Aufgabe
Stochastik – Mathe Grundkurs
Signifikanztests

Ein Hersteller von Schrauben behauptet, dass mindestens 90% seiner Schrauben rostfrei sind, wenn sie fünf Jahre lang im Außenbereich eingesetzt werden. Hierfür gibt es ein Prüfverfahren, bei dem in einem anerkannten Labor analysiert wird, ob der Anteil von Edelmetallen am Hauptbestandteil Eisen für jede Schraube ausreichend groß ist.

Stellen Sie sich vor, Sie sind als Mitarbeiter einer Firma für den Einkauf von Schrauben verantwortlich. Ihr Chef hat Sie beauftragt, zu prüfen, ob die neue Herstellerfirma als Lieferant für zukünftige Bestellungen in Frage kommt. Hierfür soll eine Stichprobe von 20 Schrauben gekauft und an das Prüflabor geschickt werden, um darüber zu entscheiden, ob das Werbeversprechen des Schraubenherstellers stimmt.

1) Entwickeln und beschreiben Sie mit den gängigen Begriffen der Hypothesentests ein Testverfahren, welches geeignet ist, die Werbeaussage des Herstellers zu prüfen ($\alpha \leq 15\%$).

2) Ihr Chef schlägt vor, dass die Firma nicht Lieferant werden soll, wenn

   a) mehr als 4 Schrauben in der Stichprobe
   b) mehr als 1 Schraube in der Stichprobe

   den Labortest nicht bestanden hat. Beurteilen Sie diese Ansätze. Welchem dieser zwei Ansätze wird der Hersteller wohl eher zustimmen?

3) Nach langen Verhandlungen mit dem Hersteller einigt man sich auf folgendes Verfahren: Der Hersteller darf Ihre Firma zukünftig beliefern, wenn maximal 15% der Schrauben dieser Stichprobe in der Laborprüfung durchfallen.

   a) Sie haben den Verdacht, dass in Wahrheit durchschnittlich nur 80% der Schrauben dieses Herstellers rostfrei sind. Wie groß ist das Risiko, dass die Stichprobe trotzdem die Laborprüfung besteht?

   b) Angenommen, das bei a) ermittelte Risiko wäre Ihnen zu hoch – welche Verbesserung des Testverfahrens wäre möglich, gegen die ein ehrlicher Hersteller nicht protestieren wird?

Lösung

1) Grundsätzlich funktionieren alle Hypothesentests nach folgendem Prinzip: Man nimmt eine Stichprobe und zählt die einzelnen Exemplare aus. Dabei erfolgt die Unterscheidung in „gut“ und „schlecht“ bzw. Treffer und Niete. Man sollte sich also zunächst überlegen, welches hier die Treffer und Nieten sind.

   Unterschieden wird in rostfreie und nicht rostfreie Schrauben. Empfehlenswert ist hier die Wahl der nicht rostfreien Schrauben als Treffer, dann gilt p=0,1. Auch die umgekehrte Festlegung mit p=0,9 wäre möglich, führt aber dazu, dass die zugehörigen kumulierten Wahrscheinlichkeiten in den klassischen Tabellenwerken zur Binomialverteilung komplizierter abzulesen sind.

   Gegeben:
   Stichprobenumfang n=20
   Trefferwahrscheinlichkeit p=0,1
   Irrtumswahrscheinlichkeit für den Fehler 1.Art $\alpha \leq 15\%$

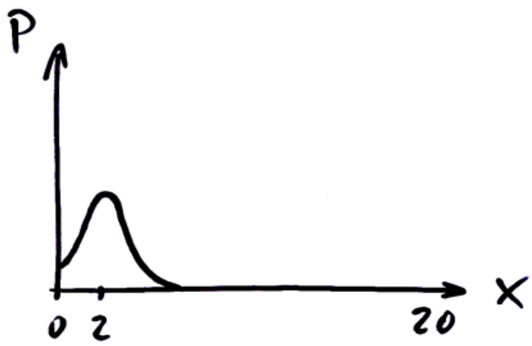
   Zu entwickeln:
   Nullhypothese H_0: $p \leq 0{,}1$
   Gegenhypothese H_1: $p > 0{,}1$

   Optional, aber hilfreich, ist die Erstellung einer kleinen Skizze mithilfe von Mittelwert μ und Standardabweichung σ:

   $\mu = n * p = 20 * 0{,}1 = 2$

   $$\sigma = \sqrt{n \bullet p \bullet (1-p)} = \sqrt{20 \bullet 0{,}1 \bullet 0{,}9} \approx 1{,}34$$

   Das sogenannte 1-Sigma-Intervall, in dem sich etwa 2/3 der Wahrscheinlichkeit konzentrieren, liegt hier also zwischen der linken Grenze 2 – 1,34 = 0,66 und der rechten Grenze 2 + 1,34 = 3,34



   Der Hersteller behauptet, dass er die Trefferwahrscheinlichkeit von p = 0,1 nicht ÜBER-schreitet. Daher werden kleine Trefferzahlen um μ = 2 zur Annahme von H_0 führen und zu große Werte zur Annahme von H_1. Es handelt sich also um einen rechtsseitigen Test.

$\alpha \le 0{,}15$ bedeutet, die Obergrenze des Annahmebereiches k muss so gewählt werden, dass das Eintreten einer höheren Trefferzahl als k unter normalen Bedingungen (d.h. wenn H_0 gilt) nur noch mit einer Wahrscheinlichkeit von höchstens 15% erfolgt.

$P(X > k) \le 0{,}15$

Die Zahl 0,15 entspricht der Fläche unter der Wahrscheinlichkeitskurve, die durch die senkrechte Grenze über k rechtsseitig abgegrenzt wird.

Da die Tabellenwerke zur kumulierten Binomialverteilung, meist mit F(n; p; k) bezeichnet, in der Regel nur Ergebnisse für $P(X \le k)$ liefern, also für Kleiner-Gleich-Ausdrücke, ist die Fragestellung mathematisch umzuwandeln.

$P(X > k)$	$\le 0{,}15$	| Gegenwahrscheinlichkeit bilden
$1 - P(X \le k)$	$\le 0{,}15$	| –1 | * (–1)
$P(X \le k)$	$\ge 0{,}85$	

Beachte die Drehung des Ungleichheitszeichens nach Multiplikation mit einer negativen Zahl! Gesucht ist also derjenige Wert für k, bei dem die kumulierte Wahrscheinlichkeitsfunktion $F(20; 0{,}1; k) \ge 0{,}85$ ist.

Ablesen in der Tabelle liefert $P(X \le 3) = 0{,}8670$, also k = 3.

Damit wird Annahme- und Ablehnungsbereich für H_0 wie folgt festgelegt:

Annahmebereich: $0 \le X \le 3$
Ablehnungsbereich: $4 \le X \le 20$

Im Klartext: Die Schraubenfirma besteht den Qualitätstest, wenn maximal 3 Schrauben in der Laborprüfung durchfallen.

2) Die Vorschläge des Chefs sind Angaben zum Annahmebereich. Zur Beurteilung wird die Wahrscheinlichkeit für den Fehler 1. Art α berechnet. (Übrigens: β kann man hier nicht berechnen, dazu müsste in der Aufgabe eine Angabe erfolgen, wie groß p tatsächlich sein soll.)

a) $\alpha = P(X > 4) = 1 - P(X \le 4) = 1 - 0{,}9568 = 0{,}0432 \approx 4{,}3\%$
b) $\alpha = P(X > 1) = 1 - P(X \le 1) = 1 - 0{,}3917 = 0{,}6083 \approx 60{,}8\%$

Ein ehrlicher Lieferant würde bei a) mit 4,3% Wahrscheinlichkeit durchfallen, bei b) mit 60,8%

Variante a) wäre für den Schraubenhersteller recht vorteilhaft. Allerdings wäre sie für den Käufer risikoreich, da der Hersteller gute Chancen auf ein Bestehen des Tests hätte, selbst wenn er seinen versprochenen Wert p = 0,1 überschreitet.

Variante b) wird wohl kein Hersteller freiwillig akzeptieren, da er hier selbst bei Einhaltung seines Werbeversprechens in den meisten Fällen durch den Test durchfällt. Für den Einkäufer hätte diese Variante den einzigen Vorteil, dass er nur sehr geringe Gefahr läuft, einen Täuschungsversuch des Herstellers nicht zu bemerken. Andererseits muss er sehr lange nach geeigneten Herstellern suchen, denn er wird vermutlich sehr viele gute Hersteller als vermeintlich ungeeignet davonjagen, bevor er einen findet.

Aus rein mathematischer Sicht ist Variante b) ungeeignet zum Nachweis der Nullhypothese H_0: $p \leq 0{,}1$, da selbst der Durchschnittswert der Verteilung $\mu = 2$ schon im Ablehnungsbereich liegt.

3) 15% ist hier keine Wahrscheinlichkeit, sondern eine Angabe zum Annahmebereich. Das Annahmeintervall errechnet sich wie folgt:

$k = 0{,}15 * 20 = 3$

→ Annahmebereich: $0 \leq X \leq 3$
→ Ablehnungsbereich: $4 \leq X \leq 20$

a) Der Verdacht besagt, dass die Trefferwahrscheinlichkeit nun 20%, also $p = 0{,}2$ ist. Wenn der Verdacht stimmt, besteht die Gefahr, dass ein Fehler 2. Art begangen wird (Annahme einer falschen Nullhypothese). Zur Annahme von H_0 führt ein Eintreten einer Trefferzahl von maximal 3. Daher ist die entsprechende Wahrscheinlichkeit aus der Tabelle abzulesen, dass eine binomialverteilte Zufallsgröße mit $n = 20$ und $p = 0{,}2$ Werte bis 3 annimmt.

$\beta = P(X \leq 3) = F(20; 0{,}2; 3) = 0{,}4114 \approx 41{,}1\%$

Das Risiko, dass bei diesem Test eine Stichprobe die Laborprüfung besteht, ist 41%. Aus Sicht des Käufers wäre es wünschenswert, ein geringeres Risiko zu haben.

b) Man kann viel darüber diskutieren, wo die Annahmegrenze eines „optimalen“ Tests liegen soll. Dabei stehen sich naturgemäß die Interessen des Herstellers, der eine hohe Wahrscheinlichkeit zum Bestehen des Tests möchte (α möglichst klein, k groß) und des Käufers, der mögliche Täuschungsversuche erkennen will (β möglichst klein, k klein), gegenüber. Eine Verkleinerung des einen Fehlers geht aber immer auf Kosten der Vergrößerung des anderen Fehlers.

Der einzige Weg, wirklich beide Fehler klein zu halten, besteht immer darin, den Stichprobenumfang n zu vergrößern. Damit werden α und β kleiner. Da solche Tests meistens Zeit und Geld kosten, ist dies in der Praxis oft nicht möglich. Rein mathematisch gesehen wäre dies jedoch eine Lösung, der beide Seiten zustimmen müssten.

Tabellenwerk

Tabellen zur Binomialverteilung P(X=k) = B(n;p;k)

n=5

k	p=0,1	p=1/6 ②	p=0,2	p=0,3	p=1/3	p=0,4	p=0,5	
0	5905	4019	3277	1681	1317	0778	0313	5
1	3281	4019	4096	3602	3292	2592	1563	4
2	0729	1608 ①	2048	3087	3292	3456	3125	3
3	0081	0322	0512	1323	1646	2304	3125	2
4	0005	0032	0064	0284	0412	0768	1563	1
5		0001	0003	0024	0041	0102	0313	0
	p=0,9	p=5/6	p=0,8	p=0,7	p=2/3	p=0,6	p=0,5	k ④

n=20

k	p=0,1	p=1/6	p=0,2	p=0,3	p=1/3	p=0,4	p=0,5	
0	1216	0261	0115	0008	0003	0000	0000	20
1	2702	1043	0576	0068	0030	0005	0000	19
2	2852	1982	1369	0278	0143	0031	0002	18
3	1901	2379	2054	0716	0429	0123	0011	17
4	0898	2022	2182	1304	0911	0350	0046	16
5	0319	1294	1746	1789	1457	0746	0148	15
6	0089	0647	1091	1916	1821	1244	0370	14
7	0020	0259	0545	1643	1821	1659	0739	13
8	0004	0084	0222	1144	1480	1797	1201	12
9	0001	0022	0074	0654	0987	1597	1602	11
10		0005	0020	0308	0543	1171	1762	10
11		0001	0005	0120	0247	0710	1602	9
12			0001	0039	0092	0355	1201	8
13				0010	0028	0146	0739	7
14				0002	0007	0049	0370	6
15		③			0001	0013	0148	5
16						0003	0046	4
17							0011	3
18							0002	2
19								1
20								0
	p=0,9	p=5/6	p=0,8	p=0,7	p=2/3	p=0,6	p=0,5	k

① 0322 bedeutet 0,0322 und ist der Wert für B(5; 1/6; 3)

② vergleiche Tabelle 1, Seite 20 und Erklärungen

③ für große k bedeuten Leerfelder B(n; p; k) = 0 (auf 4 Dezimalstellen genau)

④ für p > 0,5 wird p aus dem grau unterlegten Feld abgelesen. Das zugehörige k wird nicht aus der linken, sondern aus der grau unterlegten rechten Spalte abgelesen. z.B. B(20; 0,7; 10) = 0,0308.

Tabellen zur kumulierten Binomialverteilung / Summenverteilung $P(X \leq k) = F(n;p;k)$

n=5

k	p=0,1	p=1/6	p=0,2	p=0,3	p=1/3	p=0,4	p=0,5	
0	5905	4019	3277	1681	1317	0778	0313	4
1	9185	8038	7373	5282	4609	3370	1875	3
2	9914	9645	9421	8369	7901	6826	5000	2
3	9995	9967	9933	9692	9547	9130	8125	1
4		9999	9997	9976	9959	9898	9688	0
	p=0,9	p=5/6	p=0,8	p=0,7	p=2/3	p=0,6	p=0,5	k

⑤

n=20

k	p=0,1	p=1/6	p=0,2	p=0,3	p=1/3	p=0,4	p=0,5	
0	1216	0261	0115	0008	0003	0000	0000	19
1	3917	1304	0692	0076	0033	0005	0000	18
2	6769	3287	2061	0355	0176	0036	0002	17
3	8670	5665	4114	1071	0604	0160	0013	16
4	9568	7687	6296	2375	1515	0510	0059	15
5	9887	8982	8042	4164	2972	1256	0207	14
6	9976	9629	9133	6080	4793	2500	0577	13
7	9996	9887	9679	7723	6615	4159	1316	12
8	9999	9972	9900	8867	8095	5956	2517	11
9		9994	9974	9520	9081	7553	4119	10
10		9999	9994	9829	9624	8725	5881	9
11			9999	9949	9870	9435	7483	8
12				9987	9963	9790	8684	7
13				9997	9991	9935	9423	6
14					9998	9984	9793	5
15						9997	9941	4
16							9987	3
17							9998	2
18								1
19								0
	p=0,9	p=5/6	p=0,8	p=0,7	p=2/3	p=0,6	p=0,5	k

⑥

⑤ Für p>0,5 wird p aus dem grau unterlegten Feld abgelesen. Das zugehörige k wird nicht aus der linken, sondern aus der grau unterlegten rechten Spalte abgelesen. Bei der KUMULIERTEN Binomialverteilung wird der Ablesewert dann noch von 1 subtrahiert. z.B. F(20; 0,7; 10) = 1 – 0,9520 = 0,048.

⑥ Für große k bedeuten Leerfelder F(n; p; k) = 1 (auf 4 Dezimalstellen genau)