Teil IV:

Wahrscheinlichkeit und Statistik (WS)

Für einen meiner Enkel habe ich eine Handreichung ausgearbeitet, um ihm dabei behilflich zu sein, die schriftliche Reifeprüfung in Mathematik (österr. Zentralmatura) zu bestehen. Diese umfasst nur ein Wissen und Können, das eine Voraussetzung für die Bewältigung der 24 Aufgaben des "Kernbereichs" darstellt. (16 "Richtige" reichen für eine positive Beurteilung der ganzen Zentralmatura aus.) Dieser Lehrstoff wurde im Auftrag des BMU von einer Projektgruppe zusammengestellt, in vier Teile gegliedert und inhaltlich kurz beschrieben. Aus jedem dieser vier Teile kommen genau sechs Aufgaben zur Zentralmatura.

Dies ist Teil IV der kompletten Handreichung. Meine Arbeit ist nicht perfekt und sicher auch noch ausbaufähig. Gleichwohl halte ich sie aber schon in dieser Form für geeignet, einem größeren Kreis von AHS-Maturanten dienen zu können, nicht zuletzt deshalb, weil ich über 150 (das sind an die 40 %) der seit 2014 gestellten Maturaaufgaben hineinkopiert habe. An ihnen ist auch erkennbar, auf welches Wissen und Können ganz konkret Wert gelegt wird.

Außerdem habe ich jeweils zu Beginn die von der Projektgruppe erstellten Anforderungen aufgelistet. Wo ich in meinen Ausarbeitungen ein wenig darüber hinausgegangen bin ist das im Text vermerkt bzw. *kursiv geschrieben*. Andererseits sind Sätze und Formeln vielfach ohne Beweis angegeben. Daher nochmals: Diese meine Arbeit ist keine wissenschaftliche, den angestrebten Zweck sollte sie aber doch erfüllen.

Zuletzt: Alle seit 2014 gegebenen Maturaaufgaben finden sich einschließlich der Lösungen im Internet unter verschiedenen Adressen, z. B. unter www.mathago.at/zentralmatura.

ANFORDERUNGEN:

1. Grundkompetenzen Beschreibende Statistik

WS 1.1 Werte aus tabellarischen und elementaren grafischen Darstellungen ablesen (bzw. zusammengesetzte Werte ermitteln, d. h. aus den Grafiken ablesbare Daten zur Berechnung weiterer Kennzahlen verwenden können) und im jeweiligen Kontext angemessen interpretieren können.

Anmerkungen: (un)geordnete Liste, Strichliste, Piktogramm, Säulen-, Balken-, Linien-, Stängel-Blatt-, Punktwolkendiagramm, Histogramm (als Spezialfall eines Säulendiagramms), Prozentstreifen, Kastenschaubild (Boxplot)

WS 1.2 Tabellen und einfache statistische Grafiken erstellen, zwischen Darstellungsformen wechseln können

WS 1.3 statistische Kennzahlen (absolute und relative Häufigkeiten; arithmetisches Mittel, Median, Modus, Quartile, Spannweite, empirische Varianz/Standardabweichung) im jeweiligen Kontext interpretieren können; die angeführten Kennzahlen für einfache Datensätze ermitteln können

WS 1.4 Definition und wichtige Eigenschaften des arithmetischen Mittels und des Medians angeben und nutzen, Quartile ermitteln und interpretieren können, die Entscheidung für die Verwendung einer bestimmten Kennzahl begründen können.

Anmerkungen: Wenn auch statistische Kennzahlen (für einfache Datensätze) ermittelt und elementare statistische Grafiken erstellt werden sollen, liegt das Hauptaugenmerk auf verständigen Interpretationen von Grafiken (unter Beachtung von Manipulationen) und Kennzahlen. Speziell für das arithmetische Mittel und den Median (auch als Quartile) müssen die wichtigsten Eigenschaften (definitorische Eigenschaften, Datentyp-Verträglichkeit, Ausreißerempfindlichkeit) gekannt und verständig eingesetzt bzw. berücksichtigt werden. Beim arithmetischen Mittel sind allenfalls erforderliche Gewichtungen zu beachten („gewogenes arithmetisches Mittel“) und zu nutzen (Bildung des arithmetischen Mittels aus arithmetischen Mitteln von Teilmengen).

2. Wahrscheinlichkeitsrechnung - Grundbegriffe

WS 2.1 Grundraum (Menge der möglichen Versuchsausgänge) und Ereignisse in angemessenen Situationen verbal bzw. formal angeben können

WS 2.2 relative Häufigkeit als Schätzwert von Wahrscheinlichkeit verwenden und anwenden können

WS 2.3 Wahrscheinlichkeiten unter der Verwendung der Laplace-Annahme (Laplace-Wahrscheinlichkeit) berechnen und interpretieren können, Additionsregel und Multiplikationsregel anwenden und interpretieren können
Anmerkungen: Die Multiplikationsregel kann unter Verwendung der kombinatorischen Grundlagen und der Anwendung der Laplace-Regel (auch) umgangen werden.

WS 2.4 Binomialkoefizienten berechnen und interpretieren können

3. Wahrscheinlichkeitsverteilung(en)

WS 3.1 die Begriffe *Zufallsvariable, (Wahrscheinlichkeits-)Verteilung, Erwartungswert* und *Standardabweichung* verständig deuten und einsetzen können

WS 3.2 Binomialverteilung als Modell einer diskreten Verteilung kennen – Erwartungswert sowie Varianz/Standardabweichung binomialverteilter Zufallsgrößen ermitteln können, Normalapproximation der Binomialverteilung interpretieren und anwenden können, Arbeiten mit der Binomialverteilung in anwendungsorientierten Bereichen

WS 3.3 Situationen erkennen und beschreiben können, in denen mit Binomialverteilung modelliert werden kann

WS 3.4 Normalapproximation der Binomialverteilung interpretieren und anwenden können

Anmerkungen: Kennen und Anwenden der Faustregel, dass die Normalapproximation der Binomialverteilung mit den Parametern n und p dann anzuwenden ist und gute Näherungswerte liefert, wenn die Bedingung $n \cdot p \cdot (1-p) \geq 9$ erfüllt ist. Die Anwendung der Stetigkeitskorrektur ist nicht notwendig und daher für Berechnungen im Zuge von Prüfungsaufgaben vernachlässigbar. Kennen des Verlaufs der Dichtefunktion φ der Standardnormalverteilung **mit Erwartungswert 0 und Standardabweichung 1**. Arbeiten mit der Verteilungsfunktion Φ der Standardnormalverteilung und korrektes Ablesen der entsprechenden Werte.

4. Schließende/beurteilende Statistik

WS 4.1 Konfidenzintervalle als Schätzung für eine Wahrscheinlichkeit oder einen unbekannten Anteil p interpretieren (frequentistische Deutung) und verwenden können, Berechnungen auf Basis der Binomialverteilung oder einer durch die Normalverteilung approximierten Binomialverteilung durchführen können.

1. Beschreibende Statistik

1. Grundbegriffe der Statistik:

Die **Grundgesamtheit** G ist eine (i. A. sehr mächtige) Menge von Merkmalsträgern.
Die **Stichprobe** $S \subseteq G$ ist eine („zufällige“) endliche Teilmenge von G. S „ersetzt“ G bei der statistischen Untersuchung von G; $S = \{s_1, s_2, s_3, ...\}$, ihre Mächtigkeit ist $|S| = n$.
Das **Merkmal** (= die Variable X) ist eine bestimmte Eigenschaft der Elemente von G bzw. S und *der Gegenstand der Untersuchung.* Die Elemente von S werden als Merkmalsträger bezeichnet.
Die Merkmalsausprägung ist die Erscheinungsform $x_1, x_2, x_3, ...$des Merkmals, zusammengefasst in einer Menge A mit $|A| = k$

Die Funktion $x = X(s)$: $S \to A$ ordnet jedem Merkmalsträger $s_i \in S$ die entsprechende Ausprägung $x_i \in A$ zu.

Beispiel 1 (Teil 1): G sei die Menge aller in Österreich zugelassenen PKW; S ist die Menge aller auf einem Großparkplatz parkenden PKW (oder die Menge aller PKW, die innerhalb einer Stunde ein bestimmtes Straßenstück durchfahren). X... Automarke (oder Kennzeichen bzw. Bezirk oder Farbe), daher A = {VW, Opel, Ford, ...} oder A = {SR, SE, AM, LL, ...} oder ...

Beispiel 2 (Teil 1): G sei die Menge aller Würfe mit einem Würfel, das ist eine unendl. Menge, S sei die Menge von n Würfen, X der Ausfall des Wurfes (= die gewürfelte Punktezahl), daher $A = \{1,2,3,4,5,6\}$, daher $|A| = 6$.

Die **Absolute Häufigkeit** ist die Anzahl H_i der Merkmalsträger $\in S$ mit der Ausprägung x_i

Die **Relative Häufigkeit** $h_i = H_i : n$ $(0 \leq h_i \leq 1)$, zusammengefasst in einer Menge B. Sie wird gerne in Prozent angegeben, also in $100.h_i$.

Die Funktion $h = H(X)$: $A \to B$ heißt **Häufigkeitsverteilung**

Sie lässt sich in einer Wertetabelle zusammenfassen und (u. A.) durch ein Kreisdiagramm (eine Kreisgraphik), ein Histogramm (Streifendiagramm, Säulendiagramm, eine Balkengraphik), ein Punkt-, Linien- oder ein Stabdiagramm veranschaulichen. Beispiele dazu sind bei den Maturaaufgaben ersichtlich.

Beispiel 1 (Teil 2): Die Zählung hat $|S| = n = 213$ PKW ergeben, davon VW 47, Opel 33, Ford 26, Fiat 21, Toyota 21, Nissan 19, Renault 17, Sonstige 29

Wertetabelle:

x_i	VW	Op	Fo	Fi	To	Ni	Re	So
h_i	0,22	0,15	0,12	0,10	0,10	0,09	0,08	0,14

Beispiel 2 (Teil 2): Das Experiment (50 Würfe, also $|S| = n = 50$ hat ergeben: 1 kam 7x, 2 kam 11x, 3 kam 7x, 4 kam 6x, 5 kam 8x, 6 kam 11x

Wertetabelle:

x_i	1	2	3	4	5	6
h_i	0,14	0,22	0,14	0,12	0,16	0,22

2. Statistische Kennzahlen:

Arithmetisches Mittel: Seien $x_1, x_2, \ldots x_n$ n reelle Zahlen, so ist $a = \frac{x_1 + x_2 + \ldots + x_n}{n}$ das arithmetische Mittel dieser n Zahlen. Kommen in diesem Datensatz nur $k < n$ verschiedene Werte $x_1, x_2, \ldots, x_k$ mit den absoluten Häufigkeiten $H_1, H_2, \ldots, H_k$ vor, dann ist das arithm. Mittel

$$\boxed{a = \frac{x_1.H_1 + x_2.H_2 + \ldots + x_k.H_k}{n} = x_1.h_1 + x_2.h_2 + \ldots + x_k.h_k}$$

und wird **Mittelwert** (od. gewogenes Mittel) **der** entsprechenden **Häufigkeitsverteilung** genannt.

Median: Das ist die mittlere Zahl eines Datensatzes bzw. das arithm. Mittel aus den zwei mittleren.

Beispiel 2 (Teil 3): Der Mittelwert beträgt 3,6, der Median 3,5: Die 25. Zahl (von 50) ist 3, die 26. Zahl ist 4, (3 + 4) : 2 = 3,5.

Quartile: Das *erste* ***(untere)*** *Quartil* ist definiert als die mittlere Zahl zwischen der kleinsten Zahl und dem Median eines Datensatzes. Das *zweite Quartil* ist der Median der Daten. Das *dritte* ***(obere)*** *Quartil* ist der mittlere Wert zwischen dem Median und dem höchsten Wert des Datensatzes. Zwischen dem ersten und dem dritten Quartil liegen somit die mittleren 50 % des Datensatzes.

Modus (od. **Modalwert):** Der Modus einer Stichprobe ist jener Einzelwert, der am häufisten vorkommt.

Spannweite: Die Spannweite einer Stichprobe ist die Differenz aus dem höchsten und dem niedrigsten Wert.

Empirische Varianz und Standardabweichung:

Die ***Standardabweichung*** *(*oder *Streuung*) σ einer Häufigkeitsverteilung ist die Quadratwurzel aus der ***Varianz*** σ^2, und diese ist der Mittelwert der Quadrate der Differenzen $(x_i - m)$:

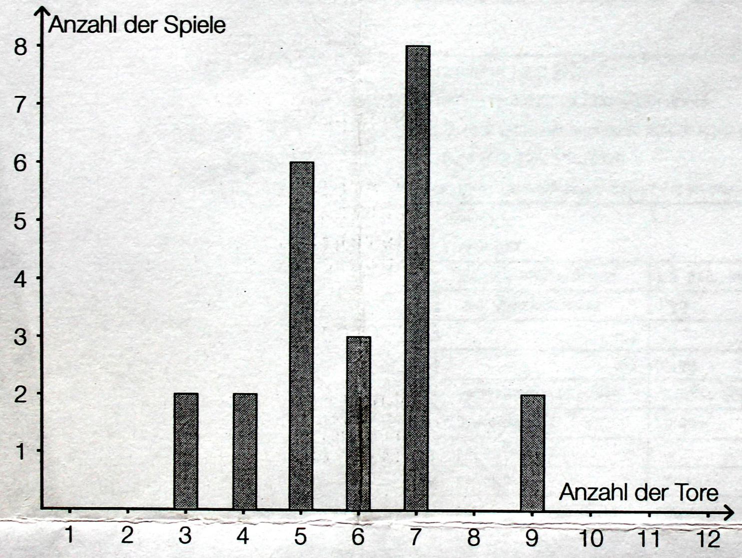
$$\boxed{\textbf{Varianz } \sigma^2 = \Sigma(x_i - m)^2.h_i}$$

Beispiel 3: Eine Schularbeit der sechsten Klasse wurde folgendermaßen benotet: S = {2, 3, 5, 2, 2, 4, 3, 1, 4, 2, 2, 3, 1, 5, 2, 4, 3, 1, 5, 3, 3, 3, 2, 3}. Um den Modus zu bestimmen, zählen wir, wie oft jeder Wert auftritt: 1 tritt 3 mal, 2 tritt 7 mal, 3 tritt 8 mal, 4 tritt 3 mal und 5 tritt auch 3 mal auf. Der Modalwert ist der häufigste Wert, also hier die Note 3. Die Spannweite ist 5 − 1 = 4. Der Median des Datensatzes ist 3, das erste Quartil ist 2 und das dritte ist 4.

Mittelwert: $m = 1/24.(1.3 + 2.7 + 3.8 + 4.3 + 5.3) \approx 2{,}8$
Varianz: $\sigma^2 \approx (-1{,}8)^2.1/8 + (-0{,}8)^2.7/24 + 0{,}2^2.1/3 + 1{,}2^2.1/8 + 2{,}2^2.1/8 = 1{,}39$
Standardabweichung: $\sigma \approx 1{,}179$.

Mat. 16/HT/Aufg. 20/1:

In der österreichischen Eishockeyliga werden die Ergebnisse aller Spiele statistisch ausgewertet. In der Saison 2012/13 wurde über einen bestimmten Zeitraum erfasst, in wie vielen Spielen jeweils eine bestimmte Anzahl an Toren erzielt wurde. Das nachstehende Säulendiagramm stellt das Ergebnis dieser Auswertung dar.



Aufgabenstellung:

Bestimmen Sie den Median der Datenliste, die dem Säulendiagramm zugrunde liegt!

Die „Datenliste, die dem gegebenen Säulendiagramm zugrunde liegt", ist offenbar jene Liste, die jedem Spiel die Anzahl der erzielten Tore zuordnet, wie beim obigen Musterbeispiel jedem Schüler „seine" Note zugeordnet wurde. Also S = {3, 3, 4, 4, 5, 5, 5, 5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 9, 9}. Vergleichsweise: Der Mittelwert liegt bei ca. 5,9.

Mat. 18/2. NT/Aufg. 20/1

In einem Gymnasium wurden in den 24 Unterstufenklassen folgende Klassenschülerzahlen er
hoben:

Klassenschülerzahl	20	21	22	23	24	25	26	27	28
Anzahl Klassen	1	2	1	2	3	2	4	6	3

Aufgabenstellung:

Ermitteln Sie den Median der Klassenschülerzahlen in der Unterstufe dieses Gymnasiums!

Median 26, Mittel ca. 25,1

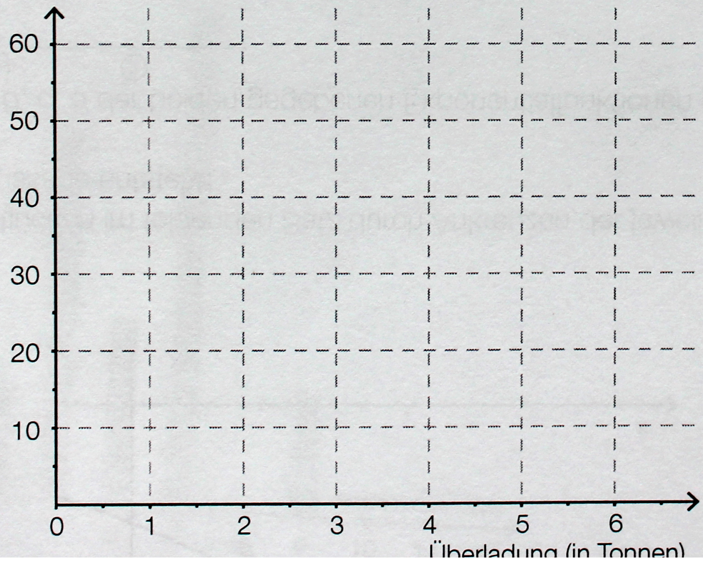
Mat. 16/HT/Aufg. 19/1:

Bei einer Verkehrskontrolle wurde die Beladung von LKW überprüft. 140 der überprüften LKW waren überladen. Details der Kontrolle sind in der nachstehenden Tabelle zusammengefasst.

Überladung $Ü$ in Tonnen	$Ü < 1\,t$	$1\,t \leq Ü < 3\,t$	$3\,t \leq Ü < 6\,t$
Anzahl der LKW	30	50	60

Aufgabenstellung:

Stellen Sie die Daten der obigen Tabelle durch ein Histogramm dar! Dabei sollen die absoluten Häufigkeiten als Flächeninhalte von Rechtecken abgebildet werden.



Mat. 18/1. NT/Aufg. 20/1: Änderung einer Datenliste

Gegeben ist eine Datenliste $x_1, x_2, ..., x_n$ mit n Werten und dem arithmetischen Mittel a. Diese Liste wird um zwei Werte x_{n+1} und x_{n+2} ergänzt, wobei das arithm. Mittel der nunmehr n + 2 Daten gleich bleibt. Aufgabenstellung: Geben Sie für diesen Fall einen Zusammenhang zwischen x_{n+1}, x_{n+2} und a mithilfe einer Formel an.

Mat. 14/HT/21/1:

Neun Athleten eines Sportvereins absolvieren einen Test; der Mittelwert der Testergebnisse ist 8. Ein zehnter holt den Test nach und erreicht ein Ergebnis von 4. Wie groß ist nun das arithm. Mittel der Testergebnisse aller zehn Athleten?

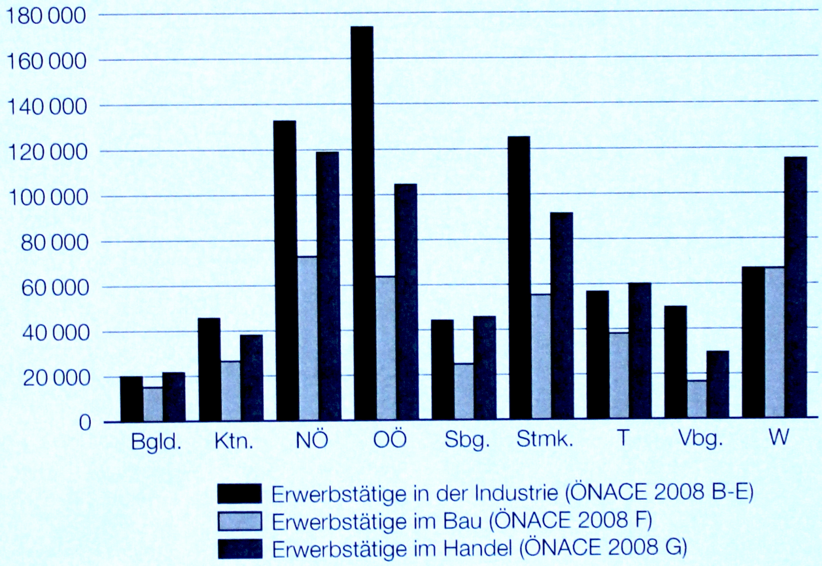
Mat. 15/2. NT/Aufg. 20/1: Median und Modus

Gegeben ist eine ungeordnete Liste von 19 natürlichen Zahlen: 5, 15, 14, 2, 5, 13, 11, 9, 7, 16, 15, 9, 10, 14, 3, 14, 5, 15, 14. Aufgabenstellung: Geben Sie den Median und den Modus dieser Liste an!

Mat. 18/2. NT/Aufg. 19/1:

Die nachstehende Grafik zeigt die Anzahl der im Jahr 2012 in Österreich Erwerbstätigen in drei Bereichen. Die Grafik weist die Daten nach Bundesländern getrennt aus.

Erwerbstätige in Industrie, Bau und Handel 2012
nach Bundesländern



Quelle: STATISTIK AUSTRIA, Mikrozensus-Arbeitskräfteerhebung 2012. Erstellt am 22.05.2013.

Aufgabenstellung:

Welche der folgenden Aussagen lässt/lassen sich aus der Grafik für das Jahr 2012 ableiten? Kreuzen Sie die zutreffende(n) Aussage(n) an!

In jedem Bundesland gab es mehr Erwerbstätige im Handel als im Bau.	☐
In der Industrie hatte Oberösterreich (OÖ) mehr Erwerbstätige als jedes andere Bundesland.	☐
Wien (W) hatte mehr Erwerbstätige im Handel als in Industrie und Bau zusammen.	☐
Vorarlberg (Vbg.) hatte in allen drei Bereichen zusammen weniger Erwerbstätige als die Steiermark (Stmk.) alleine in der Industrie.	☐
Im Handel hatte Burgenland (Bgld.) weniger Erwerbstätige als jedes andere Bundesland.	☐

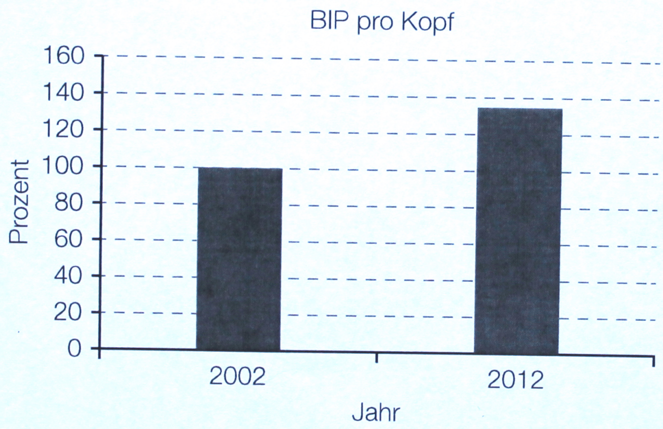
Mat. 16/2. NT/Aufg. 19/1:

In einer Schule gibt es vier Sportklassen: $S1$, $S2$, $S3$ und $S4$. S1 hat 18 Schüler, die im ersten Semester im Mittel 45,5 Unterrichtsstunden versäumt haben; in S2 lauten die entsprechenden Daten 20 Schüler und 63,2 Stunden, in S3 16 Schüler und 70,5 Stunden, in S4 15 Schüler und 54,6 Stunden. Aufgabenstellung: Berechnen Sie das arithmetische Mittel der versäumten Unterrichtsstunden aller Schüler der vier Sportklassen für den angegebenen Zeitraum!

Mat. 18/1. NT/Aufg. 19/1:

Das *nominale Bruttoinlandsprodukt* gibt den Gesamtwert aller Güter, die während eines Jahres innerhalb der Landesgrenzen einer Volkswirtschaft hergestellt wurden, in aktuellen Marktpreisen an. Dividiert man das nominale Bruttoinlandsprodukt einer Volkswirtschaft durch die Einwohnerzahl, dann erhält man das sogenannte *BIP pro Kopf*.

Die nachstehende Grafik zeigt die relative Veränderung des BIP pro Kopf in Österreich von 2012 bezogen auf 2002.



Aufgabenstellung:

Geben Sie an, ob ausschließlich anhand der Daten in der gegebenen Grafik der Wert der relativen Änderung des nominalen Bruttoinlandsprodukts in Österreich von 2012 bezogen auf 2002 ermittelt werden kann, und begründen Sie Ihre Entscheidung!

Mat. 14/1. NT/20:

Um Aussagen über die Daten einer statistischen Erhebung treffen zu können, gibt es bestimmte statistische Kennzahlen. Aufgabenstellung: Welche der nebenstehenden statistischen Kennzahlen geben Auskunft darüber, wie stark die erhobenen Daten streuen? Kreuzen Sie die beiden zutreffenden Kennzahlen an!

Median	
Spannweite	
Modus	
empirische Varianz	
arithmetisches Mittel	

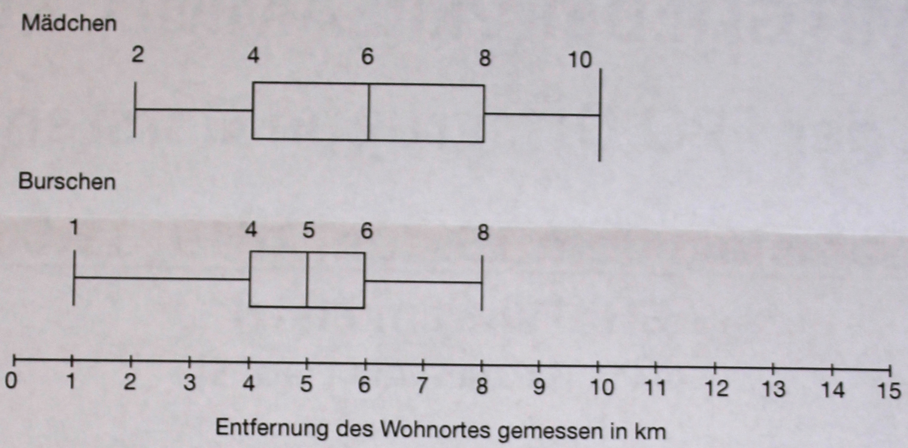
Mat. 14/2. NT/Aufg. 20/1:

Gegeben ist eine geordnete Liste mit neun Werten a_1 bis a_9. Der Wert a_1 wird um 5 vergrößert, der Wert a_9 wird um 5 verkleinert, die restlichen Werte der Liste bleiben unverändert. Durch die Abänderung der beiden Werte kann sich eine neue, nicht geordnete Liste ergeben. Aufgabenstellung: Welche statistischen Kennzahlen der Liste werden durch die genannten Änderungen in keinem Fall verändert? Kreuzen Sie die entsprechende(n) statistische(n) Kennzahl(en) an!

Median	
Spannweite	
Modus	
Standardabweichung	
arithmetisches Mittel	

Anhand der nächsten zwei Aufgaben wird das <u>Kastenschaubild</u> (der <u>Box-Plot</u>) eingeführt und erläutert.
Mat. 14/HT/Aufg. 20/1:

Alle Mädchen und Burschen einer Schulklasse wurden über die Länge ihres Schulweges befragt. Die beiden Kastenschaubilder (Boxplots) geben Auskunft über ihre Antworten.



Aufgabenstellung:

Kreuzen Sie die beiden zutreffenden Aussagen an!

Mehr als 60 % der befragten Mädchen haben einen Schulweg von mindestens 4 km.	☐
Der Median der erhobenen Daten ist bei Burschen und Mädchen gleich.	☐
Mindestens 50 % der Mädchen und mindestens 75 % der Burschen haben einen Schulweg, der kleiner oder gleich 6 km ist.	☐
Höchstens 40 % der befragten Burschen haben einen Schulweg zwischen 4 km und 8 km.	☐
Die Spannweite ist bei den Umfragedaten der Burschen genauso groß wie bei den Umfragedaten der Mädchen.	☐

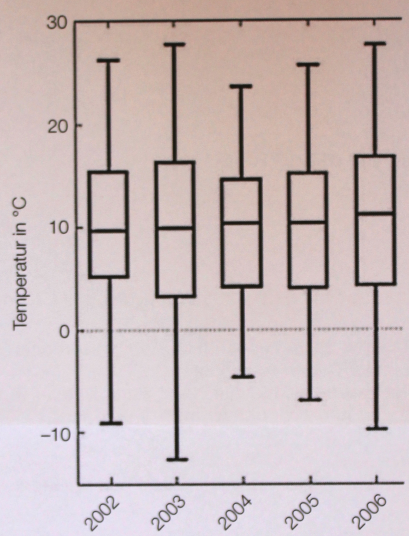
<u>Das Kastenschaubild (Box-Plot):</u>

... ist ein Diagramm, das zur *grafischen Darstellung* der Verteilung eines *der Größe nach geordneten von reellen Zahlen gebildeten Merkmals* verwendet wird. In ihm wird der **Median**, das **untere und das obere Quartil** (in einem „Kasten“) und **die beiden Extremwerte** (als Enden von zwei „Fühlern“) dargestellt. Im „Kasten“ befinden sich daher die mittleren 50 Prozent des Datensatzes.

Daraus ergeben sich bei der Aufg. 20/1 aus HT 2014 die erste und die dritte Aussage als richtig. Bei der Aufg. 19/1 aus dem 2. NT 2014 sind die erste und die vierte Aussage richtig.

Mat. 14/2. NT/Aufg. 19/1:

Die nachstehende Grafik veranschaulicht die jährlichen Temperaturaufzeichnungen der Tagesmitteltemperaturen von Braunschweig (Deutschland) im Zeitraum 2002–2006 mithilfe von Kastenschaubildern (Boxplots).



Aufgabenstellung:

Kreuzen Sie die beiden zutreffenden Aussagen an!

Im Zeitraum 2002–2006 lag der Median der jeweiligen Tagesmitteltemperaturen jeweils im Intervall [7 °C; 13 °C].	☐
Im Jahr 2006 lagen mehr als 25 % der Tagesmitteltemperaturen unter 0 °C.	☐
Das Jahr 2002 wies den größten Median der Tagesmitteltemperaturen auf.	☐
Das Jahr 2003 wies die größte Spannweite der Tagesmitteltemperaturen auf.	☐
Im Jahr 2004 betrug die Spannweite der Tagesmitteltemperaturen 10 °C.	☐

<u>Das Stängel-Blatt-Diagramm</u>

dient dazu, die in einer Urliste meist ungeordnet erfassten Daten für die weitere Arbeit aufzubereiten. Dabei erfolgt ein Aufspalten der einzelnen Daten in einen „Stängel“- und einen „Blatt“-Bestandteil. Als Stängelwerte wählt man etwa die Hunderter oder Zehnerstellen der Daten, als Blattwerte die Einer. In der linken Spalte der Darstellung werden zunächst die gewählten Stängelwerte notiert, und zwar geordnet (vom kleinsten bis zum größten Wert). In die Zeilen der rechten Spalte trägt man dann die Blattwerte ein.

Mat. 17/1. NT/Aufg. 19/1:

Die nachstehenden Stängel-Blatt-Diagramme zeigen die Anzahl der Kinobesucher je Vorstellung der Filme *A* und *B* im Lauf einer Woche. Die Einheit des Stängels ist 10, die des Blattes 1.

Film A		Film *B*	
2	0, 3, 8	2	1
3	6, 7	3	1, 4, 5
4	1, 1, 5, 6	4	4, 5, 8
5	2, 6, 8, 9	5	0, 5, 7, 7
6	1, 8	6	1, 2

		7	0

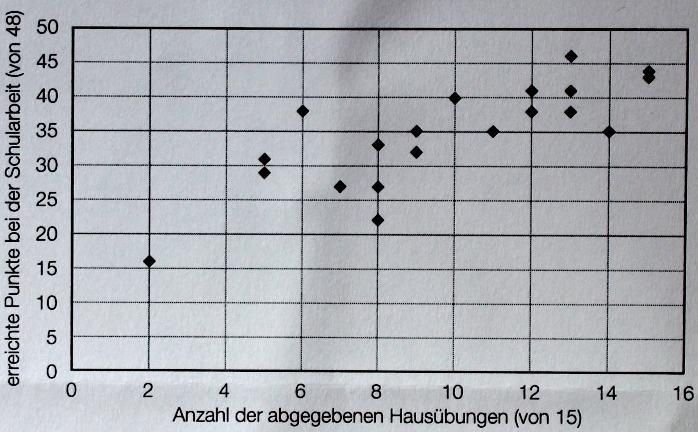
Aufgabenstellung: Kreuzen Sie diejenige(n) Aussage(n) an, die bezogen auf die dargestellten Stängel-Blatt-Diagramme mit Sicherheit zutrifft/zutreffen!

Es gab in dieser Woche mehr Vorstellungen des Films *A* als des Films *B*.	
Der Median der Anzahl der Besucher ist bei Film *A* größer als bei Film *B*.	
Die Spannweite der Anzahl der Besucher ist bei Film *A* kleiner als bei Film *B*.	
Die Gesamtanzahl der Besucher in dieser Woche war bei Film *A* größer als bei Film *B*.	
In einer Vorstellung des Films *B* waren mehr Besucher als in jeder Vorstellung des Films *A*.	

Mat. 18/HT/Aufg. 19/1:

In einer Klasse, in der ausschließlich Mädchen sind, waren bis zu einer Schularbeit 15 Hausübungen abzugeben. Bei der Schularbeit waren maximal 48 Punkte zu erreichen.

Im nachstehenden Punktwolkendiagramm werden für jede der insgesamt 20 Schülerinnen dieser Klasse die Anzahl der abgegebenen Hausübungen und die Anzahl der bei der Schularbeit erreichten Punkte dargestellt.



Aufgabenstellung:

Zwei der nachstehenden fünf Aussagen interpretieren das dargestellte Punktwolkendiagramm korrekt. Kreuzen Sie die beiden zutreffenden Aussagen an!

Nur Schülerinnen, die mehr als 10 Hausübungen abgegeben haben, konnten mehr als 35 Punkte bei der Schularbeit erzielen.	☐
Die Schülerin mit der geringsten Punkteanzahl bei der Schularbeit hat die wenigsten Hausübungen abgegeben.	☐
Die Schülerin mit den meisten Punkten bei der Schularbeit hat alle Hausübungen abgegeben.	☐
Schülerinnen mit mindestens 10 abgegebenen Hausübungen haben bei der Schularbeit im Durchschnitt mehr Punkte erzielt als jene mit weniger als 10 abgegebenen Hausübungen.	☐
Aus der Anzahl der bei der Schularbeit erreichten Punkte kann man eindeutig auf die Anzahl der abgegebenen Hausübungen schließen.	☐

2. Grundbegriffe der Wahrscheinlichkeitsrechnung

Das betreffende Wissensgebiet beschäftigt sich mit der **Wahrscheinlichkeit** $0 \leq p \leq 1$, dass ein bestimmtes Ereignis aus einem ***Grund-, Ergebnis- oder Stichprobenraum*** **G**, das ist die Menge aller möglichen Versuchsausgänge, eintritt. Diese Wahrscheinlichkeit kann auch in Prozent (= 100 p) angegeben werden. p = 0 bedeutet, dass das Ereignis nie, p = 1 bedeutet, dass das Ereignis immer eintritt. Führt man in diesem Zusammenhang eine *statistische Untersuchung* durch, so gibt die *relative Häufigkeit* der verschiedenen konkreten Versuchsausgänge einen Hinweis auf die entsprechende Wahrscheinlichkeit.

Beim Roulette ist das Ergebnis „rot/rouge" und „schwarz/noir" (oder „gerade/pair" und „ungerade/impair" oder „klein/manque" und „groß/passe") gleich wahrscheinlich, daneben gibt es aber noch den Fall, dass die Kugel auf „null/zero" fällt. Die Wahrscheinlichkeit, mit „rot" oder mit „schwarz" zu gewinnen, müsste also etwas geringer als 50 Prozent sein.

Dieser *intuitive Wahrscheinlichkeitsbegriff* wird durch die von Pierre Simon de LAPLACE (1749 - 1827) stammende „klassische" Definition von Wahrscheinlichkeit präzisiert:

> „Die Wahrscheinlichkeit eines Ereignisses ist die Anzahl der dem Ereignis günstigen Fälle, geteilt durch die Anzahl der überhaupt möglichen Fälle."

Beim Münzwurf sind zwei Fälle möglich, einer davon ist „günstig", daher ist die Wahrscheinlichkeit (engl. *probability*) p = ½ = 0,5 oder 50%. Beim Roulette sind 37 Fälle (alle ganzen Zahlen von 0 bis 36) möglich, je 18 sind für „rot" bzw. „schwarz" günstig, also $p = 18/37 \approx 0{,}486$ oder rund 48,6 %. Neben dem Symbol p wird für die Wahrscheinlichkeit auch das Symbol P(E) verwendet, worin E das abgefragte Ereignis darstellt: Z. B. P(rot) = P(schwarz) = 18/37.

Alle Teilmengen eines Grundraumes bilden den zugehörigen Ereignisraum E. Beim Münzwurf ist G = {K, Z} und E besteht aus N = {}, {K}, {Z} und {K, Z}, die entsprechenden Wahrscheinlichkeiten sind P(N) = 0, P(K) = ½, P(Z) = ½ und $P(K \vee Z) = 1$.

Ein Experiment, bei dem jeder der n möglichen Ausfälle gleich wahrscheinlich (p = 1/n) ist, heißt **LAPLACE-Experiment**. Münzwurf und Roulettespiel sind solche Experimente. (Beim Roulette ist $p = 1/37 \approx 0{,}027$ oder rund 2,7%.) Weitere LAPLACE-Experimente sind der Wurf eines Spielwürfels (p = 1/6), das Ziehen einer bestimmten Karte aus einem kompletten Spiel (n = 4x5 = 20, n = 4x8 = 32, n = 4x13 = 52 oder n = 4x8 + 22 = 54 beim Tarock), das Ziehen einer Kugel aus einer Urne mit n Kugeln oder der Pfeilwurf auf ein sich drehendes Glücksrad mit n kongruenten Sektoren. Kein LAPLACE-Experiment ist z.B. der Reißnagelwurf mit den Ausfällen „Spitze oben" oder „Spitze nicht oben". In solchen Fällen kann die Wahrscheinlichkeit nur auf experimentellem Weg mit Hilfe der zugehörigen Häufigkeitsverteilung ermittelt werden.

<u>Die Additionsregel:</u>

Die Wahrscheinlichkeit, dass beim Würfeln mit einem Würfel die Punktezahl 1 **oder** 2 auftritt, beträgt $P(1 \vee 2) = 2/6 = 1/3$. Dasselbe Ergebnis liefert aber auch die Rechung P(1) + P(2) = 1/6 + 1/6 = 1/3. *Wenn also die Wahrscheinlichkeit, dass zwei oder mehrere, voneinander unabhängige Ausfälle auftreten, abgefragt wird, dann können die einzelnen Wahrscheinlichkeiten summiert werden.*

Zu jeder Wahrscheinlichkeit p gibt es eine **Gegenwahrscheinlichkeit** $\neg p$, nämlich dass das betreffende Ereignis nicht eintritt, und für diese gilt nach der Additionsregel $\mathbf{\neg p = 1 - p}$

<u>Die Multiplikationsregel:</u>

Die Wahrscheinlichkeit, dass bei einem Wurf mit zwei Würfeln dieselbe Augenzahl auftritt, also z. B. beim ersten Würfel 6 **und** auch beim zweiten Würfel 6, ergibt sich als Produkt der einzelnen, voneinander unabhängigen Wahrscheinlichkeiten, also P(6 $\wedge$ 6) = (1/6).(1/6) = 1/36. Probe: Der Grundraum für dieses Experiment besteht aus 36 (geordneten) Zahlenpaaren: (1, 1), (1, 2), ..., (1, 6), (2, 1), (2, 2), ..., (2, 6), ..., (6, 1), (6, 2), ..., (6, 6). Davon ist nur jeweils ein Fall günstig. *Für die Wahrscheinlichkeit, dass zwei oder mehrere, voneinander unabhängige Ausfälle gleichzeitig auftreten, sind die einzelnen Wahrscheinlichkeiten miteinander zu multiplizieren.*

Binomialkoeffizienten:

Die Anzahl der verschiedenen Möglichkeiten, aus einer Menge von n Elementen k Elemente auszuwählen, ist durch den Binomialkoeffizienten „n über k“ bestimmt, welcher als Quotient von Faktoriellen (n! = n.(n-1).(n-2)...3.2.1) wie folgt definiert ist:

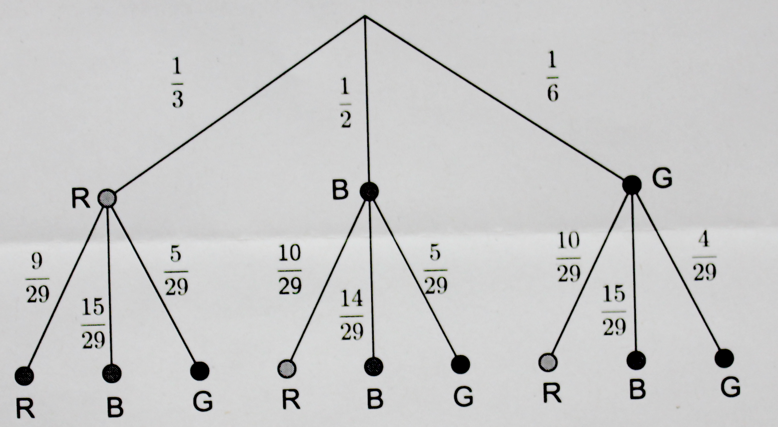
$$\binom{n}{k} = \frac{n!}{k!.(n-k)!} = \frac{n.(n-1)\ldots(n-k+1)}{k!}$$

Der zweite Bruch entsteht aus dem ersten durch Kürzen mit (n-k)!

<u>Musterbeispiel:</u> Aus einer Urne mit zwei schwarzen und drei roten Kugeln wird nacheinander (ohne Zurücklegen) gezogen. Wie groß ist die Wahrscheinlichkeit, dass die ersten zwei gezogenen Kugeln die schwarzen sind? Multiplikationsregel: (2/5).(1/4) = 1/10. Binomialkoffizienten: Der Grundraum besteht aus 5 über 2 = 10 möglichen Kombinationen, aus fünf Kugeln zwei auszuwählen. Davon ist nur eine günstig, nämlich jene mit den zwei schwarzen Kugeln. (Man darf sich die Kugeln auch „nummeriert“ denken.)

Mat. 14/2. NT/Aufg. 22/1: <u>Baumdiagramme</u>

In einem Gefäß befinden sich rote, blaue und grüne Kugeln. Es werden zwei Kugeln gezogen. Das folgende Baumdiagramm veranschaulicht die möglichen Ergebnisse des Zufallsversuchs:



Quelle: http://www.mathe-online.at/mathint/wstat1/grafiken/baumdiagramm2.gif [18.12.2014] (adaptiert)

R = rote Kugel
B = blaue Kugel
G = grüne Kugel

Aufgabenstellung:

Berechnen Sie die Wahrscheinlichkeit, dass zwei Kugeln gleicher Farbe gezogen werden!

Die Brüche geben die Wahrscheinlichkeiten für die Ziehung einer roten, blauen oder grünen Kugel an. Daraus folgt: Die **Ziehung** erfolgt **ohne Zurücklegen** und es gibt 30 Kugeln in der Urne, 10 rote, 15 blaue und 5 grüne. Das Ergebnis lautet nach der Additions- und Multiplikationsregel P(zwei gleiche Farben) = (1/3).(9/29)+(1/2).(14/29)+(1/6).4/29) = 32/87 ≈ 0,37 oder ca. 37 %. Die Summe aller möglichen „Wege“, zwei Kugeln zu ziehen, muss 1 ergeben.

Mat. 16/HT/Aufg. 21/1:

Eine Gruppe von zehn Personen überquert eine Grenze zwischen zwei Staaten. Zwei Personen führen Schmugglerware mit sich. Beim Grenzübertritt werden drei Personen vom Zoll kontrolliert. Aufgabenstellung: Berechnen Sie die Wahrscheinlichkeit, dass unter den drei kontrollierten Personen die beiden Schmuggler sind. *(Mult.satz oder Binom./Grundraum, wie beim Musterbeispiel.)*

Mat. 18/1. NT/Aufg. 21/1:

Eine Rot-Grün-Fehlsichtigkeit ist angeboren und verändert sich zeitlebens nicht. Männer sind davon zu 9 % betroffen, Frauen nur zu 0,8 %. Der Anteil der Männer an der Weltbevölkerung liegt bei 49,5 %. Aufgabenstellung: Geben Sie die Wahrscheinlichkeit an, dass eine nach dem Zufallsprinzip ausgewählte Person eine Rot-Grün-Sehschwäche hat. *(Multiplikations- und Additionsregel.)*

Mat. 18/1. NT/Aufg. 22/1:

Eine Mannschaft besteht aus n Frauen. Aus diesen wählt der Trainer an einem Tag sechs, an einem anderen Tag acht aus, wobei es auf die Reihenfolge der Auswahl nicht ankommt. An beiden Tagen ist die Anzahl der Möglichkeiten, eine solche Auswahl zu treffen, gleich groß. Aufgabenstellung: Geben Sie die Anzahl n der Frauen an, welche dieser Mannschaft angehören. *(Binomialkoeff.)*

Mat. 15/HT/Aufg. 21/1:

In einer Unterrichtsstunde sind 15 Schülerinnen und 10 Schüler anwesend. Die Lehrperson wählt für Überprüfungen nacheinander zufällig drei verschiedene Personen aus dieser Schulklasse aus. Jeder Prüfling wird nur einmal befragt. Aufgabenstellung: Kreuzen Sie die beiden zutreffenden Aussagen an!

Die Wahrscheinlichkeit, dass die Lehrperson drei Schülerinnen auswählt, kann mittels 15/25.14/25.13/25 berechnet werden.	
Die Wahrscheinlichkeit, dass die Lehrperson als erste Person einen Schüler auswählt, ist 10/25	
Die Wahrscheinlichkeit, dass die Lehrperson bei der Wahl von drei Prüflingen als zweite Person eine Schülerin auswählt, ist 4/25	
Die Wahrscheinlichkeit, dass die Lehrperson drei Schüler auswählt, kann mittels 10/25.9/24.8/23 berechnet werden.	
Die Wahrscheinlichkeit, dass sich unter den von der Lehrperson ausgewählten Personen genau zwei Schülerinnen befinden, kann mittels 15/25.14/24.23/23 berechnet werden.	

Mat. 18/2. NT/Aufg. 21/1:

In zwei Schachteln befindet sich Spielgeld. In Schachtel I sind fünf 2-Euro-Jetons und zwei 1-Euro-Jetons. In Schachtel II sind vier 2-Euro-Jetons und fünf 1-Euro-Jetons. Aus jeder der beiden Schachteln wird je ein Jeton entnommen. Dabei hat pro Schachtel jeder Jeton die gleiche Wahrscheinlichkeit, entnommen zu werden. Aufgabenstellung: Berechnen Sie die Wahrscheinlichkeit, dass nach Entnahme der Jetons in beiden Schachteln derselbe Geldbetrag vorhanden ist. *(Mult.)*

Mat. 14/1. NT/Aufg. 21/1:

In einem Adventkalender wurden versehentlich 4 der 24 vorhandenen Fenster nicht befüllt. Aufgabenstellung: Berechnen Sie die Wahrscheinlichkeit, dass Sie beim Öffnen des dritten Fensters das erste leere Fenster vorfinden! *(Mult.)*

Mat. 15/HT/Aufg. 22/1:

Ein Fußballspiel muss durch ein fünfgliedriges Elfmeterschießen entschieden werden. In einer der zwei Mannschaften stehen grundsätzlich alle elf Spieler als Elfmeterschützen zur Verfügung. Aufgabenstellung: Deuten Sie den Ausdruck (11 über 5) im gegebenen Kontext! *(Binom.)*

Mat. 15/1. NT/Aufg. 21/1:

In einem Behälter befinden sich 15 rote Kugeln und 18 blaue Kugeln. Die Kugeln sind bis auf ihre Farbe nicht unterscheidbar. Es sollen nun in einem Zufallsexperiment zwei Kugeln nacheinander gezogen werden, wobei die erste Kugel nach dem Ziehen nicht zurückgelegt wird und es auf die Reihenfolge der Ziehung ankommt. Die Buchstaben r und b haben folgende Bedeutung: r ... das Ziehen einer roten Kugel, b ... das Ziehen einer blauen Kugel. Aufgabenstellung: Ergänzen Sie die Textlücken im folgenden Satz durch Ankreuzen der jeweils richtigen Satzteile so, dass eine korrekte Aussage entsteht! Ein Grundraum G für dieses Zufallsexperiment lautet ....[1], und ....[2] ist ein Ereignis.

1)		2)	
$G = \{r, b\}$		die Wahrscheinl., dass genau eine blaue Kugel gezogen wird	
$G = \{(r, r), (r, b), (b, b)\}$		jede Teilmenge des Grundraumes	
$G = \{(r, r), (r, b), (b, r), (b, b)\}$		B	

Mat. 15/1. NT/Aufg. 22/1:

Zwei unterscheidbare, faire Würfel mit den Augenzahlen 1, 2, 3, 4, 5, 6 werden gleichzeitig geworfen und die Augensumme wird ermittelt. Das Ereignis, dass die Augensumme durch 5 teilbar ist, wird mit E bezeichnet. (Ein Würfel ist „fair“, wenn die Wahrscheinlichkeit, nach einem Wurf nach oben zu zeigen, für alle sechs Seitenflächen gleich groß ist.) Aufgabenstellung: Berechnen Sie die Wahrscheinlichkeit des Ereignisses E!

Mat. 15/2. NT/Aufg. 22/1:

Bei einem Maturaball werden zwei verschiedene Glücksspiele angeboten: ein Glücksrad und eine Tombola, bei der 1 000 Lose verkauft werden. Das Glücksrad ist in zehn gleich große Sektoren unterteilt, die alle mit der gleichen Wahrscheinlichkeit auftreten können. Man gewinnt, wenn der Zeiger nach Stillstand des Rades auf das Feld der „1“ oder der „6“ zeigt. Aufgabenstellung: Max hat das Glücksrad einmal gedreht und als Erster ein Los der Tombola gekauft. In beiden Fällen hat er gewonnen. Die Maturazeitung berichtet darüber: „Die Wahrscheinlichkeit für dieses Ereignis beträgt 3 %.“ Berechnen Sie die Anzahl der Gewinn-Lose!

WS 3.1 bis WS 3.3

3. Diskrete Wahrscheinlichkeitsverteilungen

Anhand von Beispiel 2 soll der *Zusammenhang zwischen einer Häufigkeitsverteilung (auftretende Häufigkeiten bei der Ausführung eines Experiments) und einer Wahrscheinlichkeitsverteilung* verdeutlicht werden: G ist bei diesem Beispiel die Menge aller Würfe mit einem Würfel, das ist eine unendl. Menge, S sei die Menge von n Würfen, also $|S| = n$. X sei der Ausfall des Wurfes (= die gewürfelte Punktezahl), daher besteht die Menge A der Merkmalsausprägungen x_1, x_2, ... x_k aus den natürlichen Zahlen 1 bis 6 und es ist $|A| = 6$.

Beispiel 2 lässt sich mit Hilfe des LAPLACEschen Wahrscheinlichkeitsbegriffs „mathematisieren" und so kommt man von der Häufigkeitsverteilung zur ***Wahrscheinlichkeitsverteilung***. Es ist nämlich zu erwarten, dass für alle sechs Fälle die relative Häufigkeit etwa gleich groß, also rund 0,167 sein wird. Versuche bestätigen das mit umso größerer Genauigkeit, je größer das n (also die Zahl der Würfe) gewählt wird (***„Gesetz der großen Zahlen"*** von Jakob BERNOULLI, 1654 - 1705).

Bei der Wahrscheinlichkeitsverteilung wird die Stichprobe S als ***Ergebnisraum*** (bisher Grundraum G) bezeichnet, $|S| = n$ ist die Anzahl aller (verschiedenen) möglichen Ausfälle des zugrunde liegenden Experiments (anstelle der statistisch erhobenen Ausgangsdaten). Das Merkmal X, das jedem dieser Ausfälle anhaftet und das Gegenstand der Untersuchung ist, wird als ***Zufallsvariable*** bezeichnet. Ist diese eine ganze Zahl, so liegt eine ***diskrete Wahrscheinlichkeitsverteilung*** vor. Die relative Häufigkeit h_i jeder Merkmalsausprägung x_i wird dann zur Wahrscheinlichkeit p_i, die Häufigkeitsverteilung $h = H(X)$ zur Wahrscheinlichkeitsverteilung $p = P(X)$.

Musterbeispiel: Wahrscheinlichkeitsverteilung beim Wurf von zwei Würfeln. S = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}; $n = 6^2 = 36$.

X... gewürfelte Augenzahl (Summe): A = {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

2:	Nur das Ereignis (1,1) ist „günstig"	$\Rightarrow P(X = 2) = 1/36 \approx 0{,}028$, ebenso $P(X = 12)$
3:	(1,2) und (2,1) sind „günstig"	$\Rightarrow P(X = 3) = 2/36 \approx 0{,}056$, ebenso $P(X = 11)$
4:	(1,3), (2,2), (3,1) sind „günstig"	$\Rightarrow P(X = 4) = 3/36 \approx 0{,}083$, ebenso $P(X = 10)$
5:	(1,4), (2,3), (3,2), (4,1) sind „günstig"	$\Rightarrow P(X = 5) = 4/36 \approx 0{,}111$, ebenso $P(X = 9)$
6:	(1,5), (2,4), (3,3), (4,2), (5,1)	$\Rightarrow P(X = 6) = 5/36 \approx 0{,}139$, ebenso $P(X = 8)$
7:	(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)	$\Rightarrow P(X = 7) = 6/36 \approx 0{,}167$

Der Mittelwert einer Wahrscheinlichkeitsverteilung heißt ***Erwartungswert*** **μ**; die Begriffe ***Varianz*** **σ^2** und ***Standardabweichung*** **σ** sind identisch mit jenen der Häufigkeitsverteilung.

$$\mu = x_1.p_1 + x_2.p_2 + ...+x_k.p_k = \Sigma x_i.p_i$$
$$\sigma^2 = \Sigma(x_i - \mu)^2.p_i$$

Jedes **Histogramm** einer diskreten Wahrscheinlichkeitsverteilung mit der „Balkenbreite" 1 und der „Balkenhöhe" P(X) hat einen **Flächeninhalt mit der Maßzahl 1** und die **Flächeninhalte von Teilbereichen** des Histogramms **stimmen mit den zugehörigen Wahrscheinlichkeiten überein.**

Musterbeispiel (oben): $\mu = 2.(1/36) + 3.(2/36) + 4.(3/36) + 5.(4/36) + 6.(5/36) + 7.(6/36) + 8.(5/36) + 9.(4/36) + 10.(3/36) + 11.(2/36) + 12.(1/36) = (1/36).(2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12) = 7$. $\boldsymbol{\sigma^2} = (-5)^2.(1/36) + (-4)^2.(2/36) + (-3)^2.(3/36) + (-2)^2.(4/36) + (-1)^2.(5/36) + 0.(6/36) + 1^2.(5/36) + 2^2.(4/36) + 3^2.(3/36) + 4^2.(2/36) + 5^2.(1/36) = 1/36.(25 + 32 + 27 + 16 + 5 + 5 + 16 + 27 + 32 + 25) = 210/36$, $\sigma \approx 2{,}4$. **Fertige zuletzt ein Histogramm (mit μ und σ) an!**

Die Binomialverteilung:

Unter einem ***n-stufigen BERNOULLI-Experiment*** versteht man eine Folge von n Versuchen, bei der jeder Versuch unter genau den gleichen Voraussetzungen abläuft und genau zwei Ausgänge zulässt. (Achtung: Diese n ist mit $|S|$ nicht identisch, siehe Seite 16.)

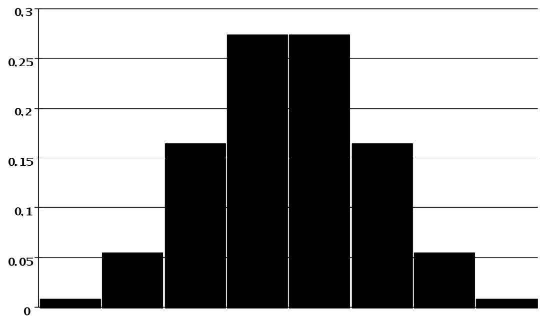
Das n-malige Ziehen aus einer Urne mit Zurücklegen und das n-malige Würfeln sind BERNOULLI-Experimente, sofern die beiden Ausgänge z. B. durch „die gezogene Kugel ist rot" und „die gezogene Kugel ist nicht rot“ bzw. durch „die gewürfelte Augenzahl ist sechs“ und „die gewürfelte Augenzahl ist nicht sechs“ definiert sind. Das einfachste BERNOULLI-Experiment ist der n-malige Münzwurf.

Tritt bei einem BERNOULLI-Experiment das Ereignis E mit der Wahrscheinlichkeit p ein, so tritt das Gegenereignis ¬E mit der Wahrscheinlichkeit q = 1 - p ein (z. B. p = 1/6 und q = 5/6 beim Würfeln, p = q = ½ beim Münzwurf). Ein Baumdiagramm eines n-stufigen BERNOULLI-Experiments enthält $2^n = |S|$ „Wege“, die eine Abfolge von k-maligem E und (n - k)-maligem ¬E aufweisen. *Jede Wegwahrscheinlichkeit ist dann ein Potenzprodukt* $p^k.q^{n-k}$. Die Anzahl der Wege, bei denen E genau k-mal vorkommt, ist nach der beim Musterbeispiel von Seite 12 angestellten Überlegung („Nummerierung“) gleich (n über k). Die Wahrscheinlichkeit, dass bei einem n-stufigen BERNOULLI-Experiment das Ereignis E genau k-mal eintritt, ist somit durch die Formel

$$\boxed{P(X = k) = \binom{n}{k}.p^k.(1-p)^{n-k}}$$

bestimmt. Das ist die Funktionsgleichung der Wahrscheinlichkeitsverteilung (= ***Wahrscheinlichkeitsfunktion***) mit der Zufallsvariablen X = „Anzahl k des Eintretens des Ereignisses E“. Die Wahrscheinlichkeitsverteilung eines n-stufigen BERNOULLI-Experiments wird (wegen des Auftretens des Binomialkoeffizienten in dieser Formel) als ***binomische Verteilung*** oder als ***Binomialverteilung,*** gelegentlich auch als ***BERNOULLI-Verteilung*** bezeichnet.

Musterbeispiel: Erstelle für n = 7 und p =1/2 (z. B. siebenmaliger Münzwurf) die Wertetabelle der entsprechenden Binomialverteilung (auf drei Dezimalen genau) und zeichne ein Histogramm.



Wie groß ist die Wahrscheinlichkeit, dass dieselbe Seite zwischen zwei- und fünfmal kommt?

Wertetabelle:

k	0	1	2	3	4	5	6	7
P(k)	0,008	0,055	0,164	0,273	0,273	0,164	0,055	0,008

$$P(2 \le X \le 5) = P(2) + P(3) + P(4) + P(5) \approx \underline{0{,}874}.$$

Die Fläche des Gesamthistogramms verhält sich zur Fläche jedes Teilbereichs wie 1 : P(a ≤ X ≤ b). *Die Fläche der Teilbereiche ist also ein Maß für die zugehörige Wahrscheinlichkeit.* Darin findet die näherungsweise Berechnung von Wahrscheinlichkeiten P(a ≤ X ≤ b) für binomialverteilte Zufallsvariable mit Hilfe der ***Normalverteilung*** ihre Begründung.

Erwartungswert und Varianz der Binomialverteilung:

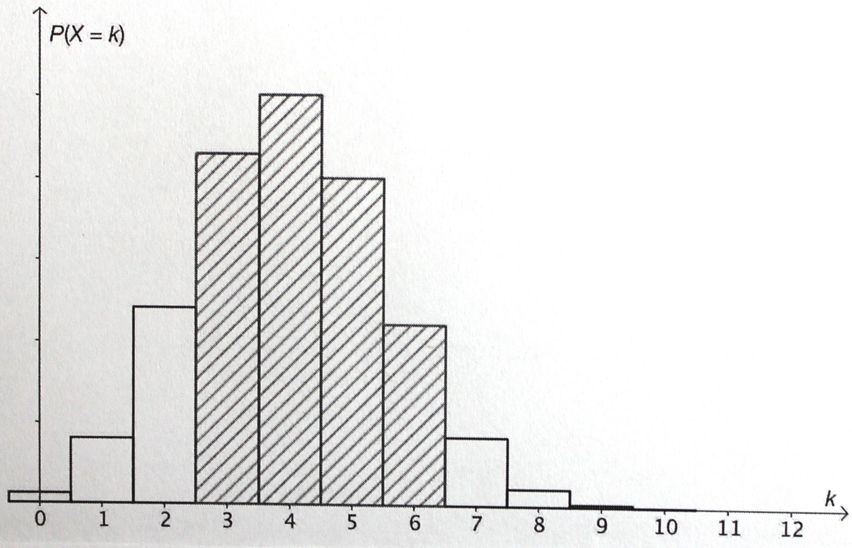
$\mu = \Sigma$ k.(n über k).$p^k.q^{n-k}$ = 0 +1.(n über 1).p.q^{n-1} + 2.(n über 2).$p^2.q^{n-2}$ + ... + n.(n über n).p^n =
n.p.q^{n-1} + 2.n.$\frac{n-1}{2}$.$p^2.q^{n-2}$ + ... + n.p^n = n.p.[q^{n-1} + (n-1).p.q^{n-2} + ... + p^{n-1}] = n.p.$(q+p)^{n-1}$ = n.p
Denn in der eckigen Klammer steht die Entwicklung von $(q+p)^{n-1}$ nach dem binomischen Lehrsatz.
In analoger Weise (allerdings recht aufwändig) bekommt man auch die folgende Formel für σ^2:

$$\boxed{\begin{array}{c}\text{Erwartungswert } \mu = n.p \\ \text{Varianz } \sigma^2 = n.p.(1-p)\end{array}}$$

Beim Musterbeispiel (Seite 16) ist demnach $\mu = 3{,}5$ und $\sigma \approx 1{,}323$. Das Histogramm einer Binomialverteilung ist nur für $p = ½$ exakt symmetrisch mit $x = \mu$ als Symmetrale, wird aber für große n auch dann annähernd symmetrisch, wenn $p \neq ½$ ist.

Mat. 14/HT/Aufg. 23/1:

Die unten stehende Abbildung zeigt die Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen X.



Aufgabenstellung:

Welcher der folgenden Ausdrücke beschreibt die Wahrscheinlichkeit, die dem Inhalt der schraffierten Fläche entspricht?
Kreuzen Sie den zutreffenden Ausdruck an!

$1 - P(X \leq 2)$	☐
$P(X \leq 6) - P(X \leq 3)$	☐
$P(X \geq 3) + P(X \leq 6)$	☐
$P(3 \leq X \leq 6)$	☐
$P(X \leq 6) - P(X < 2)$	☐
$P(3 < X < 6)$	☐

Mat. 14/HT/Aufg. 24/1:

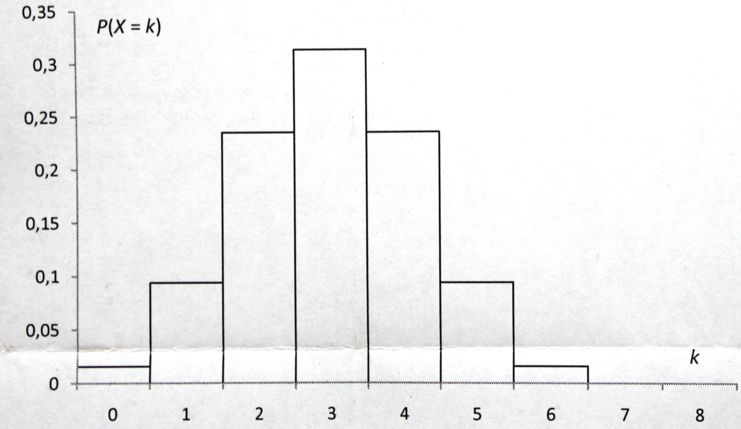
Bei einer schriftlichen Prüfung werden der Kandidatin/dem Kandidaten fünf Fragen mit je vier Antwortmöglichkeiten vorgelegt. Genau eine der Antworten ist jeweils richtig. Aufgabenstellung: Berechnen Sie die Wahrscheinlichkeit, dass die Kandidatin/der Kandidat bei zufälligem Ankreuzen mindestens viermal die richtige Antwort kennzeichnet! *(Binom.)*

Mat. 14/1. NT/Aufg. 23/1:

In der untenstehenden Abbildung ist die Wahrscheinlichkeitsverteilung einer binomialverteilten Zufallsvariablen X mit den Parametern $n = 6$ und $p = 0,5$ durch ein Säulendiagramm (Säulenbreite = 1) dargestellt. μ bezeichnet den Erwartungswert von X.

Aufgabenstellung:

Schraffieren Sie diejenigen Rechtecksflächen, die $P(X > \mu)$ veranschaulichen!



Mat. 14/1. NT/Aufg. 24/1:

In einer Urne befinden sich sieben weiße und drei rote Kugeln, die gleich groß und durch Tasten nicht unterscheidbar sind. Jemand nimmt, ohne hinzusehen, Kugeln aus der Urne. Aufgabenstellung: In welchen der folgenden Fälle ist die Zufallsvariable X binomialverteilt? Kreuzen Sie die beiden zutreffenden Aussagen an!

X beschreibt die Anzahl der roten Kugeln bei dreimaligem Ziehen, wenn jede entnommene Kugel wieder zurückgelegt wird.	
X beschreibt die Anzahl der weißen Kugeln bei viermaligem Ziehen, wenn die entnommenen Kugeln nicht zurückgelegt werden.	
X beschreibt die Anzahl der weißen Kugeln bei fünfmaligem Ziehen, wenn jede entnommene Kugel wieder zurückgelegt wird.	
X beschreibt die Anzahl der Züge, bis die erste rote Kugel gezogen wird, wenn jede entnommene Kugel wieder zurückgelegt wird.	
X beschreibt die Anzahl der Züge, bis alle weißen Kugeln gezogen wurden, wenn die entnommenen Kugeln nicht zurückgelegt werden.	

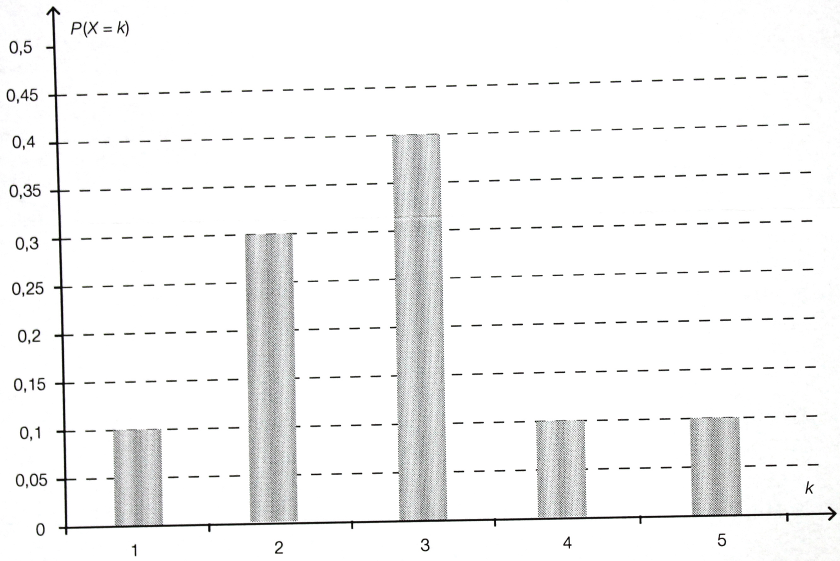
Mat. 14/2. NT/Aufg. 24/1:

Ein fairer Würfel wird zehnmal geworfen. Aufgabenstellung: Kreuzen Sie die zutreffende(n) Antwort(en) an! Welche Wahrscheinlichkeit wird durch den Term $1 - \left[\binom{10}{9}\cdot\left(\frac{1}{6}\right)^9\cdot\left(\frac{5}{6}\right)+\left(\frac{1}{6}\right)^{10}\right]$ angegeben?

Der Term gibt die Wahrscheinlichkeit an, höchstens acht Sechser zu werfen.	
Der Term gibt die Wahrscheinlichkeit an, mehr als zweimal keinen Sechser zu werfen.	
Der Term gibt die Wahrscheinlichkeit an, mindestens einmal keinen Sechser zu werfen.	
Der Term gibt die Wahrscheinlichkeit an, weniger als neun Sechser zu werfen.	
Der Term gibt die Wahrscheinlichkeit an, mehr als acht Sechser zu werfen.	

Mat. 14/2. NT/Aufg. 23/1:

Die nachstehende Abbildung zeigt die Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen X, bei der jedem Wert k (k = 1, 2, 3, 4, 5) die Wahrscheinlichkeit $P(X = k)$ zugeordnet wird.



Aufgabenstellung:

Ermitteln Sie den Erwartungswert $E(X)$ der Zufallsvariablen X!

Mat. 15/HT/Aufg. 23/1:

Bei einem Gewinnspiel gibt es 100 Lose. Der Lospreis beträgt € 5. Für den Haupttreffer werden € 100 ausgezahlt, für zwei weitere Treffer werden je € 50 ausgezahlt und für fünf weitere Treffer werden je € 20 ausgezahlt. Für alle weiteren Lose wird nichts ausgezahlt. Unter *Gewinn* versteht man *Auszahlung minus Lospreis*. Aufgabenstellung: Berechnen Sie den Erwartungswert des Gewinns aus der Sicht einer Person, die ein Los kauft!

Mat. 15/HT/Aufg. 24/1:

Stefan und Helmut spielen 5 Sätze Tennis. Stefan hat eine konstante Gewinnwahrscheinlichkeit von 60 % für jeden gespielten Satz. Aufgabenstellung: Es wird folgender Wert berechnet: $\binom{5}{3} \cdot 0{,}4^3 \cdot 0{,}6^2 = 0{,}2304$. Geben Sie an, was dieser Wert im Zusammenhang mit der Angabe aussagt! *(Binomialvert.)*

Mat. 15/1. NT/Aufg. 23/1:

Ein Glücksrad ist in acht gleich große Sektoren unterteilt, die mit gleicher Wahrscheinlichkeit auftreten. Für einmaliges Drehen des Glücksrades muss ein Einsatz von 5 € gezahlt werden. Vier Sektoren sind „Nieten“, bei zwei Sektoren werden fünf Euro, bei einem zehn Euro und bei einem 15 Euro ausbezahlt. Aufgabenstellung: Das Glücksrad wird einmal gedreht. Berechnen Sie den entsprechenden Erwartungswert des Reingewinns G (in Euro) aus der Sicht des Betreibers des Glücksrades! Der Reingewinn ist die Differenz aus Einsatz und Auszahlungsbetrag.

Mat. 15/1. NT/Aufg. 24/1:

Ein italienischer Süßwarenhersteller stellt Überraschungseier her. Das Ei besteht aus Schokolade. Im Inneren des Eies befindet sich in einer gelben Kapsel ein Spielzeug oder eine Sammelfigur. Der Hersteller wirbt für die Star-Wars-Sammelfiguren mit dem Slogan „Wir sind jetzt mit dabei, in jedem 7. Ei!". Aufgabenstellung: Peter kauft in einem Geschäft zehn Überraschungseier aus dieser Serie. Berechnen Sie die Wahrscheinlichkeit, dass Peter mindestens eine Star-Wars-Sammelfigur erhält!

Mat. 16/1. NT/Aufg. 23/1:

Ein Zufallsexperiment wird durch eine binomialverteilte Zufallsvariable X beschrieben. Diese hat die Erfolgswahrscheinlichkeit p = 0,36 und die Standardabweichung σ = 7,2. Aufgabenstellung: Berechnen Sie den zugehörigen Parameter n (= die Anzahl der Versuche)!

Mat. 17/HT/Aufg. 22/1:

Die Zufallsvariable X kann nur die Werte 10, 20 und 30 annehmen. Die nachstehende linke Tabelle gibt die Wahrscheinlichkeitsverteilung von X an, wobei a und b positive reelle Zahlen sind. Aufgabenstellung: Kreuzen Sie (rechte Tabelle) die beiden zutreffenden Aussagen an!

K	10	20	30	Der Erwartungswert von X ist 20.	
P(X=k)	a	b	a	Die Standardabweichung von X ist 20.	
				$a + b = 1$	
				$P(10 \leq X \leq 30) = 1$	
				$P(X \leq 10) = P(X \geq 10)$	

Mat. 17/1. NT/Aufg. 23/1:

Die Wahrscheinlichkeit, dass ein neuer Autoreifen einer bestimmten Marke innerhalb der ersten 10 000 Kilometer Fahrt durch einen Materialfehler defekt wird, liegt bei p %. Eine Zufallsstichprobe von 80 neuen Reifen dieser Marke wird getestet. Aufgabenstellung: Geben Sie einen Ausdruck an, mit dem man die Wahrscheinlichkeit, dass mindestens einer dieser Reifen innerhalb der ersten 10 000 Kilometer Fahrt durch einen Materialfehler defekt wird, berechnen kann!

Mat. 16/HT/aufg. 23/1:

Auf einem Tisch steht eine Schachtel mit zwei roten und acht schwarzen Kugeln. Nach dem Zufallsprinzip werden nacheinander drei Kugeln aus der Schachtel gezogen, wobei die gezogene Kugel jeweils wieder zurückgelegt wird. Aufgabenstellung: Kreuzen Sie dasjenige Ereignis an, dessen Wahrscheinlichkeit durch diesen Ausdruck $3 \cdot 0{,}8^2 \cdot 0{,}2$ berechnet wird! *(Binomialvert.)*

Es wird höchstens eine schwarze Kugel gezogen.	
Es werden genau zwei schwarze Kugeln gezogen.	
Es werden zwei rote Kugeln und eine schwarze Kugel gezogen.	
Es werden nur rote Kugeln gezogen.	
Es wird mindestens eine rote Kugel gezogen.	
Es wird keine rote Kugel gezogen.	

4. Normalapproximation der Binomialverteilung

In Abschnitt 4 werden zu allen Maturaaufgaben die Lösungswege und die Lösungen angegeben.

Für eine binomialverteilte Zufallsvariable X ist die Berechnung der Wahrscheinlichkeiten für großes n recht mühsam. Das hat die Mathematiker dazu veranlasst, nach Funktionen zu suchen, welche $P(X = k)$ für große n gut annähern und somit als (leichter handhabbare) *Ersatzfunktionen f(k)* für die Binomialverteilung dienen können.

Faustregel: Die Normalapproximation der Binomialverteilung mit den Parametern n und p durch die **GAUSS-Funktion** f(k) liefert gute Näherungswerte, wenn folgende Bedingung erfüllt ist: $\mathbf{n \cdot p \cdot (1 - p) \geq 9}$ (d. h. die Varianz beträgt mindestens 9).

$$\mathbf{P(X = k) = (n\ \text{über}\ k).p^k.(1 - p)^{n-k} \approx f(k) = \frac{1}{\sigma.\sqrt{2\pi}}.e^{\frac{-z^2}{2}}\quad \text{mit}\ z = (k - \mu)/\sigma}$$

Ohne Beweis. In einen Koordinatensystem U(z, P) liegt der Mittelwert $k = \mu$ an der Stelle $z = 0$ und somit $P(X = \mu) \approx (\sigma.\sqrt{2\pi})^{-1}$.

Musterbeispiel 1: **a)** Wie groß ist die Wahrscheinlichkeit, mit einem Würfel bei 720 Würfen genau 130mal dieselbe Augenzahl zu werfen? **b)** Wie groß ist die Wahrscheinlichkeit, genau den Erwartungswert zu erreichen?

a) $n = 720$, $p = 1/6$: $\mu = 120$, $\sigma^2 = 100 \Rightarrow \sigma = 10$, $k = 130 \Rightarrow z = (130 - 120)/10 = 1 \Rightarrow f(130) = 0{,}1.(2\pi)^{-1/2}.e^{-1/2} \Rightarrow P(X = 130) \approx 0{,}024$ oder rund 2,4% **b)** Für $k = 120$ ist $z = 0$ und $P(X = 120) \approx 0{,}1.(2\pi)^{-1/2}.e^0 \approx 0{,}03989$ oder rund 4,0%

Stetige Wahrscheinlichkeitsverteilungen:

Viele der für die Statistik relevanten Zufallsexperimente lassen sich durch *diskrete Zufallsvariable* (ganzzahlige X) nur unzureichend beschreiben, z. B. wenn es sich um die Merkmale Größe oder Gewicht handelt. Denn tatsächlich kann hier die Zufallsvariable (innerhalb gewisser Grenzen) jeden reellen Wert annehmen, und man wird sich auch mehr für die Wahrscheinlichkeit innerhalb eines Intervalls (z. B. der Körpergröße zwischen 175 und 180 cm) interessieren als für die Wahrscheinlichkeit, mit der ein ganz bestimmtes Merkmal auftritt. Mehr noch: Da auch in einer noch so kleinen Umgebung eines bestimmten Wertes $a \in \mathbf{R}$ unendlich viele andere Werte liegen, muss für jeden bestimmten Wert a der Zufallsvariablen X die Wahrscheinlichkeit $P(X = a) = 0$ sein.

In diesen Fällen spricht man von einer ***stetigen Zufallsvariablen*** und von einer ***stetigen Verteilung***. Wegen $P(X = a) = 0$ für alle a der Definitionsmenge lässt sich die Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen X nicht durch eine Wahrscheinlichkeitsfunktion (Seite 16) kennzeichnen. An deren Stelle tritt vielmehr eine ***Dichtefunktion*** f(x).

Bei beliebig vorgegebenen Werten a < b wird die Wahrscheinlichkeit P(a < X < b) durch die Maßzahl der Fläche ausgedrückt, die von der Funktionskurve der Dichtefunktion f(x) und der x-Achse zwischen den Ordinaten x = a und x = b eingeschlossen wird.

Jede Dichtefunktion f(x) muss daher der Bedingung genügen, dass der Gesamtflächeninhalt unter der Kurve den Wert 1 annimmt. In der Sprache der Integralrechnung bedeutet das: *Zur Dichtefunktion f(x) gibt es eine Stammfunktion F(x), die für alle reellen Zahlen definiert ist und die im Intervall -∞ < x < +∞ von 0 bis 1 monoton wächst.* Jede solche Dichtefunktion definiert dann eine stetige Wahrscheinlichkeitsverteilung wie folgt:

$$\mathbf{P(a \le X \le b)} = \int_a^b f(x).dx = F(b) - F(a)$$

Die Stammfunktion F(x) wird als ***Verteilungsfunktion*** bezeichnet. Wegen der Bedingung P(X = a) = P(X = b) = 0 ist es belanglos, ob die „Wahrscheinlichkeitsintervalle" als offen (a < x < b) oder als geschlossen (a ≤ x ≤ b) angesehen werden. In der praktischen Anwendung muss die berechnete Wahrscheinlichkeitsverteilung auch einer *empirischen Untersuchung* standhalten, also mit der Häufigkeitsverteilung in einer repräsentativen und ausreichend großen Stichprobe (annähernd) übereinstimmen.

Die Normalverteilung:

Achtung: Bei den Maturaaufgaben kommen konkrete Beispiele zur Normalverteilung (wie die Musterbeispiele 2 und 3) gar nicht vor, doch werden sie hier „der Vollständigkeit halber" behandelt. Für die Maturaaufgaben genügt es, die Wahrscheinlichkeitsverteilungen binomialverteilter Zufallsvariablen mit Hilfe der Stanardnormalverteilung (Inhalte unter der „Glockenkurve") berechnen zu können.

Von den in der Praxis vorkommenden stetigen Verteilungen ist jene die Wichtigste, der als Dichtefunktion die **GAUSS-Funktion** f(x) zugrunde liegt. Eine solche Wahrscheinlichkeitsverteilung wird als ***Normalverteilung*** bezeichnet. Die Parameter μ und σ werden i. A. auf empirischem Weg (d. h. als Mittelwert und Standardabweichung einer repräsentativen Stichprobe) ermittelt. Die Normalverteilung mit Erwartungswert 0 und Standardabweichung 1 heißt ***Standardnormalverteilung***.

Allen Funktionskurven der GAUSS-Funktion ist die „Glockenform" gemeinsam, mit dem Hochpunkt an der Stelle μ und zwei Wendepunkten an den Stellen μ - σ und μ + σ. Alle Kurven haben die x-Achse zur Asymptote. Die ***Glockenkurve*** ist umso höher und schmäler, je kleiner die Standardabweichung σ ist. Für großes σ entsteht eine eher flache Kurve. Im Sonderfall μ = 0 und σ = 1 (= Standardnormalverteilung, wegen $z = \frac{x-\mu}{\sigma}$ oder σ.z = x – μ ist dann x = z) bezeichnet man die GAUSS-Funktion mit φ(x) = φ(z) und die zugehörige Stammfunktion (= Verteilungsfunktion) mit Φ(x) = Φ(z). Verschiedentlich werden die Begriffe „GAUSS-Funktion" und „Glockenkurve" auf diesen Sonderfall eingeschränkt.

Aus mathematischer Sicht hat die GAUSS-Funktion den Nachteil, dass es keine Termdarstellung ihrer Stammfunktionen gibt. Der zwischen einer Glockenkurve und der x-Achse (bzw. z-Achse) befindliche Flächeninhalt kann daher nur durch numerische Integration ermittelt werden Das Ergebnis ist für jede Form der Glockenkurve gleich, nämlich A = 1 FE (ohne Beweis).

Die Werte der Standardnormalverteilung, d. h. die Werte von Φ(z) sind für z-Werte zwischen 0 und 3 in **Zahlentafeln** niedergelegt. Wie eine Zeichnung von φ(z) zeigt, sind daraus die Werte **Φ(-z)** für negative Stellen mit Hilfe der Tafelwerte **Φ(z)** für positive Stellen als

$$\mathbf{\Phi(-z) = 1 - \Phi(z)}$$

und die Werte eines symmetrischen Bereichs **Φ(z) - Φ(-z)** um den Erwartungswert μ = 0 als

$$\mathbf{\Phi(z) - \Phi(-z) = 2\Phi(z) - 1 = D(z)}$$

zu ermitteln. Die Werte dieser Zahlentafeln sind insofern ausreichend, als jede Zufallsvariable X einer Normalverteilung mit den Parametern μ und σ mittels der Formel

$$\mathbf{Z = (X - \mu) : \sigma}$$

in eine standardnormalverteilte Zufallsvariable Z übergeführt werden kann. Diesen Vorgang bezeichnet man als ***Standardisierung*** von X. Zwischen der Verteilungsfunktion F(x) einer normalverteilten Zufallsvariablen X mit den Parametern μ und σ und der Verteilungsfunktion $\Phi(z)$ der standardisierten Zufallsvariablen Z besteht somit der Zusammenhang

$$\boxed{\mathbf{F(x) = \Phi[z = (x - \mu) : \sigma]}}$$

Damit ist es möglich, die Werte der Verteilungsfunktion einer beliebigen normalverteilten Zufallsvariablen mit Hilfe der Zahlentafel für $\Phi(z)$ zu bestimmen.

Musterbeispiel 2: *Die Größe der Frauen eines bestimmten Volksstammes sei eine normalverteilte Zufallsvariable mit den (empirisch ermittelten) Parametern $\mu = 168$ cm und $\sigma = 6$ cm. Mit welchem Anteil von Frauen über 180 cm Körpergröße ist zu rechnen?*

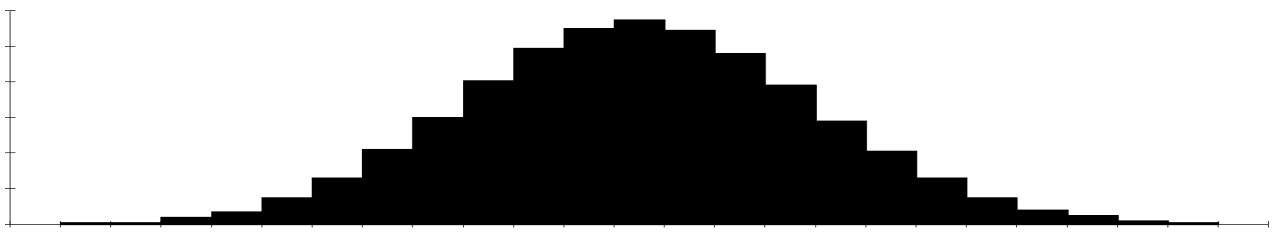
Die gesuchte Wahrscheinlichkeit $P(X > 180)$ entspricht der Fläche unter der Glockenkurve der Standardverteilung rechts von $z = (180 - 168)/6 = 2$, also $P(X > 180) = 1 - \Phi(2) = 1 - 0{,}97725 \approx 0{,}02275$ oder rund 2,3%.

Musterbeispiel 3 (Umkehraufgabe): *Die durchschnittliche Lebenserwartung eines österreichischen Mannes sei normalverteilt mit einem Mittelwert von 74 Jahre und mit einer Standardabweichung von 6 Jahren (Daten aus 1996). Man bestimme ein symmetrisches Intervall um den Mittelwert, sodass 80% aller Beobachtungswerte einer Stichprobe in diesem Intervall erwartet werden können.*

$\mu = 74$, $\sigma = 6$: Gesucht ist eine Zahl $c > 0$, sodass $P(74 - c < X < 74 + c) = 0{,}8$ wird. Standardisierung: $z_1 = (74 - c - 74)/6 = -c/6$ und $z_2 = (74 + c - 74)/6 = c/6$. Links von z_1 und rechts von z_2 darf die Wahrscheinlichkeit nur je 0,1 betragen, also $\Phi(z_1) = 0{,}1$ und $\Phi(z_2) = 0{,}9$. Nach Tabelle gehört zu $\Phi(z) \approx 0{,}9$ der Wert $z_2 \approx 1{,}28$. Also ist $c = 6.z_2 \approx 7{,}7$ Jahre $\Rightarrow \underline{66{,}3 < X < 81{,}7.}$ Oder $2\Phi(z) - 1 = D(z) = 0{,}8 \Rightarrow z \approx 1{,}28 \Rightarrow X = 6z + 74 \approx \underline{81{,}7}$. Untere Grenze $74 - 6z \approx \underline{66{,}3}$.

Binomialverteilung und Normalverteilung:

Der obere Rand der Histogramme von Binomialverteilungen nimmt für wachsende n immer mehr die Form einer Glockenkurve an. Der Umstand, dass der Flächeninhalt der Histogrammstreifen ein Maß für die zugehörigen Wahrscheinlichkeiten ist, legt die Vermutung nahe, dass die **Normalverteilung die aus der Binomialverteilung durch den Grenzübergang n → ∞ hervorgehende Wahrscheinlichkeitsverteilung** ist. Das auf Seite 21 genannte Näherungsverfahren für die Wahrscheinlichkeitsermittlung einzelner Werte einer binomialverteilten Zufallsvariablen erhärtet dies.



In der Praxis wird dieser Umstand dazu genützt, auch **bei einer binomialverteilten Zufallsvariablen X Wahrscheinlichkeiten der Art P(a < X < b) nach der Methode der Normalverteilung** zu **berechnen.** Die Parameter μ und σ ergeben sich in diesem Fall aus den bekannten Formeln $\mu = n.p$ und $\sigma^2 = n.p.(1-p)$. **Als Faustregel gilt, dass diese Methode für $\sigma > 3$ hinlänglich genau ist.**

Musterbeispiel 4: Bei der Generalversammlung eines Vereins, an der 240 stimmberechtigte Mitglieder teilnehmen, wird ein Antrag gestellt, den 30 Mitglieder befürworten, während die anderen „zufällig" (d.h. mit p = ½ dafür und q = ½ dagegen) abstimmen. Stimmenthaltung ist nicht gestattet.

Wie groß ist die Wahrscheinlichkeit, dass der Antrag mit einfacher Mehrheit, also mit mindestens 121 Stimmen, angenommen wird?

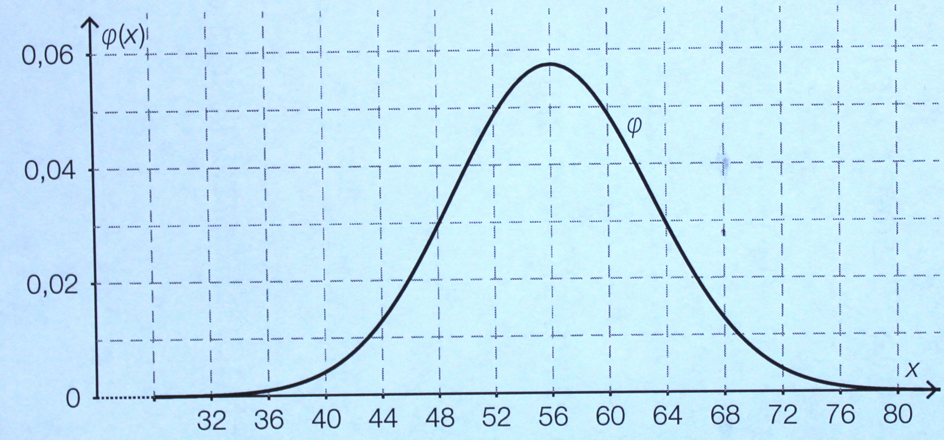
Die Zufallsvariable X ist die Anzahl der Befürworter des Antrags über die 30 sicheren Befürworter hinaus. n = 210 Mitgliedern stimmen „zufällig“ mit p = 1/2 ab, davon müssen mindestens 121 - 30 = 91 positiv abstimmen. Gesucht ist also $P(91 \leq X)$ für $\mu = 210.0{,}5 = 105$ und $\sigma^2 = 105.0{,}5 = 52{,}5$. Standardisierung: $z = (91 - 105) : \sqrt{52{,}5} \approx -1{,}93. \Rightarrow P(91 \leq X) \approx P(-1{,}93 < Z) = \Phi(1{,}93) = 0{,}9732 \approx$ <u>97 %</u>

Ein bestimmter Prozentsatz der Stöcke einer Rosensorte bringt gelbe Blüten hervor. In einem Beet wird eine gewisse Anzahl an Rosenstöcken dieser Sorte gepflanzt. Die Zufallsvariable *X* ist binomialverteilt und gibt die Anzahl der gelbblühenden Rosenstöcke an. Dabei beträgt der Erwartungswert für die Anzahl *X* der gelbblühenden Rosenstöcke 32, und die Standardabweichung hat den Wert 4. Es wird folgender Vergleich angestellt: „Die Wahrscheinlichkeit, dass sich in diesem Beet mindestens 28 und höchstens 36 gelbblühende Rosenstöcke befinden, ist größer als die Wahrscheinlichkeit, dass mehr als 32 gelbblühende Rosenstöcke vorhanden sind.“ Aufgabenstellung: Geben Sie an, ob dieser Vergleich zutrifft, und begründen Sie Ihre Entscheidung!

Lösung: $\mu = n.p = 32$, $\sigma = \sqrt{n.p.(1 - p)} = 4 \Rightarrow n = 64$, $p = 0{,}5$. Daraus folgt wegen $z_1 = (a - \mu):\sigma = -1$ und $z_2 = (a + \mu):\sigma = 1$ für $P(28 \leq X \leq 36) \approx \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = D(1) = 0{,}6827$, das sind über 68%, während die Wahrscheinlichkeit, dass die Anzahl gelb blühenden Rosen über dem Mittelwert liegt, nur 50 % ausmacht. Das Lösungsheft nennt auch noch eine „Wahrscheinlichkeit für Sigma-Umgebungen“ und gibt dafür ungefähr 0,683 an.

Mat. 16/2. NT/Aufg. 24/1:

In Europa beträgt die Wahrscheinlichkeit, mit Blutgruppe B geboren zu werden, ca. 0,14. Für eine Untersuchung wurden *n* in Europa geborene Personen zufällig ausgewählt. Die Zufallsvariable *X* beschreibt die Anzahl der Personen mit Blutgruppe B. Die Verteilung von *X* kann durch eine Normalverteilung approximiert werden, deren Dichtefunktion in der nachstehenden Abbildung dargestellt ist.



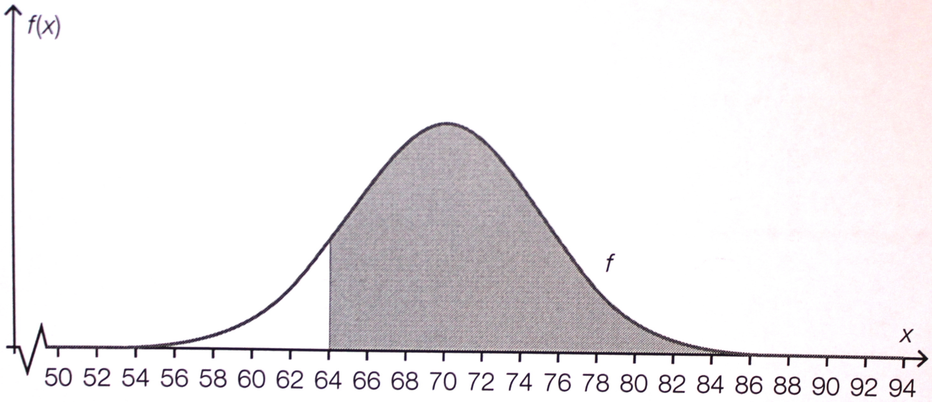
Aufgabenstellung:

Schätzen Sie anhand der obigen Abbildung den Stichprobenumfang *n* dieser Untersuchung!

Lösung: Der Mittelwert ist offenbar $56 = n.p = n.0{,}14 \Rightarrow n \approx 400$

Mat. 17/HT/Aufg. 23/1:

In nachstehender Abbildung ist die Dichtefunktion f der approximierenden Normalverteilung einer binomialverteilten Zufallsvariablen X dargestellt.



Aufgabenstellung:

Deuten Sie den Flächeninhalt der grau markierten Fläche im Hinblick auf die Berechnung einer Wahrscheinlichkeit!

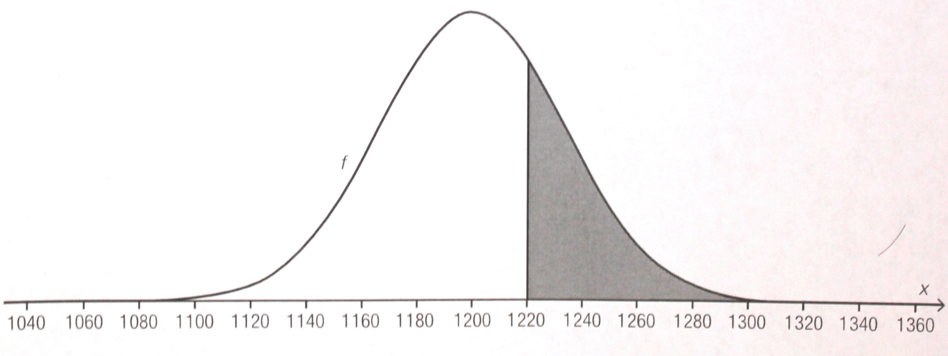
Lösung: Der Flächeninhalt A (in Prozent der Gesamtfläche unter der Kurve) gibt die Wahrscheinlichkeit $P(X \geq 64)$ an, dass also die Zufallsvariable (= Anzahl der Versuche mit dem gewünschten Ausgang) mindestens 64 beträgt.

Mat. 18/2. NT/Aufg. 23/1:

Die Rücklaufquote von Pfandflaschen einer bestimmten Sorte Mineralwasser beträgt 92 %.

In einem Monat werden 15 000 Pfandflaschen dieser Sorte Mineralwasser verkauft. Die Zufallsvariable X gibt die Anzahl derjenigen Pfandflaschen an, die nicht mehr zurückgegeben werden. Die Zufallsvariable X kann durch eine Normalverteilung approximiert werden.

Die nachstehende Abbildung stellt den Graphen der Dichtefunktion f dieser Normalverteilung dar. Der Flächeninhalt der markierten Fläche beträgt ca. 0,27.



Aufgabenstellung:

Deuten Sie den Wert 0,27 im gegebenen Kontext!

Lösung: Die Wahrscheinlichkeit, dass 1220 oder mehr Pfandflaschen nicht zurückgegeben werden, beträgt 27 %.

<u>Vertrauensintervalle (Konfidenzintervalle):</u>

Sei S eine geeignete Stichprobe für eine Grundgesamtheit G, so kann man daraus **Schätzungen** für die Parameter von G vornehmen. Denn diese stimmen sicher nicht exakt mit den entsprechenden Parametern von S überein.

Ein ***Konfidenzintervall*** oder auch ***Vertrauensintervall*** ist ein aus einer Stichprobe S bestimmter Intervallbereich. In diesem Bereich liegen die Parameter der Grundgesamtheit G dann mit einer bestimmten Wahrscheinlichkeit. Die Grenzen des Intervalls sind Zufallsvariable. Bei den *Matura-aufgaben* werden im Wesentlichen nur die Grenzen von Konfidenzintervallen für den wirklichen relativen Anteil p in einer Grundgesamtheit G aufgrund eines bekannten relativen Anteils $\hat{p}$ einer Stichprobe S von der Mächtigkeit n nach folgender **„Anteilsformel“** *(lt. Formelheft)* berechnet:

$$\boxed{\hat{p} - z\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\sqrt{\frac{\hat{p}.(1-\hat{p})}{n}}}$$

In ihr ist z ≈ 1,96 für ein Konfidenzintervall auf dem ***Konfidenzniveau*** $\gamma = 0{,}95$ (= 95 %) und z ≈ 2,575 für ein Konfidenzintervall mit der Genauigkeit $\gamma = 0{,}99$ (= 99 %). Die Werte ergeben sich aus der Formel $\Phi(z) = \frac{1+\gamma}{2}$ und den Tafeln. **Aus der Formel folgt unmittelbar: Je kleiner $|S| = n$ und je größer die Genauigkeit (γ), umso breiter ist das Konfidenzintervall und umgekehrt.**

Mat. 18/2. NT/24:

Bei einer repräsentativen Telefonumfrage mit 400 zufällig ausgewählten Personen erhält man für den relativen Anteil der Befürworter/innen von kürzeren Sommerferien den Wert 20 %. Aufgaben-stellung: Zeigen Sie durch eine Rechnung, dass das Intervall [16,0 %; 24,0 %] ein symmetrisches 95-%-Konfidenzintervall für den relativen Anteil *p* der Befürworter/innen in der gesamten Be-völkerung sein kann (wobei die Intervallgrenzen des Konfidenzintervalls gerundete Werte sind).

Lösung: $\hat{p}$ = 0,2 ist der aus der Stichprobe mit n = 400 ermittelte relative Anteil und z = 1,96. Nach der „Anteilsformel ergibt die Rechnung ungefähr $0{,}16 \leq p \leq 0{,}24$.

Mat. 15/2. NT/Aufg. 24/1:

Bei einer Meinungsbefragung wurden 500 zufällig ausgewählte Bewohner einer Stadt zu ihrer Mei-nung bezüglich der Einrichtung einer Fußgängerzone im Stadtzentrum befragt. Es sprachen sich 60 % der Befragten für die Einrichtung einer solchen Fußgängerzone aus, 40 % sprachen sich dagegen aus. Als 95-%-Konfidenzintervall für den Anteil der Bewohner dieser Stadt, die die Einrichtung einer Fußgängerzone im Stadtzentrum befürworten, erhält man mit Normalapproximation das Inter-vall [55,7 %; 64,3 %]. Aufgabenstellung: Kreuzen Sie die beiden zutreffenden Aussagen an!

Das Konfidenzintervall wäre breiter, wenn man einen größeren Stichprobenumfang gewählt hätte und der relative Anteil der Befürworter/innen gleich groß geblieben wäre.	
Das Konfidenzintervall wäre breiter, wenn man ein höheres Konfidenzniveau (eine höhere Sicherheit) gewählt hätte.	X
Das Konfidenzintervall wäre breiter, wenn man die Befragung in einer größeren Stadt durch-geführt hätte.	
Das Konfidenzintervall wäre breiter, wenn der Anteil der Befürworter/innen in der Stichprobe größer gewesen wäre.	
Das Konfidenzintervall wäre breiter, wenn der Anteil der Befürworter/innen und der Anteil der Gegner/innen in der Stichprobe gleich groß gewesen wären.	X?

Lösung laut „Korrekturheft“ wie angekreuzt. Nach meiner Rechnung mit der „Anteilsformel“ und z = 1,96 (γ = 95 %) ergibt sich aber für $\hat{p}$ = 0,6 das Intervall 0,557 < p < 0,643 (wie oben angeführt) und für $\hat{p}$ = 0,5 ergibt sich 0,457 < p < 0,543, also dieselbe Intervallbreite von 0,086.

Mat. 16/HT/Aufg. 24/1:

Auf der Grundlage einer Zufallsstichprobe der Größe n_1 gibt ein Meinungsforschungsinstitut für den Stimmenanteil einer politischen Partei das Konfidenzintervall [0,23; 0,29] an. Das zugehörige Konfidenzniveau (die zugehörige Sicherheit) beträgt γ_1. Ein anderes Institut befragt n_2 *z*ufällig ausgewählte Wahlberechtigte und gibt als entsprechendes Konfidenzintervall mit dem Konfidenzniveau γ_2 das Intervall [0,24; 0,28] an. Dabei verwenden beide Institute dieselbe Berechnungsmethode. Aufgabenstellung: Ergänzen Sie die Textlücken im folgenden Satz durch Ankreuzen der jeweils richtigen Satzteile so, dass eine korrekte Aussage entsteht: Unter der Annahme von $n_1 = n_2$ kann man aus den Angaben ...[1)] folgern; unter der Annahme von $\gamma_1 = \gamma_2$ kann man aus den Angaben ...[2)] folgern. Möglichkeiten für [1)]: $\gamma_1 < \gamma_2$, $\gamma_1 = \gamma_2$, $\gamma_1 > \gamma_2$. Möglichkeiten für [2)]: $n_1 < n_2$, $n_1 = n_2$, $n_1 > n_2$.

Lösung: $\gamma_1 > \gamma_2$ und $n_1 < n_2$ nach der Regel: Je größer das γ und je kleiner die Stichprobe, umso breiter das Intervall.

Mat. 17/1. NT/Aufg. 24/1:

Für eine Wahlprognose wird aus allen Wahlberechtigten eine Stichprobe ausgewählt. Von 400 befragten Personen geben 80 an, die Partei *Y* zu wählen. Aufgabenstellung: Geben Sie ein symmetrisches 95-%-Konfidenzintervall für den Stimmenanteil der Partei *Y* in der Grundgesamtheit an

Lösung: 16 % < p < 24 % nach der „Anteilsformel“.

Mat. 16/1. NT/Aufg. 24/1:

Bei einer repräsentativen Umfrage in Österreich geht es um die in Diskussion stehende Abschaffung der 500-Euro-Scheine. Es sprechen sich 234 von 1 000 Befragten für eine Abschaffung aus. Aufgabenstellung: Geben Sie ein symmetrisches 95-%-Konfidenzintervall für den relativen Anteil der Österreicherinnen und Österreicher, die eine Abschaffung der 500-Euro-Scheine in Österreich befürworten, an!

Lösung: 20,8 % < p < 26 % nach der „Anteilsformel“.

Mat. 17/HT/Aufg. 24/1:

Um den Stimmenanteil einer bestimmten Partei *A* in der Grundgesamtheit zu schätzen, wird eine zufällig aus allen Wahlberechtigten ausgewählte Personengruppe befragt. Die Umfrage ergibt für den Stimmenanteil ein 95-%-Konfidenzintervall von [9,8 %; 12,2 %]. Aufgabenstellung: Welche der folgenden Aussagen sind in diesem Zusammenhang auf jeden Fall korrekt? Kreuzen Sie die beiden zutreffenden Aussagen an!

Die Wahrscheinlichkeit, dass eine zufällig ausgewählte wahlberechtigte Person die Partei *A* wählt, liegt sicher zwischen 9,8 % und 12,2 %.	
Ein anhand der erhobenen Daten ermitteltes 90-%-Konfidenzintervall hätte eine geringere Intervallbreite.	X
Unter der Voraussetzung, dass der Anteil der Partei-*A*-Wähler/innen in der Stichprobe gleich bleibt, würde eine Vergrößerung der Stichprobe zu einer Verkleinerung des 95-%-Konfidenzintervalls führen.	X

95 von 100 Personen geben an, die Partei *A* mit einer Wahrscheinlichkeit von 11 % zu wählen.	
Die Wahrscheinlichkeit, dass die Partei *A* einen Stimmenanteil von mehr als 12,2 % erhält, beträgt 5 %.	

Lösung wie angekreuzt aufgrund der genannten Regeln.

Mat. 18/HT/Aufg. 24/1:

Vier Konfidenzintervalle (*A*, *B*, *C* und *D*) für einen unbekannten Anteil werden auf dieselbe Art und Weise ausschließlich unter Verwendung des Stichprobenumfangs *n*, des Konfidenzniveaus γ und des relativen Anteils berechnet, wobei der relative Anteil für alle vier Konfidenzintervalle derselbe ist. Die Konfidenzintervalle liegen symmetrisch um den relativen Anteil.

Konfidenzintervall	Stichprobenumfang *n*	Konfidenzniveau γ
A	500	90 %
B	500	95 %
C	2 000	90 %
D	2 000	95 %

Aufgabenstellung: Vergleichen Sie diese vier Konfidenzintervalle bezüglich ihrer Intervallbreite und geben Sie das Konfidenzintervall mit der kleinsten und jenes mit der größten Intervallbreite an!

Lösung: Kleinste Intervallbreite C, größte Intervallbreite B.

Das lässt sich rein verstandesgemäß lösen: Je kleiner die Stichprobe und je größer die Genauigkeit, umso breiter das Intervall.

Mat. 18/1. NT/24:

Ein Spielzeuge produzierendes Unternehmen führt in einer Gemeinde in 500 zufällig ausgewählten Haushalten eine Befragung durch und erhält ein 95-%-Konfidenzintervall für den unbekannten Anteil aller Haushalte dieser Gemeinde, die die Spielzeuge dieses Unternehmens kennen. Bei einer anderen Befragung von *n* zufällig ausgewählten Haushalten ergab sich derselbe Wert für die relative Häufigkeit. Das aus dieser Befragung mit derselben Berechnungsmethode ermittelte symmetrische 95-%-Konfidenzintervall hatte aber eine geringere Breite als jenes aus der ersten Befragung. Aufgabenstellung: Geben Sie alle $n \in \mathrm{N}$ an, für die dieser Fall unter der angegebenen Bedingung eintritt!

Lösung: n > 500 nach der fett gedruckten Regel

Mat. 19/HT/Aufg. 24/1:

Jemand möchte den unbekannten Anteil *p* derjenigen Wählerinnen und Wähler ermitteln, die bei einer Wahl für den Kandidaten *A* stimmen werden, und beauftragt ein Meinungsforschungsinstitut damit, diesen Anteil *p* zu schätzen. Im Zuge dieser Schätzung werden 200 Stichproben mit jeweils gleichem Umfang ermittelt. Für jede dieser Stichproben wird das entsprechende 95-%-Konfidenzintervall berechnet. Aufgabenstellung: Berechnen Sie die erwartete Anzahl a derjenigen Intervalle, die den unbekannten Anteil *p* enthalten!

Lösung: a = 200.0,95 = 190 laut „Korrekturheft“. **Bei dieser Fragestellung ist also die Anzahl der Stichproben mit der Genauigkeit γ = 0,95 (= 95 %) der Schätzung zu multiplizieren. Das ist einsichtig.**