**Qualitätskontrolle durch Stichproben**

nach A.F. Bissell

Ein Heimwerker kauft ein Paket Schrauben. Als er es zu Hause öffnet und einige Schrauben entnimmt, stellt er bei der ersten fest, daß sie keinen Schlitz hat. Bei einer anderen ist das Gewinde defekt. "Ich schaue mir noch ein paar andere Schrauben an", denkt er. Wenn noch eine defekt ist, werde ich das Paket reklamieren.

Dem Leser wird diese Situation bekannt vorkommen. Er braucht nur defekte Schrauben durch Heftzwecke mit losem Kopf, undichte Luftballons oder Streichhölzer ohne Zündkopf ersetzen.

In jedem dieser Fälle hat er sich eine Schachtel mit einer Anzahl von gleichen Einzelteilen gekauft und entnimmt nun einige als Stichprobe. Dadurch erhält er Informationen über die Ausprägung eines Merkmals, das er an den Einzelteilen beobachtet. Er sieht natürlich ein, daß er vom Hersteller nicht erwarten kann, daß er für die gute Qualität eines jeden Einzelteils garantiert. Aber wieviele defekte Einzelteile in der Stichprobe würden ihn davon überzeugen, daß die ganze Schachtel reklamiert werden sollte?

Mit dieser Fragestellung ist vor allem die Industrie vertraut, wenn z.B. geprüft werden muß, ob eine Sendung mit Massenartikeln akzeptiert werden kann, ohne die Brauchbarkeit jedes einzelnen Teils der Sendung zu prüfen.

Die Entscheidung, ob eine Sendung akzeptiert werden kann oder nicht, hat oft erhebliche Konsequenzen. Der Lieferant von 1000 (teuren) Maschinenersatzteilen könnte z.B. vertraglich verpflichtet sein, die gesamte Verpackung zu ersetzen oder zumindest die defekten Teile auszutauschen. Der Empfänger der Sendung könnte sich seinerseits vor die Alternative gestellt sehen, eine Lieferung mit relativ vielen defekten Teilen zu akzeptieren oder eine Produktionsunterbrechung in Kauf zu nehmen, wenn er eine Sendung dringend benötigter Teile zurückgehen läßt.

Bevor wir untersuchen, wie man die Qualität einer Sendung durch Stichproben überprüft, müssen wir zur Kenntnis nehmen, daß stets ein bestimmter Anteil defekter Teile toleriert werden muß. Die Qualität eines solchen Teils könnte z.B. bei voller Funktionstüchtigkeit lediglich durch Äußerlichkeiten (z.B. Fehldruck auf dem Deckblatt eines Steno-Blocks) eingeschränkt sein. In anderen Fällen könnte der Fehler mit einem gewissen Aufwand behoben werden. Schlecht passende Gewinde könnten mit einem Gewindeschneider gängig gemacht werden. Dadurch würde aber u.U. die Produktion unterbrochen werden. Auch ein kleiner Anteil völlig unbrauchbarer Teile könnte toleriert werden, weil z.B. das gelegentliche Aussondern von defekten Teilen den Produktionsablauf nicht stört.

Wenn aber überhaupt kein defektes Einzelteil toleriert werden kann, muß jedes Einzelteil geprüft werden.

Auf jeden Fall sollten Lieferant und Empfänger einer Sendung vertraglich definieren, wenn ein Teil als defekt anzusehen ist, wieviele defekten Teile zu tolerieren sind und welche Regreßansprüche z.B. bei Produktionsausfällen entstehen, die durch Sendungen mit zu vielen defekten Teilen verursacht werden. Darüber hinaus wird der Empfänger prüfen müssen, welche Risiken fehlerhafter Entscheidungen bei Qualitätskontrollen durch Stichproben akzeptiert werden können. Eine insgesamt gute Sendung könnte wegen einer zufällig ungünstigen Stichprobe abgelehnt werden. Eine schlechte Sendung könnte akzeptiert werden wegen der zufällig guten Qualität der Stichprobe. An Hand eines einfachen Beispiels soll dies zunächst verdeutlicht werden. Eine Sendung von 100 Teilen soll geprüft werden. Dazu werden in einer Stichprobe 20 Teile entnommen. Falls dabei nicht mehr als ein defektes Teil entdeckt wird, soll die Sendung akzeptiert werden. Bei 2 oder mehr defekten Teilen wird die Sendung zurückgewiesen.

Wenn die gesamte Sendung nicht mehr als 1 defektes Teil enthält, wird sie bei jeder Stichprobe akzeptiert. Wenn sie jedoch z.B. 2 defekte Teile enthält, ist es möglich, daß beide defekten Teile in der Stichprobe vorkommen. Wir nehmen an, daß die Stichproben "zufällig" entnommen werden und wollen zunächst ausrechnen,



mit welcher Wahrscheinlichkeit eine Stichprobe genau 2 defekte Teile enthält, wenn die ganze Sendung nur 2 defekte Teile enthält.

Von N= 100 Teilen der Sendung sind G=98 gute Teile und D= 2 defekte Teile.

Die Stichprobe enthält S= 20 Teile, davon g= 18 gute und d= 2 defekte Teile.

Zunächst ist zu beachten, daß es $\binom{100}{20} = \binom{N}{S}$ Möglichkeiten gibt, eine Stichprobe zu entnehmen.

Dabei gilt: $\binom{100}{20} = \frac{100!}{20!\ 80!}$ , allgemein: $\binom{N}{S} = \frac{N!}{S!(N-S)!}$

Soll die Stichprobe nun g= 18 gute Teile enthalten, so müssen wir weiter überlegen, wieviele Möglichkeiten es gibt, von 20 guten Teilen 18 gute Teile auszuwählen. Wie oben gibt es dafür $\binom{20}{18}$ Möglichkeiten, allgemein also $\binom{G}{g}$ Möglichkeiten. Da es nur D=2 defekte Teile in der Sendung gibt, haben wir jeweils nur eine Möglichkeit, d=2 defekte Teile für die Stichprobe auszuwählen.

Dividiert man nun die Zahl $\binom{98}{18}$ der Stichproben mit genau 2 defekten Teilen durch die Zahl $\binom{100}{20}$ aller möglichen Stichproben, so erhält man die Wahrscheinlichkeit dafür, daß wir eine Stichprobe mit genau 2 defekten Teilen entnehmen:

$$p=\binom{G}{g}:\binom{N}{S} = \frac{98!}{18!\ 80!} : \frac{100!}{20!\ 80!} = \frac{98!\ 20!\ 80}{18!\ 80!\ 100!} = \frac{19\cdot 20}{99\cdot 100} = 0{,}038 \approx 4\ \%$$

Nehmen wir einmal an, daß eine Sendung mit nur 2 defekten Teilen durchaus brauchbar ist, so wird unser Prüfverfahren nur 4 % der Sendungen mit genau 2 defekten Teilen fälschlich als unbrauch-zurückweisen, während 96 % solcher Sendungen akzeptiert werden. Unser Prüfverfahren "erkennt" also gute Sendungen. Was leistet es aber bei schlechten Sendungen?. Nehmen wir an, daß die Sendung D= 10 defekte und G= 90 gute Teile enthält. Sie wird akzeptiert, sobald die Stichprobe kein defektes Teil oder nur 1 defektes Teil enthält.

Nun gibt es $\binom{90}{20}$ Stichproben mit genau 20 guten Teilen. Darf nur 1 defektes Teil in der Stichprobe sein, so kann man dies

Teil auf $\binom{10}{1}$ Arten auswählen, die restlichen 19 guten Teile wählt man aus den 90 guten Teilen aus auf $\binom{90}{19}$ Arten, so daß es $\binom{10}{1} \cdot \binom{90}{19}$ Möglichkeiten gibt, genau 1 defektes Teil in der Stichprobe zu haben. Somit erhält man als Wahrscheinlichkeit für höchstens ein defektes Teil in der Stichprobe den Wert

$$p = \frac{\binom{90}{20} + \binom{10}{1}\binom{90}{19}}{\binom{100}{20}} = 36{,}3\ \%.$$

Mehr als 1/3 der Sendungen mit 10 defekten Teilen wird irrtümlich als akzeptabel eingestuft.

Man könnte nun versuchen, diese Schwäche des Prüfverfahrens dadurch zu beheben, daß man eine Sendung schon dann zurückweist, wenn sie mindestens 1 defektes Teil enthält. Sie wird also dann akzeptiert, wenn die Stichprobe kein defektes Teil enthält. Dies ist im 2. Beispiel der Fall mit einer Wahrscheinlichkeit von $p = \binom{90}{20} : \binom{100}{20} = 9{,}5\ \%$. Es werden also nur noch 9,5 % der der Sendungen mit 10 defekten Teilen als brauchbar eingestuft. Enthält die Sendung wie im 1. Beispiel nur 2 defekte Teile, so erhält man in diesem Fall eine Stichprobe mit keinem defekten Teil mit der Wahrscheinlichkeit von $p = \binom{98}{20} : \binom{100}{20} = 63{,}8\ \%$. Mehr als 1/3 solcher guten Sendungen werden also durch den Test zurückgewiesen.

Man hat also die Zahl der erkannten schlechten Sendungen erhöht, gleichzeitig aber auch die Zahl der guten Sendungen, die irrtümlich zurückgewiesen werden.

Eine gleichzeitige Verbesserung beider Eigenschaften könnte man durch die Entnahme größerer Stichproben erreichen, was aber mit größerem Aufwand an Zeit und Kosten verbunden sein dürfte.

<u>Fassen wir unser Vorgehen zusammen:</u>

Wir gehen also von Sendungen mit N Teilen aus und entnehmen Stichproben mit S Teilen. Dann legen wir die Höchstzahl a der defekten Teile in der Stichprobe fest, bei der wir die Sendung noch akzeptieren wollen. Dann können wir die sogenannte Operationscharakteristik des Tests bestimmen, das ist die Funktion, die



jedem Anteil D/N von defekten Teilen in der Sendung die Wahrscheinlichkeit zuordnet, daß der Test die Sendung als gut einstuft.

Für $d \leq a$ gibt $\binom{D}{d} \cdot \binom{N-D}{S-d}$ die Zahl der Stichproben mit genau d defekten Teilen an. Man erhält die Annahmewahrscheinlichkeit, indem man diese Produkte für d = 0,1... bis a berechnet, aufaddiert und die Summe durch $\binom{N}{S}$ teilt:

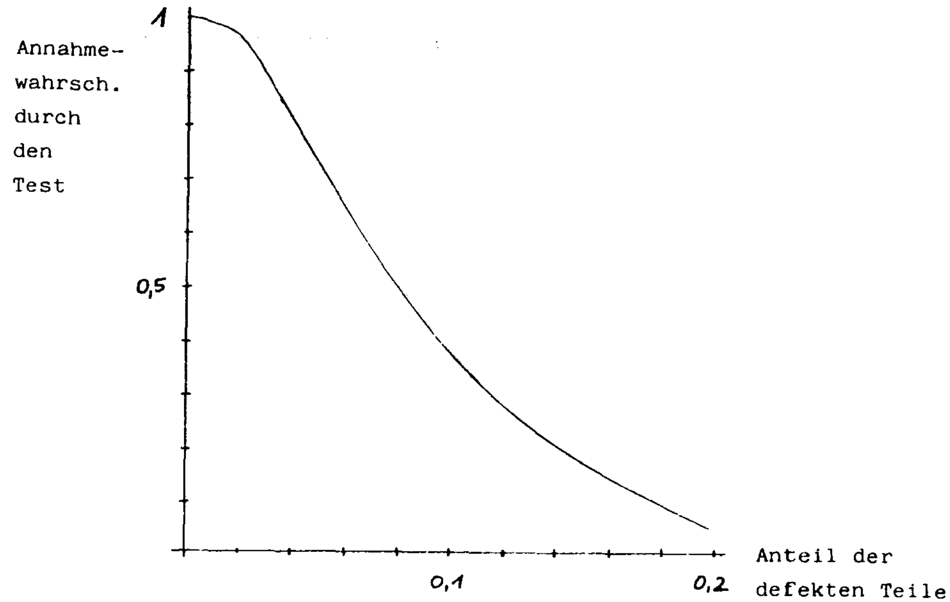
$$\text{Annahmewahrscheinlichkeit} = \left( \sum_{d=0}^{a} \binom{D}{d} \binom{N-D}{S-d} \right) : \binom{N}{S}$$

Zugehörige Operationscharakteristik:

(N=100, S= 20, a= 1)

| D/N | 0 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 |
|---|---|---|---|---|---|---|---|
| Annahme-wahrsch. p | 1,0 | 1,0 | 0,96 | 0,90 | 0,82 | 0,74 | 0,66 |

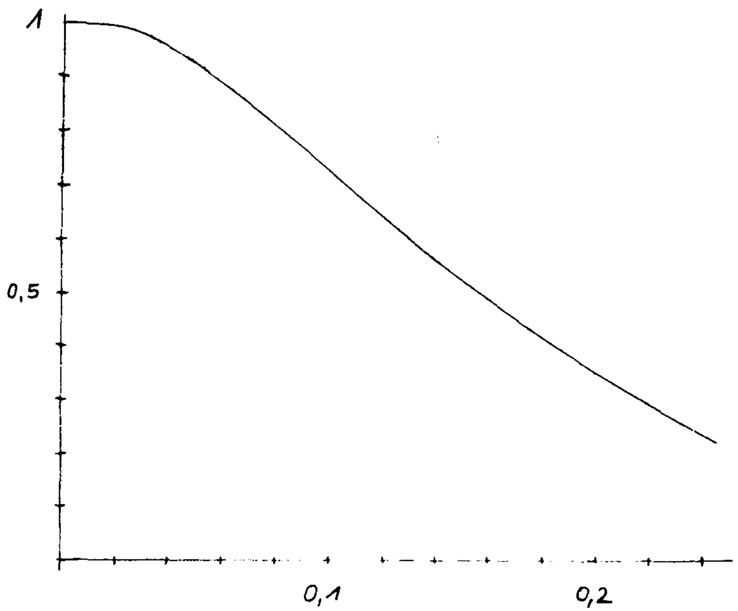
| D/N | 0,08 | 0,10 | 0,15 | 0,20 |
|---|---|---|---|---|
| p | 0,50 | 0,36 | 0,15 | 0,05 |



Dieser Kurve kann man für jede mögliche Qualitätsstufe der Produktion die Annahmewahrscheinlichkeit einer Sendung entnehmen. Eine Sendung mit 8 defekten Teilen wird demnach mit jeweils einer Wahrscheinlichkeit von 50 % angenommen bzw. abgelehnt.

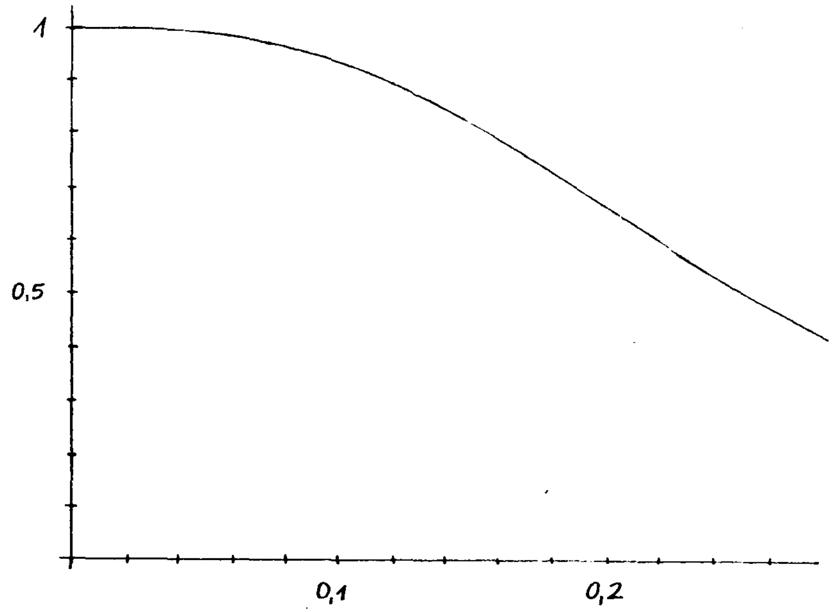
Es gibt nun mehrere Möglichkeiten, die Stichprobenkontrolle zu variieren: Änderung des Stichprobenumfangs,
Änderung der Annahmeregel oder
Änderung der Stückzahl pro Sendung.

Im folgenden Beispiel ist die Operationscharakteristik für N= 100, S= 10 und a= 1 dargestellt. Man sieht, daß die Verkleinerung des Stichprobenumfangs einen flacheren Verlauf der Kurve bewirkt. Das bedeutet insbesondere, daß nur erheblich schlechtere Qualitäten nahezu immer abgelehnt werden bzw. der Bereich der Qualitäten, dienahezu immer angenommen werden, größer wird.






Verändert man die Annahmeregel durch Erhöhung der Annahmezahl a z.B. von 1 auf 2, so vergrößern sich die Funktionswerte der zugehörigen Operationscharakteristik um den Summand $\binom{D}{2}\binom{100-D}{8}:\binom{100}{10}$ so daß der Graph scheinbar nach rechts verschoben wird:




Es bleibt zu untersuchen, welchen Einfluß die Änderung der Stückzahl pro Sendung auf die Operationscharakteristik hat. Ist die Stückzahl N genügend groß und der Stichprobenumfang S verglichen mit N klein, so ändert sich die Berechnung der Wahrscheinlichkeit kaum, wenn man die Elemente der Stichprobe einzeln entnimmt und jedesmal wieder zurücklegt. Dann lassen sich die Wahrscheinlichkeiten nach der Binomialverteilung berechnen. Ist nämlich q der Anteil der defekten Teile in der Sendung, so erhält man genau k defekte Teile in einer Stichprobe der Größe S mit der Wahrscheinlichkeit $p = q^k(1-q)^k\binom{S}{k}$. In diesem Term taucht die Elementanzahl N der Sendung nicht mehr auf, so daß p von N unabhängig ist. Geht man noch einen Schritt weiter, so wird man die Binomialverteilung durch die Poisson-Verteilung ersetzen, sofern der Anteil q der defekten Teile unter 0,1 liegt. Wir erhalten dann die Näherung $p = (Sp)^k e^{-Sp}/k!$ als Wahrscheinlichkeit für genau k defekte Teile.

In der folgenden Tabelle sind die verschiedenen Ansätze gegenübergestellt:

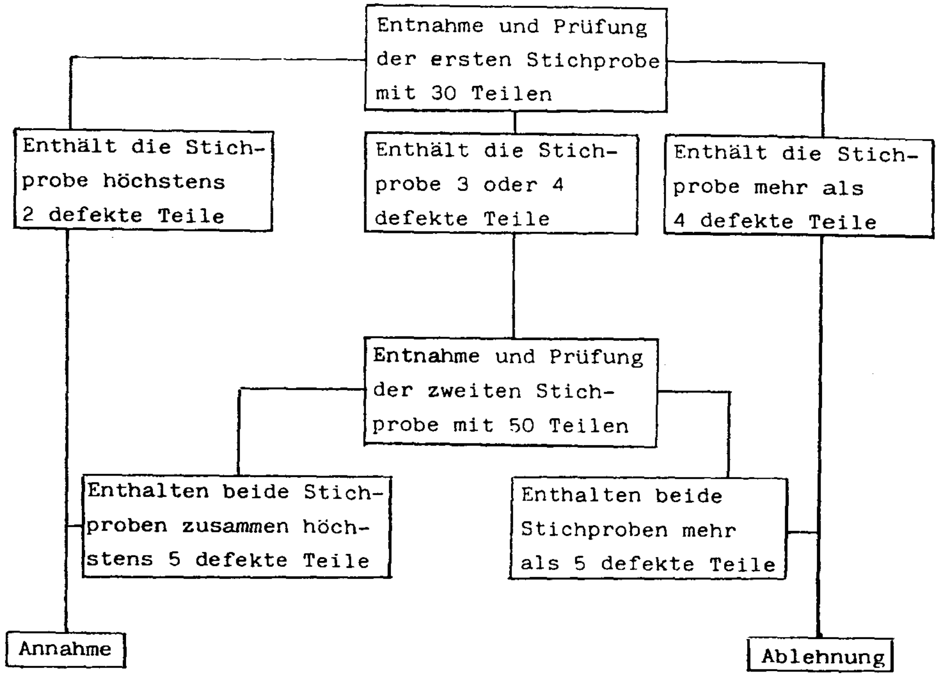
| Anteil der defekten Teile | Annahmewahrscheinlichkeit | | | |
|---|---|---|---|---|
| | N=100 | N=200 | Binom.-vert. | Poisson-vert. |
| 0 | 1,00 | 1,00 | 1,00 | 1,00 |
| 0,02 | 0,96 | 0,95 | 0,94 | 0,94 |
| 0,05 | 0,74 | 0,74 | 0,74 | 0,74 |
| 0,10 | 0,36 | 0,38 | 0,39 | 0,41 |
| 0,15 | 0,15 | 0,16 | 0,18 | |
| 0,20 | 0,05 | 0,06 | 0,07 | |

Weil Qualitätskontrollen schon seit langem Bestandteile von Verträgen sind, hat man sich bemüht, solche Verfahren zu entwickeln und zu beschreiben, die möglichst exakt die von den Vertragspartnern geforderten Eigenschaften haben. Dies hat in den Industrieländern zu Normverfahren geführt, z.B. British Standard 6001 oder DIN 40 080.

Schließlich seien noch Mehrfachstichprobenpläne erwähnt, die in Form eines einfachen Flußdiagramms das Verhalten bei allen auftretenden Fällen regeln, insbesondere, wenn man auf Grund einer Stichprobe keine eindeutige Aussage über die Qualität einer Sendung machen kann.



Beispiel eines Doppelstichprobenplans




Originaltitel in "TEACHING STATISTICS" (1980) Vol. 2 Nr. 3
(1981) Vol. 3 Nr. 1
Judgement by Sampling

Übersetzung und Bearbeitung: B. Daun