Statistische Unterhaltungen[1]

PETER PETOCZ UND ERIC SOWEY, SIDNEY

[1] Original: *„Statistical diversions“* in Teaching Statistics 35 (2013) 2, S. 107–111 bzw. 35 (2013) 1, 72–76
Übersetzung und Bearbeitung:
LAURA MARTIGNON UND JOACHIM ENGEL, LUDWIGSBURG

Vorbemerkung: *In der regelmäßigen Kolumne „Statistical Diversions“ der Zeitschrift Teaching Statistics werfen die Autoren einige Fragen auf, die im Folgeheft ausführlich diskutiert werden. Wir stellen hier die Fragen aus dem Heft 1, 2013 und die Diskussion derselben Fragen im Folgeheft vor.*

Frage 1:

Die bahnbrechenden statistischen Ideen und Methoden zum Design von Experimenten gehen auf R. A. Fisher zurück, einem der Begründer moderner statistischer Inferenz. Auf Seite 11 seines Buches *The Design of Experiments* führt Fisher erinnerungsträchtig in das Thema ein:

„Eine Lady erklärt, dass sie beim Kosten einer Tasse Tee mit Milch unterscheiden kann, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde. Wir betrachten das Problem ein Experiment zu konzipieren, mit dem diese Behauptung überprüft werden kann … Unser Experiment besteht auch acht Tassen Tee, vier auf die eine und vier auf die andere Weise gefüllt, die dann der Versuchsperson zur Beurteilung in zufälliger Reihenfolge vorgelegt werden. Die Aufgabe besteht darin, die 8 Tassen in zwei Gruppe à 4 Tassen aufzuteilen, die der Befüllung der Tassen entspricht“.

Fisher hat diesen Ansatz nicht völlig neu erfunden – er bezieht sich auf ein tatsächliches Vorkommnis. Wer war die ‚Tea tasting lady‘ ? Und was waren die tatsächlichen Umstände, auf denen Fishers Erzählung basiert?

Frage 2:

Was ist ein „Placebo“? In welchen experimentellen Kontexten ist ein Placebo nützlich? Gibt es Situationen, wo man sich hüten sollte, ein Placebo zu verabreichen?

Frage 3:

Benennen Sie drei wissenschaftliche Disziplinen aus den Natur- und Sozialwissenschaften, in denen die Beziehung zwischen interessierenden Variablen gewöhnlich mit Experimentalstudien untersucht wird. Benennen Sie drei Disziplinen, in denen die Beziehung zwischen den interessierenden Größen auf nicht-experimentelle Weise untersucht wird. Nennen Sie auch drei Disziplinen, in denen sowohl Experimental- wie auch Beobachtungsstudien üblich sind.

Frage 4:

Ihnen wird in einem einführenden Statistikkurs folgende Frage gestellt: „Warum ist es für eine valide statistische Schlussfolgerung nicht wichtiger eine repräsentative Stichprobe als eine Zufallsstichprobe zu haben?“ Wie würden Sie antworten?

Frage 5:

Ein spezieller Typus von Graph wurde ursprünglich durch ein Gedankenexperiment gerechtfertigt: Ein Mann solle all das Geld auf einen Stapel legen, das er jeden Tag verdient. Auf welche Art von Graph wird hier Bezug genommen, und wer war der Theoretiker, der die Darstellung auf diese Weise begründete? Sah dieser Theoretiker den Graph eher als statistisches oder als mathematisches Objekt?

Wie die Leserin und der Leser sicherlich bereits entdeckt haben, gibt es in der theoretischen und angewandten Wahrscheinlichkeitsrechnung viele Rätsel und Paradoxa. Im Jahre 1654 lud der französische Mathematiker Blaise Pascal seinen großen Zeitgenossen und Freund Pierre de Fermat zur gemeinsamen Untersuchung zweier grundsätzlicher Fragen ein: Wie soll einem zufälligen Ereignis eine numerische Wahrscheinlichkeit zugeordnet werden und wie soll die Wahrscheinlichkeit von zwei zusammengesetzten Ereignissen bestimmt werden? Es war zu diesen Fragen nichts Wissenschaftliches geschrieben worden seitdem – Jahrhunderte vorher – eine tiefgreifende Erkenntnis in dem Buch Prediger (9:11) der Bibel geschrieben worden war: „… Zeit und Zufall sind für alle [Menschen] gleich“. Aber während die Zeit seit ca. 4000 Jahren gemessen wurde, sind die Bemühungen den Zufall systematisch zu messen höchstens 500 Jahre alt, und die damit einhergehenden Probleme sind immer noch nicht zur Zufriedenheit aller gelöst. (Für einen kurzen Überblick über diese Fragen siehe I. J. Good (1959), Kinds of probability, *Science*, **129**, 20. Februar, 443–447).

Es sind vor allem die subtileren Aspekte dieser ungelösten Fragen, die Ungereimtheiten in der Wahrscheinlichkeitsrechnung hervorbringen: Wenn sie vernachlässigt oder nicht erkannt werden, entstehen bald Schwierigkeiten in der Interpretation. Ein weiteres Hindernis ist die menschliche Intuition. Intuition wird definiert als „Wissen ohne viel Nachdenken“ (Gigerenzer, 2008, *Bauchentscheidungen: Die Intelligenz des Unbewussten*, Berlin Verlag). Wenn man Intuitionen über gewisse Ereignisse besitzt, mag die anerkannte Theorie darüber gegen die Intuitionen verstoßen: Daraus entsteht ein Paradox. Ein Paradox ist eine Aussage, die als „wahr“ bzw. „falsch“ empfunden wird, obwohl die akzeptierte Theorie sie als „falsch“ bzw. „wahr“ bezeichnet. Ein Paradox zu „erklären“ oder „aufzulösen“ bedeutet zu klären, wo, warum und zu welchem Grade die Intuition der akzeptierten Theorie weichen muss. Zum Teil überrascht die Hartnäckigkeit unserer Intuition gegenüber manchen Aussagen der Wahrscheinlichkeitstheorie.

Viele Fachbegriffe der Stochastik sind durch Worte ausgedrückt, die auch sonst in der Umgangssprache verwendet werden. Dies kann manchmal eine Quelle der Verwirrung sein. Zwei derartige Begriffe sind „zufällig“ und „unabhängig“.

Im allgemeinen Sprachgebrauch werden die folgenden Wörter als Synonyme für „zufällig“ betrachtet: willkürlich, wahllos, chaotisch, ungeordnet, ohne Muster, unberechenbar. Der populäre Begriff des Zufalls ist offenbar ein eher unscharfer! Für einen Statistiker ist der Zufall ein Konzept, das gleichzeitig einfacher und komplexer als die umgangssprachliche Vorstellung ist. Ein Statistiker beschäftigt sich mit Zufälligkeit in zwei speziellen Situationen – (i) beim Definieren von Zufälligkeit und dem Erzeugen einer Sequenz von Zahlenwerten, die dieser Definition entsprechen (als theoretische Grundlage des Konzepts eines „Zufallsvariable“) und (ii) beim operativen Definieren von Verfahren zur Auswahl der Elemente einer „Zufallsstichprobe“ aus einer Population von Datenwerten. Für diese beiden Zwecke sind die erforderlichen Attribute von Zufälligkeit „Fehlen von Mustern“ und „Unvorhersehbarkeit“. Daher ist auf den ersten Blick die Zufälligkeit des Statistikers einfacher und gezielter als die umgangssprachliche unscharfe Verwendung.

Aber dann kommt die Notwendigkeit praktische Mechanismen zu ersinnen, von denen man hofft, dass sie Folgen ohne Muster erzeugen – sowie die Notwendigkeit Musterlosigkeit zu definieren und zu evaluieren. Denn andernfalls können wir nicht beurteilen, ob Fortschritte in Richtung auf unser ultimatives Ziel vorliegen: Die Erzeugung einwandfreier Zufallsfolgen. Die Intuition des Laien mag bei solcher Ausführlichkeit protestieren: „Ich kann doch problemlos eine Folge von Zahlen ohne Muster aufschreiben“, lautet der paradoxe Aufschrei des Laien. Aber sobald diese Bemühungen genauer geprüft werden, ist der Anspruch des Laien in der Regel schnell verworfen.

Es wird deutlich, dass effektive zufallserzeugende Mechanismen auf eine Automatisierung hinauslaufen, die mit minimalem menschlichen Input abläuft. In den letzten 80 Jahren gab es intensive Bemühungen derartige Hardware- und Software-basierten Verfahren zu realisieren. Enorme Fortschritte sind dabei erzielt worden, aber Perfektion ist noch nicht erreicht. Diese Aktivitäten schritten Seite an Seite mit dem technischen Fortschritt in der Mustererkennung voran. Um einen Eindruck von der Technisierung zu geben: Wir müssen Typen von wiederkehrenden Muster definieren, die in einer Folge von Zufallsdaten fehlen sollen. Dann müssen wir zuverlässige Tests entwickeln, die solche Muster – soweit vorhanden – auch erkennen. Da die Anzahl der denkbaren Muster überabzählbar groß ist, erkennen wir, dass die Zertifizierung perfekter Musterlosigkeit in der Praxis tatsächlich ein unerreichbares Ideal ist. Ein unvollkommener Kompromiss besteht dann darin, nur auf das Vorhandensein von „offensichtlichen“ Mustern zu testen.

Selbst wenn wir diese Art von Unvollkommenheit akzeptieren, die bestenfalls in die Nähe von Zufallsfolgen führen kann, so gibt es noch eine weitere Herausforderung. Die Unvorhersehbarkeit des nächsten zu erzeugenden Wertes (gegeben alle vorangegangenen Werte) ist nicht alleine durch das Fehlen von Mustern garantiert. Was ebenfalls benötigt wird, ist die Unabhängigkeit der erzeugten Werte. Das ist etwas qualitativ Anderes: Während Muster etwas Deterministisches sind, ist Unabhängigkeit eine stochastische Eigenschaft. Betrachten wir ein Beispiel: Angenommen, wir versuchen eine lange Zufallssequenz aus den zehn Ziffern 0, …, 9 zu erzeugen. Dann muss, um die Unabhängigkeit zu gewährleisten, der zufallserzeugende Mechanismus die Bedingung erfüllen, dass die Wahrscheinlichkeit fest ist, mit der die nächste erzeugende Zahl z. B. eine 5 ist. Die nächste Ziffer muss unabhängig von den Ziffern sein, die unmittelbar davor erzeugt wurden.

Es gibt auch tiefer gehende technische Einzelheiten in der statistischen und philosophischen Literatur zum Zufallsbegriff. Sie weisen alle auf einen Schluss hin: Perfektionistische Verfolgung der Zufälligkeit in der Datenerzeugung kann in der Praxis der eigenen Komplexität zum Opfer fallen.

Die Erwähnung der Unabhängigkeit regt eine nähere Betrachtung dieses Begriffs an. Auch hier hat die Intuition ihre eigene unbewusste Erkenntnis: „Dinge, die unabhängig voneinander sind, sind unverbunden.“ Für einen Statistiker ist diese Formulierung viel zu vage um nützlich zu sein. Tatsächlich kann diese Formulierung auch sehr verwirrend sein. Viele Lernende verwechseln unabhängige Ereignisse mit sich gegenseitig ausschließenden Ereignissen. Dies mag durch die Tatsache verursacht sein, dass sich gegenseitig ausschließende Ereignisse in einem Venn-Diagramm von zwei nicht-überlappenden (und somit unverbundenen) Kreisen dargestellt werden.

In der Wahrscheinlichkeitstheorie ist die Unabhängigkeit von zwei Ereignissen A und B ganz knapp durch die Beziehung $P(A \cap B) = P(A) \cdot P(B)$ definiert, wobei $P(A \cap B)$ die Wahrscheinlichkeit ist, dass A und B beide auftreten. Dieser Zusammenhang ist im Übrigen sowohl eine notwendige wie hinreichende Bedingung für die Unabhängigkeit. Und diese Eigenschaft ist auch gültig für Ereignisse, deren Wahrscheinlichkeit Null („unmögliche Ereignisse“) ist. Unglücklicherweise verursachen Ereignisse, deren Wahrscheinlichkeit Null, eine Reihe von Paradoxien. Zum Beispiel ist jedes unmögliche Ereignis unabhängig von sich selbst.

Wenn wir unmögliche Ereignisse ausschließen, dann kann die obige Bedingung der Unabhängigkeit überführt werden zu: $P(A|B) = P(A)$ und $P(B|A) = P(B)$. Jede dieser Versionen ist auch eine notwendige und hinreichende Bedingung für die Unabhängigkeit von A und B. Für Lernende der Statistik bieten diese neuen Formate etwas Attraktives – die Gelegenheit, ihre Intuition zu erneuern in die Richtung, wie es ein Statistiker versteht. Um die erste Version zu nehmen: Unabhängigkeit bedeutet, dass sich die Wahrscheinlichkeit für Ereignis A nicht ändert, egal ob das Ereignis B auftritt oder nicht. Aber selbst wenn sie diese Deutung aufgenommen haben, bleibt die Intuition vieler Studenten in dem gängigen Begriff der Unabhängigkeit verhaftet. Wenn somit eine Aufgabe mit den Ereignissen A und B die formale Lösung hat, dass A und B unabhängig sind, dann mögen Studierende darauf bestehen, dass dieses Resultat paradox ist, wenn es irgendeinen verbalen Hinweis innerhalb der Aufgabe gibt, der explizit oder implizit darauf hinweist, dass die beiden Ereignisse irgendwie logisch miteinander verbunden sind. Zum Beispiel seien zwei Münzen zufällig geworfen. Ereignis A sei „Die erste Münze zeigt Kopf“, Ereignis B „die beiden Münzen zeigen das Gleiche“. Dann ist $P(A) = 0{,}5$, $P(B) = P(KK) + P(ZZ) = 0{,}5$, $P(AB) = P(HH) = 0{,}25$. Also $P(A \cap B) = P(A) \cdot P(B)$ und die Ereignisse sind unabhängig. Aber die Intuition diktiert uns etwas anderes! „Wenn die erste Münze Kopf ist und wir erfahren, dass beide Münzen das Gleiche anzeigen, dann hat die zweite Münze gar keine Wahl mehr. Es muss auf ein zweites Mal Kopf hinauslaufen! Klar, dass das Ergebnis der zweiten Münze logischerweise dadurch eingeschränkt (und somit verbunden) ist, wie die erste Münze gefallen ist. Das sind abhängige Ereignisse, egal, was der Statistiker sagt!

Hier sind nun unsere Antworten auf die eingangs erwähnten fünf Fragen.

Frage 1:

Die „Tea tasting lady“, deren zuversichtliche Behauptung R. A. Fisher den Anstoß gab, über die Planung von Experimenten nachzudenken, war Dr. Muriel Bristol-Roach. Fishers Biograph beschreibt die historischen Teezeiten der Rothamsted Agricultural Research Station in Hertfordshire, offenbar im Jahre 1921 oder 1922:

„Es geschah an einem Nachmittag, als [Fisher] eine Tasse Tee der Dame neben ihm, Dr. B. Muriel Bristol, anbot. Sie lehnte ab mit dem Hinweis, dass sie eine Tasse bevorzugt, in die die Milch zuerst gegossen wird. ‚Unsinn‘, erwiderte Fisher, lächelnd, ‚Sicherlich macht es keinen Unterschied‘. Aber sie beharrte mit Nachdruck, dass es natürlich ein Unterschied sei. Von knapp hinter ihnen schlug eine Stimme vor, ‚Lassen Sie sie testen.‘ Es war William Roach, der nicht lange danach Miss Bristol heiratete. Sofort fingen sie mit den Vorüberlegungen zu einem Experiment an: Roach half mit den Tassen und frohlockte, dass Miss Bristol in hinreichend vielen Tassen das Richtige voraussagen würde.

Miss Bristols Triumph wurde nie dokumentiert, und vielleicht war Fisher mit diesem im Stegreif entstandenen Versuchsplan nicht zufrieden. Man kann jedoch sicher sein, dass er – während das Experiment ausgedacht und durchgeführt wurde – über die ausgelösten Fragen nachdachte: Wie viele Tassen sollten in dem Test verwendet werden? In welcher Reihenfolge sollten die Tassen präsentiert werden? Was sollte etwa bezüglich Zufallsschwankungen in Temperatur, Süße etc. getan werden? Welche Schlussfolgerungen könnten aus einer perfekten Anzahl richtiger Antworten gezogen werden? Welche Schlussfolgerungen, wenn ein paar Antworten falsch sind?“ (Joan Fisher Box (1978), R. A. Fisher – *The Life of a Scientist*, Wiley, S. 134)

Für weitere Erkenntnisse über die beteiligten Personen und die reichhaltigen statistischen Folgen dieses Anlass, siehe S. Senn (2012), Tea for three, *Signi-*

ficance 9 (6), 30–33, und D. V. Lindley (1993), An analysis of experimental data – the appreciation of tea and wine, *Teaching Statistics*, 15, 22–25.

Frage 2:

Auf dem Gebiet der Medizin, ist ein Placebo eine Pseudobehandlung ohne therapeutisch zu erwartende Wirkung auf den Patienten. Wenn Ärzte versuchen, die Wirksamkeit einer Behandlung für einen bestimmten Zustand (sei es ein Medikament oder eine andere Art der Therapie) statistisch zu testen, dann besteht eine Standard-Facette des experimentellen Designs darin, die Wirkung der „Behandlung" gegen eine Kontrollgruppe („keine Behandlung") zu prüfen. „Keine Behandlung" kann dabei wörtlich interpretiert werden als offene Unterlassung einer Behandlung oder die Behandlung kann durch die Verabreichung eines wirkungslosen Placebos simuliert werden. Daher besteht solch ein klinisches Experiment meistens in einem Vergleich zwischen zwei Gruppen von Patienten, bei denen die einen die Behandlung und die anderen ein Placebo verabreicht bekommen. Man hat beobachtet, dass selbst Patienten, die ein Placebo erhalten, sich oft besser fühlen und manchmal sogar eine tatsächliche physiologische Verbesserung ihres Zustandes zeigen – der so-genannte „Placebo-Effekt", dessen Mechanismus im Zusammenwirken von Geist und Körper noch nicht gut verstanden ist. Das Experiment wird daher „einfach-blind" durchgeführt, d. h. der Patient ist nicht darüber informiert, ob er oder sie die Behandlung oder das Placebo erhält. Es hat sich gezeigt, dass diese Vorgehensweise die Intensität des Placebo-Effekts – sofern er auftritt – verringert. Um zusätzlich auch eine klinische Verzerrung zu vermeiden, werden – sofern möglich – klinische Versuche „doppel-blind" durchgeführt, d. h. auch das unmittelbar mit der Intervention betraute klinische Personal weiß nicht, welcher Patient die eigentliche Behandlung und wer ein Placebo erhält.

Die Verabreichung eines Placebos statt einer klinischen Behandlung an einen Patienten, der sich seiner Art der Behandlung nicht bewusst ist, ist offensichtlich ein Akt der Täuschung durch den Arzt. Dies mag negative ethische Implikationen haben, die umso ernster sind je lebensbedrohlicher der Zustand des Patienten ist. In einem solchen Fall sollte ein Codex der Medizinethik ein „Caveat" (d. h. eine Warnung oder Vorsicht) verlangen.

Frage 3:

Wissenschaftliche Disziplinen, in denen experimentelle Studien überwiegen sind: Chemie, Physik, Psychologie, Pharmakologie und Landwirtschaft. Einige Disziplinen, in denen Beobachtungsstudien überwiegen: Kosmologie, Meteorologie, Klimatologie, Soziologie, Ornithologie. Einige Disziplinen, in denen sowohl Experimental- wie auch Beobachtungsstudien durchgeführt werden, sind: Medizin, Biologie, Geologie, Ökonomie, Bildungsforschung.

Was können wir daraus schließen? Alle Wissenschaften haben eine starke Präferenz für kontrollierte Experimente (d. h. die Wirkungen einer Intervention mit den Auswirkungen ohne Intervention zu vergleichen). Kontrolliertes Experimentieren hilft dabei (i) zu klären, welche Variable direkt die beobachteten Beziehungen beeinflussen und (ii) in der Erforschung der Richtung der Kausalität in solchen beobachteten Beziehungen. Beobachtungsstudien sind Experimentalstudien unterlegen bezüglich dieser Ziele. Gebiete, in denen Beobachtungsstudien überwiegen, sind solche, in denen Experimentalstudien entweder unmöglich oder sehr schwierig sind.

Frage 4:

Ist es für gültige statistische Schlüsse nicht wichtiger, eine repräsentative Stichprobe anstatt einer Zufallsstichprobe zu haben? Die kurze Antwort ist nein, weil die Herleitung der Schätzung eines Populationsparameters aus einer repräsentativen Stichprobe ein Grundprinzip der statistischen Inferenz verletzt. Eine Stichprobe, einmal ausgewählt, kann sich als repräsentativ herausstellen, aber zuerst und vor allem muss sie zufällig ausgewählt sein. Warum ist der Zufall so grundlegend? Wir behandeln diese allgemeine Frage in einem spezifischen Kontext.

Angenommen, die Managerin von einer Zahnarztpraxis mit 1500 erwachsenen Patienten will eine Erhebung unter ihren Patienten durchführen bezüglich ihrer Dentalpflege (Bürsten, Zahnseide etc.). Sie setzt eine Stichprobengröße von 200 erwachsenen Patienten fest und befragt die ersten 200 Patienten, die in die Praxis kommen. Sie erkennt, dass die Art der Dentalhygiene sehr wahrscheinlich vom Alter der Patienten beeinflusst ist und entscheidet daher die Patienten in vier Altersgruppen zu befragen: 16–24, 25–39, 40–59, 60 und älter. Sie geht davon aus, dass sie wohl keine zuverlässigen Ergebnisse erhält, wenn sie einfach die ersten 200 Erwachsenen, die ankommen, befragt. Warum? Weil es möglich ist, dass eine bestimmte Altersgruppe in ihrer Stichprobe sehr unterrepräsentiert ist. Daher beschließt sie eine repräsentative Stichprobe zu verwenden, d. h. der Stichprobenumfang in jeder Altersgruppe entspricht dem jeweiligen Anteil unter den 1500 Patienten. Aus den

Patientenakten findet sie die folgenden Prozentsätze für die jeweiligen Altersgruppen: 16 %, 37 %, 35 % und 12 %. Daher sucht sie Stichproben im Umfang von 32, 74, 70 und 24 Personen in den entsprechenden Altersgruppen aus.

Sie erfragt das Alter jedes Patienten bei der Ankunft und teilt dann den Fragebogen aus. Bei einem derartigen Vorgehen wird sie sehr wahrscheinlich mehr als 200 Patienten benötigen, weil die Patienten nicht unbedingt exakt in den benötigten Anzahlen eintreffen, die die Quoten für jede Altersgruppe erfüllen. Im weiteren Verlauf erstellt sie einen statistischen Bericht darüber, welcher Anteil der Patienten in jeder Teilstichprobe sich wie oft die Zähne putzt. Sie interpretiert diese Statistiken als zuverlässige Punktschätzungen der entsprechenden Proportionen für alle erwachsenen Patienten in der Praxis und berechnet Konfidenzintervalle um diese Punktschätzungen. Leider sind diese verschiedenen Schätzungen nicht sehr zuverlässig, da sie auf nicht-zufälligen Stichproben basieren.

Es mag paradox erscheinen, dass die Zufälligkeit der Stichprobe mit Blick auf die statistische Inferenz der Repräsentativität überlegen scheint. Es gibt zwei Gründe für die Priorisierung von Zufälligkeit – einer ist theoretisch, der andere praktisch. Der theoretische Grund dafür ist, dass all die optimalen Eigenschaften von Konfidenzintervallen (und anderen statistischen Rückschlüssen) auf der Basis von Wahrscheinlichkeitsverteilungen getroffen werden, die nur bei Zufallsstichproben angewandt werden können. Der praktische Grund ist, dass eine systematische Vorgehensweise bei der Stichprobenauswahl das Risiko eingeht, Daten hinzu zu nehmen oder wegzulassen mit dem Effekt verzerrter Schlussfolgerungen. Betrachten wir die Absicht der Managerin nur diejenigen zu befragen, die in die Praxis kommen. Dadurch mag sie viele Patienten ausschließen, deren Zahngesundheit entweder sehr gut (weil sie sich gewissenhaft die Zähne putzen) oder sehr schlecht ist (wegen langjähriger Vernachlässigung) ist. Bei einer einfachen Zufallsstichprobe in diesem Zusammenhang ist die richtige Vorgehensweise für die Managerin eine Tabelle von Zufallszahlen zu verwenden, um eine Zufallsstichprobe von 200 aus der Population von 1500 Patienten zu ziehen. Vielleicht hat sie sich entschieden, geschichteten Zufallsstichprobe zu ziehen, wobei die Schichten die vier Altersgruppen sind. Dann ist es ganz in Ordnung für ihre Stichprobengrößen 32, 74, 70 und 24 Personen auszuwählen, aber die Stichproben in diesen Schichten müssen per Zufall ausgewählt werden.

Frage 5:

Die betreffende Graph wurde von William Playfair (1759–1823), dem Begründer der modernen statistischen Graphik vorgeschlagen. In seinem Buch *Statistical Breviary*, erschienen im Jahre 1801, zeigte er, dass das System der Koordinatenachsen, das schon für die Darstellung mathematischer Funktionen wohl etabliert war, auch im Bereich der deskriptiven Statistik zur Darstellung realer Daten angewendet werden kann. Um seine Idee von einem Säulendiagramm zu erklären, schrieb er: „… wenn all das Geld, das ein einziger Mann im Handel erhalten hat, alles Guineen (das sind historische britische Geldmünzen) wären und er jeden Abend einen einzigen Haufen aus all den Guineen machen würde, die er an diesem Tag erhalten hat, dann würde die Höhe des Haufens proportional zu den Einnahmen dieses Tages sein“. Er verstand, wie viel graphische Darstellungen (damit meinte er seine eigenen erfundenen Liniendiagramme, Balkendiagramme, Säulen- und Kuchendiagramme) dazu beitragen können, die Ansammlungen von Statistiken mit Leben zu füllen und er bemerkte ironisch dass „… keine Studie weniger verführerisch oder trockener und langwieriger sei als Statistik, es sei denn, der Geist und die Phantasie werden eingeschaltet oder die jeweilige Person hat ein ganz spezielles Interesse am Thema, was selten der Fall ist bei jungen Männern von einem Rang im Leben“ (in E. Royston, 1956, *A note on the history of the graphical representation of data*, Biometrika, 43, 241–247). Für weitere Informationen zu Playfair und seine statistischen Beiträge, siehe P. Costigan-Eaves und M. Macdonald-Ross (1990), William Playfair (1759–1823), *Statistical Sciences*, 5, 318–326.

Anschrift der Verfasser

Peter Petocz and Eric Sowey
Macquarie University Sydney and
The University of New South Wales
Sydney
Australia
Peter.Petocz@mq.edu.au