Panini Album

Markus Kalisch

Einführung

Statistik ist die Lehre vom Umgang mit quantitativen Daten. Oft verschleiern Zufall oder Unsicherheiten einen Sachverhalt. Ein Ziel der Statistik ist, in diesen Fällen klare Aussagen zu machen.

Betrachten wir ein Beispiel. Vor der Fussball Weltmeisterschaft 2010 wurden Panini-Sammelalben für 661 verschiedene Aufkleber von Fussballstars und -stadien verkauft. Jedes Bild hat eine Nummer von 1 bis 661 auf der Rückseite. Die Aufkleber gab es entweder in einzelnen 5er Packs oder in ganzen Kisten mit hundert 5er Packs zu kaufen. Gemäss Panini sind alle Bilder gleich häufig und werden wohl zufällig auf die Tüten verteilt.

Vor dieser WM hatten Kollegen von mir den Eindruck, dass man überraschend wenige doppelte Bilder hat, wenn man nicht einzelne 5er Packs sondern eine ganze Kiste mit hundert 5er Packs kauft. Sind die Bilder in einer Kiste vielleicht nicht zufällig verteilt worden?

Beobachteter Wert: Anzahl eingeklebter Bilder

Bei der Fussball Weltmeisterschaft 2010 wollte ich diese Vermutung prüfen. Ich habe eine Kiste mit hundert 5er Packs gekaufte (500 Bilder) und alle in ein leeres Album geklebt. Ich konnte 477 Bilder einkleben, der Rest waren doppelte Bilder. Sind das nun "überraschend wenige doppelte" oder nicht? Ist es also billiger eine Kiste statt einzelne 5er Packs zu kaufen? Mit der Intuition ist diese Frage praktisch nicht zu beantworten, aber die Statistik liefert eine Lösung.

Was erwarten wir, wenn nur der Zufall am Werk ist ?

Angenommen, die Bilder werden tatsächlich zufällig auf die Tüten verteilt. Ist es plausibel, dass man unter 500 Bildern 477 oder mehr Bilder hat, die nur einmal vorkommen? Anders gefragt: Wie wahrscheinlich ist es, dass man mit zufälligem Eintüten 477 oder noch mehr aus 500 Bildern in ein leeres Album einkleben kann? Wenn diese Wahrscheinlichkeit sehr klein ist (kleiner als eine von uns festgelegte Grenze, z.B. eins zu einer Million $= 10^{-6}$), dann sind die Bilder wohl nicht zufällig eingetütet worden.

Nun, probieren wir es einfach aus! Mit dem Computer habe ich 500 Zahlen zufällig aus der Menge 1,...,661 mit Zurücklegen gezogen (jede Zahl steht für ein Panini-Bild; jedes Bild kann mehrfach gezogen werden).

```r
set.seed(123)
N <- 661 ## Anzahl aller Bilder
n <- 500 ## Aufkleber in der Box
bilder <- sample(1:N, n, replace = TRUE) ## Ziehe Bilder mit Zuruecklegen
eingeklebt <- length(unique(bilder))
```

Dabei hatte ich 346 Zahlen, die nur einmal vorkommen. In diesem zufällig gefüllten fiktiven Album hätte ich also nur 346 einfache Bilder einkleben können. Das ist weniger als die beobachteten 477 Bilder, aber vielleicht könnte ich ja bei dem nächsten zufällig gefüllten Album viel mehr Bilder einkleben.

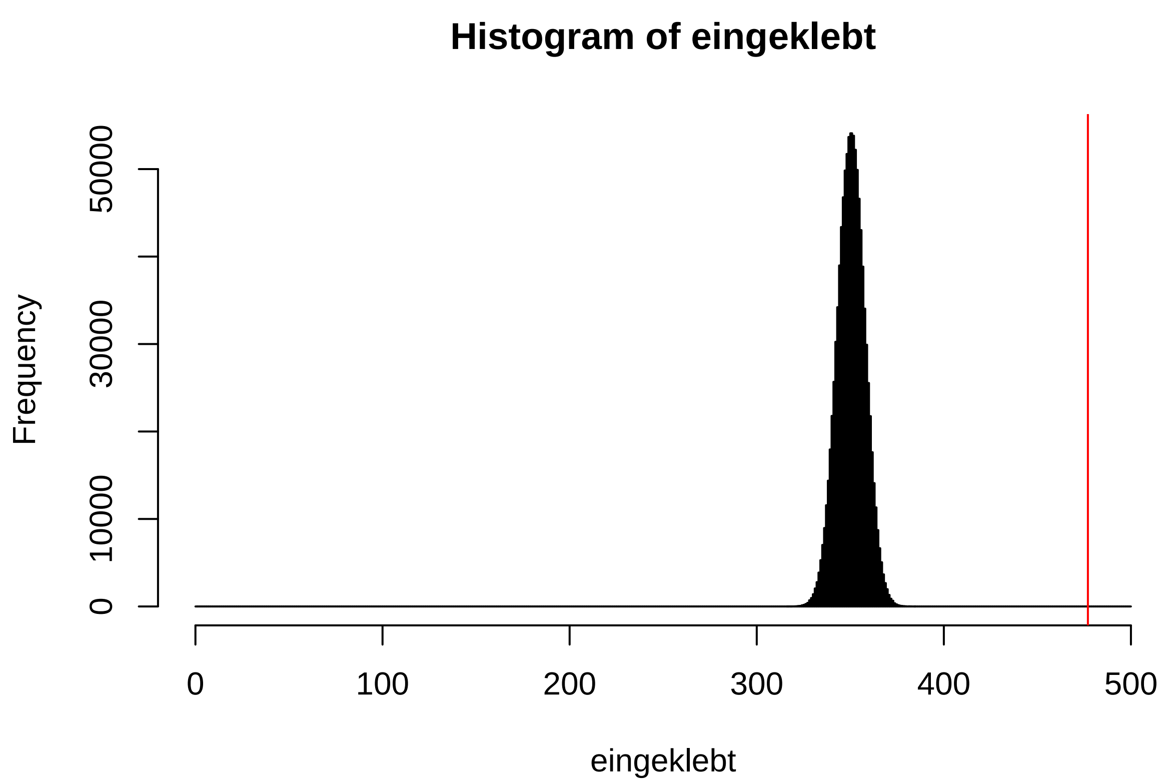
Um herauszufinden, ob 477 einfache Bilder durch Zufall entstehen können, muss ich also noch mehr fiktive Alben füllen. Mit dem Computer habe ich eine Million fiktive Alben zufällig gefüllt.

```r
set.seed(234)
N <- 661
n <- 500
reps <- 10^6
eingeklebt <- rep(NA, reps)

for (i in 1:reps) {
  bilder <- sample(1:N, n, replace = TRUE)
  eingeklebt[i] <- length(unique(bilder))
}
```

Wir stellen die gefundenen Anzahlen der eingeklebten Bilder aus der Simulation mit einem Histogramm dar. Dann zeichnen wir die Anzahl eingeklebter Bilder im realen Albung als vertikale Linie ein:

```r
hist(eingeklebt, breaks = 0:n) ## Histogramm der Anzahl eingeklebter Bilder in Simulation
abline(v = 477, col = "red") ## Zeige Anzahl Runs in der selbst erzeugten Sequenz
```



Schlussfolgerung

Die grösste Anzahl einfacher Bilder war dabei 387. Falls die Bilder wirklich zufällig eingetütet werden, ist die Wahrscheinlichkeit mit einer Kiste mehr als 387 eingeklebte Bilder zu haben also kleiner als 10^{-6}.

Das lässt folgenden Schluss zu: Wenn ich ein Album mit 500 zufällig ausgewählten Bildern füllen würde, könnte ich nicht mal in einer Million Alben so viele Bilder einkleben, wie ich in Realität beobachtet habe.

Anders gesagt: In einer Welt, in der die Panini-Bilder zufällig verteilt werden, ist die Wahrscheinlichkeit 477 oder mehr einfache Bilder in einer Kiste zu finden also kleiner als eine Million. Die Hypothese "Die 500 Bilder werden zufällig verteilt" und die Beobachtung "Ich konnte 477 Bilder einkleben" passen also nicht zusammen. Die Hypothese wird deshalb verworfen: Die Panini-Bilder werden nicht zufällig eingetütet und es scheint vorteilhaft zu sein eine ganze Kiste zu kaufen.

Ausblick

In der Statistik ist das eben geschilderte Vorgehen fundamental und wird **Hypothesentest** genannt. Ein Hypothesentest besteht aus sechs Schritten:

1. Man stellt ein **Modell** auf, das erklärt, wie die Daten entstehen. (Wir ziehen 500 Bilder mit Zurücklegen aus einer Menge von 661 Bildern.)
2. Man stellt eine **Hypothese** (häufig auch **Nullhypothese**[1] genannt) auf. ("Panini Bilder in einer Kiste werden *zufällig* aus 661 Bildern mit Zurücklegen gezogen und eingetütet".) Wenn die Nullhypothese nicht zutrifft, muss die sog. **Alternative** zutreffen. ("Panini Bilder werden *nicht zufällig*, sondern nach irgendeinem Muster gezogen und eingetütet".)
3. Man überlegt sich eine **Teststatistik** (Anzahl Bilder, die man in ein leeres Album einkleben kann, wenn man 500 Bilder hat) und die **Wahrscheinlichkeitsverteilung der Teststatistik** unter der Annahme, dass die Nullhypothese stimmt. (Welche Anzahlen von einfachen Bildern sind plausibel, wenn zufällig 500 Bilder aus 661 mit Zurücklegen gezogen werden? Ich habe die Frage mit dem Computer beantwortet; oft kann man analytische Resultate finden.)
4. Man legt ein **Signifikanzniveau** fest. (Festgelegte Grenze von 10^{-6})
5. Man bestimmt einen **Verwerfungsbereich** für die Teststatistik zum gegebenen Signifikanzniveau. (Angenommen, die Bilder werden zufällig eingetütet. In der Computersimulation haben wir gesehen, dass die Wahrscheinlichkeit mehr als 387 eingeklebte Bilder zu haben kleiner 10^{-6} ist. Der Verwerfungsbereich ist also $\{388, 389, ..., 500\}$.)[2]
6. Man **beobachtet einen Wert der Teststatistik** (477 eingeklebte Bilder) und fällt den **Testentscheid**: Verwerfe die Nullhypothese, falls der beobachtete Wert in dem Verwerfungsbereich liegt. (477 ist in der Menge $\{388, 389, ..., 500\}$. Daher wird die Nullhypothese verworfen.)

Manchmal wird eine kleine Variante des obigen Verfahrens verwendet. Man berechnet den **P-Wert**: Angenommen die Nullhypothese stimmt. Wie wahrscheinlich ist es eine Teststatistik zu beobachten, die mindestens so extrem ist wie die beobachtete Teststatistik? Die Antwort liefert der P-Wert. (Wie wahrscheinlich ist es, in meiner Computersimulation 477 oder mehr einfache Bilder zu erhalten? Ich habe es in einer Million Simulationen kein einziges Mal beobachtet. Also ist die Wahrscheinlichkeit sicher[3] kleiner als $\frac{1}{1000000} = 10^{-6}$. Also ist $p < 10^{-6}$.) Der P-Wert muss nicht immer so gross sein wie das Signifikanzniveau. Ich habe mein Beispiel so gewählt, dass beide Zahlen gleich sind. Üblicherweise wählt man als Signifikanzniveau $\alpha = 0.05$.

Wenn man eine Nullhypothese nach obiger Methode verwirft, kann man sich natürlich irren. Es könnte ja sein, dass die Bilder zufällig eingetütet werden und ich hatte einfach unglaubliches Glück und konnte 477 Bilder einkleben. Dann würde ich die Nullhypothese fälschlicherweise verwerfen. Dieses Ereignis tritt aber mit einer Wahrscheinlichkeit ein, die höchstens so gross ist wie das Signifikanzniveau bzw. der P-Wert. D.h., die Wahrscheinlichkeit, die Nullhypothese fälschlicherweise zu verwerfen, ist höchstens so gross wie das Signifikanzniveau bzw. der P-Wert. Man nennt diesen Fehler auch **Fehler 1. Art** [4].

[1] Die Namensgebung ist eher Konvention und macht für Sie im Augenblick wohl wenig Sinn. Oft wird ein Medikament getestet. Die Hypothese ist, dass das Medikament gar nicht wirkt. Der Effekt des Medikaments ist also *Null*. Daher wird die Hypothese dann "Nullhypothese" genannt. Dieser Sprachgebrauch hat sich auch in anderen Bereichen durchgesetzt.

[2] Ich habe mich hier für einen einseitigen Test entschieden; es gibt auch zweiseitige Tests. Später erfahren Sie mehr darüber.

[3] Für diejenigen, die es ganz genau wissen wollen: Da wir die Wahrscheinlichkeit mit einer Computersimulation ermittelt haben, stimmt das Ergebnis nur approximativ. In unserem Fall ist der Unterschied aber so deutlich, dass das keine praktische Rolle spielt.

[4] Es gibt auch noch den **Fehler 2. Art**. Er tritt dann auf, wenn man die Nullhypothese beibehält, obwohl sie eigentlich

falsch ist.