Der Einfluss der Stichprobengröße

(vgl. Didaktik der Stochastik, Henn)

Problem: „Den Gesundheitsminister sorgt seit langem die hohe Verbreitung von Krankheiten des Herz-Kreislauf-Systems, also der typischen Zivilisationskrankheit. Aus langjährigen Erfahrungen weiß man sicher, dass bereits 30% der Bevölkerung über 40 Jahren unter einer solchen Erkrankung leiden. Nun berichtet ein jung-dynamischer Ministerialbeamter von einer neuen empirischen Studie mit Personen über 40, die regelmäßig in ihrem Leben eine Ausdauersportart wie Jogging, Radfahren o. ä. betrieben haben. Dort habe sich gezeigt, dass eine Erkrankung des Herz-Kreislauf-Systems in dieser Gruppe signifikant (bei 5%-Niveau) weniger häufig sei als in der Gesamtbevölkerung über 40. Daher schlage er eine aufwendige Kampagne zur Förderung des Ausdauersportes vor. Angesichts der wie üblich knappen Haushaltsmittel einerseits, seinen Erfahrungen mit jung-dynamischen Mitarbeitern andererseits zögert der Minister mit der Entscheidung und fragt stattdessen zunächst einmal nach der Größe der untersuchten Stichprobe. Der Beamte antwortet:

a) n = 20, b) n = 100, c) n = 10.000.

In welchem der drei Fälle würden Sie sich als Minister am ehesten für die Kampagne entscheiden?“

Spontan neigen viele zur Antwort c), denn: je größer die Stichprobe, desto sicherer die Befunde. Indessen: Ist die „Sicherheit“ nicht in allen drei Fällen gleich, weil irgendwie im gleichen Signifikanzniveau ausgedrückt?

Und hier beginnt das Nachdenken: Bei gleichem Signifikanzniveau müsste doch die Größe der Stichprobe, die das signifikante Ergebnis liefert, gleichgültig sein. Ist also die Rückfrage des Ministers sinnlos? Was interessiert den Minister eigentlich? Vor allem doch wohl, ob die Sporttreibenden so erheblich weniger häufig erkranken, dass sich die hohen Kosten der Kampagne rechtfertigen ließen durch das erhebliche Ausmaß der erhofften Senkung des Erkrankungsrisikos. Betrüge dieses Risiko bei den Sporttreibenden immer noch 29,9%, so wäre die aufwendige Kampagne kaum sinnvoll. Diese Frage drängt nun zu einer genaueren Betrachtung des verwendeten Signifikanztestes, den die Schüler selbständig rekonstruieren:

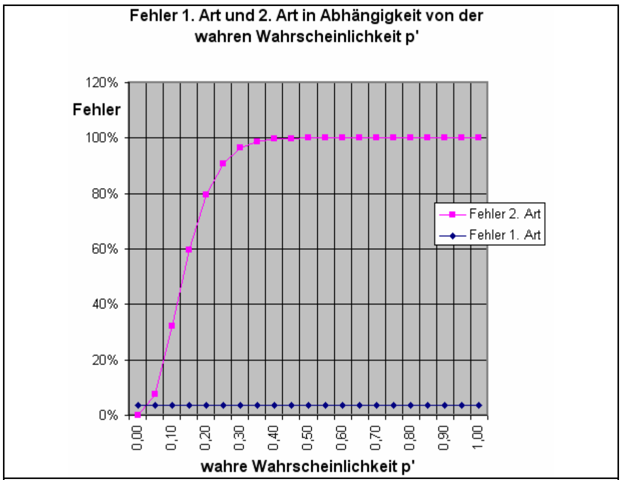
Einseitiger Test: H_0: p = 0,3; H_1: p < 0,3; X: Anzahl der Sporttreibenden mit Herz-Kreislauf-Erkrankung; a = 0,05; hieraus berechnet sich der Ablehnungssbereich {0, 1, ..., k_1} mit a) k_1 = 3,   b) k_1 = 23,   c) k_1 = 2.925. (*Nachrechnen ! Hier stimmt etwas nicht.*)

(Jetzt kann die Operationscharakteristik des Tests, die es erlaubt, den ß-Fehler zu beurteilen, mit dem Computer punktweise berechnet und graphisch dargestellt werden:)

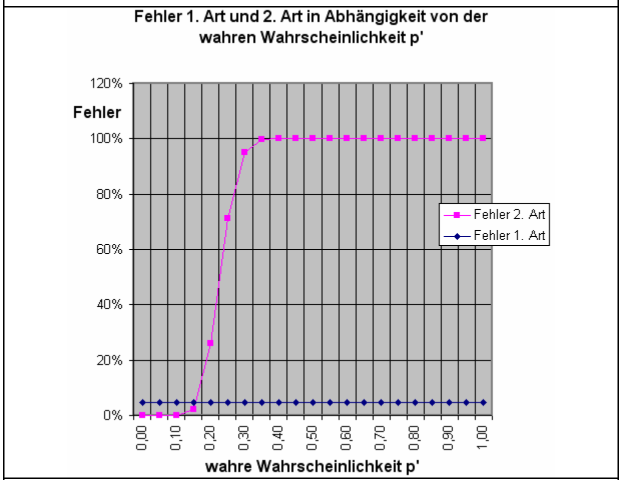
Fehler 2. Art in Abhängigkeit von der Wahrscheinlichkeit *p'*:

$$\beta(p') = 1 - binomcnf(n, p', k) = 1 - \sum_{k=0}^{k_1} b(n, p', k)$$

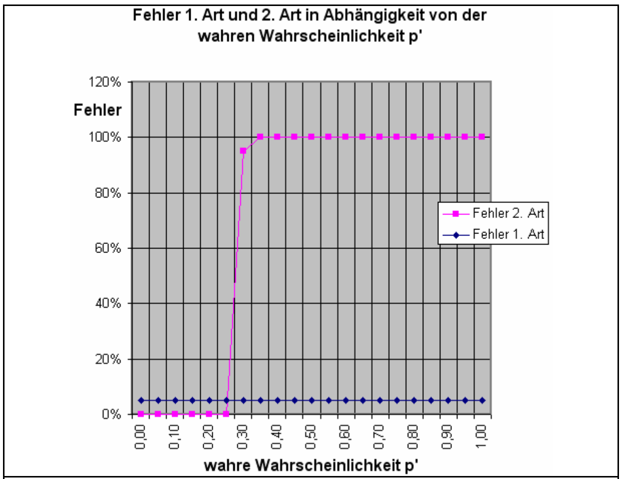
(vgl. *Gütefunktion von Tests.xls*)



$n = 20$; $p = 0,3$; $\alpha = 5\%$; $k_1 = 2$



$n = 100$; $p = 0,3$; $\alpha = 5\%$; $k_1 = 22$



$n = 10000$; $p = 0{,}3$; $\alpha = 5\%$; $k_1 = 2924$

Trennschärfe von Tests

Die Diskussion und Interpretation dieser Funktionsverläufe bringt dann die Einsicht: Bei n = 10.000 ergibt sich fast sicher ein signifikantes Ergebnis schon dann, wenn der interessierende Unterschied in der Erkrankungswahrscheinlichkeit minimal und damit für den Minister praktisch irrelevant ist; bei n = 20 können wir dagegen nur dann ein signifikantes Ergebnis erwarten, wenn sich beide Gruppen erheblich voneinander unterscheiden. Und genau darum geht es dem Minister! Entgegen dem spontanen Eindruck sollte ihn daher am ehesten noch Antwort a) zu einer Kampagne veranlassen. Fazit: Die Signifikanz eines Unterschiedes allein ist häufig irrelevant, wo es auf die absolute Größe des Unterschiedes ankommt. Die Rückfrage des Ministers ist deshalb sinnvoll, insofern sie der berechtigten Skepsis gegenüber Signifikanz in sehr großen Stichproben entspringt; noch sinnvoller wäre natürlich die direkte Frage nach der Größe des Unterschiedes.