

Prof. Antony Unwin, Alexander Pilhöfer
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
<http://stats.math.uni-augsburg.de/>

Statistik I

Übungsblatt 9

Abgabe: Dienstag 19. Juni 2012, bis spätestens 12.00 Uhr; Briefkasten: Statistik I oder per email an die Übungsleiter

Die Aufgaben können auch in 2er-Gruppen bearbeitet und abgegeben werden!

- (a) Wie erkennt man Unabhängigkeit in einem Mosaicplot? **(1P)**
- (b) Wie wäre ein Mosaicplot optimal aufgebaut, um den Einfluss von *Geschlecht* (m/w) und *Raucher* (Ja/Nein) auf eine Variable *Lungenkrebs* (Ja/Nein) zu untersuchen?**(1P)**
- (c) Nehmen Sie zu folgender Aussage Stellung: **(1P)**

“Der p -Wert eines zweiseitigen Tests ist immer doppelt so hoch, wie der des entsprechenden einseitigen Tests.”.

- (d) Interpretieren sie die folgende Aussage der Bundesagentur für Arbeit:**(1P)**

“95 Prozent aller Arbeitssuchenden bewerben sich auf die 35 Prozent aller freien Stellen, die öffentlich ausgeschrieben sind”.

2. Schlafmittel und Dioxin (5P)

- (a) An zehn Patienten wurden zwei Schlafmittel erprobt, wobei folgende zusätzliche Schlafzeiten in Stunden beobachtet wurden:

Patient	1	2	3	4	5	6	7	8	9	10
Medikament 1	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0
Medikament 2	1.9	0.8	1.1	0.1	-0.1	4.4	5.5	1.6	4.6	3.4

Führen Sie einen geeigneten t -Test sowie auch einen nicht-parametrischen Test durch, um zu überprüfen, ob die beiden Medikamente verschiedene Schlafwirkungen haben. In R ist der Datensatz im Paket MASS unter dem Namen `sleep` zu finden.

- (b) Der Datensatz `AgentOrange` (auf der Internetseite der Lehrveranstaltung) enthält Messungen von Dioxin (auch bekannt als Agent Orange) für 646 Vietnam-Veteranen und für 97 Veteranen, die nicht im Vietnam waren. Sind die Mittelwerte der beiden Gruppen signifikant voneinander verschieden? Was für einen Einfluss haben die Ausreißer? Kann man die Varianzen als gleich ansehen?

3. Schulreife (5P)

Bei einem Schulreifetest an 5005 ungefähr gleichaltrigen Kindern wurden die Ergebnisse nach Geschlecht getrennt aufgeführt. Es ergaben sich die folgenden Daten:

	Mädchen	Jungen
Schulreif	2350	2337
Nicht schulreif	121	197

- (a) Berechnen Sie per Hand einen χ^2 -Unabhängigkeitstest sowie auch einen Binomialtest und entscheiden Sie zum Test-Niveau 0.05, ob es Assoziationen zwischen Geschlecht und Reife gibt. Geben Sie bei beiden Tests alle formal wichtigen Größen an: Teststatistik, Nullhypothese, Alternativhypothese, p-Wert, Verteilung und Annahmebereich.
- (b) Wie unterscheiden sich die beiden Tests?
- (c) Kann man das Ergebnis auch schon direkt den Daten ansehen? Visualisieren Sie die Daten geeignet!

4. Sonographie (5P)

Im folgenden geht es um einen Datensatz zu 4323 Geburten im Klinikum Augsburg.

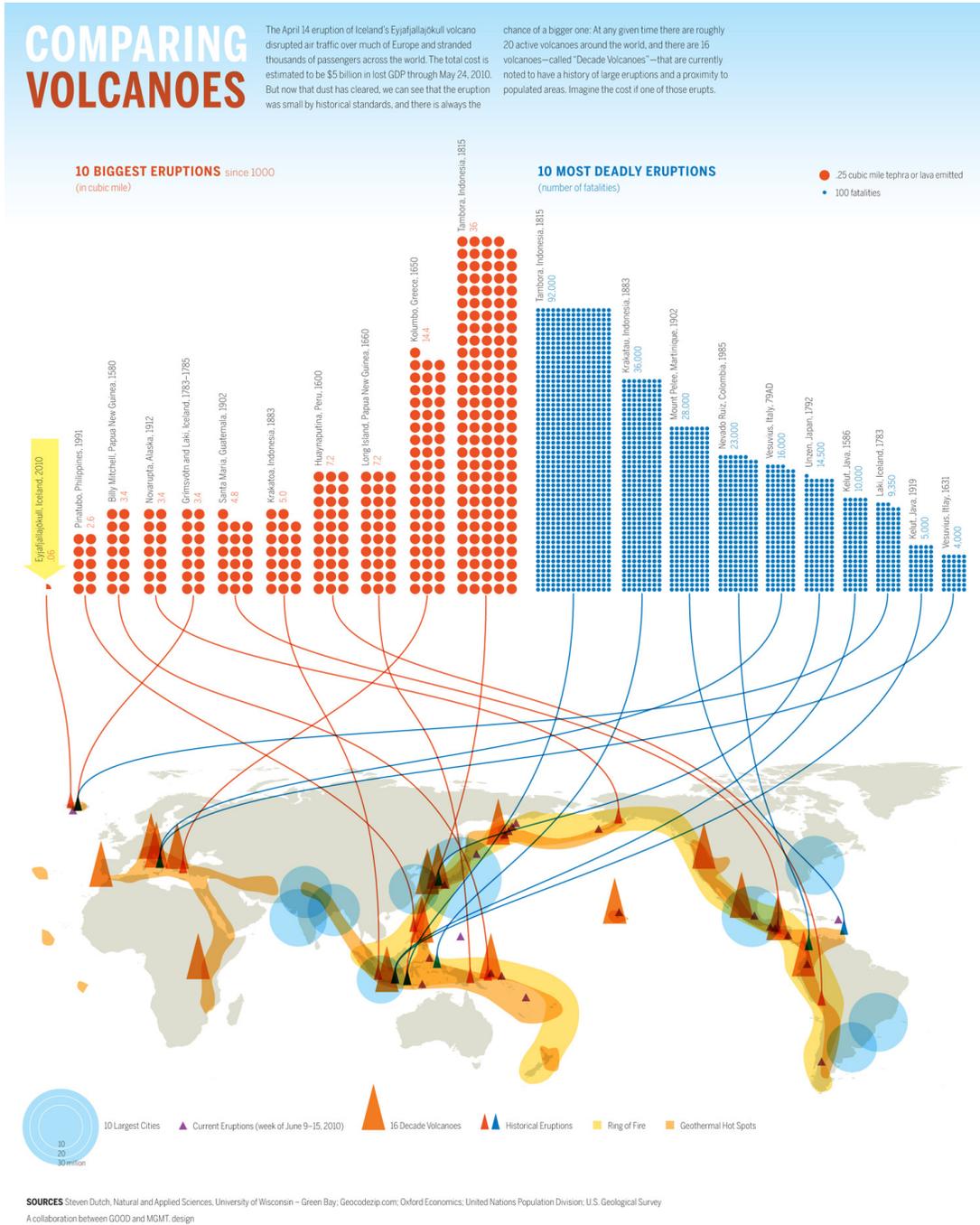
Die nachfolgende Tabelle zeigt die prozentualen Anteile der Art der Geburt V_G des Kindes in Kombination mit der Variable Frühgeburt V_p , ($1 = Ja$).

	$V_p = 0$	$V_p = 1$
Notfall-Kaiserschnitt	7.02%	1.82%
Geplanter Kaiserschnitt	9.99%	8.70%
Normale Geburt	58.14%	6.63%
Saugglocke	7.28%	0.41%

- (a) Testen Sie mit einem χ^2 -Unabhängigkeitstest, ob die Variablen voneinander unabhängig sind. Berechnen Sie die Teststatistik per Hand ohne die absoluten Häufigkeiten der Kombinationen zu berechnen! Verwenden Sie dann den p-Wert in R mit der korrekten Verteilung.
- (b) Wäre der Test immer noch signifikant, wenn man nur 100 Geburten verwendet hätte?
- (c) Berechnen Sie die Anzahl der Geburten, für die der Test erstmals einen p-Wert von 0.05 unterschreitet. Ist diese Größe eventuell nützlich?

5. Volcano (5P)

Betrachten Sie die nachfolgende Grafik zu Vulkanausbrüchen:



Analysieren und kritisieren Sie die Grafik. Gehen Sie strukturiert vor und versuchen Sie präzise Formulierungen zu finden. Als Anhaltspunkt nehmen Sie die Vorgehensweisen und Punkte der Grafikaufgaben der vorangegangenen Übungsblätter.