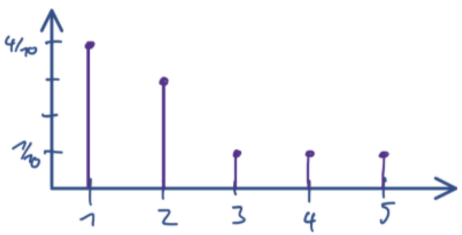
Deskriptive Statistik

Grundbegriffe

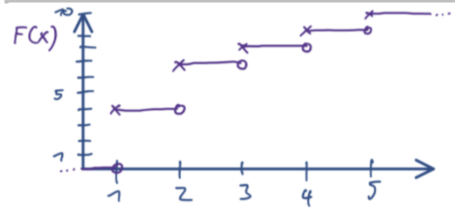
PMF:	$f(x)$ Relative Häufigkeit (Stabdiagramm)
CMF:	$F(x)$ Kumulative relative Häufigkeit (Treppendiagramm)
PDF:	$f(x)$ Höhe Balken Histogramm
CDF:	$F(x)$ Kulutaive Fläche Balken Histogramm
h_i:	Absolute häufigkeit
f_i:	Relative häufigkeit
x_{med}:	2. Quantil oder $R_{0.5}$
x_{mod}:	*Modus* oder *Modalwert* ist der häufigste Stichprobenwert
$\overline{x}$:	arithmetisches Mittel
s_x^2:	Varianz
s_x:	Standardabweichung
s_{kor}:	korrigierte Standardabweichung

Funktionstypen

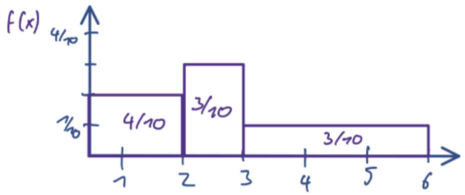
PMF



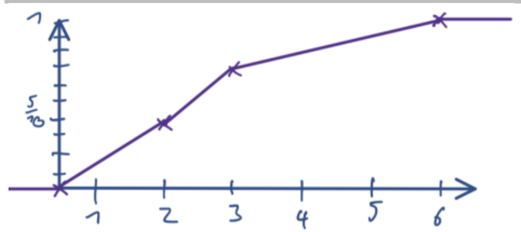
CMF



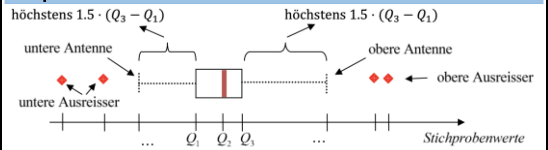
PDF



CDF



Boxplot



Untere und obere Antenne sind Stichprobenwerte!

Median bei GERADER Anzahl von Werten

5 7 9 12 | 13 13 14 30

Median: 12.5

Median bei UNGERADER Anzahl von Werten

5 7 9 12 13 13 14 16 30

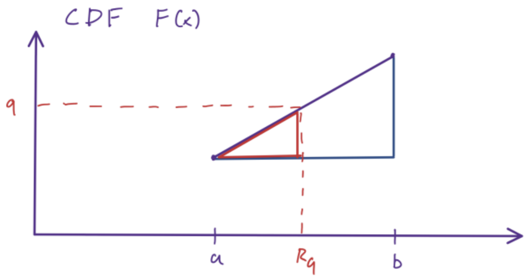
Median: 13

Eingabe in Taschenrechner

Lists & Spreadsheet öffnen
Spalte oben beschriften
Werte in Spalte eintragen
Doc → 4 Einfügen → 7 Data & Statistics
Unten auf «Klicken für mehr Variablen» und auswählen
Menu → 1 Plot-Typ → 2 Box-Plot

Quantil

Quantile CDF:



Zur Berechnung von R_q sucht man diejenige Klasse $[a, b[$ mit $F(a) \leq q \leq F(b)$ b

- $m = \frac{q-F(a)}{R_q-a}$
- $R_q = a + (b-a) \cdot \frac{q-F(a)}{F(b)-F(a)}$
- $q = F(a) + \frac{F(b)-F(a)}{b-a} \cdot (R_q - a)$
- $q = F(a) + f(R_q) \cdot (R_q - a)$

1!	1	6!	720
2!	2	7!	5040
3!	6	8!	40320
4!	24	9!	362880
5!	120	10!	3628800

Aufgaben

Klassierte Daten

Klasse	AbsH	SummeAbsH	RelH	PDF	CDF
20-30	8	8	$\frac{8}{40}$	$\frac{8}{400}$	$\frac{8}{40}$
30-50	10	18	$\frac{10}{40}$	$\frac{10}{800}$	$\frac{18}{40}$
50-60	18	36	$\frac{18}{40}$	$\frac{18}{400}$	$\frac{36}{40}$
60-80	4	40	$\frac{4}{40}$	$\frac{4}{800}$	$\frac{40}{40}$

AbsH -> Absolute Häufigkeit
SummeAbsH -> Summe Absolute Häufigkeit
RelH -> Relative Häufigkeit

P

AbsH : (Anzahl * Klassenbreite)

CDF

Mittelwert $\bar{x}$

$$\bar{x} = \frac{1}{40}(25 \cdot 8 + 40 \cdot 10 + 55 \cdot 18 + 70 \cdot 4) = 46.75$$

Varianz s^2

$$s^2 = \frac{1}{40}(25^2 \cdot 8 + 40^2 \cdot 10 + 55^2 \cdot 18 + 70^2 \cdot 4) - \bar{x}^2 = 190.577$$

Korrigierte Varianz s_{korr}^2

$$\frac{n}{n-1} \cdot s^2 = \frac{40}{39} \cdot 190.688 = 195.577$$

Standardabweichung s

$$\sqrt{s^2} = \sqrt{19.577} = 13.809$$

Korrigierte Standardabweichung S_{korr}

$$\sqrt{\frac{n}{n-1}} \cdot s = \sqrt{\frac{40}{39}} \cdot 13.809 = 13.9849$$

Median -> Q_2

$40 \cdot \frac{1}{2} = 20$ -> 20 ist in Klasse **50-60 (zwischen 18 - 36)**

a=50 b=60 **F(a)=$\frac{18}{40}$** **F(b)=$\frac{36}{40}$**

$$R_q = a + (b-a) \cdot \frac{q - F(a)}{F(b) - F(a)} =$$

$$50 + (60-50) \cdot \frac{0.50 - \frac{18}{40}}{\frac{36}{40} - \frac{18}{40}} = 51.11$$

Interquartilsabstand -> Q_3 - Q_1

$$56.67 - 34 = 22.67$$

Q_1

$40 \cdot \frac{1}{4} = 10$ -> 10 ist in Klasse **30-50 (zwischen 8 - 18)**

a=30 b=50 **F(a)=$\frac{8}{40}$** **F(b)=$\frac{18}{40}$**

$$30 + (50-30) \cdot \frac{0.25 - \frac{8}{40}}{\frac{18}{90} - \frac{8}{40}} = 34$$

Q_3

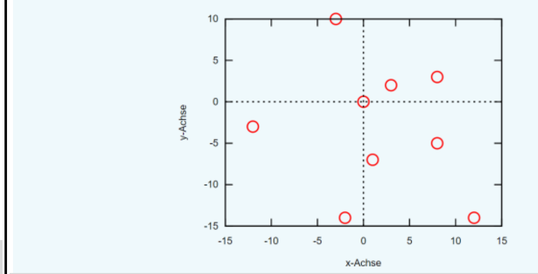
$40 \cdot \frac{3}{4} = 30$ -> 30 ist in Klasse **50-60 (zwischen 18 - 36)**

a=50 b=60 **F(a)=$\frac{18}{40}$** **F(b)=$\frac{36}{40}$**

$$50 + (60-50) \cdot \frac{0.75 - \frac{18}{40}}{\frac{36}{40} - \frac{18}{90}} = 56.67$$

Multivariante Statistik

Gegeben ist die folgende bivariate Datenliste:
$[[8,-5],[-2,-14],[1,-7],[-12,-3],[8,3],[-3,10],[12,-14],[0,0],[3,2]]$



x-Koordinatenliste

Jeweils erster Wert in den Klammern
[8, -2, 1, -12, 8, -3, 12, 0, 3]

Sortierte x-Koordinatenliste

Einfach sortieren → [-12, -3, -2, 0, 1, 3, 8, 8, 12]

Rangliste x-Koordinate

Sortierte Liste nehmen und Rangliste erstellen
Kommt ein Wert mehrmals vor
→ Ränge addieren und durch Anzahl dividieren
[-12, -3, -2, 0, 1, 3, 8, 8, 12]
1 2 3 4 5 6 7 8 9
Der Wert 8 kommt zwei Mal vor, daher
$\frac{7+8}{2}$ → Ränge addieren und durch Anzahl dividieren
[-12, -3, -2, 0, 1, 3, 8, 8, 12]
1 2 3 4 5 6 7.5 7.5 9
Diese Ränge dann in die unsortierte x-Koorinatenliste
[8, -2, 1, -12, 8, -3, 12, 0, 3]
7.5 3 5 1 7.5 2 9 4 6 → Lösung

y-Koordinatenliste

Jeweils zweiten Wert in den Klammern
[-5, -14, -7, -3, 3, 10, -14, 0, 2]

Sortierte y-Koordinatenliste

Sortieren → [-14, -14, -7, -5, -3, 0, 2, 3, 10]

Rangliste y-Koordinate

Wie oben → [4, 1.5, 3, 5, 8, 9, 1.5, 6, 7]

Kovarianz (nicht korrigiert)

Ähnlich zu berechnen wie Varianz

$$\frac{\sum(xi - \bar{x})(yi - \bar{y})}{n}$$

$$\bar{x} = \frac{8 - 2 + 1 - 12 + 8 - 3 + 12 + 0 + 3}{9} = \frac{5}{3}$$

$$\bar{y} = \frac{-5 - 14 - 7 - 3 - 3 + 10 - 14 + 0 + 2}{9} = \frac{-34}{9}$$

$$\frac{\left(8-\frac{5}{3}\right)\left(-5+\frac{34}{9}\right)+\left(-2-\frac{5}{3}\right)\left(-14+\frac{34}{9}\right)+\cdots+\left(3-\frac{5}{3}\right)\left(2+\frac{34}{9}\right)}{9} = -11.59$$

Pearson Korrelationskoeffizient

$$\frac{Kovarianz}{Standartabweichung\ x * Standartabweichung\ y}$$

Taschenrechner → x und y in Tabelle eingeben (benennen)
Menu → Statistik → Stat. Ber. → Stat. Mit zwei Var
Spalten auswählen, Rest lassen, ok → runterscrollen bis «r»

Pearson Korrelationseffizient ist stets im Bereich [0,1]. Desto näher an 1, desto grösser ist die Korrelation.
[-1,0] je näher an -1 desto stärker der negative lineare zusammenhang

Kovarianz der Ranglisten

Gleich wie Kovarianz, nur mit Ranglisten statt normaler Liste

Spearman Korrelationskoeffizient

Misst die Stärke und Richtung des linearen Zusammenhangs.

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

$$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \overline{x^2} - \bar{x}^2$$

$$s_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2} = \overline{y^2} - \bar{y}^2$$

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

Bsp.:

i	1	2	3	4	5	6	
x_i	59	35	43	23	42	27	
y_i	14.6	11.8	14.3	13.0	14.2	11.0	
$\mathrm{rg}(x_i)$	6	3	5	1	4	2	$\overline{\mathrm{rg}(x)} = 3.5$
$\mathrm{rg}(x_i) - \overline{\mathrm{rg}(x)}$	2.5	-0.5	1.5	-2.5	0.5	-1.5	
$\mathrm{rg}(y_i)$	6	2	5	3	4	1	$\overline{\mathrm{rg}(y)} = 3.5$
$\mathrm{rg}(y_i) - \overline{\mathrm{rg}(y)}$	2.5	-1.5	1.5	-0.5	0.5	-2.5	

Dabei wird dieser Koeffizient berechnet wie die Pearson-Korrelation mit dem Unterschied, dass die Ränge statt der Originaldaten verwendet werden.

$$r_{Sp} = \frac{\sum_{i=1}^{n}\left(\mathrm{rg}(x_i) - \overline{\mathrm{rg}(x)}\right)\left(\mathrm{rg}(y_i) - \overline{\mathrm{rg}(y)}\right)}{\sqrt{\sum_{i=1}^{n}\left(\mathrm{rg}(x_i) - \overline{\mathrm{rg}(x)}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(\mathrm{rg}(y_i) - \overline{\mathrm{rg}(y)}\right)^2}}$$

$$\sqrt{\sum_{i=1}^{n}\left(\mathrm{rg}(x_i) - \overline{\mathrm{rg}(x)}\right)^2} = \sqrt{\sum_{i=1}^{n}\left(\mathrm{rg}(y_i) - \overline{\mathrm{rg}(y)}\right)^2}$$

$$= \sqrt{2.5^2 + (-0.5)^2 + 1.5^2 + (-2.5)^2 + 0.5^2 + (-1.5)^2} = \sqrt{17.5}$$

CDF Wert berechnen

In der folgenden Tabelle stehen die PDF Werte einer klassierten Stichprobe:

Klassengrenzen	27-48	48-77	77-103	103-140	140-207
PDF	$\frac{11}{1281}$	$\frac{13}{1769}$	$\frac{4}{793}$	$\frac{20}{2257}$	$\frac{9}{4087}$

CDF Wert F(153) → 153 ist in der letzten Klasse
207-140 = 67 → 67 * 9 = 603 → 603/4087 in letzter Klasse
4087-603 = 3484 → 3484/4087 bis zur letzten Klasse
153-140 = 13 → 13*9 = 117 → 117+3484 = 3061 →
3061/4087

PDF Wert berechnen

In der folgenden Tabelle stehen die absoluten Häufigkeiten einer klassierten Stichprobe:

Klassengrenzen	13-30	30-101	101-209	209-397	397-839
Absolute Häufigkeit	14	3	11	22	8

PDF Wert f(759) → in der letzten Klasse
Relative Häufigkeit ausrechnen → 14+3+11+22+8 = 58
Letzte Klasse → 8/58
8/58 ist relative Häufigkeit der ganzen Klasse
→ 839-397 = 442 → 8/58/442 = **8/25636**

PMF Wert berechnen

Berechnen Sie den folgenden PMF Wert für die Stichprobe
$x = [0, 2, -2, -1, 3, -2, 2, 3, 0, -1, 3, -2, -3, -1, 1, 2, 1, 1]$:
$f(0) = ?$

Anzahl 0 / Gesamte Anzahl = **2/18**
Achtung: Wenn z.B. f(2.5), dann 0, da 2.5 nie vorkommt

CMF Wert berechnen

Berechnen Sie den folgenden CMF Wert für die Stichprobe
$x = [-8, 1, -2, 6, 0, 1, -8, -2, 3, 3, 3, 0, 3, -1, 3, -2, 6, 2]$:
$F(-3) = ?$

Stichprobe sortieren → Anzahl Werte bis gesuchter Wert
Im Beispiel → -8, -8
Anzahl Werte / Gesamte Anzahl = **2/18**

Praktikum 1

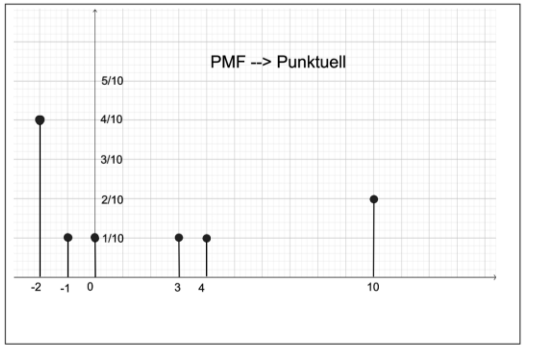
Aufgabe 1

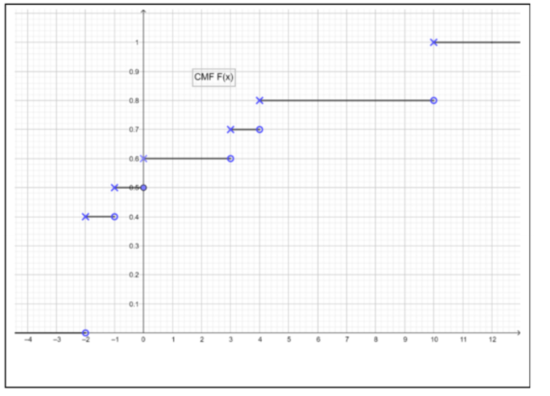
Aufgabe 1

a) Bestimmen Sie die PMF und CMF der Stichprobe: X=(-2,10,0,-2,4,-1,3,-2,10,-2).

Merkmale	-2	-1	0	3	4	10
PMF f(x)	4/10	1/10	1/10	1/10	1/10	2/10
CMF F(x)	4/10	5/10	6/10	7/10	8/10	10/10

b) Zeichnen Sie die PMF mit einer vernünftigen Skala.





Kombinatorik

Probleme

Variation(mit Reihenfolge) – Mit Wiederholung n^k

Zahlenschlossproblem:
Sie haben den Code ihres **6** stelligem Zahlenschloss**(0-9)** vergessen. Wie viele Versüche brauchen Sie im schlimmsten Fall?
$\mathbf{10^6}$

Bitproblem:
Wie viele Zahlen können mit **64 Bits** dargestellt werden?
$\mathbf{2^{64}}$

Buchstabenproblem:
Möglichkeiten für **5-Stellige** Wörter aus **10 Buchstabenplättchen** legen:
$\mathbf{10^5}$

Variation(mit Reihenfolge) – Ohne Wiederholung

Schwimmwettkampf:
10 Schwimmer, wie viele Möglichkeiten für das **Podest**(1-3) gibt es?
$$\frac{10!}{(10-3)!}$$

Buchstabenproblem:
Aus Anzahl **5-stelliger Wörtern** aus **10 verschiedenen Buchstaben**(mehrmals verwendbar):
$$\frac{10!}{(10-5)!}$$

Generäle:
Anzahl Möglichkeiten um **11** Generäle um einen runden Tisch zu verteilen:
11!
General-x immer neben General-y:
(11-1)!

Kombination(ohne Reihenfolge) – Mit Wiederholung

Zahnarztproblem:
3 Spielzeuge ziehen aus **5 Kisten** ziehen. Anzahl Möglichkeiten:
$$\binom{n+k-1}{k} = \binom{7}{3} = 35$$

Tellschiessen:
3 Pfeile auf Scheibe mit **10 Bereichen**. **(+1 daneben)** Anzahl Möglichkeiten:
$$\binom{n+k-1}{k} = \binom{13}{3} = 286$$

Kombination(ohne Reihenfolge) – Ohne Wiederholung

Lotto:
6 Ziehungen aus 49 Kugeln Chance die richtige kombination zu treffen:
$$\frac{n!}{(n-k)*k!} = \frac{49}{6}$$

Fussballmannschaft:
Anzahl Möglichkeiten bei 11 von 20.
$$\binom{20}{11}$$
Die Klasse besteht aus **8 Frauen** und **12 Männern**.
Anzahl Möglichkeiten um im **Team 6 Frauen und 5 Männer** sind.
$$\binom{8}{6} \cdot \binom{12}{5}$$

Teilmengenproblem:
Anzahl **3-elementige Teilmengen** hat die **Menge {1,2,3,4}**:
$$\binom{4}{3} = 4$$
Anzahl Teilmengen der **Menge {1,2,3,4}**
$\mathbf{2^4}$

Auswahlen von k Objekten aus einer Gesamtheit von n Objekten			
Variation (=mit Reihenfolge)		**Kombination** (=ohne Reihenfolge)	
Mit Wiederholung	***Ohne Wiederholung***	***Mit Wiederholung***	***Ohne Wiederholung***
Zahlenschloss (1) Bitproblem (4) Buchstabenproblem (7b)	Schwimmwettkampf (2) Buchstabenproblem (7a) Napoleon (9)	Zahnarztproblem (5) Tellschiessen (8)	Lotto (3) Fussballmannschaft (6a) Teilmengenproblem (11a)
n^k	$\frac{n!}{(n-k)!}$	$\binom{n+k-1}{k}$	$\frac{n!}{(n-k)!k!} = \binom{n}{k}$

- **Variation** von k aus n Objekten **mit** Wiederholung.
- **Variation** von k aus n Objekten **ohne** Wiederholung.
- **Kombination** von k aus n Objekten **mit** Wiederholung.
- **Kombination** von k aus n Objekten **ohne** Wiederholung.

$\binom{n}{k}$ **-> nCr(n,k)**

Aufgabe 2:
a) Bei der Trio-Wette müssen die drei bestplatzierten Pferde eines Pferderennens in der richtigen Reihenfolge vorausgesagt werden.
Wieviele verschiedene Trio-Wetten gibt es bei einem Pferderennen bei dem sechs Pferde teilnehmen?
b) Bei der Tiercé-Wette müssen die drei bestplatzierten Pferde eines Pferderennens vorausgesagt werden. Die Reihenfolge wird dabei nicht berücksichtigt.
Wieviele verschiedene Tiercé-Wetten gibt es bei einem Pferderennen bei dem sechs Pferde teilnehmen?

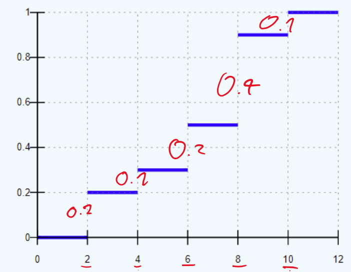
Lösung:
a) Variation ohne Wiederholung, $\frac{6!}{3!}=120$ (b) Kombination ohne Wiederholung, $\binom{6}{3}=20$

Aufgabe 4:
Beim deutschen Fussballtoto muss man die Ergebnisse aus 11 Fussballspielen vorhersagen. Wie das Ergebnis getippt werden muss, sei für ein Spiel erklärt.
Hamburger SV gegen Schalke 04
1 = Hamburger SV gewinnt
2 = Schalke 04 gewinnt
0 = unentschieden.
Wie viele mögliche Tippergebnisse gibt es?

Lösung:
Variation mit Wiederholung, $3^{11}=177147$

Aufgaben Quizzes

Diese Graphik zeigt die **CMF** der Zufallsvariablen X:



Bestimmen Sie mithilfe der Graphik die folgenden Kenngrössen der Zufallsvariablen X.

$E(X) =$ 6.2 $\quad 2 \cdot 0.2 + 4 \cdot 0.1 + 6 \cdot 0.2 + 8 \cdot 0.4 + 10 \cdot 0.1 = 6.2$

$E(X^2) =$ 45.2 $\quad 2^2 \cdot 0.2 + 4^2 \cdot 0.1 + 6^2 \cdot 0.2 + 8^2 \cdot 0.4 + 10^2 \cdot 0.1 = 45.2$

$V(X) =$ 6.76 $\quad E(X^2) - E(X)^2$

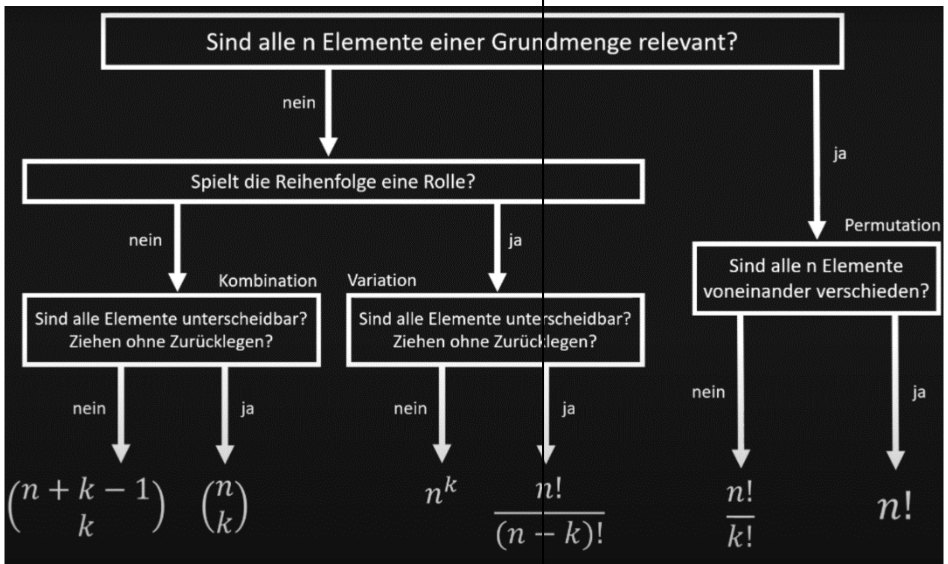
x	1	2	3	5
PMF	0.5	0.2	0.2	0.1

Berechnen Sie die folgenden Kenngrössen:

$$E(X) = 1 \cdot 0.5 + 2 \cdot 0.2 + 3 \cdot 0.2 + 5 \cdot 0.1 = 2$$

$$E(X^2) = 1^2 \cdot 0.5 + 2^2 \cdot 0.2 + 3^2 \cdot 0.2 + 5^2 \cdot 0.1 = 5.6$$

$$V(X) = E(X^2) - E(X)^2 = 1.6$$



Anzahl Möglichkeiten einer 6-8 stelligen Zahl mindestens 1 4 zu haben

Aufgabe 5:
Auf wie viele Arten lässt sich aus 5 Ehepaaren eine Gruppe von 4 Personen auswählen, die keines der Ehepaare enthält.

Lösung:
Kombination ohne Wiederholung: $\binom{5}{4}=5$ ist die Anzahl der Möglichkeiten aus 5 Ehepaaren 4 Ehepaare auszuwählen. Von jedem der 4 Ehepaare die man ausgewählt hat kann man nun je einen der Partner auswählen; dies kann auf $2^4=16$ Arten erfolgen. Die gesuchte Anzahl ist $5 \cdot 16=80$.

Aufgabe 6:
Beim Rommé spielt man mit 110 Karten: sechs davon sind Joker. Zu Beginn eines Spiels erhält jeder Spieler genau 12 Karten. In wieviel Prozent aller möglichen Fälle sind darunter
a) genau zwei bzw. b) mindestens ein Joker?

Lösung:
a) $\frac{\binom{6}{2}\cdot\binom{104}{10}}{\binom{110}{12}} \cdot 100\% \approx 11.13\%$, b) $100\% - \frac{\binom{104}{12}}{\binom{110}{12}} \cdot 100\% \approx 50.85\%$

Aufgabe 7:
Sind in mehr als 60% aller Fälle von vier (nicht gleichaltrigen) Geschwistern mindestens zwei im gleichen Monat geboren?

Lösung:
Nein, $100\% - \frac{12 \cdot 11 \cdot 10 \cdot 9}{12^4} \cdot 100\% \approx 42.71\%$

Aufgabe 8:
Wie viele Worte lassen sich aus den Buchstaben des Wortes ABRAKADABRA bilden? (Nur Worte in denen alle Buchstaben vorkommen!)

Lösung:
$\binom{11}{5} \cdot \binom{6}{2} \cdot \binom{4}{2} \cdot \binom{2}{1} \cdot \binom{1}{1} = 83160$

Aufgabe 10:
Vor ihnen stehen 5 verschiedene Objekte A, B, C, D, E. Sie sollen neunmal auf irgendeines der Objekte zeigen. Wieviele Möglichkeiten gibt es, wenn die Reihenfolge keine Rolle spielen soll?

Lösung:
Kombination mit Wiederholung, $\binom{5+9-1}{9} = \binom{13}{4} = 715$

Aufgabe 11:
Ein moderner Künstler erhält den Auftrag ein Bild mit drei Farben zu malen. Der Künstler hat sich entschlossen, daß das Bild aus 12 nebeneinander angeordneten Streifen bestehen soll und 4 Streifen silber, 4 Streifen gold und 4 Streifen bronze sein sollen. Unter wievielen Möglichkeiten kann der Künstler auswählen?

Lösung:
$\binom{12}{4} \cdot \binom{8}{4} \cdot \binom{4}{4} = 34650$

Prüfung 2020

Aufg 2

	[200-400)	[400-500)	[500-600)	[600-1000)
Klassenbreite b_i	200	100	100	400
Klassenmitte m_i	300	450	550	800
Abs. Häufigk. h_i	4	6	8	22
Rel. Häufigk f_i	$\frac{4}{40}$ =0.1	$\frac{6}{40}$ =0.15	$\frac{8}{40}$ =0.2	$\frac{22}{40}$ =0.55
Rel. Häufigk dichte(PDF) $f(x)$	$\frac{4}{40\cdot 200}$	$\frac{6}{40\cdot 100}$	$\frac{8}{40\cdot 100}$	$\frac{22}{40\cdot 400}$
Kummulierte Vert funktion F(x)	0.1	0.25	0.45	1

Stichprobenmittel -> $\bar{x}$

$$\bar{x} = \Sigma\, m_i \cdot h_i =$$

$$300 \cdot 0.1 + 450 \cdot 0.15 + 550 \cdot 0.2 + 800 \cdot 0.55 = 647.5$$

Stichprobenvarianz -> s^2

$$s^2 = \overline{x^2} - \bar{x}^2 =$$

$$(300^2 \cdot 0.1 + 450^2 \cdot 0.15 + 550^2 \cdot 0.2 + 800^2 \cdot 0.55) - 647.5^2$$

$$= 32618.75$$

Median der klassierten Daten -> Q_2

$40 \cdot \frac{1}{2} = 20$ -> 20 ist in Klasse **600-1000 (zwischen 0.45 - 1)**

a=600 b=1000 **F(a)=0.45** **F(b)= 1**

$$R_{0.5} = 600 + (1000 - 600) \cdot \frac{0.5 - 0.45}{1 - 0.45} = \mathbf{636.364}$$

Boxplot 1 Aufg 3

Daten: {2700, 2850, 2930, 3040, 2600, 4160} n=6

Sortiert: {2600, 2700, 2850, 2930, 3040, 4160}

Q_1

2700

Q_2

$\frac{2850+2930}{2}$ =2890

Q_3

3040

IQR *1.5

$$IQR \cdot 1.5 = Q_3 - Q_1 = 340 \cdot 1.5 = \mathbf{510}$$

Untere Antenne (alles zwischen Q1 bis Q1-510)

2600

Obere Antenne (alles zwischen Q3 bis Q3+510)

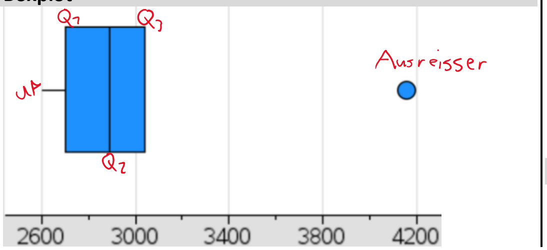
keine

Unterer Ausreisser - keine

Oberer Ausreisser

4160

Boxplot



Aufg 4

Tag i	1	2	3	4	5	Mittelwert
Preis x_i	4.7	4.3	3.8	4.5	5.2	4.5
Tagesmenge y_i	70	75	80	75	50	70

Pearson-Korrelationskoeffizient 1.

Tag i	1	2	3	4	5	Mittelwert
Preis x_i	4.7	4.3	3.8	4.5	5.2	4.5
Tagesmenge y_i	70	75	80	75	50	70
x_i^2	22.09	18.49	14.44	20.25	27.04	20.462
y_i^2	4900	5625	6400	5625	2500	5010
$x_i y_i$	329	322.5	304	337.5	260	310.6

$$s_x^2 = \overline{x^2} - \bar{x}^2 = 20.462 - 4.5^2 = 0.212$$

$$s_y^2 = \overline{y^2} - \bar{y}^2 = 5010 - 70^2 = 110$$

$$s_{xy} = \overline{xy} - \bar{x}\bar{y} = 310.6 - 4.5 \cdot 70 = -4.40$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-4.5}{\sqrt{0.212} \cdot \sqrt{0.110}} = -0.911$$

➔ **-0.911 = starker negativer Zusammenhang zwischen Tagesmenge und Preis**

Pearson-Korrelationskoeffizient 2

$$\frac{Kovarianz}{Standartabweichung\ x * Standartabweichung\ y}$$

$$kov = \frac{\sum(xi - \bar{x})(yi - \bar{y})}{n}$$

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

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

Lineare regression in y Richtung

$$m = \frac{s_{xy}}{s_x^2} = \frac{-4.40}{0.212} = -20.755$$

$$d = \bar{y} - m\bar{x} = 70 - (-20.755) \cdot 4.5 = 163.40$$

➔ Zunahme des Preises um 1CHF/kg senkt verkauf um 20.755kg/d

Bestimmtheitsmass

$$R^2 = r_{xy}^2 = -0.911^2 = 0.8299$$

82.99% der Gesamtvarianz in den yDaten kann durch die regressionsgerade erklärt werden.

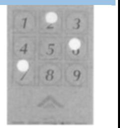
Preis damit 90kg/d verkauft werden

$$\hat{y} = 90 \Rightarrow \quad \hat{x} = \frac{\mathbf{1}}{\boldsymbol{n}}(\hat{y} - \boldsymbol{d})$$

$$= \frac{\mathbf{1}}{\mathbf{-20.755}} \cdot (\mathbf{90 - 163.40}) = \mathbf{3.54\ CHF}$$

Aufg 5

Das abgebildete Ticket wird mit einem Stanzmuster entwertet.



a) (1 Punkt) Dabei werden bei der Entwertung drei Löcher in die neun Felder des Tickets gestanzt. Wie viele solcher Karten müsste man sammeln, um alle möglichen Stanzmuster zu besitzen?

b) (2 Punkte) Um die Wiederverwendbarkeit zu erschweren, werden weitere Lochmuster hinzugenommen, so dass nun das Muster aus 2, 3 oder 4 Löchern bestehen kann. Wie viele solcher Karten müsste man nun sammeln, um alle möglichen Stanzmuster zu besitzen?

Um einem Betrug vorzubeugen wird nun neu stattdessen ein Code auf die Karte gestempelt.
Wie viele solcher Karten müsste man nun sammeln,

c) (1 Punkt) wenn der Code aus fünf verschiedenen Ziffern (0,1,...,9) besteht?

d) (1 Punkt) wenn der Code aus fünf beliebigen Ziffern besteht?

e) (1 Punkt) wenn der Code an den ersten beiden Stellen aus beliebigen Buchstaben des Alphabets (insgesamt 26 Buchstaben, Gross- und Kleinschreibung ohne Bedeutung) und den drei letzten Stellen aus beliebigen Ziffern besteht?

f) (2 Punkte) wenn der Code aus genau zwei, unterschiedlichen Buchstaben des Alphabets und einer Zahl bestehend aus mindestens 2, maximal 4 Ziffern besteht?

a)

$$\binom{n}{k} = \binom{9}{3} = 84$$

b)

$$\binom{9}{2} + \binom{9}{3} + \binom{9}{4} = 246$$

c)

$$\frac{n!}{(n-k)!} = \frac{10!}{(10-5)!} = 30'240$$

d)

$$10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 100'000$$

e)

$$26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 676'000$$

f)

$$26 \cdot 25 \cdot (10^2 + 10^3 + 10^4) = 7'215'000$$

Aufg 6

Roche hat einen «Rapid Antigen Test» herausgegeben, mit welchem man eine akute Erkrankung an SARS-Cov-2 nachweisen kann.
Die Firma schreibt, dass der Test eine an SARS-Cov-2 erkrankte Person mit einer Wahrscheinlichkeit von 96.52% korrekt als an der Krankheit erkrankt diagnostiziert (Testergebnis positiv, gegeben dass die Person tatsächlich erkrankt ist).
Eine gesunde Person wird mit einer Wahrscheinlichkeit von 99.68% als gesund erkannt (Testergebnis negativ, gegeben dass die Person tatsächlich gesund ist).
In dieser Aufgabe gehen wir davon aus, dass in einer Region mit ca. 1'000'000 Einwohnern 1% aktuell an Covid-19 erkrankt sind und die Angaben der Firma zutreffen.

a)

(3 Punkte) Befüllen Sie die untenstehende Kontingenztabelle (graue Felder) mit den genannten Zahlen.

	Testergebnis positiv	Testergebnis negativ	
Akut erkrankt			
Akut nicht erkrankt			
			1'000'000

	Test positiv	Test negativ	
Akut krank	9652	348	10'000
Nicht krank	3168	986'832	990'000
	12820	987'180	1'000'000

b)

(1 Punkt) Wie gross ist die Wahrscheinlichkeit, dass eine Person akut an SARS-Cov-2 erkrankt ist und ein negatives Testergebnis erhält?

$$p(E \wedge \bar{T}) = \frac{10'000}{100'000} \cdot \frac{348}{10'000} \cdot 100 = 0.0348\%$$

c)

(2 Punkte) Berechnen Sie die Wahrscheinlichkeit, dass eine Person, die ein positives Testergebnis erhält, tatsächlich an Covid-19 erkrankt ist.

$$p(E \wedge T|T) = \frac{9652}{12'820} \cdot 100 = 75.29\%$$

d)

$$P(Erkrankt) = 1\%$$

$$P(Positiv) = 1.282\%$$

$$P(EP) = \frac{9652}{10000} = 0.97$$

$$P(EP) \neq P(E) * P(p)$$

== Nicht stochastisch unabhängig!

Aufgabe 8

Präsident Romp möchte wissen wie gross sein Wähleranteil für die nächsten Wahlen in etwa sein wird. Hierfür werden nun 2000 zufällig ausgewählte Wahlberechtigte bestimmt und unabhängig voneinander befragt. Davon würden im Moment 108 für ihn stimmen.

Gesucht ist ein Konfidenzintervall zum Konfidenzniveau 0.99 für den Anteil derjenigen Wahlberechtigten, auf die sich Präsident Romp verlassen kann. Geben Sie das Resultat dieser Intervallschätzung inkl. aller Zwischenschritte an und runden Sie die Grenzen des Vertrauensintervalls auf 3 Nachkommastellen.

$$\hat{p} = \frac{108}{2000}$$

$$q = \frac{1+y}{2} = \frac{1+0.99}{2} = 0.995$$

c = 0.975 aus Tabelle: u_p aus Liste -> 2.576

$$\bar{x} = \hat{p}$$

$$\theta_u = \bar{x} - c \cdot \sqrt{\frac{\bar{x} \cdot (1 - \bar{x})}{n}}$$

$$\theta_o = \bar{x} + c \cdot \sqrt{\frac{\bar{x} \cdot (1 - \bar{x})}{n}}$$

Von 0.041 – 0.067

Prüfung 2018

Aufg 1

X von...bis..	Absolute Häufigkeit h_i	Relative Summenhäufigkeit $F_i(CDF)$
200-500	600	$\frac{600}{1500} = 0.4$
500-600	300	$\frac{900}{1500}$
600-800	100	$\frac{1000}{1500}$
800-1000	500	$\frac{1500}{1500}$

1500

0.3 Quantil

$$\frac{0.4}{500-200} = \frac{0.3}{R_{0.3}-200} \rightarrow R_{0.3} = \frac{500-200}{0.4}\cdot 0.3 + 200 = \mathbf{425}$$

% <= 880

$$600 + 300 + 100 + \left(\frac{500}{200}\cdot 80\right) = 1200$$

$$\frac{100\cdot 1200}{1500} = \mathbf{80\%}$$

Aufg 2

Gegeben ist die folgende Stichprobe von Längendaten x [Einheit cm]

9.6 6 8.5 3 0.2 1 4.7 8 6.8 1.3

Klassieren Sie die Daten in 3 aneinander liegende Klassen I, II, und III, so dass die zugehörigen PDF und die CDF Werte die folgende Tabelle erfüllen:

x von … bis…	PDF Wert auf der Klasse	CDF Wert am rechten Klassenrand
Klasse I:	0.1	0.1
Klasse II:	0.1	0.7
Klasse III:	0.1	1

Begründen Sie ausführlich ihre Klassierung und die Gültigkeit der Tabelle.

Gegeben ist die folgende Stichprobe von Längendaten x [Einheit cm]

9.6 6 8.5 3 0.2 1 4.7 8 6.8 1.3

Klassieren Sie die Daten in 3 aneinander liegende Klassen I, II, und III, so dass die zugehörigen PDF und die CDF Werte die folgende Tabelle erfüllen:

x von … bis…	PDF Wert auf der Klasse	CDF Wert am rechten Klassenrand
Klasse I:	0.1	0.1
Klasse II:	0.1	0.7
Klasse III:	0.1	1

Stichprobe ordnen

0,2 1 1.3 3 4.7 6 6.8 8 8.5 9.6

Klassenbreite = $\frac{CDF\ der\ Klasse}{PDF\ der\ Klasse}$

Klasse 1 = $\frac{0.1}{0.1} = 1$

Klasse 2 = $\frac{0.6}{0.1} = 6$

Klasse 3 = $\frac{0.3}{0.1} = 3$

0,2 | 1 1.3 3 4.7 6 6.8 | 8 8.5 9.6 → Klassengrenzen

Mögliche Klassengrenzen

I = [0,1[, *II* = [1,7[, *III* = [7,10[

Aufg 3 – Boxplot

X = {2, -1, 3, 4, 9, 4, -1, 2, 2}

X_sort= {-1, -1, 2, 2, 2, 3, 4, 4, 9}

Q1

$Q_1 = \frac{-1+2}{2} = \mathbf{0.5}$

Q2 – Median

$Q_2 = \mathbf{2}$

Q3

$Q_1 = \frac{4+4}{2} = \mathbf{4}$

$IQR\cdot 1.5$

$IQR\cdot 1.5 = (Q_3 - Q_1)\cdot 1.5 = (4-0.5)\cdot 1.5 = \mathbf{5.25}$

Untere Antenne

0.5 – 5.25 = -4.75 -> **-1**

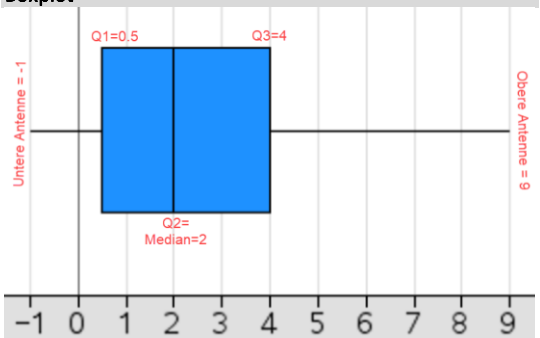
Obere Antenne

4 + 5.25 = 9.25 -> **9**

Untere Ausreisser - keine

Obere Ausreisser - keine

Boxplot

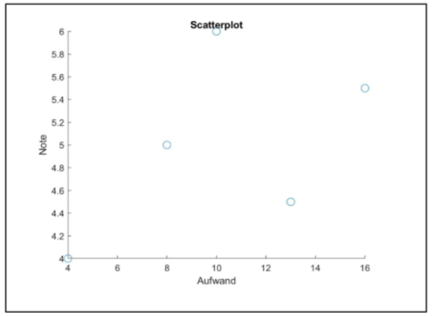


Aufg 4

Eine Klasse wird nach der Prüfung zu ihrem Lernverhalten befragt. In den Daten ergibt sich der folgende Zusammenhang zwischen dem Aufwand in Stunden für die Prüfungsvorbereitung und der Note in der Prüfung.

Student	1	2	3	4	5
Aufwand	10	16	8	13	4
Note	6	5.5	5	4.5	4

a) Skizzieren Sie einen Scatterplot der Daten. Wie interpretieren Sie diesen Plot? (2 Punkte)



Ein leichter linearer (positiver) Zusammenhang besteht zwischen Aufwand und Note.

b) Bestimme **arithmetische Mittel** und die **Varianz** von Aufwand und Note sowie die **Kovarianz** (inkl. Tabelle).

						Summe	Mittel
A	10	16	8	13	4	51	51/5
N	6	5.5	5	4.5	4	25	5
A^2	100	256	64	169	16	605	121
N^2	36	30.25	25	20.25	16	127.5	25.5
$A\cdot N$	60	88	40	58.5	16	262.5	52.5

$$V(A) = \overline{A^2} - \overline{A}^2 = 121 - \left(\frac{51}{5}\right)^2 = 16.96$$

$$V(N) = \overline{N^2} - \overline{N}^2 = 25.5 - 5^2 = 0.5$$

$$Kov(A,N) = \overline{AN} - \overline{A}\cdot\overline{N} = 52.5 - \frac{51}{5}\cdot 5 = 1.5$$

c) Berechne den **Korrelationskoeffizienten**. Was bedeutet dies? Was würden Sie den Studierenden empfehlen?

$$r_{A,N} = \frac{Kov(A,N)}{\sqrt{V(A)}\cdot\sqrt{V(N)}} \approx \frac{1.5}{9.2087\cdot 0.7071} = 0.515$$

Leichter linearer Zusammenhang besteht. Nach dem Bestimmtheitsmass $R^2 = r_{A,N}^{\ 2} \approx 0.2652$ sind etwa 27% der auftretenden N Werte durch die linear prognostizierten N Werte erklärt. Viel arbeiten!

Aufgabe 5

Gegeben: 5 Programme -> Anz Auswahlen z.B.(P1, P3, P5) von MINDESTENS 2 gibt es

$$\binom{5}{5} + \binom{5}{4} + \binom{5}{3} + \binom{5}{2} = 26 \text{ oder } 2^5 - \binom{5}{1} - \binom{5}{0} = 26$$

Die 5 Programme werden nacheinander ausgeführt. Wie viele verschiedene reihenfolgen gibt es?

5! = 120

Zu jeder Kombination aus Programmen hintereinander von mind 1 Programm wird zu jeder Möglichkeit ein test generiert. Wie viele solche Tests gibt es?

$$\frac{5!}{(5-1)!} + \frac{5!}{(5-2)!} + \frac{5!}{(5-3)!} + \frac{5!}{(5-4)!} + \frac{5!}{(5-5)!} = 325$$

2 von den 5 programmen sind jetzt identisch, wie sehen die 3 berechnungen jetzt aus?

$$a)\ \binom{4}{4} + \binom{4}{3} + \binom{4}{2} = 11$$

$$b)\ \binom{5}{2} * 3! = 10 * 6 = 60$$

c) Es gibt 4 unterschiedliche Fälle

zu a) $\sum_{k=2}^{4}\binom{4}{k} = 2^4 - \sum_{k=0}^{1}\binom{4}{k} = 16 - \binom{4}{0} - \binom{4}{1} = 16-1-4 = 11$

zu b) $\binom{5}{2}3! = 10\cdot 6 = 60$

zu c) 4 Fälle:

A) Auswahlen ohne P1 und ohne P2:

$$\sum_{k=1}^{3}\binom{3}{k}\cdot k! = \sum_{k=1}^{3}\frac{3!}{(3-k)!\cdot k!}\cdot k! = \sum_{k=1}^{3}\frac{3!}{(3-k)!} = 3+6+6 = 15$$

B) Auswahlen ohne P1 und mit P2:

$$\sum_{k=0}^{3}\binom{3}{k}\cdot(k+1)! = \sum_{k=0}^{3}\frac{3!}{(3-k)!\cdot k!}\cdot(k+1)! = 3!\sum_{k=0}^{3}\frac{(k+1)}{(3-k)!} = 1+6+18+24 = 49$$

C) Auswahlen mit P1 und ohne P2: wie Fall B), ergibt dieselben Reihenfolgen, da P1 und P2 identisch sind.

D) Auswahlen mit P1 und mit P2, d.h. zweimal dasselbe Programm:

$$\sum_{k=0}^{3}\binom{3}{k}\cdot k!\cdot\binom{2+k}{2} = \sum_{k=0}^{3}\frac{3!}{(3-k)!\cdot k!}\cdot k!\cdot\frac{(2+k)!}{2!\cdot k!} = \sum_{k=0}^{3}\frac{3\cdot(2+k)!}{(3-k)!\cdot k!} = 3\left(\frac{1}{3}+3+12+20\right) = 106$$

Total erhält man: $15+49+106 = 170$.

Aufgabe 6

Datenpaket aus 4 Übertragungsbits und 2 Korrekturbits.

Jedes Bit mit Wahrscheinlichkeit p fehlerhaft.

Ein korrektes Korrekturbit kann ein fehlerhaftes Übertragungsbit korrigieren, ein fehlerhaftes Korrekturbit jedoch nicht.

Bits sind voneinander stochastisch unabhängig.

Ein Paketfehler tritt auf, wenn mind. 1 Bit fehlerhaft ist.

Berechne die Paketfehlerwahrscheinlichkeit.

U = Anzahl korrekter Übertragungsbits

K = Anzahl korrekter Korrekturbits

$$P(U = k) = \binom{4}{k}(1-p)^k * p^{4-k}$$

Und

$$P(K = n) = \binom{2}{n}(1-p)^n * p^{2-n}$$

$$\begin{aligned} P(korrektes\ Paket) &= P(U=4) + P(U=3) \\ &\quad * \left(P(K=1) + P(K=2)\right) \\ &\quad + P(U=2) * P(K=2) \\ &= (1-p)^4 + 4(1-p)^3 p\left(2(1-p)p + (1-p)^2\right) \\ &\quad + 6(1-p)^2 p^2 (1-p)^2 \end{aligned}$$

P(Paketfehler) = 1 – P(korrektes Paket)

Aufgabe 7

2 Runden, fairer 6-seitiger Würfel

1. Runde = Win bei 1-3 / Loose bei 4-6

Wenn Win in 1. Runde: 2. Runde gleiche Chancen

Wenn Loose in 1. Runde: 2. Runde = 1-4 = Win / 5-6 = Loose

R1 = G und R2 = G → Gewinn

R1 = $\bar{G}$ und R2 = $\bar{G}$ → Verloren

a) Erstelle Wahrscheinlichkeitstabelle

	$R_2 = G$	$R_2 = \bar{G}$
$R_1 = G$	1/4	1/4
$R_1 = \bar{G}$	1/3	1/6

b) Erstelle Tabelle mit Randwahrscheinlichkeiten

	$R_2 = G$	$R_2 = \bar{G}$	PDF von R_1
$R_1 = G$	1/4	1/4	1/2
$R_1 = \bar{G}$	1/3	1/6	1/2
PDF von R_2	7/12	5/12	1

Sind R1 = G und R2 = G abhängig oder unabhängig?

Die beiden Ereignisse sind stochastisch abhängig, denn z.B.:

$$P(R_1 = G \wedge R_2 = G) = \frac{1}{4} \neq \frac{1}{2}\cdot\frac{7}{12} = P(R_1 = G)\cdot P(R_2 = G)$$

Aufgabe 8

Für einen Spurhalteassistenten stehen zwei Systeme: Typ1, Typ2 zur Auswahl, wobei das System vom Typ1 eine Testfahrt in 98 Prozent der Fälle fehlerlos bewältigt und das System vom Typ2 die Testfahrt nur in 75 Prozent der Fälle fehlerlos bewältigt.

Eine Serie von 100 Fahrzeugen wurde mit dem System Typ1 ausgestattet und danach wurde eine zweite Serie von 500 Fahrzeugen (aus Kostengründen) mit dem System Typ2 bestückt. Sie haben (zufällig) eines dieser 600 Fahrzeuge erhalten und der Spurhalteassistent hat soeben die ersten beiden Testfahrten fehlerlos bewältigt. Verschiedene Testfahrten eines Fahrzeugs werden dabei als stochastisch unabhängig angesehen!

a) Bestimme Wahrscheinlichkeit, dass das Auto von T1 ist.

Ereignisse: $T1: Fahrzeug$ vom Typ1 , $T2: Fahrzeug$ vom Typ1 , $F1: Fehler$ bei erster Testfahrt , $F2: Fehler$ bei zweiter Testfahrt .

$$P(T1) = \frac{1}{6},\ P(T2) = \frac{5}{6},\ P(\neg F1 \mid T1) = 0.98,\ P(\neg F1 \mid T2) = 0.75.$$

$$P(T1 \mid \neg F1 \wedge \neg F2) = \frac{P(T1 \wedge \neg F1 \wedge \neg F2)}{P(\neg F1 \wedge \neg F2)} = \frac{\frac{1}{6}\cdot 0.98^2}{\frac{1}{6}\cdot 0.98^2 + \frac{5}{6}\cdot 0.75^2} \approx 0.255$$

b) Bestimme Wahrscheinlichkeit, dass 3. Fahrt auch fehlerlos

$$P(\neg F3 \mid \neg F1 \wedge \neg F2) = \frac{P(\neg F1 \wedge \neg F2 \wedge \neg F3)}{P(\neg F1 \wedge \neg F2)} = \frac{\frac{1}{6}\cdot 0.98^3 + \frac{5}{6}\cdot 0.75^3}{\frac{1}{6}\cdot 0.98^2 + \frac{5}{6}\cdot 0.75^2} \approx 0.81$$