INVI Übung

Philip Burggasser, Paul Vonbank, Adelina Teofanescu

# Skewness - Schiefe

Rechtsschief **> 0**Linksschief **< 0**Symmetrisch **= 0**

# Kurtosis - Wölbung

**Exzess = 0** : Normalgipflig od. Mesokurtisch. Die Normalverteilung hat die Kurtosis β2 = 3 und entsprechend den Exzess 0.  
**Exzess > 0** : Steilgipflig, supergaußförmig od. Leptokurtisch. Es handelt sich hierbei um im Vergleich zur Normalverteilung spitzere Verteilung, d. h. Verteilung mit starken Peaks.  
**Exzess < 0** : Flachgipflig, subgaußförmig od. platykurtisch. Man spricht von einer im Vergleich zur Normalverteilung abgeflachten Verteilung

# Übung 1

***Informationen über Datensatz***

Der Datensatz mit 47 (französischsprachige Provinzen der Schweiz) Beobachtungen zu 6 Variablen, jeweils in Prozent, d.h. in[0, 100].

|  |  |  |  |
| --- | --- | --- | --- |
| [,1] | Fertility | Ig, ‘gemeinsame standardisierte Fruchtbarkeitsmessung’ | Metric ratio |
| [,2] | Agriculture | % von Männern, die in der Landwirtschaft tätig sind. | Metric ratio |
| [,3] | Examination | % Wehrpflichtige mit Bestnote bei der Heeresprüfung | Metric ratio |
| [,4] | Education | % Ausbildung über die Grundschule hinaus für Wehrpflichtige. | Metric ratio |
| [,5] | Catholic | % ‘katholisch’ (im Gegensatz zu ‘protestantisch'). | Metric ratio |
| [,6] | Infant.Mortality | Lebendgeburten, die weniger als 1 Jahr leben. | Metric ratio |

***Sinnvolle Schätzer***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Lokation | Variation | Schiefe | Gewicht |
| Fertility | Median, Harmonic mean | MAD from the median | Annähernd symmetrisch, fast ein bisschen linksschief | skewness = -0.4706269 kurtosis = 3.403232 |
| Agriculture | Arithmetic mean | Standard deviation | Leicht linksschief, Multimodal | skewness = -0.330867 kurtosis = 2.207406 |
| Education | Median | MAD from the median | Rechtsschief | skewness = 2.34281 kurtosis = 9.541434 |
| Catholic | Mean (wegen Gesamtbevölkerung) | Range | Bimodal | skewness = 0.4946274 kurtosis = 1.393236 |
| Infant | Mean (wegen Gesamtbevölkerung) | Standard deviation, Range | Annähernd symmetrisch, Leicht linksschief | skewness = -0.3422987 kurtosis = 3.943302 |

# Übung 2

***Informationen über Datensatz***

Datensatz State.x77 – Matrix mit 50 Zeilen (je Staat) und 8 Spalten:

|  |  |  |  |
| --- | --- | --- | --- |
| [,1] | Population | population estimate as of July 1, 1975 | metric ratio |
| [,2] | Income | per capita income (1974) | metric ratio |
| [,3] | Illiteracy | illiteracy (1970, percent of population) | metric ratio |
| [,4] | Life Exp | life expectancy in years (1969–71) | metric ratio |
| [,5] | Murder | murder and non-negligent manslaughter rate per 100,000 population (1976) | metric ratio |
| [,6] | HS Grad | percent high-school graduates (1970) | metric ratio |
| [,7] | Frost | mean number of days with minimum temperature below freezing (1931–1960) in capital or large city | metric ratio |

***Sinnvolle Schätzer***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Lokation | Variation | Schiefe | Gewicht |
| Population | arithmetic mean  schwere Ränder | Variance (bc of outliners) | bimodal, rechtsschief, stark steilgipflig | skewness = 1.981395 kurtosis = 7.030555 |
| Income | Median – middle value of the ordered data, nv  schwere Ränder | Standard deviation | unimodal, eher symmetrisch, leicht rechtsschief, steilgipflig | skewness = 0.2109882 kurtosis = 3.378353 |
| Illiteracy | Mode – value with the greatest frequency,  schwere Ränder | Standard deviation | bimodal, rechtsschief, steilgipflig | skewness = 0.8437669 kurtosis = 2.632838 |
| Life Exp | median, nv  leichte Ränder | MAD from median (bc robust) | unimodal, leicht linksschief, steilgipflig | skewness = -0.1582224 kurtosis = 2.427038 |
| Murder | mean  schwere Ränder | Standard deviation | bimodal, leicht rechtsschief, leicht steilgipflig | skewness = 0.1333186 kurtosis = 1.862559 |
| HS Grad | mean (non robust)  leichte Ränder | Standard deviation | bimodal, leicht linksschief, leicht steilgipflig | skewness = -0.3290666 kurtosis = 2.209573 |
| Frost | median  schwere Ränder | Standard deviation | unimodal, linksschief, steilgipflig | skewness = -0.3776493 kurtosis = 2.140944 |

Regression:

Aufgrund der starken Korrelation von HS Grad, Life Exp und Frost mit den anderen erklärenden Variablen, wurden diese aus dem Modell genommen (Population, Income, Illiteracy sind drinnen).

Residual vs. Fitted: Mittel der Residuen liegt um 0 -> lineare Zusammenhang möglich

Normal Q-Q: Residuen sind normal verteilt, schwere Ränder auf beide Seiten

Scale-Location: Residuen sind ungefähr homoskedastisch

Residuals vs. Leverage: Alaska ist ein Hebelpunkt -> nicht lineares Modell

# Übung 3 – LakeHuron

|  |  |
| --- | --- |
| Lokation | Mean, vielleicht harmonic Mean falls Extremwerte nicht beachtet werden sollen. |
| Variation | StDev oder MAD, Beides liefert ähnliche Resultate, MAD berücksichtigt Extremwerte weniger.  Wenn die Daten für Hochwasserschutz o.ä. verwendet werden ist Range nützlicher. |
| Schiefe | skewness = -0.1397719 -> links schief |
| Gewicht | kurtosis = 2.499163 -> mesokurtic, leichte Ränder |

Voraussetzungen für lineares Modell:

1. Fehler[E(i)] = 0
2. E(i) iidindepent, identically, distributed) => Klare Kurve erkennbar
3. E(i) homostedastisch => Varianz wächst
4. E(i) annähernd normalverteilt => nicht sehr sehr grob Normalverteilt, eher nicht

* Voraussetzungen für lineares Modell nicht gegeben

# Übung 4