SBP Mathe Aufbaukurs 1

Diese Lernkarten sind sorgfältig erstellt worden, erheben aber weder Anspruch auf Richtigkeit noch auf Vollständigkeit.

Das Lernen mit Lernkarten funktioniert nur wenn die Inhalte bereits einmal verstanden worden sind. Ich warne davor diese Lernkarten nur stur auswendig zu lernen.

Diese und andere Lernkarten können von

`http://www.clifford.at/zettelkasten/`

heruntergeladen werden.

Viel Erfolg bei der **SBP Mathe Aufbaukurs 1** Prüfung!

Clifford Wolf <clifford@clifford.at>

Absolute und relative Häufigkeit

Bei n beobachteten Objekten mit k verschiedenen Merkmalsausprägungen:

$$h_i = \frac{H_i}{n} \quad \text{mit } i = 1, 2, \ldots, k$$

H_i	absolute Häufigkeit	$(H_i \in [0; n])$
h_i	relative Häufigkeit	$(h_i \in [0; 1])$

$$\sum_{i=1}^{k} H_i = n \quad \text{und} \quad \sum_{i=1}^{k} h_i = 1$$

Das arithmetische Mittel und seine Eigenschaften

Das arithmetische Mittel $\overline{x}$ ist ein Zentralmass. Es sei x_1, x_2, $\ldots$, x_n eine Liste reeller Zahlen:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

$$\Rightarrow \quad n \cdot \overline{x} = x_1 + x_2 + \cdots + x_n \quad \Rightarrow \quad \sum_{i=1}^{n} (x_i - \overline{x}) = 0$$

Von keinem Wert ist die Summe der Abstandsquadrate der x_i geringer als vom arithmetischen Mittel, d.h.:

$$\sum_{i=1}^{n} (x_i - x)^2 \text{ hat an der Stelle } x = \overline{x} \text{ ein Minimum.}$$

Das arithmetische Mittel und Häufigkeit

Das arithmetische Mittel $\overline{x}$ kann nicht nur direkt aus den beobachteten Objekten sondern auch indirekt aus der Häufigkeit der Merkmalsausprägungen ermittelt werden:

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} H_i \cdot a_i \qquad \text{bzw.} \qquad \overline{x} = \sum_{i=1}^{k} h_i \cdot a_i$$

n	Anzahl der beobachteten Objekte
k	Anzahl der verschiedenen Merkmalsausprägungen
a_i	Zahlenmässige Werte der Merkmalsausprägungen
H_i	absolute Häufigkeit
h_i	relative Häufigkeit

Das gewogene arithmetische Mittel

Das gewogene arithmetische Mittel ist ein Verfahren um ein arithmetisches Mittel aus bereits gemittelten Werten zu errechnen:

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot \overline{x}_i \quad \text{mit } n = \sum_{i=1}^{k} n_i$$

k	Anzahl der bereits gemittelten Werte
x_i	Arithmetisches Mittel einer Liste mit n_i Elementen
n_i	Anzahl der beobachteten Objekte im Wert $\overline{x}_i$

Das geometrische Mittel und das harmonische Mittel

Für Größen die multiplikativ miteinander verknüpft werden (z.B. Wachstumsraten wie Zinsen) wird als Zentralmass das geometrische Mittel $\hat{x}$ verwendet:

$$\hat{x} = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n}$$

Für Größen die indirekt Proportional zu anderen Größen sind (z.B. Geschwindigkeit zur Zeit) wird als Zentralmass das harmonische Mittel $\tilde{x}$ verwendet:

$$\tilde{x} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_1} + \cdots + \frac{1}{x_n}}$$

Median und Modus

In einer Liste von Variablenwerten bezeichnet man den wert mit der höchsten Häufigkeit als **Modus** (oder **Modalwert**).

Ordnet man eine Liste $x_1, x_2, \ldots, x_n$ der Größe nach, so heisst der in der Mitte stehende Wert m **Median**. Bei einer geraden Anzahl von Werten wählt man das arithmetische Mittel der beiden in der Mitte stehenden Werte.

Der Median ist jenes Zentralmass, bei dem die Summe der Beträge der Abweichungen am kleinsten ist, d.h.:

$$\sum_{i=1}^{n} |x_i - x| \text{ hat an der Stelle } x = m \text{ ein Minimum.}$$

Statistische Skalen

Nominaldaten: Die Variablenwerte (Begriffe oder über Zahlen codiert). Drücken lediglich eine Verschiedenartigkeit aus. z.B.: Namen, Geschlecht, Haarfarbe, usw.

Ordinaldaten: Die Variablenwerte bringen neben der Verschiedenartigkeit auch eine Rangordnung zum Ausdruck. z.B.: Güteklassen, Schulnoten u.Ä.

Metrische Daten: Die Variablenwerte lassen sich nicht nur ordnen, sondern es lassen sich auch Abstände zwischen Werten angeben und sinnvoll interpretieren. z.B.: Jahreszahlen, Längen, Mengenangaben, usw.

Spannweite und Quartile

Bei einer Liste $x_1, x_2, \ldots, x_n$ nennt man die Differenz zwischen dem grössten Wert $x_{\max}$ und dem kleinsten Wert $x_{\min}$ die Spannweite.

Der Wert des Elements aus der Liste der den unteren $a$ten Teil der Elemente vom oberen $(1-a)$ten Teil der Elemente trennt wird x_a genannt. So ist z.B. $x_{0,50}$ der Median der Liste.

Den Wert $x_{0,25}$ nennt man auch unteres Quartil Q_1.
Den Wert $x_{0,50}$ nennt man auch mittleres Quartil Q_2 (oder Median).
Den Wert $x_{0,75}$ nennt man auch oberes Quartil Q_3.

Die Differenz $Q_3 - Q_1$ nennt man Interquartil-Spannweite. Sie umfasst (annähernd) die mittlere Hälfte der Daten.

Mittlere absolute und quadratische Abweichung und die empirische Standardabweichung

Sei $x_1, x_2, \ldots, x_n$ eine Liste und z ein Zentralmass:

Mittlere absolute Abweichung $s^\star$ (minimal beim Median):	$s^\star = \frac{1}{n} \cdot \sum_{i=1}^{n} \lvert x_i - z \rvert$
Mittlere quadratische Abweichung oder empirische Varianz s^2 (minimal beim arithmetischen Mittel):	$s^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - z)^2$
Empirische Standardabweichung s:	$s = \sqrt{s^2} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - z)^2}$

Berechnung der empirischen Standardabweichung

Sei $x_1, x_2, \ldots, x_n$ eine Liste, $\overline{x}$ das arithmetische Mittel, a_i die auftretenden Werte, H_i die absolute Häufigkeit der Werte, h_i die relative Häufigkeit der Werte und s die empirische Standardabweichung:

$$s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2} = \sqrt{\frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i^2 \right) - \overline{x}^2}$$

$$s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{k} H_i \cdot (a_i - \overline{x})^2} = \sqrt{\frac{1}{n} \cdot \left(\sum_{i=1}^{k} H_i \cdot a_i^2 \right) - \overline{x}^2}$$

$$s = \sqrt{\sum_{i=1}^{k} h_i \cdot (a_i - \overline{x})^2} = \sqrt{\left(\sum_{i=1}^{k} h_i \cdot a_i^2 \right) - \overline{x}^2}$$

Berechnung der mittleren absoluten Abweichung

Sei $x_1, x_2, \ldots, x_n$ eine Liste, z ein Zentralmass, a_i die auftretenden Werte, H_i die absolute Häufigkeit der Werte, h_i die relative Häufigkeit der Werte und $s^\star$ die mittlere Abweichung von z:

$$s^\star = \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - z| \implies$$

$$s^\star \;=\; \frac{1}{n} \cdot \sum_{i=1}^{k} H_i \cdot |a_i - z| \;=\; \sum_{i=1}^{k} h_i \cdot |a_i - z|$$

Vergleich von Streuungen

Sei $\overline{x}$ das arithmetische Mittel einer Liste mit der Standardabweichung s und der mittleren absoluten Abweichung $s^\star$ vom Zentralmass z:

$$V = \frac{s}{\overline{x}} \quad = \; \text{Variationskoeffizient}$$

$$v = \frac{s^\star}{z} \; = \; \text{Variabilitätskoeffizient}$$

Variationskoeffizient und Variabilitätskoeffizient sind dimensionslos und werden meist in Prozent angegeben.

Wahrscheinlichkeit und relativer Anteil

Sei $\mathbb{G}$ die endliche Grundgesamtheit mit g Elementen und $\mathbb{A} \subset \mathbb{G}$ jene Teilmenge von $\mathbb{G}$ mit a Elementen die eine bestimmte Eigenschaft aufweisen.

$$h(\mathbb{A}) = \frac{a}{g} \qquad = \text{der } \textbf{relative Anteil} \text{ von } \mathbb{A} \text{ in } \mathbb{G}$$

Wird nun ein Element aus $\mathbb{G}$ zufällig ausgewählt und das Ereignis $\mathbf{E}$ bezeichnet dabei die zufällige Auswahl eines Elements aus $\mathbb{A}$:

$$P(\mathbf{E}) = h(\mathbb{A}) \qquad = \text{die } \textbf{Wahrscheinlichkeit} \text{ von } \mathbf{E}$$

Wertebereich von Wahrscheinlichkeiten

Sei $\mathbf{E}$ ein Ereignis und $P(\mathbf{E})$ die Wahrscheinlichkeit des eintretens:

$$0 \;\leq\; P(\mathbf{E}) \;\leq\; 1$$

$$P(\mathbf{E}) + P(\neg\mathbf{E}) = 1 \quad\Longrightarrow\quad P(\neg\mathbf{E}) = 1 - P(\mathbf{E})$$

Additionsregel für einander ausschließende Ereignisse:

$$P(\mathbf{E}_1 \vee \mathbf{E}_2) = P(\mathbf{E}_1) + P(\mathbf{E}_2)$$

Bedingte Wahrscheinlichkeiten

Seien $\mathbf{E}_1$ und $\mathbf{E}_2$ Ereignisse:

$\mathbf{E}_2$ unter der Annahme das $\mathbf{E}_1$ bereits eingetreten ist:

$$\mathbf{E}_2|\mathbf{E}_1$$

Die Wahrscheinlichkeit vom Eintreten von $\mathbf{E}_2$ unter der Annahme das $\mathbf{E}_1$ bereits eingetreten ist:

$$P(\mathbf{E}_2|\mathbf{E}_1)$$

Wenn $\mathbf{E}_1$ und $\mathbf{E}_2$ voneinander unabhängig sind:

$$P(\mathbf{E}_2|\mathbf{E}_1) = P(\mathbf{E}_2|\neg\mathbf{E}_1) = P(\mathbf{E}_2)$$

Multiplikationsregel für Wahrscheinlichkeiten

Seien $\mathbf{E}_1$ und $\mathbf{E}_2$ Ereignisse:

$$P(\mathbf{E}_1 \wedge \mathbf{E}_2) \quad = \quad P(\mathbf{E}_1) \cdot P(\mathbf{E}_2|\mathbf{E}_1)$$

Bei unabhängigen Ereginissen: $P(\mathbf{E}_1 \wedge \mathbf{E}_2) \quad = \quad P(\mathbf{E}_1) \cdot P(\mathbf{E}_2)$

Additionsregel für Wahrscheinlichkeiten

Seien $\mathbf{E}_1$ und $\mathbf{E}_2$ Ereignisse:

$$P(\mathbf{E}_1 \vee \mathbf{E}_2) \quad = \quad P(\mathbf{E}_1) + P(\mathbf{E}_2) - P(\mathbf{E}_1 \wedge \mathbf{E}_2)$$

Für einander
ausschliessende Ereginisse: $P(\mathbf{E}_1 \vee \mathbf{E}_2) \quad = \quad P(\mathbf{E}_1) + P(\mathbf{E}_2)$

Empirisches Gesetz der großen Zahlen

Die relative Häufigkeit $h(\mathbf{E})$ eines Ereignisses $\mathbf{E}$ “stabilisiert” sich mit zunehmender Versuchsanzahl um einen festen Wert.

Wenn $\mathbf{E}$ bei k von n Versuchen eintritt, dann ist:

$$h(\mathbf{E}) = \frac{k}{n}, \qquad h(\mathbf{E}) \approx P(\mathbf{E})$$

Definition: Zufallsgröße und Wahrscheinlichkeitsverteilung

Eine Funktion, die jedem möglichen Ausgang eines Zufallsexperiments einen zahlenmäßigen Wert zuordnet, nennt man **Zufallsgröße**.

Eine Funktion, die jedem möglichen Wert a_i der (diskreten) Zufallsgröße B eine Wahrscheinlichkeit $P(B = a_i)$ zuordnet, nennt man **Wahrscheinlichkeitsverteilung von** B.

Definition: Fakulät

Die Fakultät von n, geschrieben als $n!$, gesprochen als "n Faktorielle", ist das Produkt aller Zahlen von 1 bis n:

$$n! \;=\; 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n \;=\; \prod_{i=1}^{n} i$$

Spezialfall $n = 0$:

$$0! \;=\; 1$$

Anzahl der Permutationen (mögliche voneinander verschiedene Anordnungen der Elemente) einer Menge mit n Elementen: $n!$

Permutationen von k aus n

Permutationen von k aus n = die Anzahl nPr(n, k) aller Möglichen Anordnungen von k verschiedenen Elementen aus einer Menge mit n Elementen.

$$\mathrm{nPr}(n, k) \;=\; \frac{n!}{(n-k)!} \;=\; n \cdot (n-1) \cdot \ldots \cdot (n-k+1) \;=\; \prod_{i=n-k+1}^{n} i$$

Kombinationen von k aus n

Kombinationen von k aus n = die Anzahl $\mathrm{nCr}(n,k)$ aller Möglichen Auswahlen von k verschiedenen Elementen aus einer Menge mit n Elementen, unabhänig von der Reihenfolge der Elemente.

$$\mathrm{nCr}(n,k) \quad = \quad \binom{n}{k} \quad = \quad \frac{n!}{k! \cdot (n-k)!} \quad = \quad \mathrm{nPr}(n,k) \cdot \frac{1}{k!}$$

Für $\binom{n}{k}$ sagt man "k aus n" oder auch "n über k".
Die Funktion $\binom{n}{k}$ wird Binomialkoeffizient genannt.

Binomialverteilte Zufallsgrößen

Sei **E** ein Ereignis mit der Wahrscheinlichkeit p in einem Versuch der n mal durchgeführt wird und B die Anzahl der Male in denen **E** eintritt, dann ist die Wahrscheinlichkeit das B gleich der Zahl k ist:

$$P(B = k) = \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k}$$

Ordnet man jedem möglichen Wert k der Zufallsgröße B eine Wahrscheinlichkeit $P(B = k)$ zu, so ist dadurch eine Wahrscheinlichkeitsverteilung gegeben, die man **Binomialverteilung** mit den Parametern n und p ("n-p-binomialverteilung") nennt; entsprechend nennt man B eine n-p-binomialverteilte Zufallsgröße.

Definition: γ-Schätzbereich

Man nennt das Intervall, in dem die (relativen) Häufigkeiten der Stichprobe mit der Wahrscheinlichkeit γ liegen, den γ-**Schätzbereich** für (relative) Häufigkeiten.

Erwartungswert und Varianz von diskreten Zufallsgrößen

Sei Z eine diskrete Zufallsgröße mit den Werten $a_1, a_2, \ldots, a_k$, die mit den Wahrscheinlichkeiten $p_1, p_2, \ldots, p_k$ angenommen werden.

$$\mu = \sum_{i=1}^{k} a_i \cdot p_i \quad = \text{Erwartungswert von } Z$$

$$\sigma^2 = \left(\sum_{i=1}^{k} a_i^2 \cdot p_i \right) - \mu^2 \quad = \text{Varianz von } Z$$

$$\sigma \quad = \text{Standardabweichung von } Z$$

Erwartungswert und Varianz von binomialverteilten Zufallsgrößen

Sei B eine binomialverteilte Zufallsgröße mit den Parametern n und p.

$$\mu = n \cdot p \quad = \text{Erwartungswert von } B$$

$$\sigma^2 = n \cdot p \cdot (1 - p) \quad = \text{Varianz von } B$$

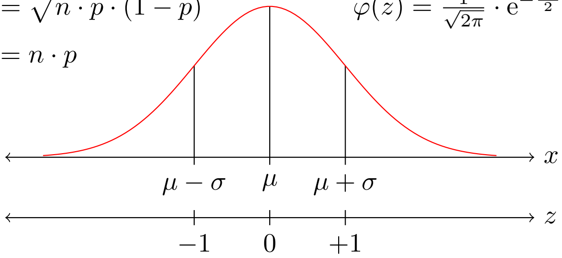
$$\sigma \quad = \text{Standardabweichung von } B$$

Die Gaußsche Glockenkurve

$$\sigma = \sqrt{n \cdot p \cdot (1-p)}$$

$$\mu = n \cdot p$$

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{z^2}{2}}$$



$$\Phi(z_0) = \int_{-\infty}^{z_0} \varphi(z)\mathrm{d}z \qquad z = \frac{x-\mu}{\sigma}$$

Normalverteilte Zufallsgrößen

Eine kontinuierliche Zufallsgröße X heißt **normalverteilt** mit den Parametern μ und σ (μ-σ-normalverteilt), wenn die Wahrscheinlichkeitsverteilung durch die Gaußsche Glockenkurve beschrieben wird. Für eine μ-σ-normalverteilte Zufallsgröße X gilt:

$$P(X \leq \mu + z_0 \cdot \sigma) = \Phi(z_0) = \int_{-\infty}^{z_0} \varphi(z)\mathrm{d}z$$

$$P(\mu + z_1 \cdot \sigma \leq X \leq \mu + z_2 \cdot \sigma) = \Phi(z_2) - \Phi(z_1)$$

$$P(\mu - z_0 \cdot \sigma \leq X \leq \mu + z_0 \cdot \sigma) = 2\Phi(z_0) - 1$$

Laplace-Bedingung

Sei B eine n-p-binomialverteilte Zufallsgröße (mit Erwartungswert μ und Standardabweichung σ), so kann die Gaußsche Glockenkurve als Wahrscheinlichkeitsverteilung angenommen werden, wenn die Laplace-Bedingung

$$n \cdot p \cdot (1 - p) \geq 9 \quad \iff \quad \sigma \geq 3$$

erfüllt ist. Es gilt in diesen Fällen also:

$$P(B \leq \mu + z_0 \cdot \sigma) \approx \Phi(z_0)$$

γ-Schätzbereich für normalverteilte Zufallsgrößen

Zur Ermittlung des γ-Schätzbereiches für eine (näherungsweise) normalverteilte Zufallsgröße muss zunächst ein z_0 ermitteln werden, so dass

$$\gamma = 2\Phi(z_0) - 1 \;\Rightarrow\; \frac{\gamma + 1}{2} = \Phi(z_0) \quad \text{ist.}$$

Dann ist der γ-Schätzbereich für die absolute Häufigkeit H:

$$[\mu - z_0 \cdot \sigma;\;\; \mu + z_0 \cdot \sigma]$$

Und der γ-Schätzbereich für die relative Häufigkeit h:

$$\left[\frac{\mu - z_0 \cdot \sigma}{n};\;\; \frac{\mu + z_0 \cdot \sigma}{n}\right] \;=\; \left[p - z_0 \cdot \sqrt{\frac{p \cdot (1 - p)}{n}};\;\; p + z_0 \cdot \sqrt{\frac{p \cdot (1 - p)}{n}}\right]$$

Testen von Hypothesen über relative Anteile

Gegeben ist eine Hypothese über die relative Häufigkeit p einer Merkmalsausprägung in einer Grundgesamtheit. Diese Hypothese soll anhand einer Stichprobe getestet werden.

Dazu wird Anhand der der Stichprobengröße, der vermuteten relativen Häufigkeit (= Wahrscheinlichkeit des Auftretens der Merkmalsausprägung) und der gewünschten Sicherheit (bzw. Irrtumswahrscheinlichkeit) ein γ-Schätzbereich ermittelt.

Liegt die vermuteten Häufigkeit innerhalb des γ-Schätzbereichs so ist die Vermutung mit dem Stichprobenergebnis (mit der gewählten Irrtumswahrscheinlichkeit) vereinbahr. Andernfalls ist die Vermutung mit dem Stichprobenergebnis (mit der gewählten Irrtumswahrscheinlichkeit) unvereinbahr.

Konfidenzintervall

Sei h die beobachtete relative Häufigkeit einer Merkmalsausprägung in einer Stichprobe mit n Elementen und γ eine (grosse) Wahrscheinlichkeit.

Die Menge aller p deren, deren γ-Schätzbereich den Wert h enthält, nennt man **γ-Konfidenzintervall** (γ-Vertrauensintervall) für p:

$$h - z_0\sqrt{\frac{h \cdot (1-h)}{n}} \quad \leq \quad p \quad \leq \quad h + z_0\sqrt{\frac{h \cdot (1-h)}{n}}$$

$$\text{mit} \quad 2\Phi(z_0) - 1 \quad \approx \quad \gamma$$

Erforderlicher Stichprobenumfang

Sei ϵ die vorgegebene halbe Länge des gesuchten Konfidenzintervalls, h die erwartete relative Häufigkeit der zu untersuchenden Merkmalsausprägung, $\gamma = 2\Phi(z_0) - 1$ die geforderte Sicherheit und n die Anzahl der benötigten Elemente in der Stichprobe:

$$n = \frac{z_0^2}{\epsilon^2} \cdot h \cdot (1 - h)$$

Bzw. die maximal benötigte Stichprobengrösse wenn es keine Abschätzungen für h gibt:

$$n = \frac{z_0^2}{\epsilon^2} \cdot \left(\frac{1}{2}\right)^2 = \frac{z_0^2}{4 \cdot \epsilon^2}$$