Eindimensionale Häufigkeitsverteilung

Allg.

a_i Merkmalsausprägung

x_j Stichprobenwerte

n_i absolute Häufigkeit

f_i relative Häufigkeit

k Anzahl Merkmalsausprägungen

n Anzahl aller Werte

m_j KlassenMitte

N_i kumulierte absolute Häufigkeit („Wie viele sind … und jünger")

F_i kumulierte relative Häufigkeit

$$\sum_{i=1}^{6} x_i$$

obere Summationsgrenze

untere Summationsgrenze

Index

$$\sum_{\text{von}}^{\text{bis}} \text{über}$$

Fechnersche Lageregel

2) $\bar{x} = \tilde{x} = x_{mod}$
symmetrische Verteilung

3) $x_{mod} < \tilde{x} < \bar{x}$
rechtsschiefe Verteilung

1) $\bar{x} < \tilde{x} < x_{mod}$
linksschiefe Verteilung

statistische Maßzahlen

Lagemaße

arithmetisches Mittel $\bar{x}$

1. n Stichprobewerten x_j

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$$

2. a_i Merkmalsausprägungen, n_i absolute Häufigkeiten

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} a_i n_i$$

3. a_i Merkmalsausprägungen, f_i relative Häufigkeiten

$$\bar{x} = \sum_{i=1}^{k} a_i f_i$$

! Ausreißer (nicht robust)

klassierte Daten:

$$\boxed{\bar{x} = \frac{1}{n}\sum_{j=1}^{k} m_j \cdot n_j = \sum_{j=1}^{k} m_j \cdot f_j}$$

Median $\tilde{x}$

Ungerade Anzahl Stichprobenwerte

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)}$$

Gerade Anzahl Stichprobenwerte

$$\tilde{x} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right)$$

Verwendung bei:

- sehr kleinem n
- Asymmetrisch
- bei Ausreißern

Modus x_{mod}

höchste Häufigkeitsdichte

geometrisches Mittel

$$x_g = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdots x_n}$$

Streuungsparameter

Spannweite R

Differenz: größte und kleinster Merkmalsausprägung

$$R = x_{(n)} - x_1$$

Median von MAD

Median absoluter Abweichungen vom Median

$\Rightarrow |x_1 - \tilde{x}|, |x_2 - \tilde{x}|, \ldots, |x_n - \tilde{x}|$

ausrechnen + ordnen, dann Median

Bsp. siehe S. 29

Mittlere lineare (absolute) Abweichung MAD

Mittlere Abweichung von Mittelwert a
↳ z.B. $\bar{x}$ oder $\tilde{x}$

$$d_a = \frac{1}{n}\sum_{j=1}^{n} |x_j - a|$$

mit Merkmalsausprägungen:

$$d_a = \frac{1}{n}\sum_{i=1}^{k} |a_i - a| \cdot n_i$$

Varianz und Standardabweichung

Varianz

1. n Stichprobenwerten x_j

$$s^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})^2 = \frac{1}{n}\sum_{j=1}^{n} x_j^2 - \bar{x}^2$$

siehe Bsp. Blatt 31

2. a_i Merkmalsausprägungen, n_i absolute Häufigkeiten

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(a_i-\bar{x})^2 \cdot n_i = \frac{1}{n}\sum_{i=1}^{k} a_i^2 n_i - \bar{x}^2$$

siehe Bsp. Blatt 32/A4

3. a_i Merkmalsausprägungen, f_i relative Häufigkeiten

$$s^2 = \sum_{i=1}^{k}(a_i-\bar{x})^2 \cdot f_i = \sum_{i=1}^{k} a_i^2 f_i - \bar{x}^2$$

siehe Blatt 32/A3

$\bar{x}^2$ EINMAL am Ende abziehen

Standardabweichung

$$s = \sqrt{s^2}$$

Variantskoeffizient

$$V = \frac{s}{\bar{x}}$$

Themen Klausur 2

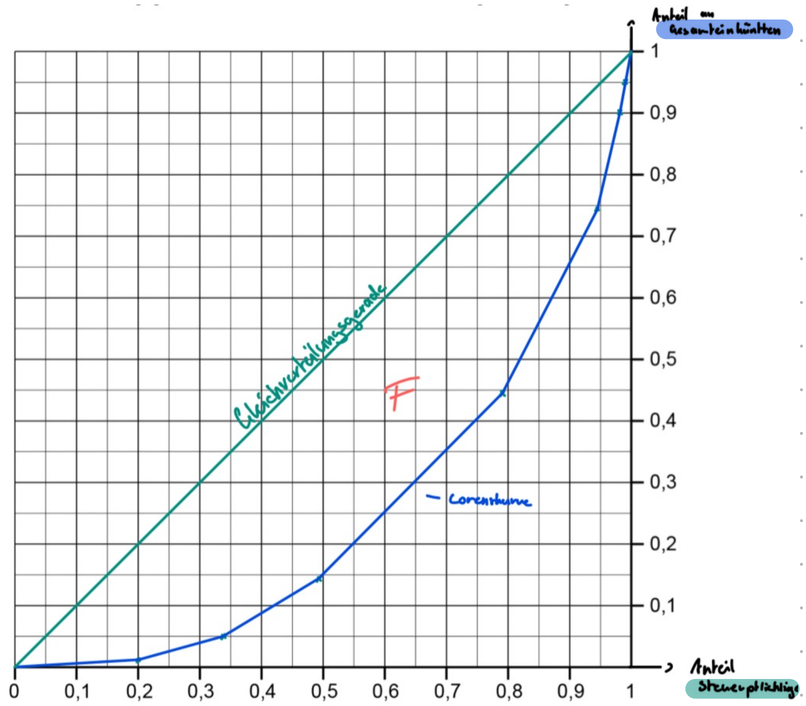
1. Quantile und ihre Bedeutung
2. Lorenzkurve und ihre Bedeutung
Zweidimensionale Häufigkeitsverteilung
3. Ab- und Unabhängigkeit bei ?
4. Parameter zweidimensionaler Verteilungen -> $\bar{x}$ ... $n_{ij} \sim n_{i.}$ ...
5. Kovarianz

Lorenzkurve

beschreibt (grafisch) wie sich die Merkmalssumme, Summe aller Beobachtungswerte, auf die einzelnen Beobachtungswerte auswirkt

=> in einem Diagramm wird die kum. Anteile der Merkmalssumme gegen die kum. Anteile der Beobachtungen gestellt

Bsp:



F: Fläche zwischen Normalverteilung und Lorenzkurve drückt Konzentration aus
je größer desto konzentrierter

Quantile

Ein x% Quantil wird von x% der Werte unterschritten und von 100% − x% überschritten
Bsp: 1. Quartil (25% Quantil) 25% drunter 75% drüber

Quartil: unteres (25%), oberes (75%) } zwischen den Quartilen liegen 50% der Werte
=> daraus: Quartilsabstand: $\tilde{x}_{75\%} - \hat{x}_{25\%}$

$$\tilde{x}_a = \begin{cases} x_{([na]+1)} & \text{falls } n \cdot a \text{ keine natürliche Zahl} \\ \frac{1}{2}\left(x_{(na)} + x_{(na+1)}\right) & \text{falls } na \text{ natürliche Zahl} \end{cases}$$

→ x% // [na] => Komma abschneiden [7,5] → 7

Bsp: 10 Werte : unteres Quartil
↳ 25%
=> $n \cdot a = 10 \cdot 0{,}25 = 2{,}5$ => k. natürliche Zahl
=> $x_{([2,5]+1)} = x_{2+1} = x_3$
=> $\tilde{x}_{25\%} = x_3$

zweidimensionale Häufigkeitsverteilung (2dHv)

allg. Kontingenztafel

$x \backslash y$	y_1	y_j	…	y_l	Zeilensumme
x_1	x_{11}	x_{1j}		x_{1l}	$n_{1\cdot}$
x_2 ⋮					
x_i	x_{i1}	x_{ij}		x_{il}	$n_{i\cdot}$
⋮					
x_m	x_{m1}	x_{mj}		x_{ml}	n_{m}
Spaltensumme	$n_{\cdot 1}$	$n_{\cdot j}$		$n_{\cdot l}$	n

(Un)Abhängigkeit bei 2dHv

Zwei Merkmale sind X und Y sind genau dann unabhängig verteilt, wenn alle bedingten Verteilungen von X bzw Y mit den jeweiligen Randverteilungen übereinstimmen

	Raucher	Gelegenheitsraucher	Nichtraucher	Summe
Weiblich	10 $\frac{10}{40}$ 0,25	10 0,25	20 0,5	40 1 (Basis)
Männlich	15 $\frac{15}{60}$ 0,25	15 0,25	30 0,5	60 1 (Basis)
Summe	25 0,25	25 0,25	50 0,5	100

	Raucher	Gelegenheitsraucher	Nichtraucher
Weiblich	$\frac{4}{16}$ = 0,25	0,4	0,4375
Männlich	$\frac{12}{16}$ = 0,75	0,6	0,5625
Summe	1	1	1

→ Geschlecht abhängig von Rauchen?
=> falsche „Richtung"

Vorgehen:
1. relative Häufigkeit (variables x festes y)
2. relative Häufigkeit (variables y festes x)

siehe S. 58

3. Randverteilungen gleich? // Spalte/Zeile gleiche Häufigkeit
(4. Kausalität => welche Richtung Abhängig)

Parameter zweidimensionaler Verteilungen

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

$$\bar{y} = \bar{y} = \frac{1}{n} \sum_{j=1}^{l} y_j n_{\cdot j}$$

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{m} x_i^2 \cdot n_{i\cdot} - \bar{x}^2$$

$$s_y^2 = \frac{1}{n} \sum_{j=1}^{l} y_j^2 \cdot n_{\cdot j} - \bar{y}^2$$

$l = 3$, $m = 3$

X Y	$y_1 = 2$	$y_2 = 3$	$y_3 = 4$	
$x_1 = 2$	1	1	0	2 $n_{1\cdot}$
$x_2 = 3$	2	3	1	6 $n_{2\cdot}$
$x_3 = 4$	0	0	2	2
	3 $n_{\cdot 1}$	4 $n_{\cdot 2}$	3	10 = n

$$s = \sqrt{s^2}$$

Kovariane

Formel

$$\mathrm{COV}(X,Y)=\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{l}(x_i-\overline{x})(y_j-\overline{y})\cdot n_{ij}=\frac{1}{n}\overbrace{\sum_{i=1}^{m}\underbrace{\sum_{j=1}^{l}x_iy_jn_{ij}}_{\text{for-schleife } i=1}}^{\text{for-schleife } i\le m}-\overline{x}\cdot\overline{y}$$ Kontingenztabelle

$$\mathrm{COV}(X,Y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\overline{x}\cdot\overline{y}$$ keine Kontingenztabelle

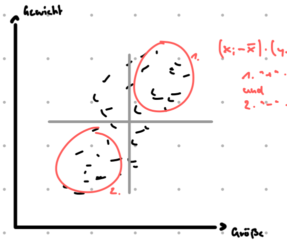
X Y	$y_1 = 2$	$y_2 = 3$	$y_3 = 4$	
$x_1 = 2$	1	1	0	2
$x_2 = 3$	2	3	1	6
$x_3 = 4$	0	0	2	2
	3	4	3	10

Zusammenhänge von Merkmalen können mit dem Begriff der Variation beschrieben werden.

=> **Variation:**

gleichläufige Variation: je mehr desto mehr

Werden die Ausprägungen des einen Merkmals größer, nehmen im Allgemeinen auch die des anderen Merkmals zu.

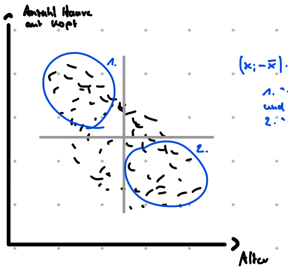


$(x_i-\overline{x})\cdot(y_j-\overline{y})$ positiv, da

1. "+" · "+" = +
und
2. "−" · "−" = +

gegenläufige Variation: je mehr desto weniger

Werden die Ausprägungen des einen Merkmals größer, nehmen im Allgemeinen die des anderen kleinere Werte an.



$(x_i-\overline{x})\cdot(y_j-\overline{y})$ negativ, da

1. "−" · "+" = −
und
2. "+" · "−" = −

weder noch: $\mathrm{cov}(x,y) = 0$ oder sehr klein

Maßstahl für die Variation: Kovariant

Die Kovarianz ist ein Parameter, der angibt, inwieweit die beiden Untersuchungsmerkmale gleichläufig oder gegenläufig variieren

SkriptNr
SeitenNr

53 Begriffe
58+59 61+62 Übungen lineare Regression
63+73 − 68 Korrelation
69 − 72 nichtlineare Regression
79 − 88 Zeitreihen 3er, 4er Mittel unterschied wissen
96 − 104 operations Research

Aufgabe Moodle

Regression

Einführung

1) Welche Stärke hat der Zusammenhang zwischen beiden Merkmalen?

ganz stark → funktionaler Zusammenhang
gar nicht → alle Ausprägungen möglich

Eine Maßzahl für diese Stärke zu bestimmen, ist Gegenstand der Korrelationsanalyse

Die Maßzahl nennt man Korrelationskoeffizient

2) Welche Form hat das Zusammenwirken der Merkmale?

Kann der Zusammenhang durch eine durchschnittliche Funktion beschrieben werden?

Bestimmung derartiger Funktionen ist Gegenstand der Regressionsanalyse

Die Regressions- bzw. Korrelationsanalyse gibt rein formal eine zahlenmäßige Auskunft über Art und Grad des Zusammenhangs zwischen zwei Variablen. Sie sagt aber nichts aus über die Kausalität dieses Zusammenhangs. Deshalb muss das abschließende Urteil von den Fachwissenschaften gefüllt werden.

Lineare Regression

1. Regressionsgerade

$y = a + bx$ Formelsammlung

$a = \bar{y} - b\bar{x}$; $b = \frac{cov(x,y)}{s_x^2}$

Korrelation

Pearson Korrelationskoeffizient (S. 73)

Bezeichnung r

$r = \frac{cov(x,y)}{s_x \cdot s_y}$ Formelsammlung

verbale Beschreibung der Größe:

- 0 keine Korrelation
- $0 < |r| < 0,2$ sehr geringe Korrelation
- $0,2 < |r| < 0,5$ geringe Korrelation
- $0,5 < |r| < 0,7$ mittlere Korrelation
- $0,7 < |r| < 0,9$ hohe Korrelation
- $0,9 < |r| < 1$ sehr hohe Korrelation

auswendig

Spearmanscher Korrelationskoeffizient (S. 76)

$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}$

Rangzuordnung: wenn 2 oder mehr gleich dann Mittleren Rang für alle
→ Rang 5+6 → 5,5

! Ränge vergeben, dann $\underbrace{\text{Differenzen der Ränge}}_{d_i}$ berechnen

nichtlineare Regression

Wiederholung Logarithmus

$\log_{10}(100) = 2$ „10 hoch was ist 100"
$\log_2(8) = 3$

Der Logarithmus x einer Zahl a zur Basis b
$\log_b(a) = x$ ist die Zahl für die gilt
$b^x = a$

$\log_2(2^3) = 3$

$\log_a(u \cdot v) = \log_a(u) + \log_a(v)$
$\log_a(\frac{u}{v}) = \log_a(u) - \log_a(v)$
$\log_a(u^k) = k \cdot \log_a(u)$

Vorgehen:

1. Allgemeine Form der homogenen Potenzfunktion
   $y = a \cdot x^b$ | vorgegeben

je nach Funktion anders

2. Logarithmieren der Funktion $\log(x^b) = b \cdot \log x$

   $\log y = \log a + b \cdot \log x$
   $y^* = a^* + b \cdot x^*$

3. Die lineare Regressionsfunktion

   ↳ Tabelle $x_i^* | y_i^* | x_i^* \cdot y_i^*$
   ↳ a^*, b herausfinden
   ↳ Funktion $y = a + bx$

4. zurückwandeln

   $a = 10^{a^*}$ wegen $\log_{10}(a)$

5. $y = a \cdot x^b$

$y = a b^x$
↓
$\log y = \log a + x \cdot \log b$
$y^* = a^* + x \cdot b^*$
$\downarrow a = 10^{a^*}$
$y = 10^{a^*} + (10^{b^*})^x$

$y = a \cdot e^{bx}$
↓
$\ln y = \ln a + b \cdot x$
$y^* = a^* + b \cdot x$
$\downarrow a = e^{a^*}$
$y = e^{a^*} + e^{b \cdot x}$

Zeitreihen

6.1 Grundbegriffe

$y = f(t)$

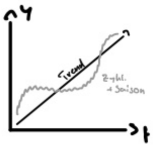
6.1.1 Die Komponenten einer Zeitreihe:

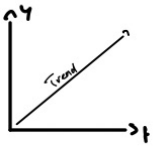
1. **Der Trend** gibt die langfristige Entwicklung wieder
   unbeeinflusst von Schwankungen
   Soll durch möglichst einfache mathematische Funktion dargestellt werden

2. **Die zyklische Komponente**
   gibt mittelfristige konjunkturelle Einflüsse wieder
   (z.B. Schweinezyklus, Apple entl.)
   Stichworte: Aufschwung, Depression.
   Problematisch ist die zuverlässige Herausrechnung zyklischer Komponente.

3. **Die Saisonkomponente**
   gibt jahreszeitliche (z.B. Skiumsatz, Grippe),
   monatliche (z.B. Kontostand),
   wöchentliche (z.B. Alkoholabsatz),
   tägliche (z.B. Strom, Wasserverbrauch) wieder

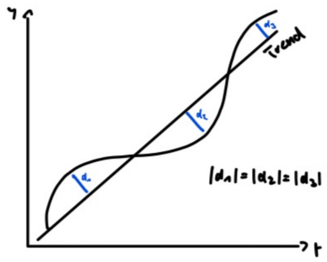
4. **Die Restkomponente**
   beinhaltet einmalige Einflüsse, z.B. technische Neuerungen, Krieg, Naturkatastrophen, Pandemien





6.1.2 Die Verknüpfung der Komponenten

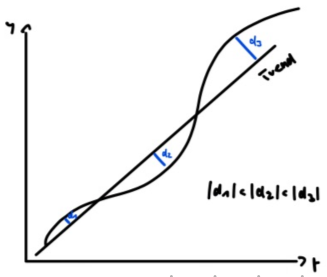
Additive Verbundenheit



Eine additive Verknüpfung von Trend + Saison liegt vor, wenn die Saison nicht von der Größe des Trends abhängt.
Schwankungen immer gleich groß

$y = T(t) + Z(t) + S(t) + R(t)$

Multiplikative Verbundenheit



Man kann bei ökonomischen Reihen davon ausgehen, dass insbesondere Konjunktur- und Saisonausschläge bei steigendem Niveau der Zeitreihen zunehmen.

$y = T(t) \cdot Z(t) \cdot S(t) \cdot R(t)$

Trendbestimmung

Methode gleitende Mittelwerte

Bei Saisons:
Beim gleitenden 3er Durchschnitt werden mögliche Saisonbewegungen ausgeglichen.
Alle Saisonperioden werden miteinberechnet.

5er-Durchschnitt:
Prinzipiell wird der ermittelte Trend gleichmäßiger je größer der gleitende Durchschnitt
Aber, hier kommt es nicht zum Saisonausgleich, da 2 Drittelfahre doppelt einberechnet werden. 5er → kein Saisonausgleich 6er → Saisonausgleich

! gerade (4er, ...) Durchschnitte:
Um der Mitte einen Wert zuordnen zu können und alle Saisons mit einzuberechnen, nimmt man 5 Werte (einen mehr) und halbiert den ersten und letzten Wert
[siehe auch Formel]
↳ t_0 Mitte
$3+4 : m=1$
$5+6 : m=2$
⋮

Nachteile:
1) Die Methode führt nicht zu einer mathematisch fassbaren Funktion.
2) Die Reihen werden bei höheren m immer kürzer.

Vorteile
- Die Kurve glättet hervorragend Verläufe und passt sich gut an
- Je höher das m desto stärker wird geglättet
- konkrete Verwendung bei Aktienkursen (z.B. Vergleich 200er und 50er)
- Saisonbereinigung gut möglich

Methode der kleinsten Quadrate

Anwendung der Regressionsrechnung auf Zeitreihen
Trendlinie ist Funktion, die von der Zeit abhängt

y_t abhängige Variable
t unabhängige Variable [siehe FSammlung]

$y_t = a + b \cdot t$ Trendgerade

$b = \frac{cov(t, y_t)}{s_t^2}$ $a = \bar{y}_t - b\bar{t}$

→ lineare Reg mit t statt x

Saisonindizes

geben an um wie viel die Werte einer Zeitreihe über oder unter einem Durchschnittswert liegen.

Werte werden auf Normalwert bezogen (z.B. Jahresdurchschnitt), dieser wird gleich 100 gesetzt, Abweichung davon ist der Saisonindex

$$S_{i.j} = \frac{U_{i.j}}{\bar{x}_i}$$

z.B. QuartalU

z.B. JahresU

$$\bar{S}_{i.j} = \frac{1}{n}\sum_{i=1}^{n} S_{i.j}$$

im jten Quartal beträgt der Umsatz nur ... % des durchschnittlichen QUmsatzes des iten Jahres

Prognosen mit Saisonindizes

Dazu berechnet man den durchschnittlichen Saisonindex für den Unterzeitraum, z.B. für jedes Quartal.

$$\bar{S}_j = \frac{1}{n}\sum_{i=1}^{m} S_{i.j} \quad \text{z.B. } S_1 = \frac{1}{2}(66,7 + 71,4)$$

Wenn man ausschließlich den Wert $\bar{x}_i$ der Trendfunktion für das Jahr i mit S_j multipliziert, erhält man den geschätzten Beobachtungswert $\hat{x}$ → Dach

$$\hat{x}_{i.j} = \bar{x}_i \cdot \frac{\bar{S}_j}{100} \Rightarrow \hat{U}_{i.j} = \bar{x}_i \cdot \frac{\bar{S}_j}{100}$$

Prognostizieren von $\bar{x}_i$:

$T_g: x = a + bt \Rightarrow \hat{x}_6 = a + b \cdot 6$

$a = \bar{x}_t - b\bar{t}$

$$b = \frac{cov(t, x_t)}{s^2_t}$$

Operations Research

1.1 Definition

Unter Operations Research versteht man die Anwendung von wissenschaftlichen Erkenntnissen auf das Problem der Entscheidungsfindung in der Unsicherheits- oder Risikosituation, mit dem Ziel, den Entscheidungsträgern bei der Suche nach optimalen Lösungen eine quantitative Basis zu liefern.

Prinzipielle Arbeitsweise bei OR-Verfahren:

Reales Problem

| Abstraktion

mathematisches Modell

| Rechnung

Modell-Lösung

| Interpretation

Lösungsvorschlag für reales Problem

Ein Modell als gedankliches Hilfsmittel dient zur Abstraktion und Vereinfachung des Realproblems, da eine detaillierte Beschreibung der Realität mit allen Ursachen und Zusammenhängen wegen ihrer Komplexität nicht möglich ist. Im Modell müssen wesentliche Eigenschaften und Beziehungen der Realität erhalten sein.

Lineare Optimierung

Lösungsmethode für die lineare Optimierung

a) Nebenbedingungen aus Aufgabenstellung aufstellen → umformen und zulässigen Lösungsbereich einzeichnen.

b) Durch Umstellung der Gleichung der Zielfunktion $G = 500x_1 + 800x_2$ erhält man

$$x_2 = -\frac{5}{8}x_1 + \frac{1}{800}G$$

eine Geradenschar mit G als Parameter

$\frac{1}{800}G$ ist Achsenabschnitt

c) Man erreicht den größtmöglichsten Wert von $\frac{1}{800}G$ (und damit auch von G), wenn man eine Gerade der Geradenschar auswählt, einzeichnet und so lange parallel nach oben verschiebt, bis sie den Rand des Lösungsbereiches oder eine Ecke erreicht.

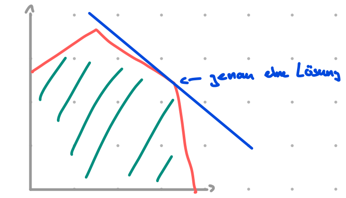
Der Punkt ist dann die optimale Lösung

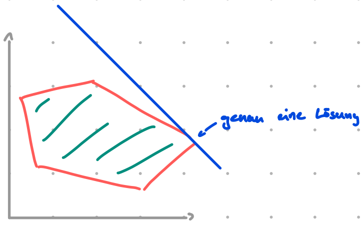
Lineare Optimierung

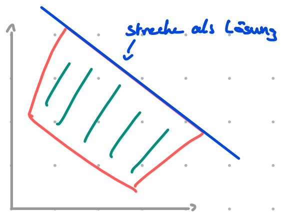
Optimierung, d.h. Maximierung oder Minimierung einer linearen Funktion, der sog. Zielfunktion, deren Variablen einem System von linearen Ungleichungen, den sog. Nebenbedingungen genügen müssen.

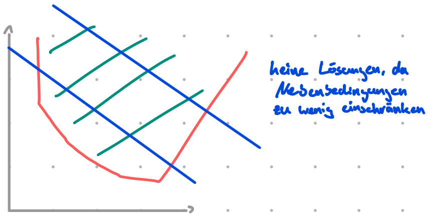
Lösungsmöglichkeiten bei Maximum- bzw. Minimum-Optimierung

Maximum-Optimierung

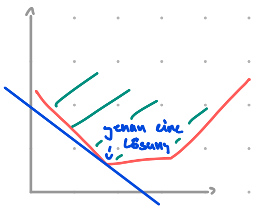


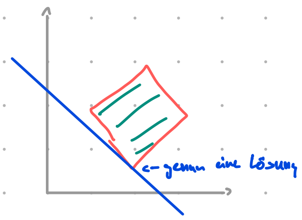


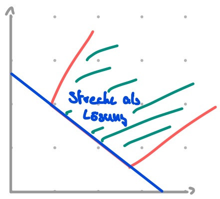


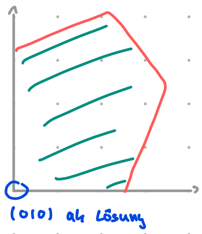


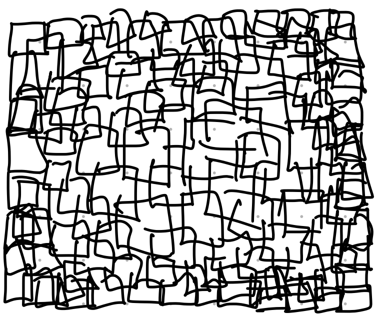
Minimum-Optimierung











- Rechnerische Lösung von Problemen der linearen Optimierung S.105 + 110, 112 (115), 134
    - Maximierungs- und Minimierungsprobleme
    - Idee des Lösungsverfahren erläutern können / vergleichen Sie dazu den ausgeteilten Text aus Runzheimer, der auch als Scan im Kapitel 02 Lineare Optimierung zu finden ist
    - Steepest Unit Ascent Kriterium zur Auswahl des Pivotelements
    - Greatest Change-Verfahren zur Auswahl des Pivotelements } S.117
    - Zwei-Phasen-Verfahren S.119
    - Dual-Simplexverfahren S.135
- Transportprobleme S.138
    - Möglichkeiten zur Auswahl eines Starttableaus (Nord-West-Ecken-Verfahren, aufsteigende Kostenwerte, Vogelsche Approximationsmethode) S.140, 141
    - Optimierung mit dem Stepping Stone-Verfahren S.141
    - Anmerkung: https://chat.openai.com/ liefert auf die Fragen "Was ist die Vogelsche Approximationsmethode" und "Was ist das Stepping Stone Verfahren beim Transportproblemen" tolle Erläuterungen.
    - Bitte lassen Sie sich aber NICHT über ChatGpt ein Beispiel vorrechnen, da kam bei mir gerade totaler Unsinn raus.

Rechnerische Lösung von LO (Lineare Optimierungs) – Problemen

Da nach den Simplex-Theorien eine optimale Lösung – falls eine existiert – in einem Eckpunkt des zulässigen Lösungsbereich liegt, wäre es prinzipiell möglich, alle Schnittpunkte aller Geraden zu berechnen, mit diesen Wertpaaren die Zielfunktion zu prüfen und die optimale Lösung auszuwählen.

Problem: relativ aufwendig
nicht alle Eckpunkte im Lösungsbereich

=> Lösungsverfahren (Iterationsverfahren) das nur die Eckpunkte überprüft und nach Finden der optimalen Lösung abbricht.

Simplex-Algorithmus

→ Jede Ungleichung der Nebenbedingung kann durch Einfügen einer zusätzlichen Variable (sog. Schlupfvariablen) in eine Gleichung überführt werden.

Bsp:

Struktur- oder Hauptvariablen | Schlupfvariablen

1) $3x_1 + 2x_2 + s_1 = 1200$
2) $5x_1 + 10x_2 + s_2 = 3000$
3) $\frac{1}{2}x_1 + s_3 = 125$

3 Nebenbedingungen
3 Schlupfvariablen

s_1, s_2, s_3 dürfen nicht negativ sein

Das Gleichungssystem besteht jetzt aus drei Gleichungen und 5 Variablen
=> Nur Eckpunkte sind von Interesse
Eckpunkte zeichnen sich dadurch aus, dass zwei Variablen 0 sind

Vorgehen Simplex (Steepest Ascent Unit):

1. Simplexkriterium *Wahl Pivotspalte*
   Jetzt wird diejenige nicht Basisvariable (NBV), die in der Zielfunktion den größten positiven Koeffizienten hat, als Basisvariable angenommen
   x_2 wird Basisvariable

   → 1. unten größte

	-	-	-	-
	-	-	-	-
	1	8	0	-

→ Tabelle größer, da Schlupfvariablen

2. Simplexkriterium *Wahl Pivotzeile*
   Welche bisherige Basisvariable (s_1, s_2, s_3) wird im Austausch dafür Nichtbasisvariable?
   Kriterium: Wähle diejenige Variable als aus der Basis austretende Variable aus, für die der nicht negative Quotient aus rechte Seite (RS) und den Koeffizienten minimal wird.
   (Anschaulich: Man darf nicht aus Lösungsbereich herausklappen)

   → 2. rechts kleinste (nicht negativ)

Pivotelement

	1	3	4	1
	2	6	9	8
	1	8	0	7

3. Pivotelement auf 1 bringen (hier: 3)
   ↳ ganze Zeile durch ← teilen →

	$\frac{1}{3}$	1	$\frac{4}{3}$	$\frac{1}{3}$
	2	6	9	8
	1	8	0	7

4. oberhalb und unterhalb von Pivot soll '0' stehen
   => umformen

   ↳

	$\frac{1}{3}$	1	$\frac{4}{3}$	$\frac{1}{3}$
	-	0	-	-
	-	0	-	-

6 zu 0 machen:
$6 + (-6) \cdot 1 = 0$
$\underbrace{11 + (-6) \cdot 1}$
auf ganze Zeile anwenden
z.B. $2 + (-6) \cdot \frac{1}{3} = \ldots$
$8 + (-8) \cdot 1 = 0$
$111 + (-8) \cdot 1$

5. wiederholen bis untere Spalte (-G) nur negative

6. optimale Lösung: linke Spalte = rechte Spalte
   => z.B. $x_1 = 2$
   $x_2 = 5$
   $G = 100$

x_1				2
x_2				5
$-G$		-8	-5	-100

! nicht vergessen:
Anfang $x_1 = 0 \wedge x_2 = 0$
nach Tausch
hier: $x_1 = 0 \wedge s_1 = 0$

	x_1	x_2	s_1	—
s_1		?		
s_2				
$-G$				

x_2			
s_2	/	/	/
$-G$			

Vorgehen (Greatest-Change)

→ Zielfunktionskoeffizienten → untere Zeile

1. Für jede Spalte mit positiven ZFK Quotient bilden
2. Jeweils kleinsten nehmen
3. Jeweils Quotienten mit ZFK multiplizieren
4. Größeren nehmen
5. Zeile + Spalte des größeren = Pivotelement (hier x_1, zeile s_2)
6. umformen
7. wenn nur 1 positiv, wie gehabt Steepest-Ascent

BV	x_1	x_2	s_1	s_2	s_3	RS	Quotient x_1-Spalte	Quotient x_2-Spalte	Greatest Change
s_1	1	2	1	0	0	170	170	85	
s_2	1	1	0	1	0	150	150	150	$150 \cdot 300 = 45\,000$
s_3	0	3	0	0	1	150	/	60	$60 \cdot 500 = 30\,000$
$-G$	300	500	0	0	0	36000			

Unzulässige Ausgangslösung → 2-Phasen

Phase 1: Erreichen des zulässigen Lösungsbereiches, da der Startwert nicht im Lösungsbereich ist. Die negativen Einträge auf der rechten Seite entfernen, indem man negative Pivotelemente nimmt

→ negative Pivotelemente (auf 1 bringen) bis RS nicht mehr negativ
<> P1 abgeschlossen

Phase 2: Optimierungsphase → wie üblich

Lösung von Minimierungsproblem

Minimieren Sie $K = 20x_1 + 40x_2$

unter folgenden Nebenbedingungen

1) $x_1 \geq 2$
2) $x_1 + x_2 \geq 3$
3) $3x_2 \geq 2$

u_1 1) $1x_1 + 0x_2 \geq 2$
u_2 2) $1x_1 + 1x_2 \geq 3$
u_3 3) $0x_1 + 3x_2 \geq 2$
$20x_1 + 40x_2$

1) Minimierung mit dem Zwei-Phasenverfahren

K minimieren heißt
$-K = G$ maximieren unter NB
1) $-x_1 \leq -2$
2) $-x_1 - x_2 \leq -3$
3) $-3x_2 \leq -2$

I Tableau $x_1 = 0 \wedge x_2 = 0$

	x_1	x_2	s_1	s_2	s_3	RS
s_1	-1	0	1	0	0	-2
s_2	-1	-1	0	1	0	-3
s_3	0	-3	0	0	1	-2
$-G$	-20	-40	0	0	0	0

II T $x_2 = 0 \wedge s_2 = 0$

	x_1	x_2	s_1	s_2	s_3	RS
s_1	0	1	1	-1	0	1
x_1	1	1	0	-1	0	3
s_3	0	-3	0	0	1	-2
$-G$	0	-20	0	-20	0	60

III T $s_2 = 0 \wedge s_3 = 0$

	x_1	x_2	s_1	s_2	s_3	RS
s_1	0	0	1	-1	$\frac{1}{3}$	$\frac{1}{3}$
x_1	1	0	0	-1	$\frac{1}{3}$	$2\frac{1}{3}$
x_2	0	1	0	0	$-\frac{1}{3}$	$\frac{2}{3}$
$-G$	0	0	0	-20	$-\frac{20}{3}$	$73\frac{1}{3}$

Phase 1 abgeschlossen
alle positiv

Phase 2 abgeschlossen
alle negativ
keine optimierung möglich

Optimale Lösung:
$x_1 = 2\frac{1}{3}$
$x_2 = \frac{2}{3}$
mit $K = 73\frac{1}{3}$

2) Minimierung mit dem Dual-Simplex-Verfahren

I $u_1 = 0 \wedge u_2 = 0 \wedge u_3 = 0$

	u_1	u_2	u_3	s_1	s_2	RS	Quotient
s_1	1	1 (PE)	0	1	0	20	20 ← 2. Simplex
s_2	0	1	3	0	1	40	40
$-G$	2	3	2	0	0	0	

↑ 1. Simplex

II $u_1 = 0 \wedge u_3 = 0 \wedge s_1 = 0$

	u_1	u_2	u_3	s_1	s_2	RS
u_2	1	1	0	1	0	20
s_2	-1	0	3	-1	1	20 ←
$-G$	-1	0	2	-3	0	-60

III $u_1 = 0 \wedge s_1 = 0 \wedge s_2 = 0$

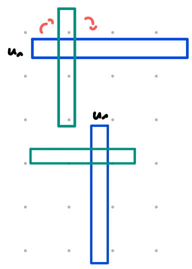
	u_1	u_2	u_3	s_1	s_2	RS
u_2	1	1	0	1	0	20
u_3	$-\frac{1}{3}$	0	1	$-\frac{1}{3}$	$\frac{1}{3}$	$\frac{20}{3}$
$-G$	$-\frac{1}{3}$	0	0	$\underbrace{-2\frac{1}{3}}_{\hat{=}-x_1}$	$\underbrace{-\frac{2}{3}}_{\hat{=}-x_2}$	$\underbrace{-73\frac{1}{3}}_{\hat{=}-K}$

optimale Lösung
$x_1 = 2\frac{1}{3}$
$x_2 = \frac{2}{3}$
$K = 73\frac{1}{3}$

=> Schlupfvariable = Lösung

Zusatz:

Nebenbedingungen:



$$\Rightarrow \begin{array}{l} 4 + 2 + 3 \geq 2 \\ 3 + 2 + 7 \geq 3 \\ 600 + 320 + 840 = K \end{array}$$

	u_1	u_2	s_1	s_2	s_3	RS
s_1	4	3	pro NB eine Schlupfvariable			600
s_2	2	2				320
s_3	3	7				840
$-G$	2	3				0

Dual-Simplex-Verfahren. Nach dem Dualitätssatz lassen sich aus dem Optimaltableau des einen Modells die optimalen Lösungen des anderen Modells ablesen

Die Werte der Strukturvariablen aus dem primalen Modell sind gleich den Werten der Schlupfvariablen in der Zielfunktion des dualen Modells mit umgekehrten Vorzeichen

oben s_1, s_2
=> unterste Zeile
$-x_1, -x_2$

Transportprobleme

Starttableau

Nord-West-Ecken-Verfahren

- Von oben links bis unten rechts
- immer alles was geht in aktuelles Feld

Reihenfolge aufsteigender Kostenwerte

- Kosten pro Feld eintragen
- niedrigster beginnen (alles was geht)
- nächst niedriger
  ⋮

Vogel'sche Approximationsmethode

- Kostendifferenz zweitgünstigster - günstigster (Zeile + Spalte)
- größte nehmen
- günstigere Kosten auswählen
- alles was geht
- wiederholen

Optimierung

Stepping-Stone

- Variable die 0 ist erhöhen, sodass eine andere zu 0 wird
- immer rechteckig / gerade Linien + nicht über Null-Felder
- bis nicht mehr möglich bzw. alle möglichen Akkumulationen positiv -> teurer
  ↳ $\underbrace{+ \text{startKosten} - xK + yK - zK \ldots}_{\text{immer abwechselnd } +/-}$

  => opt. Lösung erreicht

Wie / wie viel „mitnehmen"?

↳ - auf Max A/N schauen

- wie viel kann pro Feld verändert werden?
  ↳ kleinsten Wert nehmen (":")
- alle "+"-Felder x+i
  alle "-"-Felder x-i

von \ nach	B_1	B_2	B_3	B_4	Angebot
A_1	+5/+2 – (90) +2	-2 2 (150) -2 max	7 (40)	– (80)	9
A_2	5 (30) -2	+2 1 (80) -5 -> +5 ? +2 max	– (70)	4 (10)	10
A_3	– (20)	da 3 (40)	– (20)	– (60)	3
Nachfrage	5	6	7	4	

↳

von \ nach	B_1	B_2	B_3	B_4	Angebot
A_1	2 (90)	– (150)	7 (40)	– (80)	9
A_2	3 (30)	3 (80)	– (70)	4 (10)	10
A_3	– (20)	3 (40)	– (20)	– (60)	3
Nachfrage	5	6	7	4	