



Statistik und Wahrscheinlichkeitsrechnung



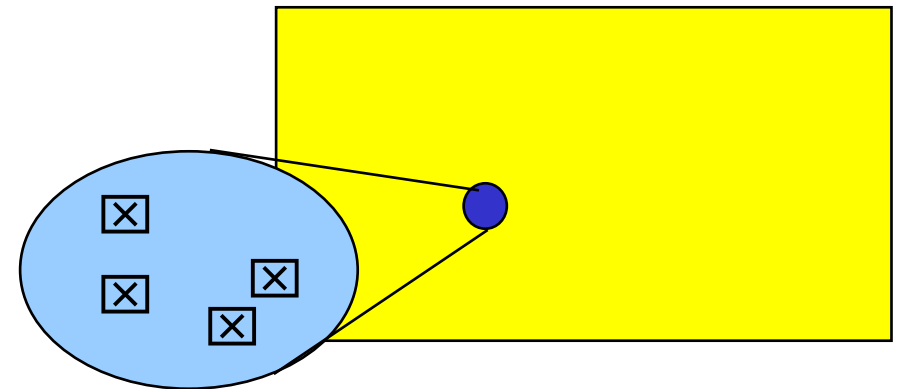
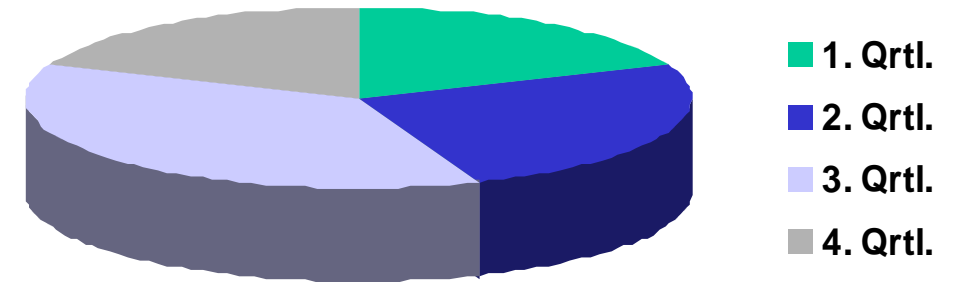
Stochastik

Wahrscheinlichkeitsrechnung und **Statistik** sind eng miteinander verbunden und werden üblicherweise unter dem Namen **Stochastik** (Lehre von den zufälligen Prozessen) zusammengefasst.



Teilgebiete der Statistik

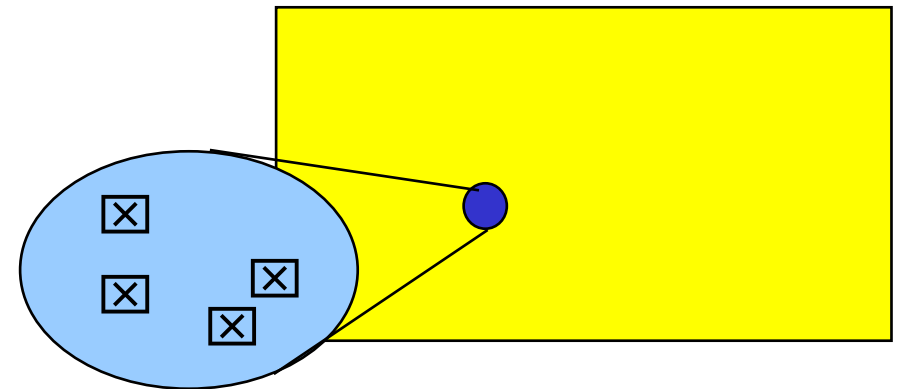
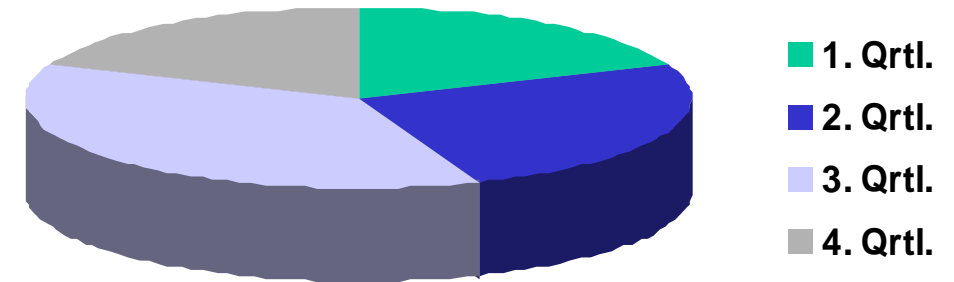
- **Deskriptive Statistik**
 - Darstellung von Daten
 - Berechnung von Kenngrößen
- **Induktive Statistik**
 - Stichprobenuntersuchungen
 - Übertragung auf die Grundgesamtheit mit den Methoden der Wahrscheinlichkeitsrechnung





Teilgebiete der Statistik

- **Deskriptive Statistik**
 - Darstellung von Daten
 - Berechnung von Kenngrößen
- **Induktive Statistik**
 - Stichprobenuntersuchungen
 - Übertragung auf die Grundgesamtheit mit den Methoden der Wahrscheinlichkeitsrechnung





Deskriptive Statistik

Grundbegriffe



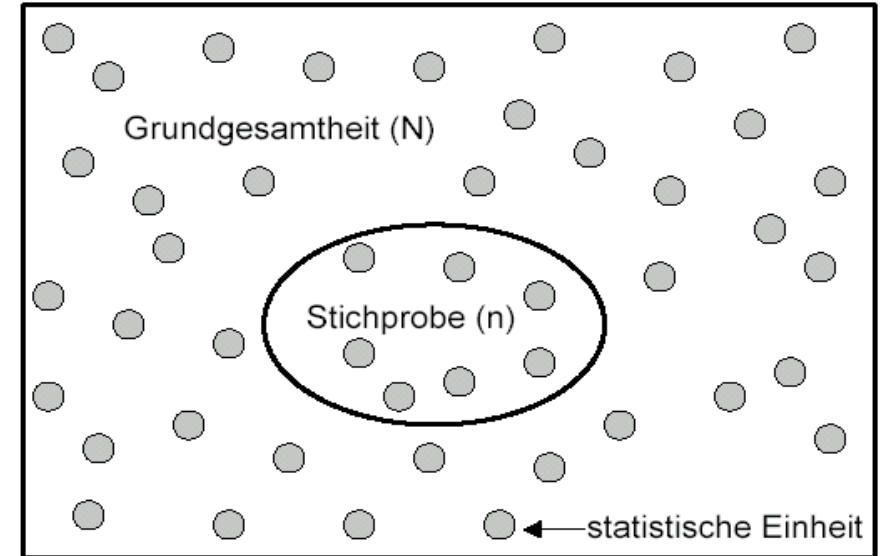
Grundgesamtheit und Stichprobe

Die Gruppe aller Untersuchungseinheiten, in denen Merkmale von Interesse sind, nennt man **Grundgesamtheit (Population oder statistische Masse)**. Da man nicht alle Elemente der Population untersuchen kann, befasst man sich meistens mit einer Auswahl aus der Grundgesamtheit, einer **Stichprobe**.

Die **Abgrenzung einer statistischen Masse** erfolgt über die Identifikationskriterien:

- sachlich
- räumlich
- zeitlich

z.B.: **Einwohner in Münster 2014**

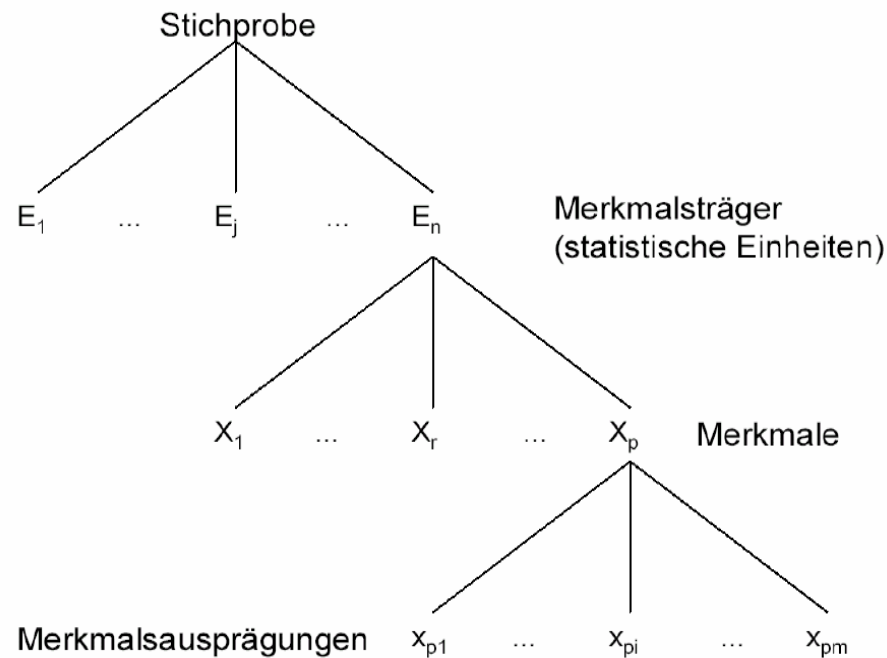


Grundgesamtheit	: Gesamtheit aller statistischen Einheiten, die bestimmte Identifikationsmerkmale aufweisen.
Totalerhebung	: Erfassung und Aufbereitung aller statistischen Einheiten.
(Zufalls-)Stichprobe	: Zufällig ausgewählter Teil der Grundgesamtheit.



Grundgesamtheit und Stichprobe

Die Elemente der untersuchten Stichprobe (z.B. Objekten, Personen oder Ereignissen) werden auch **Merkmalsträger** genannt. Die Merkmale, die verschiedene **Ausprägungen** annehmen, nennt man **Variablen**; die Merkmalsausprägungen selbst werden Variablenwerte oder Kategorien genannt.



Merkmalsträger	Merkmal (X)	Merkmalsausprägungen (x)
PKW	Navi	ja / nein
PKW	Farbe	schwarz, weiß, rot, ...
PKW	Preis	20.000 €, ...
PKW	Motor	Benzin, Diesel, Gas, ...

Die Festlegung der Skala/Maßeinheiten für die Merkmalsausprägungen nennt man **Skalierung**.



Skalenniveaus / Skalentypen

geringstes
Skalenniveau

Informationsgehalt

höchstes
Skalenniveau

Skalentyp	Datenniveau	Eigenschaft	log./math. Operationen	Mittelwert	Beispiele
Nominalskala	nominal qualitativ, diskret	<u>keine</u> Rangordnung,	$= / \neq$	Modalwert (Modus, häufigster Wert)	Geschlecht (m/w), Kinder (ja/nein), Wohnsitz (Bonn, Rom,...)
Ordinalskala	ordinal qualitativ, diskret	Rangordnung (Abstände unbekannt)	$= / \neq$ $< / >$	Modalwert, Median	Dienstgrad, Apfel (süß, ..., sauer) Schulnoten
Intervallskala	quantitativ, kardinal (metrisch), diskret oder stetig	<u>kein</u> natürlicher Nullpunkt und <u>keine</u> natürliche Maßeinheit	$= / \neq$ $< / >$ $+ / -$	Modalwert, Median, arithmetisches Mittel	<u>stetig</u> : Temperatur in Celsius, Kalenderzeit, <u>diskret</u> : Schulnoten*
Verhältnisskala (Rationalskala)	quantitativ, kardinal (metrisch), diskret oder stetig	natürlicher Nullpunkt, aber <u>keine</u> natürliche Maßeinheit	$= / \neq$ $< / >$ $+ / - \cdot / :$	Modalwert, Median, arithmetisches, geo- metrisches Mittel	<u>stetig</u> : Temperatur in Kelvin, Längenmaß, <u>diskret</u> : Alter (in ganzen Jahre)
Absolutskala	quantitativ, kardinal (metrisch), diskret	natürlicher Nullpunkt und natürliche Maßeinheit	$= / \neq$ $< / >$ $+ / - \cdot / :$	Modalwert, Median, arithmetisches, geo- metrisches Mittel	Stückzahl an prod. PKW, Einwohnerzahl

* unter Annahme gleicher Abstände zw. den Noten



Skalentypen

Übung

Merkmal	Skalentyp	diskret / stetig
Name	Nominalskala	diskret
Geschlecht	Nominalskala	diskret
Alter	Verhältnisskala	stetig
Einkommen in €	Verhältnisskala	diskret (nur ganze Cent)
Temperatur in Celsius	Intervallskala	stetig
Temperatur in Kelvin	Verhältnisskala	stetig
Schulnote	Ordinalskala *	diskret
Obstsorte	Nominalskala	diskret
Ränge bei einem Sportverein	Ordinalskala	diskret
Jahreszahlen	Intervallskala	diskret

* unter Annahme gleicher Abstände zw. den Noten: Intervallskala



Grundgesamtheit und Stichprobe

Datengewinnung

Primärerhebung / Primärstatistik

Daten werden erstmalig / neu für die Untersuchung erhoben.

Erhebungsmethoden: mündliche oder schriftliche Befragung, Beobachtungen, Experimente, elektronische / automatische Erfassung

Sekundärerhebung

Rückgriff auf vorhandene Daten



Häufigkeitsverteilungen

Nach einer Datenerhebung liegen die Messwerte eines Merkmals zunächst in Form einer sog. **Urliste** vor. In dieser Zahlenliste werden die Messwerte der einzelnen Untersuchungseinheiten hintereinander aufgelistet.

Urliste

Die Urliste ist die Liste der Einzeldaten, so wie erhoben, noch nicht sortiert oder gruppiert.

$X_1, X_2, X_3, X_4, \dots, X_n$

sortierte Urliste

Hier werden die Ausprägungen in eine aufsteigende Reihenfolge gebracht.

$X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}, \dots, X_{(n)}$

d.h.

$X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq X_{(4)} \leq \dots \leq X_{(n)}$



Häufigkeitstabellen

quantitativ diskrete Merkmale

Personenzahl pro Haushalt	absolute Häufigkeit	relative Häufigkeit	kumulierte relative Häufigkeit (Verteilungsfunktion)
1	120	7,2%	7,2%
2	500	29,9%	37,1%
3	680	40,7%	77,8%
4	260	15,6%	93,4%
5	80	4,8%	98,2%
6 und mehr	30	1,8%	100,0%
Summe	1670	100,0%	



Häufigkeitstabellen

quantitativ diskrete Merkmale

Personenzahl pro Haushalt	absolute Häufigkeit	relative Häufigkeit	kumulierte relative Häufigkeit (Verteilungsfunktion)	arith. Mittel
1	430	43,0%	43,0%	0,43
2	360	36,0%	79,0%	0,72
3	140	14,0%	93,0%	0,42
4	60	6,0%	99,0%	0,24
5	10	1,0%	100,0%	0,05
Summe	1000	100,0%		1,86



Häufigkeitstabellen

quantitativ stetiger Merkmale werden zu Klassen zusammengefasst

Körpergröße in m	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion
über 1,4 - 1,5	150	4,5%	4,5%
über 1,5 - 1,6	200	6,1%	10,6%
über 1,6 - 1,7	680	20,6%	31,2%
über 1,7 - 1,8	960	29,1%	60,3%
über 1,8 - 1,9	990	30,0%	90,3%
über 1,9 - 2,0	320	9,7%	100,0%
Summe	3300	100,0%	



Häufigkeitstabellen

die **Häufigkeitsdichte** ist definiert als die relative Häufigkeit geteilt durch die Klassenbreite (hier 0,1)

Körpergröße	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion	Häufigkeitsdichte
über 1,4 - 1,5	150	4,5%	4,5%	0,45
über 1,5 - 1,6	200	6,1%	10,6%	0,61
über 1,6 - 1,7	680	20,6%	31,2%	2,06
über 1,7 - 1,8	960	29,1%	60,3%	2,91
über 1,8 - 1,9	990	30,0%	90,3%	3,00
über 1,9 - 2,0	320	9,7%	100,0%	0,97
Summe	3300	100,0%		



Häufigkeitstabellen

Übung

Preis pro kg	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion	Häufigkeitsdichte
über 1,00 - 1,20	200	14,3%	14,3%	0,71
über 1,20 - 1,30	160	11,4%	25,7%	1,14
über 1,30 - 1,40	220	15,7%	41,4%	1,57
über 1,40 - 1,60	280	20,0%	61,4%	1,00
über 1,60- 1,80	240	17,1%	78,6%	0,86
über 1,80 - 2,20	300	21,4%	100,0%	0,54
Summe	1400	100,0%		



Aufgabe 1

a)	Nationalität	nominal
b)	Geschlecht	nominal
c)	Postleitzahl	nominal
d)	Größe	rational / stetig
e)	Gewicht	rational / stetig
f)	Alter	rational / stetig
g)	Kinderzahl	rational / diskret
h)	Militärdienstgrad	ordinal
i)	Rangplatz im Schachverein	ordinal
j)	Körpertemperatur	intervall / stetig



Aufgabe 2

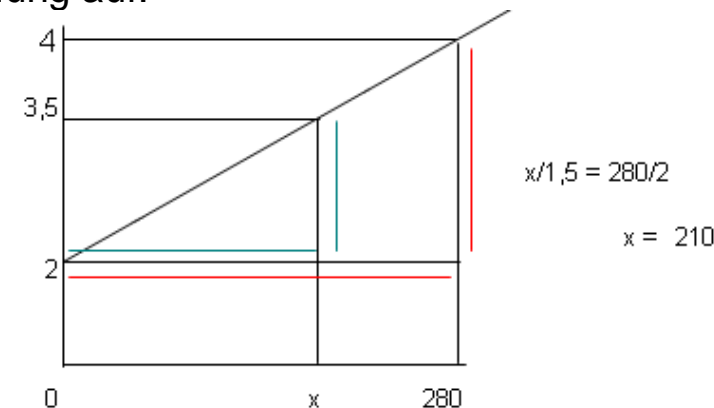
- a) Einkommen von alleinlebenden Erwerbspersonen im Jahr 2000.
es fehlt die räumliche Abgrenzung!
- b) Verkehrsunfälle im Saarland am 24.12.2000
o.K.
- c) Familienstand der erwachsenen Einwohner in Münster.
es fehlt die zeitliche Abgrenzung!



Aufgabe 3

1000 Motoren eines bestimmten Typs weisen folgende Lebensdauerverteilung auf:

Lebensdauer (in Jahren)	Anzahl (der Motoren)
bis 2	40
über 2 bis 4	280
über 4 bis 6	400
über 6 bis 8	220
über 8 bis maximal 10	60



- Wieviele Motoren haben eine Lebensdauer von höchstens 3,5 Jahren?
- Bestimmen Sie den Anteil der Motoren mit einer Lebensdauer von über 5 Jahren.
- Wieviele Motoren haben eine "normale" Lebensdauer (zwischen 3 und 7 Jahren)?

- Wieviele Motoren haben eine Lebensdauer von höchstens 3,5 Jahren?

$$40 + 210 = \mathbf{250}$$

- Bestimmen Sie den Anteil der Motoren mit einer Lebensdauer von über 5 Jahren.

$$200 + 220 + 60 = 480 \quad \Rightarrow \mathbf{48\%}$$

- Wieviele Motoren haben eine "normale" Lebensdauer (zwischen 3 und 7 Jahren)?

$$140 + 400 + 110 = \mathbf{650}$$



Aufgabe 4

Eine statistische Untersuchung hat folgende rationalskalierten Merkmalsausprägungen der betrachteten Einheiten ergeben:

4, 4, 5, 5, 3, 8, 7, 7, 8, 4, 5, 5, 4, 6, 8, 6, 4, 5, 5, 4.

Erstellen Sie eine Häufigkeitstabelle

Merkmalsausprägung	Absolute Häufigkeit	Relative Häufigkeit	Relative Summenhäufigkeit
3	1	5%	5%
4	6	30%	35%
5	6	30%	65%
6	2	10%	75%
7	2	10%	85%
8	3	15%	100%
Summe	20	100%	



Verteilungsmaße

▶ **Lagemaße**

geben einen Mittelwert der Merkmalswerte an

▶ **Streuungsmaße**

geben die Abweichung vom Lageparameter an



Verteilungsmaße

► Lageparameter

- Median (Zentralwert)
- Modus (Modalwert)
- arithmetisches Mittel

- Streuungsparameter
 - Spannweite
 - Varianz
 - Standardabweichung
 - Variationskoeffizient



Median

Bei einer **gruppierten** Verteilung nimmt im Median die Verteilungsfunktion den Wert 50 % an.

Median
→

Körpergröße	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion
über 1,4 - 1,5	150	4,5%	4,5%
über 1,5 - 1,6	200	6,1%	10,6%
über 1,6 - 1,7	680	20,6%	31,2%
über 1,7 - 1,8	960	29,1%	60,3%
über 1,8 - 1,9	990	30,0%	90,3%
über 1,9 - 2,0	320	9,7%	100,0%
Summe	3300	100,0%	



Median

Übung

Gewicht	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion
über 40 - 50	200		
über 50 - 60	300		
über 60 - 70	980		
über 70 - 80	960		
über 80 - 90	750		
über 90 - 100	310		
Summe			



Median

Übung

Median



Gewicht	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion
über 40 - 50	200	5,7%	5,7%
über 50 - 60	300	8,6%	14,3%
über 60 - 70	980	28,0%	42,3%
über 70 - 80	960	27,4%	69,7%
über 80 - 90	750	21,4%	91,1%
über 90 - 100	310	8,9%	100,0%
Summe	3500	100,0%	



Modus oder Modalwert

Bei einer **stetigen** und damit **gruppierten** Verteilung ist der Modalwert als die Mitte der Klasse mit der größten Häufigkeitsdichte definiert.

Körpergröße	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion	Häufigkeitsdichte
über 1,4 - 1,5	150	4,5%	4,5%	0,45
über 1,5 - 1,6	200	6,1%	10,6%	0,61
über 1,6 - 1,7	680	20,6%	31,2%	2,06
über 1,7 - 1,8	960	29,1%	60,3%	2,91
über 1,8 - 1,9	990	30,0%	90,3%	3,00
über 1,9 - 2,0	320	9,7%	100,0%	0,97
Summe	3300	100,0%		

1,85



Modus oder Modalwert

Übung

Körpergröße	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion	Häufigkeitsdichte
über 1,3 - 1,5	300			
über 1,5 - 1,6	400			
über 1,6 - 1,7	800			
über 1,7 - 1,8	700			
über 1,8 - 2,0	900			
über 2,0 - 2,3	400			
Summe	3500			



Modus oder Modalwert

Übung

Körpergröße	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion	Häufigkeitsdichte
über 1,3 - 1,5	300	8,6%	8,6%	0,43
über 1,5 - 1,6	400	11,4%	20,0%	1,14
über 1,6 - 1,7	800	22,9%	42,9%	2,29
über 1,7 - 1,8	700	20,0%	62,9%	2,00
über 1,8 - 2,0	900	25,7%	88,6%	1,29
über 2,0 - 2,3	400	11,4%	100,0%	0,38
Summe	3500	100,0%		

1,65



Arithmetisches Mittel

Bei **gruppierten** Daten berechnet man das **gewogene arithmetische Mittel** aus den jeweiligen Klassenmitten, wobei die Gewichte durch die relativen Häufigkeiten bestimmt werden.

Körpergröße	absolute Häufigkeit	relative Häufigkeit	arithmetisches Mittel
über 1,4 - 1,5	150	4,5%	0,066
über 1,5 - 1,6	200	6,1%	0,094
über 1,6 - 1,7	680	20,6%	0,340
über 1,7 - 1,8	960	29,1%	0,509
über 1,8 - 1,9	990	30,0%	0,555
über 1,9 - 2,0	320	9,7%	0,189
Summe	3300	100,0%	1,753





Verteilungsmaßzahlen

- ▶ Lageparameter
 - Median
 - Modus
 - arithmetisches Mittel

- ▶ **Streuungsparameter**
 - Spannweite
 - Varianz
 - Standardabweichung
 - Variationskoeffizient



Aufgabe 5 a)

Ermitteln Sie Modus, Median und arithmetisches Mittel für die Daten aus Aufgabe 3 und Aufgabe 4.

Lebensdauer	absolute Häufigkeit	relative Häufigkeit	Verteilungsfunktion	Häufigkeitsdichte	Modus	Median	ar. Mittel
bis 2	40	4,0%	4,0%	0,02			0,04
über 2 bis 4	280	28,0%	32,0%	0,14			0,84
über 4 bis 6	400	40,0%	72,0%	0,20	5,00	5,00	2,00
über 6 bis 8	220	22,0%	94,0%	0,11			1,54
über 8 bis 10	60	6,0%	100,0%	0,03			0,54
Summe	1000	100,0%					4,96



Aufgabe 5 b)

Werte	Werte sort.	Modus	Median	ar. Mittel
4	3			
4	4			
5	4			
5	4			
3	4			
8	4			
7	4	4 und		
7	5	5		
8	5			
4	5		5	
5	5		5	
5	5			
4	5			
6	6			
8	6			
6	7			
4	7			
5	8			
5	8			
4	8			
	107			5,35



Aufgabe 6)

Ein PKW fährt auf der Hinreise konstant mit 120 km/h, auf der Rückreise (gleiche Strecke) konstant mit 80 km/h.

Lösung:

96 km/h – und zwar für jede beliebige Strecke!

Beispiel: Die (einzelne) Strecke beträgt 1200 km. Dann benötigt der PKW für die Hinfahrt 10 Stunden, für die Rückfahrt 15 Stunden. Insgesamt also 25 Stunden für 2.400 km. Das entspricht einer Durchschnittsgeschwindigkeit von 96 km/h.

x [km]	v [km/h]	t
1.200	120	10
1.200	80	15
2.400		25

$$2 \cdot x / (x/120 + x/80)$$

96

$$\frac{2x}{x/120 + x/80}$$



Aufgabe 7)

Berechnen Sie die Varianz für folgende Altersverteilung:
20, 25, 40, 35, 30, 29, 31, 26, 30, 34.

	Wert 1	Wert 2	Wert 3	Wert 4	Wert 5	Wert 6	Wert 7	Wert 8	Wert 9	Wert 10	Mittel Summe Varianz
	20	25	40	35	30	29	31	26	30	34	30
Abweichung	-10	-5	10	5	0	-1	1	-4	0	4	
zum Quadrat	100	25	100	25	0	1	1	16	0	16	284
Varianz											28,4



Aufgabe 8

Gegeben seien die folgenden Einzelwerte:

5, 3, 2, 1, 1, 2, 2, 4, 5, 3.

a) Bestimmen Sie Modus, Median und arithmetisches Mittel.

b) Bestimmen Sie Spannweite, Varianz, Standardabweichung und Variationskoeffizient.

	Wert 1	Wert 2	Wert 3	Wert 4	Wert 5	Wert 6	Wert 7	Wert 8	Wert 9	Wert 10	Mittel Summe Varianz
	5	3	2	1	1	2	2	4	5	3	2,8
Abweichung zum Quadrat	2,2	0,2	-0,8	-1,8	-1,8	-0,8	-0,8	1,2	2,2	0,2	19,6
Varianz											1,96
Standabw											1,4
Varkoeff.											0,5



Aufgabe 8

Gegeben seien die folgenden Einzelwerte:

5, 3, 2, 1, 1, 2, 2, 4, 5, 3.

a) Bestimmen Sie Modus, Median und arithmetisches Mittel.

b) Bestimmen Sie Spannweite, Varianz, Standardabweichung und Variationskoeffizient.

Modus	2	= Modalwert: häufigster Wert
Median	2,5	Wert in der Mitte einer geordneten Reihe
arithmetisches Mittel	2,8	$\mu = \frac{1}{n} \cdot \sum_{i=1}^n x_i$
Spannweite	4	Spannweite = $x_{\max} - x_{\min}$
Standardabweichung	1,400	σ
Varianz	1,960	$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$
Variationskoeffizient	0,500	$\text{VarK} = \frac{\sigma}{\mu}$



Aufgabe 9

Gegeben seien jeweils die folgenden 5 Einzelwerte:

1, 1, 2, 3, 4

sowie

100, 100, 200, 300, 400

- a) Bestimmen Sie jeweils Modus, Median und arithmetisches Mittel.
- b) Bestimmen Sie jeweils Spannweite, Varianz, Standardabweichung und Variationskoeffizient.



Aufgabe 9

Modus	1	= Modalwert: häufigster Wert
Median	2	Wert in der Mitte einer geordneten Reihe
arithmetisches Mittel	2,2	$\mu = \frac{1}{n} \cdot \sum_{i=1}^n x_i$
Spannweite	3	Spannweite = $x_{\max} - x_{\min}$
Standardabweichung	1,166	σ
Varianz	1,36	$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$
Variationskoeffizient	0,5	$\text{VarK} = \frac{\sigma}{\mu}$

Modus	100	= Modalwert: häufigster Wert
Median	200	Wert in der Mitte einer geordneten Reihe
arithmetisches Mittel	220	$\mu = \frac{1}{n} \cdot \sum_{i=1}^n x_i$
Spannweite	300	Spannweite = $x_{\max} - x_{\min}$
Standardabweichung	116,6	σ
Varianz	13.600,00	$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$
Variationskoeffizient	0,5	$\text{VarK} = \frac{\sigma}{\mu}$



Mehrdimensionale Häufigkeitsverteilungen



Mehrdimensionale Häufigkeitsverteilungen

▶ wie lässt sich der Zusammenhang zwischen Merkmalen durch eine Funktion beschreiben?



Regressionsrechnung

▶ wie stark ist der Zusammenhang?



Korrelationsrechnung



Regressionsrechnung

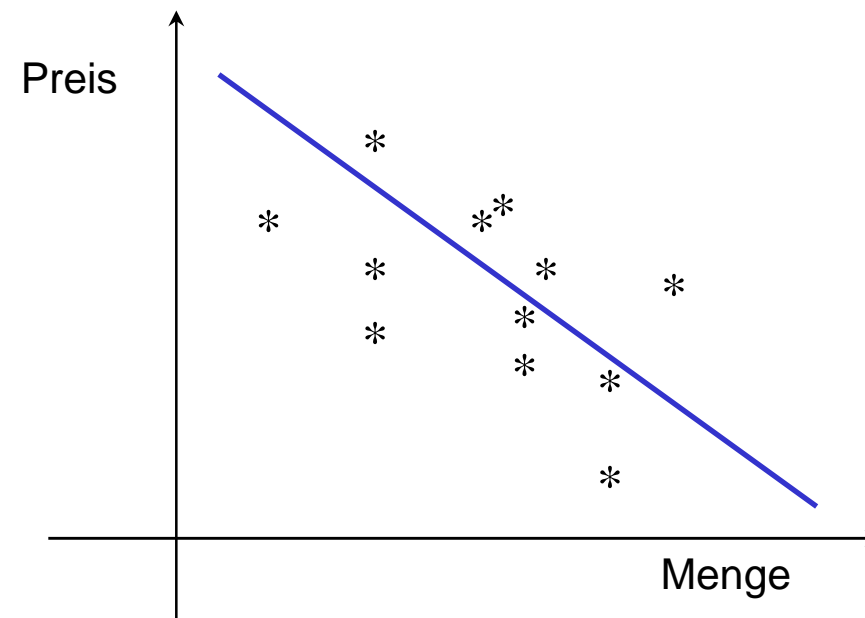
Eine Reihe von statistischen Einheiten ist durch zwei Merkmale, z.B. Absatzmenge und Verkaufspreis, gekennzeichnet.

Da ein Zusammenhang nahe liegt, soll dieser durch eine - im einfachsten Fall lineare - Funktion beschrieben werden.



Regressionsrechnung

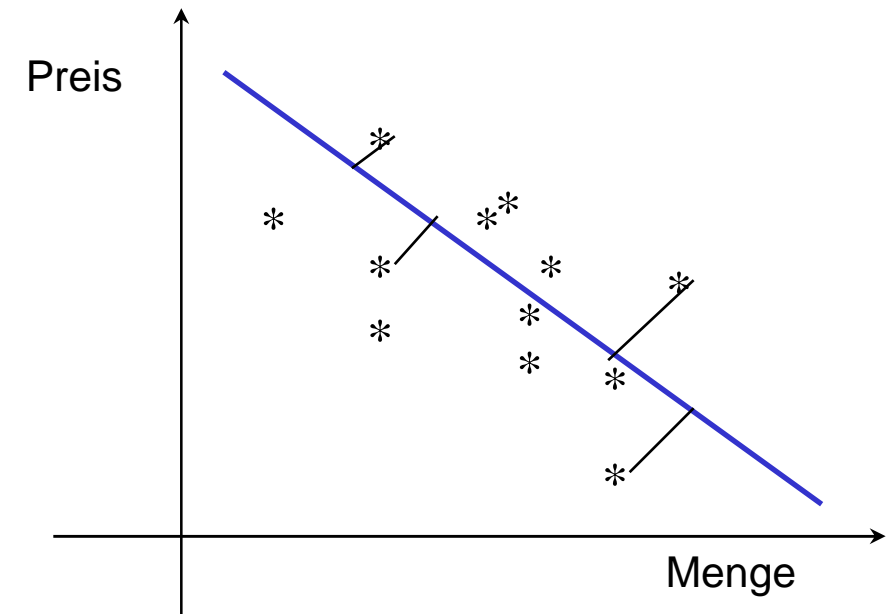
Grafisch bedeutet das, dass man eine Gerade sucht, die den gemessenen Werten möglichst nahe kommt, d.h. die Summe der Abweichungen von der Geraden soll minimiert werden.





Regressionsrechnung

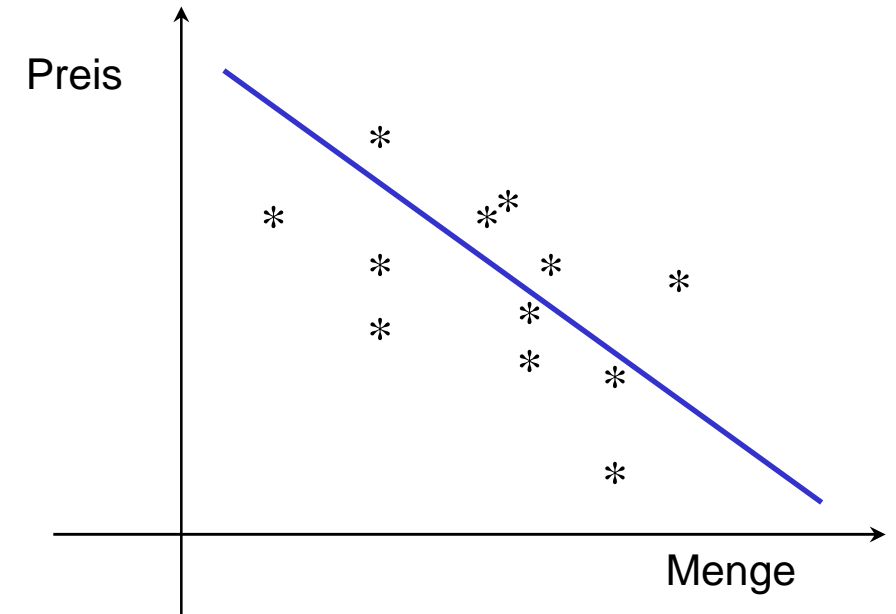
Größere Abweichungen werden stärker gewichtet, da die Summe der Quadrate der Abweichungen minimiert wird, um vorzeichenwechselbedingte Neutralisationen der Abweichungen zu vermeiden.





Korrelationsrechnung

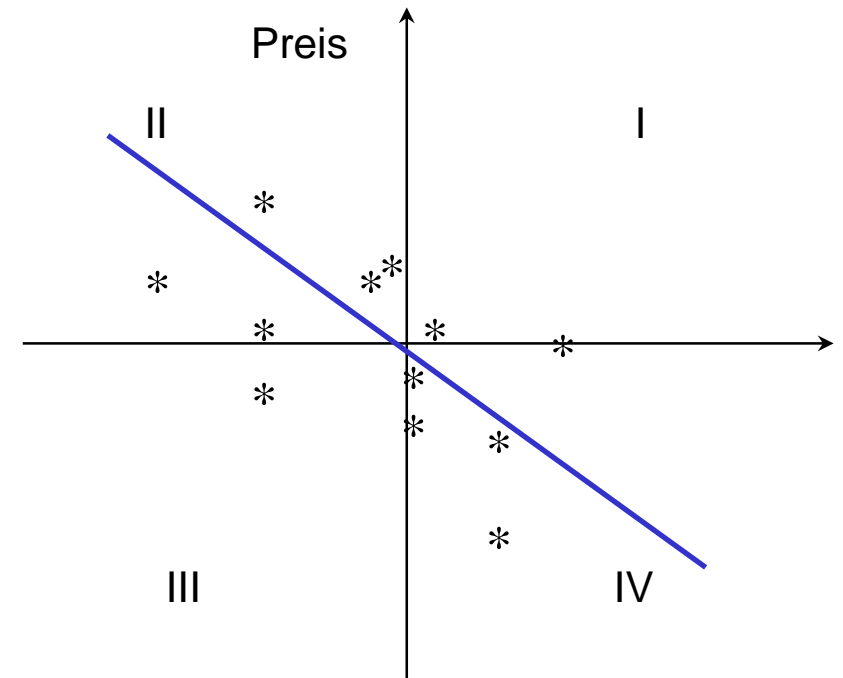
Der Korrelationskoeffizient gibt an, wie eng der Zusammenhang zwischen der berechneten Regressionsgeraden und den beobachteten Werten ist, indem er die Verteilung und den Abstand in Bezug auf die Regressionsgerade misst.





Korrelationsrechnung

Der Korrelationskoeffizient beträgt bei exaktem negativem oder positivem linearem Zusammenhang -1 bzw. $+1$. Bei schwächerem Zusammenhang liegt er zwischen -1 und 0 bzw. 0 und $+1$. Bei 0 gibt es überhaupt keinen Zusammenhang (Korrelation) zwischen den untersuchten Parametern.





Aufgabe 1 Für sechs Monate liegen die Daten über den Hypothekenzinssatz X sowie über den saisonbereinigten Auftragseingang Y im Privatwohnungsbaugewerbe vor:

Wer bauen will, braucht i.d.R. Fremdkapital. Je günstiger die Hypothekenzinsen, desto günstiger ist die Finanzierung und desto mehr Bauvorhaben werden c.p. realisiert.

b)

Monat i	Zinssatz x_i	Aufträge y_i	$x_i y_i$	x_i^2	y_i^2
1	6	3,0	18,0	36	9,00
2	5	3,2	16,0	25	10,24
3	7	2,5	17,5	49	6,25
4	7	2,3	16,1	49	5,29
5	8	2,0	16,0	64	4,00
6	9	2,0	18,0	81	4,00
Summe	42	15,0	101,6	304	38,78

Regressionsfunktion: $y = 4,88 - 0,34x$

Prognose: bei 4% ist **3,52**, bei 7% ist **2,33** Mio. GE Auftragsvolumen zu erwarten.



Aufgabe 2

Monat i	Menge x_i (in TME)	Kosten y_i (in TGE)	$x_i y_i$	x_i^2	y_i^2
1	20	800	16.000	400	640.000
2	30	900	27.000	900	810.000
3	15	750	11.250	225	562.000
4	25	830	20.750	625	688.900
5	30	920	27.600	900	846.400
6	20	780	15.600	400	608.400
Summe	140	4.980	118.200	3.450	4.156.200

Kostenfunktion (Regressionsfunktion): $K = 575 + 10,91x$



Zusammenfassung

In der Regressionsrechnung geht es darum, eine Funktion zu bestimmen, die den zahlenmäßigen Zusammenhang zwischen zwei Merkmalen möglichst gut wiedergibt. Diese Funktion läßt sich für die Beobachtungswertepaare zweier quantitativer Merkmale immer berechnen, auch wenn der lineare Zusammenhang nur schwach oder gar nicht vorhanden ist. Sie gibt also keine Auskunft darüber, wie stark der statistische Zusammenhang ist. Genau darum geht es in der Korrelationsrechnung: der Korrelationskoeffizient ist ein Maß für die Stärke des linearen Zusammenhangs zweier quantitativer Merkmale.



Wahrscheinlichkeitsrechnung



Wahrscheinlichkeitsrechnung

Gegenstand der Wahrscheinlichkeitsrechnung ist die Untersuchung zufälliger Ereignisse mit den klassischen Annahmen:

- die Elementarereignisse schließen einander aus
- alle Elementarereignisse sind gleich möglich
- genau ein Elementarereignis tritt ein



Wahrscheinlichkeitsrechnung

Die Wahrscheinlichkeit ist definiert als:

$$P = \frac{\text{Anzahl der günstigen Elementarereignisse}}{\text{Anzahl aller Elementarereignisse}}$$



Tabellenform

Alle Alternativen eines Zufallsexperiments werden aufgezeichnet

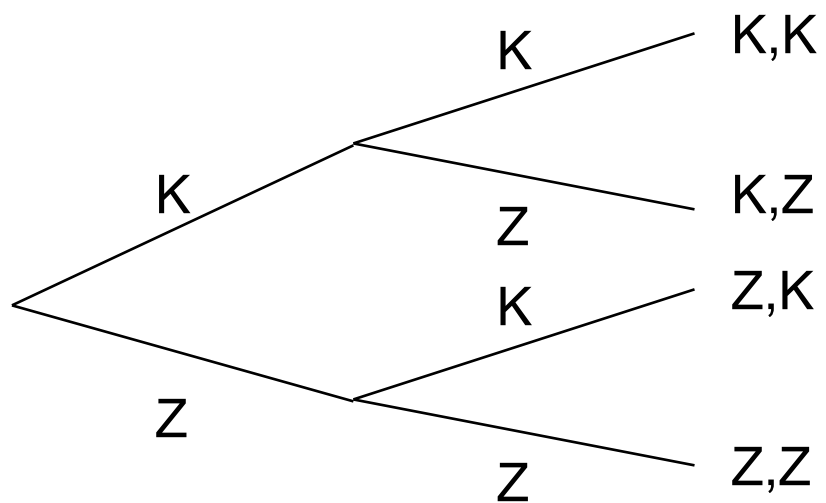
Beispiel: zweimaliger Münzwurf
(Kopf/Zahl)

1. Wurf	2. Wurf	
K	KK	KZ
Z	ZZ	ZK



Baumdiagramme

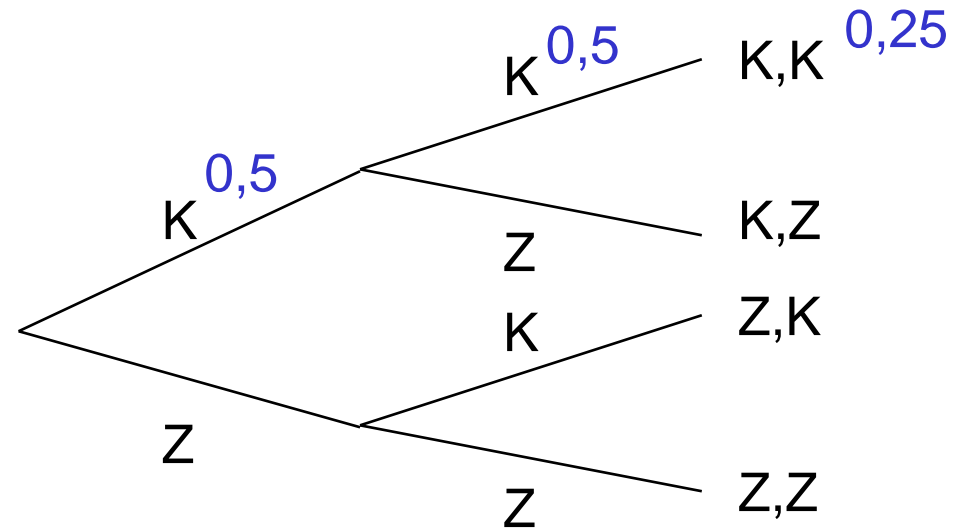
Alle Alternativen eines Zufallsexperiments werden aufgezeichnet





Baumdiagramme

Die Wahrscheinlichkeit eines Ereignisses ist gleich dem Produkt der Wahrscheinlichkeiten längs des zugehörigen Pfades.

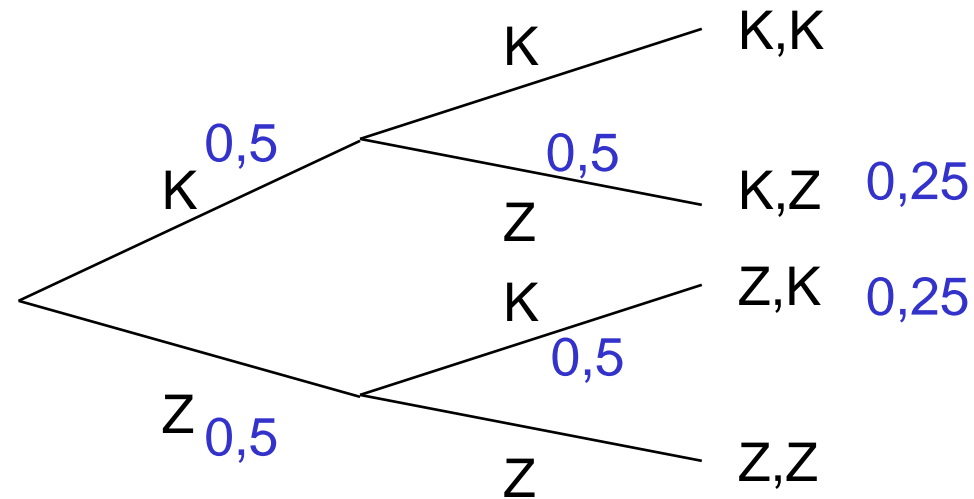


Die Wahrscheinlichkeit für „Kopf“ beim 1. und 2. Wurf ist:
 $0,5 \times 0,5 = 0,25$



Baumdiagramme

Die Wahrscheinlichkeit eines Ereignisses, das sich aus mehreren Pfaden zusammen setzt, ist gleich der Summe der einzelnen Pfadwahrscheinlichkeiten



Die Wahrscheinlichkeit für „Kopf“ und „Zahl“ ist:
 $0,25 + 0,25 = 0,5$



Aufgabe 1

Ein idealer Würfel wird geworfen. Wie hoch ist die Wahrscheinlichkeit, eine Augenzahl von

- a) genau 4
- b) höher als 1
- c) höchstens 3 zu erzielen?

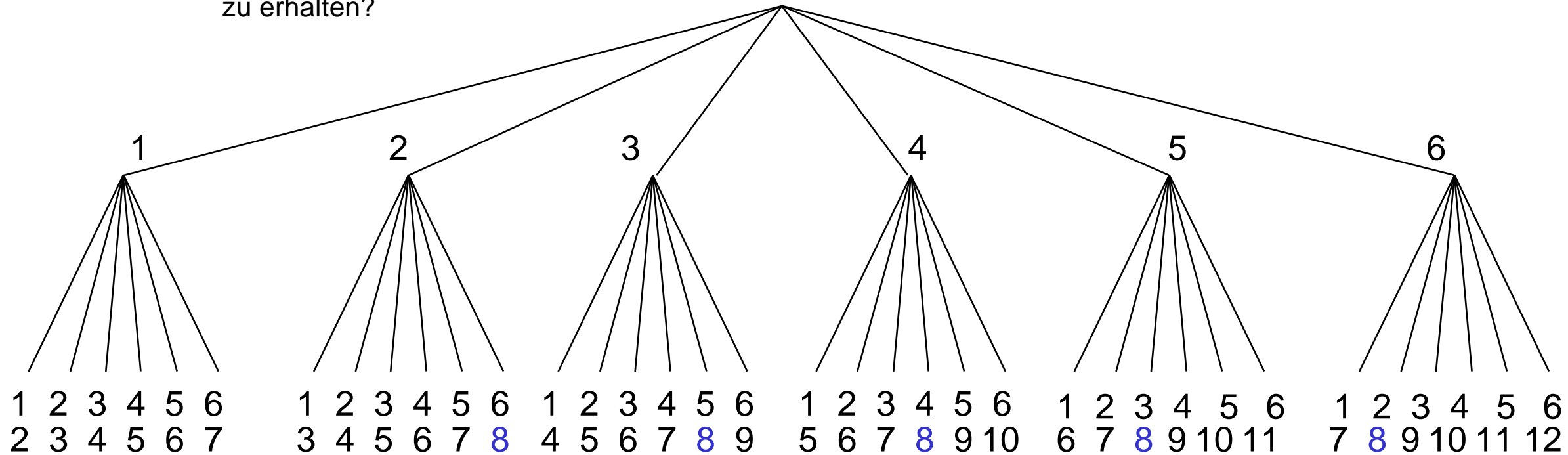
- a) **$1/6$** b) **$5/6$** c) **$3/6$**



Aufgabe 2a)

Zwei ideale Würfel werden geworfen. Wie hoch ist die Wahrscheinlichkeit, eine Augensumme von

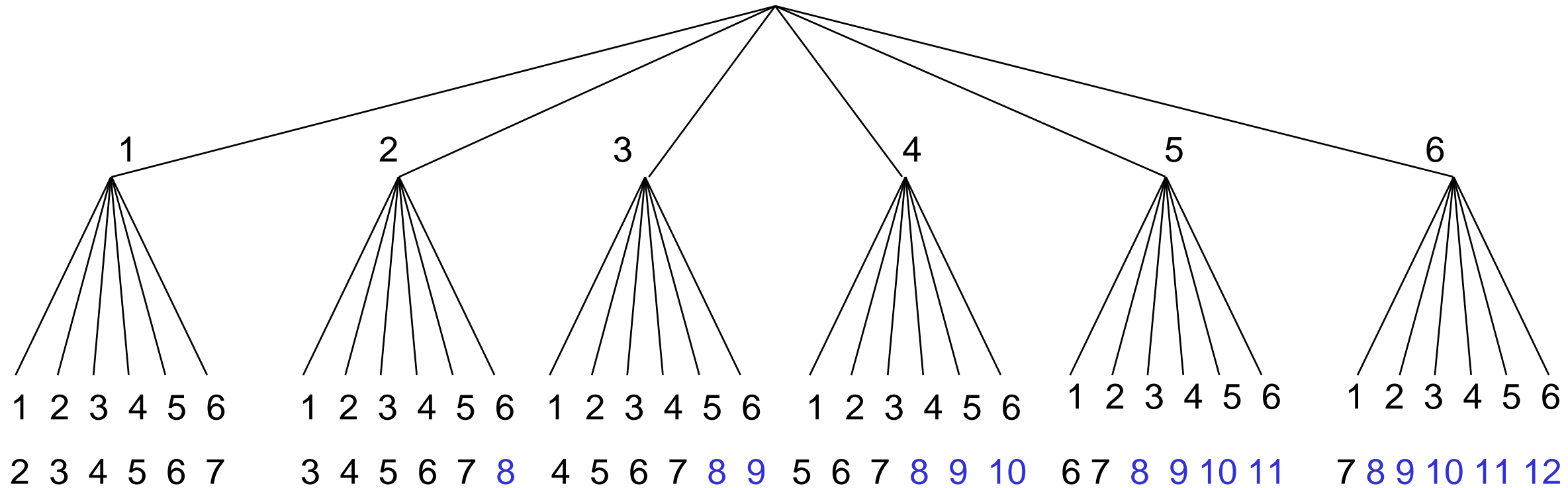
- genau 8
- mindestens 8
- höchstens 8
zu erhalten?



$$p = 5/36$$



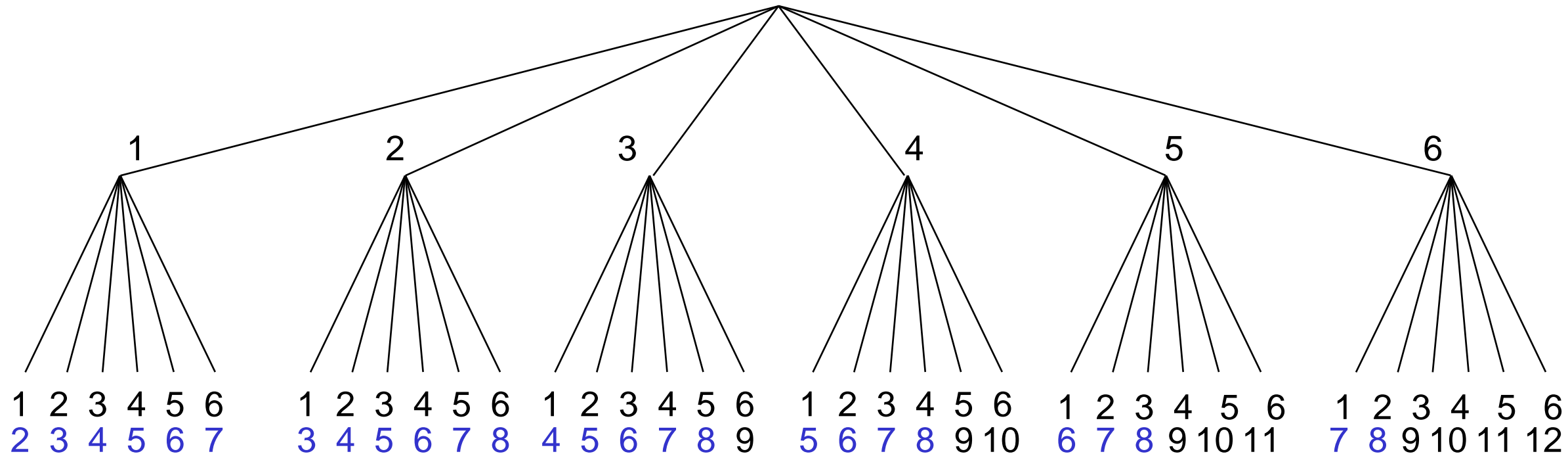
Aufgabe 2b)



$$p = 15/36$$



Aufgabe 2c)

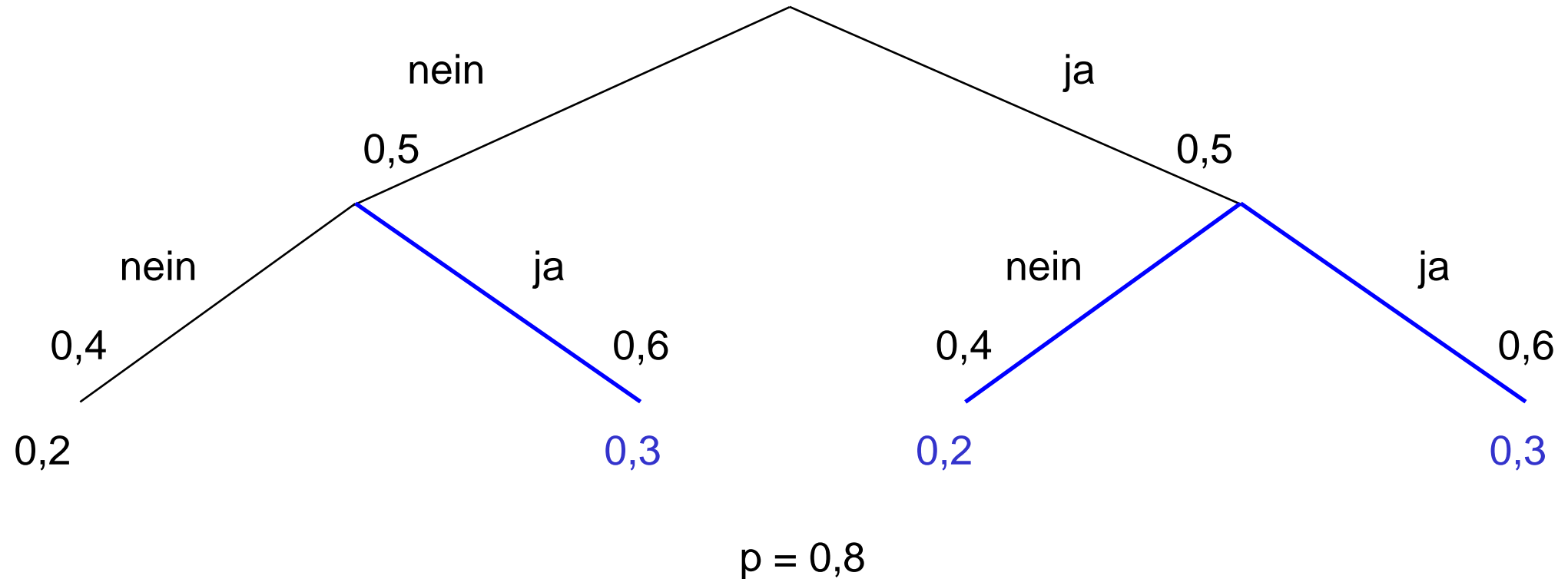


$$p = 26/36$$



Aufgabe 3

Jemand bewirbt sich bei zwei Firmen (A und B). Die (unabhängigen) Wahrscheinlichkeiten dafür, angenommen zu werden, schätzt er bei A auf 0,5, bei B auf 0,6. Wie groß ist die Wahrscheinlichkeit, von wenigstens einer der beiden Firmen eine Zusage zu erhalten?



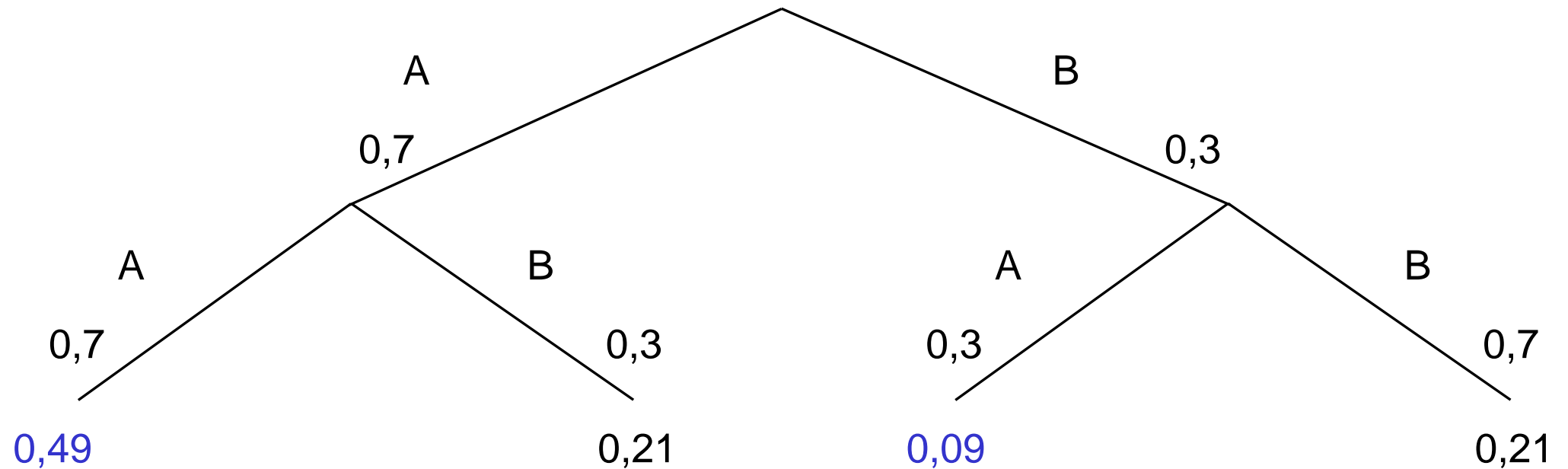


Aufgabe 4a)

Ein Vertreter kauft jedes Jahr ein neues Kraftfahrzeug. In Frage kommen für ihn nur die Typen A und B. Die Wahrscheinlichkeit, dass er im nächsten Jahr genau den gleichen Typ fährt, den er schon im Jahr zuvor fährt, liegt bei 0,7. Im Moment fährt er Typ A. Wie groß ist die Wahrscheinlichkeit, dass er

a) im übernächsten Jahr ebenfalls Typ A fährt

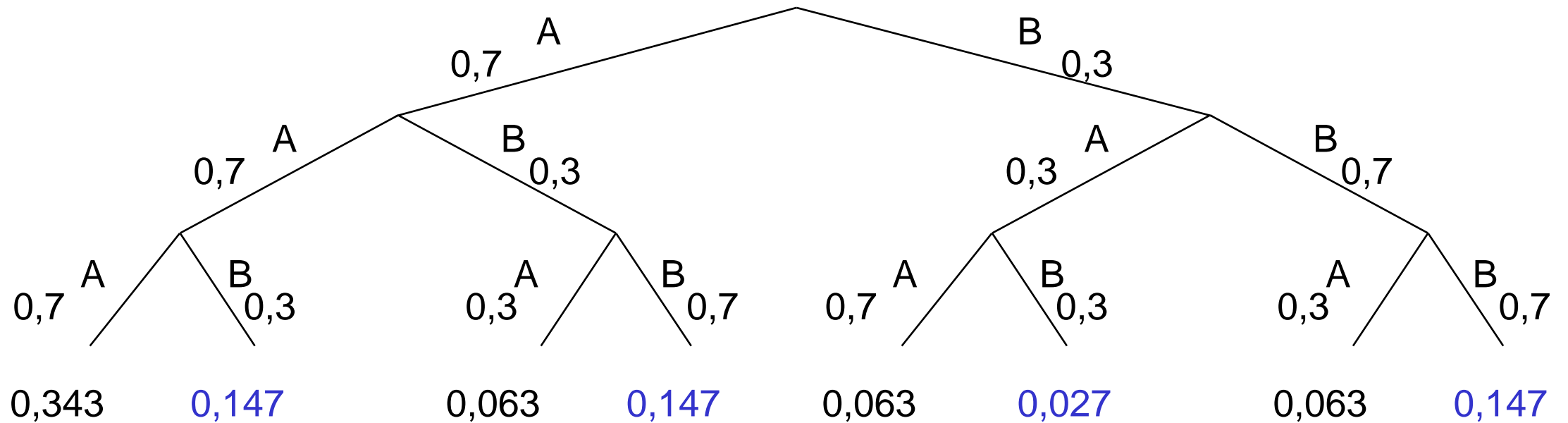
b) in drei Jahren Typ B fährt?



$$p = 0,58$$



Aufgabe 4b)

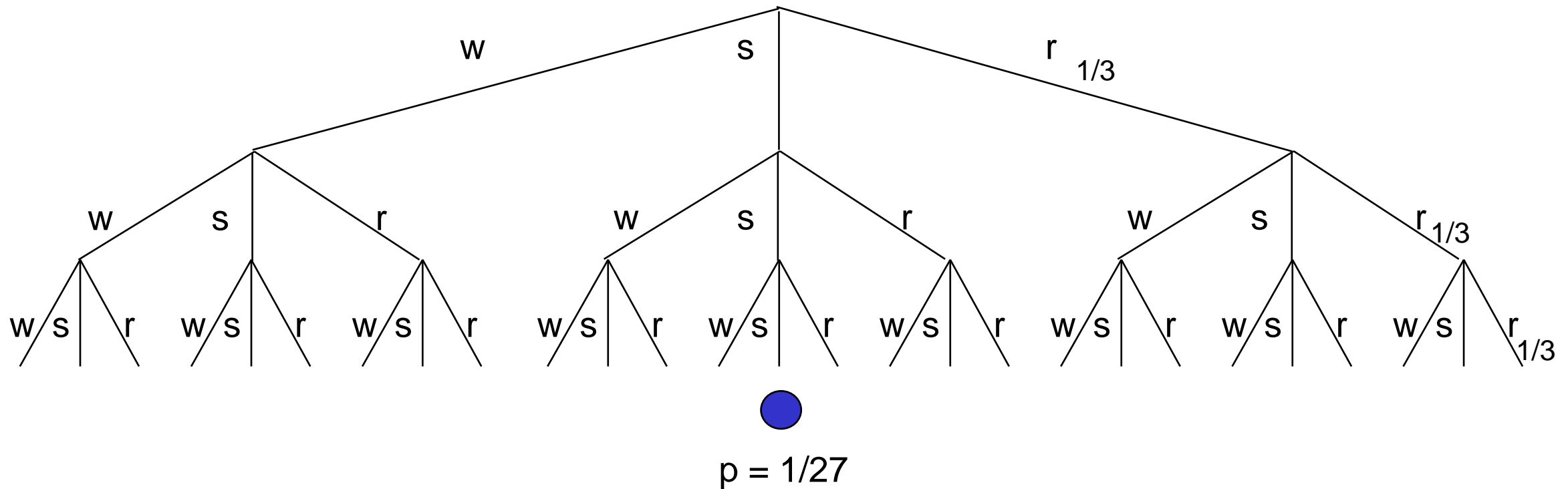


$$p = 0,468$$



Aufgabe 5a)

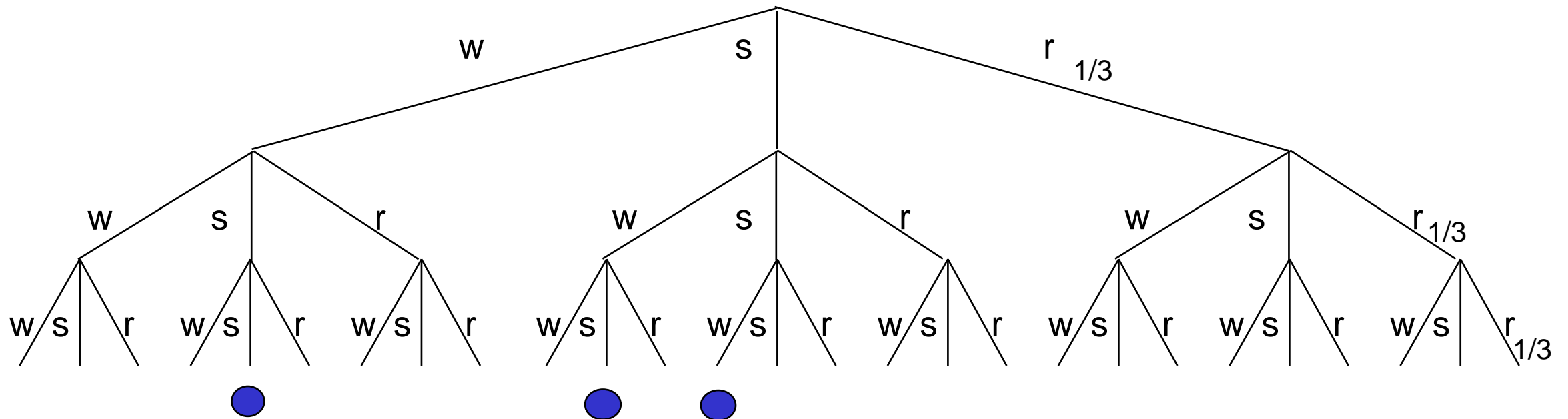
In einer Urne befinden sich zwei weiße, zwei schwarze und zwei rote Kugeln. Es werden nacheinander drei Kugeln gezogen; jede einzelne gezogene Kugel wird nach dem Ziehen sofort wieder in die Urne zurückgelegt. Wie hoch ist die Wahrscheinlichkeit dafür, a) drei schwarze Kugeln,





Aufgabe 5b)

b) zwei schwarze und eine weiße Kugel,

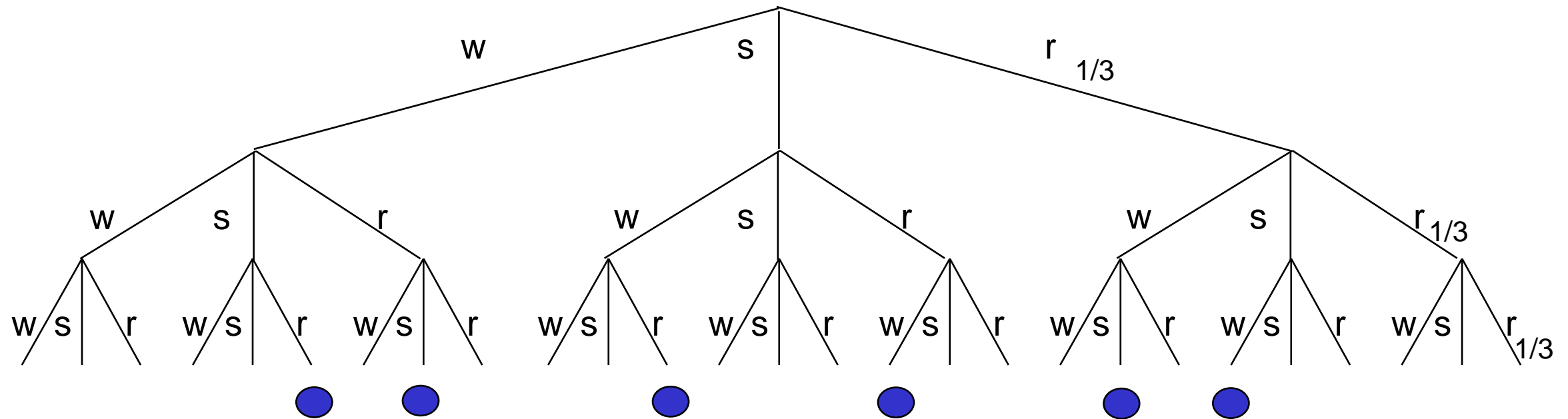


$$p = 3/27$$



Aufgabe 5 c)

eine schwarze, eine weiße und eine rote Kugel zu ziehen?



$$p = \frac{6}{27}$$