

Skalierungen

Skalierung	Beschreibung	Beispiel
Nominal	Ausprägung der Merkmale haben keine Reihenfolge	Zivilstand, Geschlecht, Autofarbe
Ordinalskala	Die Merkmalausprägung hat eine Reihenfolge	Preis, Leistung
Intervallskala	Es existiert kein echter Nullpunkt. Wird in Masseinheit gemessen	Zeit, Temperatur
Verhältnisskala	Intervallskala mit absoluten Nullpunkt	Gewinn, Kosten, Anzahl Mitarbeiter
Metrisch	Hat Intervall- oder verhältnisskalierte Ausprägungen. Nomal- und ordinalskalierte Merkmale sind nichtmetrische Merkmale.	
Diskret	Metrisch skalierte erkmal kann nur bestimmte Werte annehmen.	1..n Kinder, nicht 1,5
Stetig	Jeder Wert ist möglich	BIP in CHF
qualtiativ	Nominal	
quantitativ	Ordinal, Intervall, Verhältnis	

Mittelwerte

Mittelwerte

Durchschnitte

Median  
(mittlerer Wert)

Modus  
(häufigster Wert resp. grösste Dichte)

arithmetisches Mittel  
(additiv)

geometrisches Mittel  
(multiplikativ)

Histogramm

$$\text{Dichte} = \frac{\text{relative Häufigkeit}}{\text{Klassenbreite}}$$

Klassenbreite wird definiert.

$$\text{relative Häufigkeit} = \frac{\text{absolute Häufigkeit}}{\text{gesamte Anzahl Werte}}$$

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

Geometrisches Mittel (Aufeinanderfolgende Änderungen)

$$\bar{R} = \sqrt[n]{\prod_{t=1}^n (1 + R_t)} - 1$$

Mit Klassen

$$\bar{x} = \frac{\sum_{i=1}^n (\text{Klassenmitte} * \text{abs. Häufigkeit})}{\sum_{i=1}^n (\text{abs. Häufigkeit})}$$

Gleitender Durchschnitt

Median

Der Durchschnitt der benachbarten Werte aus einer Reihe.  
3. Ordnung: (vorangehender Wert + aktueller Wert + folgender Wert) / 3

$$\text{Untergrenze der Medianklasse} + \frac{j}{f} * \text{Medianklassenbreite}$$

$j = \text{Anzahl Werte in der Medianklasse bis zur Mitte}$   
 $f = \text{Anzahl Werte in der Medianklasse}$

Quantile

Modus

rechtsschief

linksschief

symmetrische Verteilung

Sind Perzentile, Dezile und Quartile. Median ist das 0.5-Quantil.

Merkmalsausprägung einer Messreihe die am häufigsten vorkommt, bzw. die höchste Dichte hat.

Modus < Median < arithmetisches Mittel

arithmetisches Mittel < Median < Modus

aithmetisches Mittel = Median = Modus

Verteilung mit Klassen

	L1	L2	L3			
Altersklasse	Klassenmitte	Anzahl Personen	rel. Häufigkeit	Klassenbreite	kum. rel. Häufigkeit	Dichte
18 < 20	19	3	9.38%	2	9.38%	4.69%
20 < 25	22.5	15	46.88%	5	56.25%	9.38%
25 < 30	27.5	5	15.63%	5	71.88%	3.13%
30 < 40	35	5	15.63%	10	87.50%	1.56%
40 < 65	52.5	4	12.50%	25	1	0.50%
		32				

rel. Häufigkeit=Anzahl Pesonen/32  
Dichte= rel. Häufigkeit/Klassenbreite  
Berechnung durchschnittliches Alter: 1-Var Stat L1, L2  
x=28,66

L1	L2
Alterobergrenze	kum. rel. Häufigkeit
20	9.38%
25	56.25%

Berechnung mittlere Alter: LinReg L2, L1  
f(50)=24.33

Streuung

Beschreibt die Streuung von gemessenen Daten.		
Spannweite	$\text{Spannweite} = \text{Maximum} - \text{Minimum}$	
Quartilsabstand	$\text{Quartilsabstand} = 3. \text{Quartil} - 1. \text{Quartil}$ Quartil ist das mediale Merkmal an 1/4 Stelle einer Datenreihe	
Varianz und Standardabweichung	Beschreibt den Durchschnitt von Abständen einer Messreihe vom Mittelwert dieser Messreihe.	Empirische Varianz $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$ Empirische Standardabweichung $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$ Empirisch -> Stichproben
	Varianz $\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$ Standardabweichung: $\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$	
Variationskoeffizient	$V_{\sigma} \approx \frac{s}{\bar{x}}$	
Kursrendite	Logarithmierte Rendite ist die stetige Rendite Monatsrenditen annualisieren: $\bar{x}_{ann} = n * \bar{x}$ und $s_{ann} = \sqrt{n} * s$ wobei $n = 12$	

Indexzahlen

Jahr	1999	2000	2001	2002	2003	2004	2004
Arbeitslose	98602	71987	67197	100504	145687	153091	148537
Index 2000=100	137.0	100	93.3	139.6	202.4	212.7	206.3

Berechnung	Formel	Beispiel
Prozentuale Veränderung	$p = \frac{\sqrt[n]{\text{Wert}_n}}{\text{Wert}_0} - 1$ $n = \text{Anzahl Jahre}$	Senkung Arbeitslosenzahlen von 1999-2001 $p = \sqrt[2]{\frac{67197}{98602}} - 1 = -17.4\%$

Der zusammengesetzte Index  
 $I_t = \frac{\sum_{i=1}^n (\text{aktueller Preis}_i * \text{Basismenge})}{\sum_{i=1}^n (\text{Basispreis}_i * \text{Basismenge}_i)} * 100$  n Güter i = 1, ..., n; aktuelle Periode = t  
Basispreis ist der Preis t0

Janik von Rotz (<http://janikvonrotz.ch>)

1

Janik von Rotz (<http://janikvonrotz.ch>)

2

Bivariate Statistik

Kovarianz	Verhältniss aus der durchschnittlichen Standardabweichung
$\text{cov}(x,y) = \sigma_{xy} = \frac{i}{n} \cdot \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})$	
Korrelationskoeffizient	Kovarianz ohne Merkmalsprägung.
Bestimmtheitsmass	R^2
Variant	totale Varianz = erklärbare Varianz + nicht erklärbare Varianz

r (linearer Zusammenhang)	Stärke des funktion. Zusammenhangs	R^2
0		0%
...	kein nachweisbarer funktionaler Zusammenhang (sehr schwach)	
+0.25		5%
...	schwach	
+0.5		25%
...	mittel	
+0.75		50%
...	stark	
+1		100%

Regression

Modell	Formel	TS
Lineares Modell	$y = ax + b$	LinReg ax+b
Polynom-Modell	$y = \sum_{i=0}^n a_i x^i$	Excel
Quadratisches Modell	$y = ax^2 + bx + c$	Quadratic Reg
Kubisches Modell	$y = ax^3 + bx^2 + cx + d$	Cubic Reg
Logarithmisches Modell	$y = a + b \cdot \ln(x)$	LnReg a+blnx
Potenz-Modell	$y = a \cdot x^b$	PwrReg ax^b
Exponentielles Modell	$y = a \cdot b^x$	ExpReg ab^x
Exponentielles Modell	$y = a \cdot e^{c \cdot x} = a(e^c)^x$	Excel

Zeitreihen

Der zusammengestzte Index	$I_t = \frac{\sum_{i=1}^n \text{Preisindex}_i \cdot \text{Wert}_i}{\sum_{n=1}^n \text{Wert}_i}$
Komponenten der Zeitreihen	Y: Zeitreihe F: Konjunkturkomponente S: Saisonkomponente E: Restkomponente
Additive Verbundheit	$Y_t = F_t + S_t + E_t$
Multiplikative Verbundheit	$Y_t = F_t \cdot S_t \cdot E_t$

Saisonbereinigung

Berechnung		3. Ordnung		LinReg	L2-	Vergleich
				L1,L2	Saisonkonstanten	Vortrimester
Liste	L1	L2		L3=f(L1)	L1=L2-L3	
Jahr	Trimester	Nr.	Anzahl Fahrgäste	gleitender Durchschnitt	linearer Trend	Abweichung
2012	Haupt	1	639	...	484	155
2012	Zwischen	2	396	433	499	-103
2012	Neben	3	265	508	515	-250
2013	Haupt	4	864	534	530	334
2013	Zwischen	5	472	542	545	-73
2013	Neben	6	289	514	560	-271
2014	Haupt	7	782	...	575	207
						407
						19%
						9%
						20%
						-11%
						-2%
						0%

Saisonkonstante		
H	232	(155+334+207)/3
Z	-88	...
N	-260.5	...

Damit Trendwerte Berechnen: Anzahlfahrfgäste=Saisonkomponente\*linearer Trend

Im multiplikativen Modell gilt:

Abweichung: L1=L2/L3

Saisonkonstanten: H:  $\sqrt[3]{155 \cdot 344 \cdot 207} = 1.431$

Saisonbereinigt: Anzahl Fahrgäste/Saisonkonstante

Taschenrechner

<Actions>; Keys and values

Variationskoeffizient berechnen

Data + <Eingabe Daten> + 2nd + Data + 1-Var Stats

<Select> Sx + Enter

2nd + Data + 1-Var Stats

<Select> X + Enter

<Calculate> Sx/x

Berechnen Median

Data + <Eingabe Daten relative Häufigkeit; letzter Wert unter 50% erster Wert über 50%>

2nd + Data + 4: LingReg L1, L2, One, YES

Berechnen des Durchschnitts Einkommen

Data + <Eingabe Daten> + 2nd + Data + 2-Var + Freq: L2 + Enter

Berechnen der relativen Häufigkeit in ganze Promille

Data + <L3 wählen> + sto-> + data + 2 + <Formel einfügen, bsp L2/Durchschnitt> + Enter

Löschen Data

Data + Data + <Auswahl treffen>

interpolieren mit Zweisatz / Dreisatz

Data + <Eingabe Daten>

2nd + Data + 4: LinReg + <X und Y Achsen anpassen>

Anstatt Achsen anpassen

2nd + Data + StatVars + x'( + <Wert eingeben>

Standardabweichung

<Daten eingeben> + 1-Var stats l1,l2

Listen kopieren

L1 auswählen + sto -> + Data + L1 auswählen

Prozentuale Veränderung

L2 = L1, dann Formeln löschen (data, Pfeiltaste rechts, 5)  
den ersten Eintrag in L1 sowie den letzten Eintrag in L2 löschen  
L3 = L1 / L2 - 1, dann Formeln löschen (data, Pfeiltaste rechts, 5)