Signifikanztest: Signifikanzniveau und allg. Vorgehen

30.03.2020

Wiederholung

Korrektur: Ich danke David J. für den Hinweis auf einen Tippfehler von mir:
In der Tabelle der „vier möglichen Fälle“ muss in der letzten Zeile ganz links stehen:
$p > \frac{1}{6}$ (Ich habe das korrigiert, zuvor stand hier fälschlich: $p \geq \frac{1}{6}$). Im Zusammenhang:

In Wirklichkeit: H_0 trifft nicht zu, d. h. $p > \frac{1}{6}$ (Kandidatin kann „Farbe-Fühlen“)	H_0 wird beibehalten (Falsche Entscheidung, denn H_0 trifft ja nicht zu: sog. Fehler 2. Art)	H_0 wird abgelehnt (Richtige Entscheidung)

Zusammenfassung des bisher Gelernten:

Ein Signifikanztest ist ein Verfahren, mit dem entschieden werden kann, ob eine bestimmte Annahme über eine „Trefferwahrscheinlichkeit“ p (oder einen Anteil p) abgelehnt oder beibehalten wird.

Die bestimmte Annahme über p bezeichnet man als Nullhypothese H_0.

Im Einführungsbeispiel wollten wir testen, ob die Kandidatin „bloß rät“, d. h. H_0: $p = 1/6$.

Man führt eine Stichprobe mit einer bestimmten Länge n durch und ermittelt, wie oft das interessierende Merkmal („Treffer“) dabei vorkommt (möglich: 0 bis n mal).

Dazu verwendet man die Zufallsgröße X: „Anzahl der ‚Treffer‘ bei der Stichprobe der Länge n“

Vor der Durchführung der Stichprobe legt man eine Entscheidungsregel fest: Die möglichen Werte von X werden aufgeteilt in einen Annahmebereich A und einen Ablehnungsbereich $\overline{A}$.

Im Einführungsbeispiel:
$A = \{0; 1; 2; 3\}$ (wenige richtig bestimmte Stoffe: Annahme von H_0 („Sie rät“))
$\overline{A} = \{4; 5; \dots; 10\}$ (viele richtig bestimmte Stoffe: Ablehnung von H_0 (d. h. „kann Farbe fühlen“))

Nun führt man die Stichprobe durch und erhält dabei einen bestimmten Wert für die Zufallsgröße X:

Wenn $X \in A$ dann wird H_0 beibehalten,
wenn $X \in \overline{A}$ dann wird H_0 abgelehnt.

Bei der Durchführung des Tests können zwei Fehler auftreten:

Fehler 1. Art: H_0 trifft (in Wirklichkeit) zu, wird aber (fälschlich) abgelehnt.
Fehler 2. Art: H_0 trifft (in Wirklichkeit) nicht zu, wird aber (fälschlich) beibehalten.

Die Wahrsch. für den Fehler 1. Art (H_0 trifft zu, d. h. p ist bekannt) lässt sich (immer) berechnen.

Die Wahrsch. für den Fehler 2. Art (H_0 trifft nicht zu, d. h. p ist nicht bekannt) lässt sich nur dann berechnen, wenn man eine Annahme über die unbekannte „Trefferwahrscheinlichkeit“ p macht (oder wenn die Aufgabenstellung hierfür eine Vorgabe macht).

Nach der Bearbeitung von S. 118/2 „Telepathie“ und dem Lesen von S. 115 (unten) bis S. 116 (oben) wissen wir:

Durch die Veränderung der Entscheidungsregel (bei gleichem Stichprobenumfang)
kann man zwar die Wahrscheinlichkeit für den Fehler der einen Art verringern.
Dadurch erhöht man aber die Wahrscheinlichkeit für den Fehler der anderen Art.

Durch eine Erhöhung des Stichprobenumfangs
können die Wahrscheinlichkeiten für den Fehler 1. Art und für den Fehler 2. Art verringert werden.
Allerdings bedeutet eine Erhöhung des Stichprobenumfangs auch einen höheren Aufwand.

Bsp.:
Eine Umfrage vor einer Wahl bei n = 1.000 Personen liefert mit Fehlern behaftete Ergebnisse.
Bei n = 10.000 Personen ist das Ergebnis genauer, die Umfrage ist jedoch viel aufwändiger.

Im Einführungsbeispiel haben wir die Entscheidungsregel „beliebig, aber sinnvoll“ festgelegt.

In der Realität bestimmt man die Entscheidungsregel so,
dass die Wahrscheinlichkeit für den Fehler 1. Art einen bestimmten Wert nicht überschreitet.

Diesen bestimmten Höchstwert nennt man Signifikanzniveau α .

Wahl der Nullhypothese und Signifikanzniveau

Beispiel: **„Lieferant und Händler“**

Ein Händler bezieht von einem Lieferanten eine große Menge eines bestimmten Produkts.

Ein bestimmter Anteil p der gelieferten Exemplare ist dabei fehlerhaft.

Weil aus Kostengründen nicht alle „sehr vielen“ Exemplare getestet werden können, beschränkt man sich beim Test auf eine Stichprobe der Länge $n = 100$.

Der Lieferant schlägt vor:

Die Lieferung wird als „in Ordnung“ angesehen und muss bezahlt werden, wenn der Anteil fehlerhafter Exemplare höchstens 20 % ist.

Dann ist die Nullhypothese: H_0: $p \leq 0{,}2$ („wenige fehlerhafte Exemplare“)

Der Lieferant möchte nur mit einem Risiko von 5 % „sein Geld nicht erhalten“ (H_0 wird abgelehnt), obwohl die Lieferung in Ordnung ist (d. h. H_0 trifft zu).

Der Fehler 1. Art (H_0 trifft zu, wird aber fälschlich abgelehnt) soll also höchstens 5 % sein.

Aus dieser Forderung ergibt sich die Entscheidungsregel:

Die möglichen Werte für die Zufallsgröße X: „Anzahl fehlerhafter Exemplare“ werden aufgeteilt:

$$\underbrace{0 \quad \dots \quad \dots \quad k}_{\substack{\text{Annahmebereich } A \\ \text{(„wenige fehlerhaft“)}}} \quad \underbrace{k+1 \quad \dots \quad \dots \quad 100}_{\substack{\text{Ablehnungsbereich } \bar{A} \\ \text{(„viele fehlerhaft“)}}}$$

Aus der Forderung $P(\text{„Fehler 1. Art“}) \overset{!}{\leq} 5\,\%$ ergibt sich:

Fehler 1. Art: H_0 trifft zu, $\rightarrow$ $p \leq 0{,}2$
wird (fälschlich) abgelehnt. $\rightarrow$ $X \in \bar{A}$, d. h. $X \geq k + 1$

Für den Lieferanten ist der riskanteste Fall, dass der Fehleranteil genau 20 % ist, denn in diesem Fall möchte er trotzdem noch „sein Geld erhalten“.
Wir rechnen deshalb im Folgenden mit $p = 0{,}2$.

Also:

$$
\begin{aligned}
P_{0,2}^{100}(X \in \bar{A}) &\overset{!}{\leq} 5\,\% \\
P_{0,2}^{100}(X \geq k+1) &\leq 0{,}05 && \text{| Tafelwerk: nur „}X \leq k\text{“ tabelliert} \\
1 - P_{0,2}^{100}(X \leq k) &\leq 0{,}05 && \text{| Auflösen nach „P“} \\
1 - 0{,}05 &\leq P_{0,2}^{100}(X \leq k) && \text{| Seitentausch} \\
P_{0,2}^{100}(X \leq k) &\geq 0{,}95
\end{aligned}
$$

Wir müssen nun im Tafelwerk bei der Binomialverteilung für $p = 0{,}2$ und $n = 100$ nachsehen, ab welchem k die Summenwahrscheinlichkeit (rechte Spalte) erstmals größer oder gleich 0,95 ist.

Tafelwerk (S. 18): $P_{0,2}^{100}(X \leq 27) = 0{,}96585$ ($\geq 0{,}95$ „passt“) und zum Vergleich:

$P_{0,2}^{100}(X \leq 26) = 0{,}94417$ ($< 0{,}95$ „zu klein“)

Also: $k = 27$, folglich ergibt sich als Entscheidungsregel:

$X \in A = \{0; 1; \dots; 27\}$ $\rightarrow$ H_0 wird beibehalten („Lieferant bekommt sein Geld“)

$X \in \bar{A} = \{28; 29; \dots; 100\}$ $\rightarrow$ H_0 wird abgelehnt („Händler muss nicht zahlen“)

Wie groß ist bei dieser Entscheidungsregel das Risiko für den Händler, dass er die Lieferung bezahlen muss, obwohl 25 % der gelieferten Exemplare fehlerhaft sind?

Also: Gesucht ist die Wahrscheinlichkeit für den Fehler 2. Art:

Fehler 2. Art: H_0 trifft nicht zu, $\rightarrow$ $p = \mathbf{0{,}25}$ *[durch die Aufgabenstellung gegeben]*
wird (fälschlich) beibehalten. $\rightarrow$ $X \in A$, d. h. $X \leq 27$

$$P_{\mathbf{0,25}}^{100}(X \in A) = P_{0,25}^{100}(X \leq 27) \overset{\text{[Tafelwerk S. 18]}}{=} 0{,}72238 \approx 72{,}2\,\%$$

Damit ist der Händler natürlich nicht zufrieden, denn er muss bei dieser Entscheidungsregel mit einer Wahrsch. von über 72 % eine Lieferung bezahlen, die zu 25 % fehlerhafte Exemplare enthält.

Der Händler schlägt deshalb vor:

Die Lieferung wird als „unbrauchbar“ angesehen und muss nicht bezahlt werden, wenn der Anteil fehlerhafter Exemplare mindestens 20 % ist.

Dann ist die Nullhypothese: H_0: $p \geq 0{,}2$ („viele fehlerhafte Exemplare“)

Der Händler möchte nur mit einem Risiko von 5 % „zu unrecht zahlen“ (H_0 wird abgelehnt), obwohl die Lieferung unbrauchbar ist (d. h. H_0 trifft zu).

Der Fehler 1. Art (H_0 trifft zu, wird aber fälschlich abgelehnt) soll also höchstens 5 % sein.

Aus dieser Forderung ergibt sich die Entscheidungsregel:

Die möglichen Werte für die Zufallsgröße X: „Anzahl fehlerhafter Exemplare“ werden aufgeteilt:

$$\underbrace{0 \quad \dots \quad \dots \quad k}_{\text{Ablehnungsbereich } \overline{A}\ (\text{„wenige fehlerhaft“})} \qquad \underbrace{k+1 \quad \dots \quad \dots \quad 100}_{\text{Annahmebereich } A\ (\text{„viele fehlerhaft“})}$$

Aus der Forderung P(„Fehler 1. Art“) $\overset{!}{\leq}$ 5 % ergibt sich:

Fehler 1. Art: H_0 trifft zu, $\rightarrow$ $p \geq 0{,}2$
wird (fälschlich) abgelehnt. $\rightarrow$ $X \in \overline{A}$, d. h. $X \leq k$

Für den Händler ist der riskanteste Fall, dass der Fehleranteil genau 20 % ist, denn bereits ab 20 % fehlerhaften Exemplaren möchte er die Lieferung nicht bezahlen.
Wir rechnen deshalb mit $p = 0{,}2$.

Also: $P_{0,2}^{100}(X \in \overline{A}) \overset{!}{\leq} 5\ \%$

$P_{0,2}^{100}(X \leq k) \leq 0{,}05$

Wir müssen nun im Tafelwerk bei der Binomialverteilung für p = 0,2 und n = 100 nachsehen, bis zu welchem k die Summenwahrscheinlichkeit (rechte Spalte) noch kleiner oder auch gleich 0,05 ist.

Tafelwerk (S. 17): $P_{0,2}^{100}(X \leq 13) = 0{,}04691$ ($\leq 0{,}05$ „passt“) und zum Vergleich:

$P_{0,2}^{100}(X \leq 14) = 0{,}08044$ ($> 0{,}05$ „zu groß“)

Also: $k = 13$, folglich ergibt sich nun als Entscheidungsregel:

$X \in A = \{14; 15; \dots; 100\}$ $\rightarrow$ H_0 wird beibehalten („Händler muss nicht zahlen“)

$X \in \overline{A} = \{0; 1; \dots; 13\}$ $\rightarrow$ H_0 wird abgelehnt („Händler muss zahlen“)

Wie groß ist bei dieser Entscheidungsregel das Risiko für den Lieferanten, dass er die Lieferung nicht bezahlt bekommt, obwohl nur 15 % der gelieferten Exemplare fehlerhaft sind?

Also: Gesucht ist die Wahrscheinlichkeit für den Fehler 2. Art:

Fehler 2. Art: H_0 trifft nicht zu, $\rightarrow$ $p = \mathbf{0{,}15}$ *[durch die Aufgabenstellung gegeben]*
wird (fälschlich) beibehalten. $\rightarrow$ $X \in A$, d. h. $X \geq 14$

$$P_{\mathbf{0,15}}^{100}(X \in A) = P_{0,15}^{100}(X \geq 14) = 1 - P_{0,15}^{100}(X \leq 13) \overset{\text{[Tafelwerk S. 13]}}{=} 1 - 0{,}34743 = 0{,}65257 \approx$$

$$\approx 65{,}3\ \%$$

Damit ist nun aber der Lieferant nicht zufrieden, denn er muss bei dieser Entscheidungsregel mit einer Wahrsch. von über 65 % auf „sein Geld verzichten“, obwohl nur 15 % der Exemplare fehlerhaft sind.

[Eine weitere Seite noch, nicht aufgeben!]

Wahl der Nullhypothese und Signifikanzniveau

Das Beispiel „Lieferant und Händler“ hat gezeigt:

Die Wahl der Nullhypothese hat für den Lieferanten und für den Händler unterschiedliche Konsequenzen:

Der Lieferant möchte verhindern, dass er eine „gute“ Lieferung (H_0: $p \leq 0{,}2$) nicht bezahlt bekommt.

Die Wahrscheinlichkeit dafür möchte er auf 5 % beschränken, d. h. die Wahrsch. für den Fehler 1. Art
[H_0 trifft zu („gute Lieferung“), wird aber fälschlich abgelehnt ($X \in \overline{A}$, d. h. $X \geq k + 1$)]
darf das Signifikanzniveau $\alpha = 0{,}05$ nicht überschreiten.

Die sich aus dieser Forderung ergebende Entscheidungsregel hat aber z. B. zur Folge, dass der Händler mit einer Wahrsch. von über 72 % eine Lieferung bezahlen muss, die 25 % fehlerhafte Exemplare enthält.

Der Händler möchte verhindern, dass er eine „schlechte“ Lieferung (H_0: $p \geq 0{,}2$) bezahlen muss.

Die Wahrscheinlichkeit dafür möchte er auf 5 % beschränken, d. h. die Wahrsch. für den Fehler 1. Art
[H_0 trifft zu („schlechte Lieferung“), wird aber fälschlich abgelehnt ($X \in \overline{A}$, d. h. $X \leq k$)]
darf das Signifikanzniveau $\alpha = 0{,}05$ nicht überschreiten.

Die sich aus dieser Forderung ergebende Entscheidungsregel hat aber z. B. zur Folge, dass der Lieferant mit einer Wahrsch. von über 65 % kein Geld bekommt, obwohl nur 15 % der Exemplare fehlerhaft sind.

Zusammenfassung: Wahl der Nullhypothese und Signifikanzniveau:

> Die Nullhypothese wird so gewählt,
> dass der Fehler 1. Art von größerer Bedeutung ist als der Fehler 2. Art.
>
> Bei der Bestimmung der Entscheidungsregel lässt sich die Wahrsch. für den Fehler 1. Art gering halten:
>
> Man legt einen Höchstwert α fest, den der Fehler 1. Art nicht überschreiten darf,
> und ermittelt darauf beruhend die Entscheidungsregel.
>
> Diesen Höchstwert α, den die Wahrscheinlichkeit für den Fehler 1. Art nicht überschreiten darf, bezeichnet man als **Signifikanzniveau α** . *[Siehe auch: Merkhilfe innen-unten-rechts]*

Zusammenfassung: Vorgehen bei einem Signifikanztest:

> **Allgemeines Vorgehen bei einem Signifikanztest:**
>
> 1. Formulierung der Nullhypothese H_0
> 2. Festlegung der Zufallsgröße X bei einer Stichprobe der Länge n *(„Wofür steht X?“)*
> 3. Vorgabe des Signifikanzniveaus α und Bestimmung der darauf beruhenden Entscheidungsregel
> 4. Durchführung des Test und Entscheidung aufgrund des Stichprobenergebnisses

[Kern vieler (Abitur-)Aufgaben ist die Bestimmung der Entscheidungsregel bei vorgegebenem Signifikanzniveau α.]

Damit haben wir alles Wesentliche zum Thema „Signifikanztest“ behandelt. (Juhu!)

Als Übungen empfehle ich die Bearbeitung von S. 119/5 „Qualitätskontrolle“ und S. 119/6 „Ausbau der Stadtautobahn“.

Bitte unbedingt selbstständig bearbeiten!

Die Lösungen werden spätestens am Do., 02.04., auf www.mathe12.mathebauer.de veröffentlicht werden.

Anmerkungen für „STARK-Bücher-LeserInnen“ und „Erklärvideo-AnschauerInnen“:

Wir dürfen uns in Bayern in der Schule auf „einseitige Binomialtests“ beschränken:

Binomialtest: Die Testgröße X ist binomialverteilt (→ Tafelwerk).

Einseitiger Test: Die Nullhypothese ist entweder

H_0: $p \leq p_0$ (Der Ablehnungsbereich $\overline{A}$ liegt dann „rechts“ → sog. „rechtsseitiger Test“.)
oder (siehe oben: Lieferant: H_0: $p \leq 0{,}2$ → $\overline{A}$ „liegt rechts“)

H_0: $p \geq p_0$ (Der Ablehnungsbereich $\overline{A}$ liegt dann „links“ → sog. „linksseitiger Test“.)
(siehe oben: Händler: H_0: $p \geq 0{,}2$ → $\overline{A}$ „liegt links“)

Daneben gibt es noch zweiseitige Tests ($\overline{A} = \{0; 1; \ldots; k_1\} \cup \{k_2; \ldots; n\}$), Alternativtests sowie Tests, bei denen die Testgröße eine andere Wahrscheinlichkeitsverteilung hat als die Binomialverteilung.