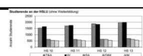
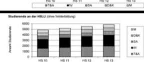
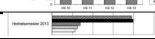
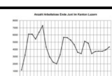
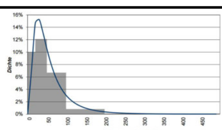
Univariate Statistik	
Grundlagen A1	Nominal-S. tiefste Hierarchiestufe, Verhältnis-S. höchste. Es wird auf basis tiefstem Messniveau analysiert.
Nominalskala Merkmal Zivilstand — Ausprägungen a: ledig b: verheiratet c: verwitwet d: geschieden	(nicht-metrisch, qualitativ) keine Reihenfolge od. Abstand, somit kein Durchschnitt berechenbar. Codierung wird benutzt zum Abkürzen 1 / w = weiblich
Ordinalskala Preis — a: sehr teuer b: teuer c: günstig d: sehr günstig	(nicht-metrisch, qualitativ) hat Reihenfolge, unbekannter Abstand, benutzt auch Codierung
Intervallskala in °C gemessen, Nullpunkt = Gefrierpunkt des Wassers in °F gemessen, Nullpunkt = tiefste Temperatur des "kältesten Winters 1709" 60°C warmes Wasser ist nicht doppelt so warm wie Wasser mit 30°C, denn: 60°C = 140°F, 30°C = 86°F. z.B. in Jahren gemessen. 1999 ist mehr Zeit verflossen als 1990. Zwischen 1960 und 1999 ist gleich viel Zeit vergangen wie zwischen 1860 und 1899, nämlich 1999-1960 = 39 Jahre. Die Festlegung des Jahres 0 ist aber willkürlich. Deshalb ist die Aussage, dass im Jahre 2000 doppelt so viel Zeit vergangen sei als im Jahre 1000, sinnlos.	(metrisch, quantitativ) Ausprägungen in Masseiheit gemssen, kein Nullpunkt (nur relativ), Abstände = Intervalle werden durch Zuordnungszahlen richtig wiedergegeben und somit können sie addiert und subtrahiert werden, aber kein Verhältnis 60° doppelt so heiss wie 30° geht nicht weil relativ
Verhältnisskala Ein Gewinn von 2 Mio. Fr. ist doppelt so gross wie ein Gewinn von 1 Mio. Fr.	(metrisch, quantitativ) Wie Intervallskala nur zusätzlich einen absoluten Nullpunkt somit kann ich Verhältnis angeben (doppelt so gross). Nur bei Verhältnisskalen kann ich %-Veränderungen angeben.
nicht metrisch, qualitativ	kann nicht mit Masseinheit gemessen werden
metrisch, quantitativ	kann diskret oder stetig sein, kann mit Masseinheit gemessen werden
diskretes Merkmal	kann auf Skala nur bestimmten Wert annehmen, z.B. Familie hat 1,2,3 Kinder nicht 1.5 oder ich bin x Jahre alt
stetiges Merkmal	kann auf Skala jeden Wert annehmen, z.B. ich bin x Jahre, y Monate und z Tage,… alt
Häufigkeitstabelle für nominale, ordinale oder metrisch diskrete Merkmale	Häufigkeitstabelle mit Merkmalsausprägungen, den **absoluten Häufigkeiten** (W) und/oder den **relativen Häufikeiten** (w)
Häufigkeitstabelle für metrisch stetige Merkmale	Ausprägungen werden als **Klassen** angegeben "von..bis unter", geschieht auch bei diskreten Merkmalen mit vielen Ausprägungen dann werden Klassen "von…bis" gemacht
relative Häufigkeit	$\text{relative Häufigkeit} = \frac{\text{absolute Häufigkeit}}{\text{gesamte Anzahl Werte}}$ oder via Dichte: $\text{Dichte} * \text{Klassenbreite} = \text{relative Häufigkeit}$
Gliederungstabellen	Um ein oder mehrere Merkmale zu vergleichen, zwei- oder multidimensional. Exakte Beschriftung für eindeutigkeit, Masseinheit angeben, Gesamtzahl muss ersichtlich sein (relativ!), Quellen angeben
amtliche Statistik Symbole - anstelle einer Zahl bedeutet null (nichts). 0 Der Zahlenwert beträgt zwar mehr als "nichts", doch ist sein Wert kleiner als die Hälfte der verwendeten Zähleinheit. (Ist die Zähleinheit z.B. 1 Mio., so bedeutet 0 weniger als 500'000 und 0.0 weniger als 50'000.) . anstelle einer Zahl bedeutet, dass die Zahlenangabe nicht möglich ist, weil die begrifflichen Voraussetzungen fehlen. ... Daten sind nicht erhältlich oder sind ohne Bedeutung oder sind aus anderen Gründen weggelassen.	Grafische Darstellungen können Tabellen ergänzen jedoch nicht ersetzen.
Säulendiagramm	für qualitative und quantitative Daten, Achtung Häufigkeitsachse (senkrecht) sollte bei 0 beginnen, 3D auch nicht so gut
Gruppensäulendiagramm 	Einzelne Komponenten sind vergleichbar, bei vielen Komponenten unübersichtlich, Total nur mit Totalsäule ersichtlich
Komponentensäulendiagramm 	Anteile der Komponenten ersichtlich, kein Total notwendig, Veränderung ist nur bei der untersten Komponente klar die anderen sind nicht auf einer Linie somit unleserlich wie verändert.
Balkendiagramm 	ist ein waagrechtes Säulendiagramm, ist super für lange Merkmalsbeschriftungen
Kreisdiagramm	Gut für Gliederungszahlen, wie ist Total gegliedert, "Kuchen"
Liniendiagramm 	für Zeitreihen, wie ist die Entwicklung während Zeit, einzelne Punkte gemessen, die werden dann verbunden so sieht es wie konstante Messung aus. Achtung Massstab von Häufigkeit richtig wählen sonst verzieht es die Zahlen und sieht dramatisch aus.
Lineares Liniendiagramm	Benutzt wie Liniendiagramm für Entwicklung während Zeit
Logarithmiertes Liniendiagramm	Benutzt wenn uns relative resp. Prozentuale Veränderung interessiert
Kurvendiagramm	z.B. Thermograph nimmt dauernd auf
Histogramm 	Um stetige Merkmale graphisch darzustellen machen wir Klassen, Klassen sind nicht alle gleich breit darum sinnvoll Klassenhäufigkeit darstellen = Dichte. Dichte steht in sekrechter Achse, Klassen in waagrechter; rel. Häufikeit ist Rechteck eben (Dichte * Klassenbreite), gesamte Fläche ist 100%. Histogramm ist annäherung an Kurve die ich bekäme wenn ich einzelwerte hätte.
Dichte	$\text{Dichte} = \frac{\text{relative Häufigkeit}}{\text{Klassenbreite}}$
Summen- und Produktzeichen	
$\sum$ Summenzeichen Sigma $\sum_{i=1}^{5} 2x_i = 2x_1 + 2x_2 + 2x_3 + 2x_4 + 2x_5 = 2(x_1 + x_2 + x_3 + x_4 + x_5) = 2\sum_{i=1}^{5} x_i$	Lässt Summen kurz und übersichtlich darstellen, ich addiere jede Runde für so viele "Runden" wie n angibt $\sum_{\text{bei welchem Wert beginne ich}}^{\text{wie viel mal,ab i}} Wert_{welcher}$ $\sum_{i=i}^{n} x_i$
$\prod$ Produktzeichen $\prod_{i=1}^{n}(1+R_i) = (1+R_1)\cdot(1+R_2)\cdot\ldots\cdot(1+R_n)$	Genau gleich wie beim Summenzeichen nur das hier jede Runde multipliziert wird und nicht addiert

Univariate Statistik

Mittelwerte A2

k = Anzahl versch. Merkmalswerte od. Anzahl Klassen	n = Anzahl Messwerte	w = relative Häufigkeit (als Faktor 0.8 sind 80%) W = absolute Häufigkeit
Klassenmitte	$\frac{von - Klass}{bis - Klasse} = Klassenmitte$	Einkommensklassen in CHF von 20'000 bis 50'000, Klassenmittel = 35'000, ich rechne mit Klassenmitte bei stetigem Merkmal zu Klassen zusammengefasst wurde.
relative Häufigkeit	$\text{relative Häufigkeit} = \frac{\text{absolute Häufigkeit}}{\text{gesamte Anzahl Werte}}$	oder via Dichte: Dichte * Klassenbreite = relative Häufigkeit
Dichte	$\text{Dichte} = \frac{\text{relative Häufigkeit}}{\text{Klassenbreite}}$	
Arithmetisches Mittel $\bar{x}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ mit $\sum_{i=1}^{k} W_i = n,\ w_i = \frac{W_i}{n};\ \sum_{i=1}^{k} w_i = 1$ $\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot W_i}{\sum_{i=1}^{k} W_i} = \sum_{i=1}^{k} x_i \cdot w_i$	Summe aller Messwerte (beobachteter Merkmalsausprägungen) x_1, x_2, x_3 geteilt durch deren Anzahl Messwerte n, nur für metrisch skalierte Merkmale sinnvoll. Achtung hier wird jeder Merkmalswert gleich berücksichtigt, dh. Extreme beeinflussen den Mittelwert stark. Entspricht Schwerpunkt Häufigkeitsverteilung.

Einkommensklasse (in CHF)			Klassenmitte (in CHF)	abs. Häufigkeit W_i	rel. Häufigkeit w_i
0	bis unter	20'000	10000	7514	7514/37573 = 20.0%
20'000	bis unter	50'000	35000	13645	36.3%
50'000	bis unter	100'000	75000	12540	33.4%
100'000	bis unter	200'000	150000	3063	8.2%
200'000	und mehr		500000 *	811	2.2%
Total				37573	100.0%

**Klassenmitte von halboffener Klasse kann nur berrechnet werden wenn Durchschnitt bekannt. Sonst muss Annahme getroffen werden wie in Tabelle unten.*

Das durchschnittliche Einkommen in CHF beträgt:

$$\bar{x} = \frac{10000 \cdot 7514 + 35000 \cdot 13645 + 75000 \cdot 12540 + 150000 \cdot 3063 + 500000 \cdot 811}{37573} = 62'762.- \quad \text{oder}$$

$$\bar{x} = 10000 \cdot 0.200 + 35000 \cdot 0.363 + 75000 \cdot 0.334 + 150000 \cdot 0.082 + 500000 \cdot 0.022 = 63'055.-$$

Median

$$\text{Median} = \text{Untergrenze der Medianklasse} + \frac{j}{f} \cdot \text{Medianklassenbreite}$$

mit j = Anzahl (Anteil) Werte in der Medianklasse bis zur Mitte

f = Anzahl (Anteil) Werte in der Medianklasse

Zentralwert (Median) teilt die, nach Grösse sortierten Merkmalswerte in zwei Hälften. 50% höher als Median, 50% kleiner. Z.B. mittlerer bezahlter Preis.
Der Median ist das 0.5-Quantil, das 2. Quartil (2/4), das 5. Dezil (5/10) oder das 50. Perzentil (50/100).

Einkommensklasse (in CHF)			rel. Häufigkeit	rel. H. kumuliert
0	bis unter	20'000	20.0%	20.0%
20'000	bis unter	50'000	36.3%	56.3%
50'000	bis unter	100'000	33.4%	89.7%
100'000	bis unter	200'000	8.2%	97.8%
200'000	und mehr		2.2%	100.0%
Total			100.0%	

$$\text{Median} = 20000 + \frac{50 - 20.0}{36.3} \cdot 30000 = \text{CHF } 44'793.-$$

$$3.\text{Quartil} = 50000 + \frac{75 - 56.3}{33.4} \cdot 50000 = 77'994.-$$

$$0.9\text{-Quantil} = 100000 + \frac{90 - 89.7}{8.2} \cdot 100000 = 103'659.- \approx \text{CHF } 104'000.- \quad \text{resp.}$$

$$0.9\text{-Quantil} = 100000 + \frac{90 - 89.7}{8.1} \cdot 100000 = 103'704.- \approx \text{CHF } 104'000.-$$

Modus

Häufigster Messwert. Wenn in Klasseneinteilung dann Mitte der Klasse mit höchster Dichte.

Einkommensklasse (in CHF)			rel. Häufigkeit	Klassenbreite (in CHF 10000)	Dichte
0	bis unter	20'000	14%	2	7%
20'000	bis unter	50'000	30%	3	**10%**
50'000	bis unter	100'000	35%	5	7%
100'000	bis unter	200'000	19%	10	1.9%
200'000	und mehr		2%	...	...

Modus: 35'000 Modalklasse: 20T bu. 50T. (10% Dichte)

Modusklasse: Klasse mit höchster Dichte, Achtung nicht unbedingt grösste relative Häufigkeit.

rechtsschiefe Verteilung	meistens Einkommensverteilungen	Modus < Median < arithmetisches Mittel
linksschiefe Verteilung		arithmetisches Mittel < Median < Modus
symmetrische Verteilung		arithmetisches Mittel = Median = Modus

Geometrisches Mittel

$$x_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

Wir benutzt wenn Renditen zeitlich aufeinander folgen und ich einen Durchschnitt will

1. Alle Prozente in Faktoren umrechnen x1, x2 (-10% sind 0.9, +60% sind 1.6)
2. Anzahl Faktoren berrechnen = n
3. Faktoren miteinander multiplizieren = x
4. n-te Wurzel von x

59 Jahre	Produktionswachstum USA [in %]
1948 – 1972 24	4.0
1972 – 1995 23	3.4
1995 – 2007 12	3.5

Ausnahme-Beispiel über mehrere Jahre:

$$\sqrt[59]{1.040^{24} \cdot 1.034^{23} \cdot 1.035^{12}} - 1$$

kumulierte Rendite

$$R_{kumuliert} = R_{tA} \cdot R_{tB} \cdot R_{tC} \cdot R_{tn}$$

Alle Faktoren multiplizieren zum kummulieren

durchschnittliche Wachstumsrate

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

Sind absolute Zahlen bekannt, dann gilt:

$$\bar{R} = \sqrt[n]{\frac{\text{Endwert}}{\text{Anfangswert}}} - 1$$

Variante 1 gibt $\bar{R}$ = Faktor z.B. 1.22 für +22%
Variante 2 gibt $\bar{R}$ = Prozent z.B. 0.2 für 20%

$$\bar{R} = \sqrt[5]{1.05 \cdot 1.12 \cdot 0.96 \cdot 0.85 \cdot 1.2} = 1.0286 - 1 = 0.0286 = 2.86\%$$

Eine andere Firma steigert ihren Umsatz in 5 Jahren von 10 Mio. auf 16 Mio. Die durchschnittliche jährliche Umsatzsteigerung ist

$$\bar{R} = \sqrt[5]{\frac{16}{10}} - 1 = 1.6^{0.2} - 1 = 9.86\%$$

stetige Rendite (logarithmierte)

$$\ln(x_G) = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln(x_i) \quad \text{bzw.} \quad \bar{r} = \ln(1 + \bar{R}) = \frac{1}{n} \cdot \sum_{t=1}^{n} \ln(1 + R_t)$$

$$\bar{r} = \ln(1 + \bar{R}) = Ergebnis \cdot 100 = stetige\ Rendite\ in\ \%$$

Ist R eine effektive (oder einfache) Rendite, r-Oberstrich ist dann die stetige Rendite, engl. Log-Return.
Logarithmus des geometrischen Mittels ist das arithmetische Mittel der Logarithmen der Faktoren, für die das geometrische Mittel berechnet werden soll.

Eine Aktie steigt im ersten Jahr 20%, im zweiten Jahr um 30%. Die durchschnittliche Zunahme pro Jahr rechnen wir entweder mit

$x_G = \sqrt{1.2 \cdot 1.3} = 1.249$ durchschnittliche effektive Rendite $\bar{R}$ 24.9%

arithmetischen Mittel der stetigen Renditen

$$\bar{r} = \frac{\ln(1.2) + \ln(1.3)}{2} = 0.222 = 22.2\% \quad = \text{durchschnittliche stetige Rendite}$$

>>> Die durchschnittliche effektive Rendite ist dann: $\bar{R} = e^{0.2223} - 1 = 24.9\%$

Streuungsmasse A3

Spannweite

Spannweite = Maximum – Minimum

Einfach in Bestimmung, Nachteil beide Extremwerte sind berücksichtigt (Ausreisser)

Quartilsabstand

Quartilsabstand = 3. Quartil – 1. Quartil

Spannweite der mittleren 50% der Messwerte

A: 5.0 – 4.0 = 1.0
B: 5.0 – 4.0 = 1.0
C: 5.5 – 3.5 = 2.0
D: 4.0 – 4.0 = 0.0 *(!)*

Note	relative Häufigkeiten (in %)				relative Häufigkeiten kumuliert (in %)			
	Gruppe A	Gruppe B	Gruppe C	Gruppe D	Gruppe A	Gruppe B	Gruppe C	Gruppe D
1.0	0	1	3	0	0	1	3	0
1.5	0	0	2	0	0	1	5	0
2.0	0	0	3	0	0	1	8	0
2.5	0	0	5	0	0	1	13	0
3.0	1	1	9	0	1	2	22	0
3.5	6	5	5	23	7	7	27	23
4.0	25	23	10	55	32	30	37	78
4.5	36	35	10	21	68	65	47	99
5.0	24	28	11	1	92	93	58	100
5.5	8	6	22	0	100	99	80	100
6.0	0	1	20	0	100	100	100	100

Varianz
(sigma-Quadrat)

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} \quad \text{resp.} \quad \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2\cdot W_i}{\sum_{i=1}^{n} Wi}$$

Man berrechnet von jedem Messwert die Abweichung vom Mittelwert der Messwerte. Anschliessend berrechnet man den Durchschnitt dieser Abweichungen. NUR MIT ABSOLUTEN, realen WERTEN

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Note x_i	W_i	$x_i \cdot W_i$	$x_i-\bar{x}$	$(x_i-\bar{x})\cdot W_i$	$\lvert x_i-\bar{x}\rvert$	$\lvert x_i-\bar{x}\rvert\cdot W_i$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^2\cdot W_i$
3.0	1	3	-1.5	-1.5	1.5	1.5	2.25	2.25
3.5	6	21	-1	-6	1	6	1	6
4.0	25	100	-0.5	-12.5	0.5	12.5	0.25	6.25
4.5	36	162	0	0	0	0	0	0
5.0	24	120	0.5	12	0.5	12	0.25	6
5.5	8	44	1	8	1	8	1	8
Summe	100	450		0		40		28.5
Durchschnitt		$\bar{x} = 4.5$		0		0.4		0.285

$\sqrt{0.285} = 0.534$

1) Ausprägungen
2) Gewichtung
3) Gewichtete Werte (Summe 3 : Summe 2 = $\bar{x}$)
4) Abweichung von $\bar{x}$
5) Gewichtete Abweichung (Summe davon = 0 wenn nicht $\bar{x}$ nicht arith. Mittel)
6) Spalte 4 aber Minus weggelassen
7) Spalte 6 mal Gewichtung (Summe 7 : Summe 2 = **durchschnittliche Abweichung**)
8) Abweichung Spalte 4 hoch 2 so geht Minus auch weg
9) Abweichung Spalte 8 gewichtet

Summe Spalte 9 : Summe 2 = durchschnittliche quadratische Abweichung "Varianz" $\sigma^2 = \frac{28.5}{100} = 0.285$

Wurzel Varianz = Standardabweichung $\sigma = \sqrt{\frac{28.5}{100}} = 0.534$

Standardabweichung
(sigma)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}} \quad \text{resp.} \quad \sqrt{\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2\cdot W_i}{\sum_{i=1}^{n} Wi}}$$

Man berrechnet einfach die Wurzel der Varianz. ACHTUNG NUR MIT ABSOLUTEN, realen WERTEN!

empirische Varianz

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \quad \text{Bsp.} \quad s^2 = \frac{28.5}{99} = 0.288$$

Varianz welche bei Stichproben Erhebungen verwendet wird und empirisch die Varianz schätzt

empirische Standardabweichung

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \quad \text{Bsp.} \quad s = \sqrt{\frac{28.5}{99}} = 0.537$$

Standardabweichung welche bei Stichproben Erhebungen verwendet wird und empirisch die Standardabweichung schätzt

s via Sigma

$$s = \sqrt{\frac{n}{n-1}}\cdot\sigma$$

Wenn der TR nur sigma liefert aber nicht s kann ich so die empirische Standardabweichung (s) berrechnen.

Standardabweichung und Varianz sind nützlich für: Risikomessung bei Finanzanlagen, Volatilität: Standardabweichung der stetigen Renditen; oder Qualitätsmessungen in der Produktion: Standardabweichung oder Varianz von Längen, Gewichten, Reissfestigkeiten

Variationskoeffizient

$$V_\sigma = \frac{\sigma}{\bar{x}} \approx \frac{s}{\bar{x}}$$

Streuung im Verhältnis zum Mittelwert. Wird benutzt wenn Streuungen von verschiedenen Merkmalen mit untersch. Grössenordnungen verglichen werden. Variationskoeffizient ergibt ein Dimensionsloses Resultat man kann es in % angeben.

TR Kontrolle

0) TR data
1) L1 $\mathbf{x_i}$ eingeben (dh. absolute Werte oder Klassenmitte)
2) L2 $\mathbf{W_i}$ eingeben (dh. absolute Häufigkeit)
3) TR 2nd 1-Var Stats L1"data", L2 "menge" ergibt **n** und $\bar{x}$ und **s** oder **sigma** (wenn w_i und nicht W_i als Häufigkeit dann nur sigma entnehmen, s stimmt dann nicht) kann auch Varianzschätzer $\mathbf{s^2}$ entnehmen (s_x^2)
4) L3 = L2 / n * 100 = relative Häufigkeit = $\mathbf{w_i}$
5) data Pfeil-Rechts 5 um Formeln zu löschen
6) L1 neu mit Klassenobergrenze befüllen
7) L2 neu mit Klassenbreite befüllen
8) **Median** = TR LinReg L2, L1, one, Yes dann f(50)
9) **Quartilsabstand** = TR LinReg L2, L1, one, Yes dann f(25)=Q1 und f(75)=Q3 anschliessend Q3-Q1=Quartilsabstand
10) L1 löschen und neu mit L3 : L2 = L1 **Dichte** befüllen
11) **Modalklasse** bestimmen, **Modus** entnehmen

Volatilität

$$\bar{x}_{ann} = n\cdot\bar{x} \quad \text{und} \quad s_{ann} = \sqrt{n}\cdot s$$

wobei n = 4, 12, 52 oder 250

Annualisierte (auf ein Jahr hochgerechnet) Standardabweichung der stetigen (log.) Renditen. Benutzt um Risiko zu berrechnen.

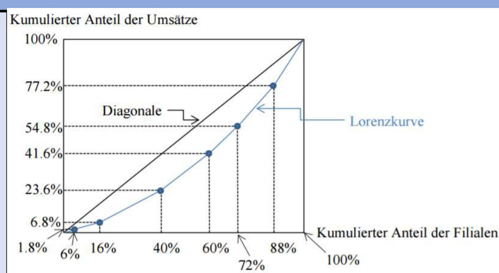
Monat	Jun	Jul	Aug	Sep	Okt	Nov	Dez	Jan	Feb	Mär	Apr	Mai	Jun
Kurs [CHF] oder Index	100.0	108.0	113.4	111.7	116.5	117.9	110.0	105.6	109.3	105.8	102.0	107.1	114.6
effektive Rendite	…	8.0%	5.0%	-1.5%	4.3%	1.2%	-6.7%	-4.0%	3.5%	-3.2%	-3.6%	5.0%	7.0%
stetige Rendite	…	7.7%	4.9%	-1.5%	4.2%	1.2%	-6.9%	-4.1%	3.4%	-3.3%	-3.7%	4.9%	6.8%

Wir haben Kurs oder eben Rendite und berrechnen die effektive Rendite (% Wachstum) und davon die stetige Rendite in Prozent (ln(1+R)*100). Wir berrechne dann das arith. Mittel der stetigen Renditen $\overline{x}_{ann}$ (Summe aller durch 12 wenn Monate hier = 1.1%) sowie die Standardabweichung der stetigen Renditen (hier = 4.9%) danach können wir die Formeln anwenden und die Volatilität s_{ann} ableiten.

TR Kontrolle	1) Kurse in L1	5) L1 = ln(L3) = **stetige Rendite**
	2) L2 = L1, dann Formeln löschen (data, Pfeil-rechts, 5)	6) 2nd Data 1-Var-Stats L1 => arith. Mittel und
	3) erster Eintrag in L1 sowie letzter Eintrag in L2 löschen	7) $\bar{x} \cdot n$ und $Sx \cdot \sqrt{n}$ (Monate: n = 12)
	4) L3 = L1 : L2, dann Formeln löschen ergibt Faktoren	

Konzentrationsmasse A4

Lorenzkurve



Eigenet sich zur Darstellung der Konzentration/Verteilung eines metrischen Merkmals (stetig od. diskret). Ziel ist es die Konzentration vom Merkmal auf ihre Merkmalsträger darzustellen. Auf der x Achse (waagrecht) der Anteil Haushalte, auf der y Achse (senkrecht) Anteil des Vermögens.
Auf Diagonale hätten 100% der Haushalte 100% des Vermögens also perfekt verteilt. Je weiter weg von Diagonale desto ungleicher verteilt. Je ungleicher verteilt, desto konzentrierter. Weil Werte in grösseren Päckli sind (konzentriert), wenn überall gleich verteilt sind Päckli überall gleich gross somit unkonzentriert

Lorenzkurve Interpretation

Lorenzkurve immer schwach konvex

Wenn keine weiteren Informationen vorliegen und man sich z.B. für den Vermögensanteil der 30% vermögensärmsten Haushalte interessiert, dann muss man linear interpolieren (TR: LinReg).

Lorenzkurve Darstellung

Umsatz (in 1000 CHF)		Klassenmitte	Häufigkeit			Umsatz		
			absolut	relativ	rel. kum.	absolut*	relativ	rel. kum.
10 bis unter	20	15	3	0.06	**0.06**	45	0.018	**0.018**
20	30	25	5	0.10	**0.16**	125	0.050	**0.068**
30	40	35	12	0.24	**0.40**	420	0.168	**0.236**
40	50	45	10	0.20	**0.60**	450	0.180	**0.416**
50	60	55	6	0.12	**0.72**	330	0.132	**0.548**
60	80	70	8	0.16	**0.88**	560	0.224	**0.772**
80	110	95	6	0.12	**1.00**	570	0.228	**1.000**
Total			50	1.00		2500	1.00	

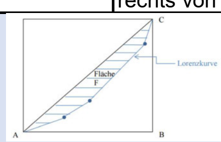
1) Aufsteigend sortieren (gegeben)
2.1) Habe Umsatzklassen, davon Klassenmitte ableiten wenn keine Detaildaten
2.2) Häufigkeit absolut, relativ und rel. kumuliert auflisten
2.3) Umsatz absolut (absolute H. * Klassenmitte Umsatz), relativ und rel. kumuliert auflisten
3.1) waagrecht x-Achse (Menge) also hier Anteil Filialen
3.2) senkrecht y-Achse (Wert) also hier Anteil Umsatz
>> Zeichnen wie oben unter Lorenzkurve Definition

Lorenzkriterium

Zwei Lorenzkurven übereinander wenn Kurve B immer rechts von Kurve A dann ist B definitiv ungleicher als A

Zwei Lorenzkurven übereinander wenn Kurve B und Kurve A sich schneiden, kann nicht eindeutig gesagt werden wo ungleicher.

Gini-Koeffizient



$$\text{Gini-Koeffizient} = \frac{\text{Fläche F}}{\text{Dreiecksfläche ABC}}$$

Wir benutzt weil aus Lorenzkurve eben nicht immer genau Ungleichheit abgelesen werden kann. Gini-Koeffizient misst das Verhältnis der Fläche zwischen der Lorenzkurve und der Diagonalen zur Dreiecksfläche unterhalb der Diagonalen. **Gibt Ungleichheit an.** Wenn Gini steigt, steigt Ungleichheit.

Gini-Koeffizient Interpretation

0 ≤ Gini-Koeffizient < 1

Gini-Koeffizient ist dimensionslos

Grosser Gini impliziert grosse Konzentration (z.B. grosse Ungleichheit im Einkommen) und umgekehrt

Gini-Wert von 0 gleich keine Konzentration z.B. vollkommene Gleichheit der Einkommen

Achtung mit Interpretation gleiche Lorenzkurve aber bspw. gespiegelt ergibt ähnliche Verteilung aber gleichen Gini-Koeffizient trotzdem ist das Vermögen sehr unterschiedlich konzentriert.

Wenn nur 2 Unternehmen sich je 50% des vermögens teilen ist der Gini = 0, dh. es wäre unkonzentriert (gleich verteilt) aber eigentlich rein logisch wäre das Vermögen sehr stark konzentriert.

absolute Konzentration	Konzentrationsrate also logische Konzentration (Gegenteil von Gini-Koeffizient und Lorenzkurve
relative Konzentration	Gini-Koeffizient und Lorenzkurve messen relative Konzentration, darum ist es aus dem Bsp. Mit den beiden Unternehmen wenig konzentriert.

Bivariate Statistik

Korrelation B1

Kovarianz

Je grösser/kleiner X, desto grösser/kleiner Y (positive Korrelation) ODER Je grösser/kleiner X, desto kleiner/grösser Y (negative Korrelation)

$$\mathrm{Cov}(X,Y) = \sigma_{XY} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

Untersucht Zusammenhänge und quantifiziert diese. Benötigt metrische Merkmale. Koordinaten (Abweichung von Durchschnitt) werden zuerst miteinander multipliziert anschliessend werden die Produkte summiert und durch die Anzahl Datenpaare dividiert

empirische Kovarianz

$$s_{XY} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$$

empirisch geschätzt, analog zu empirischer Streumassen auch Nenner -1, ist für Stichproben

Korrelationskoeffizient

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{s_{XY}}{s_X \cdot s_Y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Abweichung zum Mittelwert wird im Gegensatz zur Kovarianz noch zusätzlich durch die jeweilige Standardabweichung dividiert. Das standardisiert die Daten. (Gleiche Formel wie oben nur zusätzlich dividiert durch Standardabweichung.) **Es gilt -1 ≤ r ≤ 1 Werte nahe bei 1 zeigen einen starken Zusammenhang!** Korrelationskoeffizient misst nur Lineare Zusammenhänge darum auch TR LinReg und keine quadratischen. r=0 heisst kein linearer Zusamamenhang aber vielleicht ein quadratischer das sieht man beim Zeichnen

funktionaler Zusammenhang: Wenn Korrelationskoeffizient nahe bei +/- 1 dann ist ein funktionaler Zusammenhang gegeben aber nicht unbedingt kausaler Z.

kausaler Zusammenhang: Ursache-Wirkung Zusammenhang, wegem dem ist das so. Wenn ein Kausal-zhang vermutet wird rechnet man Korrelation.

r	Stärke funktionaler Zusammenhang	R^2
±0		0%
	kein / sehr schwach	
±0.25		ca. 5%
	schwach	
±0.5		25%
	mittel	
±0.75		ca. 50%
	stark	
±1		100%

Beispiel:

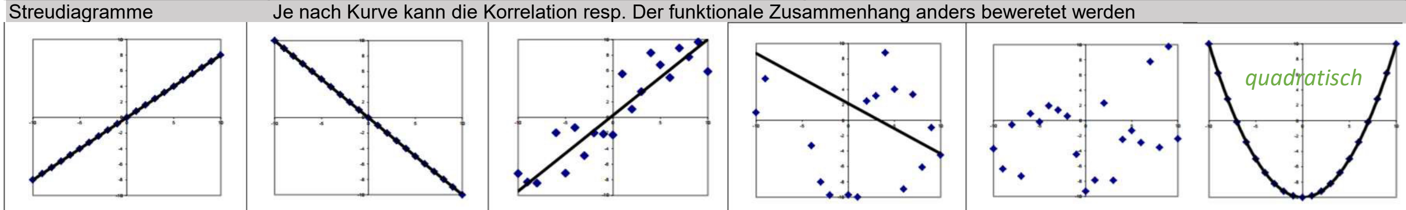
Land	Unabhängigkeit x_i	Inflation (in %) y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
Australien	2	6.5	-0.36	0.8	-0.288
Belgien	2	4.2	-0.36	-1.5	0.54
CH	4	3.6	1.64	-2.1	-3.444
Dänemark	2.5	6.6	0.14	0.9	0.126
Deutschland	4	3.3	1.64	-2.4	-3.936
Frankreich	2	6.2	-0.36	0.5	-0.18
GB	2	6.7	-0.36	1.0	-0.36
Italien	1.75	7.3	-0.61	1.6	-0.976
Japan	2.5	5.0	0.14	-0.7	-0.098
Kanada	2.5	4.6	0.14	-1.1	-0.154
Neuseeland	1	7.8	-1.36	2.1	-2.856
Niederlande	2.5	4.3	0.14	-1.4	-0.196
Norwegen	2	6.2	-0.36	0.5	-0.18
Schweden	2	6.2	-0.36	0.5	-0.18
Spanien	1.5	8.4	-0.86	2.7	-2.322
USA	3.5	4.3	1.14	-1.4	-1.596
Summe					-16.1
Durchschnitt	$\bar{x} = 2.36$	$\bar{y} = 5.7\%$			
Standardabweichung	0.837	1.54%			

Wenn jetzt die Werte in Tabelle grösser wären wäre auch Kovarianz dementsprechen grösser… Somit ist sie abhängig von den Merkmalen und kann kaum interpretiert werden… Dazu gibt es die Standardisierung mithilfe des Korrelationskoeffizienten

$$s_{xy} = \frac{-16.1}{16-1} = -1.07$$

$$r = \frac{-1.07}{0.837 \cdot 1.54} = -0.83$$

Streudiagramme: Je nach Kurve kann die Korrelation resp. Der funktionale Zusammenhang anders beweretet werden



a) $r = 1$: positiver funktionaler Zusammenhang

b) $r = -1$: negativer funktionaler Zusammenhang

c) $r \approx 0.9$: starker positiver Zusammenhang

d) $r \approx -0.6$: mittlerer negativer Zusammenhang

e) $r \approx 0$: kein Zusammenhang (Unabhängigkeit)

f) $r = 0$: kein linearer Zusammenhang

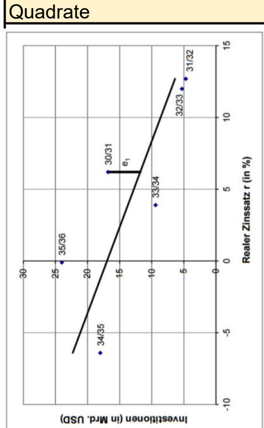
TR Kontrolle

1) x-Achse Werte eingeben (L1)
2) y-Achse Werte eingeben (L2)
3) TR: data, L1, L2, one, yes, LinReg
4) Korrelationskoeffizient r auslesen

Regression B2

Methode der kleinsten Quadrate

Man geht von einem Zusammenhang zwischen x-Achse und y-Achse aus (Zusammenhang ist in Form von Modell linar, exponentiell etc.) Ausgehend von den Punkten auf Streudiagramm versucht man dann eine Gerade (Funktionsgleichung) zu



$$S = \sum_{i=1}^{n} e_i^2 \text{ minimieren.}$$

1) S minimieren

$$S = \sum_{i=1}^{n} (y_i - (ax_i + b))^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \rightarrow \text{Min.}$$

2) Funktionsgleichung einsetzen, für welches a und b ist S minimal?

Und damit das Gleichungssystem:

$$399.91a + 28.3b = 146.51 \qquad \text{(Gleichung 1)}$$
$$28.3a + 6b = 78.2 \qquad \text{(Gleichung 2)}$$

3) ergibt 2 Gleichungen

`TR: sys-solv` liefert die Regressionsgleichung:
Für die Daten x_i und y_i heisst dies:
Mit Hilfe der Regressionsgeraden schätzen wir die y-Werte $\hat{y}_i$

4) Gleichungen auflösen

$$y = -0.83x + 16.97$$
$$y_i = -0.83x_i + 16.97 + e_i$$
$$\hat{y}_i = -0.83x_i + 16.97$$

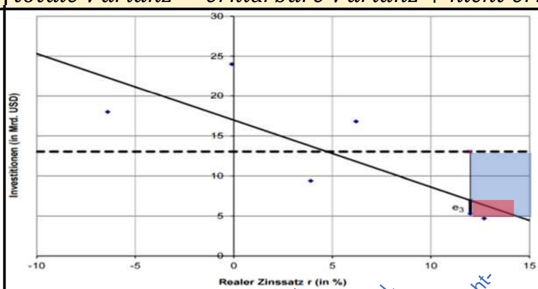
TR Berrechnung

Jahr	Realer Zins (in %) x_i	Investition (in Mrd. $) y_i	x_i^2	$x_i \cdot y_i$
30/31	6.2	16.8	38.44	104.16

1) x-Werte in L1, y-Werte in L2
2) 2nd, data, Modell wählen (hier LinReg)
3) a und b ablesen ergibt Regressionsgleichung (siehe oben)

31/32	12.7	4.7	161.29	59.69
32/33	12.0	5.3	144.00	63.60
33/34	3.9	9.4	15.21	36.66
34/35	–6.4	18.0	40.96	–115.20
35/36	–0.1	24.0	0.01	–2.40
Summe	*28.3*	*78.2*	*399.91*	*146.51*

Lineares Modell	$y = ax + b$	(TR: *Lin*Reg ax+b)
Polynom Modell	$y = \sum_{i=0}^{n} a_i x^i$	(Excel)
	$y = ax^2 + bx + c$	(TR: QuadraticReg)
	$y = ax^3 + bx^2 + cx + d$	(TR: CubicReg)
Logarithmisches Modell	$y = a + b \cdot \ln(x)$	(TR: *Ln*Reg a+blnx)
Potenz Modell	$y = a \cdot x^b$	(TR: PwrReg ax^b)
Exponentielles Modell	$y = a \cdot b^x$	(TR: ExpReg ab^x)
	$y = a \cdot e^{c \cdot x} = a \cdot \left(e^c\right)^x$	(Excel)

Bestimmtheitsmass R^2	$R^2 = \frac{Erklärbare\ Varianz}{Totale\ Varianz}$	**Anteil erklärbarer Varianz** Man fragt sich wie viel der Streuung man auf den Input zurückführen kann resp. Auf etwas anderes? Achtung schauen ob Stichprobe ode rnicht wegen Varianz
Varianz Zusammensetzung	$totale\ Varianz = erklärbare\ Varianz + nicht\ erklärbare\ Varianz$	
Bestimmtheitsmass Bedeutung in zusammenhang Regressionsmodell		Wie viel von dieser Varianz kann durch die Schwankung des realen Zinssatzes erklärt werden? Wir setzen die **Summe der quadratischen Abstände zwischen den Datenpunkten und der Regressionsgeraden** ins Verhältnis zur **Summe der quadratischen Abstände zwischen den Datenpunkten und dem Durchschnitt der y-Werte** (vgl. Abb. 27). Damit erhalten wir den Anteil der nicht erklärbaren Varianz.ẏ 13.0

Berrechnungs Beispiel

L3 = f(L1) L1 = L2 - L3

Jahr n ⇩	Realer Zins (in %) L1 x_i	Investition (in Mrd. $) L2 y_i	Schätzer (in Mrd. $) $\hat{y}_i$	Differenz (in Mrd. $) $y_i - \hat{y}_i$	Total für tot. Varianz $(y_i - \bar{y})^2$	Beitrag erkl. Varianz $(\hat{y}_i - \bar{y})^2$	Beitrag nicht-erkl. Varianz $(y_i - \hat{y}_i)^2$
30/31	6.2	16.8	11.8	5.0	14.2	1.5	25.0
31/32	12.7	4.7	6.4	-1.7	69.4	44.4	2.8
32/33	12.0	5.3	7.0	-1.7	59.8	36.9	2.7
33/34	3.9	9.4	13.7	-4.3	13.2	0.5	18.6
34/35	-6.4	18	22.3	-4.3	24.7	86.1	18.6
35/36	-0.1	24	17.1	6.9	120.3	16.2	48.3
Summe		78.2 = !!!!	78.2	0.0	301.6	185.5	116.0

durchschnittliche Investitionen $\bar{y}$: n ⇨ $\frac{78.2}{6} = 13.0$ Mrd. USD

Totale Varianz s_Y^2 (empirisch): n -1 ⇨ $\frac{301.6}{5} = 60.3$

Erklärbare Varianz: $\frac{185.5}{5} = 37.1$

Nicht erklärbare Varianz: $\frac{116.0}{5} = 23.2$

Anteil erklärbarer Varianz: $R^2 = \frac{37.1}{60.3} = 61.5\%$

Anteil nicht erklärbarer Varianz: $1 - R^2 = 1 - 61.5\% = 38.5\%$

Interpretation	61.5% der Varianz der Investitionstätigkeit ist durch die Zinsentwicklung erklärbar, 38.5% sind nicht erklärbar. Es gibt noch andere Gründe für die Schwankung der Investitionen. R2 heisst Bestimmtheitsmass. ***Je grösser R^2, desto besser passt sich die Regressionskurve der Punktewolke an.*** $R^2 = 1$: Der vermutete Zusammenhang trifft absolut zu. $R^2 = 0$: Der vermutete Zusammenhang trifft absolut *nicht* zu. (siehe Zusammenhangstabelle oben)	
Hinweis	Im linearen Regressionsmodell ist das Bestimmtheitsmass gleich dem Korrelationskoeffizienten im Quadrat: $R^2 = r^2$. In unserem Beispiel: 0.615 = (–0.7844)2 ✓	
TR Kontrolle	1) x-Werte in L1	4) a und b ablesen für Methode der kleinsten Quadrate
	2) y-Werte in L2	5) r^2 oder R^2 ablesen für Bestimmtheitsmass
	3) 2nd, data, passendes Modell wählen z.B. LinReg	
Lineare Mehrfachregression	3 Spalten L1, L2, L3 ich muss Methode der kleinsten Quadrate von L1 & L3 sowie von L2 & L3 machen wenn ich für L3 schätzen will. Dann Formel zusammensetzen (LinReg)	Eine Zielgrösse y hänge von mehreren Einflussfaktoren x1, x2, x3 ab ist die Behauptung.
Voraussetzungen für Mehrfachregression	Damit ich die Schätzung durch Mehrfachregression machen kann dürfen keine Einflussfaktoren korreliert sein (Praxis nur schwach) sonst gibt es Fehler.	
Umsatz [1000 CHF] = 2.93·Grösse [in km^2] – 0.141· Anz. Einw. [in 1000] + 557		Weil die Faktoren stark korreliert sind zeigt diese Regressionsgleichung dass **je mehr Einwohner** da sind der **Umsatz abnimmt…** Was keinen Sinn macht…

Zeitreihen

Indexzahlen C1

Einfacher Index

$$\frac{absolute\ Zahl\ neues, jüngeres\ Jahr}{absolute\ Zahl\ Basis\ Jahr}\cdot 100 = Index$$

Indexzahl eines anderen Jahres anhand der Basis ausrechnen. Damit kann ich auch für frühere Jahre Index berrechnen und so z.B. verschiedene Warenkorb LIKs verketten.

Veränderung

$$\sqrt[Verstrichen\ \ Jahre]{\frac{absolute\ Zahl\ neues, jüngeres\ Jahr}{absolute\ Zahl\ vorheriges, älteres\ Jahr}}\cdot 100 - 1 = \%Veränderung\ geg. Vorjahr$$

Jahr	BIP real in Mio. Fr.	Index Basis 1990	Index Basis 2000	Veränderung gegenüber Vorjahr in %
1990	447 604	**100.0**	88.7	
2000	504 356	112.7	**100.0**	1.2 *
2005	542 803	121.3	107.6	1.5 *
2006	565 012	126.2	112.0	4.1
2007	588 376	131.5	116.7	4.1
2008	601 466	134.4	119.3	2.2
2009	588 760	131.5	116.7	−2.1
2010	605 706	135.3	120.1	2.9
2011	617 161	137.9	122.4	1.9
2012	623 940	139.4	123.7	1.1
2013	635 990	142.1	126.1	1.9
2014	648 552	144.9	128.6	2.0
2015	652 555	145.8	129.4	0.8

10 Jahre

Beispiel:

Durchschnittliche Veränderung 1990 bis 2000 : $\sqrt[10]{\frac{112.7}{100.0}} = \sqrt[10]{1.127} = 1.012$

durchschnittliche Zunahme von 1.2% pro Jahr.

Index Basis 1990 = 447'604 um Index von 2007 zu berrechnen einfach BIP 2007 getelt durch Basiswert $\frac{588376}{447604}\cdot 100 = 131.5$

Veränderung 2010 bis 2011: $\frac{617\,161}{605\,706} = 1.0189118 \approx 1.019$ oder $\frac{137.9}{135.3} = 1.019$

also eine Zunahme um 1.9% $= (1.019-1)\cdot 100\%$

Umbasierung

$$\frac{Indexzahl\ Jahr\ x}{Indexzahl\ Jahr\ Basis}\cdot Jahr\ Basis\ Indexzahl\ in\ umbasiertem\ Index$$

$$= Indexzah\ \ Jahr\ x\ in\ umbasiertem\ Index$$

Indexzahl eines Jahres mit Basis x auf eine neue Basis y umbasieren/umrechnen. Ähnlich wie einfacher Index ausrechnen nur andere Basis.

Zusammengesetzter Index (Laspeyres Index)

Erkärung:

$$I_t = \frac{\sum_{i=1}^{n}(\text{aktueller Preis}_i \cdot \text{Basismenge}_i)}{\sum_{i=1}^{n}(\text{Basispreis}_i \cdot \text{Basismenge}_i)}\cdot 100 \quad \text{n Güter } i = 1, \dots, n \ ; \ \text{aktuelle Periode} = t$$

Mathematische Notation:

$$I_t = \frac{\sum_{i=1}^{n} p_t^{(i)}\cdot q_0^{(i)}}{\sum_{i=1}^{n} p_0^{(i)}\cdot q_0^{(i)}}\cdot 100$$

$p_0^{(i)}$	Preis des Gutes i in der Basisperiode
$p_t^{(i)}$	Preis des Gutes i in der Periode t
$q_0^{(i)}$	Menge des Gutes i in der Basisperiode
$w_0^{(i)} = p_0^{(i)}\cdot q_0^{(i)}$	Wert des Gutes i in der Basisperiode

$$I_t = \frac{\sum_{i=1}^{n}\text{Preisindex}_i\cdot\text{Wert}_i}{\sum_{i=1}^{n}\text{Wert}_i}$$

Der Laspeyres Index ist nichts anderes als das gewichtete arithmetische Mittel der einfachen Preisindizes der einzelnen Güter. Gewichtet wird mit den wertmässigen Anteilen der einzelnen Güter in der Basisperiode.

Trick 77:
Übers Kreuz multiplizieren, durch restlichen, gegebener Wert dividieren

Monat t	Gut 1 Preis (Fr./Stück)	Gut 1 Preisindex
0	31	100.00
1	48	154.84

$$\frac{48\cdot 100}{31} = 154.84$$

$$\frac{154.84\cdot 31}{100} = 48$$

Beispiel:

Monat t	Gut 1 Preis (Fr./Stück)	Gut 1 Menge (Stück)	Gut 2 Preis (Fr./m)	Gut 2 Menge (m)	Gut 3 Preis (Fr./kg)	Gut 3 Menge (kg)
0	31	18	11	340	8.7	520
1	48	…	15	…	9.0	…
2	55	…	20	…	9.0	…

Monat 0 $\quad I_0 = 100.0$

Monat 1 $\quad I_1 = \frac{48\cdot 18 + 15\cdot 340 + 9.0\cdot 520}{31\cdot 18 + 11\cdot 340 + 8.7\cdot 520}\cdot 100 = 120.7$

Monat 2 $\quad I_2 = \frac{55\cdot 18 + 20\cdot 340 + 9.0\cdot 520}{31\cdot 18 + 11\cdot 340 + 8.7\cdot 520}\cdot 100 = 141.4$

einfache Preisindizes der drei Güter

Monat t	Gut 1 Preis (Fr./Stück)	Gut 1 Preisindex	Gut 2 Preis (Fr./m)	Gut 2 Preisindex	Gut 3 Preis (Fr./kg)	Gut 3 Preisindex
0	31	100.00	11	100.00	8.7	100.00
1	48	154.84	15	136.36	9.0	103.45
2	55	177.42	20	181.82	9.0	103.45

Gewichtung der 3 Güter in der Basisperiode

Monat 0	Gut 1	Gut 2	Gut 3	Total
Preis	31	11	8.7	
Menge	18	340	520	
Wert, Gewicht absolut	558	3740	4524	8822
Wert, Gewicht relativ	0.0633	0.4239	0.5128	1.0000

Monat 0 $\quad I_0 = 100.0$

Monat 1 $\quad I_1 = \frac{154.84\cdot 558 + 136.36\cdot 3740 + 103.45\cdot 4524}{8822} = 120.7$ oder $I_1 = 154.84\cdot 0.0633 + 136.36\cdot 0.4239 + 103.45\cdot 0.5128 = 120.7$

Monat 2 $\quad I_2 = \frac{177.42\cdot 558 + 181.82\cdot 3740 + 103.45\cdot 4524}{8822} = 141.4$ oder $I_2 = 177.42\cdot 0.0633 + 181.82\cdot 0.4239 + 103.45\cdot 0.5128 = 141.4$

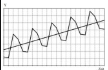
Diese Berechnung eines Laspeyres Preisindexes hat den Vorteil, dass sie nicht nur für die Zusammensetzung eines Indexes von Preisen von verschiedenen Gütern verwendet werden kann, sondern auch für die Zusammensetzung eines Gesamtindexes aus verschiedenen Teilindizes. So wird beispielsweise der schweizerische Landesindex der Konsumentenpreise (LIK) aus verschiednen Teilindizes hierarchisch zusammengesetzt.

Zeitreihenanalyse C2

Zeitreihe: Entwicklung einer Grösse im Verlaufe der Zeit, gemessen in regelmässigen Zeitabständen

deskriptive Zeitreihenanalyse	Befasst sich mit der Zerlegung einer Zeitreihe in folgende Komponenten:
Trend	Grundrichtung Zeitreihe = Trend, Trend ist langfr. Entwicklungsrichtung ➚➘➔
Konjunkturkomponente	Wellenförmige Komponente (Konjunkturschwankungen)
Trendkomponente F_t	Trend und Konjunkturkomponente kombiniert. Auch glatte Komponente genannt.
Saisonkomponente S_t	Beschreibt saisonale Schwankungen während e. Jahr (Arbeitslosenquote), Monats (Umsätze), Woche, Tages etc
Restkomponente E_t	Einmalige und zufällige Einflüsse die schwer abzuschätzen sind

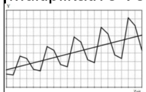
Additive Verbundenheit



Die sich periodisch wiederholenden Abweichungen vom Trend sind unabhängig von der Grösse des Trendwertes; der saisonale Term und der Fehlerterm sind additiv.

$$Y_t = F_t + S_t(+E_t)$$

Multiplikative Verbundenheit



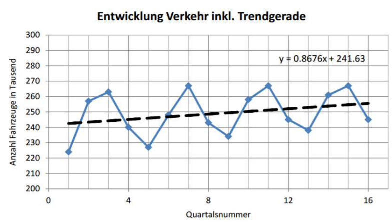
Die sich periodisch wiederholenden Abweichungen vom Trend sind proportional abhängig von der Grösse des Trendwertes. Absolut gesehen entstehen immer grössere Ausschläge vom Trend, wenn die Trendkomponente wächst. Die saisonalen Schwankungen sind multiplikativ und können in Prozenten des Trends dargestellt werden.

$$Y_t = F_t \cdot S_t(\cdot E_t)$$

Ermittlung Trendkomponente F_t:

Methode der kleinsten Quadrate

Jahr	Quartal	Nr. (L1)	Anzahl (in Tausend Fahrzeuge) (L2)
2005	I	1	224
	II	2	257
	III	3	263
	IV	4	240
2006	I	5	227
	II	6	248
	III	7	267



1) TR LinReg
2) Trendfunktion aus Ergebnissen ableiten
3) Ft = a x + b resp. Ft = a * t + b

F_t = 0.868 * t + 242

r = 0.289 ➔ kein gr. Zusammenhang zwischen Zeit und Anzahl Fahrzeuge

a = 0.867
b = 241.63
r^2 = 0.0835
r = 0.289

Gleitender Durchschnitt

Man nimmt den Durchschnitt von 3 Kursen für 3. Ordnung oder 4 Kursen für 4. Ordnung
3. Ordnung = (114+115+102)/3 = 110
4. Ordnung = ((0.5*128)+114+115+102+(0.5*103)) / 4 = 112
Die Methode des gleitenden Durchschnitts eignet sich gut um Trendänderungen sichtbar zu machen.
Nachteil: geglättet ist Zeitreihe kürzer, Trend ohne

Woche	Kurs	Gl. D. 3. Ordnung	Gl. D. 4. Ordnung
1	109	…	…
2	115	117	…
3	128	119	117
4	114	119	116
5	115	*110*	*112*
6	102	107	106
7	103	101	99.6
8	97	92.7	92.4

x-te Ordnung

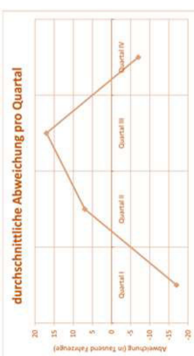
Je nach Zeitspanne andere Ordnung, für Quartale 4. Ordnung für Monatswerte 12. Ordnung. Das was sich wiederholt zusammenfassen. Je tiefere Ordnung desto schwächere Glättung; und umgekehrt.

Ermittlung Saisonkomponente S_t:

additiv

$S_t = Y_t - F_t$ oder $Y_t = F_t + S_t$

Saisonfigur sieht so aus, anhand durchschnittlicher Abweichung über alle Jahre.



L1 L2 L3 = f(L1) S

Jahr	Quartal	Nr.	Anzahl (in 1000) Y	Trend * (in 1000) F	Abweichung im Quartal (in 1000) I	II	III	IV
2005	I	1	224	242	-18			
	II	2	257	243		14		
	III	3	263	244			19	
	IV	4	240	245				-5
2006	I	5	227	246	-19			
	II	6	248	247		1		
	III	7	267	248			19	
	IV	8	243	249				-6
2007	I	9	234	249	-15			
	II	10	258	250		8		
	III	11	267	251			16	
	IV	12	245	252				-7
2008	I	13	238	253	-15			
	II	14	261	254		7		
	III	15	267	255			12	
	IV	16	245	256				-11
durchschnittliche Abweichung					-17	7	17	-7

muss = 0!!!

1) TR LinReg L1 als x, L2 als y, one, yes
2) -> sto, data rechts 5
3) data, -> sto, L3 = f(L1), enter
4) Abweichungen im Quartal errechnet Anzahl - Trendwert = Abweichung (weil Differenz genommen darum additiv!!)
5) Abweichungen können gemittelt werden

*(gleich wie oben) F_t = 0.868 * t + 242*

Prognose

f(17) = 256
f(21) = 260
f(22) = 261

Trendwert + durchschn. Abweichung = Prognosewert

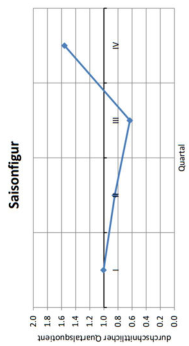
Jahr	Quartal	Nr.	Trendwert (in Tausend Fahrzeuge)	*durchschnittliche* Abweichung (in 1000 Fahrz.)	Prognosewert (in Tausend Fahrzeuge)
2009	I	17	256	−17	239
	II	18	257	+7	264
	III	19	258	+17	275
	IV	20	259	− 7	252
2010	I	21	260	−17	243
	II	22	261	+7	268

Einfach die Funktionsgleichung für die weiteren Quartale (x) ausrechnen.
Die durchschnittliche Abweichung nehme ich von vorher weil ich ja keine neuen Angaben habe. Dadurch erhalte ich Prognosewerte.

multiplikativ

$S_t = Y_t \div F_t$ oder $Y_t = F_t \cdot S_t$

Saisonfigur sieht so aus, anhand durchschnittlicher Quartalsquotient über alle Jahre.



Jahr	Quartal	Nr. (l 1)	Umsatz (in Mio. CHF) Y (L2)	Trendwert (in Mio. CHF) F (L3 = f(L1))	S Quartals*quotient* I	II	III	IV
1	I	1	1.30	1.33	0.977			
	II	2	1.27	1.37		0.927		
	III	3	1.05	1.41			0.746	
	IV	4	2.16	1.45				1.492
2	I	5	1.54	1.49	1.036			
	II	6	1.34	1.53		0.878		
	III	7	0.92	1.56			0.588	
	IV	8	2.46	1.60				1.534
3	I	9	1.61	1.64	0.980			
	II	10	1.38	1.68		0.820		
	III	11	0.99	1.72			0.575	
	IV	12	2.91	1.76				1.653
4	I	13	1.85	1.80	1.028			
	II	14	1.40	1.84		0.762		
durchschnittlicher Quartalsquotient (*geometrisches* Mittel)					1.005	0.845	0.632	1.558

1) TR LinReg L1 als x, L2 als y, one, yes
2) -> sto, data rechts 5
3) data, -> sto, L3 = f(L1), enter
4) Abweichungen im Quartal errechnet Anzahl : Trendwert = Abweichung (weil Quotient genommen darum multiplikativ!!)
5) Abweichungen können gemittelt werden

Ft = a * x + b (TR: LinReg)

$F_t = 0.0391 * t + 1.2912$

Prognose

f(15) = 1.88

Trendwert * durchschn. Quotient = Prognosewert

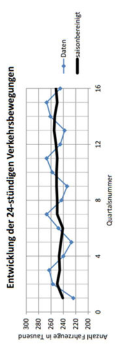
Jahr	Quartal	Nr.	Trendwert (in Mio. CHF)	*durchschnittlicher* Quotient	Prognosewert (in Mio. CHF)
4	III	15	1.88	0.632	1.19
4	IV	16	1.92	1.558	2.99

Gleich wie bei additiv. Durch Funktionsgleichung für andere Nr. (t) die Trendwerte ausrechnen. Als durchschnittlicher Quotient nehme ich die wie oben berrechnet (achtung richtiges Quartal nehmen). Und erhalte Prognose.

Saisonbereinigung:

additiv

Saisonfigur:
blau = Y
schwarz = Yber



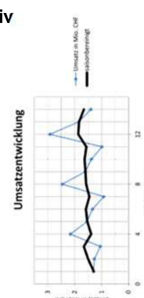
Jahr	Quartal	Nr.	Anzahl (in Tausend) Y	durchschn. Abweichung (in Tausend) S	Anzahl saisonbereinigt (in Tausend) Y_{ber}	Veränderung in % saisonbereinigt
2005	I	1	224	-17	241	...
	II	2	257	7	250	3.7
	III	3	263	17	246	-1.6
	IV	4	240	-7	247	0.4
2006	I	5	227	-17	244	-1.2
	II	6	248	7	241	-1.2
	III	7	267	17	250	3.7
	IV	8	243	-7	250	0.0
2007	I	9	234	-17	251	0.4
	II	10	258	7	251	0.0
	III	11	267	17	250	-0.4
	IV	12	245	-7	252	0.8
2008	I	13	238	-17	255	1.2
	II	14	261	7	254	-0.4
	III	15	267	17	250	-1.6
	IV	16	245	-7	252	0.8

$$Y_{ber,t} = Y_t - \bar{S}_t$$

Ich rechne die Anzahl minus (-) die durchschnittliche Abweichung und erhalte die Anzahl saisonbereinigt

multiplikativ

Saisonfigur:
blau = Y
schwarz = Yber



Jahr	Quartal	Nr.	Umsatz (in Mio. CHF) Y	durchschn. Abweichung S	Umsatz saisonbereinigt (in Mio. CHF) Y_{ber}	Veränderung saisonbereinigt
1	I	1	1.30	1.005	1.29	...
	II	2	1.27	0.845	1.50	16%
	III	3	1.05	0.632	1.66	11%
	IV	4	2.16	1.558	1.39	-17%
2	I	5	1.54	1.005	1.53	11%
	II	6	1.34	0.845	1.59	4%*
	III	7	0.92	0.632	1.46	-8%
	IV	8	2.46	1.558	1.58	8%
3	I	9	1.61	1.005	1.60	1%
	II	10	1.38	0.845	1.63	2%
	III	11	0.99	0.632	1.57	-4%
	IV	12	2.91	1.558	1.87	19%
4	I	13	1.85	1.005	1.84	-1%
	II	14	1.40	0.845	1.66	-10%

$$Y_{ber,t} = Y_t / \bar{S}_t$$

Ich rechne die Anzahl (Umsatz) geteilt durch (:) die durchschnittliche Abweichung und erhalte die Anzahl saisonbereinigt