

Bevölkerung und Energie

Semesterarbeit WSWP, FFHS BSc INF 2011

Felix Rabe

8. Juli 2013

Inhaltsverzeichnis

1	Einleitung	2
2	Weltbevölkerung	3
2.1	Wachstum	3
2.1.1	Korrelationsrechnung	4
2.1.2	Regressionsrechnung	4
2.1.3	Längerer Abschnitt	5
2.2	Altersverteilung	6
2.3	Durchschnittswachstum der Weltbevölkerung von 2000 bis 2010	6
3	Energiekonsum	9
3.1	USA	9
4	Nachnamen in den USA	10
4.1	Einleitung	10
4.2	Histogramm über Top-10	10
4.3	Histogramm über den gesamten Datensatz	11
4.4	Wahrscheinlichkeiten	12
5	Schliessende Statistik	14
5.1	Rauchende Schweiz	14
5.2	Berechnungen Konfidenzintervall	14
6	Anhang	16
6.1	Quellen	16
6.1.1	Bevölkerungszahlen	16
6.1.2	Energiekonsum	16
6.1.3	Nachnamen der USA	16
6.2	Interna	16

Kapitel 1

Einleitung

Diese Arbeit betrachtet Bevölkerungszahlen und Energiekonsum mehrerer Länder im Zeitraum von 1965 bis 2010, inklusive.

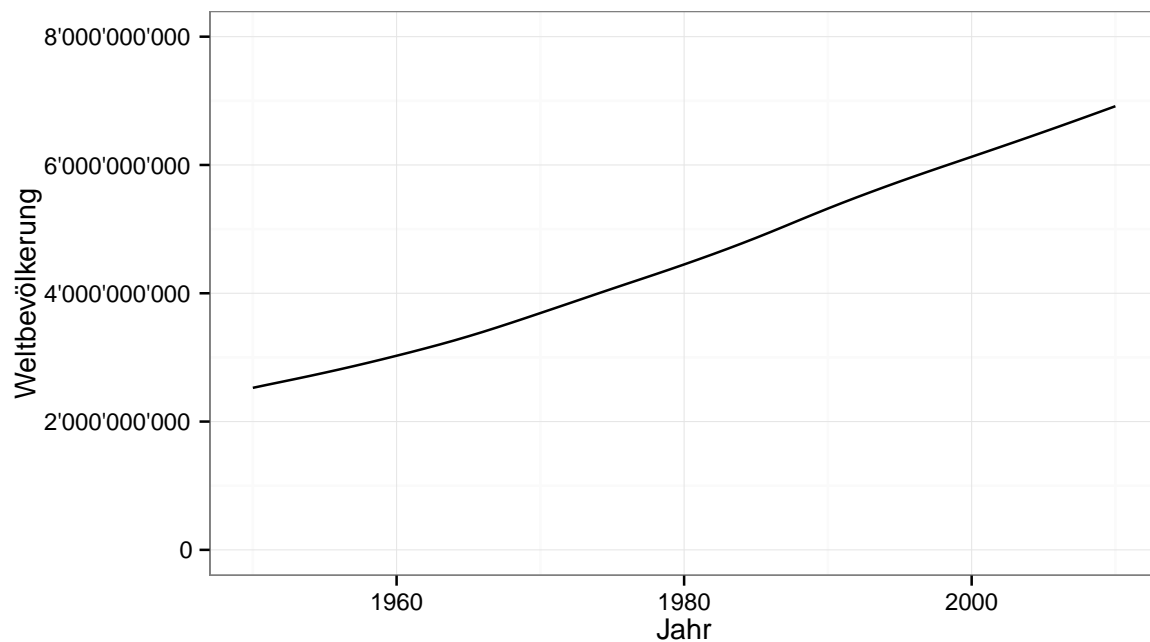
Dieses Thema ist interessant, weil unsere Ressourcen beschränkt sind, während unser Verbrauch weiterhin exponentiell zunimmt.

Kapitel 2

Weltbevölkerung

2.1 Wachstum

Das Wachstum der Weltbevölkerung lässt sich anhand der verwendeten Daten grafisch folgendermaßen darstellen:



Im betrachteten Zeitraum, besonders seit 1970, scheint sich ein ungefähr lineares Wachstum abzuzeichnen. Um diese Behauptung zu überprüfen, können die Daten in den Jahren 1970 (inkl.) bis 2010 (exkl., ergibt $n = 40$ Beobachtungen) mit einer Regressionsgeraden von Jahr nach Weltbevölkerung angenähert werden.¹

¹Die Bevölkerungsdaten gelten jeweils per 1. Juli eines gegebenen Jahres. Vereinfachend wird in dieser Arbeit davon ausgegangen, dass die Jahre gleich lang sind. Korrekterweise müssten Schaltjahre berücksichtigt werden, z.B. indem als x-Werte Tage (Abstand 365 bzw. 366) statt Jahre (Abstand 1) verwendet werden.

2.1.1 Korrelationsrechnung

Zunächst wird überprüft, wie stark Jahr und Bevölkerung miteinander in Zusammenhang stehen. Das verdeutlicht die Korrelationsrechnung.²

Die Formel für den Korrelationskoeffizienten r_{xy} lautet:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Die dazu benötigten Werte, arithmetisches Mittel \bar{x} , arithmetisches Mittel \bar{y} , Standardabweichung s_x , Standardabweichung s_y , Kovarianz s_{xy} , haben die folgenden Definitionen und Werte:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1989.5$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 5'258'342'429$$

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 11.6905$$

$$s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} = 958'171'156$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 11'198'337'193$$

Daraus ergibt sich:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = 0.9997$$

Das Resultat nahe bei +1 bedeutet, dass sich die Bevölkerungskurve praktisch auf einer Geraden mit positiver Steigung befindet.

2.1.2 Regressionsrechnung

Es wird im Folgenden die "Methode der kleinsten Quadrate" angewendet:

$$\bar{y} - a - b\bar{x} = 0$$

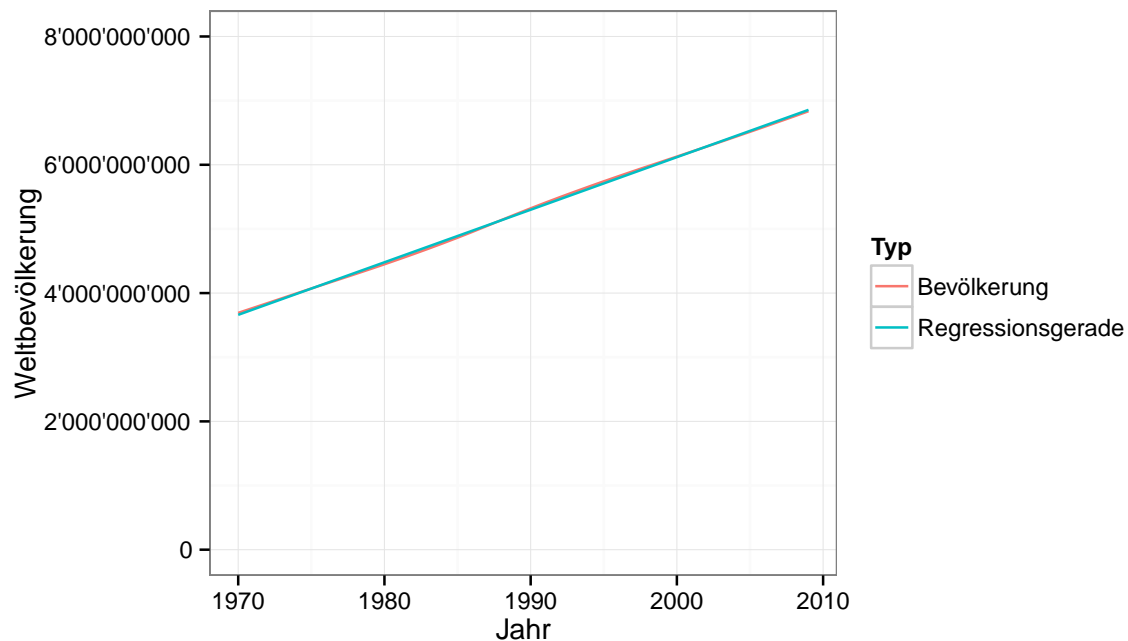
$$a\bar{x} + b \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{1}{n} \sum_{i=1}^n x_i y_i$$

²Die konkreten Werte, die für die folgenden Berechnungen verwendet werden, sind in Tabelle 2.1 auf Seite 8 aufgelistet. Für x wird das Jahr verwendet, für y die Bevölkerung.

Resultate des aufgelösten Gleichungssystems für a und b, unter Verwendung der oben ermittelten Werte:

$$b = \frac{s_{xy}}{s_x^2} = 81'939'053$$

$$a = \bar{y} - b\bar{x} = -157'759'402'784$$

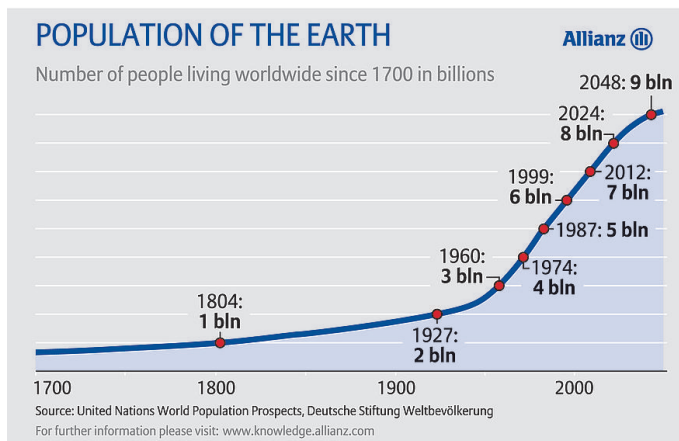


Interessanterweise ist der Wert a deutlich negativ. Das bedeutet, dass der y-Achsenabschnitt und damit die theoretische Weltbevölkerung im Jahr 0 negativ gewesen wäre, wenn sich das Wachstum der Bevölkerung tatsächlich linear entwickelt hätte. Das festgestellte lineare Wachstum ist jedoch eine rein lokale Beobachtung im angegebenen Zeitabschnitt. Im Folgenden noch eine Grafik zur historischen Entwicklung.

2.1.3 Längerer Abschnitt

Eine Suche auf Google zeigt die Bevölkerungsentwicklung in einem längeren Zeitraum, mit teilweise exponentiellem Wachstum:³

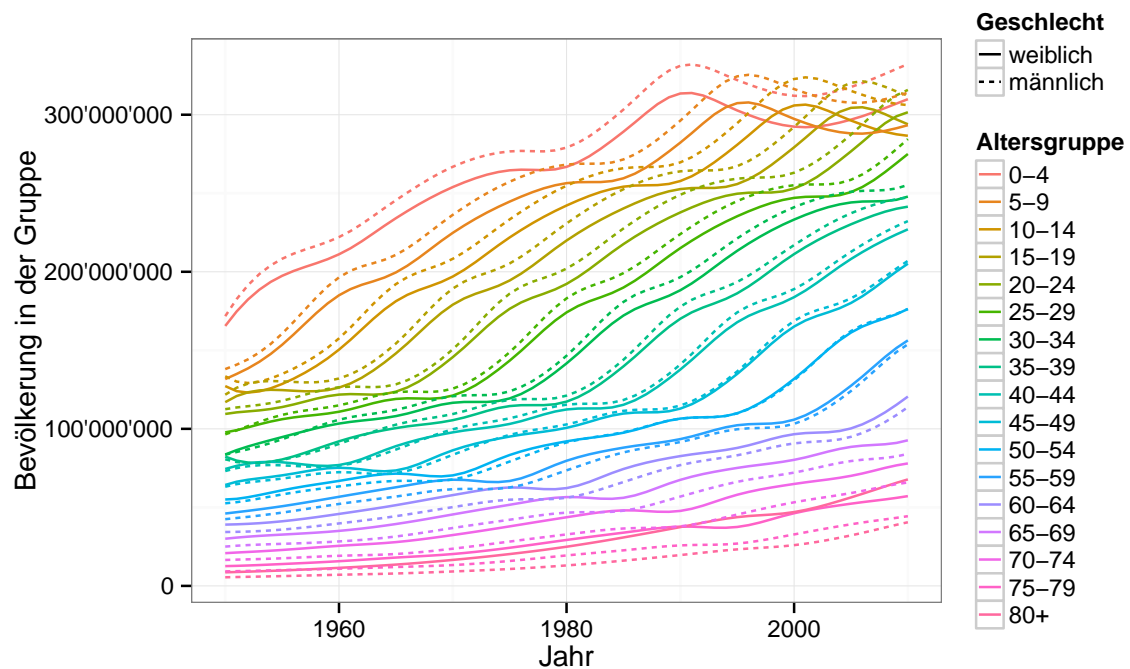
³Quelle: <http://ecache.ilbe.com/files/attach/new/20130615/377678/220411312/1412886170/ea09e52b28dc96677352e6e7d4aca053.jpg>, abgerufen 6. Juli 2013.



The publication of this graphic is free of charge provided that users credit Allianz SE.
Graphics are available in the media section of the Allianz Knowledge Site.
www.knowledge.allianz.com/en/media/graphics

2.2 Altersverteilung

Eine Übersicht über die gesamte Entwicklung der Weltbevölkerung, aufgegliedert nach Altersgruppe und Geschlecht, lässt sich wie folgt darstellen:



Man sieht daran sehr schön die “Baby Booms”, die längere Lebenserwartung von Frauen gegenüber Männern, und wie wir in absehbarer Zeit ein echtes Problem mit der Altersvorsorge haben werden.

2.3 Durchschnittswachstum der Weltbevölkerung von 2000 bis 2010

Weltbevölkerung in 2000: 6'127'700'428

Weltbevölkerung in 2010: 6'916'183'482

Durchschnittliches jährliches Wachstum: 1.218%

Kontrolle:

$$6'127'700'428 \cdot \left(\frac{1.2178}{100} + 1\right)^{10} = 6'916'183'482$$

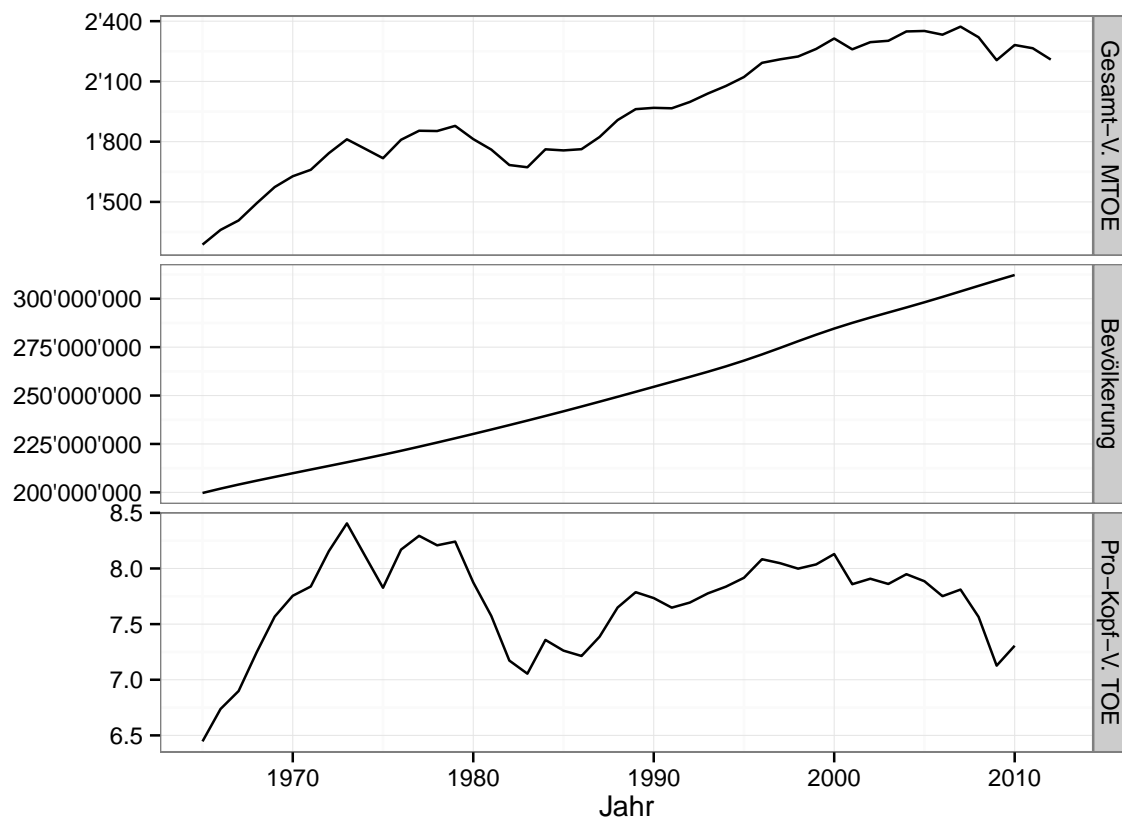
Jahr	Bevölkerung
1970	3'691'172'616
1971	3'766'754'337
1972	3'842'873'611
1973	3'919'182'343
1974	3'995'304'940
1975	4'071'020'434
1976	4'146'135'814
1977	4'220'816'714
1978	4'295'664'819
1979	4'371'527'911
1980	4'449'048'798
1981	4'528'234'631
1982	4'608'962'384
1983	4'691'559'863
1984	4'776'392'846
1985	4'863'601'517
1986	4'953'376'686
1987	5'045'315'901
1988	5'138'214'730
1989	5'230'452'474
1990	5'320'816'667
1991	5'408'908'713
1992	5'494'899'563
1993	5'578'865'122
1994	5'661'086'374
1995	5'741'822'412
1996	5'821'016'709
1997	5'898'688'295
1998	5'975'303'618
1999	6'051'478'026
2000	6'127'700'428
2001	6'204'147'007
2002	6'280'853'822
2003	6'357'991'755
2004	6'435'705'624
2005	6'514'094'605
2006	6'593'227'944
2007	6'673'105'894
2008	6'753'649'232
2009	6'834'721'990

Tabelle 2.1: Tabelle der Bevölkerungszahlen von 1970 bis 2009, die den Berechnungen zugrunde liegen.

Kapitel 3

Energiekonsum

3.1 USA



Die grafische Auswertung des Pro-Kopf-Verbrauchs der USA widerlegt die Behauptung teilweise, Amerikaner lebten auf grossem Fuss. Der Graph zeigt uns, wie seit den 70er-Jahren der Pro-Kopf-Energiekonsum in den USA gar rückläufig ist.

Kapitel 4

Nachnamen in den USA

4.1 Einleitung

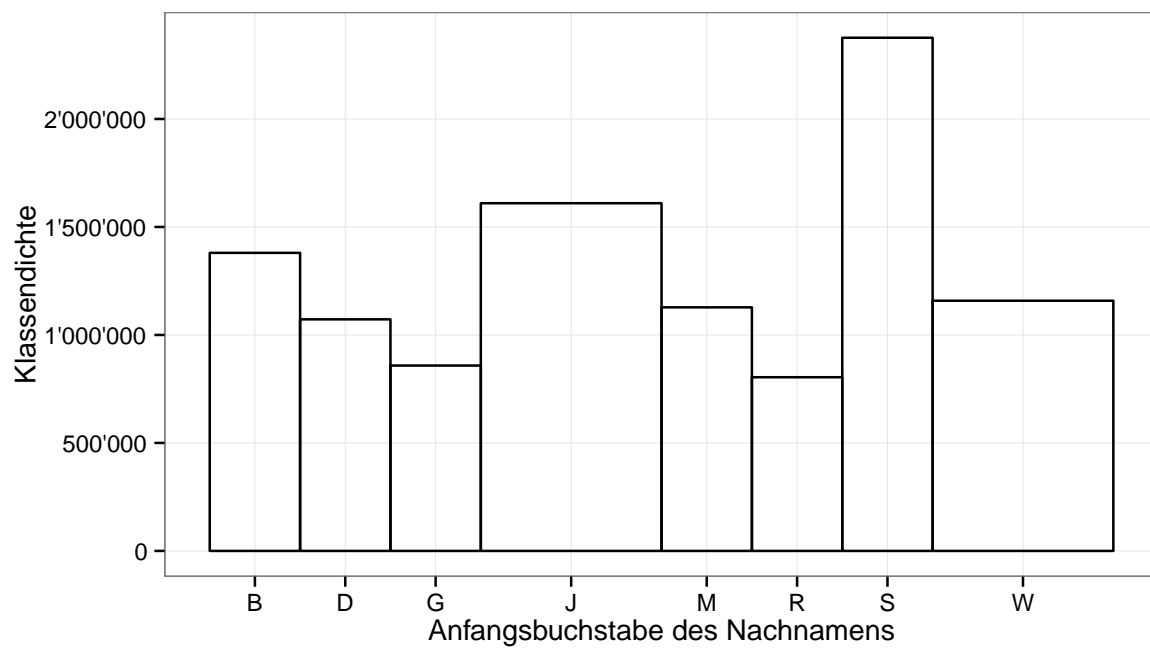
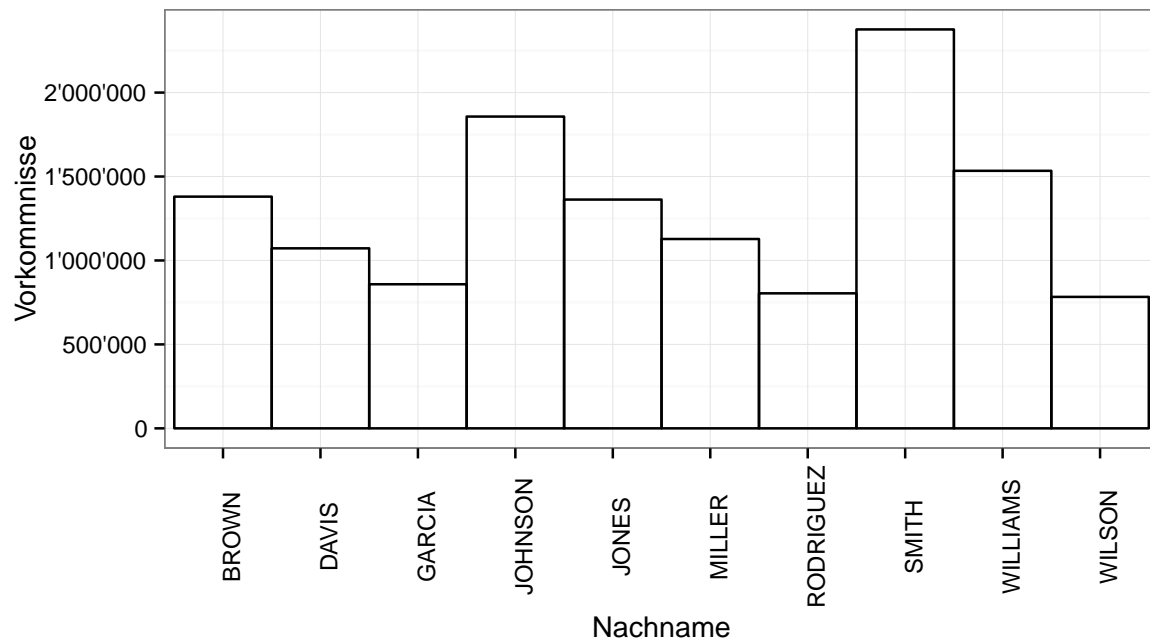
Die folgenden Auswertungen basieren auf den Nachnamen in den USA, die im Jahr 2000 mindestens 100 mal vorkamen.

Rang	Name	Anzahl
1	SMITH	2'376'206
2	JOHNSON	1'857'160
3	WILLIAMS	1'534'042
4	BROWN	1'380'145
5	JONES	1'362'755
6	MILLER	1'127'803
7	DAVIS	1'072'335
8	GARCIA	858'289
9	RODRIGUEZ	804'240
10	WILSON	783'051

Tabelle 4.1: Top 10 US-Namen im Jahr 2000.

4.2 Histogramm über Top-10

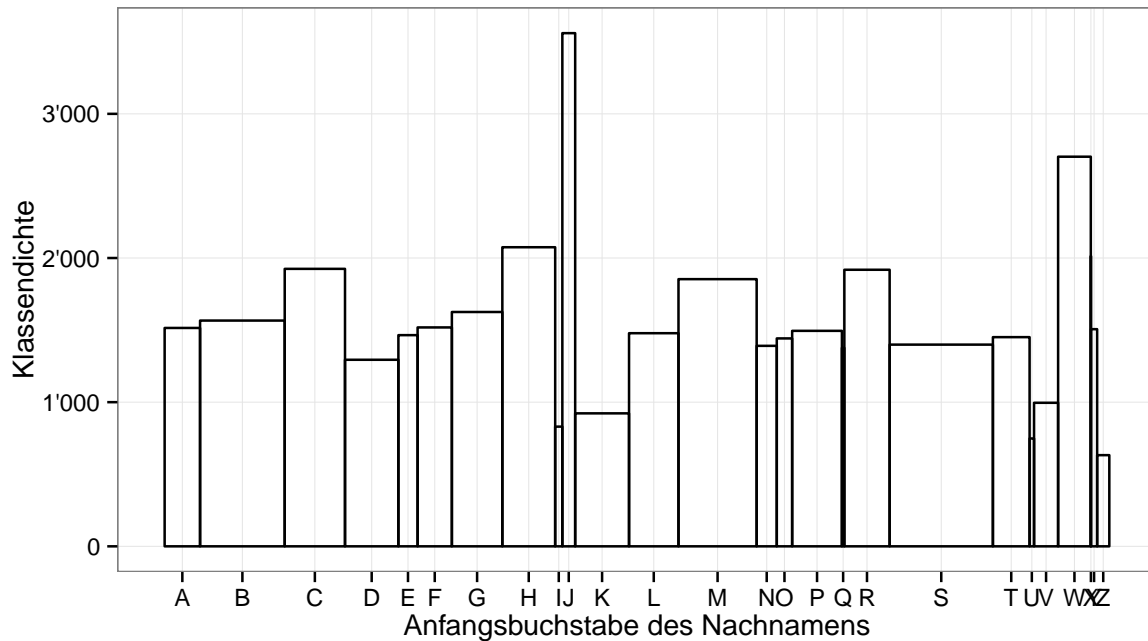
Die Nachnamen können nun nach ihren Vorkommnissen pro Anfangsbuchstabe untersucht werden. Zunächst werden die Top-10 gerechnet, um das Vorgehen zu erläutern und zu verifizieren, im nächsten Abschnitt werden dann alle Vorkommnisse gerechnet.



Soweit die Idee.

4.3 Histogramm über den gesamten Datensatz

Nun das selbe mit der gesamten Tabelle, welche insgesamt 151'671 Einträge umfasst:



Anhand der Histogrammflächen lässt sich nun ablesen, wie häufig ein bestimmter Buchstabe den Anfang eines amerikanischen Nachnamens bildet.

4.4 Wahrscheinlichkeiten

Mit den Bevölkerungszahlen und der US-Namensliste aus 2000 lassen sich nun folgende Berechnungen anstellen:

Ω := Ergebnisraum, Weltbevölkerung (jede Person ist ein Elementarereignis)

$$|\Omega| = 6'127'700'428$$

B := ist Amerikaner

$$|B| = 284'594'395$$

$$P(B) = \frac{|B|}{|\Omega|} = 0.0464$$

A := heisst "Smith"

$A \cap B$ = heisst "Smith" und ist Amerikaner

$$|A \cap B| = 2'376'206$$

$$P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = 3.8778 \times 10^{-4}$$

$P(A|B)$ = Wahrscheinlichkeit, dass ein Amerikaner "Smith" heisst

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0.0083$$

Da die Ereignisse A und B abhängig sind (A unterscheidet sich je nach Land), gilt nicht (zwingend): $P(A \cap B) = P(A) \cdot P(B)$. Somit sind weitere Resultate wie beispielsweise $P(A)$ und $P(B|A)$ nicht

rechenbar, da die verwendeten Daten keine Informationen zum Auftreten des Namens “Smith” ausserhalb der USA enthalten.

Kapitel 5

Schliessende Statistik

5.1 Rauchende Schweiz

Eine Google-Anfrage nach “anteil raucher in der schweiz” ergibt eine NZZ-Schlagzeile mit der Zahl 27%.¹

Angenommen, wir führen 1000 Befragungen durch, wobei 248 angeben, dass sie rauchen. Bei welchem Vertrauensniveau liegt die NZZ-Angabe von 27% gerade noch im Konfidenzintervall? Je kleiner das Vertrauensniveau, umso repräsentativer war die Umfrage.

5.2 Berechnungen Konfidenzintervall

Gesucht: $1 - \alpha$.

$$n = |\Omega| = 1000$$

$A :=$ Person ist Raucher

$$|A| = 270$$

$$X = 248$$

$$\hat{p} = \frac{X}{n} = \frac{248}{1000} = 0.248$$

$$p = \frac{|A|}{n} = \frac{270}{1000} = 0.27$$

$$k = p - \hat{p} = \frac{|A| - X}{n} = \frac{22}{1000} = 0.022$$

¹<http://www.nzz.ch/aktuell/startseite/27-prozent-der-schweizer-bevoelkerung-raucht-1.5968845>

$$\begin{aligned}
k &= 0.022 = c \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= \sqrt{\frac{0.1865}{1000}} = 0.0137 \\
c &= \frac{k}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{0.022}{0.0137} = 1.611 \\
\Phi(c) &= 1 - \frac{\alpha}{2} \\
\alpha &= 2 - 2\Phi(c) \\
1 - \alpha &= 2\Phi(c) - 1 = 2\Phi(1.611) - 1 = 2 \cdot 0.9464 - 1 = 0.8928
\end{aligned}$$

Die Befragung war also durchaus repräsentativ, der Wert \hat{p} lag innerhalb des 90%-Vertrauensintervalls von p .

Kapitel 6

Anhang

6.1 Quellen

6.1.1 Bevölkerungszahlen

Quelle: United Nations, Department of Economic and Social Affairs, Population Division (2013). World Population Prospects: The 2012 Revision, CD-ROM Edition.

<http://esa.un.org/wpp/Excel-Data/population.htm> ⇒ “Annual Population by Age Groups”.

Unter <http://esa.un.org/wpp/index.htm> ist einsehbar, auf welchen Quellen die Bevölkerungszahlen im Detail beruhen. Es wurden die aktuellen, im Juni/Juli 2013 veröffentlichten, Daten verwendet.

6.1.2 Energiekonsum

Quelle: <http://www.bp.com/statisticalreview>. Es wurden die aktuellen, im Juni/Juli 2013 veröffentlichten, Daten verwendet.

6.1.3 Nachnamen der USA

Quelle: “Genealogy Data: Frequently Occurring Surnames from Census 2000”.

<http://www.census.gov/genealogy/www/data/2000surnames> ⇒ “File B: Surnames Occurring 100 or more times”.

6.2 Interna

Dieses Dokument wurde mit \LaTeX , Javascript und R erarbeitet: R version 3.0.1 (2013-05-16), x86_64-apple-darwin10.8.0.

Installierte R Pakete: doBy, ggplot2, knitr, reshape2, xtable.

Der vollständige Quellcode mit allen Datensätzen der vorliegenden Arbeit ist unter <https://github.com/felixrabe/ffhs-wswp-projekt> einsehbar.