

MATHEMATIK MTA 12
SCHULJAHR 07/08
STATISTIK

PROF. DR. CHRISTINA BIRKENHAKE

INHALTSVERZEICHNIS

1. Merkmale	2
2. Urliste und Häufigkeitstabellen	9
3. Graphische Darstellung von Daten	10
4. Lageparameter	13
4.1. Modus	13
4.2. Median	15
4.3. Quantile	18
4.4. Arithmetisches Mittel	18
5. Streuungsmaße	20
5.1. Spannweite	20
5.2. Varianz und Standardabweichung	20
5.3. Variationskoeffizient	23
6. Korrelation und Regression	24
7. Aufgaben	28

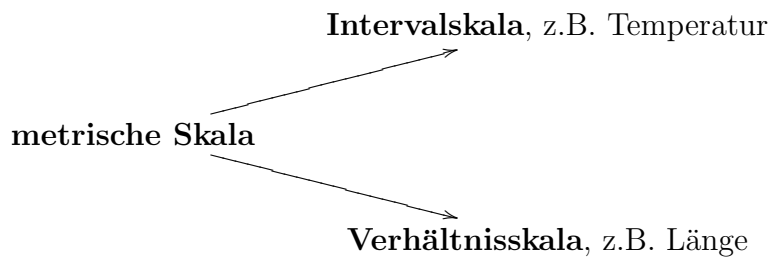
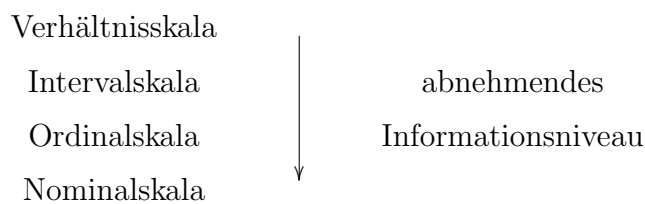
1. MERKMALE

Klassenfragebogen

Merkmal:	die Eigenschaft, die bei einer statistischen Untersuchung abgefragt wird. z.B. Geschlecht,...
Merkmalsträger	Objekt/Lebewesen, das das Merkmal trägt. z.B. Schüler dieser Klasse, Schüler in Nürnberg
Merkmalswert	die Angabe, die bei der Befragung eines Merkmals festgestellt wurde. z.B. Merkmal: Geschlecht \Rightarrow Merkmalswerte: m. und w.

Skalierung/Ordnung von Merkmalen:

Skala	Merkmal	Beispiele
Nominalskala	qualitative Merkmale	Geschlecht, Beruf
Ordinalskala	ordinale Merkmale	Noten
Metrische Skala	quantitative Merkmale	Größe, Datum

**Skalierung \Leftrightarrow Ordnung**

Fragebogen

Klasse: _____ Datum: _____

Geschlecht m. w.

Blutgruppe 0 A B AB

Ihr Schlafbedürfnis
viel mittel wenig

Wann sind Sie gestern zu Bett gegangen?

Wählen Sie eine ganze Zahl zwischen 0 und 100

Alter Jahre

Größe cm

Wieviel Münzgeld haben Sie dabei? Cent

URLISTE: KLASSE

DATUM

Nr.	Geschlecht	Blutgruppe	Schlafbedürfnis	Bettruhe
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				

URLISTE: KLASSE

DATUM

Nr.	Zahl	Alter	Größe	Münzgeld
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				

HÄUFIGKEITSTABELLEN ZUM FRAGEBOGEN VOM

MERKMAL: GESCHLECHT

i	x_i	h_i	f_i	H_i	F_i
1	m.				
2	w.				

MERKMAL: BLUTGRUPPE

i	x_i	h_i	f_i	H_i	F_i
1	0				
2	A				
3	B				
4	AB				

MERKMAL: SCHLAFBEDÜRFNIS

i	x_i	h_i	f_i	H_i	F_i
1	1 = viel				
2	2 = mittel				
3	3 = wenig				

MERKMAL: BETTRUHE

i		h_i	f_i	H_i	F_i
1					
2					
3					
4					
5					
6					
7					
8					
9					

MERKMAL: ZAHL

i		h_i	f_i	H_i	F_i
1					
2					
3					
4					
5					

MERKMAL: ALTER

i	$]x_i^u, x_i^o]$	h_i	f_i	H_i	F_i
1					
2					
3					
4					
5					
6					
7					

MERKMAL: GRÖSSE

i	$]x_i^u, x_i^o]$	h_i	f_i	H_i	F_i
1					
2					
3					
4					
5					
6					
7					

MERKMAL: MÜNZGELD

i		h_i	f_i	H_i	F_i
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					

2. URLISTE UND HÄUFIGKEITSTABELLEN

Bezeichnungen in Häufigkeitstabellen

n	Stichprobenumfang
x_1, \dots, x_ν	Merkmalswerte
h_i	absolute Häufigkeit des Merkmals x_i

$$h_1 + \dots + h_\nu = n$$

$f_i = \frac{h_i}{n} \cdot (100\%)$ relative Häufigkeit des M.W. x_i

$$f_1 + \dots + f_\nu = 100\%$$

nur bei ordinalen und metrischen Merkmalen:

H_i absolute Summenhäufigkeit des M.W. x_i

$$H_i = h_1 + \dots + h_i$$

$$H_1 = h_1, \quad H_\nu = h_1 + \dots + h_\nu = n$$

$F_i = \frac{H_i}{n}$ relative Summenhäufigkeit des M.W. x_i

$$F_i = f_1 + \dots + f_i$$

$$F_1 = f_1, \quad F_\nu = \frac{H_\nu}{n} = 100\%$$

Klasseneinteilung

Bei quantitativen Merkmalen: Spektrum der Merkmalswerte wird in Intervalle aufgeteilt.

Beispiel Merkmal: Münzgeld, Spektrum: 0 – 10 €

Klassen: $[0, 100[, [100, 200[, [200; 300[, \dots$

Klassenbreite nicht notwendig konstant!

⇒ klassifizierte Häufigkeitsverteilungen

x_i^u Klassenuntergrenze

x_i^o Klassenobergrenze

$[x_i^u; x_i^o[$ oder $]x_i^u; x_i^o]$ i-te Klasse

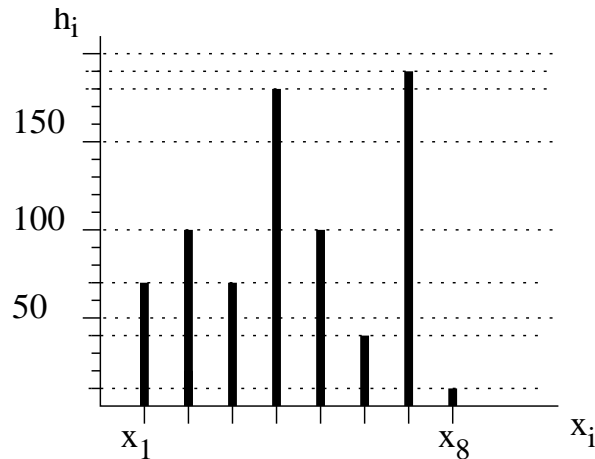
Typ einer Häufigkeitsverteilung:

diskrete Verteilung ⇔ **nichtklassifizierte Verteilung**

3. GRAPHISCHE DARSTELLUNG VON DATEN

Stabdiagramm

(Arbeitsblatt Stabdiagramm und Histogramm)

**Eignung:** qualitative, ordinale, diskrete quantitative Häufigkeitsverteilungen**Kreisdiagramm****Eignung:** qualitative, ordinale, diskrete quantitative Häufigkeitsverteilungen, wie bei Stabdiagramm.Winkel proportional zur Häufigkeit h_i .Winkel für Merkmal x_i : $360^\circ \cdot f_i$ **Histogramm****Eignung:** Klassifizierte quantitative Häufigkeitsverteilungen.x-Achse: Klassengrenzen $x_1^u, x_2^u, \dots, x_\nu^o$

Über jeder Klasse: Rechteck, Rechteckfläche proportional zur Häufigkeit

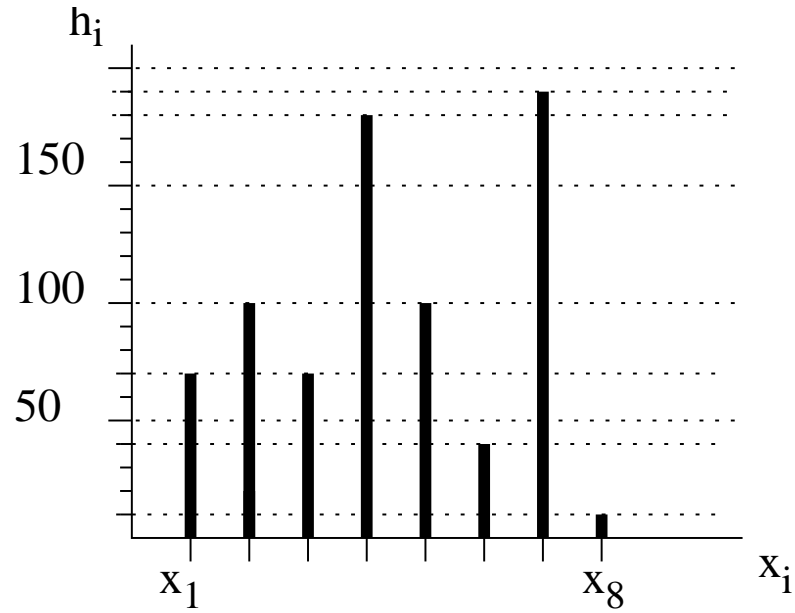
$$h_i \sim \text{Rechteckfläche} = \text{Höhe} \cdot \text{Breite}$$

$$\text{Klassenbreite} = x_i^o - x_i^u$$

$$\Rightarrow \text{Rechteckhöhe} = \frac{\text{Fläche}}{\text{Breite}} \sim \frac{h_i}{(x_i^o - x_i^u)}$$

$\text{Klassendichte } d_i := \frac{h_i}{(x_i^o - x_i^u)}$	$\text{Rechteckhöhe} \sim d_i$
------------------------------------------------------------	--------------------------------

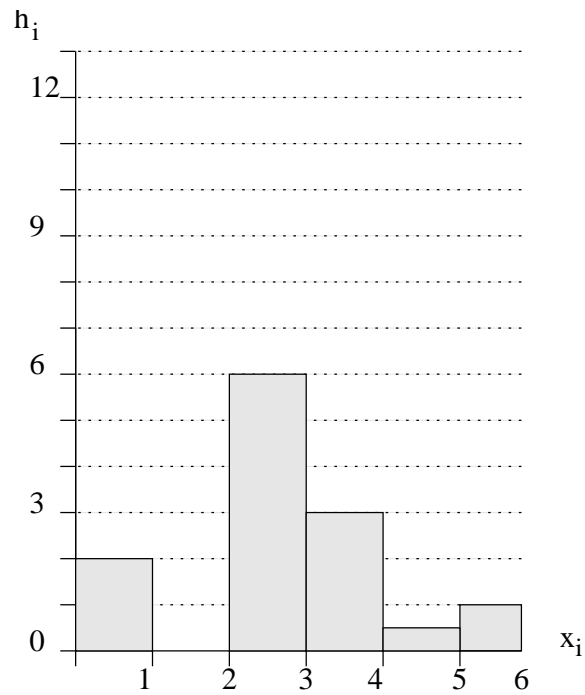
Stabdiagramm und Histogramm



Fall I: Stabdiagramm ist nominal skaliert, d.h. x_1, \dots, x_8 sind Namen.

Fall II: Stabdiagramm ist quantitativ skaliert, z. B. mit

$$x_1 = 10, x_2 = 20, \dots, x_8 = 80$$

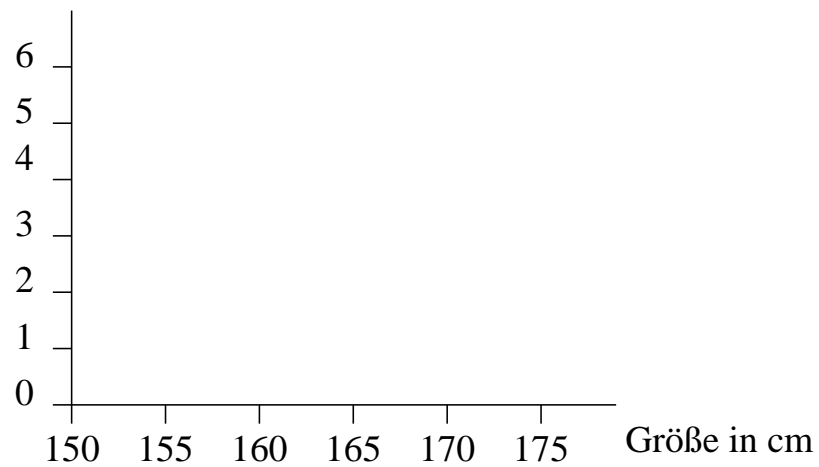


Histogramme

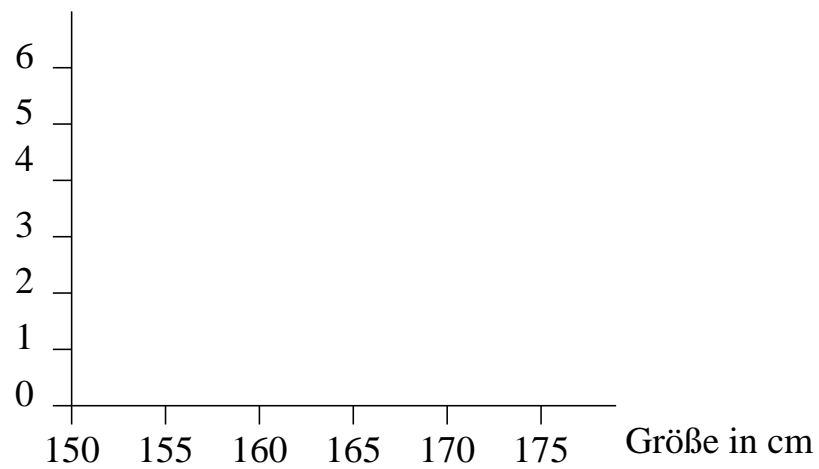
Häufigkeitstabelle: Größe der Schüler einer Klasse

Nr.	Größe x in cm	h_i	Kastenhöhe	Kastenhöhe
1	[150; 155[1		
2	[155; 160[3		
3	[160; 165[2		
4	[165; 170[5		
5	[170; 175[1		

Histogramm, bei konstanter Klassenbreite



Histogramm, bei nicht konstanter Klassenbreite



4. LAGEPARAMETER

Lageparameter sind Werte, die die Häufigkeitsverteilung beschreiben. Hier sollen Modus, Median, Quantile und arithmetisches Mittel beschrieben werden.

4.1. Modus.

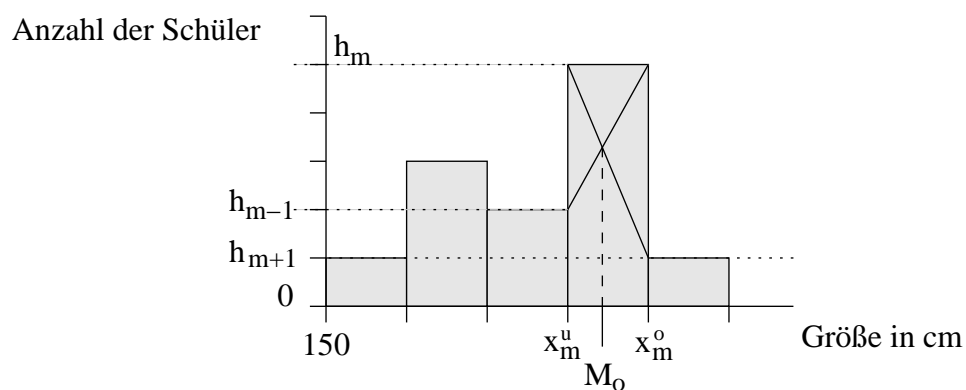
Modus M_o der Merkmalswert, der am häufigsten beobachtet wurde.

Eignung: Jede Skala

Bei klassifizierten (quantitativen) Häufigkeitsverteilungen:

Modusklasse $[x_m^u, x_m^o[$ die Klasse mit der größten Klassenhäufigkeit h_m bzw. Klassendichte d_m

Graphische Ermittlung des Modus:



Konstante Klassenbreite:

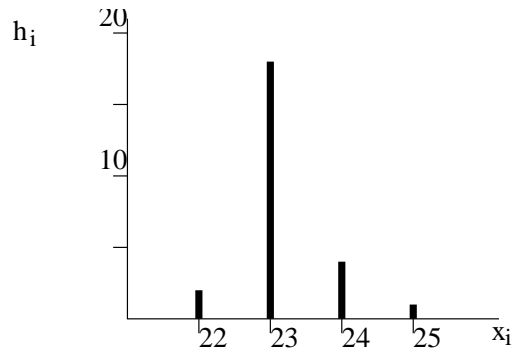
$$M_o = x_m^u + \frac{h_m - h_{m-1}}{(h_m - h_{m-1}) + (h_m - h_{m+1})} \cdot (x_m^o - x_m^u)$$

Nichtkonstante Klassenbreite:

$$M_o = x_m^u + \frac{d_m - d_{m-1}}{(d_m - d_{m-1}) + (d_m - d_{m+1})} \cdot (x_m^o - x_m^u)$$

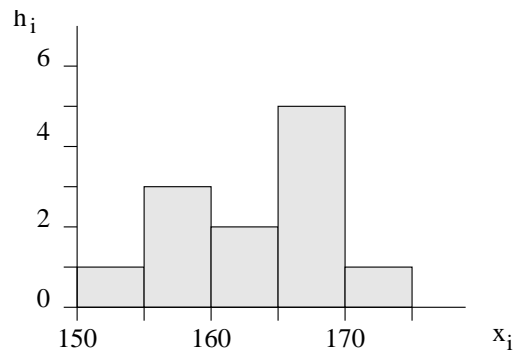
Modus M_0

diskrete Verteilung



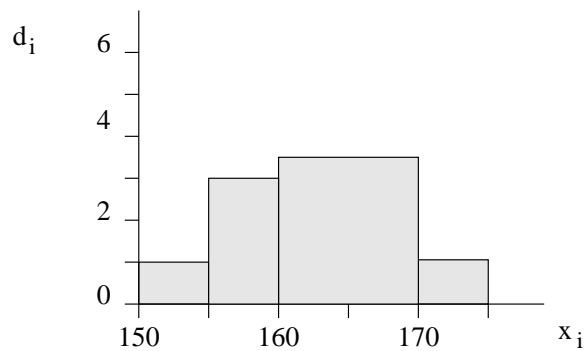
Klassifizierte Verteilung, konstante Klassenbreite

$$M_0 = x_m^u + \frac{h_m - h_{m-1}}{(h_m - h_{m-1}) + (h_m - h_{m+1})} \cdot (x_m^o - x_m^u)$$



Klassifizierte Verteilung, nicht konstante Klassenbreite

$$M_0 = x_m^u + \frac{d_m - d_{m-1}}{(d_m - d_{m-1}) + (d_m - d_{m+1})} \cdot (x_m^o - x_m^u)$$



4.2. Median.

Median M_e der Merkmalswert, der unter Berücksichtigung der Häufigkeiten in der Mitte steht.

= **Zentralwert**.

Eignung: Skala muß mindestens ordinalskaliert sein.

Beispiele aus der Klassenstatistik:

Merkmal Größe: stellen Sie sich der Größe nach hin. Die Größe, des in der Mitte Stehenden Schülers ist der Median.

Merkmal Schlafbedürfnis: stellen Sie sich der Reihe nach hin, zuerst die mit wenig Schlafbed., dann die mit mittel und dann die mit viel. Der Median ist wieder der zum mittleren Schüler gehörige Merkmalswert.

Median bei klassifizierten Häufigkeitsverteilungen:

Der Median halbiert die Fläche des Häufigkeiten-Histogramms

Median bei einer klassifizierten Verteilung:

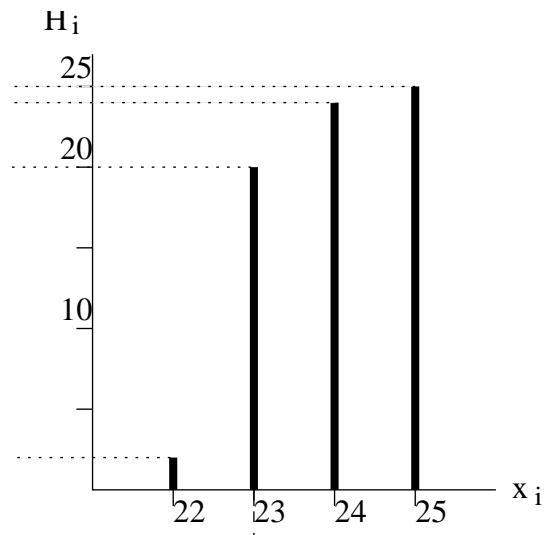
Medianklasse: hier $[x_m^u, x_m^o[= [165, 170[$

$$\begin{aligned} M_e &= x_m^u + \frac{\frac{n}{2} - H_{m-1}}{H_m - H_{m-1}} (x_m^o - x_m^u) \\ &= x_m^u + \frac{\frac{n}{2} - H_{m-1}}{h_m} (x_m^o - x_m^u) \end{aligned}$$

Median M_e

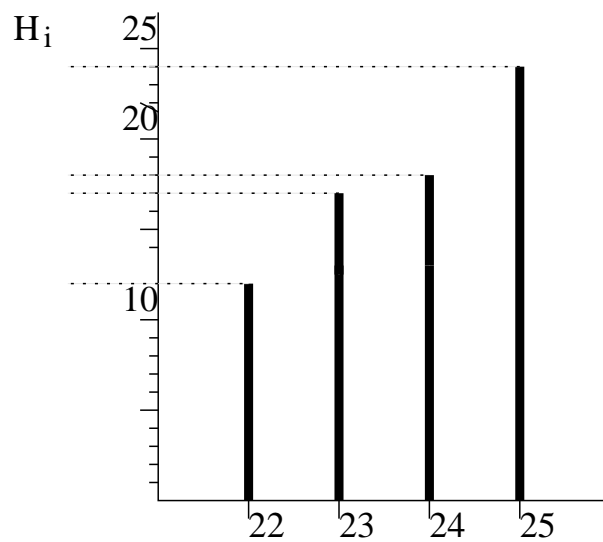
Diskrete Verteilung

Fall: n ungerade



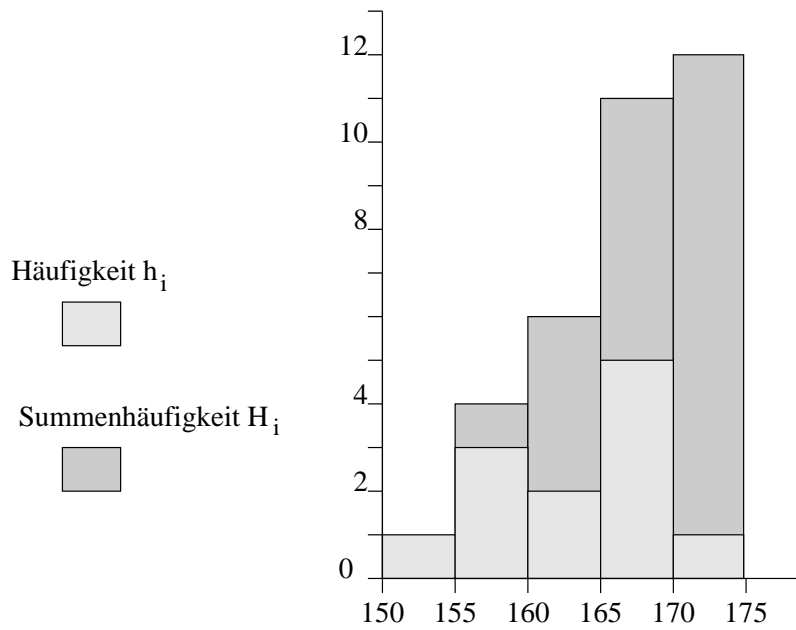
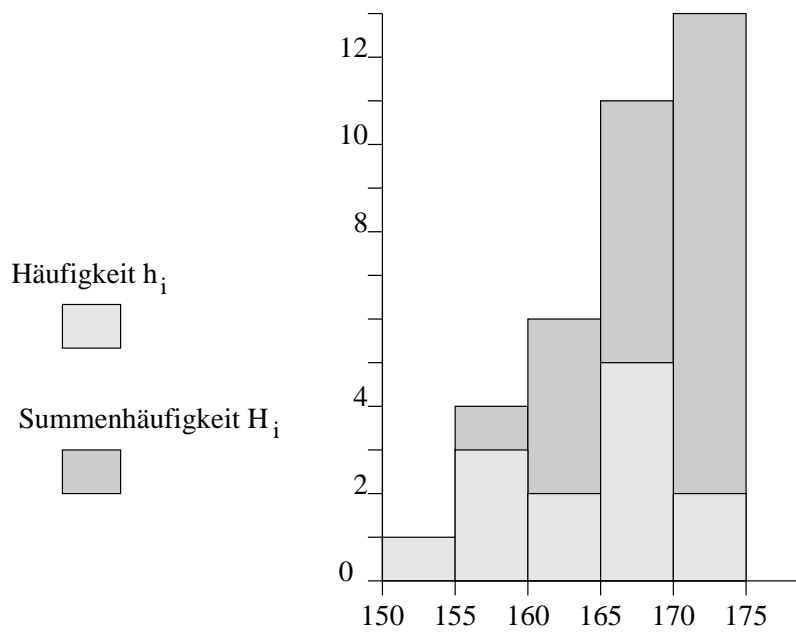
i	x_i	h_i	H_i
1	22	2	2
2	23	18	20
3	24	4	24
4	25	1	25

Fall: n gerade



i	x_i	h_i	H_i
1	22	12	12
2	23	5	17
3	24	1	18
4	25	6	24

Median bei einer Klassifizierten Verteilung mit konstanter Klassenbreite



4.3. Quantile.

Median M_e ist gleich dem 50%-Quantil, die Hälfte der Merkmalswerte mit Vielfachheiten liegen davor und die andere Hälfte danach.

Analog definiert man:

$$25\text{-Quantil} = 1\text{-tes Quartil} = x_{25\%}$$

$$75\text{-Quantil} = 3\text{-tes Quartil} = x_{75\%}$$

$$10\text{-Quantil} = x_{10\%}$$

$$90\text{-Quantil} = x_{90\%}$$

4.4. Arithmetisches Mittel.

nichtklassifizierte Häufigkeitsverteilung: $\bar{x} = \frac{1}{n}(x_1 h_1 + \dots + x_\nu h_\nu)$

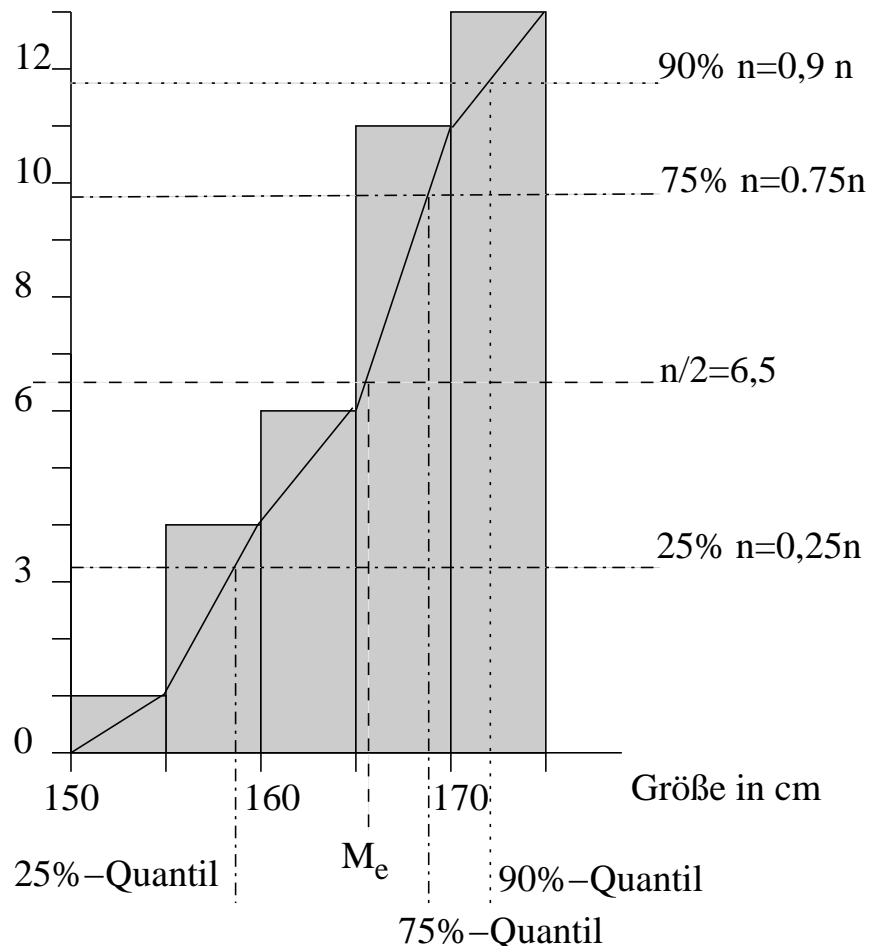
klassifizierte Häufigkeitsverteilung: $\bar{x} = \frac{1}{n}(x'_1 h_1 + \dots + x'_\nu h_\nu)$

mit Klassenmitten $x'_i = \frac{x_i^u + x_i^o}{2}$

Quantile

Anzahl der Schüler

Summenhäufigkeit H_i



5. STREUUNGSMASSE

Während die Lageparameter markante Merkmalswerte einer Verteilung angeben, beschreiben die Streuungsparameter den groben Verlauf der Verteilung. Hier sollen Spannweite, Varianz, Standardabweichung und der Variationskoeffizient behandelt werden.

5.1. Spannweite.

Definition 5.1. Die **Spannweite** R ist die Differenz aus größtem und kleinstem beobachteten Merkmalswert.

Voraussetzung: metrische Skala, mindestens intervallskaliert

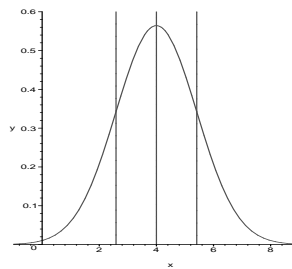
Spannweite $R = \text{größter Merkmalswert} - \text{kleinster Merkmalswert}$

$$R = \begin{cases} x_{max} - x_{min} & \text{falls nicht klassifiziert} \\ x_{max}^o - x_{min}^u & \text{falls klassifiziert} \end{cases}$$

5.2. Varianz und Standardabweichung.

Voraussetzung: metrische Skala, mindestens intervallskaliert

Die Grundidee ist, die Verteilung mit der Standard-Normalverteilung zu vergleichen. Gelättet sieht die Normalverteilung wie die Gaußsche Glockenkurve aus:



Die Varianz σ^2 ist ein Maß für die Abweichung vom arithmetischen Mittel. Im Bild der Glockenkurve ist das Maximum am arithmetischen Mittel \bar{x} und der Abstand der Wendepunkte vom Maximum ist die Standardabweichung σ .

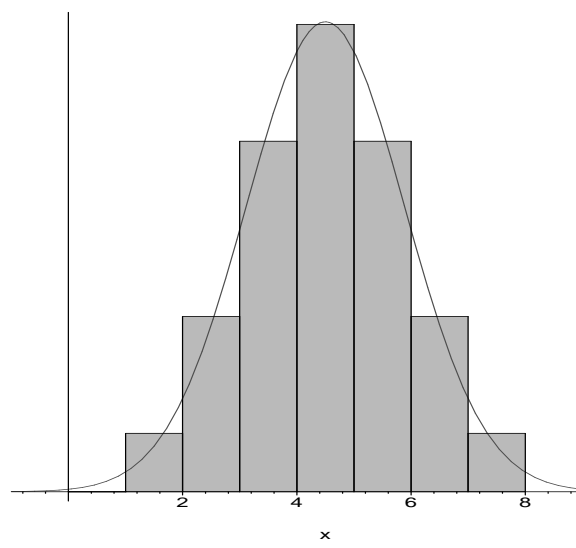
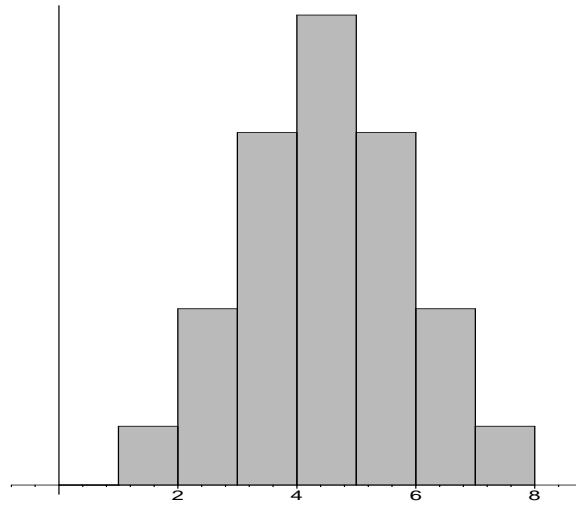
Varianz:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{\nu} (x_i - \bar{x})^2 \cdot h_i = \frac{1}{n} \sum_{i=1}^{\nu} x_i^2 \cdot h_i - \bar{x}^2$$

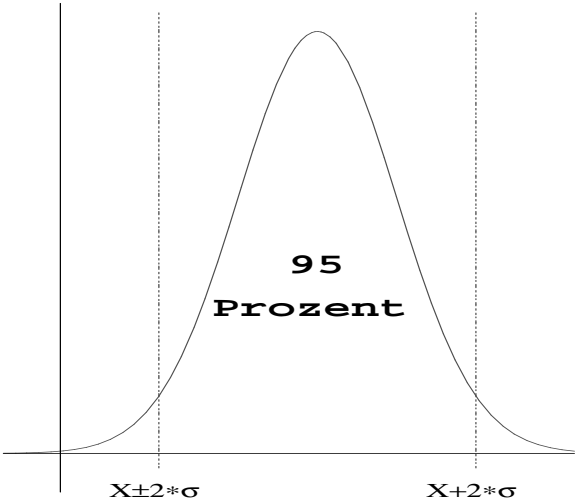
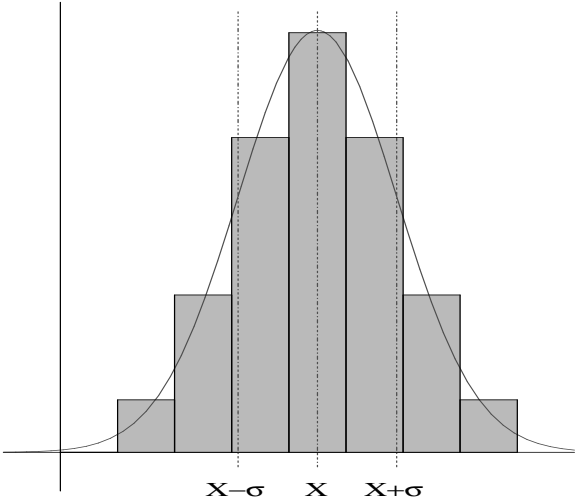
Standardabweichung:

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

Normalverteilung



Standardabweichung



5.3. Variationskoeffizient.

Voraussetzung: Verhältnisskala

Der Variationskoeffizient mißt die relative Streuung.

Variationskoeffizient:

$$VK = \frac{\sigma}{\bar{x}} \cdot 100\%$$

Der Variationskoeffizient ist nicht anschaulich interpretierbar.

6. KORRELATION UND REGRESSION

Größe von Geschwistern (in cm)

Nr.	Familie	Schwester	Bruder
1	1	69	71
2	2	64	68
3	3	65	66
4	4	63	67
5	5	65	70
6	6	62	71
7	7	65	70
8	8	64	73
9	9	66	72
10	10	59	65
11	11	62	66
12	12	65	59
13	13	62	64

Erstellen Sie ein **Stuediagramm**:

- Zu welcher Familie gehört der größte Junge / das kleinste Mädchen?
- Wie groß ist die Schwester des größten Jungen/ der Bruder des kleinsten Mädchens?
- Trage Sie die Gerade $x = y$ ein, Was bedeutet wenn ein Punkt ober- bzw. unterhalb dieser Gerade liegt?

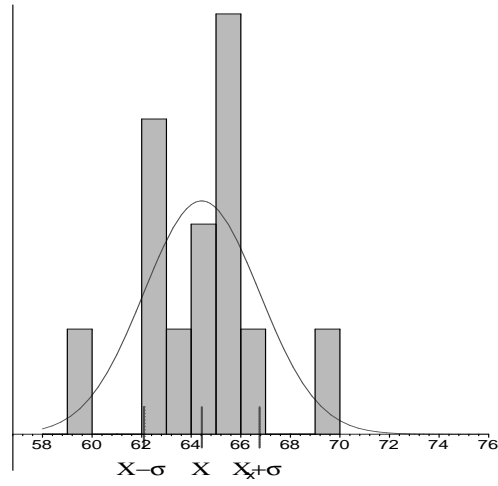
d) Vervollständigen Sie die folgende Tabelle:

Nr.	Schwester (x_i)	Bruder (y_i)	$x_i y_i$	x_i^2	y_i^2
1	69	71			
2	64	68			
3	65	66			
4	63	67			
5	65	70			
6	62	71			
7	65	70			
8	64	73			
9	66	72			
10	59	65			
11	62	66			
12	65	59			
13	62	64			
Summe					
$\frac{1}{n}$ Summe					

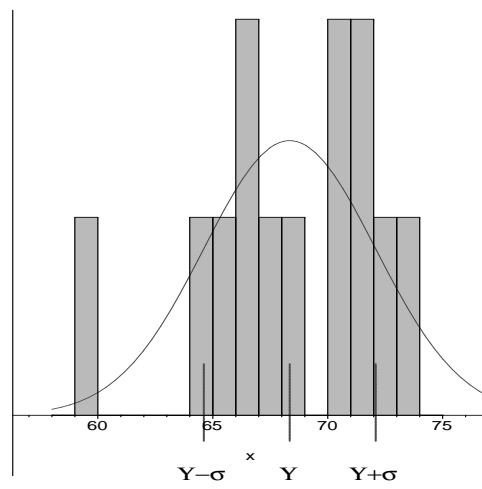
e) Berechnen Sie die Standardabweichungen σ_x und σ_y sowie $\sigma_x \pm \bar{x}$, $\sigma_y \pm \bar{y}$ und $\sigma_{xy} := \frac{1}{n} \sum_i x_i y_i - \bar{x} \bar{y}$.

f) Fertigen Sie für beide Merkmale: Größe Schwester(x) bzw. Größe Bruder (y) Histogramme an.

Schwestern



Brüder



g) Berechnen Sie die **Regressionsgeraden** nach den Formeln:

$$B_y = \frac{\frac{1}{n} \sum_i x_i y_i - \bar{x} \bar{y}}{\frac{1}{n} \sum_i x_i^2 - \bar{x}^2} = \frac{\sigma_{xy}}{\sigma_x^2}$$
$$A_y := \frac{\bar{x} \cdot \frac{1}{n} (\sum_i x_i y_i) - \frac{1}{n} \sum_i x_i^2 \cdot \bar{y}}{\frac{1}{n} \sum_i x_i^2 - \bar{x}^2} = \bar{y} - B_y \bar{x}$$

h) Berechnen Sie den **Korrelationskoeffizienten** r nach den Formeln:

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

Interpretation der Regressionsgeraden:

Die Regressionsgerade G_y beschreibt die Abhängigkeit des Merkmals Y vom Merkmal X . Zu jedem Merkmalswert x_i kann ein durchschnittlicher zugehöriger Merkmalswert $y(x_i) = B_y x_i + A_y$ berechnet werden. Analog mit G_X .

Interpretation des Korrelationskoeffizienten r :

Der Korrelationskoeffizient mißt die Stärke des linearen Zusammenhangs der Merkmale X und Y . Es gilt:

$$-1 \leq r \leq 1$$

Vorzeichen von r

$r > 0$: Wird x groß, so wird auch y tendenziell groß. Der Zusammenhang der Merkmale ist positiv!

$r = +1$: Die Datenpunkte liegen auf einer Geraden mit positiver Steigung.

$r = 0$: Die Datenpunkte zeigen keinen linearen Zusammenhang.

$r < 0$: Wird x groß, so wird y tendenziell kleiner. Der Zusammenhang der Merkmale ist negativ!

$r = -1$: Die Datenpunkte liegen auf einer Geraden mit negativer Steigung.

Stärke des Zusammenhangs

$|r|$ **nahe Null**: Der Zusammenhang der Merkmale ist klein, die Regressionsgeraden sind nahezu senkrecht zueinander.

$|r|$ **nahe Eins**: Der Zusammenhang der Merkmale ist groß. Die Regressionsgeraden fallen nahezu zusammen. Die Merkmale liegen fast auf der Regressionsgeraden.

7. AUFGABEN

Aufgabe 1:

Gebe von folgenden Merkmalen die Skala an:

Gehaltsgruppe

Einkommen

Autofarbe

Kundenzufriedenheit

Berufsbezeichnung

Dienstgrad

Mietpreis

Klausurergebnis

Cholesterinspiegel

Wartezeiten von Patienten einer Arztpraxis am 18.9.2003

Uhrzeit des Behandlungsbeginns der Patienten dieser Praxis am 18.9.2003

Schultypen

Anzahl von Kindern in Schulklassen

Rauchgewohnheiten

Schulnoten

Benzinverbrauch von PKW's

Körpertemperatur eines Patienten

Lebensdauer

Alter

Geburtsdatum

Aufgabe 2:

Im Schulungszentrum eines Medizinischen Labors wurde in Parallelkursen das Alter der Teilnehmer ermittelt. Folgende Häufigkeitstabellen haben sich dabei ergeben:

Kurs I:

Alter	22	23	24	25
absolute Häufigkeit h_i	2	18	4	1

Kurs II:

Alter	22	23	24	25
absolute Häufigkeit h_i	3	13	2	2

Kurs III:

Alter	22	23	24	25
absolute Häufigkeit h_i	9	12	6	3

- (1) Bestimme in den drei Fällen die Anzahl der Teilnehmer n .
- (2) Erstelle in den drei Fällen die Häufigkeitstabellen.
- (3) In welchem Kurs ist die Anzahl der ≤ 23 -jährigen am größten?
- (4) Was ist der prozentuale Anteil der 22-jährigen in jedem Kurs?

Aufgabe 3:

In einem medizinischen Forschungsinstitut mit 180 Beschäftigten findet man die folgende Altersstruktur:

5	bis 20 Jahre
59	über 20 bis 30 Jahre
31	über 30 bis 40 Jahre
39	über 40 bis 50 Jahre
46	über 50 bis 60 Jahre

Erstelle die Häufigkeitstabelle.

Aufgabe 4:

Welche graphischen Darstellungen eignen sich für die Merkmale unserer Klassenstatistik?

Geschlecht

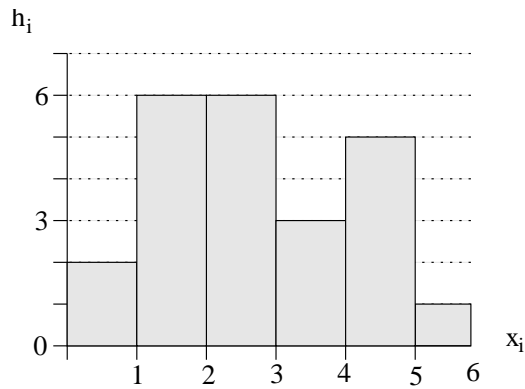
Blutgruppe

Geburtsjahr

Alter

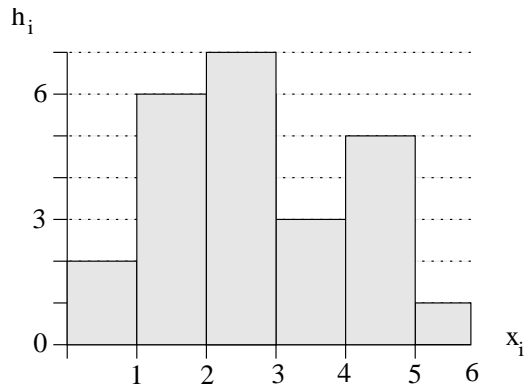
Größe
Schlafbedürfnis
WBC
HGB
PLT

Aufgabe 5:



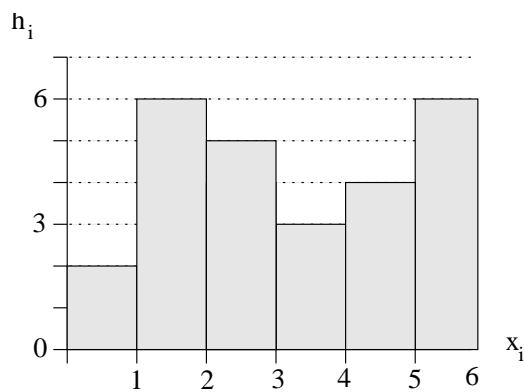
i	$]x_i^u, x_i^o]$	h_i	H_i
1	$]0, 1]$		
2	$]1, 2]$		
3	$]2, 3]$		
4	$]3, 4]$		
5	$]4, 5]$		
6	$]5, 6]$		

Aufgabe 6:



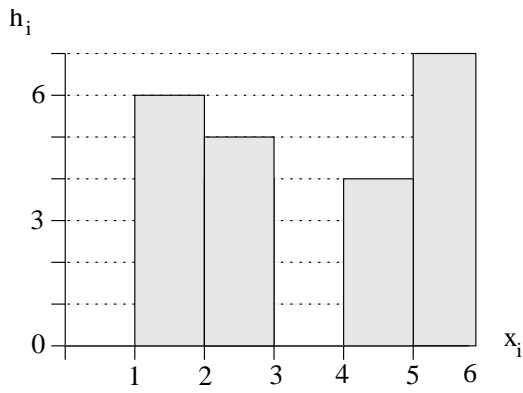
i	$]x_i^u, x_i^o]$	h_i	H_i
1	$]0, 1]$		
2	$]1, 2]$		
3	$]2, 3]$		
4	$]3, 4]$		
5	$]4, 5]$		
6	$]5, 6]$		

Aufgabe 7:



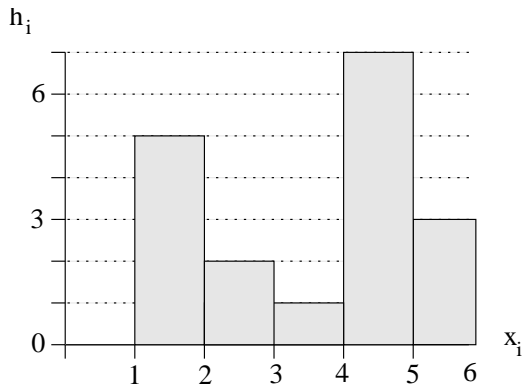
i	$]x_i^u, x_i^o]$	h_i	H_i
1	$]0, 1]$		
2	$]1, 2]$		
3	$]2, 3]$		
4	$]3, 4]$		
5	$]4, 5]$		
6	$]5, 6]$		

Aufgabe 8:



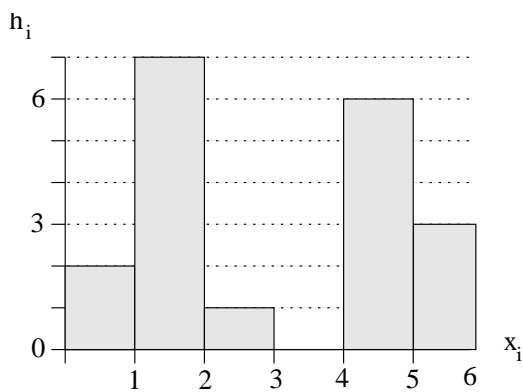
i	$]x_i^u, x_i^o]$	h_i	H_i
1	$]0, 1]$		
2	$]1, 2]$		
3	$]2, 3]$		
4	$]3, 4]$		
5	$]4, 5]$		
6	$]5, 6]$		

Aufgabe 9:



i	$]x_i^u, x_i^o]$	h_i	H_i
1	$]0, 1]$		
2	$]1, 2]$		
3	$]2, 3]$		
4	$]3, 4]$		
5	$]4, 5]$		
6	$]5, 6]$		

Aufgabe 10:



i	$]x_i^u, x_i^o]$	h_i	H_i
1	$]0, 1]$		
2	$]1, 2]$		
3	$]2, 3]$		
4	$]3, 4]$		
5	$]4, 5]$		
6	$]5, 6]$		