Prof. Dr. Klaus Eckhardt

Binomialtest

veröffentlicht im Internet unter aufgabomat.de

Inhalte: Bernoulli-Experiment, Bernoulli-Kette, binomialverteilte Zufallsvariable, Erwartungswert einer binomialverteilten Zufallsvariable, Nullhypothese, Alternativhypothese, Signifikanzniveau, Annahmeintervall, links- und rechtsseitiger Test

Gliederung:

1 Binomialverteilte Zufallsvariablen

Ein Zufallsexperiment mit genau zwei möglichen, sich gegenseitig ausschließenden Ausgängen bezüglich eines Ereignisses A wird **Bernoulli-Experiment** genannt. Die Wahrscheinlichkeit für das Eintreten des Ereignisses wird üblicherweise abkürzend mit p bezeichnet, die Wahrscheinlichkeit dafür, dass das Ereignis nicht eintritt, mit q: $P(A) = p$, $P(\overline{A}) = 1 - p = q$.

Eine Folge voneinander unabhängiger Bernoulli-Experimente mit konstanter Wahrscheinlichkeit p bildet eine so genannte **Bernoulli-Kette**. Die Häufigkeit, mit der das Ereignis A in einer Bernoulli-Kette der Länge n eintritt, ist eine **binomialverteilte Zufallsvariable**, die hier mit K bezeichnet wird. Ihre **Wahrscheinlichkeitsfunktion** berechnet sich als

$$\begin{aligned} f(k) &= P(K = k) \\ &= \binom{n}{k} p^k q^{n-k} \end{aligned} \qquad (1)$$

mit $n \in \mathbb{N}^+$ und $k \in \mathbb{N}$. Die Werte f(k) der Wahrscheinlichkeitsfunktion entsprechen der Wahrscheinlichkeit, dass ein Ereignis der Wahrscheinlichkeit p bei n-maliger Wiederholung eines Bernoulli-Experiments genau k-mal eintritt.

Die Werte F(k) der **Verteilungsfunktion** geben dagegen die Wahrscheinlichkeit an, dass eine binomialverteilte Zufallsvariable K einen Wert annimmt, der bis zu k beträgt:

$$F(k) = P(K \le k) = \sum_{i=0}^{k} f(i) = \sum_{i=0}^{k} \binom{n}{i} p^i q^{n-i}. \quad (2)$$

Die Werte beider Funktionen werden von den Parametern n und p der Binomialverteilung bestimmt. Der Binomialtest dient dazu, aus Messwerten bzw. Beobachtungs- oder Befragungsdaten auf den Parameter p zu schließen.

2 Linksseitiger Binomialtest

2.1 Aufgabe

Eine Wahlumfrage wird durchgeführt. Zweihundert der befragten wahlberechtigten Personen geben an, wählen gehen zu wollen. Darunter sprechen sich fünfzehn Personen für die Partei XYZ aus. Das letzte Wahlergebnis der Partei XYZ lag bei 10 %. Deutet sich eine Änderung ihres Stimmenanteils an?

In der Stichprobe beträgt die Zustimmung 7,5 %. Es interessiert aber nicht speziell das Wahlverhalten in dieser relativ kleinen Stichprobe von lediglich zweihundert Personen, sondern das Wahlverhalten in der Grundgesamtheit der gesamten Wählerschaft von mehreren Millionen Menschen. Von der Stichprobe auf die Grundgesamtheit zu schließen, ist die Aufgabe eines statistischen Tests.

2.2 Wie verteilt ist die Zufallsvariable?

Würde die Befragung mit einer anderen Gruppe von zweihundert Personen durchgeführt, wäre nicht unbedingt dasselbe Ergebnis zu erwarten. Die Anzahl derjenigen Personen, die sich in einer beliebigen Stichprobe des Umfangs $n = 200$ für die Partei XYZ aussprechen werden, ist eine Zufallsvariable, die grundsätzlich die Werte von 0 bis 200 annehmen kann.

Um den geeigneten statistischen Test auszuwählen, muss zuerst geklärt werden, welcher Typ von Zufallsvariable vorliegt bzw. durch welche Wahrscheinlichkeitsverteilung die Zufallsvariable beschrieben wird. Im Beispiel stellt die Befragung einer einzelnen Person ein Bernoulli-Experiment dar, denn die betreffende Person wird entweder für die Partei XYZ stimmen oder nicht. Unter der Voraussetzung, dass die befragten Personen ihre Meinung unabhängig voneinander äußern, liegt eine Bernoulli-Kette der Länge $n = 200$ vor. Die Anzahl derjenigen Personen, die sich bei einer Befragung für die Partei XYZ aussprechen, stellt damit eine binomialverteilte Zufallsvariable dar.

Die binomialverteilte Zufallsvariable mit $n = 200$ und $p = 0{,}10$ wird nachfolgend mit K bezeichnet. Ihre Wahrscheinlichkeitsfunktion ist in Abbildung 1 dargestellt, ihre Verteilungsfunktion in Abbildung 2. Die Abszisse ist in beiden Fällen nur bis $k = 40$ ausgeführt, da sich für größere k keine sichtbaren Veränderungen der Funktionswerte ergeben.

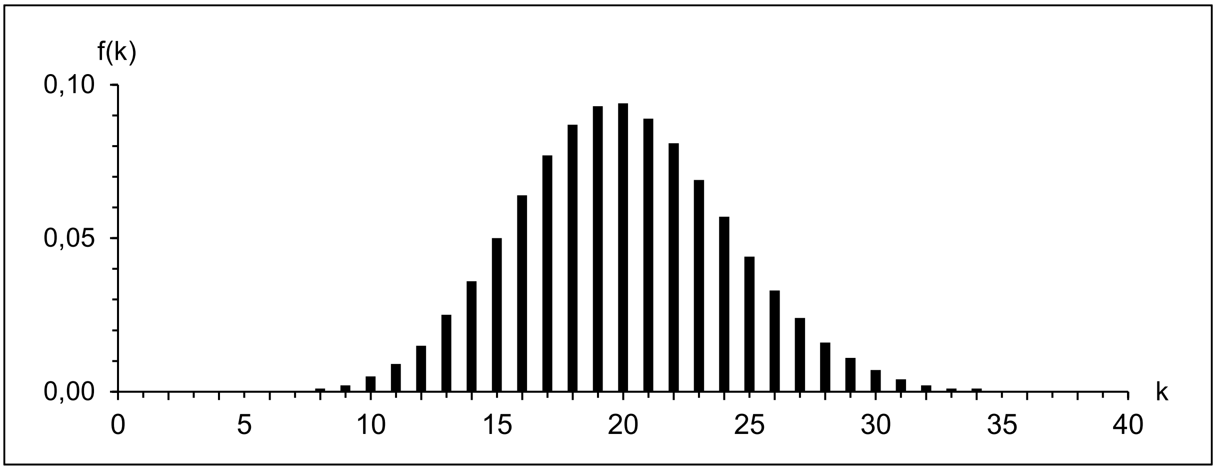

Abbildung 1: Wahrscheinlichkeitsfunktion der binomialverteilten Variable K mit n = 200 und p = 0,10.

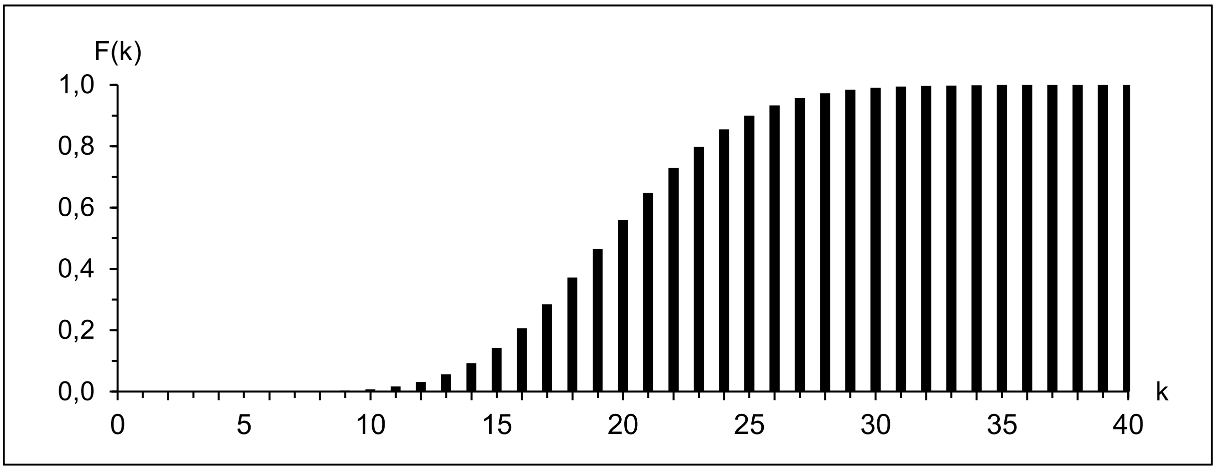

Abbildung 2: Verteilungsfunktion der binomialverteilten Variable K mit n = 200 und p = 0,10.

2.3 Wie lauten Null- und Alternativhypothese des Tests?

Wie viele Stimmen für die Partei XYZ hätte man in der vorliegenden Stichprobe erwartet? Unter der Annahme, dass die Zustimmungsrate immer noch 10 % beträgt, wären es 10 % der 200 Stimmen gewesen, also 20 Stimmen. Dies ist der so genannte **Erwartungswert** E(K) der Zufallsvariable K.

Stattdessen waren es nur 15 Stimmen. Dies erweckt den Verdacht, dass sich die Zustimmung für die Partei verringert haben könnte. Dieser Verdacht ($p < 0{,}10$) wird zur so genannten **Alternativhypothese** des Tests gemacht. Jeder statistische Test ist mit der Angabe von Null- und Alternativhypothese einzuleiten. Anstatt diese beiden Begriffe auszuschreiben, ist es auch üblich, die abkürzenden Bezeichnungen H_0 und H_1, manchmal auch H_A, zu verwenden (mit H für Hypothese). Die **Nullhypothese** ist das logische Gegenteil der Alternativhypothese. Das logische Gegenteil von "kleiner" ist "größer oder gleich", sodass sich im vorliegenden Fall ein statistischer Test mit den folgenden Hypothesen ergibt:

H_0: $p \geq 0{,}10$
H_1: $p < 0{,}10$.

2.4 Anhand welchen Kriteriums wird zwischen Null- und Alternativhypothese entschieden?

Je geringer die Zustimmung in der Stichprobe ist, desto wahrscheinlicher ist es, dass sich der Stimmenanteil der Partei XYZ in der Grundgesamtheit der Gesamtwählerschaft verringert hat. Umgekehrt gilt: Je größer der Wert der Zufallsvariable K ist, desto wahrscheinlicher ist es, dass die Nullhypothese des Tests zutrifft.

Wie groß muss K mindestens sein, damit man von der Gültigkeit der Nullhypothese ausgehen kann?

Um diese Frage zu entscheiden, wird die Verteilungsfunktion der Variable K herangezogen (Abschnitt 1). Die Verteilungsfunktion ist, mathematisch gesprochen, streng monoton wachsend. Je größer k ist, desto größer ist der zugehörige Wert F(k) (Abbildung 2). Dem Kriterium „Die Nullhypothese wird beibehalten, falls die Zufallsvariable K einen gewissen Mindestwert annimmt.“ entspricht das Kriterium „Die Nullhypothese wird beibehalten, falls die Verteilungsfunktion F(k) einen gewissen Mindestwert annimmt.“.

Dieser Mindestwert der Verteilungsfunktion muss zu Beginn des Tests festgelegt werden. Man nennt ihn das **Signifikanzniveau** des Tests und bezeichnet ihn üblicherweise abkürzend mit dem kleinen griechischen Alpha (α). Der in der Praxis meist verwendete Wert für α ist 0,05, ohne dass dies objektiv begründet wäre. Das Signifikanzniveau kann auch größer oder kleiner als 0,05 gewählt werden. Dieser Aspekt wird hier allerdings nicht weiter beleuchtet.

Ist das Signifikanzniveau festgelegt, so ist im nächsten Schritt derjenige Wert k_α zu bestimmen, den die Zufallsvariable K mindestens annehmen muss, damit man wegen $F(k) \geq \alpha$ bei der Nullhypothese bleibt. Dies ist die untere Begrenzung des so genannten **Annahmeintervalls**. Dieser Name rührt daher, dass immer dann, wenn der beobachtete Wert der Zufallsvariable innerhalb dieses Intervalls liegt, die Nullhypothese angenommen bzw. als gültig angesehen wird.

Als Signifikanzniveau wird im vorliegenden Beispiel $\alpha = 0{,}05$ gewählt. Die Verteilungsfunktion F(k) hat unter anderem die folgenden Werte: F(12) = 0,032, F(13) = 0,057, F(14) = 0,093, ….[1] Der kleinste Wert, für den $F(k) \geq 0{,}05$ ist, ist folglich 13. Die untere Begrenzung des Annahmeintervalls ist $k_{0,05} = 13$.[2]

Die obere Begrenzung des Annahmeintervalls ist durch den Maximalwert der Variable K von 200 festgelegt. Es ergibt sich somit das Annahmeintervall [13; 200]. Das heißt, immer dann, wenn sich bei der Befragung von zweihundert wählenden Personen mindestens 13 Personen für die Partei XYZ aussprechen, schließt man daraus, dass die Zustimmung für die Partei nicht geringer als 0,10 geworden ist bzw. dass die Nullhypothese des Tests gilt. Da sich im Beispiel k = 15 und damit $k \geq k_{0,05}$ ergeben hat, ist dieses Kriterium erfüllt. Als Ergebnis des Tests wird gefolgert, dass die Nullhypothese gilt oder, wie man auch sagt, beibehalten werden kann.

Da die untere bzw. linke Begrenzung des Annahmeintervalls zu bestimmen war, liegt ein so genannter **linksseitiger Test** vor.

In Kurzform ließe sich der Test wie folgt dokumentieren:

K: binomialverteilte Zufallsvariable mit n = 200, p = 0,10

[1] Werte der Wahrscheinlichkeits- und der Verteilungsfunktion der Binomialverteilung lassen sich mithilfe eines Tabellenkalkulations- oder Statistikprogramms berechnen. In Excel beispielsweise dient dazu die Funktion BINOM.VERT.

[2] In Excel beispielsweise kann dieser Wert auch mit der Funktion BINOM.INV berechnet werden. Formel: =BINOM.INV(200;0,10;0,05).

$k = 15$

Linksseitiger Binomialtest:

H_0: $p \geq 0{,}10$
H_1: $p < 0{,}10$
$\alpha = 0{,}05$

Untere Begrenzung des Annahmeintervalls: $k_{0,05} = 13$
$k \geq k_{0,05} \Rightarrow H_0$ wird beibehalten.

3 Rechtsseitiger Binomialtest

3.1 Aufgabe

Eine Krebserkrankung konnte bislang mit einer Wahrscheinlichkeit von 70 % geheilt werden. An vierhundert erkrankten Personen wird eine neue Therapie erprobt. Von diesen vierhundert Personen genesen dreihundertzwanzig, ein Anteil von 80 %. Kann die neue Therapie als wirksamer angesehen werden?

3.2 Durchführung des Tests

Die Behandlung einer einzelnen Person ist ein Bernoulli-Experiment: Sie bleibt krank oder wird gesund. Die Anzahl der genesenden Personen in einer Gruppe von vierhundert Erkrankten ist eine binomialverteilte Zufallsvariable. Die binomialverteilte Zufallsvariable mit $n = 400$ und $p = 0{,}70$ wird nachfolgend mit K bezeichnet.

Die Wahrscheinlichkeit p einer Heilung scheint mit der neuen Therapie im Vergleich zum bisherigen Wert von 0,70 gestiegen zu sein. Dieser Verdacht wird zur Alternativhypothese H_1 des Tests gemacht, die Nullhypothese H_0 ist das logische Gegenteil:

H_0: $p \leq 0{,}70$
H_1: $p > 0{,}70$.

Je geringer die Anzahl der geheilten Personen in der Stichprobe, desto wahrscheinlicher ist es, dass keine Verbesserung des Behandlungserfolgs erzielt wurde. Mit anderen Worten: Je kleiner der Wert der Zufallsvariable K ist, desto wahrscheinlicher ist es, dass die Nullhypothese des Tests zutrifft.

Bis zu welchem Maximalwert von K bleibt man bei der Nullhypothese? Je größer ein Wert ist, desto geringer ist die Wahrscheinlichkeit, dass er überschritten wird. Um den gesuchten Maximalwert zu definieren, wird zu Beginn des Tests die Wahrscheinlichkeit festgelegt, mit der er höchstens überschritten werden darf. Dieser Höchstwert der Überschreitungswahrscheinlichkeit ist das Signifikanzniveau α des Tests.

Die obere Begrenzung des Annahmeintervalls ist also dadurch gekennzeichnet, dass der betreffende Wert mit der Wahrscheinlichkeit α oder einer noch geringeren Wahrscheinlichkeit überschritten wird. Ein Wert, der mit der Wahrscheinlichkeit α überschritten wird, wird aber mit der Wahrscheinlichkeit $1 - \alpha$

unterschritten. Das Annahmeintervall wird nach oben durch den größten Wert $k_{1-\alpha}$ der Zufallsvariable K begrenzt, für den $F(k) \leq 1 - \alpha$ ist.

Als Signifikanzniveau wird in diesem Beispiel $\alpha = 0{,}01$ gewählt ($\Rightarrow 1 - \alpha = 0{,}99$). Die Verteilungsfunktion F(k) hat unter anderem die folgenden Werte: F(300) = 0,988, F(301) = 0,991, F(302) = 0,994, …[3]. Der größte Wert, für den $F(k) \leq 0{,}99$ ist, ist 300. Die obere Begrenzung des Annahmeintervalls ist $k_{0,99} = 300$.[4]

Wegen $k > k_{0,99}$ wird die Nullhypothese verworfen. Die Daten werden also dahingehend gedeutet, dass die neue Therapie eine Verbesserung darstellt.

Da die obere bzw. rechte Begrenzung des Annahmeintervalls zu bestimmen war, liegt ein so genannter **rechtsseitiger Test** vor.

In Kurzform ließe sich der Test wie folgt dokumentieren:

K: binomialverteilte Zufallsvariable mit n = 400, p = 0,70
k = 320

Rechtsseitiger Binomialtest:

H_0: $p \leq 0{,}70$
H_1: $p > 0{,}70$
$\alpha = 0{,}01$

Obere Begrenzung des Annahmeintervalls: $k_{0,99} = 300$
$k > k_{0,99} \Rightarrow H_0$ wird verworfen.

4 Weitere Anwendungsbeispiele

Ein linksseitiger Binomialtest wird durchgeführt, wenn der Verdacht besteht, dass sich die Wahrscheinlichkeit p für das Eintreten des interessierenden Ereignisses im Vergleich zu einem Referenzwert p_0 verringert hat.

H_0: $p \geq p_0$
H_1: $p < p_0$

Beispiele:

- Ein Hersteller von Saatgut gibt als Keimfähigkeit p_0 an. Nach einer gewissen Lagerungsdauer werden n Saatkörner ausgebracht. Von diesen keimen k Körner, wobei $k < p_0\, n$ ist. Hat die Keimfähigkeit des Saatguts abgenommen?

---

[3] In Excel beispielsweise zu berechnen mit der Funktion BINOM.VERT.

[4] In Excel beispielsweise kann dieser Wert auch mit der Funktion BINOM.INV berechnet werden. Formel: =BINOM.INV(400;0,70;0,99)−1. Die Funktion BINOM.INV(n;p;1–α) gibt den kleinsten Wert k aus, für den $F(k) \geq 1 - \alpha$ ist. Der kleinste Wert k, für den $F(k) \geq 0{,}99$ ist, ist im vorliegenden Fall aber 301, sodass vom Ergebnis noch 1 subtrahiert werden muss.

- Um die Ausschussrate bei der Herstellung eines Produkts gegenüber dem bisherigen Wert p_0 zu senken, wird eine Schulung des Betriebspersonals durchgeführt. Anschließend sind von n neu hergestellten Exemplaren des Produkts noch k Exemplare fehlerhaft, wobei $k < p_0\, n$ ist. Hat die Schulung Erfolg gehabt?

Ein rechtsseitiger Binomialtest wird durchgeführt, wenn der Verdacht besteht, dass sich die Wahrscheinlichkeit p für das Eintreten des interessierenden Ereignisses im Vergleich zu einem Referenzwert p_0 erhöht hat.

H_0: $p \leq p_0$
H_1: $p > p_0$

Beispiele:

- Ein fairer bzw. homogener Spielwürfel zeigt die Augenzahl Sechs mit der Wahrscheinlichkeit $p_0 = 1/6$. Bei n Würfen mit einem Würfel erscheint k-mal die Sechs, wobei $k > p_0\, n$ ist. Ist der Würfel so manipuliert worden, dass die Sechs mit einer erhöhten Wahrscheinlichkeit auftritt?
- Bislang hat der Anteil der Studienanfängerinnen in einem Studienfach p_0 betragen. Durch die gezielte Ansprache von Schulabsolventinnen wird versucht, diesen Anteil zu erhöhen. Im folgenden Semester sind unter n neu Studierenden k Frauen, wobei $k > p_0\, n$ ist. War die Maßnahme wirksam?