23 Statistik

Ziel dieses Kapitels ist es eine Intuition für grundlegende statistische Begriffe und deren Eigenschaften zu bilden. Sie sollen begreifen, dass jede Messung und damit jede Statistik grundsätzlich ungenau ist. Die Ungenauigkeit einer Messung kann ebenfalls abgeschätzt werden und *muss* angegeben werden.

> Statistik «ist die Lehre von Methoden zum Umgang mit quantitativen Informationen» (Daten). Sie ist eine Möglichkeit, «eine systematische Verbindung zwischen Erfahrung (Empirie) und Theorie herzustellen».[1] Unter Statistik versteht man die Zusammenfassung bestimmter Methoden zur Analyse empirischer Daten. [...] Die Statistik wird als Hilfswissenschaft von allen empirischen Disziplinen und Naturwissenschaften verwendet, wie zum Beispiel der Medizin (Medizinische Statistik), der Psychologie (Psychometrie), der Politologie, der Soziologie, der Wirtschaftswissenschaft (Ökonometrie), der Biologie (Biometrie), der Chemie (Chemometrie) und der Physik. Die Statistik stellt somit die theoretische Grundlage aller empirischen Forschung dar. Da die Menge an Daten in allen Disziplinen rasant zunimmt, gewinnt auch die Statistik und die aus ihr abgeleitete Analyse dieser Daten an Bedeutung. Andererseits ist die Statistik ein Teilgebiet der reinen Mathematik.

Quelle: https://de.wikipedia.org/wiki/Statistik, 18. Oktober 2018

Aufgabe 455 Man findet manchmal Werbungen für Anlagefonds, die sich damit brüsten, in den letzten fünf Jahren immer besser als der Börsendurchschnitt gewirtschaftet zu haben, im Gegensatz zu anderen Fonds.

Wir wollen Anlagefonds mit Münzwürfen nachstellen. Zahl heisst jeweils «der Fonds erwirtschaftet mehr Gewinn als der Börsendurchschnitt», Kopf heisst «weniger».

Nehmen Sie zwei Münzen und werfen Sie jede fünf mal. Notieren Sie sich, wie viel mal Sie «Zahl» geworfen haben. Wer hat die «beste» Münze?

Diskutieren Sie das Resultat und den Zusammenhang mit obiger Werbung für Anlagefonds.

Aufgabe 456 Eine Münze wird 50 mal geworfen und die Anzahl Würfe mit «Zahl» gezählt.

a) Wie viel mal «Zahl» erwarten Sie durchschnittlich?

b) Führen Sie den Versuch mindestens einmal durch und notieren Sie an die Wandtafel, wieviel mal Sie «Zahl» geworfen haben. Notieren Sie daneben auch die prozentuale Abweichung vom erwarteten Wert.

c) Kopieren Sie alle Resultate von der Wandtafel in Ihr Heft und berechnen Sie folgende Grössen: Total der Anzahl Würfe, total der Würfe mit «Zahl», die erwartete Anzahl Würfe «Zahl» und die prozentuale Abweichung davon.

d) Machen Sie eine Grafik, aus der ersichtlich ist, welche Anzahl «Zahlwürfe» wie viel mal vorgekommen ist.

Aufgabe 457 Mit einer Münze kann ein zufälliges Ergebnis Kopf/Zahl erzeugt werden. Das kann dann als Ja/Nein, 0/1 oder z.B. als eine der beiden Mannschaften interpretiert werden, die das Anspiel erhält.

a) Geben Sie eine Methode an, wie eine natürliche Zahl zwischen 0 und 3 (inklusive) mit wiederholten Münzwürfen erzeugt werden kann.

b) Suchen Sie in der Klasse Personen, die in a) die gleiche Methode wie Sie vorgeschlagen haben. Überprüfen Sie dann durch mehrmaliges Wiederholen der Methode, ob alle Zahlen auch in etwa gleich häufig erzeugt werden. Tragen Sie dazu Ihre Resultate zusammen und zeichnen Sie die Häufigkeiten grafisch auf.

c) Erklären Sie die gemessenen Häufigkeiten.

d) Schlagen Sie eine Methode mit wiederholten Münzwürfen vor für die Erzeugung zufälliger natürlicher Zahlen von 0 bis und mit 15. Alle Zahlen sollen mit gleicher Wahrscheinlichkeit erzeugt werden.

e) Schlagen Sie eine Methode vor für die gleichwahrscheinliche zufällige Erzeugung der natürlichen Zahlen von 0 bis und mit 2.

[1] Horst Rinne (2008): *Taschenbuch der Statistik.* (4. Auflage), Harri Deutsch Verlag, Frankfurt am Main, S. 1

Aufgabe 458 Wie gross ist die Wahrscheinlichkeit einen «Joland zu pullen», d.h. bei 50 Münzwürfen genau 4 mal «Zahl» zu werfen?

Grundidee: Wirft man eine Münze 50 mal, kann das Resultat mit Nullen und Einsen aufgeschrieben werden (0 ist «Kopf», 1 ist «Zahl»). Man erhält also ein «Wort» mit 50 Zeichen.

Jedes dieser Wörter erhält man mit gleicher Wahrscheinlichkeit. Es gibt genau ein Wort mit 0 Einsen, 50 Wörter mit genau einer Eins und noch viel mehr Wörter mit mehr Einsen.

Die erste Frage, die wir beantworten wollen, ist die Frage, wie viele Wörter der Länge 50 es überhaupt gibt. Danach zählen wir, wie viele Wörter es gibt mit Länge 50 mit genau vier Einsen. Der Quotient der beiden Zahlen ergibt die Wahrscheinlichkeit.

Gesamte Anzahl Wörter mit Länge 50 Für die gesamte Anzahl Wörter mit Länge 50 betrachten wir erst einmal Wörter der Länge 1, 2, usw.

Länge	Anzahl	Systematische Aufzählung aller Wörter
1		
2		
3		
4		
50		
n		

Um einzusehen, dass sich die Anzahl jedesmal verdoppelt, helfen folgende Überlegungen:

✎

Anzahl Wörter mit Länge 50 mit genau vier Einsen Um die Anzahl zu zählen, fragen wir uns, auf wie viele Arten können solche Wörter gebildet werden. Wir stellen uns dazu 50 Stellen mit Nullen gefüllt vor. Dann platzieren wir vier Einsen.

Für die Platzierung der ersten Eins haben wir 50 Möglichkeiten. Für die zweite Eins haben wir noch 49 Möglichkeiten (auf den Platz der ersten Eins dürfen wir diese nicht setzen). Für die dritte Eins bleiben noch 48 und für die vierte Eins noch 47 Möglichkeiten. Total haben wir also

$$50 \cdot 49 \cdot 48 \cdot 47 = 5'527'200 \text{ Möglichkeiten.}$$

Aber *Achtung*! Wir haben jedes einzelne Wort mehrfach gezählt. Als Beispiel soll das Wort 10110010. . . 0 dienen. Dieses Wort wurde z.B. mit der Platzierung 10230040. . . 0 erreicht (wobei die Ziffern für die Nummer der Platzierung stehen). Aber natürlich auch mit der Platzierung 40320010. . . 0. Wir müssen also folgende Frage beantworten:

Auf wie viele Arten können vier Einsen auf vier Positionen platziert werden?

Für die erste Eins haben wir 4 Möglichkeiten, für die zweite noch 3, für die dritte 2 und die letzte nur noch 1 Möglichkeit. Total also

$$4 \cdot 3 \cdot 2 \cdot 1 = 24 \text{ Möglichkeiten.}$$

Wir haben oben also jedes Wort 24-fach gezählt. Damit ist die Anzahl Wörter mit Länge 50 und genau 4 Einsen

$$\frac{50 \cdot 49 \cdot 48 \cdot 47}{4 \cdot 3 \cdot 2 \cdot 1} = 230'300.$$

Die Wahrscheinlichkeit, einen «Joland zu pullen» ist also genau so gross, wie aus 2^{50} Wörtern zufällig eines der $230'300$ Wörtern zu erwischen:

$$\frac{230300}{2^{50}} = \frac{57'575}{281'474'976'710'656} \approx 2.045 \cdot 10^{-10} \quad \text{also etwa 1 zu 5 Milliarden.}$$

Aufgabe 459 Wie gross ist die Wahrscheinlichkeit, mit einem Tipp im Schweizer Zahlenlotto zu gewinnen?

1 : 8 Millionen $\quad \frac{45 \cdot 44 \cdots 40}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}$

a) Früher musste für einen Tipp sechs (unterschiedliche) Zahlen aus 45 ausgewählt werden.

b) Heute müssen für eine Tipp sechs (unterschiedliche) Zahlen aus 42 ausgewählt werden, plus eine Glückszahl aus 6. $\approx$ 1 : 30 Millionen bei 2.50 pro Tipp 75 Mio

Aufgabe 460 Wir betrachten ein Leiterspiel bei dem mit einem normalen Würfel mit Zahlen 1 bis 6 gewürfelt wird. Es wurde behauptet, dass das Feld 28 oft ein Spezialfeld ist, weil die Wahrscheinlichkeit, dieses zu erreichen grösse sei als für andere Felder. Der Grund sei, dass man im Durchschnitt 3.5 würfelt, dass also Vielfache von 7 mit grösserer Wahscheinlichkeit besucht werden.

a) Simulieren Sie die Situation mit einer Tabellenkalkulation und vergleichen Sie die Besuchshäufigkeiten der Felder 27, 28 und 29. Können Sie zuverlässig Unterschiede feststellen? Was ist mit den Feldern 5, 6 und 7? Können Sie sich die Resultate plausibel erklären?

b) Berechnen Sie die exakten Wahrscheinlichkeiten P_i für den Besuch der Felder 0 bis 30.

23.1 Lage- und Streuungsmasse

Wir gehen im Folgenden davon aus, dass wir einen reellen Wert messen möchten und dass die Messung mehrmals wiederholt wird. Als Resultat hätten wir nicht nur gerne eine Schätzung für den wahren Wert, sondern auch noch ein Intervall, worin der wahre Wert mit einer bestimmten Wahrscheinlichkeit (typischerweise 95%) liegen soll.

Die Anzahl Messwerte wird im Folgenden mit n und die einzelnen Messwerte mit $x_1, x_2, \ldots, x_n$ bezeichnet.

Beispiele solcher Messungen sind z.B. die Zeugnisnote in Mathematik, Blutdruckwerte, CO_2-Werte in der Athmosphäre, die Grösse von Naturkonstanten, etc.

23.1.1 Lagemasse

Das verbreitetste Lagemass ist der **Durchschnitt**, auch **Mittelwert** oder **arithmetisches Mittel** genannt.

Merke Mittelwert

$$\mu = \frac{1}{n}(x_1 + x_2 + x_3 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Anstatt μ (mü, griechisches 'm') ist auch die Notation $\overline{x}$ gebräuchlich.

Ein weiteres Lagemass, das ebenfalls oft angetroffen wird, ist der sogenannte **Median**, auch **Zentralwert** genannt. Er wird meist wie folgt berechnet:

Merke Median

Erst werden die Messwerte aufsteigend sortiert. Der Median ist dann der Wert in der Mitte der Liste (bei ungeradem n), bzw. der Mittelwert der beiden Werte in der Mitte der Liste (bei geradem n).

$\tilde{x}$

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}} & \text{wenn } n \text{ ungerade} \\ \frac{1}{2}\cdot\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \end{cases}$$

Tilde $\quad$ x_4 $\quad$ 4 $\quad$ 7 $\quad$ $n=4$ $\quad$ $\frac{x_2+x_3}{2}$

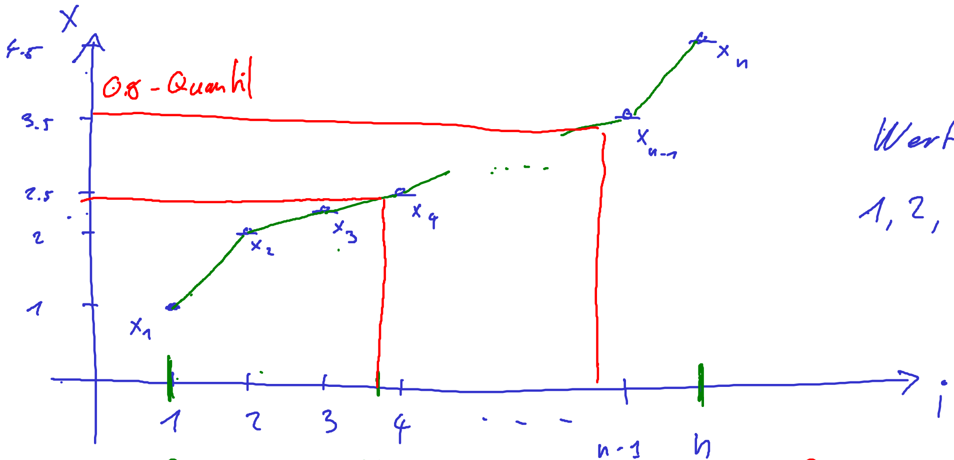
Berechnen μ und $\tilde{x}$ von

a) $\overset{6}{7},\ \overset{1}{2},\ \overset{2}{4},\ \overset{4}{5},\ \overset{3}{4},\ \overset{7}{17},\ \overset{5}{6}$ $\quad n = 7$

b) $7,\ 2,\ 4,\ 5,\ 4,\ 17$

a) $\mu \cong 6.42 \qquad \tilde{x} = 5$

b) $\mu \cong 6.5 \qquad \tilde{x} = 4.5$



Werte sortiert:

1, 2, 2.25, 2.5, ...

... 3,5, 4,5

0.25-Quantil $\leadsto$ $i = 3.75$ $\leadsto$ Quantil ≈ 2.4

Aufgabe 461 In einer Bar sind 20 «normale» Leute. Wie gross schätzen Sie den Median und den Mittelwert der Einkommen der anwesenden Gäste?

Der CEO eines Schweizer Grosskonzerns betritt die Bar. Wie sieht es jetzt mit dem Median und dem Mittelwert der Einkommen der anwesenden Gäste aus?

Welches der beiden Lagemasse ist in diesem Fall angemessener?

Aufgabe 462 In einer Klasse mit 20 Schülern war der Notendurchschnitt bei einer Prüfung 4.42, wobei eine Schülerin die Prüfung nicht geschrieben hatte. Nach der Nachprüfung war der Notendurchschnitt eine 4.48.
Die genaue Aufgabenstellung fehlt da absichtlich. Diskutieren und berechnen Sie!

Streuungsmasse

Ein Streuungsmass gibt an, wie sehr eine Wertereihe «fluktuiert», d.h. wie verschieden die Werte sind, oder wie sehr «verstreut» die Werte sind.

Warum ein Streuungsmass interessant ist, soll wie folgt motiviert werden: Man nimmt an, man hat zwei unterschiedliche Messungen durchgeführt, die aber das Gleiche messen sollen. Beide Messungen produzieren zwei Wertereihen mit gleich vielen Werten. Die Streuung der ersten Wertereihe ist aber grösser als die der zweiten. Welcher Mittelwert ist «vertrauenswürdiger»? Bzw. bei welcher Messung kann ein kleineres «Vertrauensintervall» angegben werden?

Es gibt verschiedene Varianten, die Streuung zu messen. Eine geläufige Variante (weil mathematisch einfach zu analysieren) ist die «Standardabweichung» (mit 'd'!).

Merke Standardabweichung

Die (empirische) Standardabweichung einer Wertereihe $x_1, x_2, \ldots, x_n$ ist

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2},$$

wobei μ der Mittelwert und σ der griechische Kleinbuchstaben «sigma» ('s') ist.

Wie der Mittelwert ist die Standardabweichung **nicht robust**, d.h. deren Wert wird stark durch **Ausreisser** beeinflusst.

Aufgabe 463 Berechnen Sie den Mittelwert und die Standardabweichung folgender Wertereihen:

a) 105, 100, 100, 95, 105, 98. $\sum_{i=1}^{6}(x_i-100.5)^2 = 77.5$ $\sigma = \sqrt{\frac{1}{5}\cdot 77.5} \approx 3.937$

b) Eine Folge von 100 Einsen. $\sigma = 0$

c) Eine Folge mit 785 Einsen und 215 Nullen. Zusatzfrage: welches Experiment könnte diese Daten geliefert haben? $\sigma \approx 0.41$

$\sigma = C.3$

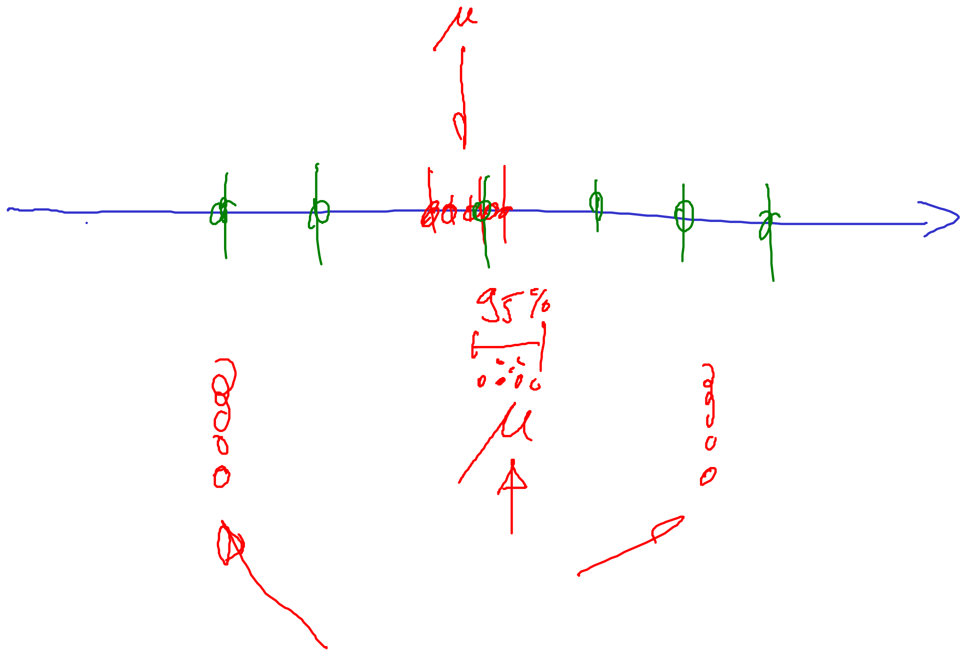
d) 1, 2, 3, …, 101. Verwenden Sie dazu die Summenzeichen-Funktion des TR, zu erreichen mit «menu, 4, 5». Auf dem Ti-89 mit ~~sum~~ Σ`(Ausdruck, Laufvariable, untere Grenze, obere Grenze)`.

Ein weiteres Streuungsmass ist der sogenannte «**Interquartilsabstand**». Dabei sind die «**Quartile**» ein Spezialfall der Quantile, die informell wie folgt definiert sind:

> Zu einer Wertereihe $x_1, x_2, \ldots, x_n$ ist der p-Quantil ein Wert q so, dass ca. $100 \cdot p\%$ der Werte kleiner oder gleich q ist.

Ein Spezialfall ist der Median, der dem 0.5-Quantil, bzw. dem 50%-Quantil entspricht. Wie auch beim Median, gibt es verschiedene Defintionen für die Berechnung der Quantile. Wir verwenden hier die gleiche Definition, wie gängige Tabellenkalkulationsprogramme.

Streuung



$p \in [0,1]$

Definition 60 p-Quantil z.B. 0.25 $0 \leq p \leq 1$

Erst werden die Werte aufsteigend sortiert: $x_1 \leq x_2 \leq \ldots \leq x_n$. Aus p wird eine Position mit einer linearen Funktion $f(p)$ zwischen 1 und n berechnet, so dass $f(0) = 1$ und $f(1) = n$. Ist $f(p)$ eine natürliche Zahl, ist das p-Quantil gleich $x_{f(p)}$.
Sonst rundet man $f(p)$ auf die nächste natürliche Zahl i ab. Man bestimmt eine lineare Funktion $g(p)$ so, dass $g(i) = x_i$ und $g(i+1) = x_{i+1}$. Das p-Quantil ist dann $g(p)$.

$f(p)$ 0,25-Quantil

Merke

Weiter definiert man das 1. Quartil als das 0.25-Quantil und das 3. Quartil als das 0.75-Quantil.
Oft spricht man auch von Perzentilen, wenn p ein ganzzahliges Vielfache von $\frac{1}{100}$ ist.

Aufgabe 464 Gegeben sind 2 reelle Zahlen a und b. Bestimmen Sie die Funktionsgleichung f jener linearen Funktion $f(t)$ für die $f(0) = a$ und $f(1) = b$ gilt.

Aufgabe 465

a) Berechnen Sie das p-Quantil für $p = 0.1$, $p = 0.25$ und $p = 0.5$ für die Wertereihe 4, 6, 5, 4, 2.

b) Für vorhergehende Aufgabe, sortieren Sie die Werte aufsteigend und zeichnen Sie die Punkte (i, x_i) in ein Koordinatensystem ein verbinden Sie benachbarte Punkte mit einer Strecke. Zeichnen Sie von $x = 1$ bis $x = 5$ eine zweite Skala von 0 bis 1 ein. Wie können damit die Quantile grafisch bestimmt werden?

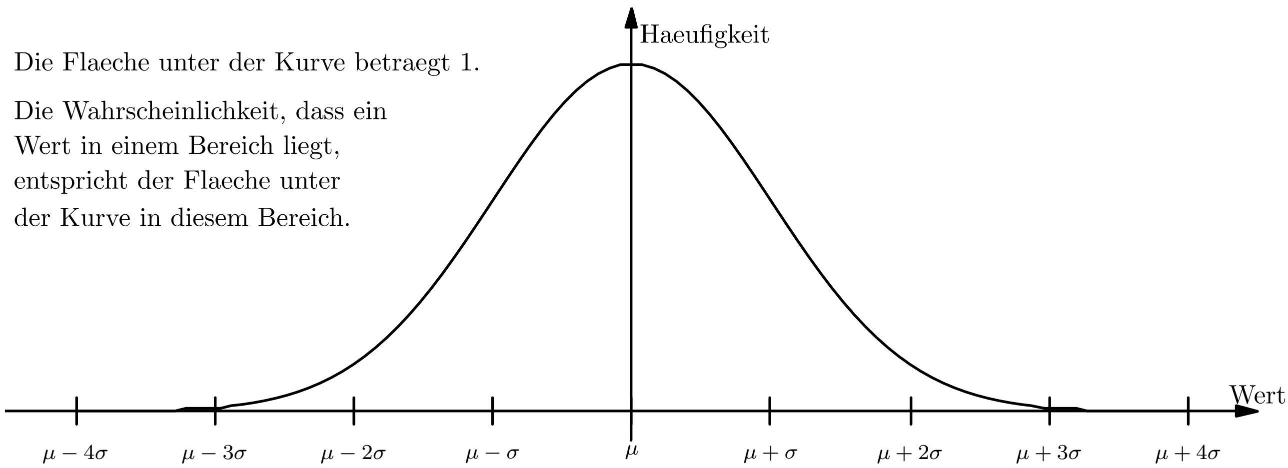
c) Berechnen Sie das 0.25-Quantil (das erste Quartil) für die Wertereihe 1, 2, 3, ..., 10.

23.2 Streuung des Mittelwerts

Dem Mittelwert einer Messreihe kann ebenfalls eine Streuung zugeordnet werden. Diese Streuung könnte man empirisch ermitteln, indem das Experiment viele Male wiederholt wird. Das ist aber meistens nicht praktikabel. Man kann aber mathematisch Folgendes für die Standardabweichung σ_μ des Mittelwerts μ nachweisen:

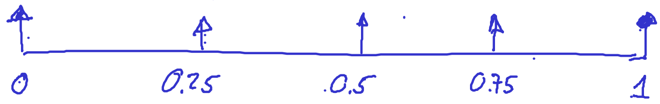
$$\sigma_\mu \approx \frac{1}{\sqrt{n}}\sigma. \qquad \text{in guter Näherung für } n \geq 50.$$

Hat man weniger als 50 Messwerte, wird mit obiger Formel die Streuung unterschätzt, d.h. ein davon abgeleitetes Vertrauensintervall ist zu klein (es ist dann die sogenannte t-Verteilung zu benutzen).
Weiter kann mathematisch gezeigt werden, dass **Mittelwerte** gemäss einer **Glockenkurve** verteilt sind (was man durch vielmaliges Wiederholen des Experiments anschaulich machen könnte).

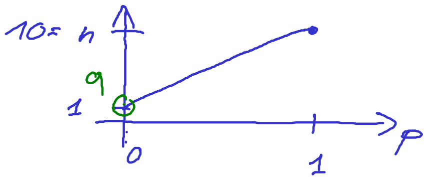


A 4 (5 c) (x_n): $1, 2, 3, 4, 5; 6, 7, 8, 9, 10$

1. & 3. Quartil?



Position: lineare Fkt. f mit $f(0) = 1$, $f(1) = 10 = n$



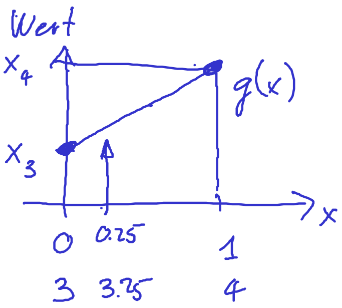
$$f(p) = \overbrace{m}^{\text{Steigung}} \cdot p + \overbrace{q}^{\text{y-Achsenabschnitt}}$$

$$f(p) = 9 \cdot p + 1 = (n-1) \cdot p + 1$$

$$f(0.25) = \frac{9}{4} + 1 = 3.25$$

Wert: lin. Fkt. $g(x)$ so, dass $g(0) = x_3$

$g(1) = x_4$



$$g(x) = (x_4 - x_3) \cdot x + x_3$$

1. Quartil: $g(0.25) = (4-3) \cdot \frac{1}{4} + 3 = \underline{\underline{3.25}}$

Position für 3. Quartil: $f(0.75) = 0.75 \cdot 9 + 1 = 7.75$

3. Quartil: 7.75 $\left(= 0.75 \cdot (8-7) + 7 \right.$
$\left. \quad 0.75 \cdot (x_8 - x_7) + x_7 \right)$

Für die Glockenkurve (Normalverteilung) mit der Spitze bei μ und der «Breite» σ gelten folgende Faustregeln:

Merke Faustregeln für die Glockenkurve

Im Intervall $[\mu - \sigma, \mu + \sigma]$ liegen ungefähr $\frac{2}{3}$ (68.3%) der Werte.
Im Intervall $[\mu - 2\sigma, \mu + 2\sigma]$ liegen ungefähr 95% (95.45%) der Werte.

Weitere Werte sind für $\pm 3\sigma$ 99.73%, für $\pm 4\sigma$ 99.993% und für $\pm 6\sigma$ 99.9999998% (damit arbeiten die Physiker am CERN).
Daraus lassen sich jetzt für den Mittelwert **Vertrauensintervalle** bzw. **Konfidenzintervalle** bilden, d.h. Intervalle in der der wahre Wert mit grosser Wahrscheinlichkeit liegt (typischerweise 95%).

Merke Vertrauensintervall für den Mittelwert

Bei 50 Messungen oder mehr, liegt der wahre Wert mit Wahrscheinlichkeit 95% im Intervall $[\mu - 2\sigma_\mu, \mu + 2\sigma_\mu]$.

23.3 Repetitionsaufgaben

Aufgabe 466 Gegeben sind die Wertereihen (x_i):

a) 13, 7, 6, 10, 8, 12, 7, 9

b) Eine Folge aus 100 Nullen und 200 Einsen.

Berechnen Sie den Mittelwert μ, den Median $\tilde{x}$, die Standardabweichung σ und das erste und dritte Quartil $q_{0.25}$ und $q_{0.75}$.

Aufgabe 467 Mit einem Würfel wird 600 mal gewürfelt. Dabei werden 120 Sechsen gezählt.
Daraus erhält man eine Wertereihe aus Nullen (keine Sechs gewürfelt) und Einsen (Sechs gewürfelt). Berechnen Sie den Mittelwert μ sowie die Standardabweichung des Mittelwerts σ_μ. Geben Sie ein 95%-Vertrauensintervall für den «wahren» Wert an. Würden Sie sagen, dass der Würfel gezinkt ist?

Aufgabe 468 In einer online-Umfrage wurde gefragt: «Haben Sie eine Sehhilfe»?
Es antworten 80 Teilnehmer mit folgendem Resultat: 20 Teilnehmer antworten «Ja», der Rest «Nein». Es wird getitelt «25% der Schweizer benötigen eine Sehhilfe».
Wie beurteilen Sie die Zuverlässigkeit der Umfrage und der Schlussfolgerung?

Aufgabe 469 Im Euromillions müssen erstens 5 Zahlen aus 50 ausgewählt werden. Zusätzlich dazu müssen zweitens 2 Sterne aus 12 ausgewählt werden.
Wie viele Möglichkeiten gibt es für die erste Wahl? Wie viele für die zweite? Wie viele total? Was ist die Wahrscheinlichkeit die 5 richtigen Zahlen und die 2 richtigen Sterne auszuwählen?

Aufgabe 470 Wir betrachten ein Poker-Spiel mit 52 Karten (4 Farben, Werte 2-10,J,Q,K,A). Eine Poker-Hand besteht aus 5 Karten. Wie viele mögliche Poker-Hände

a) gibt es insgesamt?

b) gibt es mit vier Assen?

c) gibt es mit *genau* vier gleichen Werten?

d) gibt es mit *genau* drei gleichen Werten?

e) gibt es mit einem Full-House (d.h. 2 gleiche und 3 gleiche Werte)?

Aufgabe 471 Die Schulzimmerwetterstation (CO_2 Messgerät) im H21 überträgt die Daten drahtlos über das LoRa-Protokoll ins Informatiklabor im E-Stock. Cédric Solenthaler hat dabei für verschiedene Antennenkonfigurationen die gemessene Empfangsqualität aufgezeichnet, um herauszufinden, welche Konfiguration die beste Qualität liefert. Werten Sie die Messungen aus und beurteilen Sie.

23.4 Lösungen

Hinweise zu den Symbolen:

Diese Aufgaben könnten (mit kleinen Anpassungen) an einer Prüfung vorkommen. Für die Prüfungsvorbereitung gilt: "If you want to nail it, you'll need it".

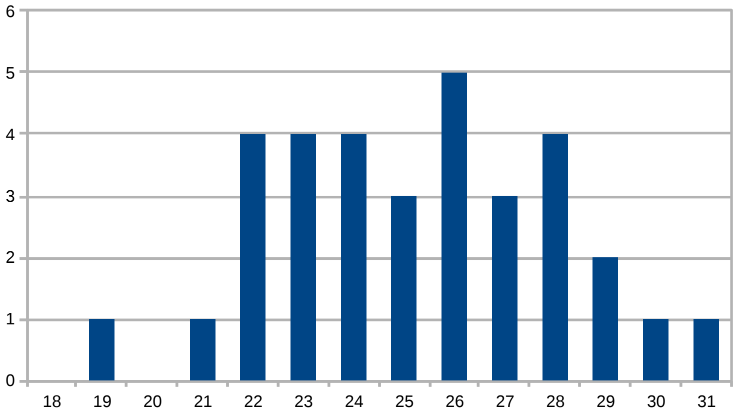
Diese Aufgaben sind wichtig, um das Verständnis des Prüfungsstoffs zu vertiefen. Die Aufgaben sind in der Form aber eher nicht geeignet für eine Prüfung (zu grosser Umfang, nötige «Tricks», zu offene Aufgabenstellung, etc.). **Teile solcher Aufgaben können aber durchaus in einer Prüfung vorkommen!**.

Diese Aufgaben sind dazu da, über den Tellerrand hinaus zu schauen und oder die Theorie in einen grösseren Kontext zu stellen.

Lösung zu Aufgabe 455 ex-bester-muenzwurf-fonds

Lösung zu Aufgabe 456 ex-muenze-durchschnitt

d) Ein mögliches Histogram sieht wie folgt aus:



Lösung zu Aufgabe 457 ex-wuerfeln-mit-muenze

a) Möglichkeit 1: 3 Würfe, Anzahl «Zahl» zählen. Möglichkeit 2: Münze 1 entscheidet gerade/ungerade, Münze 2 entscheidet 0/1 oder 2/3. Man könnte auch eine Binärzahl mit zwei Stellen erzeugen.

b) Bei Möglichkeit 1 kommen die 1 und die 2 drei mal so häufig vor wie die 0 und die 3. Bei Möglichkeit zwei sind die Häufigkeiten ausgeglichen.

c) Bei Möglichkeit 2 gibt es 8 gleich wahrscheinliche Varianten, wobei 0 für Kopf und 1 für Zahl steht:

   000, 001, 010, 011, 100, 101, 110, 111.

   In $\frac{1}{8}$ der Fälle erhält man eine 0, in $\frac{3}{8}$ der Fälle eine 1, ebenso für die 2 und die 3 erhält man wieder nur in $\frac{1}{8}$ der Fälle.

   Bei Möglichkeit 1 gibt es 4 gleich wahrscheinliche Varianten, die den Zahlen 0 bis 3 zugeordnet werden.

d) Binärzahl mit 4 Stellen, oder Intervallhalbierung. Erste Münze entscheidet < 8 oder ≥ 8. Die nächste Münze halbiert die verbleibenden Zahlen etc.

e) Wie Möglichkeit 1, aber man wiederholt den Prozess, wenn eine 3 herauskommt.

Lösung zu Aufgabe 458 ex-wahrscheinlichkeit-4-von-50

Lösung zu Aufgabe 459 ex-lotto-6-aus-45

a) Wir wählen 6 Zahlen nacheinander aus, dafür haben wir

$$45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40 = 5'864'443'200 \text{ Möglichkeiten.}$$

Wir haben haben aber wieder mehrfach gezählt, so kann z.B. die Kombination 1,2,3,4,5,6 auf verschiedene Arten erreicht werden. Für die erste Auswahl haben wir 6 Möglichkeiten, für die zweite 5, etc.

Damit gibt es also

$$\frac{45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 8'145'060 \text{ Möglichkeiten.}$$

Von all diesen Möglichkeiten gewinnt nur genau eine. Die Wahrscheinlichkeit ist also

$$\frac{1}{8'145'060}$$

Die Wahrscheinlichkeit ist also ungefähr 1 zu 8 Millionen.

b) Analog zu a) haben wir für die sechs Zahlen

$$\frac{42 \cdot 41 \cdot 40 \cdot 39 \cdot 38 \cdot 37}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 5'245'786 \text{ Möglichkeiten.}$$

Für jede dieser Möglichkeiten haben wir sechs Möglichkeiten für die Auswahl der Glückszahl, also total

$$6 \cdot 5'245'786 = 31'474'716 \approx \pi \cdot 10^7 \text{ Möglichkeiten.}$$

Da nur genau eine dieser Möglichkeiten gewinnt, ist die Gewinnwahrscheinlichkeit mit einem Tipp

$$\frac{1}{31'474'716}.$$

Lösung zu Aufgabe 460 ex-leiterspiel-feld-28

a) Ein Unterschied zwischen den Feldern 27, 28 und 29 ist empirisch nicht festzustellen. Der Unterschied zwischen den Feldern 5, 6 und 7 aber sehr wohl. Besonders das Feld 7 wird klar weniger besucht als das Feld 6.

Das Feld 6 kann direkt erreicht werden. Feld 7 nur indirekt. Die Wahrscheinlichkeit, dass das Feld 6 indirekt erreicht wird, ist aber nur wenig kleiner als jene für das Feld 7.

b) Die Wahrscheinlichkeiten entsprechen den relativen Besuchshäufigkeiten. Die Grundidee ist folgende: Die Häufigkeit eines Feldes i ergibt sich aus den Häufigkeiten der Felder, von denen das Feld i erreicht werden kann, d.h. aus den Feldern $i-6$, $i-5$, …, $i-1$. Z.B. ist die Wahrscheinlichkeit das Feld i vom Feld $i-6$ zu erreichen genau $\frac{1}{6}$.

Die Berechnung der Wahrscheinlichkeiten erfolgt damit rekursiv wie folgt:

$$\begin{cases} P_0 &= 1 & \text{Startfeld} \\ P_i &= 0 & \text{für } i < 0 \\ P_i &= \displaystyle\sum_{k=i-6}^{i-1} \frac{1}{6} P_k & \text{für } i \geq 1 \end{cases}$$

Diese Rekursion kann z.B. in Excel programmiert werden. Man erhält folgende Werte:

0	1	2	3	4	5	6	7	8	9
1.00000	0.16667	0.19444	0.22685	0.26466	0.30877	0.36023	0.25360	0.26809	0.28037
10	11	12	13	14	15	16	17	18	19
0.28929	0.29339	0.29083	0.27926	0.28354	0.28611	0.28707	0.28670	0.28559	0.28471
20	21	22	23	24	25	26	27	28	29
0.28562	0.28597	0.28594	0.28576	0.28560	0.28560	0.28575	0.28577	0.28574	0.28570

Zum Vergleich $\frac{1}{3.5} \approx 0.285714$, was der durchschnittlichen Besuchswahrscheinlichkeit entspricht. Damit ist die Besuchswahrscheinlichkeit des Feldes 28 1.0000746 mal grösser als der Durchschnitt, aber kleiner als die Besuchswahrscheinlichkeit vom Feld 27.

Lösung zu Aufgabe 461 ex-unterschied-mittelwert-median

Lösung zu Aufgabe 462 ex-note-nachpruefung

Seien $x_1, \ldots, x_{20}$ alle Prüfungsnoten. Es gilt

$$\frac{1}{19}\sum_{i=1}^{19} x_i \approx 4.24 \qquad \text{und} \qquad \frac{1}{20}\sum_{i=1}^{20} x_1 \approx 4.48$$

Wir müssen davon ausgehen, dass die Notenschnitte gerundet und nicht exakt sind (besonders bei 19 Noten).

Multipliziert man die erste Gleichung mit 19 und die zweite mit 20, kann die erste von der zweiten subtrahiert werden und es bleibt

$$x_{20} \approx 20 \cdot 4.48 - 19 \cdot 4.24 = 5.62$$

Die Nachprüfung war wohl eine 5.6. Weil dies aus den Angaben berechnet werden kann, sollte die Lehrperson den neuen Schnitt nicht kommunizieren.

Lösung zu Aufgabe 463 ex-stdev-von-hand

a) $\mu = 100.5$, $\sigma = 3.937$.

b) $\mu = 1$, $\sigma = 0$.

c) Das könnte (ein schon fast gemogelt gutes) Resultat des Experiments sein, bei dem π durch zufälliges Wählen eines Punktes ermittelt wurde.

$\mu = 0.785$. $\sigma^2 = \frac{1}{999} \cdot (215 \cdot 0.785)^2 + (785 \cdot 0.215)^2 \approx 0.1689$. Und damit $\sigma \approx 0.411$.

d) Entweder durch Überlegen oder mit der Summenformel für arithmetische Reihen: $\mu = \frac{1}{n} \cdot 101 \cdot \frac{1+101}{2} = 51$.

$$\sigma = \sqrt{\frac{1}{100}\sum_{i=1}^{101}(i-51)^2} \approx 29.3$$

Lösung zu Aufgabe 464 ex-lineare-01-funktion
Der Achsenabschnitt ist a, die Steigung $\frac{\Delta y}{\Delta x} = \frac{b-a}{1} = b - a$. Also

$$f(t) = a + t \cdot (b - a).$$

Lösung zu Aufgabe 465 ex-quantile-von-hand

a) Median (0.5-Quantil) ist 4, das erste Quartil (0.25-Quantil) ist 4.

Das 10%-Quantil entspricht der Position $i = 1 + 0.1 \cdot (10 - 1) = 1.9$. Wir suchen also den richtigen Wert zwischen x_1 und x_2. Dazu berechnen wir den Wert an der Stelle 0.9 der linearen Funktion $g(t)$, die für $t = 0$ den Wert x_1 und für $t = 1$ den Wert x_2 liefert: $g(t) = x_1 + t \cdot (x_2 - x_1) = 2 + 0.9 \cdot (4 - 2) = 3.8$. Damit ist der 0.1-Quantil gleich 3.8.

b) Betrachtet man den Streckenzug als Graph einer (stückweise linearen) Funktion, kann auf zweiten Skala von 0 bis 1 auf dieser Funktion direkt der entsprechende Quantil abgelesen werden.

c) Position $i = 1 + 0.25 \cdot (10 - 1) = 3.25$. Das 0.25-Quantil ist also $x_3 + \frac{1}{4}(x_4 - x_3) = 3 + \frac{1}{4} \cdot (4 - 3) = 3.25$.

Lösung zu Aufgabe 466 ex-repe-diverse-masse
a) $\mu = 9$, $\tilde{x} = 8.5$, $\sigma = \sqrt{\frac{1}{7} \cdot 44} =\approx 2.507$.
Quartile: Erst die Werte aufsteigend sortieren.
$q_{0.25}$ entspricht der Position $1 + 0.25 \cdot 7 = 2.75$. Da beide Werte 7 sind, ist $q_{0.25} = 7$.
$q_{0.75}$ entspricht der Position $1 + 0.75 \cdot 7 = 6.25$. Der 3. Quartil ist also $0.75 \cdot x_6 + 0.25 \cdot x_7 = 10.5$
b) $\mu = \frac{2}{3}$, $\tilde{x} = 1$, $\sigma = \sqrt{\frac{1}{299}\left(100 \cdot \left(\frac{2}{3}\right)^2 + 200\left(\frac{1}{3}\right)^2\right)} = \sqrt{\frac{1}{299} \cdot \frac{600}{9}} = \sqrt{\frac{1}{299} \cdot \frac{200}{3}} \approx 0.472$.
Das erste Quartil ist 0, das dritte Quartil ist 1.

Lösung zu Aufgabe 467 ex-repe-gezinkter-wuerfel
$\mu = \frac{120}{600} = \frac{1}{5}$.
$\sigma = \sqrt{\frac{1}{599} \cdot \left(120 \cdot \left(\frac{4}{5}\right)^2 + 480 \cdot \left(\frac{1}{5}\right)^2\right)} = \sqrt{\frac{96}{599}} \approx 0.4003$. Daraus folgt
$\sigma_\mu = \frac{\sigma}{\sqrt{n}} \approx 0.01634$.
Damit ergbit sich das 95%-Vertrauensintervall von $[\mu - 2\sigma_\mu, \mu + 2\sigma_\mu] \approx [0.1673, 0.2327]$.
Der erwartete Wert von $\frac{1}{6}$ für einen ungezinkten Würfel ist gerade nicht mehr im Vertrauensintervall. Ist lässt sich sagen, dass so ein Resultat mit einem ungezinkten Würfel unwahrscheinlich ist (weniger wahrscheinlich als 2.5% (die anderen 2.5% wären für ein Resultat von ca. 80 Sechser oder weniger).

Lösung zu Aufgabe 468 ex-repe-umfrage-verlaesslichkeit
Es wird «Nein» mit 0 und «Ja» mit 1 dargestellt.
$\mu = 0.25$, $\sigma = \sqrt{\frac{1}{159}\left(20 \cdot \left(\frac{3}{4}\right)^2 + 60 \cdot \left(\frac{1}{4}\right)^2\right)} = \sqrt{\frac{15}{79}} \approx 0.4357$
$\sigma_\mu = \frac{\sigma}{\sqrt{n}} \approx 0.04872$
Daraus ergibt sich ein 95%-Vertrauensintervall von ca. 15%-35%. Schon rein statistisch ist die Genauigkeit der Umfrage schlecht.
Weiter ist zweifelhaft, ob die Online-Umfrage-Teilnehmer representativ für die Schweizer Bevölkerung sind.

Lösung zu Aufgabe 469 ex-repe-lotto
5 aus 50: $\frac{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2'118'760$.
2 aus 12: $\frac{12 \cdot 11}{2 \cdot 1} = 66$.
Total Möglichkeiten: $66 \cdot 2'118'760 = 139'838'160$.
Die Wahrscheinlichkeit, in EuroMillions alles richtig zu tippen, beträgt $1 : 139'838'160 \approx 7.151 \cdot 10^{-9}$.