R. Krell (www.r-krell.de), 17./18. Januar 2018

Hypothesentest paranormaler Fähigkeiten

Am 19. Oktober 2017 hatte ich die Fernsehsendung der Reihe „Planet Wissen" mit dem Titel „Übersinnliche Phänomene – Was ist dran?" gesehen, die in mehreren ARD-Programmen, u.a. im WDR, ausgestrahlt worden war, und z.Z. noch in der Mediathek vorhanden ist.

U.a. wurde dort von einer Frau berichtet, die behauptete, verschmutzte Blumenerde durch Pendeln erkennen zu können. Sie wollte damit den von der GWUP -- Gesellschaft zur wissenschaftlichen Untersuchung von Parawissenschaften -- ausgesetzten Preis von 10.000,- € gewinnen, wenn sie ihre Fähigkeiten unter Laborbedingungen nachweisen könne.

Sie war mit folgendem Test einverstanden: Auf insgesamt 13 langen Tischen standen in ausreichendem Abstand jeweils 10 Schalen mit Blumenerde, wobei jeweils genau eine Schale pro Tisch durch Beimischung von Zucker verunreinigt war. Das Preisgeld sollte ausgezahlt werden, wenn sie an mindestens 10 der 13 Tische die Schale mit der verunreinigten Erde richtig identifizieren könnte.

Die Dame, die von ihren besonderen Fähigkeiten überzeugt war, lies einen Gegenstand über den Schalen pendeln und bezeichnete pro Tisch eine Schale als verunreinigt. Sie war überzeugt, an allen 13 Tischen die richtige Schale gefunden zu haben. Tatsächlich hatte sie allerdings nur an zwei Tischen die verunreinigte Erde gefunden, hielt das aber dann immer noch für einen Beweis ihrer übersinnlichen Fähigkeiten, auch wenn sie den Preis nicht gewonnen hatte. Der Versuchsleiter konnte sie anscheinend nicht wirklich davon überzeugen, dass jeder beliebige Teilnehmer allein durch zufälliges Raten mit hoher Wahrscheinlichkeit zwei Treffer erzielen würde.

Die GWUP, die offenbar schon seit Jahrzehnten einmal jährlich in den Räumen der Uni Würzburg Tests mit Pendlern, Wünschelrutengängern und weiteren Kandidatinnen und Kandidaten durchführt, hat übrigens das ausgelobte Geld noch nie bezahlen müssen: kein selbsternanntes Medium konnte seine angeblichen Fähigkeiten unter den mit den Kandidaten abgesprochenen, arrangierten und kontrollierten Bedingungen unter Beweis stellen.

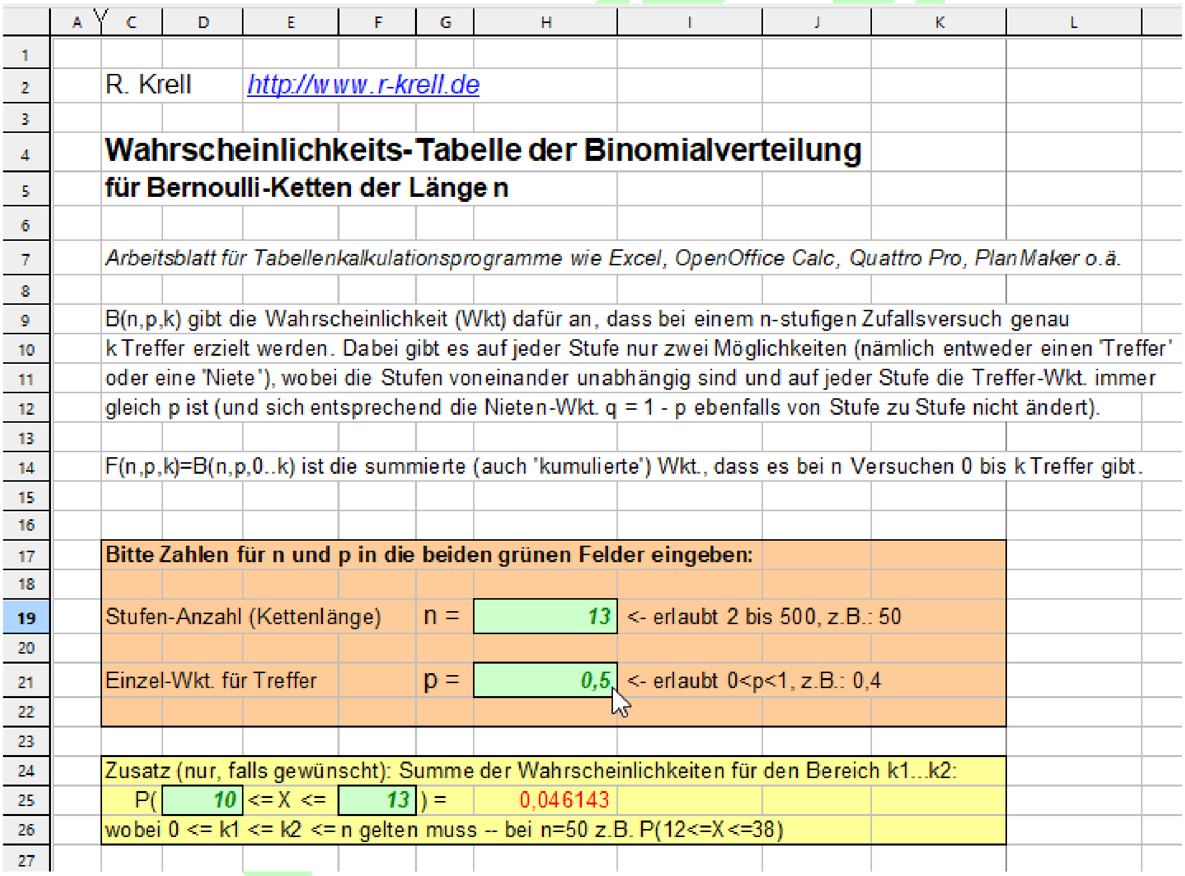
Insoweit hatte mich die Fernsehsendung nicht weiter überrascht. Umso erstaunter war ich, als gegen Ende der Sendung einer der beiden Studiogäste plötzlich über den Test schimpfte: Walter von Lucadou, Leiter der Parapsychologischen Beratungsstelle in Freiburg, behauptete unvermittelt und ohne weitere Erklärung, einen solchen Test könne niemand bestehen, und es sei unethisch, die naive Teilnehmerin, die keine Ahnung von Stochastik habe, so vorzuführen.

Im Fernsehbericht hatte sich aber bis dahin niemand über die Pendlerin lustig gemacht oder abfällig über sie gesprochen. Nur Herr von Lucadou bezeichnete sie jetzt öffentlich als naiv. Wer an einem solchen Test teilnimmt, um Geld zu gewinnen und seine behaupteten Fähigkeiten zu demonstrieren, und sich mit den Versuchsbedingungen einverstanden erklärt hat, muss meiner Meinung nach auch hinnehmen, dass sachlich über eine eventuelle Niederlage berichtet wird. Jeder Sportler im Wettkampf muss von Fans und Berichterstattung weit Schlimmeres ertragen. Die Teilnehmerin muss auch keinen Stochastikkurs absolviert haben, um behaupten zu können, dass sie alle dreizehn (auf jeden Fall aber mindestens zehn) Tische richtig bewerten könne. Der Vorwurf, ein solcher Test sei viel zu hart, um bestanden werden zu können, erschien mir auf Anhieb ebenfalls ungerechtfertigt, blieb aber im Gedächtnis und regte mich nachträglich zu folgendem Nachrechnen an:

Entsprechend ihrer eigenen Behauptung muss man von der Vermutung H_1 ausgehen, die Teilnehmerin verfüge über besondere („paranormale“) Erkenntnismöglichkeiten (ihre Erkenntnisrate p sei also höher als eine willkürlich gesetzte Grenze p_0, d.h. H_1: $p > p_0$). Entsprechend muss man versuchen, zum Beweis von H_1 das Gegenteil, nämlich die Nullhypothese H_0: $p \leq p_0$ zu widerlegen. Geht man – mangels mitgeteilter Informationen – davon aus, dass auch hier mit der oft üblichen Irrtumswahrscheinlichkeit von $\alpha \leq 5$ % für den maximalen Fehler 1. Art gearbeitet wird, und die Entscheidungsregel ja bekannt ist (Preisgeld und damit Ablehnung von H_0 bei 10 bis 13 richtigen Tischen, sonst – bei 0 bis 9 Richtigen – kein Preis bzw. Vereinbarkeit mit H_0), so sollte sich p_0 aus der aufsummierten Binomialverteilung ermitteln lassen, weil

$$(*)\quad B_{n,p,k}\,(13;\,p_0;\,10..13) = P\,(X \leq 13) - P\,(X \leq 9) \leq 0{,}05 = 5\ \%$$

gefordert ist. Mit einem modernen Taschenrechner oder auch durch Probieren (also Einsetzen verschiedener p-Werte) z.B. in meinem Rechenblatt „binomialtabelle.xls“ [http://www.r-krell.de/m.htm#Stochastik] findet man, dass bei $p_0 = 0{,}5$ die Bedingung (*) erfüllt ist (bei $p_0 = 0{,}51$ hingegen nicht mehr):


R. Krell http://www.r-krell.de

Wahrscheinlichkeits-Tabelle der Binomialverteilung
für Bernoulli-Ketten der Länge n

Arbeitsblatt für Tabellenkalkulationsprogramme wie Excel, OpenOffice Calc, Quattro Pro, PlanMaker o.ä.

B(n,p,k) gibt die Wahrscheinlichkeit (Wkt) dafür an, dass bei einem n-stufigen Zufallsversuch genau k Treffer erzielt werden. Dabei gibt es auf jeder Stufe nur zwei Möglichkeiten (nämlich entweder einen 'Treffer' oder eine 'Niete'), wobei die Stufen voneinander unabhängig sind und auf jeder Stufe die Treffer-Wkt. immer gleich p ist (und sich entsprechend die Nieten-Wkt. q = 1 - p ebenfalls von Stufe zu Stufe nicht ändert).

F(n,p,k)=B(n,p,0..k) ist die summierte (auch 'kumulierte') Wkt., dass es bei n Versuchen 0 bis k Treffer gibt.

Bitte Zahlen für n und p in die beiden grünen Felder eingeben:

Stufen-Anzahl (Kettenlänge) n = 13 <- erlaubt 2 bis 500, z.B.: 50

Einzel-Wkt. für Treffer p = 0,5 <- erlaubt 0<p<1, z.B.: 0,4

Zusatz (nur, falls gewünscht): Summe der Wahrscheinlichkeiten für den Bereich k1...k2:
P(10 <= X <= 13) = 0,046143
wobei 0 <= k1 <= k2 <= n gelten muss -- bei n=50 z.B. P(12<=X<=38)

Die GWUP verteilt das Preisgeld also im Prinzip schon, wenn die Pendlerin erkennbar mindestens die Hälfte der Tische richtig beurteilt. Wegen der unvermeidlichen zufälligen Ergebnis-Abweichungen bei kleinen Versuchs- bzw. Tischzahlen reicht es aber nicht, etwa auf 7 (nächste ganze Zahl über dem Erwartungswert von $6{,}5 = 0{,}5 \cdot 13$) Tischen die Schale mit der verunreinigten Erde genau zu lokalisieren. Erst bei 10 richtig gelösten Tischen liegt das Ergebnis soweit über dem zufälligen ‚statistischen Rauschen‘, dass die 10.000 € ausgezahlt werden sollen – obwohl auch bei dieser Entscheidungsregel noch in rund einem von zwanzig Tests (5 %) die Prämie insofern zu Unrecht gewährt würde, weil das gute Abschneiden eben doch nur durch glücklichen Zufall von einem Kandidaten mit in Wirklichkeit unter 50 % Erkennungsrate bei den Tischen erreicht wird.

50 % richtige Tische zu fordern, erscheint auf den ersten Blick sehr fair. Andererseits soll nicht vergessen werden, dass auf einem Tisch nicht etwa nur zwei, sondern zehn Schalen stehen. Die Chance, durch reinen Zufall einen ganzen Tisch richtig zu beurteilen, liegt also nicht bei 50 %, sondern nur bei 10 %!
Und um 10 von 13 Tischen richtig zu beurteilen, müsste der einzelne Tisch auch nicht mit 50 %, sondern mit einer Sicherheit von 76,9 % richtig erkannt werden (erst dann liegt der Erwartungswert bei 10 Tischen) – also 7,7 mal besser als beim unbedarften Raten. Und selbst für ein Medium mit 76,9-%-iger Tischerkennung ist das Risiko 2. Art mit etwa 35 % recht hoch und 7 mal so groß wie das Risiko 1. Art. Das ist ungünstig für die Kandidaten.
(Zur Erinnerung: Ein Fehler 2. Art entstünde, wenn durch ein zufällig schlechtes Stichprobenergebnis ein in Wirklichkeit fähiges Medium fälschlich nicht als solches erkannt würde. Das Risiko 2. Art wäre die Wahrscheinlichkeit für das Auftreten des Fehlers 2. Art beim Test. – Beim Fehler 1. Art wird hier hingegen ein schlechtes Medium wegen zufällig gutem Stichprobenergebnis/glücklichem Raten irrtümlich für eine Person mit besonderer Begabung gehalten. Die Wahrscheinlichkeit hierfür ist das Risiko 1. Art, auch Irrtumswahrscheinlichkeit genannt).

Legt man das Anforderungsniveau für Abiturnoten in NRW zu Grunde, müsste die Pendlerin für 76,9-%-ige Tischerkennung also schon „gut“ sein, um den Preis zu erreichen (ab 75 % der Rohpunkte soll es für eine Abiturarbeit eine glatte Zwei geben; ab 80 % gibt es eine Zweiplus, ab 85 % eine Einsminus, usw.). Einer Person mit nur „ausreichenden“ oder „befriedigenden“ besonderen Eigenschaften will die GWUP die ausgelobten 10.000 € offenbar möglichst nicht zahlen. Liegt hier der Grund für die abfällige Äußerung seitens Herrn von Lucadou? Wenn er den Test als überzogen und als nicht bestehbar verdammt, geht er wohl davon aus, dass es keine Personen mit guten paranormalen Fähigkeiten gibt – wer wollte da widersprechen.

Die Freiburger Beratungsstelle zeigt sich sonst allerdings nicht so skeptisch: nach eigenen Angaben erreichen sie pro Jahr einige Tausend Anrufe von Leuten, die vor allem nicht für verrückt gehalten werden wollen, weil sie für sie unerklärliche Phänomene beobachtet haben. Dr. Dr. von Lucadou, der nicht nur in Physik, sondern auch in Psychologie promoviert hat, nimmt diese Anrufer ernst, hört ihnen unvoreingenommen zu und beruhigt, dass auch andere Ähnliches erlebt hätten. Auch wenn im weit überwiegenden Teil der geschilderten Fälle bald natürliche Erklärungen gefunden werden, bleibt wohl ein kleiner Prozentsatz, wo dies nicht gelingt. Herr von Lucadou berichtet von spukhaften Ereignissen, die zwar abgestellt (weil Lebensbedingungen verändert oder uneingestandene Wünsche erkannt und beachtet wurden), aber nicht wirklich erklärt werden konnten. Letzteres sei allerdings auch nicht sein Ziel; vielmehr gehe es darum, den Anrufern zu helfen. Und solange man noch nicht genau wisse, durch welchen exakten physikalischen Mechanismus bei psychosomatischen Krankheiten z.B. negative Gedanken ein Magengeschwür im eigenen Körper verursachen können, hält er es auch für möglich, dass Gedanken sogar nach außen wirken. Die von ihm nebulös verwendeten Begriffe der Verschränkung (der in der Physik allerdings nur bei Paaren isolierter Elementarteilchen unter ganz bestimmten Bedingungen vorkommt) und der Musterübereinstimmung überzeugen nicht wirklich und ersetzen keine fehlende Erklärung. Hier wird meiner Meinung nach der von ihm erhobene naturwissenschaftliche Anspruch klar verlassen.

Die GWUP veranstaltet die jährlichen Tests übrigens, um zu zeigen, dass es keine paranormalen Fähigkeiten gibt (auch wenn viele der Kandidaten sich für übersinnlich begabt halten – es aber unter Testbedingungen nie nachweisen konnten). Ab welcher ‚Qualität‘ die GWUP einen Preis anbietet, bleibt ihr natürlich überlassen. Und eine Kandidatin, die bis zum Schluss des Tests zu 100 % überzeugt war, alle 13 Tische richtig erkannt zu haben, weil sie ja klare Signale beim Pendeln erfahren habe, hätte sicher auch am Test teilgenommen, wenn sie die vorstehende Rechnung durchgeführt und von der 76,9-%-Marke gewusst hätte.

Zur Abrundung noch die zu erwartende Tischzahl für den völlig unbegabten Rater (der blind auf eines der 10 Schälchen pro Tisch tippt und damit nur eine 10-%-ige Chance auf eine richtige Tischerkennung hat): Der Erwartungswert liegt dann wegen $0{,}1 \cdot 13 = 1{,}3$ bei 1,3 richtig geratenen Tischen – wobei natürlich kein Tisch zu einem Bruchteil richtig sein kann, sondern nur ganze Zahlen als Ergebnisse möglich sind. Wegen $B_{n,p,k}\ (13;\ 0{,}1;\ 1) \approx 0{,}3672$ und $B_{n,p,k}\ (13;\ 0{,}1;\ 2) \approx 0{,}2448$ liegt die Wahrscheinlichkeit, bei zufälligem Raten 1 bis 2 Tische richtig zu haben, bei über 61 %. Nur in einem Viertel der Fälle (25,4 %) wird ein ratender Laie gar keinen richtigen Tisch benennen können. Will man zufälliges Raten mit höchstens 5-%-iger Irrtumswahrscheinlichkeit ausschließen, darf dies erst ab 4 (bis 13) richtig geratenen Tischen passieren. Die 2 von der Kandidatin im Film richtig erpendelten Tische ergeben also wirklich keinerlei Hinweis auf besondere Fähigkeiten!

Quellen und Verweise / Links

- Sendung „Planet Wissen: Übersinnliche Fähigkeiten – Was ist dran?“, gesendet von WDR, SWR und ARD-alpha am 19.10.2017, mit Moderator Jo Hiller und den Studiogästen Thorsten Havener und Walter von Lucadou (Webseite https://www.planet-wissen.de/sendungen/sendung-parapsychologie-100.html mit Link zum 58-minütigen Video https://www.planet-wissen.de/video-uebersinnliche-phaenomene--was-ist-dran-100.html bzw. alternative Video-URL http://wdrmedien-a.akamaihd.net/medp/ondemand/weltweit/fsk0/127/1274826/1274826_14851826.mp4
- Gesellschaft zur wissenschaftl. Untersuchung von Parawissenschaften (GWUP) in Würzburg: https://www.gwup.org/
- Parapsychologische Beratungsstelle in Freiburg: http://www.parapsychologische-beratungsstelle.de
- Videos zum Thema „Geister und Grusel“ auf planet-wissen: https://www.planet-wissen.de/kultur/fabelwesen/geister_und_grusel/pwvideoplanetwissenvideowastunwennmanvonaliensentfuehrtwird100.html, darunter als 6. Video (Parapsychologie – Die Erforschung des Unerklärlichen) ein knapp 4-minütiges Interview mit Walter von Lucadou
- Tabellenblatt zur Berechnung binomialverteilter Wahrscheinlichkeiten „binomialtabelle.xls“ erreichbar unter http://www.r-krell.de/m.htm#Stochastik
- Weitere Hinweise zur Testtheorie
  - in meinem pdf-Dokument „Stochastik in der SII“ (stoch_s2.pdf), erreichbar über http://www.r-krell.de/m.htm#Stochastik
  - in meiner Zusammenfassung „Hinweise für das richtige Testen von Hypothesen“ bei http://www.r-krell.de/m.htm#Stochastik
  - in Klausuren zur Stochastik auf http://www.r-krell.de/m-klausuren.htm

Weiterführende Aufgaben

1. Rechne nach, dass – wie im letzten Absatz des vorstehenden Textes angegeben – beim einseitigen Test erst bei 4 und mehr richtig geratenen Tischen (à 10 Schalen, von denen genau eine verunreinigte Erde enthält) eine signifikante Abweichung vom zufällig zu erwar-

tenden Ergebnis vorliegt, sodass mit $\alpha = 5$ % Irrtumswahrscheinlichkeit eine besondere Fähigkeit attestiert wird. Begründe auch, warum nicht zweiseitig getestet wird.

2. Führe die Rechnung von 1. für eine Irrtumswahrscheinlichkeit von nur 1 % durch.

3. Nimm an, das Preisgeld würde schon bei 8 (und mehr) richtig geratenen Tischen ausgezahlt. Welches p_0 wird dann gefordert?

4. Jetzt soll das Preisgeld schon ausgezahlt werden, wenn H_0: $p \leq 0{,}2$ widerlegt werden kann (der Kandidat also signifikant mehr als doppelt so gut wie ein zufälliger Rater sein soll). Stelle die Entscheidungsregel für $\alpha = 5$ % auf und bestimme den Fehler 2. Art für ein hypothetisches Medium, das in Wirklichkeit bzw. auf lange Sicht a) 40 %, b) 50 %, c) 76,9 % der Tische richtig erkennen kann.

5. Nimm für einen Moment an, es gäbe wirklich Personen mit paranormalen Fähigkeiten, die im Mittel 20 % (oder mehr) der Tische richtig erkennen könnten. Der Stichprobenumfang sei wieder n=13, d.h. jede Person soll wieder 13 Tische bewerten. Danach muss entschieden werden, ob die Person ein begabtes Medium oder ein blind ratender Scharlatan ist. Formuliere die Hypothesen und finde die Entscheidungsregel so, dass ein solches Medium höchstens mit einer Irrtumswahrscheinlichkeit von 5 % für einen Scharlatan gehalten wird. Ermittle auch den Fehler 2. Art, dass bei diesem Test ein wirklicher Scharlatan als vermeintliches Medium mit paranormalen Fähigkeiten durchgeht.

6. Die GWUP-Bedingung, dass pro Tisch genau eine Schale verunreinigt ist, macht es den Kandidaten eigentlich sogar sehr leicht: Glaubt beispielsweise ein Kandidat, bei verunreinigter Erde mache das Pendel eine Kreisbewegung, während es sonst in einer Ebene schwingt, so kann er, selbst wenn das Pendel gelegentlich elliptisch schwingt, einfach die Schale mit der rundesten Ellipse aussuchen, um auf die verunreinigte Schale zu tippen. Außerhalb des Labors und des GWUP-Tests weiß man aber nicht, wo wie viel verunreinigte Erde ist. Ein realitätsnaher Test müsste daher eigentlich pro Tisch eine beliebige, unbekannte Anzahl an Schälchen verunreinigen dürfen, was den Test allerdings wesentlich erschwert. Ermittle, mit welcher Sicherheit dann jede einzelne Schale richtig erkannt werden muss, damit im Mittel 20 % aller Tische richtig erkannt werden können.

7. Die Probleme beim Hypothesentest liegen bekanntlich nicht in der Mathematik, sondern in den vorher getroffenen willkürlichen Festsetzungen. Gehe wieder vom einfachen Test mit genau einer verunreinigten von 10 Schalen pro Tisch aus, und lege p_0 und/oder α sinnvoll fest:
   Diskutiere und begründe, ab welcher Prozentzahl p_0 für einen richtig erkannten Tisch ein Kandidat deiner Meinung nach als Medium mit besonderen Fähigkeiten anerkannt werden sollte und welche Irrtumswahrscheinlichkeit sinnvoll ist, wenn es a) ‚nur' um ein Preisgeld von 10.000 € geht, b) wenn die Kandidatin ihr Geld damit verdient, gutgläubigen Bauherren zum Kauf eines Grundstücks (auch auf einem ehemaligen Deponie-Gelände) zu raten / vom Kauf eines Grundstücks abzuraten / zu einem Umzug zu raten, wenn der Boden ihrer Meinung nach nicht verseucht bzw. verseucht ist, c) wenn der Kandidat für seine nicht verunreinigte Erde besondere heilerische Fähigkeiten behauptet und man befürchten muss, dass ihm im Falle seiner Anerkennung möglicherweise weitere Hunderte Patienten vertrauen, ihm in der Hoffnung auf Heilung ihr Erspartes für ein Schälchen heilende Erde opfern und auf das rechtzeitige Aufsuchen eines ausgebildeten Arztes verzichten.

Ergebnisse der Aufgaben
(wie immer bitte erst nach eigenen schriftlich festgehaltenen Lösungsversuchen lesen):

1. Es soll zufälliges Raten ausgeschlossen werden, d.h. man hofft/vermutet, dass der Kandidat besser als der zufällige Rater ($p_0 = 0{,}1$) ist. Deshalb ist die Vermutung H_1: $p > 0{,}1$ bzw. die zu widerlegende Nullhypothese H_0: $p \le 0{,}1$. Weil H_0 eine kleine Zahl ($p \le$ ..) behauptet, kann sie nur widerlegt werden, wenn in der Stichprobe viele Treffer auftreten, d.h. $X > k$ ist (Grundsätzlich kann man auch $X \ge k_1$ fordern. Nur erweist es sich als geschickter, mit $P(X > k) = 1 - P(X \le k)$ zu arbeiten, weil $P(X \le k)$ tabelliert bzw. in Rechnern implementiert ist, $P(X < k_1)$ hingegen nicht. Hat man k, nutzt man leicht $k_1 = k + 1$. X ist die Trefferzahl).
Das Widerlegen soll nur mit geringer Wahrscheinlichkeit ($\le \alpha$) durch reinen Zufall erfolgen können (schließlich will man auf eine Fähigkeit des Kandidaten schließen), d.h. es wird $P(X > k) \le \alpha$ gefordert. Wegen $P(X > k) = 1 - P(X \le k) \le \alpha$ ergibt sich aus der hinteren Ungleichung die äquivalente Forderung $P(X \le k) \ge 1 - \alpha$. Jetzt kann in der Tabelle oder mit dem Rechner für n=13 und für die größte in H_0 noch erlaubte Wahrscheinlichkeit $p = 0{,}1$ das kleinste k gesucht und gefunden werden, für das gerade $P(X \le k) \ge 0{,}95$ gilt. So erhält man hier $k = 3$. Wegen $X > k$ beginnt der Ablehnungsbereich für H_0 erst bei 4, d.h. erzielt ein Kandidat beim Versuch 4 oder mehr Treffer (d.h. kann 4 oder mehr Tische richtig lösen), so ist es ziemlich unwahrscheinlich, dass sich dieser Stichprobenausgang auf bloßen Zufall zurückführen lässt (Risiko $\le \alpha$) und man will an die besondere Fähigkeit des Kandidaten glauben bzw. hält es für ausreichend wahrscheinlich, um ihn – mit 95-%-iger Sicherheit – als begabtes Medium anzuerkennen.
Erzielt der Kandidat hingegen im Versuch nur 0 bis 3 Richtige, weiß man nichts. Er kann geraten haben oder doch über besondere Fähigkeiten verfügen, hatte aber zufällig Pech bei der Stichprobe. Zur Unterscheidung wären weitere Tests, am besten mit erhöhtem Stichprobenumfang, also deutlich mehr Tischen, nötig.

   Ein zweiseitiger Test würde feststellen, ob der Kandidat besser oder schlechter als ein zufälliger Rater ist. Von einem Medium erwartet man aber, dass es besser ist. Deswegen darf hier nur diese Richtung interessieren.

2. Es gelten die gleichen Umformungen wie oben, nur ist jetzt $1 - \alpha = 0{,}99$. Es wird also k für $P(X \le k) \ge 0{,}99$ gesucht. Man findet $k = 4$; der Ablehnungsbereich für H_0 verringert sich also auf 5 bis 13 Treffer. Immer zieht eine Verringerung der Irrtumswahrscheinlichkeit eine Verkleinerung des Ablehnungsbereiches nach sich. Wer 100% sicher sein will und mit 0% Irrtumswahrscheinlichkeit ablehnen will, darf nie ablehnen und kann gar nichts entscheiden!

3. Wie vor ist auch hier die Vermutung, dass der Kandidat etwas besonderes kann, also H_1: $p > p_0$ für seine Trefferwahrscheinlichkeit p gilt. Anders als in den Aufgaben 1 und 2 ist aber nicht k, sondern p_0 gesucht. Die Gegenhypothese zu H_1 ist wieder H_0: $p \le p_0$ und wird widerlegt, wenn die Anzahl X der Treffer hoch ist: $X > k$ bzw. $X \ge k{+}1 = 8$. Es ist also $k = 7$ gegeben. Nun muss gesucht/probiert werden, für welches möglichst große p (bzw. p_0) bei n=13 noch die Bedingung $P(X \le 7) \ge 0{,}95$ gilt. Dies ist bei $p_0 = 0{,}35$ der Fall (wenn man nur ganze Prozente zulässt). Die gegeben Entscheidungsregel gilt also für einen Test, bei dem eine signifikant bessere Treffer- bzw. Erkennungsrate als 35 % nötig ist, um den Test zu bestehen.

4. Hier ist schon H_0: $p \le 0{,}2$ gegeben (die Vermutung war also H_1: $p > 0{,}2$), ebenso sind n=13 und $\alpha = 0{,}05$ bekannt. Wie in den Aufgaben 1 und 2 wird k bzw. der Ablehnungsbereich und

damit die Entscheidungsregel gesucht. Wegen $P(X \leq 5) \approx 0{,}97$ (aber $P(X \leq 4) \approx 0{,}90$ noch zu klein) ist $k = 5$ und H_0 wird abgelehnt (und damit H_1 indirekt bestätigt), wenn der Kandidat 6 bis 13 Tische richtig erkennt.
Der Fehler 2. Art wird begangen, wenn ein tatsächlich fähiger Kandidat (d.h. mit in Wirklichkeit bzw. auf lange Sicht $p > 0{,}2$) irrtümlich nicht erkannt wird, weil gerade die vorgeführte Stichprobe zufällig schlecht ausfällt, d.h. dort die Trefferzahl $X \leq k$ ist.
Mit dem schon berechneten $k = 5$ muss man also nur für $n=13$ und die in a) bis c) angegebenen Wahrscheinlichkeiten die Werte von $P(X \leq 5)$ berechnen und erhält a) 57,4 %, b) 29 % und c) 0,33 %. Je besser der Kandidat in Wirklichkeit ist, desto geringer ist sein Risiko, im Test zu versagen bzw. falsch beurteilt zu werden.

5. Jetzt geht es darum, nicht irrtümlich (bzw. mit höchstens einer Wahrscheinlichkeit von 5 %) für einen Scharlatan gehalten zu werden. Ein Scharlatan kann (mangels besonderer Fähigkeiten) nur wenige Tische richtig erkennen, d.h. für ihn gilt die Vermutung H_1: $p < 0{,}2$. Dann ist natürlich auch das Ordnungszeichen in H_0 genau umgekehrt wie in allen vorstehenden Aufgabe, d.h. H_0: $p \geq 0{,}2$. Dieses H_0 wird abgelehnt, wenn nur wenige Treffer erzielt werden, also $X \leq k$ gilt (Weil $P(X \leq k)$ direkt nachgesehen/berechnet werden kann, nimmt man hier das Gleichheitszeichen besser dazu – vgl. Anmerkung in der Lösung 1). Das Risiko der irrtümlichen Ablehnung soll wieder durch $\alpha = 5$ % beschränkt sein. Es wird für $n = 13$ und $p_{(0)} = 0{,}2$ also das größte k gesucht mit $P(X \leq k) \leq 0{,}05$. Weil schon $P(X \leq 0) \approx 0{,}055$ über α liegt, gibt es kein solches k, also keinen Ablehnungsbereich! Man müsste bei diesem Test jedem glauben, dass er ein begabtes Medium sein kann, wenn man die Irrtumswahrscheinlichkeit auf 5 % beschränkt. Das bedeutet nicht, dass jetzt jeder Kandidat ein Medium ist – man weiß es nicht und kann es mit diesem Test weder bestätigen noch ausschließen. Eine wirkliche Entscheidung ist also (wenn man α nicht vergrößern will) nur mit einer größeren Zahl von Tischen möglich (Bei $n = 20$ Tischen würden immerhin Kandidaten, die gar keinen Tisch richtig erkennen, als Scharlatane entlarvt. Bei dem Stichprobenumfang 20 wäre aber auch schon $P(X \leq 1) \approx 0{,}069$ größer als α. Für trennschärfere Aussagen müssten also noch deutlich mehr Tische genommen werden!)
Der Vergleich der Aufgaben bzw. Lösungen 4 und 5 zeigt, wie wichtig die Richtung des Tests bzw. die Abhängigkeit von der ursprünglichen Vermutung und der zugehörigen Nullhypothese ist: n, $p_{(0)}$ und α sind ja beidesmal gleich. Nur die Umkehrung des Ordnungszeichens in den Hypothesen sorgt entweder hier für den (unmöglichen) Ablehnungsbereich mit weniger als 0 Treffern oder für die Ablehnung bei 6 bis 13 Treffern (in Aufgabe 4). Dazwischen – hier bei 0 bis 5 Treffern – ist beides, aber eben keine Entscheidung möglich, weil sich das Ergebnis nicht deutlich genug (signifikant) von der möglichen zufälligen Streuung unterscheidet.
P.S.: Neben Scharlatanen, die wissentlich vorgeben, nicht vorhandene Fähigkeiten zu besitzen (und sich mit einem Rest von Gewissen in wichtigen Fällen – wie etwa in Aufgabe 7 b) und c) geschildert – hoffentlich zurückhalten), gibt es leider offenbar nicht wenige Menschen, die gegen alle Vernunft von ihren paranormalen Fähigkeiten überzeugt sind und noch glauben, anderen mit ihrer vermeintlichen Gabe helfen zu können oder zu müssen, obwohl sie die behaupteten Fähigkeiten in Wirklichkeit gar nicht haben. Die letztere Sorte ist noch gefährlicher als die Scharlatane.

6. Ein Tisch ist nur richtig, wenn alle zehn Schalen auf dem Tisch richtig beurteilt wurden, d.h. ${p_{Schale}}^{10} \geq 0{,}2$ gilt. Damit muss $p_{Schale} \geq 0{,}2^{0,1} \approx 0{,}85$ sein, d.h. jede einzelne Schale müsste mit mindestens 85-%-iger Sicherheit richtig beurteilt werden. (Für die GWUP-Bedingung, dass jeder Tisch mit 50-%-iger Wahrscheinlichkeit richtig sein soll, müsste sogar $p_{Schale} \geq 0{,}5^{0,1} \approx 0{,}93$ sein).

7. Bei a) könnte – weil es ‚nur' um Geld im überschaubaren Rahmen geht – durchaus wie in den Aufgaben 3 und 4 mit $p_0 = 0{,}35$ oder sogar nur $p_0 = 0{,}2$ sowie $\alpha = 0{,}05$ gearbeitet werden (wobei der Testanbieter riskiert, ganz gelegentlich Geld auch an einen zufällig glücklichen, aber unbegabten Teilnehmer auszahlen zu müssen; dieses Risiko sinkt mit höherem p_0 oder geringerem α). Selbst $p_0 = 0{,}1$ wäre noch – vielleicht mit geringerer Prämie – möglich, wenn der Anbieter überzeugt ist, dass es gar keine Kandidaten mit besonderen Fähigkeiten gibt, und eine irrtümliche Gewinnauszahlung in jedem 20. oder zumindest 100. Fall verschmerzen kann (Aufgaben 1 und 2 mit $\alpha = 0{,}05$ bzw. $\alpha = 0{,}01$).
   Anders sieht es hingegen in den Fällen b) und c) aus. Hier wären Gefahren für das Leben und die Gesundheit vieler Menschen möglich und große Summen etwa für Umzüge, den Abbruch nicht verwendbarer Häuser o.ä., wenn sich das angebliche Medium irrt. Insofern müssen hier höhere Anforderungen gestellt werden: p_0 sollte schon deutlich über 0,5 (besser bei 0,95 oder noch höher) liegen und die Irrtumswahrscheinlichkeit dafür, dass ein Nichtskönner als vermeintliches Medium anerkannt wird, höchstens1 % oder noch kleiner sein. Dafür muss man dann in Kauf nehmen, dass man auch ein eigentlich recht gutes Medium (wenn es ein solches gäbe) nicht erkennt. Das würde zwar die Person mit besonderen Fähigkeiten ärgern (und beschränkt evtl. zu Unrecht ihre Verdienstmöglichkeiten), ist aber insgesamt gesehen wohl der geringere Schaden als der mögliche hohe Gesundheits- und Vermögensschaden von ganz vielen Kunden eines unfähigen Mediums.

(Fehler in meiner Lösung? – Bitte Nachricht an mail@r-krell.de. Danke!)