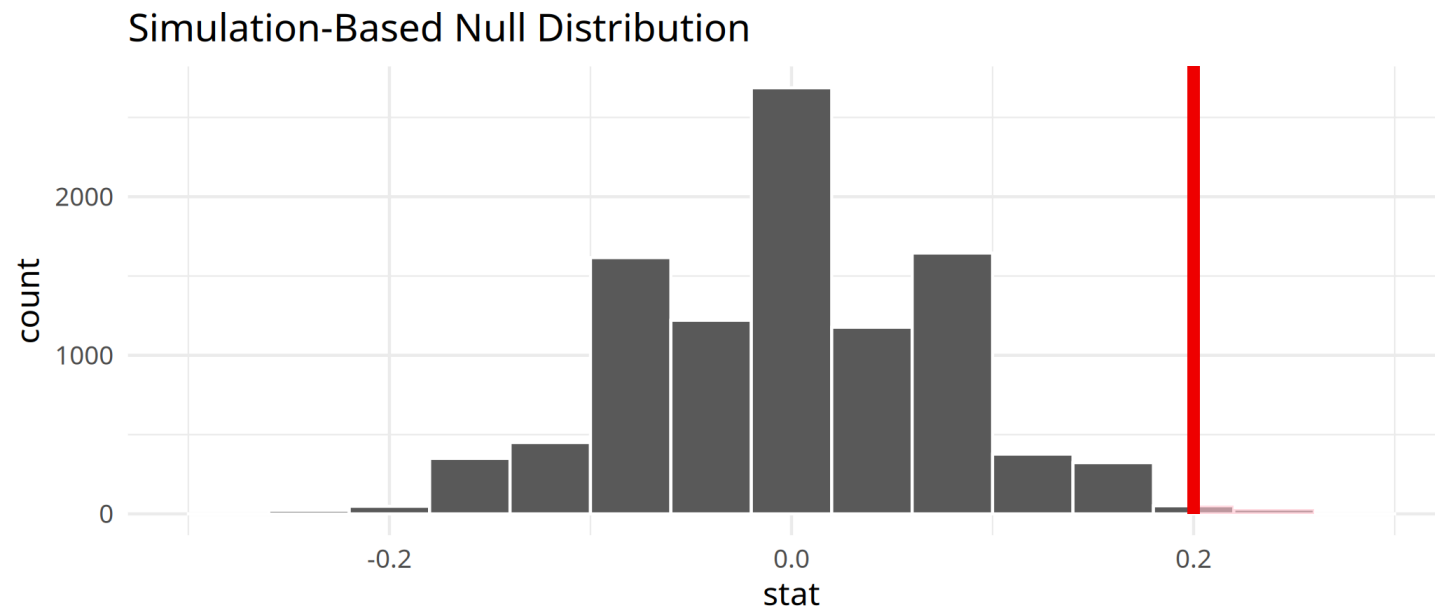
Nun haben Sie die Zuteilung zu den Gruppen *einmal* randomisiert. Jedoch möchten Sie gerne eine Verteilung erhalten, wie bei unserem Beispiel der Stichprobe und dem Bootstrap, die sogenannte *Nullverteilung*

Anhand dieser Verteilung (10000 Züge) können Sie abschätzen, wie wahrscheinlich der gemessene Effekt rein statistisch ist, d.h. wie oft dieser Effekt allein durch Zufall auftritt:

# Hypothesentest

Nun haben Sie die Zuteilung zu den Gruppen *einmal* randomisiert. Jedoch möchten Sie gerne eine Verteilung erhalten, wie bei unserem Beispiel der Stichprobe und dem Bootstrap, die sogenannte *Nullverteilung*

Anhand dieser Verteilung (10000 Züge) können Sie abschätzen, wie wahrscheinlich der gemessene Effekt rein statistisch ist, d.h. wie oft dieser Effekt allein durch Zufall auftritt:



# Eingefärbte Fläche

Die eingefärbte Fläche in der vorherigen Grafik zeigt, wo der tatsächlich gemessene Wert in unserer Verteilung liegt.

- + Visualisierung wie wahrscheinlich die beobachtete Differenz von 0.2 rein zufällig gemessen wurde
- + Effekt ist sehr unwahrscheinlich nur durch Zufall gemessen worden, d.h. wir können  $H_0$  verwerfen

Durch den sogenannten p-Wert können Sie abschätzen, wie wahrscheinlich es ist eine solchen Wert nur durch Zufall zu erhalten, gegeben das die Nullhypothese zutrifft.

# p-Werte

Der p-Wert quantifiziert die Evidenz gegen die Nullhypothese.

# p-Werte

Der p-Wert quantifiziert die Evidenz gegen die Nullhypothese.

## Exakter p-Wert:

```
null_distribution %>%  
  get_p_value(obs_stat = obs_diff_prop, direction = "right")
```

```
# A tibble: 1 × 1  
  p_value  
  <dbl>  
1 0.0082
```

**Statistische Signifikanz:** Das Signifikanzlevel zu dem wir eine Hypothese testen wird meist mit  $\alpha$  abgekürzt. Wir sprechen von einer Signifikanz zum x% Level, wenn der p-Wert geringer ist als x.

# p-Werte

Der p-Wert quantifiziert die Evidenz gegen die Nullhypothese.

## Exakter p-Wert:

```
null_distribution %>%  
  get_p_value(obs_stat = obs_diff_prop, direction = "right")
```

```
# A tibble: 1 × 1  
  p_value  
  <dbl>  
1 0.0082
```

**Statistische Signifikanz:** Das Signifikanzlevel zu dem wir eine Hypothese testen wird meist mit  $\alpha$  abgekürzt. Wir sprechen von einer Signifikanz zum x% Level, wenn der p-Wert geringer ist als x.

Ein Effekt wird meist als statistischer Signifikanz angesehen, wenn der p-Wert kleiner als  $(\alpha) = 0.05$



# Klassische nicht parametrische Test

Anstatt die zuvor gezeigte Methode des Bootstrap finden Sie in vielen empirischen Ausarbeitungen sogenannte *nichtparametrische* Tests, wie z.B. den sogenannten t-Test. Mit Hilfe dieses t-Tests können Sie untersuchen ob die Mittelwerte der beiden Gruppen signifikant voneinander unterschiedlich sind:

```
experiment %>%  
  mutate(dvd_f = ifelse(dvd == "kaufen", 1, 0)) %>%  
  t.test(dvd_f ~ gruppe, data = .)
```

```
Welch Two Sample t-test  
  
data:  dvd_f by gruppe  
t = 2.6027, df = 145.38, p-value = 0.01021  
alternative hypothesis: true difference in means between group kontrolle and group treatment is r  
95 percent confidence interval:  
 0.04812293 0.35187707  
sample estimates:  
mean in group kontrolle mean in group treatment  
      0.7466667      0.5466667
```

# Klassische nicht parametrische Test

Hier erhalten wir eine t-Statistik von 2,6 mit einem p-Wert von 0,01. D.h. es ist sehr unwahrscheinlich, dass unser Effekt durch Zufall zustande kam.

# Zusammenhang von Konfidenzintervallen, t-Statistik und p-Werten

Die Verbindung von p-Werten zu Konfidenzintervallen ergibt sich indirekt.

Wenn das Konfidenzintervall von 95% *nicht* die 0 beinhaltet, dann muss der p-Wert kleiner als 0,05 sein.

Die t-Statistik des Koeffizienten ist in diesem Fall größer als 1,96 (ergibt sich aus der Normalverteilung).

# Probleme des p-Wertes

Sie sollten sich **nicht** allein auf den p-Wert verlassen!

p-Werte sind ein Indikator dafür, in wie fern die Daten zu einem spezifischen statistischen Modell passen

# Probleme des p-Wertes

Sie sollten sich **nicht** allein auf den p-Wert verlassen!

p-Werte sind ein Indikator dafür, in wie fern die Daten zu einem spezifischen statistischen Modell passen

- + Ein p-Wert ist *nicht* die Wahrscheinlichkeit das die Nullhypothese wahr ist, oder die Alternativhypothese falsch
  - + Der p-Wert zeigt an, wie kompatibel die Daten mit einer bestimmten Hypothese sind
- + Der p-Wert gibt an, ob die Ergebnisse durch Zufall erklärt werden können, er gibt jedoch keine Auskunft über die Korrektheit der Hypothese
- + Das Signifikanzlevel von 0.05 ist eine Konvention
  - + Gerade noch signifikante sind qualitativ nicht besser wie gerade noch insignifikante Resultate, auch wenn diese Grenze in der Literatur oft verwendet wird
- + Der p-Wert sagt nichts über die ökonomische Bedeutung eines Effekts aus.

# Probleme des p-Wertes

Sie sollten sich **nicht** allein auf den p-Wert verlassen!

p-Werte sind ein Indikator dafür, in wie fern die Daten zu einem spezifischen statistischen Modell passen

- + Ein p-Wert ist *nicht* die Wahrscheinlichkeit das die Nullhypothese wahr ist, oder die Alternativhypothese falsch
  - + Der p-Wert zeigt an, wie kompatibel die Daten mit einer bestimmten Hypothese sind
- + Der p-Wert gibt an, ob die Ergebnisse durch Zufall erklärt werden können, er gibt jedoch keine Auskunft über die Korrektheit der Hypothese
- + Das Signifikanzlevel von 0.05 ist eine Konvention
  - + Gerade noch signifikante sind qualitativ nicht besser wie gerade noch insignifikante Resultate, auch wenn diese Grenze in der Literatur oft verwendet wird
- + Der p-Wert sagt nichts über die ökonomische Bedeutung eines Effekts aus.

Je größer die Stichprobe, desto einfacher finden wir statistisch signifikante Effekte.

# Fehlertypen

Sie können die Nullhypothese auf Grund von statistischen Tests auf einem bestimmten Signifikanzniveau entweder ablehnen oder auch nicht.

Hierbei können Sie jedoch Fehler begehen:

<b>Entscheidung</b>	<b><math>H_0</math> ist korrekt</b>	<b><math>H_1</math> ist korrekt</b>
Entscheidung für $H_0$	Korrekte (Spezifität)	Fehler 2. Art
Entscheidung für $H_1$	Fehler 1. Art	Korrekte (Sensitivität)

Fehler 1. Art: Sie lehnen die Nullhypothese ab, obwohl sie korrekt ist.

Fehler 2. Art: Sie lehnen die Nullhypothese *nicht* ab, obwohl sie falsch ist

# Fehlertypen

Beispiel zu Fehler 1. Art:

- ✚ Im Beispiel mit den Studenten würde ein Fehler 1. Art bedeuten, dass Sie schlussfolgern die Erinnerung der Studenten an die Sparmöglichkeit hat keinen Einfluss auf deren Verhalten, trotz des Ergebnisses aus den Daten.
- ✚ Eine Bank vergibt einen Kredit an einen eigentlich nicht kreditwürdigen Kunden. Der Kredit fällt aus und die Bank hat einen Fehler 1. Art begangen. (Tatsächlicher Verlust)



# Fehlertypen

Beispiel zu Fehler 1. Art:

- Im Beispiel mit den Studenten würde ein Fehler 1. Art bedeuten, dass Sie schlussfolgern die Erinnerung der Studenten an die Sparmöglichkeit hat keinen Einfluss auf deren Verhalten, trotz des Ergebnisses aus den Daten.
- Eine Bank vergibt einen Kredit an einen eigentlich nicht kreditwürdigen Kunden. Der Kredit fällt aus und die Bank hat einen Fehler 1. Art begangen. (Tatsächlicher Verlust)

Beispiel zu Fehler 2. Art:

- Wenn die Erinnerung der Studenten an die Sparmöglichkeit in Wirklichkeit keine Auswirkung auf ihre Entscheidung hat, wir jedoch auf Grund der Daten die Nullhypothese verwerfen begehen wir einen Fehler 2. Art.
- Eine Bank kommt nach ihrer Analyse zu dem Schluss, dass der Kunde nicht kreditwürdig ist und vergibt somit den Kredit nicht. Hätte die Bank jedoch besser hin geschaut, dann hätte Sie erkannt, dass der Kunde kreditwürdig gewesen ist Die Bank begeht hier einen Fehler 2. Art. (Opportunitätskosten)

# Fehlertypen

Ziel ist es ein Testverfahren mit hoher Sensitivität und hoher Spezifität zu erhalten. Die Konsequenzen eines Fehlers 1. Art sind meist größer als die eines Fehlers 2. Art.

# Signifikanzlevel und Fehler 1. und 2. Art

- ✚ Wenn Sie besonders über den Fehler 1. Art besorgt sind, dann sollten Sie ein hohes Signifikanzniveau wählen, z.B.  $\alpha < 0.01$ .
  - ✚ In dem Fall wollen Sie besonders vorsichtig sein die Nullhypothese zu verwerfen. D.h. Sie brauchen sehr starke Evidenz für  $H_1$  um  $H_0$  zu verwerfen
- ✚ Wenn Sie eher darüber besorgt sind einen Fehler 2. Art zu machen, dann sollten Sie ein niedriges Signifikanzniveau wählen, z.B.  $\alpha = 0.10$ 
  - ✚ Hier wollen Sie vorsichtig sein  $H_0$  nicht zu verwerfen, wenn die Nullhypothese eigentlich falsch ist