

Statistik–Skalen (Gurtner 2004)

Nominalskala:

Daten haben nur **Namen(Nomen)** und (eigentlich) keinen Zahlenwert
Es kann nur der **Modus** („ofteste“ Wert) berechnet werden

Beispiel 1:

Die Befragung von 48 Personen nach dem beim Weg zur Arbeit benutzten Verkehrsmittel ergab die Urliste:

1 2 3 3 4 4 3 3 4 2 2 3
1 1 1 2 3 4 3 2 5 4 5 4
3 4 5 4 1 3 1 1 1 1 5 4
1 4 5 1 2 2 3 3 5 4 1 1

Dabei bedeutet: 1–Eigener PKW 2–Öffi 3–Fahrrad 4–Zu Fuß 5–Mitfahrer PKW

- a) Wieso ist das Merkmal „Verkehrsmittel“ **nominal** skaliert obwohl dort Zahlen stehen?
- b) Erstellen Sie ein „**Strichliste**“
- c) Erstellen Sie eine Tabelle der **absoluten** und **relativen Häufigkeiten**
- d) Erstellen Sie ein **Histogramm (Staffelbild)**
- e) Ermitteln Sie den **Modus** (häufigster Wert)

Lösung:

a) Die Zahlen bedeuten **keine Reihenfolge**, sondern dienen nur zur Vereinfachung der Ergebnisse der Befragung

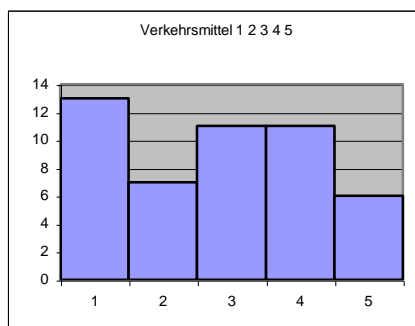
b)

```

1 – |||||  |||||  |||
2 – |||||  ||
3 – |||||  |||||  |
4 – |||||  |||||  |
5 – |||||  |
    
```

c)

Nummer	abs. Häufigkeit	relative Häufigkeit
1	13	$13/48 = 0,27 = 27\%$
2	7	$7/48 = 0,15 = 15\%$
3	11	$11/48 = 0,23 = 23\%$
4	11	$11/48 = 0,23 = 23\%$
5	6	$6/48 = 0,12 = 12\%$
SUMME	48	$48/48 = 1 = 100\%$



- d)
- e) **Modus** (häufigster Wert) ist 1 = Eigener PKW

Übung 1) Ermitteln Sie alles wie in Beispiel 1 für die Urliste der Haarfarben (1=braun 2=brünett 3=blond 4=rot 5=schwarz) von 30 Studentinnen:

2 3 1 5 5 5 2 5 3 2 5 2 5 2 3 4 3 2 2 5 2 1 2 4 5 2 5 2 3 1

Ordinalskala:

Daten können geordnet werden, die Abstände zwischen den einzelnen Werten sind allerdings unklar (Noten, Güteklassen) – es kann nur der **Median** (mittlere Wert der Liste) und die **Quartile** berechnet werden (und natürlich auch der Modus).

Beispiel 2:

Die Noten von 60 Studenten sind folgendermaßen aufgeteilt:

Note	Anzahl der Studenten
1	8
2	11
3	10
4	18
5	13

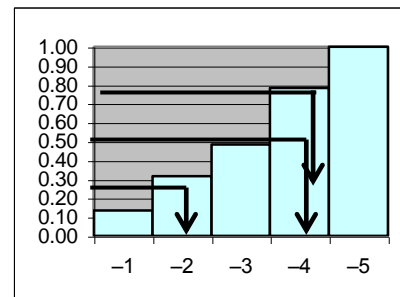
Bestimmen Sie die folgenden Größen:

- die **Summenhäufigkeitsverteilung** in Tabellenform und in Diagrammform
- Minimum, Maximum** und **Spannweite** der Daten
- Median** (Zentralwert) und **1. und 3. Quartil**
- Nehmen Sie an, die Note 1 wird mit 30 Punkten, die Note 2 mit 27 Punkten, die Note 3 mit 22 Punkten, die Note 4 mit 15 Punkten, die Note 5 mit 0 Punkten bewertet – welcher gewichtete Mittelwert ergibt sich. Vergleichen Sie mit dem gewichteten Mittelwert der Noten

Lösung:

a)

Note	Studentenanzahl	laufende Summe	relative Summe
-1	8	8	0.13
-2	11	19	0.32
-3	10	29	0.48
-4	18	47	0.78
-5	13	60	1.00



b)

Minimum = 1, **Maximum** = 5 und **Spannweite** = Maximum-Minimum = 4

c) Der Median wird ermittelt, in dem man den Wert sucht, der in der Mitte der Liste steht.

Der Wert in der Mitte ist der Mittelwert vom 30. und 31. Studenten. Der 30. Student hat die Note 4 und der 31. Student hat die Note 4 → **Median** ist 4

Im Diagramm ist es auch ersichtlich. Schaut man unter 0,5 auf der y-Achse und geht nach rechts, so erreicht man den Wert 4 → Median ist 4

Das 1.Quartil ergibt sich aus der 4-Teilung der Liste. Der 15. und 16.Student sind in der Mitte der ersten Hälfte aller Studenten. Beide haben die Note 2 → **1.Quartil** ist 2 (im Diagramm unter 0,25 waagrecht zu sehen)

Das 3.Quartil ergibt sich aus dem Mittelwert des 45. und 46. Studenten (beide Male Note 4), also ergibt sich → **3.Quartil** ist 4 (im Diagramm unter 0,75 zu sehen)

d) gewichteter Mittelwert = $(30 \cdot 8 + 27 \cdot 11 + 22 \cdot 10 + 15 \cdot 18 + 0 \cdot 13) / 60 \approx 17$ (entspräche Note 4 !!!)

ungewichteter Mittelwert = $(1 \cdot 8 + 2 \cdot 11 + 3 \cdot 10 + 4 \cdot 18 + 5 \cdot 13) / 60 \approx 3,3$ (entspräche Note 3)

Die **Mittelwertbildung ist hier eigentlich nicht zulässig**, da die Abstände der Noten untereinander nicht gleich sind, wie man bei der Punkteverteilung sehen kann.

Übung 2) Die Eierproduktion erfolge in 6 Güteklassen. Ermitteln Sie alles wie in Beispiel 2 a)-c) für die folgenden Daten:

Güteklasse	Anzahl
1	6
2	23
3	12
4	51
5	42
6	16
Summe	150

Beispiel 3:

Bei einer Stichprobe durch die Polizei wurde folgende Urliste für die Profiltiefen (in mm) von 14 Reifen ermittelt:

1,5 3,5 3,5 4,0 4,5 7,5 5,5 5,5 2,0 2,5 4,5 3,5 3,5 4,5

a) Machen Sie ein „**Stengel-Blatt-Diagramm**“

b) Ermitteln Sie Min, Max, Spannweite, Median und die Quartile und erstellen Sie damit ein „**BOX-PLOT**“ (Kastenschaubild)

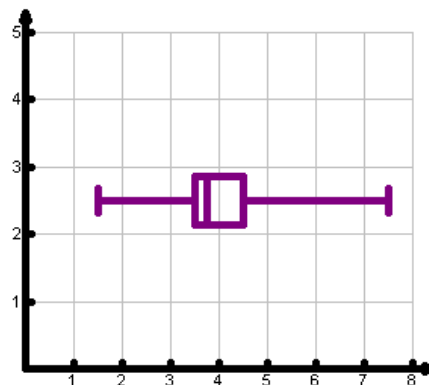
Lösung:

```

1, | 5
2, | 0, 5
3, | 5, 5, 5, 5
4, | 0, 5, 5, 5
5, | 5, 5
6, |
7, | 5
  
```

1.Quartil ←
 Median ←
 3.Quartil ←

b) Min = 1,5 Max = 7,5 Spannweite = 6 Median = 3,75 1.Quartil = 3,5 3.Quartil = 4,5 ergibt das BOX-PLOT:



Hier erkennt man, dass die Quartile sehr nahe beieinander liegen, was darauf hindeutet, dass es links und rechts Ausreißer (1,5 und 7,5) gibt, die Hauptverteilung aber in der Umgebung [3,5; 4,5] liegt.

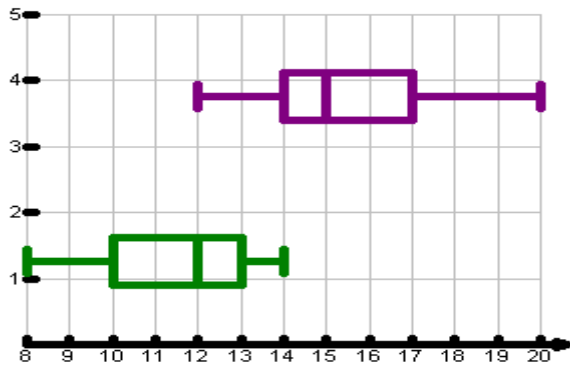
Übung 3) wie Beispiel 3 für die Daten der Testpersonen eines Alkoholisierungstests:

0,5 ‰ 0,8 ‰ 1,7 ‰ 0,9 ‰ 0,7 ‰ 1,1 ‰ 1,4 ‰ 1,5 ‰ 0 ‰ 0,5 ‰ 0,4 ‰ 0,8 ‰

Übung 4) wie Beispiel 3 für die Daten des Treibstoffverbrauchs von 20 Autos auf einer Teststrecke von 500 km (in Liter):

43,2 47,2 51,4 45,6 45,7 54,8 48,5 43,2 44,0 50,1
46,2 45,1 47,3 48,2 43,2 49,3 43,1 47,2 45,9 49,1

Übung 5) In folgendem Diagramm sind die mittleren Studiendauern als BOX-PLOT von 2 Studien eingezeichnet. Lesen Sie die Quartile ab und interpretieren Sie die Verteilung.



Metrische Skalen:

Daten können Zwischenwerte annehmen ohne absolutem Nullpunkt (Intervallskala: Temperatur, Zeit) oder mit absolutem Nullpunkt (Verhältnisskala: Körpergröße,...)
Jetzt können alle Mittelwerte berechnet werden!

Klasseneinteilung:

Beispiel:

Am Bauernhof werden 30 Eier eingesammelt und gewogen (in g):

69 63 55 57 62 56 59 62 63 51 53 59 49 64 65 52 68 58
 57 68 64 63 66 57 64 59 55 67 65 60

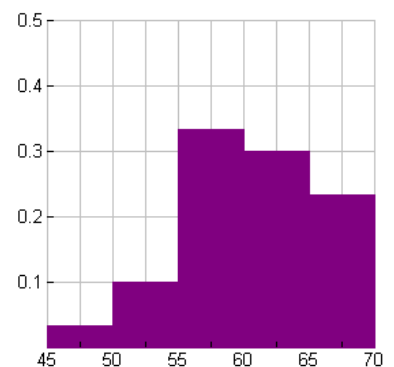
Sortieren Sie die Eier nach Gewichtgruppen (<45, 45-49, 50-54, 55-59, 60-64, 65-69, ≥70) und erstellen Sie eine **Klasseneinteilung** mit Strichliste, absoluter und relativer Häufigkeit

und erstellen Sie ein **Histogramm** (Streifen/Staffel-schaubild)

Berechnen Sie den **Mittelwert** der Eiermassen – mit den **Urdaten** und mit den **Klassenmitten** und absoluten Häufigkeiten der Klasseneinteilung

Lösung:

Masse (in g)	Strichliste	abs.Häufigkeit	rel. Häufigkeit
< 45		0	0
45-49		1	$1/30 = 0,03 = 3\%$
50-54		3	$3/30 = 0,10 = 10\%$
55-59		10	$10/30 = 0,33 = 33\%$
60-64		9	$9/30 = 0,30 = 30\%$
65-69		7	$7/30 = 0,23 = 23\%$
≥ 70		0	0
SUMME		30	$30/30 = 1 = 100\%$



mittl.Masse	abs.Häuf.	Produkt
47.5	1	47.5
52.5	3	157.5
57.5	10	575
62.5	9	562.5
67.5	7	472.5
SUMME	30	1815
geteilt durch 30:		60.5 = Mittelwert (\approx Mittelwert aus der Urliste = 60,3 g)

Übung 6) Von 30 Autos wurden im Ortsbereich die Geschwindigkeiten gemessen (in km/h):

55 63 48 103 57 48 72 35 42 60 52 59 71 80 83
63 70 51 83 75 60 67 56 40 105 49 53 57 70 46

Machen Sie eine **Klasseneinteilung** in vorschriftsmäßig (≤ 50), leicht überhöht (51-60), stark überhöht (61-75) und Raser (>75) und ermitteln Sie ein **Histogramm** und die **Mittelwerte** auf beide Arten

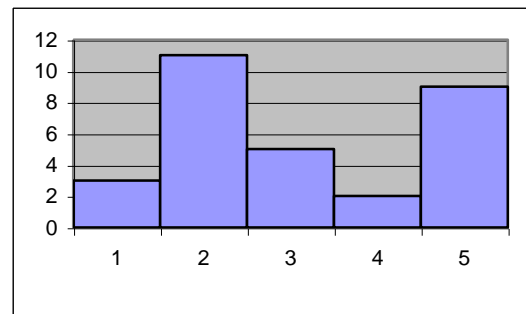
Übung 7) Folgende Gehälter gelangen in einem Betrieb zur Auszahlung:

582 582 598 598 634 634 680 687 720 740 1130 1310 1310 1383
1390 1390 1450 1480 1480 1870 2100 2170 2225 2245 2245 2245 2388

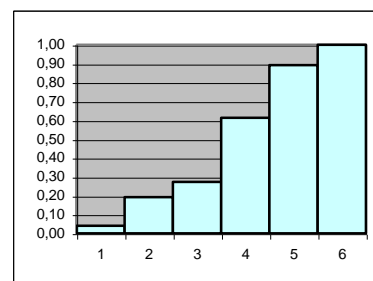
- Machen Sie eine **Klasseneinteilung** im Bereich 0 bis 2400 in 3 gleich breite Klassen und zeichnen Sie das Histogramm
- Machen Sie eine **Klasseneinteilung** im Bereich von 0 bis 2400 in 4 gleich breite Klassen und zeichnen Sie das Histogramm
- Vergleichen** Sie beide Darstellungen und begründen Sie, welches Histogramm die Arbeitgeberseite und welches die Arbeitnehmerseite bei Lohnverhandlungen wahrscheinlich verwenden würde.
- Machen Sie ein **Boxplot** der Urdaten

Lösungen:

	NR	Strichliste	abs	rel.Häuf.
1)	1		3	0,10
Modus = 2	2		11	0,37
	3		5	0,17
	4		2	0,07
	5		9	0,30
	SUMME		30	1

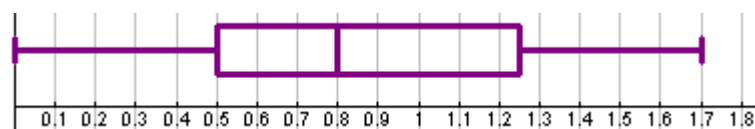


2)	Güteklasse	Anzahl	laufende Summe	relative Summe
	1	6	6	0,04
	2	23	29	0,19
	3	12	41	0,27
	4	51	92	0,61
	5	42	134	0,89
	6	16	150	1,00



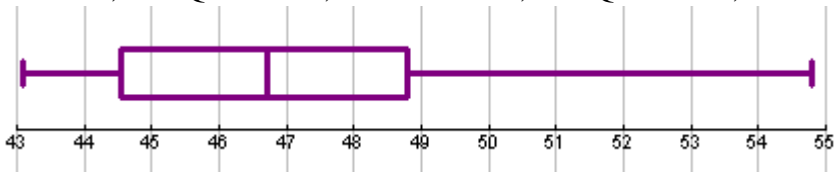
MIN = 1, MAX = 6, Spannweite = 5, 1.Quartil = 3, Median = 4, 3.Quartil = 5

3) 0, | 0 4 5 5 7 8 8 9
1, | 1 4 5 7



MIN = 0 MAX = 1,7 Spannweite = 1,7 1.Quartil = 0,5 Median = 0,8 3.Quartil = 1,25

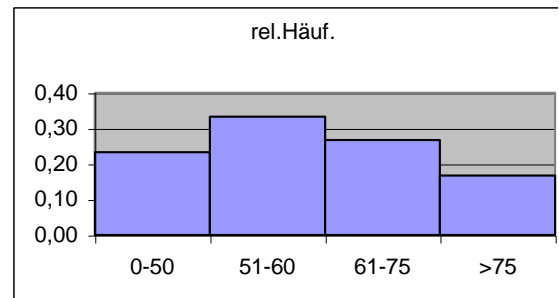
4) Median = 46,7
 1.Qu.=44,55 3.Qu.=48,8
 Min=43,1 Max=54,8 Sp=11,7



5) **1.Studium:** Median =15, 1.Qu. ist nahe bei, 3.Qu. auch, die Hälfte der Studenten ist zw. 14 und 17 Semestern fertig. Frühestens ist man mit 12 und spätestens mit 20 Semestern fertig
2.Studium: Median ist bei 12. Die Hälfte der Studenten ist zwischen 10 und 13 Semestern fertig, Ausreißer sind: mit 8, bzw. 14 Semestern fertig

6)

eschwindigkeit	abs.Häuf.	rel.Häuf.	Klassenmitte	Produkt
0-50	7	0,23	25	5,83333333
51-60	10	0,33	55	18,33333333
61-75	8	0,27	67,5	18
>75	5	0,17	90	15
SUMME	30	1	237,5	57,1666667
				<u>Mittelwert</u>

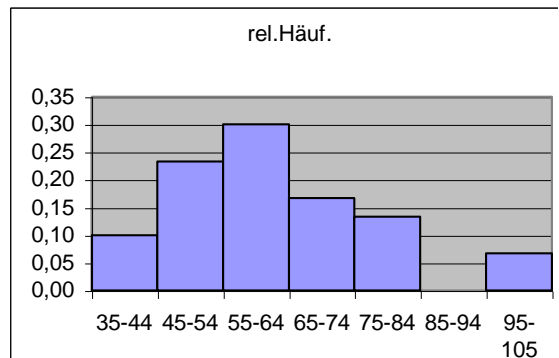


Mittelwert aller Ursprungsdaten = 62,4

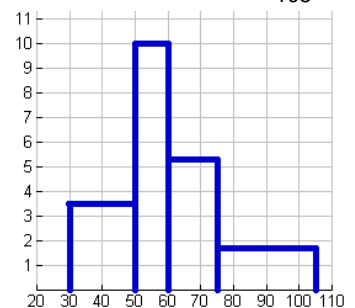
Hier ist das Problem der unterschiedlichen Klassenbreiten und des unklaren Beginns und Endes

Besser angepasst ist folgende Klasseneinteilung: Spannweite= 105-35 = 70, also 7 Klassen mit Klassenbreite 10

Gehälter	abs.Häuf.	rel.Häuf.	Klassenmitte	Produkt
35-44	3	0,10	40	4,00
45-54	7	0,23	50	11,67
55-64	9	0,30	60	18,00
65-74	5	0,17	70	11,67
75-84	4	0,13	80	10,67
85-94	0	0,00	90	0,00
95-105	2	0,07	100	6,67
SUMME	30	1	220	45,33
				<u>Mittelwert</u>

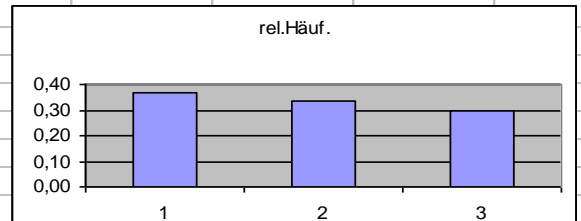


Andere Lösung mit Flächendarstellung ist

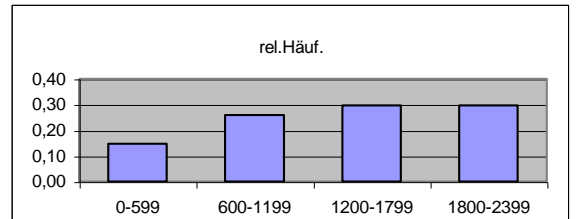


7)

Gehälter	abs.Häuf.	rel.Häuf.	Klassenmitte	Produkt
0-799	10	0,37	400	148,148148
800-1599	9	0,33	1200	400
1600-2399	8	0,30	2000	592,592593
SUMME	27	1	3600	1140,74074
			Klassen -	Mittelwert



Gehälter	abs.Häuf.	rel.Häuf.	Klassenmitte	Produkt
0-599	4	0,15	300	44,4444444
600-1199	7	0,26	900	233,333333
1200-1799	8	0,30	1500	444,444444
1800-2399	8	0,30	2100	622,222222
SUMME	27	1,00	4800	1344,44444
			Klassen -	Mittelwert



Mittelwert aus den Ausgangsdaten: 1343,19

