

Stochastik für die Naturwissenschaften

Dr. C.J. Luchsinger

Hip hip hooray - Jetzt beginnt die eigentliche Statistik!

8. Schätzen von Parametern und Konfidenzintervalle

Literatur Kapitel 8

- * Statistik in Cartoons: Kapitel 6 und Kapitel 7
- * Stahel: Kapitel 7 und 9
- * Storrer: Kapitel 42 und 43

8.0 Grundbegriffe der beurteilenden Statistik

Kapitel 42 in Storrer wird *in der Vorlesung* übersprungen. Sie sollten es sicher zwei mal lesen: einmal nach dieser Doppelstunde, dann am Schluss der Vorlesung nochmals. Lesen Sie dazu auch in diesem Skript 5.3 nochmals durch. Kurz zusammengefasst (Storrer Kapitel 42):

Statistik:

aus *Stichprobe* Rückschlüsse auf Gesamtheit ziehen

(Stichprobe, weil zu wenig Zeit und Geld, n endlich und gewöhnlich "klein")

- * Küken > Durchschnittsgewicht
- * neue Kartoffelsorte > besser als alte? Unten überlappen leider die Daten, womit der Entscheid nicht so klar ist:

Alte Sorte	410	420	430	440	450	450	480
Neue Sorte	440	450	455	480	490	505	

- * Münze > fair?
- * Anzahl Zerfälle > poissonverteilt? Vgl Schluss Kapitel 7. Statistische Überprüfung einer einleuchtenden Theorie (Mathematik und Physik) anhand von empirischen Daten (Kapitel 9: χ^2 -Anpassungstest).

Im ersten Beispiel wird die Zufallsgrösse (zufällig ist die Auswahl des Kükens) X das Gewicht sein und $E[X]$ das durchschnittliche Gewicht aller Küken.

Wir haben normalerweise nicht nur *ein* Küken in der Stichprobe, sondern allgemein n , zum Beispiel $n = 50$. Wie ist das zu verstehen? Die Daten sind dann (Studis nach Feldversuch)

$$x_1, x_2, \dots, x_{50}. \quad (8.1)$$

Das zugrunde liegende Experiment (Auswahl eines Kükens) wird hierzu 50 mal (unabhängig) wiederholt (mit "Zurücklegen", siehe Storrer Kapitel 42). Wir haben damit eine Folge von Zufallsgrössen (bei Luchs im Büro)

$$X_1, X_2, \dots, X_{50}, \quad (8.2)$$

welche im konkreten Fall (8.1) hervorgebracht haben. Diese Folge in (8.2) wird mit iid Zufallsgrössen modelliert - ausser wir vereinbaren etwas anderes.

Lesen Sie dazu vor allem Seite 202 in Storrer aufmerksam durch - vielleicht werden Ihnen damit Kapitel 5 und 7 eher klar.

8.1 Schätzen von Parametern (Estimation of parameters)

8.1.1 Schätzen des Erwartungswerts (μ ist ein Parameter!)

Gegeben sei eine Stichprobe vom Umfang 9. Der Mittelwert der Grundgesamtheit ist leider unbekannt. Wir wollen diesen Mittelwert schätzen. Die Stichprobe sieht folgendermassen aus (bereits der Reihe nach geordnet und auf 2 Stellen nach dem Komma gerundet):

-0.21, 0.11, 0.39, 0.64, 1.24, 1.46, 1.51, 2.89, 3.53

Wie könnten wir μ schätzen?

Wir bezeichnen einen Schätzer für einen Parameter μ mit $\hat{\mu}$ (für σ^2 mit $\hat{\sigma}^2$). Dabei ist - wegen der akademischen Forschungsfreiheit ! - ein Schätzer eine *beliebige* ! Funktion der Daten (siehe oben). Man kann die Daten auch vollständig ignorieren ("Rapport beim CEO").

Welchen Vorschlag sollten wir warum auswählen?

MathematikerInnen haben mit gutem Grund das arithmetische Mittel

$$\frac{1}{n} \sum_{i=1}^n X_i$$

als *den* Schätzer für den Mittelwert identifiziert.

Warum und Einwände:

1. Einer von vielen Gründen ist die Erwartungstreue: Wie in Kapitel 5 berechnet, gilt nämlich $E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \mu$. Also wird man im Mittel (wenn wir immer wieder erneut mit neuen Daten mit dem arithmetischen Mittel den Erwartungswert berechnen) den Erwartungswert genau treffen.
2. Wenn wir zudem immer mehr Daten erhalten, also das n erhöhen, wird wegen des LLN der Schätzwert immer näher an den Erwartungswert kommen (siehe Kapitel 7.1).
3. Leider hat das arithmetische Mittel einen grossen Nachteil: die starke Abhängigkeit von Ausreissern, also extremen Werten. Wenn auch nur ein Datenwert völlig falsch ist (Übermittlungs- oder Ablesefehler), dann wird das arithmetische Mittel eventuell auch total falsch. Der Median ist robuster.

Das heisst für die Praxis:

8.1.2 Schätzen der Varianz

Wir haben bereits in der beschreibenden Statistik den folgenden Schätzer für die Varianz kennengelernt:

$$\frac{1}{(n-1)} \sum_{j=1}^n (x_j - \bar{x})^2. \quad (\text{kleine } x_j \text{ da Daten aus Feldversuch})$$

Warum wird vorgeschlagen, durch $(n-1)$ zu teilen? Die Antwort findet sich in (43.6), wobei die konkreten Rechnungen analog 8.1.1 - aber viel komplizierter - sind. Es geht wieder um die Erwartungstreue:

$$E \left[\frac{1}{(n-1)} \sum_{j=1}^n (X_j - \bar{X})^2 \right] = \sigma^2. \quad (\text{grosse } X_j \text{ da bei Luchs})$$

d.h. insbesondere, dass $\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2$ *nicht* erwartungstreu ist!

8.1.3 Durchschnitt als Zufallsgrösse, der Standardfehler

Wir sind bis jetzt (vgl Teil 8.0 und 5.3) locker zwischen kleinen x und grossen X hin und her gesprungen.

$$\frac{1}{n} \sum_{i=1}^n x_i$$

ist eine Zahl. Aber bei anderer Auswahl würde diese Zahl wohl leicht anders aussehen.

Wenn wir den Ausdruck

$$\frac{1}{n} \sum_{i=1}^n X_i$$

untersuchen, erfahren wir etwas über den erwarteten Wert des arithmetischen Mittels und über die Varianz des arithmetischen Mittels. Wir repetieren nicht zum ersten Mal von Kapitel 5: $E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \mu$ und $V[\bar{X}] = V\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sigma^2$ und damit

$$sd[\bar{X}] = \frac{\sigma}{\sqrt{n}}.$$

Damit können wir interessante Schlussfolgerungen ziehen:

1. [schwache Voraussetzungen - muss nicht \mathcal{N} sein] Wenn wir eine Folge von n nicht unbedingt normalverteilten - iid Zufallsgrößen X_1, \dots, X_n haben mit Standardabweichung σ , dann hat das arithmetische Mittel (als Zufallsgröße) eine Standardabweichung von

$$\sigma_{\bar{X}} := sd[\bar{X}] = \frac{\sigma}{\sqrt{n}}. \quad (\text{sog. Standardfehler (engl standard error)})$$

Dieser kombinierte Ausdruck kommt häufig vor und hat deshalb einen eigenen Namen! Das σ ist normalerweise nicht bekannt und wir werden es mit

$$s := \sqrt{\frac{1}{(n-1)} \sum_{j=1}^n (x_j - \bar{x})^2}$$

schätzen. Zusammen erhalten wir als *Schätzung* des Standardfehlers folgenden Ausdruck:

$$s_{\bar{X}} := \frac{s}{\sqrt{n}} = \sqrt{\frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n(n-1)}}.$$

In der Praxis wird bereits diese *Schätzung* des Standardfehlers häufig kurz als Standardfehler bezeichnet. Beispiel Storrer p 207 unten schauen Sie als Hausaufgabe an.

Warnung und Konsequenzen: Verwechseln Sie nicht Standardfehler ($sd[\bar{X}] = \sigma/\sqrt{n}$) und die Standardabweichung ($sd[X] = \sigma$) selber: Im Standardfehler wird durch \sqrt{n} geteilt und damit wird er immer kleiner, je grösser wir das n (=Stichprobengröße) machen - wegen des LLN. Dies sollte ja auch so sein: je mehr Daten wir haben, desto präziser können wir schätzen. Die Präzision ist aber derart, dass man zum Beispiel vier mal mehr Daten braucht, um eine doppelte Genauigkeit (Standardfehler halbieren) zu erhalten.

Warum schaut man diesen kombinierten Ausdruck überhaupt an? Nachfolgend ein paar Begründungen aus verschiedenen, ergänzenden Blickwinkeln, zum Teil doppelspurig, aus der Praxis. Wir benutzen $2 \doteq 1.96$:

$$Z \sim \mathcal{N}(0, 1) \Rightarrow 95 \% \text{ der Daten } z_1, \dots, z_n \text{ in } [-2, 2].$$

$$X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow 95 \% \text{ der Daten } x_1, \dots, x_n \text{ in } [\mu - 2\sigma, \mu + 2\sigma].$$

$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n) \Rightarrow$ wenn man mehrere (!) Realisationen von \bar{X} anschaut: 95 % der Daten (also der $\bar{x}_i : \bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ - jedes \bar{x}_i ist der Durchschnitt von n Versuchen) liegen in $[\mu - 2\sigma/\sqrt{n}, \mu + 2\sigma/\sqrt{n}]$.

Wieder auf der Ebene von Zufallsgrössen (Theorie): in 95 % der Fälle liegt \bar{X} in

$$\left[\mu - 2\frac{\sigma}{\sqrt{n}}, \mu + 2\frac{\sigma}{\sqrt{n}}\right] = [\mu - 2 \cdot \text{Standardfehler}, \mu + 2 \cdot \text{Standardfehler}]$$

Ich betreue 20 Studis mit ihren Arbeiten und weiss, dass im Schnitt bei einem von 20 (5 %) das \bar{x} ausserhalb dieser Bandbreite liegen wird - Sie als Studi wissen aber nicht, ob Sie der Pechvogel sind. Sie nehmen Ihr \bar{x} (wegen der investierten Zeit) hoffentlich sehr ernst - für mich ist es einfach einer von 20 Werten, der da um den wahren Mittelwert μ herumhüpft - wenn auch immer enger. Mit ein bisschen grösserer Empathie kann ich Ihnen aber immerhin sagen: Sie erhalten ein \bar{x} und ein $s_{\bar{x}}$. Wenn $s_{\bar{x}}$ klein ist (kleines σ , grosses n), gibt das Sicherheit, dass Ihr \bar{x} nahe bei μ ist, denn mit 95 % Wahrscheinlichkeit gilt ja

$$\bar{X} \in [\mu - 2 \cdot \text{Standardfehler}, \mu + 2 \cdot \text{Standardfehler}]$$

oder anders formuliert:

$$\mu \in [\bar{X} - 2 \cdot \text{Standardfehler}, \bar{X} + 2 \cdot \text{Standardfehler}],$$

denn \bar{X} und μ liegen maximal zwei Standardfehler voneinander entfernt. Mit konkreten Daten operational formuliert: in 95 % der Fälle haben Sie

$$\mu \in [\bar{x} - 2 \cdot s_{\bar{x}}, \bar{x} + 2 \cdot s_{\bar{x}}]$$

und sie wollen ja wissen, wo das wahre μ liegt.

2. [starke Voraussetzungen - muss \mathcal{N} sein - alles Bisherige bleibt richtig]: Von Kapitel 6: Summe von Normal ist Normal (Summanden müssen nicht mal identisch verteilt sein) $\sum_{i=1}^n \mathcal{N}(\mu_i, \sigma_i^2) = \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$. Wenn wir also sogar eine Folge X_1, \dots, X_n von iid $\mathcal{N}(\mu, \sigma^2)$ -Zufallsgrößen haben, so hat

$$\bar{X}$$

eine

$$\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

-Verteilung (Mittelwert μ und Varianz $\frac{\sigma^2}{n}$ galt ja schon vorher). Dies gibt das Bild in Storrer p 208. Weil die Varianz immer kleiner wird, wird die Dichtefunktion immer enger und damit die Schätzung immer präziser. Wir rechnen hierzu noch das Zahlenbeispiel in Storrer p 209 oben.

8.2 Konfidenzintervalle (Intervall-Schätzer)

Man kann ab hier wieder neu einsteigen, falls man abgehängt wurde. Es lohnt sich, denn dies sind einfache Prüfungsaufgaben. Das einzige, was man wissen muss, ist, dass bei normalverteilten Daten gilt:

$$\bar{X} \sim$$

Eigentlich interessieren wir uns für einen unbekanntem Parameter, den wir schätzen wollen. Konfidenzintervalle geben uns eine Vorstellung von der Präzision eines Schätzers / einer Schätzung:

Ziele von 8.2: Die StudentInnen wissen, was ein Konfidenzintervall (KI) ist - und was es *nicht* ist. Sie erkennen Formeln für KI in 4 wichtigen Fällen wieder und können diese anwenden.

8.2.1 Was ist ein Konfidenzintervall (KI) - was ist es *nicht*

Was ist es *nicht* - drei Beispiele: 1. Sozialwissenschaften: In den Nachrichten liest man oft Sätze der Art (Zahlen frei erfunden): "Aufgrund einer Befragung mit einer Stichprobe von 10'000 Personen kam man zum Schluss, dass der Anteil der Anhänger von Bundeskanzlerin Merkel mit 95 % Wahrscheinlichkeit in einem Konfidenzintervall von [46%, 48%] liegt." Was ist hier falsch?

2. Physik:

3. Stochastik:

Definition 8.1 [Konfidenzintervalle] Ein Konfidenzintervall KI für μ mit Konfidenzoeffizient $(1 - \alpha)$ ist eine zufällige Teilmenge von \mathbb{R} mit der Eigenschaft, dass

$$P[\mu \in KI] = 1 - \alpha$$

für alle $\mu \in \mathbb{R}$ (siehe auch Storrer p 217).

Bemerkungen zu Definition 8.1: Im Teil "was ist es *nicht*" haben wir bereits betont, dass μ nicht zufällig ist. Das Konfidenzintervall KI ist zufällig (*vor* der Realisation). Wenn danach die Realisation mit konkreten Zahlen vorliegt, sollte man Formulierungen brauchen wie: "[46%, 48%] ist eine Realisation eines 95 % Konfidenzintervalles". Dies machen nur MathematikerInnen; Sie dürfen sagen:

"[46%, 48%] ist ein 95 % Konfidenzintervall".

8.2.2 Konstruktion von KI's

Jetzt geht es noch um die Frage, wie wir es hinkriegen, dass das Konfidenzintervall (mit grossen X_i) genau die richtige Grösse hat.

8.2.2.1 Daten aus $\mathcal{N}(\mu, \sigma^2)$: KI für μ wenn σ^2 bekannt (selten - aber gut für Aufbau)

Daten x_1, \dots, x_n , berechnen arithmetisches Mittel \bar{x} . Bei Luchs im Büro; vor Versuch: Wechseln zu Zufallsgrössen: betrachten \bar{X} . Verteilung von \bar{X} ist:

Also ist

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

wegen der "Z-Transform" eine $\mathcal{N}(0, 1)$ -Zufallsgrösse. Aber das ist ja fantastisch: wir kennen die kritischen Werte bei einer $\mathcal{N}(0, 1)$ -Verteilung:

$$P\left[-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq 1.96\right] = 95\%. \quad (8.3)$$

Ziel ist offensichtlich ein $(1 - \alpha) = 95\%$ -KI! Aber: wir können offensichtlich (8.3) (noch) nicht als KI verkaufen. Machen wir ein paar algebraische Umformungen (Resultat ist dann (8.8)):

$$P\left[-1.96\sqrt{\frac{\sigma^2}{n}} \leq \bar{X} - \mu \leq 1.96\sqrt{\frac{\sigma^2}{n}}\right] = 95\%. \quad (8.4)$$

Wir addieren μ :

$$P\left[\mu - 1.96\sqrt{\frac{\sigma^2}{n}} \leq \bar{X} \leq \mu + 1.96\sqrt{\frac{\sigma^2}{n}}\right] = 95\%, \quad (8.5)$$

formen es noch um zu

$$P\left[\bar{X} \in \left[\mu - 1.96\sqrt{\frac{\sigma^2}{n}}, \mu + 1.96\sqrt{\frac{\sigma^2}{n}}\right]\right] = 95\%. \quad (8.6)$$

Dies ist aber gleichbedeutend mit

$$P\left[\mu \in \left[\bar{X} - \frac{1.96\sigma}{\sqrt{n}}, \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right]\right] = 95\%. \quad (8.7)$$

Wir vergleichen voller Stolz und Befriedigung (8.7) mit Definition 8.1.

Also gilt (nach Versuch, mit Daten): *ein* 95 % KI für μ bei bekanntem σ^2 und Daten x_1, \dots, x_n in Modell $\mathcal{N}(\mu, \sigma^2)$ ist:

$$\left[\bar{x} - \frac{1.96\sigma}{\sqrt{n}}, \bar{x} + \frac{1.96\sigma}{\sqrt{n}} \right]. \quad (8.8)$$

99 % KI in dieser Situation ist:

Was geschieht wenn σ^2 wächst?

Was geschieht wenn n wächst?

Kleine Aufgabe zu KI I: Es wird angenommen, dass die Durchmesser der auf einer bestimmten Anlage hergestellten Stahlkugeln durch die Realisationen einer normalverteilten Zufallsgrösse mit $\sigma = 1.04$ mm beschrieben werden können. Aus einer Stichprobe vom Umfang $n = 30$ ergab sich $\bar{x} = 12.14$ mm. Bestimmen Sie für die Vertrauenswahrscheinlichkeit von 0.95 und 0.99 die Grenzen des KI für den mittleren Durchmesser dieser Kugeln.

8.2.2.2 Daten aus $\mathcal{N}(\mu, \sigma^2)$: KI für μ wenn σ^2 unbekannt (realistisch)

Daten x_1, \dots, x_n , berechnen das arithmetische Mittel \bar{x} . Bei Luchs im Büro; vor Versuch: Wechseln auf die Ebene der Zufallsgrößen: betrachten \bar{X} . Verteilung von \bar{X} ist:

Also hat wegen der "Z-Transform" der Ausdruck

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

eine $\mathcal{N}(0, 1)$ -Verteilung. Oooooops? σ^2 ist unbekannt! Keine Panik: Wir werden es einfach schätzen:

$$\hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Das Tier

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{\bar{X} - \mu}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} \quad (8.9)$$

hat eine t_{n-1} -Verteilung! Die t -Verteilung ist "fast" eine Normalverteilung. Das "fast" kommt daher, dass wir im Nenner von (8.9) nicht das (eben unbekannt) σ einsetzen können, sondern nur eine Schätzung hiervon. Die Schätzung von σ bedeutet mehr Unsicherheit, womit die t -Verteilungen mehr Gewicht in den Enden haben. In Storrer (43.7) haben Sie hierzu mehr Informationen. In meinem Skript ist in 6.2.6 ebenfalls die t -Verteilung aufgeführt.

Mit den gleichen Schritten wie in 8.2.2.1 folgt dann analog zu (8.8); nach Versuch, mit Daten:

$$\left[\bar{x} - \frac{CV \hat{\sigma}}{\sqrt{n}}, \bar{x} + \frac{CV \hat{\sigma}}{\sqrt{n}} \right], \quad (8.10)$$

wo

$$\hat{\sigma} := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (8.11)$$

Der kritische Wert (Critical Value CV) ist jetzt nicht mehr nur von $(1 - \alpha)$ abhängig (wie in 8.2.2.1), sondern auch von n (vgl. z.B. Tabelle in (51.4)). Zum Beispiel mit $n = 20$ sind die kritischen Werte für ein 95%-KI:

zu vergleichen mit den 1.96 von 8.2.2.1.

Kleine Aufgabe zu KI II: gleiche Ausgangslage wie bei Aufgabe I, aber wir kennen σ nicht und haben es mit (8.11) genau auf 1.12 mm geschätzt (**„geschätzt“ ist ein Signalwort für die Prüfung**: σ unbekannt, also geschätzt: t_n -Verteilung - auch wenn wir mit \mathcal{N} gestartet sind - und zwar mit " $df - 1$ "). Berechnen Sie das 0.95-KI nochmals in dieser Situation.

Was passiert wenn n grösser wird?

Was wenn σ^2 grösser wird?

Storrer A 43-8:

8.3. Die 4 wichtigsten KI aus der Praxis (beyond Storrer)

Orientierungsschema:

In den bisherigen Situationen (8.2.2.1 und 8.2.2.2) stellen wir fest, dass dortige 95 % Konfidenzintervalle approximativ immer nach dem gleichen Schema aufgebaut sind:

$$[\bar{x} - 2\text{sd}(\bar{X}), \bar{x} + 2\text{sd}(\bar{X})]. \quad (\text{generic})$$

Die Zahl "2" ist im Fall der Normalverteilung eigentlich präziser unser bekanntes 1.96. Den Unterschied zwischen 1.96 und 2 ignorieren Praktiker/innen gerne - wir jetzt auch. Wegen des CLT haben wir bei grossen n approximativ in der Summe (\bar{x} beinhaltet eine Summe!) ebenfalls fast eine Normalverteilung (technisch: solange die X_i 's eine endliche Varianz besitzen). Deshalb können wir in den noch folgenden 3 Fällen A, B und C obige generische Formel (generic) benutzen.

A (Bernoulli/Binomial): In HHS (43.9) wird dieser Fall auch behandelt - mit unserer Methode ist es einfacher. Wenn Sie aber mit der Mathematik total auf Kriegsfuss stehen, springen Sie direkt zu Formel (Fall A operational) - give your brain a chance: Um

$$[\bar{x} - 2\text{sd}(\bar{X}), \bar{x} + 2\text{sd}(\bar{X})]. \quad (\text{generic})$$

anzuwenden, müssen wir eine Übersetzungsleistung erbringen, denn bisher ging es bei den KI um normalverteilte Ausgangsdaten x_1, \dots, x_n und nicht um Wahrscheinlichkeiten. Obige Formel kann aber dank CLT auch hier für ein 95 % KI für die wahre Wahrscheinlichkeit p eingesetzt werden: Sei x_i gleich 1 bei Erfolg und gleich 0 bei Misserfolg bei Versuch i . Dann ist $\sum_{i=1}^n x_i$ die gesamte Anzahl Erfolge und damit $\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$ der Anteil der Erfolge (zB bei $n = 100$, $\sum_{i=1}^{100} x_i = 57$, $\bar{x} = 0.57$). Weil wir jetzt ein Konfidenzintervall für die Erfolgswahrscheinlichkeit wollen, ändern wir gleich die Bezeichnung und setzen (denken Sie an die Analogie: Anteil - Proportion - Wahrscheinlichkeit):

$$\hat{p} := \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i.$$

(generic) wird jetzt von

$$[\bar{x} - 2\text{sd}(\bar{X}), \bar{x} + 2\text{sd}(\bar{X})]$$

zu

$$[\hat{p} - 2\text{sd}(\hat{p}), \hat{p} + 2\text{sd}(\hat{p})]$$

Aber wir kennen $\text{sd}(\hat{p})$ nicht! Um das zu berechnen, muss man \hat{p} als Zufallsgrösse auffassen (beides wird mit \hat{p} bezeichnet: $\hat{p} := \bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ und $\hat{p} := \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$ - es gibt keine Verwirrung, zumindest nicht unter Mathematiker/innen :-). Wir haben dann:

$$V[\hat{p}] = V\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} V\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n V[X_i] = \frac{1}{n^2} n V[X_1] = \frac{1}{n} p(1-p).$$

Damit gilt $\text{sd}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$. Leider ist p unbekannt; wir ersetzen es deshalb durch die Schätzung und erhalten $\text{sd}(\hat{p}) \doteq \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Damit erhalten wir für das KI

$$\left[\hat{p} - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]. \quad (\text{Fall A operational})$$

Das ist Formel (4) in HHS, Seite 220. Praktikerregel: $n\hat{p}(1 - \hat{p}) > 9$ und $\hat{p} \in [0.1, 0.9]$, sonst sogenannte Methode nach Wilson (das ist im Storrer) anwenden. Wir rechnen zur Kontrolle obiger Formel gleich mal das Beispiel aus 8.2.1 nach, bei der es um das KI zur Anhängerschaft von Angie ging:

Die Formel sieht für die USA (329 M Einwohner), Schweiz (8550000 Einwohner) und Seldwyla (10013 Einwohner) genau gleich aus!?

B (Bernoulli/Binomial; KI für Differenz bei Proportionen): Wir können jetzt - nach den Vorbereitungsarbeiten in A - schneller vorgehen. Wenn Sie Fragen zu Zwischenschritten haben, bitte nach dem Unterricht melden.

Wir haben jetzt 2 Stichproben (zB Erfolg von 2 Operationsmethoden oder bei Abstimmungen ob Romandie anders stimmen wird als Deutschschweiz). Analog zu A haben wir m Objekte in Gruppe 1 und n Objekte in Gruppe 2. Sei x_i gleich 1 bei Erfolg und gleich 0 bei Misserfolg bei Versuch i in Gruppe 1. Dann ist $\sum_{i=1}^m x_i$ die gesamte Anzahl Erfolge und damit $\bar{x} := \frac{1}{m} \sum_{i=1}^m x_i$ der Anteil der Erfolge. Analog sei y_j gleich 1 bei Erfolg und gleich 0 bei Misserfolg bei Versuch j in Gruppe 2. Dann ist $\sum_{j=1}^n y_j$ die gesamte Anzahl Erfolge und damit $\bar{y} := \frac{1}{n} \sum_{j=1}^n y_j$ der Anteil der Erfolge. Weil wir wieder ein Konfidenzintervall für die Erfolgswahrscheinlichkeiten (jetzt aber für die Differenz ebendieser Wahrscheinlichkeiten) wollen, ändern wir gleich die Bezeichnung und setzen:

$$\hat{p}_1 := \bar{x} := \frac{1}{m} \sum_{i=1}^m x_i \quad , \quad \hat{p}_2 := \bar{y} := \frac{1}{n} \sum_{j=1}^n y_j.$$

Wir wollen ein 95 % KI für die Differenz der Wahrscheinlichkeiten $p_1 - p_2$ konstruieren: von (generic) her haben wir - auf der Ebene von Zufallsgrössen $(\hat{p}_1 - \hat{p}_2) \pm 2\text{sd}(\hat{p}_1 - \hat{p}_2)$. Sie berechnen bitte analog zum Fall A $\text{sd}(\hat{p}_1 - \hat{p}_2)$. Wir erhalten schlussendlich für das KI:

$$\left[(\hat{p}_1 - \hat{p}_2) - 2\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}} \quad , \quad (\hat{p}_1 - \hat{p}_2) + 2\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}} \right].$$

Praktikerrregel: $\hat{p}_i \in [0.1, 0.9]$ und m, n je grösser als 30; sonst googeln.

Achtung für die Prüfungseinsicht, falls wir es übersehen haben: man kann statt $(\hat{p}_1 - \hat{p}_2)$ auch alles mit $(\hat{p}_2 - \hat{p}_1)$ aufziehen - ist gleichwertig und sollte die gleiche Punktzahl geben!

C (2 Stichproben; Normalverteilung; KI für Differenz der Mittelwerte): Wir haben wieder 2 Stichproben (zB Ertrag Düngemittel von 2 verschiedenen Düngern). Analog zu B haben wir m Objekte in Gruppe 1 und n Objekte in Gruppe 2. Sei x_i der Wert bei Versuch i in Gruppe 1 - wir modellieren mit einer $\mathcal{N}(\mu_1, \sigma^2)$ -Zufallsgrösse. Dann ist $\bar{x} := \frac{1}{m} \sum_{i=1}^m x_i$ der Durchschnittswert in Gruppe 1. Analog sei y_j der Wert bei Versuch j in Gruppe 2 - wir modellieren mit einer $\mathcal{N}(\mu_2, \sigma^2)$ -Zufallsgrösse. Dann ist $\bar{y} := \frac{1}{n} \sum_{j=1}^n y_j$ der Durchschnittswert in Gruppe 2. **Die Varianzen der einzelnen Messungen in Gruppe 1 und 2 müssen gleich sein!** Wir wollen ein 95 % KI für die Differenz der Mittelwerte $\mu_1 - \mu_2$ konstruieren: von (generic) her haben wir - auf der Ebene von Zufallsgrössen $(\bar{X} - \bar{Y}) \pm 2\text{sd}(\bar{X} - \bar{Y})$. Analog zu A und B wird auch hier $\text{sd}(\bar{X} - \bar{Y})$ berechnet, wobei für die Schätzung von σ^2 ein hier nicht begründeter ungewöhnlicher Ausdruck verwendet wird: Wir erhalten für das KI zuerst:

$$\left[(\bar{x} - \bar{y}) \pm 2\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \frac{\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{j=1}^n (y_j - \bar{y})^2}{m + n - 2}} \right].$$

Hier wird *in Abweichung des bisherigen Vorgehens* nicht 2 benutzt, sondern der entsprechende Term aus der t_{m+n-2} -Verteilung, da man σ^2 geschätzt hat (**Fehlerquelle Prüfung**).

$$\left[(\bar{x} - \bar{y}) \pm \text{CV} \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \frac{\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{j=1}^n (y_j - \bar{y})^2}{m + n - 2}} \right].$$

Alle 4 Fälle für KI müssen für die Prüfung beherrscht werden. In den Übungen werden wir aber nicht in jedem Semester zu jedem Fall ein Beispiel lösen. Schauen Sie stattdessen die 10 letzten Prüfungen im Archiv an.

8.4. KI: wie gross muss n sein?

Wichtig:

1. Lesen Sie jetzt das komplette Kapitel im Storrer II selber durch (Kapitel 42-43).
2. Lösen Sie danach mindestens 5 Aufgaben hinten im Kapitel und vergleichen Sie mit den Lösungen am Schluss des Buches. Bei Bedarf lösen Sie mehr Aufgaben.
3. Gehen Sie in die Übungsstunde. Drucken Sie das Übungsblatt dazu *vorher* aus, lesen Sie *vorher* die Aufgaben durch und machen sich erste Gedanken dazu (zum Beispiel, wie man sie lösen könnte).
4. Dann lösen Sie das Übungsblatt: zuerst immer selber probieren, falls nicht geht: Tipp von Mitstudi benutzen, falls immer noch nicht geht: Lösung von Mitstudi anschauen, 1 Stunde warten, versuchen, aus dem Kopf heraus wieder zu lösen, falls immer noch nicht geht: Lösung von Mitstudi abschreiben (und verstehen - also sollte man insbesondere keine Fehler abschreiben!).
5. Lösen Sie die entsprechenden Prüfungsaufgaben im Archiv.