Thema: Skalenniveaus

Aufgabe 1:

Ordnen Sie die folgenden Merkmale den verschiedenen Skalierungen und Merkmalstypen zu:
Einkommen – Haarfarbe – Alter – soziale Stellung – Körpergröße – Geschlecht – Beruf – Schultypen – Anzahl von Kindern in Schulklassen – Raucher / Nichtraucher.

Skala / *Typen*	nominal	ordinal	metrisch (intervall / rational)
diskret			
stetig			

Lösung:

Skala / *Typen*	nominal	ordinal	metrisch (intervall / rational)
diskret	Haarfarbe, Beruf, Geschlecht, Raucher/Nichtraucher	Soz. Stellung, Schultypen	Anzahl von Kindern in Schulklassen
stetig			Alter, Einkommen, Körpergröße

Thema: Häufigkeitsverteilungen & Kreuztabellen

Aufgabe 1:

Antworten nach der Frage der allgemeinen wirtschaftlichen Lage (30 Befragte):
(1=sehr gut; 2=gut; 3=teils/teils; 4=schlecht; 5=sehr schlecht; 6=weiß nicht)

1, 3, 2, 2, 6, 5, 1, 3, 3, 4, 2, 1, 6, 5, 4, 2, 3, 4, 5, 1, 3, 3, 4, 2, 6, 1, 4, 3, 5, 1

a) stellen Sie die Antworten in einer Häufigkeitstabelle dar!
[absolute Häufigkeit f(x), relative Häufigkeit p(x), prozentuale Häufigkeit Proz(x), absolute kumulierte Häufigkeit, relative kumulierte Häufigkeit, prozentuale kumulierte Häufigkeit]

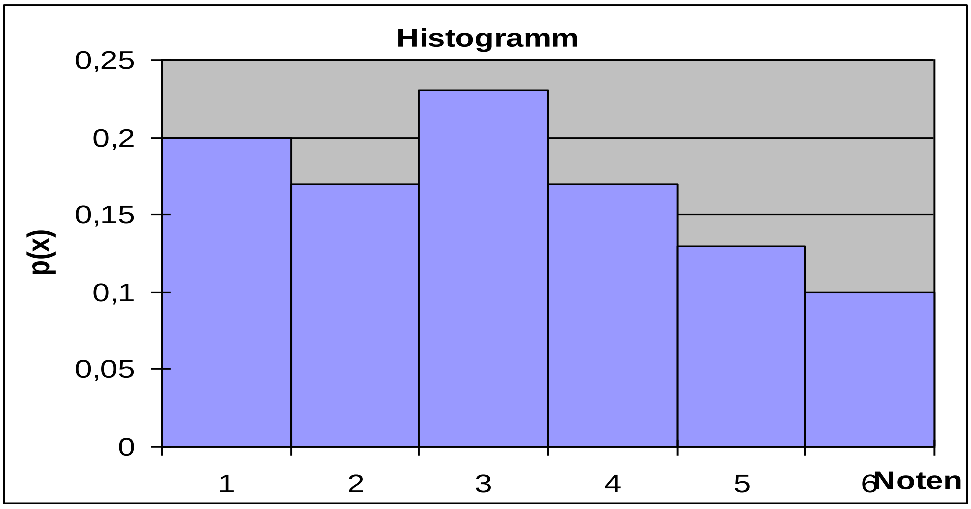
Lösung:

X	f(x)	p(x)	Proz(x)	f(x) kum.	p(x) kum.	Proz(x) kum.
1	6	0.2	20	6	0,2	20
2	5	0,17	17	11	0,37	37
3	7	0,23	23	18	0,6	60
4	5	0,17	17	23	0,77	77
5	4	0,13	13	27	0,90	90
6	3	0,1	10	30	1,00	100
Σ	30	1,00	100	-	-	-

b) wählen Sie eine geeignete graphische Darstellungsform und stelle das Ergebnis graphisch dar (1. nicht kumuliert, 2. kumuliert)!

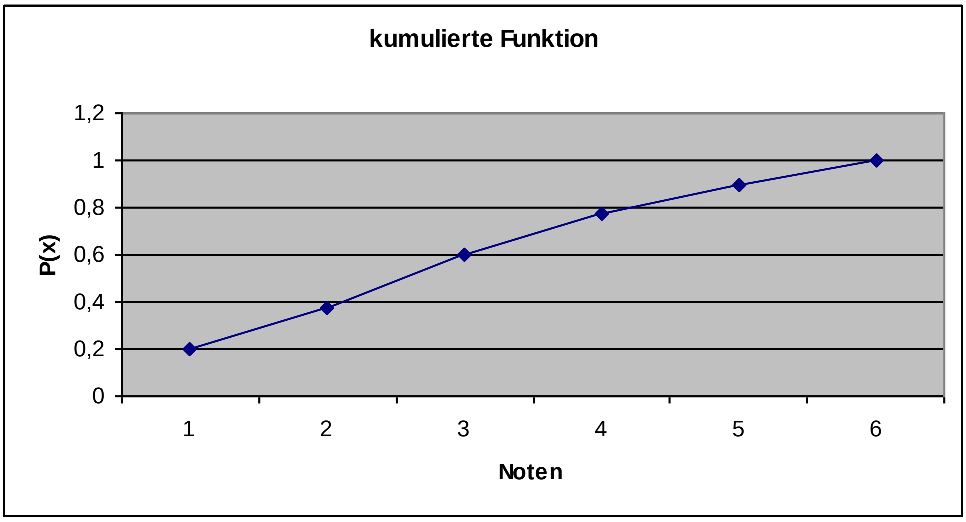
Lösung:

Nicht kumuliert:



#

Kumuliert:



Alternativ kann auch die „Weiß nicht“-Kategorie bei der Prozentuierung und Kumulation **nicht** berücksichtigt werden!

Aufgabe 2:

Folgende Alter wurden bei einer Umfrage festgestellt:
16, 56, 46, 48, 23, 32, 79, 19, 27, 30, 41, 38, 23, 51, 34, 65, 37, 54, 69, 17, 32, 43, 81, 24, 47, 42, 63, 21, 51, 36

a) Gruppieren Sie die Daten in geeignete Merkmalsklassen!

Lösung:

0-25 ; 25-35 ; 35-45 ; 45-55 ; >55

b) stellen Sie die Antworten in einer Häufigkeitstabelle dar! [f(x), p(x), Proz(x)]

Lösung:

	f(x)	p(x)	Proz(x)
0-25	7	0,23	23
>25-35	5	0,17	17
>35-45	6	0,2	20
>45-55	6	0,2	20
>55	6	0,2	20
Σ	30	1,00	100

Aufgabe 3:

Die verschiedenen Personen (Alter) gaben die folgenden Antworten auf die Frage der allgemeinen wirtschaftlichen Lage (die zweite Zahlenreihe gibt die Antwort auf einer Skala von 1 bis 6 nach dem Schulnotensystem an):

Alter	16	56	46	48	23	32	79	19	27	30	41	38	23	51	34	65	37	54	69	17	32	43	81	24	47	42	63	21	51	36
Antwort	1	3	2	2	6	5	1	3	3	4	2	1	6	5	4	2	3	4	5	1	3	3	4	2	6	1	4	3	5	1

a) stellen Sie die Daten in einer Kreuztabelle dar (Alter als gruppierte Daten aus Aufgabe 2)!

Lösung:

Alter ↓ 4Lage →	1	2	3	4	5	6	Σ
0-25	2	1	2	0	0	2	7
> 25 - 35	0	0	2	2	1	0	5
> 35 - 45	3	1	2	0	0	0	6
> 45 – 55	0	2	0	1	2	1	6
> 55	1	1	1	2	1	0	6
Σ	6	5	7	5	4	3	30

b) Annahme: Die verschieden Personen werden zusätzlich zwischen *männlich* und *weiblich* unterschieden. Wie könnte eine mögliche Kreuztabelle aussehen?

Lösung:

Alter ↓ Lage→	0-25		>25-35		>35-45		>45-55		>55		>55		Σ
	Mann	Frau	Mann	Frau	Mann	Frau	Mann	Frau	Mann	Frau	Mann	Frau	
0-25													
>25-35													
>35-45													

>45-55													
> 55													
Σ													

Aufgabe 4:

Folgende Hypothese wurde aufgestellt: „Frauen haben mehr Angst im Dunkeln das Haus zu verlassen!“

Auf die Aussage „Ich habe Angst davor im Dunkeln das Haus / die Wohnung zu verlassen“ machten die 93 Befragten (männlich und weiblich) folgende Angaben:

	Stimmt nicht-stimmt wenig	Stimmt teilweise	Stimmt ziemlich-stimmt sehr	Gesamt
männlich	24		13	
weiblich		9		
Gesamt	46		34	93

Erstellen Sie eine Kreuztabelle (=Kontingenztafel) und ermitteln Sie Zeilen- sowie Spaltenprozente und interpretieren Sie das Ergebnis.

Lösung:

Absolute Häufigkeiten:

	Stimmt nicht – stimmt wenig	Stimmt teilweise	Stimmt ziemlich – stimmt sehr	Σ
männlich	24	4	13	41
weiblich	22	9	21	52
Σ	46	13	34	93

Zeilenprozente:

	Stimmt nicht – stimmt wenig	Stimmt teilweise	Stimmt ziemlich – stimmt sehr	Σ
männlich	58	10	32	100
weiblich	42	17	41	100
Σ				

Spaltenprozente:

	Stimmt nicht – stimmt wenig	Stimmt teilweise	Stimmt ziemlich – stimmt sehr	Σ
männlich	52	31	38	
weiblich	48	69	62	
Σ	100	100	100	

Interpretation: In „Zeilenprozente“ Diagonale betrachten oder Ursachen vergleichen. Hier Diagonale links oben- rechts unten gering stärker besetzt, ebenfalls geringe Unterschiede zw. Männern und Frauen: „58% der Männer stimmen nicht oder wenig zu gegenüber 42% der Frauen; dagegen sagen nur 32% der Männer, dass es stimmt, gegenüber 41% der Frauen.

Aufgabe 5:

Aus der folgenden Tabelle ist die Religionszugehörigkeit (C: Christliche Religion; S: Sonstige) sowie die Zugehörigkeit zu den Berufsgruppen A: Angestellte, B: Beamte, F: Freiberufler ersichtlich.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Beruf	A	F	B	B	A	A	B	F	F	F	A	B	A	A	A	A	F	B	B	A
Religion	C	S	S	C	C	C	C	S	C	S	C	C	S	C	C	C	S	C	C	S

Erstellen Sie eine Kreuztabelle. Berechnen Sie Zeilen- sowie Spaltenprozente. Interpretieren Sie das Ergebnis.

Lösung:

Absolute Häufigkeiten:

Beruf ↓, Religion→	C	S	Σ
A	7	2	9
B	5	1	6
F	1	4	5
Σ	13	7	20

Spaltenprozente:

Beruf ↓, Religion→	C	S
A	54	29
B	38	14
F	8	57
Σ	100	100

Zeilenprozente:

Beruf ↓, Religion→	C	S	Σ
A	78	22	100
B	83	17	100
F	20	80	100

Interpretation: Man kann sehen, dass unter den Angestellten und Beamten mehr Christen sind, während die Freiberufler eher „sonstige" Religionszugehörigkeit haben. Das heißt, es gibt einen Zusammenhang zwischen Berufsgruppen und Religionszugehörigkeit (Zeilenprozente betrachten).

Aufgabe 6:

Die 50 Beschäftigten eines Betriebes haben während eines bestimmten Zeitraumes folgende Überstunden geleistet:

40	32	4	18	30	46	10	10	28	20
36	44	24	42	6	2	34	34	8	26
8	14	22	40	38	34	32	24	36	18
16	26	16	34	46	20	26	36	44	26
42	4	26	44	12	42	36	14	16	32

a) Stellen Sie aus diesen Angaben eine Häufigkeitsverteilung in den Größenklassen (von ... bis unter ...) 0-10, 10-20, 20-30, 30-40 und 40-50 auf.

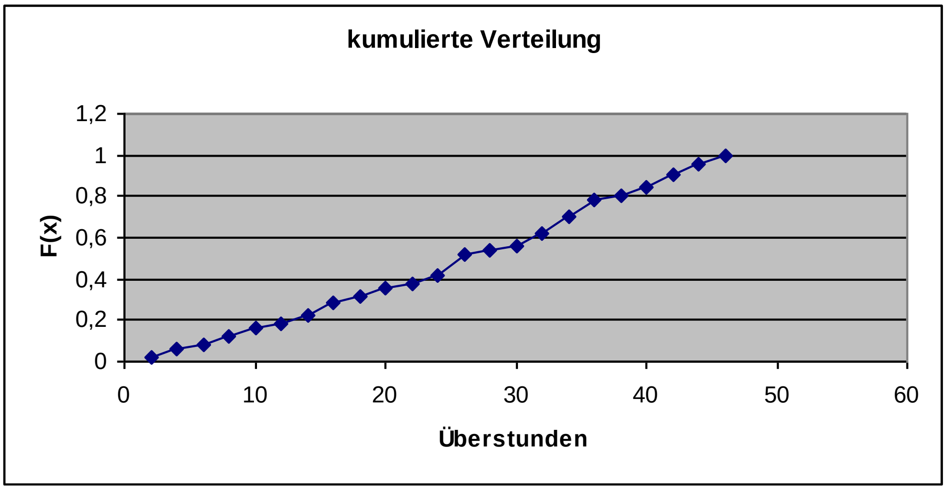
Lösung:

	f(x)	p(x)	Proz(x)	f(x) kum.	p(x) kum.	Proz(x) kum.

0-10	6	0,12	12	6	0,12	12
10-20	10	0,20	20	16	0,32	32
20-30	11	0,22	22	27	0,54	54
30-40	13	0,26	36	40	0,8	80
40-50	10	0,2	20	50	1	100
Σ	50	1,00	100	-	-	-

b) Zeichnen Sie die kumulierte Verteilung der Überstunden möglichst genau.

Lösung:



b) Bestimmen Sie den Anteil der Beschäftigten, die
- weniger als 20
- mindestens 20
- von 20 bis unter 40

Überstunden geleistet haben.

Lösung:

(1) $\mathbf{\frac{16}{50} = 0{,}32}$

(2) $\frac{34}{50} = 0{,}68$

(3) $\frac{24}{50} = 0{,}48$

Aufgabe 7:

In einer Bildungssoziologischen Studie wurde das Engagement der Schüler in Arbeitsgemeinschaften außerhalb der Schulzeit untersucht. Gefragt wurde dabei, ob die Schüler in keiner, einer, mehr als einer AG aktiv sind. Weiterhin wurde die subjektive Schichteinstufung der Eltern erhoben (Unterschicht, Mittelschicht, Oberschicht).
Folgende Ergebnisse sind aus der Studie bekannt:

- insgesamt 600 Schüler wurden befragt
- nur 120 von ihnen sind der Oberschicht zuzuordnen, dafür 200 der Unterschicht
- 280 der befragten Schüler nehmen an keiner AG teil
- 80 besuchen mehr als eine AG außerhalb der Schule

Erstellen Sie aus diesen Informationen eine Kreuztabelle besetzen Sie die leeren Zellen (absolute Zahlen und Prozentwerte) dabei so, dass gilt:

a) Je höher die subjektive Schichteinstufung, an umso mehr AG' s beteiligt sich der Schüler. (starker Zusammenhang)

Lösung:

	Nie	1	> 1	Σ
Unterschicht	200	0	0	200
Mittelschicht	80	200	0	280
Oberschicht	0	40	80	120
	280	240	80	600

b) Je höher die subjektive Schichteinstufung, an umso mehr AG' s beteiligt sich der Schüler. (schwacher Zusammenhang)

Lösung:

	Nie	1	> 1	Σ
Unterschicht	100	60	40	200
Mittelschicht	160	110	10	280
Oberschicht	20	70	30	120
	280	240	80	600

c) Besteht kein Zusammenhang zwischen den Variablen.

Lösung:

	Nie	1	> 1	Σ
Unterschicht	93	80	27	200
Mittelschicht	131	112	37	280
Oberschicht	56	48	16	120
	280	240	80	600

Thema: Lagemaße & Verteilungsformen

Aufgabe 1:

An einer bundesdeutschen Universität sind in den einzelnen Fakultäten im ersten Fachsemester folgende Anzahlen von Studierenden eingeschrieben:

Ingenieurwesen: 326
Recht: 224
Theologie: 60
Wirtschaftswissenschaften: 532
Philosophie: 89
Pädagogik: 128

Charakterisieren Sie diese Häufigkeitsverteilung durch geeignete Lagemaße!

Lösung:

Modus: Wirtschaftswissenschaften (532 Fälle)

Aufgabe 2:

In einem Unternehmen wurden bei 9 Mitarbeitern folgende Fehlzeiten, gemessen in Tagen, festgestellt: 9, 3, 13, 62, 12, 4, 12, 7, 2. Wie groß ist der Mittelwert?

Lösung:

Mittelwert: 13,78

Aufgabe 3:

Gegeben ist folgende Häufigkeitstabelle:

Altersklasse	Anzahl n
18 - <30,5	634
30,5 - <43	1233
43 - <55,5	934
55,5 - <68	723

Berechnen Sie geeignete Lagemaße!

Lösung:

Modus: Altersklasse 30,5 - <43 Jahren,

Median: Ebenfalls Altersklasse 30,5 - <43 Jahren
Interpolierter Median: Länge der Klasse: 12,5
Nr. des Medianfalls: insgesamt n = 3524, Medianfall = 1762,5
Distanz des Medianfalls in der Medianklasse: 1762,5 – 634 = 1128,5
Median: 30,5 + 1128,5/1233×12,5 = 30,5 + 11,44 = 41,94

Mittelwert: (24,25×634+36,75×1233+49,25×934+61,75×723) / 3524 = 42,94

Aufgabe 4:

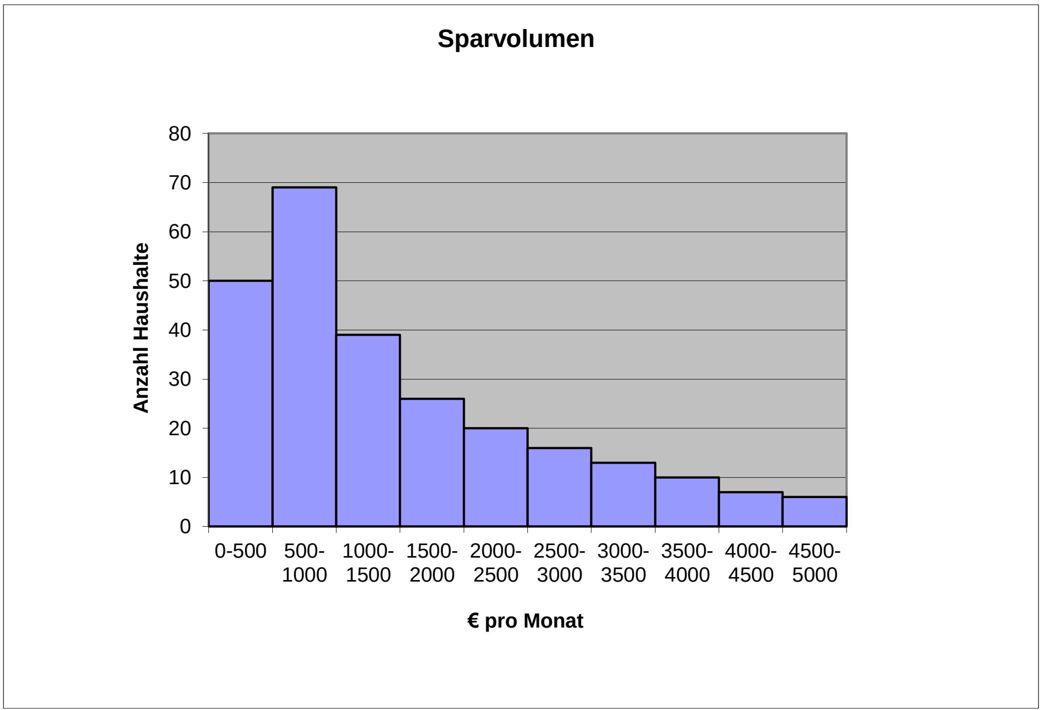
Aus einer Umfrage zur monatlichen „Sparsumme“ verschiedener Haushalte ergaben sich folgende Umfrageergebnisse (fiktiv):

„Sparsumme“ in €	**Häufigkeit**
0 - 500	50
>500 - 1000	69

>1000 - 1500	39
>1500 - 2000	26
>2000 - 2500	20
>2500 - 3000	16
>3000 - 3500	13
>3500 - 4000	10
>4000 - 4500	7
>4500 - 5000	6

Stellen Sie das Umfrageergebnis graphisch dar, berechnen Sie alle geeigneten Lagemaße und geben Sie die Art der Verteilungsform an!

Lösung:



Linkssteil, unimodal
Modus: Sparvolumen von 500–1000 €
Median: Klasse < 1000 – 1500, interpoliert: 1000 + 500 × (128,5 – 50 – 69) / 39 = 1121,79
Mittelwert: (250×50+750×69+…+4750×6) / 256 = 1505, 86

Aufgabe 5:

Die Lageparameter haben unterschiedliche Eigenschaften. Gehen Sie kurz auf diese ein (Skalenniveau, Empfindlichkeit gegenüber Ausreißern etc.).

Versuchen Sie sich anhand des folgenden Beispiels die Empfindlichkeit gegenüber Ausreißern zu verdeutlichen:

Folgende Werte stellen das Alter von 20 Befragten dar:

27	50	31	45	41	33	56	35	39	35	34	26	13	44	54	18	34	49	98	38

Berechnen Sie Mittelwert, Median und Modus und formulieren Sie jeweils einen Ergebnissatz ohne statistische Begriffe zu verwenden. Geben Sie die Art der Verteilungsform an.

Lösung:

Modus: geeignet für nominalskalierte Daten; bei gruppierten Daten muss auf die Messklasseneinteilung geachtet werden.
Median: Nur für Daten, die min. Ordinalskalenniveau haben. Ist robust gegenüber Ausreißern.
Mittelwert: Intervallskalenniveau nötig. Nicht robust gegenüber Ausreißern.

Beispiel: Der Größe nach ordnen:
13 18 26 27 31 33 34 34 35 35 38 39 41 44 45 49 50 54 56 98
Mittelwert: 40
Median: 36,5
Modus: bei 34 und 35 Jahren

Antwortsatz: Die Hälfte der Befragten ist jünger als 36 Jahre. Im Durchschnitt sind die Befragten 40 Jahre alt (Ausreißer Beeinflussung: 13 und 98 Jahre).

Art der Verteilungsform: Mittelwert > Median > Modus: linksssteil.

Aufgabe 6:

Folgende Werte stellen das Alter von 20 Befragten dar:

23	45	86	34	42	52	30	40	71	59	29	24	19	45	19	58	61	26	23	40

Berechnen Sie Mittelwert, Median und Modus und formulieren Sie jeweils einen Ergebnissatz ohne statistische Begriffe zu verwenden.

Lösung:

Beispiel: Der Größe nach ordnen:
19 19 23 23 24 26 29 30 34 40 40 42 44 45 52 58 59 61 71 86
Mittelwert: 41,25
Median: 40
Modus: bei 19; 23 und 40 Jahren

Antwortsatz: Die Hälfte der Befragten ist jünger als 40 Jahre. Im Durchschnitt sind die Befragten 41 Jahre alt.

Aufgabe 7:

Die Arbeitsgemeinschaften „Physik für Anfänger" und „Physik für dumme Anfänger" treten in dem Wettstreit „Auf zum Atom" gegeneinander an. Das Ergebnis der Aufgaben wird in Schulnoten zusammengefasst; hier der Notenspiegel für beide AG's:

Ergebnisse (Note) „Auf zum Atom"	**Absolute Häufigkeit „Physik für Anfänger"**	**Absolute Häufigkeit „Physik für dumme Anfänger"**
1	15	3
2	8	5
3	11	22
4	10	19
5	6	4
6	11	1
$\sum$	61	54

Berechnen Sie geeignete Lagemaße sowie den Quartilabstand. Interpretieren Sie die Ergebnisse vergleichend.

Lösung:

Physik für Anfänger:

Modus: Note 1
Median: Mittlerer Wert = 30,5. Liegt bei Note 3.
Mittelwert: 3,28
Q1 liegt beim 16. Wert = 2 ; Q3 liegt beim 46. Wert = 5: Quartilsabstand: 1,5

Physik für dumme Anfänger:

Modus: Note 3
Median: Liegt bei Note 3.
Mittelwert: 3,35
Q1= 3; Q3= 4: Quartilsabstand: 0,5

Die AG „Physik für dumme Anfänger" zeigt eine im Durchschnitt leicht schlechtere Leistung als die AG „Physik für Anfänger". Zudem weisen die „dummen Anfänger" eine homogenere Leistungsstruktur als die andere AG auf, bei der es zwar relativ viele sehr gute Mitglieder, aber eben auch im Vergleich zu den „dummen Anfängern" viele sehr schlechte und ein weniger stark ausgeprägtes Mittelfeld gibt.

Aufgabe 8 :

In der Praxis von Dr. Müller wurde die Wartezeit der Patienten beobachtet. Unter 10 Patienten ergab sich folgende Häufigkeitstabelle:

Wartezeit in Minuten	Absolute Anzahl
0 bis 5 Minuten	1
5 bis 10 Minuten	2
10 bis 12 Minuten	4
12 bis 14 Minuten	2
14 bis 20 Minuten	1

Berechnen Sie den Median. Benutzen Sie Interpolation.

Lösung:

a) nicht interpoliert: Median liegt in der Kategorie „10-12 Min".
b) interpoliert: exakte „mittlere" Person: 5,5te; exakte untere Fallgrenze der Kategorie „10-12Min.": 3,5te Person; Länge des zu interpolierenden Intervalls: 4 Fälle, 2 Min.
Median = 10 Min. + (5,5 – 3,5)/4 * 2 Min.= 11 Min.
Oder auch rein symmetrische Betrachtung: da die Verteilung vollkommen symmetrisch ist, muss der Median genau die Mitte der mittleren Klasse sein, und das sind 11 Min.

Aufgabe 9 :

Folgende Mittelwerte wurden bei einer Umfrage zur Einflussnahme durch Parteimitarbeit (Antwort-Skala: 1-7) ermittelt:

Neue Bundesländer: Anzahl 1005; Mittelwert: 2,92
Alte Bundesländer: Anzahl 2167; Mittelwert 3,51

Berechnen Sie den Mittelwert für „Gesamtdeutschland".

Lösung:

Mittelwert(Gesamtdeutschland) = (1005*2,92+2167*3,51)/(1005+2167) = 3,32

Aufgabe 10 :

Sie haben für Ihre Abschlussarbeit berufstätige Mütter und Väter hinsichtlich ihrer Meinung zum Thema „Berufstätigkeit und Zeit nehmen für die Kinder", befragt.

Folgende Daten haben Sie erhalten:

„ Es ist wichtig sich, trotz Berufstätigkeit für die Kinder Zeit zu nehmen."	**Codierung**	**Anzahl Väter**	**Anzahl Mütter**
Lehne völlig ab	1	5	2
Lehne ab	2	10	5
Lehne ein wenig ab	3	11	9
Weder noch	4	6	17
Stimme ein wenig zu	5	19	23
Stimme zu	6	29	31
Stimme voll und ganz zu	7	20	45

a) Berechnen Sie Modus und Median für Männer und Frauen. Formulieren Sie je einen Antwortsatz, ohne statistische Begriffe zu verwenden.
b) Welche Verteilungsform liegt jeweils vor? Begründen Sie ihre Antwort.
c) Bestimmen Sie den Quartilsabstand für Männer und Frauen und interpretieren Sie den Unterschied.

Lösung:

a) Männer: n = 100: D = 6 = Stimme zu, Median = 5 = stimme ein wenig zu
Frauen: n = 132: D = 7 =„ Stimme voll und ganz zu"; Median = 6 = „ stimme zu"
Die Hälfte der Männer stimmt mindestens „ein wenig" zu, während die Hälfte der Frauen mindestens zustimmt. Am häufigsten ist bei den Männern die einfache Zustimmung, dagegen bei den Frauen die Zustimmung „voll und ganz". Im Vergleich insgesamt stimmen also die Frauen stärker zu als die Männer.

b) Es liegt eine in beiden Fällen eine rechtssteile Verteilungsform vor, da Median < Modus. (Bimodal ist auch möglich [für Väter])

c) Männer. Q1 = 3, Q3 = 6, QA = 1,5. Frauen: Q1 = 5, Q3 = 7, QA = 1. Die Frauen sind sich in dieser Frage einiger als die Männer, da diese eine um eine halbe Kategorie größere Streuung aufweisen. (möglich für Frauen ist auch QA=1,25; oder: Formel und Hauptsache kleiner als bei Vätern!)

Aufgabe 11

Eine biologische Untersuchung beschäftigt sich mit der Zahl der vorkommenden Schmetterlingsarten in verschiedenen Lebensraumtypen in Schleswig Holstein. Es wurden jeweils 20 Waldgebiete, 20 Wiesen- bzw. Weidenflächen sowie 20 Moor- und Sumpfgebiete untersucht.
Die Urliste zeigt folgende Werte für die Zahl der Schmetterlingsarten in den verschiedenen Lebensräumen.

Fall	Wiesen/Weiden	Wald	Moore/ Sümpfe
1	5	0	0
2	5	0	0
3	7	1	1
4	10	1	1
5	10	1	2
6	10	2	2
7	11	2	2
8	11	2	3
9	11	2	4
10	11	3	5
11	12	3	6
12	13	3	7
13	13	3	7
14	13	4	7
15	13	5	8
16	15	6	9
17	15	7	9
18	15	6	9
19	15	6	9
20	16	14	9

a) Welches Skalenniveau liegt vor?
b) Berechnen Sie bitte dem Skalenniveau entsprechend alle möglichen Lagemaße.
c) Welche Verteilungsform liegt vor? Bitte begründen Sie Ihre Antwort.
d) Wenn Sie bei einem Sonntagsausflug möglichst viele verschiedene Schmetterlingsarten beobachten wollen, welchen Lebensraumtyp würden Sie aufsuchen? Bitte begründen Sie Ihre Antwort.

Lösung:

Die Aufgabe ist etwas unklar, weil nicht genau gesagt wird, ob b) und c) pro Lebensraum oder insgesamt berechnet werden sollen. Hier wird von pro Lebensraum ausgegangen.

a) Skalenniveau:
Merkmal „Anzahl Schmetterlingsarten“ = metrisch.
Merkmal „Lebensraum“ = nominal (Wiesen, Wald, Moor)

b) „Anzahl Schmetterlinge“
Wiese: Modus D = 11, 13 und 15 (alle gleich häufig); Median X~ = 11,5 (zwischen 10. und 11. Wert); Mittelwert X- = 231/20 = 11,55
Wald: Modus D = 3 und 2; Median X~ = 3 ; Mittelwert X- = 71/20 = 3,55

Moor: Modus D = 9; Median X~ = 5,5 ; Mittelwert X- = 100/20 = 5

c) Verteilungsform:
Wiese: symmetrisch (alle3 Lagemaße etwa gleich)
Wald: linkssteil (D < X~ < X-)
Moor: rechtssteil (X- < X~ < D)

d) Wald, weil im Durchschnitt dort die meisten sind und die Verteilung gleichmäßiger ist.

Thema: Streuungsmaße & Verteilungsmodelle

Aufgabe 1:

Von 13 Studenten wurde die Körpergröße ermittelt. Die Merkmalsausprägungen (Körpergröße in Meter) wurden bereits der Größe nach geordnet.

1,60	1,67	1,67	1,68	1,68	1,70	1,70	1,72	1,73	1,75	1,76	1,78	1,84

a) Berechnen Sie Median und Mittelwert.

Lösung:

Mittelwert: 1,7138 m; Median: 1,70 m

b) Berechnen Sie den Quartilabstand. Formulieren Sie einen Antwortsatz.

Lösung:

QA = 0,035 m → Die Körpergröße der mittleren 50%-Studenten differiert maximal 0,07m und weicht im Durchschnitt um 0,035 m vom Median ab.

Aufgabe 2:

Ein Produzent eines Golfschlägers hat eine verbesserte Ausgabe hergestellt. In einem Experiment wurde bei 10 Testschlägen die Abschlagweite notiert. Die Weite in Metern ist der folgenden Tabelle zu entnehmen:

150	149	155	145	153	154	147	152	149	153

Berechnen Sie Standardabweichung und Varianz.

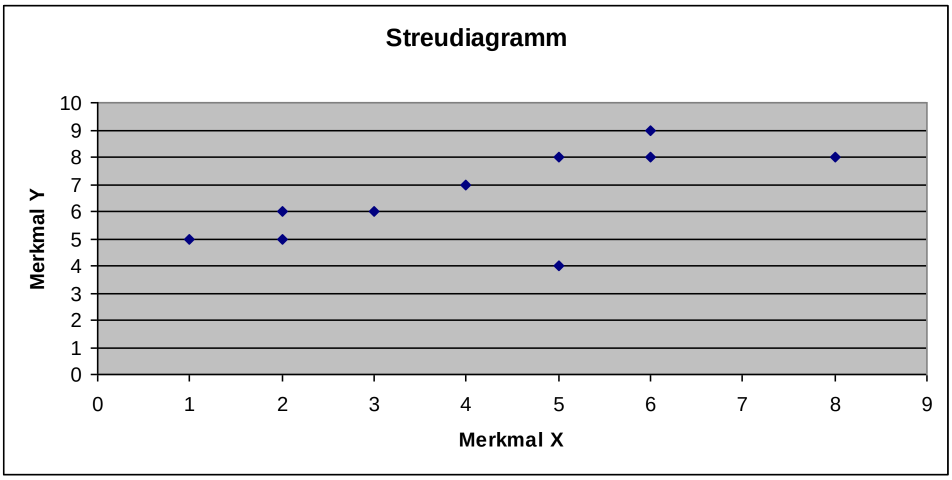
Lösung:

$s = 3{,}23$
$s^2 = 10{,}46$

Aufgabe 3:

Gegeben sind die folgenden Beobachtungspaare zweier metrisch messbarer Merkmale X und Y: (4;7), (1;5), (2;5), (2;6), (3;6), (5;4), (5;8), (6,8), (6;9), (8;8).
Zeichnen Sie ein Streuungsdiagramm und berechnen Sie die Kovarianz sowie den Korrelationskoeffizienten. Interpretieren Sie die Ergebnisse.

Lösung:



r = 0,668
s_{xy} = 2,422

Es besteht ein starker, gleichgerichteter Zusammenhang zwischen dem Merkmal X und dem Merkmal Y.

Aufgabe 4: Die Korrelation zwischen der Statistik-Note und der Germanistiknote betrage im Durchschnitt 0.4. An der Universität Flensburg ist die Standardabweichung der Statistiknote doppelt so groß wie die durchschnittliche Standardabweichung der Statistiknote, sonst sind alle Werte wie im Durchschnitt. Wie groß ist die Korrelation zwischen den beiden Noten an der Uni Flensburg?

Lösung: sei sx = Standardabweichung Statistiknote, sy = Standardabweichung Germanistiknote, sxy = Kovarianz zwischen Statistiknote und Germanistiknote. Dann ist

r (allgemein) = sxy / (sx * sy) = 0,4; r (Flensburg) = sxy / (2sx * sy)= 0,4* 1/2 = 0,2

Aufgabe 5:

In der Praxis von Dr. Müller wurde die Wartezeit der Patienten beobachtet. Unter 10 Patienten ergab sich folgende Häufigkeitstabelle:

Wartezeit in Minuten	Absolute Anzahl
5 Minuten	1
10 Minuten	2
12 Minuten	4
14 Minuten	2
20 Minuten	1

Berechnen Sie den Quartilabstand. Formulieren Sie einen Antwortsatz.

Lösung:

Wartezeit in Minuten	Absolute Anzahl	F(x)
5 Minuten	1	1
10 Minuten	2	3
12 Minuten	4	7
14 Minuten	2	9
20 Minuten	1	10

$$Q_1 Pos:3 \rightarrow Q_1 = 10 Minuten$$
$$Q_3 Pos:8 \rightarrow Q_3 = 14 Minuten$$
$$Q_A = 2 Minuten$$

Aufgabe 6

Von 3000 Personen wurde ihr durchschnittliches Einkommen (X) und die Anzahl der Krankheitstage (Y) in den letzten 5 Jahren erfasst. Es ergaben sich folgende Kennwerte:

$$\bar{x} = 2500\, Euro, s_x = 1000\, Euro; \;\; \bar{y} = 30\, Tage, s_y = 20\, Tage$$

Welche der beiden Merkmale variiert stärker?

Lösung:

Variationskoeffizienten berechnen:

$$V_x = \frac{1000}{2500} = 0{,}4 = 40\,\%; \quad V_y = \frac{20}{30} = 0{,}67 = 67\%$$

Die StdAbw von X beträgt 40% des Mittelwerts, von Y dagegen 67%. >Y variiert stärker.

Thema: Normalverteilung & zentraler Grenzwertsatz

Aufgabe 1:

Die Köpergröße eines bestimmten Jahrgangs ist normalverteilt mit den Werten μ = 95 cm und σ=7cm.
(Man sagt dazu auch „die Körpergröße ist N(95cm, 7cm) verteilt.)
Wie viel Prozent dieser Kinder sind im Mittel

a) kleiner als 1 m,

Lösung:
X: Körpergröße
X~ N(95;7)

1 m = 100 cm

F(100) = F((100-95)/7) = 0,76 → 76% sind kleiner als 1 m.

b) größer als 1,05m,

Lösung:

1,05 m = 105 cm

F(105) = F((105-95)/7) = 0,92 → 1-0,92 = 0,08 → 8% sind größer als 1,05 m.

c) zwischen 88 cm und 103 cm?

Lösung:

F(103) - F(88) = F((103-95)/7) – F ((88-95)/7) = 0,87 – 0,16 =0,71 → 71% sind zwischen 88 cm und 103 cm groß.

Aufgabe 2:

Die erwartete Segelfahrtdauer zwischen Frederikshavn und Göteborg ist 3 Stunden (180 Minuten) und eine Standardabweichung von 15 Minuten. Es wird angenommen, dass die Segelfahrtdauer eine Normalverteilung hat.

a) Berechnen Sie den Anteil der Segelfahrten, die mehr als 215 Minuten dauern.
Ein Vertreter fährt im Durchschnitt 10 Mal pro Vierteljahr mit der Fähre.

Lösung:

X : Segelfahrtdauer
X~N(180;15)

F(215) = F((215-180)/15) = 0,99 → 1- 0,99 = 0,01 → 1% der Fahrten dauert länger als 215 Minuten

b) Berechnen Sie den Anteil von Stichproben der Größe 10, in denen die durchschnittliche Fahrdauer mehr als 215 Minuten beträgt.

Lösung:

$$\overline{X} \sim N(180; \frac{15}{\sqrt{10}})$$

$$\overline{X} \sim N(180;4,74)$$

F(215) = F((215 -180)/4,74) = 1 → 1-1 = 0 → 0% beträgt der Anteil der Stichproben

c) Sie nehmen viele Stichproben der Größe 10 und berechnen jeweils die durchschnittliche Segelfahrtsdauer. In welchem Schwankungsintervall liegen 95% dieser durchschnittlichen Segelfahrtsdauer.

Lösung:

SI= [$z_{\alpha/2}$*σ+μ; $z_{1-\alpha/2}$*σ+μ]
SI= [-1,96*15/√10+180; 1,96*15/√10+180]
SI= [173,8019; 186,1981]

Aufgabe 3:

Bei einer Klausur mit einer maximalen Punktzahl von 100 seien die Ergebnisse (näherungsweise) normalverteilt mit μ=60 und σ=10. Bestimmen Sie den Anteil der Studenten

a) die durchgefallen sind, wenn zum Bestehen der Klausur mindestens 50 Punkte erforderlich sind.

Lösung:

X: Ergebnisse einer Klausur
X~N(60;10)
F(50) = F((50-60)/10) = 0,16 → 16% sind durchgefallen

b) die die Note „gut“ erhalten, wenn diese für Punktzahlen von 80 bis 95 (jeweils einschließlich) vergeben wird.

Lösung:

Lösung:
F(95) – F(80) = F((95-60)/ 10) – F((80-60)/10) = F(3,5) – F(2,0) = 1 – 0,980 = 0,02 → 2%

Oder interpoliert: 0,9975 – 0,9775 = 0,02 → 2%

c) Auf welchen Wert muss die Mindestpunktzahl festgelegt werden, wenn nicht mehr als 10% der Studenten durchfallen sollen?

Lösung:

$F(X_{0,1}) = 0,1 \rightarrow X_{0,1} = 60+Z_{0,1}*10 = 60 + (-1,28 *10) = 47,2 \rightarrow$ die Mindestpunktzahl müsste 47,2 Punkte betragen.

Thema: Wahrscheinlichkeiten & Binomialverteilung

Aufgabe 1:

Von 13 Schülern gehen 4 zum Gymnasium, 5 zur Realschule und 4 zur Hauptschule. Die Intelligenzquotienten (IQ) dieser Schüler mögen lauten:

Schulart	Schüler-Nr.	IQ
Gymnasium (G)	1	100
	2	104
	3	110
	4	120
Realschule (R)	5	90
	6	92
	7	95
	8	99
	9	102
Hauptschule (H)	10	86
	11	87
	12	93
	13	95

Bilden Sie folgende Ereignisse und berechnen Sie die Wahrscheinlichkeiten:

a. Geht zum Gymnasium. **Lösung: G = {1,2,3,4}; P(G) = 4/13**
b. Geht zur Realschule. **Lösung: R={5,6,7,8,9}; P(R) = 5/13**
c. Geht zur Hauptschule. **Lösung: H={10,11,12,13}; P(H) = 4/13**
d. Hat IQ größer gleich 100. **Lösung: GIQ={1,2,3,4,9}; P(GIQ) = 5/13**
e. Hat IQ kleiner gleich 90. **Lösung: KIQ={5,10,11}; P(KIQ) = 3/13**
f. Geht zur Realschule und hat IQ größer gleich 100
Lösung: (R ∩ GIQ)={9}; P(R ∩ GIQ) = 1/13
g. Geht zur Realschule oder hat IQ größer gleich 100
Lösung: (R U GIQ)={1,2,3,4,5,6,7,8,9}; P(R U GIQ) = 9/13
h. Geht zur Hauptschule und hat IQ größer gleich 100
Lösung: (H ∩ GIQ)={Φ}; P(H ∩ GIQ) = 0
i. Geht zur Hauptschule oder hat IQ größer gleich 100

Lösung: $(H \cup GIQ)=\{1,2,3,4,9,10,11,12,13\}$; $P(H \cup GIQ) = 9/13$

j. Geht zum Gymnasium und hat IQ größer gleich 100
Lösung: $(G \cap GIQ)= G$; $P(G \cap GIQ) = 4/13$

k. Geht zum Gymnasium oder hat IQ größer gleich 100
Lösung: $(G \cup GIQ)= GIQ$; $P(G \cup GIQ) = 5/13$

l. Geht zum Gymnasium und Hauptschule
Lösung: $(G \cap H)=\{\Phi\}$; $P(G \cap H) = 0$

m. Geht zum Gymnasium oder Realschule
Lösung: $(G \cup R)=\{1,2,3,4,5,6,7,8,9\}$; $P(G \cup R) = 9/13$

n. Geht nicht zur Realschule
Lösung: $(\neg R)=\{1,2,3,4,10,11,12,13\}$; $P(\neg R) = 8/13$

o. Geht zum Realschule oder Gymnasium und hat IQ größer gleich 100
Lösung: $(GIQ \cap (G \cup R))=[\{1,2,3,4,9\} \cap \{1,2,3,4,5,6,7,8,9\}] = \{1,2,3,4,9\}$;
$P(GIQ \cap (G \cup R)) = 5/13$

p. Ereignisraum (Ω). **Lösung: $\Omega = \{1,2,3,4,5,6,7,8,9,10,11,12,13\}$; $P(\Omega) = 1$**

Aufgabe 2:

Sie wählen 30 Personen zufällig aus. Wie groß ist die Wahrscheinlichkeit, dass mindestens 2 davon den gleichen Geburtstag haben?

Lösung: A={Mindestens 2 der 30 ausgewählte Personen haben Geburtstag an der gleichen Tag}
B_i={Ausgewählte Person i hat ein anderer Geburtstag als der i-1 anderer die schon ausgewählt wurden}

$P(A) = 1- P(B_1 \cap B_2 \cap B_3 \ldots\ldots\ldots\ldots \cap B_{29} \cap B_{30})$
$P(A) = 1- P(B_1)^* P(B_2)^* P(B_3)^* \ldots\ldots\ldots\ldots P(B_{29})^* P(B_{30})$
$P(A) = 1-[1*364/365*363/365*362/365*\ldots\ldots*337/365*336/365) =$
$P(A) = 1 – [(365*364*363*362*\ldots\ldots\ldots*337*336)/365^{30}]$
$P(A) = 1 – 0.29 = 0.71$

Aufgabe 3:

In einer kleinen Stadt gibt es 15366 erwachsene Frauen und 14634 erwachsene Männer. Von den Männer sind 6% und von den Frauen 8% arbeitslos.

a) Ein Erwachsener ist ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass diese Person arbeitslos ist?

Lösung:

A={Person ist arbeitslos}
M={Es ist einen Mann}
 # Frauen = 15366
 # Männer = 14634
 # Erwachsene = 15366+14634 = 30000
% Arbeitslose (Frauen) = 0.08 → # Arbeitslose (Frauen) = 0.08*15366 = 1229
% Arbeitslose (Männer) = 0.06 → # Arbeitslose (Männer) = 0.06*14634 = 878
 # Arbeitslose = 1229+878 = 2107
$P(A) = P(A \cap M) + P(A \cap \neg M) = 878/30000+1229/30000 = 0.07$

Lösung mit Baumdiagramm:

30000 Personen

14634 M **15366 F**

878 AL **13756** **1229 AL** **14137**

P(Person ist arbeitslos) = (878+1299)/30000 = 0,07

b) Wenn man eine arbeitslose Person auswählt, wie groß ist die Wahrscheinlichkeit, einen Mann ausgewählt zu haben? Und eine Frau?

Lösung:

C={Der arbeitlose ist einen Mann}

P(C) = 878/2107 = 0.42

D={Der arbeitlose ist eine Frau}

P(D) = 1229/2107 = 0.58

Aufgabe 4:

Es sei bekannt, dass es in einem Studentenwohnheim 41 IIM, 76 BA, 50 EW und 10 MAG Studenten gibt. Man wählt 10 Studenten zufällig aus. Wie groß ist die Wahrscheinlichkeit, dass man gar keinen BA Student auswählt? Wählen Sie aus mit Zurücklegen!

Lösung:

 # IIM = 41; # BA = 76; # EW = 50; # MAG = 10 → Total=177

Experiment: 10 Studenten werden ausgewählt

A={Man wählt kein BA Student}

X=# der BA Studenten die ausgewählt werden (von 10)

X~B(10,76/177) → X~B(10,0.43)

P(A) = P(X=0) = 0.0025

Binomial Tabelle

	X~B(n,p) ; P(X=k)								
n	**10**	***p***							
k	**0,08**	**0,09**	**0,1**	**0,15**	**0,2**	**0,25**	**0,3**	**0,45**	**0,5**
0	0,4344	0,3894	0,3487	0,1969	0,1074	0,0563	0,0282	0,0025	0,0010
1	0,3777	0,3851	0,3874	0,3474	0,2684	0,1877	0,1211	0,0207	0,0098
2	0,1478	0,1714	0,1937	0,2759	0,3020	0,2816	0,2335	0,0763	0,0439
3	0,0343	0,0452	0,0574	0,1298	0,2013	0,2503	0,2668	0,1665	0,1172
4	0,0052	0,0078	0,0112	0,0401	0,0881	0,1460	0,2001	0,2384	0,2051
5	0,0005	0,0009	0,0015	0,0085	0,0264	0,0584	0,1029	0,2340	0,2461
6	0,0000	0,0001	0,0001	0,0012	0,0055	0,0162	0,0368	0,1596	0,2051
7	0,0000	0,0000	0,0000	0,0001	0,0008	0,0031	0,0090	0,0746	0,1172
8	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0014	0,0229	0,0439
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0042	0,0098
10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0010

Aufgabe 5:

Es sei bekannt, dass 20% aller Mathe Studenten, die die Abschluss-Klausur mitmachen, durchfallen. Nach der Klausur man trifft eine Gruppe von 5 Leuten, die gerade die Klausur gemacht habe.

a) Wie hoch ist die Wahrscheinlichkeit, dass alle Studenten in der Gruppe bestanden haben?

Lösung:

A={Mathe Student fällt durch} P(A) = 0.20;

N=5

X= # der Mathe Studenten die durchgefallen sind (von 5)

X~B(5,0.2)

B={Alle 5 Studenten haben bestanden}

P(B) = P(X=0) = 0.3277

b) Wie hoch ist die Wahrscheinlichkeit, dass alle durchgefallen sind?

Lösung:

C={Alle 5 Studenten sind durchgefallen}

P(C) = P(X=5) = 0.0003

c) Berechnen Sie die Wahrscheinlichkeit, dass maximal 2 Leute durchgefallen sind.

Lösung:

D={Maximal 2 von den 5 Studenten sind durchgefallen}

P(D) = P(X=0) + P(X=1) + P(X=2) = 0.3277 + 0.4096 + 0.2048 = 0.9421

Binomial Tabelle

	X~B(n,p) ; P(X=k)								
n	**5**	***p***							
k	**0,08**	**0,09**	**0,1**	**0,15**	**0,2**	**0,25**	**0,3**	**0,45**	**0,5**
0	0,6591	0,6240	0,5905	0,4437	0,3277	0,2373	0,1681	0,0503	0,0313
1	0,2866	0,3086	0,3281	0,3915	0,4096	0,3955	0,3602	0,2059	0,1563
2	0,0498	0,0610	0,0729	0,1382	0,2048	0,2637	0,3087	0,3369	0,3125
3	0,0043	0,0060	0,0081	0,0244	0,0512	0,0879	0,1323	0,2757	0,3125
4	0,0002	0,0003	0,0005	0,0022	0,0064	0,0146	0,0284	0,1128	0,1563
5	0,0000	0,0000	0,0000	0,0001	0,0003	0,0010	0,0024	0,0185	0,0313

Aufgabe 5:

Die Wahrscheinlichkeit, dass ein AKW einen Supergau erlebt, ist nach Berechnungen der Reaktorsicherheitskommission 1 Mal in 10000 Jahren. Es werde angenommen, dass ein Supergau jederzeit passieren kann (z.B. wg. Naturkatastrophen), es liege also eine Gleichverteilung vor, in jedem Jahr sei es gleich wahrscheinlich. Es laufen weltweit 400 Atomreaktoren. Wie groß ist die Wahrscheinlichkeit, dass innerhalb von 25 Jahren mindestens eines dieser AKWs einen Supergau hat?

Lösung:

mit Binomialverteilung:
P(Mindestens eins) = 1 – P(keins).
Zu berechnen: P (keins innerhalb von 25 Jahren).

Für ein AKW gilt:
Erwartungswert eines Supergaus ist 1 innerhalb von 10000 Jahren
P (Supergau innerhalb eines Jahres) = 1/10000
P (Supergau innerhalb von 25 Jahren) = 25 * 1/10000 = 1/400
P (kein Supergau innerhalb von 25 Jahren) = 399/400

Für 400 AKWs, die unabhängig sind, gilt dann:
P (AKW1 „und“ AKW2 „und“…AKW400 jeweils kein Supergau in 25 Jahren) = $(399/400)^{400}$=0,3674

Für die Gegenw., dass mindestens 1 AKW explodiert, gilt dann
P(„mindestens 1“) = 1 – 0,3674 = 0,6326 ≈ 63,26 %

Anmerkung: Korrekt müsste diese Wahrscheinlichkeit mit der Poissonverteilung berechnet werden. Beide Berechnungen führen aber zu demselben Ergebnis, weil die Poissonverteilung die Grenzverteilung der Binomialverteilung für große n und kleine p ist.

Thema: Konfidenzintervalle

Aufgabe 1:

Eine bildungsökonomische Studie, die sich unter anderem mit materiellen Investitionen der Eltern in die Bildung ihrer Kinder im Bereich Sekundarstufe I beschäftigt, hat das Folgende ergeben: Eltern von Kindern, die das Gymnasium besuchen, geben im Durchschnitt 150 Euro pro Jahr für Arbeitsmaterialien, Schulbücher und Nachhilfeunterricht aus. Bei den Eltern der Realschüler sind dies im Durchschnitt 125 Euro mit einer Standardabweichung von 40 Euro.
Es wurden für beide Schulformen je 500 Elternteile befragt.

a) Berechnen Sie für die Realschüler das Konfidenzintervall zum üblichen Vertrauensniveau!

Lösung:

$$\bar{x} = 125$$

$$KI = [-1{,}96 * \frac{40}{\sqrt{500}} + 125; 1{,}96 * \frac{40}{\sqrt{500}} + 125]$$

$$95\% - KI[121{,}49; 128{,}51]$$

b) Unterstützen die Ergebnisse die weit verbreitete Vermutung, dass eine gymnasiale Ausbildung den Eltern pro Jahr finanziell mehr abverlangt als der Besuch einer Realschule?

Lösung:

Ja, denn die 150 € von Eltern deren Kinder das Gymnasium besuchen liegen nicht im Konfidenzintervall. Eine gymnasiale Ausbildung verlangt von den Eltern finanziell mehr ab als der Besuch einer Realschule

---

Aufgabe 2:

In einer großen Realschule in Musterstadt wurden 195 Schüler zum Thema „Alkohol unter Jugendlichen" gemacht. 39 Schüler weigerten sich an der Umfrage teilzunehmen.
Frühere Schätzungen ergaben, dass 25% aller Schüler sich zu diesem Thema nicht äußern wollen.
Ist die Behauptung gerechtfertigt, dass an der Realschule in Musterstadt mehr Schüler an der Umfrage teilnahmen (95%-KI)?

Lösung:

$$\hat{p} = \frac{39}{195} = 0{,}2$$

$$KI = \left[-1{,}96 \cdot \sqrt{\frac{0{,}2 \cdot 0{,}8}{195}} + 0{,}2; 1{,}96 \cdot \sqrt{\frac{0{,}2 \cdot 0{,}8}{195}} + 0{,}2\right]$$

$$95\% - KI : [0{,}144; 0{,}256]$$

Nein die Behauptung ist nicht gerechtfertigt, denn die 25 % der früheren Schätzung liegen ebenfalls in dem Konfidenzintervall.

Aufgabe 3:

In einer Untersuchung über die Körpergrößen von Frauen wurde eine repräsentative Stichprobe von 25 Frauen gezogen. Es ergab sich ein Mittelwert von 168cm und eine Standardabweichung von 5 cm. Berechnen Sie das 95%-Konfidenzintervall, sowie das 99%-Konfidenzintervall für den Mittelwert der Stichprobe, vergleichen Sie die Länge der beiden Intervalle und fassen Sie den Vergleich in Worte.

Lösung:

95%-KI: [166;170]
99%-KI: [165,4;170,6]
Durch Verringerung der Irrtumswahrscheinlichkeit (100-95 bzw. 100-99) wird das Intervall, indem der Wert liegt größer.