STOCHASTIK

Testen von Hypothesen

Teil 1

Grundlagen der Signifikanztests

Berechnungen mit der Binomialverteilung

Datei Nr. 35010

Stand: 7. Februar 2019

FRIEDRICH W. BUCKEL

INTERNETBIBLIOTHEK FÜR SCHULMATHEMATIK

https://mathe-cd.de

Vorwort

Das Testen von Hypothesen gehört zur Beurteilenden Statistik. Ich zeige hier, wie man die vier wichtigen Fälle behandelt, und welche grundlegenden Fragestellungen es dazu gibt.

Es wurde auch noch ein Abschnitt angefügt, in dem $\sigma-$Intervalle des Erwartungswertes als Annahmebereich einer Hypothese verwendet werden. Dazu benötigt man aber oft Tabellen.

Ein Problem für diesen Text sind die unterschiedlichen Fähigkeiten der benötigten Rechner. Ich biete daher oftmals mehrere Erklärungen dazu an.

Grundkenntnisse über die Binomialverteilung werden vorausgesetzt. (Texte 34010 ff.)

Inhalt

§ 1 Um was geht es? – Wichtige Infos…

In unserem Alltag begegnen uns vielfach Behauptungen bzw. Angaben, die wir glauben können, die man aber auch auf ihre Richtigkeit hin überprüfen kann.

a) Ein Medikament gegen Bluthochdruckt wirkt bei <u>mindestens</u> 60% der an Bluthochdruck erkrankten Personen.

b) In einer Kiste mit Glaskugeln beträgt der Anteil der roten Kugeln <u>entweder</u> 25% <u>oder</u> 40%.

c) Der Anteil der defekten Schrauben in einer Kiste beträgt <u>höchstens</u> 5%.

d) Der Spielautomat liefert mit <u>genau</u> 50% Wahrscheinlichkeit Gewinnspiele.

Da eine totale Kontrolle aller Patienten / Glaskugeln / Schrauben / Gewinnspiele nicht realisierbar bzw. zu kostenintensiv ist, führt man immer einen Test durch. Dazu wählt man eine **Stichprobe** aus und zählt dabei die „Objekte“, welche die Behauptung bestätigen oder widerlegen.

Ein solches Testergebnis wird immer vom Zufall gesteuert. Wenn im Beispiel b) in einer Kiste mit Glaskugeln 40% rote sind, und wir greifen zufällig (also blind) 200 Kugeln (das ist dann der Umfang der Stichprobe) heraus, dann **kann man erwarten, dass 40% davon rot sind, also 80**. Diese Zahl nennt man den **Erwartungswert.** Man berechnet ihn durch die Formel: $E = n \cdot p = 200 \cdot 0{,}4 = 80$.
80 rote Kugeln unter 200 zu finden, ist natürlich reine Spekulation, denn wir können <u>zufälligerweise</u> auch 95 rote finden, oder aber auch nur 70, je nachdem wie wir hineingreifen.

Wie wahrscheinlich solche Ziehungsergebnisse sind, kann man berechnen. Da beim Ziehen aus der Kiste die roten Kugeln laut Angabe mit der konstanten Wahrscheinlichkeit $p = 0{,}4$ vorkommen, stellt das Ziehen von 200 Kugeln eine **Bernoulli-Kette** dar. Die Wahrscheinlichkeit der einzelnen Ergebnisse wird daher mit der **Binomialverteilung** berechnet. <u>Rechnen wir also nach:</u>
X sei die Anzahl der roten Kugeln unter 200 gezogenen mit $p_{rot} = 0{,}4$. Dann erhält man

a) genau 80 rote Kugeln mit $P(X = 80) = 0{,}0575 \approx 6\%$

b) genau 95 rote Kugeln mit $P(X = 95) = 0{,}0056 \approx 0{,}6\%$

c) genau 70 rote Kugeln mit $P(X = 70) = 0{,}0205 \approx 2\%$

```
binomialPDf(80,200,0.4)
                         0.0575
binomialPDf(95,200,0.4)
                         0.0056
binomialPDf(70,200,0.4)
                         0.0205
```

Daraus erkennt man, dass man rein zufällig mehr als 80 rote Kugeln finden kann, aber auch weniger.

Annahme oder Ablehnung?

Weicht die Anzahl der gefundenen Objekte deutlich (die Mathematiker sagen dazu „signifikant“) vom Erwartungswert ab, dann wird man die Annahme oder Behauptung ablehnen, also der Behauptung, dass 40% der Kugeln rot sind, nicht glauben. Und dennoch kann sie wahr sein, auch wenn uns der Zufall eine deutliche Abweichung beschert hat. Zufällige Schwankungen bringen es also mit sich, dass dabei **Fehler** passieren können, und zwar folgende:
Man kann eine wahre Behauptung versehentlich ablehnen, man kann aber auch eine falsche Behauptung versehentlich glauben! Dies werden wir uns genauer ansehen, und später werden wir sogar lernen, die Fehlerwahrscheinlichkeit zu berechnen.

Spiegelt das Testergebnis die Realität wider?

Wir wollen eine Lieferung testen. Die Abbildung (eine Vierfeldertafel) zeigt, dass es dabei zu vier unterschiedlichen Situationen kommen kann. Dazu definiere ich diese Ereignisse:

G = Wir haben eine gute Lieferung erhalten, $\overline{G}$ = …eine schlechte Lieferung.

Und das ist möglich:

Test–ergebnis \ Realität	G	$\overline{G}$
G	Überein-stimmung	Fehler 2. Art Versehentliche Annahme
$\overline{G}$	Fehler 1. Art Versehentliche Ablehnung	Überein-stimmung

(1) **Realität und Testergebnis können übereinstimmen.**
Dann ist alles in Ordnung.
Das kann in zwei Fällen passieren:
Die Ware ist gut und der Test bestätigt es.
Die Lieferung ist schlecht, und wir finden das auch durch das Testergebnis heraus.

(2) **Liefert der Test jedoch ein von der Realität abweichendes Ergebnis**, dann ist einer dieser beiden Fehler passiert:

Beim **Fehler 1. Art** lehnt man eine gute Lieferung versehentlich ab.
Beispiel: Die Stichprobe aus einer guten Lieferung enthält zufällig zu viele defekte Schrauben.

Beim **Fehler 2. Art** nimmt man die Lieferung versehentlich an, obwohl sie in Wirklichkeit sehr fehlerhaft ist.
Beispiel: Das Medikament gegen Bluthochdruck doch nicht so wirksam ist wie behauptet, aber das Testergebnis präsentiert uns zufällig viele positive Wirkungen.

Es gibt dazu im Wesentlichen zwei Aufgabenstellungen:

Führt man einen Test durch, benötigt man eine ***Entscheidungsregel***, die uns sagt, welches Testergebnis zur Annahme bzw. Ablehnung der zu testenden Hypothese führt.

1. Aufgabe: Die Entscheidungsregel (für Annahme bzw. Ablehnung) ist gegeben. Dann wird verlangt, dass man die Güte dieses Tests überprüft, indem man ausrechnet, mit welchen Wahrscheinlichkeiten die Fehler 1. Art bzw. 2. Art auftreten. Das sagt dann auch etwas über die Güte des Tests aus.

2. Aufgabe: Man soll selbst die Grenzen für Annahme bzw. Ablehnung ermitteln, die Entscheidungsregel also festlegen. Dabei soll man darauf achten, dass der Fehler 1. Art nicht zu groß wird.

Dabei kommt es zu interessanten Überlegungen, denn der Verkäufer hat andere Erwartungen an ein Testergebnis als der Kunde, er wird also seinen Test anders bewerten!

2.2 Rechtsseitige Signifikanztests

Beim **<u>linksseitigen</u>** <u>Signifikanztest</u> lautet eine Nullhypothese etwa $p \geq 0{,}8$ (mindestens 80%).
Diese wird also angenommen, wenn man hinreichend große Testergebnisse findet, und folglich wird sie abgelehnt, wenn man **signifikant wenige** findet.

<u>Der Ablehnungsbereich liegt also **links**</u>: z. B. so: $S = \{\underbrace{0,\ldots,35}_{\bar{A}} \mid \underbrace{36,\ldots,50}_{A}\}$

Beim **<u>rechtsseitigen</u>** <u>Signifikanztest</u> lautet eine Nullhypothese etwa $p \leq 0{,}6$ (höchstens 60%).
Diese wird also angenommen, wenn man hinreichend wenige Testergebnisse findet, und folglich wird sie abgelehnt, wenn man **signifikant viele** findet.

<u>Der Ablehnungsbereich liegt also **rechts**</u>: z. B. so: $S = \{\underbrace{0,\ldots,14}_{A} \mid \underbrace{15,\ldots,20}_{\bar{A}}\}$

Man kann es sich so merken:

Behauptet man, dass p <u>mindestens</u> ($\geq$) …. ist,
wird die Nullhypothese für kleine Werte von X abgelehnt: <u>Linksseitiger</u> Test.
Behauptet man, dass p <u>höchstens</u> ($\leq$) …. ist,
wird die Nullhypothese für große Werte von X abgelehnt: <u>Rechtsseitiger</u> Test.

Auch hier gibt es verschiedene Aufgabenstellungen:

1. Fall: Die Entscheidungsregel ist gegeben, Fehler 1. Art (und ggf. 2. Art) wird berechnet,

2. Fall: Die Entscheidungsregel ist so aufzustellen, dass ein vorgegebenes Signifikanzniveau eingehalten wird.

3. Fall: Die Entscheidungsregel ist aufzustellen, so dass z. B. das Gesamtrisiko (also die Summe der Fehler 1. und 2. Art) minimal wird.

Hinweis: Es gibt immer wieder Aufgabenstellungen, die sehr unklar formuliert sind. Dann weiß man nicht sofort, welchen Ansatz man machen soll, einen rechts-seitigen oder linksseitigen Test.

Im Abschnitt 2.3 zeige ich einige Beispiele dazu und gebe Tipps, wie man dann einen Lösungsweg findet.

Beispiel 4: Gegebene Entscheidungsregel – Bewertung des Tests

Bei Spielautomaten der Sorte K dürfen laut Verordnung nur höchstens 60% der Spiele verlorene Spiele sein. Ein amtlicher Prüfer testet einen Spielapparat so: Er will 20-mal spielen und den Automat dann sperren lassen, wenn er mehr als 15-mal verliert.

a) Berechne die Annahmewahrscheinlichkeit sowie die Wahrscheinlichkeit für den Fehler 1. Art.

b) Wie groß ist der Fehler 2. Art, wenn der Apparat tatsächlich in 70 % aller Fälle keinen Gewinn ausschüttet?

Wie würden Sie den Test einschätzen?

Musterlösung

1. Testumfang: n = 20
2. Nullhypothese: $p \le 0,6$
3. Testvariable: X = Zahl der verlorenen Spiele

   X ist binomial verteilt mit p = 0,6
4. Erwartungswert: $E(X) = n \cdot p = 12$
5. Definitionsbereich für X: $S = \{\underbrace{0\,;\,1\,;\,...\,;\,\boxed{12}\,;\,...\,;\,15}_{A} \,|\, \underbrace{16\,;\,...\,;\,20}_{\bar{A}}\}$
6. Bewertung des Tests:

   Annahmewahrscheinlichkeit: $P(A) = P(X \le 15) = F_B(15;20;0,6) = 0,9490$

   Fehler 1. Art: $\alpha = P(\bar{A}) = P(X \ge 16) = 0,0510$

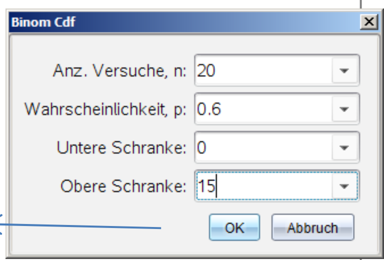
   (Versehentliche Ablehnung).

   Fehler 2. Art: (Versehentliche Annahme bei p = 0,7)

   $\beta = P(A) = P(X \le 15) = F_B(15;20;0,7) = 0,7625$.

   (TI Nspire:)

binomCdf(20,0.6,0,15)	0.949048
binomCdf(20,0.6,16,20)	0.050952
binomCdf(20,0.7,0,15)	0.762492



Bei diesem Test spielt der Fehler 1. Art keine große Rolle, da es um den Schutz der Kunden geht. Für sie ist es wichtig, dass sie eine möglichst hohe Sicherheit darin haben, dass kein betrügerischer Automat ihre Gewinne behält. Unter diesem Aspekt ist der Test sehr schlecht, denn β ist viel zu groß. Also müsste der Annahmebereich kleiner gemacht werden.

Beispiel 5: Entscheidungsregel festlegen

Beim 2. Automaten spielt er 50-mal. Wie muss er die Grenzen des Ablehnungsbereiches festlegen, wenn das Signifikanzniveau 10 % betragen soll?

Muss er bei 38 verlorenen Spielen den Apparat einziehen?

Musterlösung

1. Nullhypothese H_O: $p \le 0{,}6$

   Man geht bei dieser Nullhypothese davon aus, dass der Apparat in Ordnung ist.
2. Testumfang: $n = 50$
3. Testvariable: X = Zahl der verlorenen Spiele

   X ist binomialverteilt mit p = 0,6 und hat die Definitionsmenge $S = \{0\,;\,...\,;\,50\}$
4. Erwartungswert: $E(X) = n \cdot p = 50 \cdot 0{,}6 = 30$
5. Annahme- und Ablehnungsbereich: $S = \{\underbrace{0\,;1\,;...\,;\boxed{30}\,;...\,;L}_{A}\,|\,\underbrace{R\,;...\,;50}_{\bar{A}}\}$

   In dieser Menge stehen zwei gesuchte Zahlen L (links und rechts), man kann auch k und k+1 verwenden. Sie bilden dann die Entscheidungsregel.
6. Festlegung von R durch die Bedingung: Das **Signifikanzniveau soll 10 %** betragen, d.h. für den Fehler 1. Art soll gelten: $\alpha = P(\bar{A}) = P(X \ge R) \le 0{,}1$

   Man benötigt hierzu die Umkehrung der Binomialverteilungsfunktion, denn man kennt den Funktionswert 0,1 und sucht den X-Wert, also R.

Verschiedene Wege zur Lösung dieser Ungleichung.

1. Berechnung mit einer Tabelle

Man benötigt die Tabelle der Verteilungsfunktion $F_B(x,n,p)$ zu n = 50. Und speziell für p = 0,6.
Das ist nun gerade eine etwas schwierigere Aufgabe, denn die Tabellen sind meist so angelegt, dass man für p-Werte über 0,5 den unteren und rechten Tabelleneingang verwenden muss:
Und dann muss man den gesuchten wert durch Subtraktion von 1 berechnen.

Verteilungsfunktion $F_B(x;n;p)$ der Binomialverteilung

n	x \ p	0,05	0,10	1/6	0,20	0,25	0,30	0,35	0,40	0,45	0,50	
	16	1,0000	1,0000	0,9978	0,9856	0,9017	0,6839	0,3889	0,1561	O,0427	0,0077	33
	17	1,0000	1,0000	0,9992	0,9937	0,9449	0,7822	0,5060	0,2369	0,0765	0,0164	32
	18	1,0000	1,0000	0,9997	0,9975	0,90713	0,8594	0,3356	0,3356	0,1273	ß.ß325	31

p = 0,6

Man liest ab, dass $P(X \le 33) > 0{,}1$ aber $P(X \le 34) < 0{,}1$ ist. (Ist hier nicht zu sehen.)

2. Möglichkeit mit einem Rechner, der Binomialfunktionen berechnen kann.

a) Lösung durch Probieren (Screenshot TI Nspire CAS)

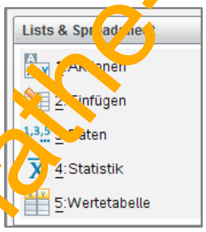
Dann kann man beispielsweise so vorgehen, dass man sich so lange einige Werte berechnen lässt, bis man auf die Zahl den Wert 0,1 stößt, was zwischen 34 und 35 passiert.
Hier eine Liste von solchen Funktionswerten:

binomCdf(50,0.6,33,50)	0.236876
binomCdf(50,0.6,34,50)	0.156091
binomCdf(50,0.6,35,50)	0.095502

b) Lösung mit einer Rechner-Wertetafel (Screenshot: TI Nspire CAS)

Man definiert eine Binomialverteilungsfunktion
wechselt dann zu „Lists & Spreadsheet" und
lässt sich dort eine Wertetafel anzeigen

Define $f(x)$=binomCdf(50,0.6,x,50) *Fertig*



x	f(x):=
	binomCdf..
32.	0.335613
33.	0.236876
34.	0.156091
35.	0.095502
36.	0.053955
37.	0.027988

Wählt man diese Methode für CASIO ClassPad II,
dann empfehle ich eine explizite Folge zu erzeugen, so wie ich das auf Seite 11 gezeigt habe.

3. Möglichkeit mit einem Rechner, der die inverse Binomialfunktion kennt.

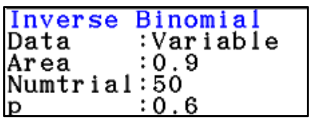
Neue Rechner besitzen eine Berechnungsmöglichkeit für die **inverse Funktion** zu BinomialCDf.
Damit kann man eine <u>**Gleichung**</u> **(nicht Ungleichung) der Form** $P(X \le k) = \alpha$ **lösen,**
d. h. man gibt n, p und α **ein und erhält dann k.**

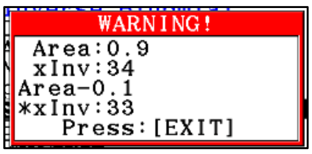
WARNUNG:

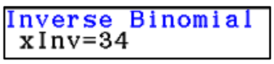
Diese inverse Binomialfunktion bezieht sich auf die Funktion, die $P(X \le k)$ **berechnet. In unserer Aufgabe heißt sie aber** $P(X \ge R)$**.**
Also muss man zuvor diese Umformung machen:
$P(X \ge R) < 0{,}1 \quad \Leftrightarrow \quad 1 - P(X \le L) < 0{,}1 \quad \Leftrightarrow \quad P(X \le L) > 0{,}9$

Wir wollen also die <u>Gleichung</u> $\text{BinomialCDf}(0, x, 50, 0.6) = 0.9$ lösen lassen:

Screenshot vom Grafikrechner GTR CASIO fx CG 20:







Da hier zwei L-Werte ein Ergebnis nahe bei 0,1 liefern, muss man eine Probe machen und zwei Werte berechnen (siehe ganz oben). Man findet: L = 34.

$$\text{Erg.:} \quad A = \{0; ..; \underset{=L}{34}\} \text{ bzw. } \overline{A} = \{\underset{=R}{35}; ...; 50\}$$

Ergebnis: Man erkennt, dass ab $R \ge 35$ diese Wahrscheinlichkeit kleiner als 10% ist.
Die angegebenen 38 verlorenen Spiele führen also zum Einzug des Gerätes.