Schliessende Statistik (Inferenzstatistik)

1 **Das Schätzproblem**

2 **Vertrauensintervalle**

3 Das Testproblem

Die schliessende Statistik beschäftigt sich mit diesen drei Fragen. Sie hat die Aufgabe, eine Brücke zwischen den konkreten **Daten** einer Stichprobe und der **Theorie** herzustellen. Eine Verbindung ermöglichen die aus der Wahrscheinlichkeitsrechnung bekannten Modelle (Binomialverteilung, Normalverteilung, t-Verteilung,).
In vielen Fällen ist nämlich die Wahrscheinlichkeitsverteilung einer Stichprobe bekannt, sie enthält aber noch unbekannte Parameter. Dies führt zur Frage, wie aus einer konkreten Stichprobe diese Parameter (Wahrscheinlichkeiten, Erwartungswert oder Varianz) geschätzt oder getestet werden können?

1 Das Schätzproblem

1.1 Punktschätzungen

Das Prinzip besteht darin, die unbekannten Parameter mit den entsprechenden Kennzahlen der Daten zu schätzen. Zu Beginn des Kapitels Wahrscheinlichkeitsrechnung wurden schon nach diesem plausiblen Verfahren die Wahrscheinlichkeiten geschätzt (vgl. etwa das Reisnagelbeispiel). Diese Methode heisst:

Momentenmethode

Sind in einer Stichprobe vom Umfang n genau x Ergebnisse für das Ereignis günstig, dann wird die unbekannte Wahrscheinlichkeit p geschätzt durch die relative Häufigkeit
$\hat{p} = \frac{x}{n}$

Beispiel:
Jeder Chip eines bestimmten Herstellers sei mit der Wahrscheinlichkeit p in Ordnung.
Findet man in einer Stichprobe von 1000 Chips 79 defekte, dann ist die plausibelste Schätzung für p gerade die relative Häufigkeit $\frac{1000-79}{1000} = 0.921$.

Binomialverteilung:

Ein Bernoulliversuch mit der Erfolgswahrscheinlichkeit p wir n-mal wiederholt. Für die Wahrscheinlichkeit in n Wiederholungen genau x Erfolge zu erzielen, gilt dann bekanntlich:
$P_n(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$
Die unbekannte Erfolgswahrscheinlichkeit p und die unbekannte Varianz σ^2 werden dann geschätzt durch
$\hat{p} = \frac{x}{n}$ $\hat{\sigma}^2 = \frac{\hat{p}\cdot(1-\hat{p})}{n}$

Beispiel: Mädchengeburt.
Im Jahr 2022 wurden in der Schweiz $n = 82381$ Kinder geboren, davon waren $x = 39907$ Mädchen. Als Schätzwert für eine Mädchengeburt ergibt sich damit
$\hat{p} = \frac{x}{n} = \frac{39907}{82371} \approx 0.484$

Poissonverteilung.

Als Schätzwert für den unbekannten Erwartungswert λ kann das arithmetische Mittel verwendet werden

Als Beispiel sei erwähnt:
«Tote durch Hufschlag» (→ Poissonverteilung, Bortkiewicz 1893)

Anzahl der Kavalleristen der preussischen Armee, die durch Hufschlag getötet wurden. Dazu wurden in 10 Kavalleriekorpsteilen während 20 Jahren die folgenden Zahlen ermittelt:

Anzahl k der Getöteten	Anzahl der Jahre pro Korps und pro Dienstjahr
0	109
1	65
2	22
3	3
4	1
Summe	200

Der Mittelwert $\bar{x} = \frac{1}{200} \cdot (0 \cdot 109 + 1 \cdot 65 + 2 \cdot 22 + 3 \cdot 3 + 4 \cdot 1) = 0.61$
wird als Schätzwert für das unbekannte λ verwendet.

Der Erwartungswert λ bedeutet, dass in den 20 Jahren in den 10 Korps insgesamt 122 Kavalleristen starben, also pro Korps in 2 Jahren ungefähr 1 Kavallerist.

Normalverteilung

Schätzer für Erwartungswert und Standardabweichung einer normalverteilten Gesamtheit

Schätzer für den Erwartungswert: $\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ empirischer Mittelwert
Anderer Schätzer: Median

Schätzer für die Varianz bzw.
Standardabweichung: $\hat{\sigma}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$ empirische Varianz

Dazu ein Beispiel auf der folgenden Seite: Körpergrössen von Jugendlichen:
gesucht sind Punktschätzungen für den Erwartungswert μ und die Standardabweichung σ.

$\hat{\mu} = \bar{x} = 106.25$ $\hat{\sigma} = s = 10.6$

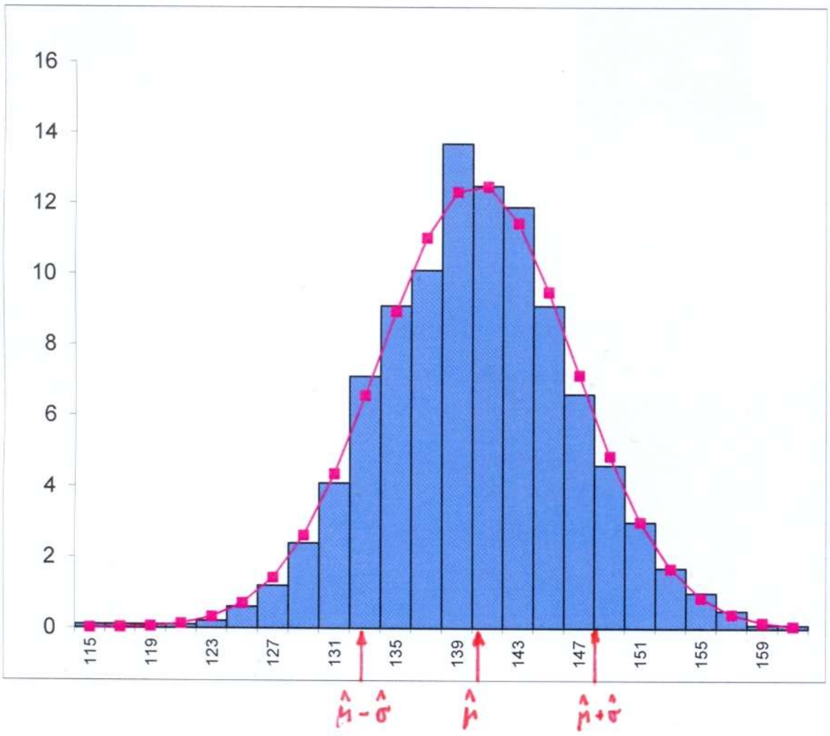
Diese Schätzwerte sind auch die Parameterwerte für die eingezeichnete Normalverteilung.

Übungsaufgabe:
Bei einem Test wurden folgende Zeiten in Millisekunden gemessen:
28.4, 24.2, 29.1, 25.4, 27.6
Lösung:
$\hat{\mu} = \bar{x} = 26.94 \approx 26.9$ $\hat{\sigma} = s \approx 2.1$

Körpergrössen von 4922 Jugendlichen im Alter zwischen 10 und 11 Jahren

Die Daten stammen von Jugendgesundheitsuntersuchungen an Volksschulen Bayerns im Jahre 1965/66

Körpergrösse in cm	rel. Häufigkeit in %	Dichte der Normalverteilung
115	0.1	0.0
117	0.1	0.0
119	0.1	0.0
121	0.1	0.1
123	0.2	0.3
125	0.6	0.7
127	1.2	1.4
129	2.4	2.6
131	4.1	4.4
133	7.1	6.6
135	9.1	8.9
137	10.1	11.0
139	13.7	12.3
141	12.5	12.5
143	11.9	11.4
145	9.1	9.5
147	6.6	7.1
149	4.6	4.9
151	3.0	3.0
153	1.7	1.7
155	1	0.8
157	0.5	0.4
159	0.1	0.2
161	0.1	0.1
		100.0



emp. Mittelwert	$\hat{\mu} =$ **140.246**
emp. Varianz	40.31
emp. Standardabweichung	$\hat{\sigma} =$ **6.35**

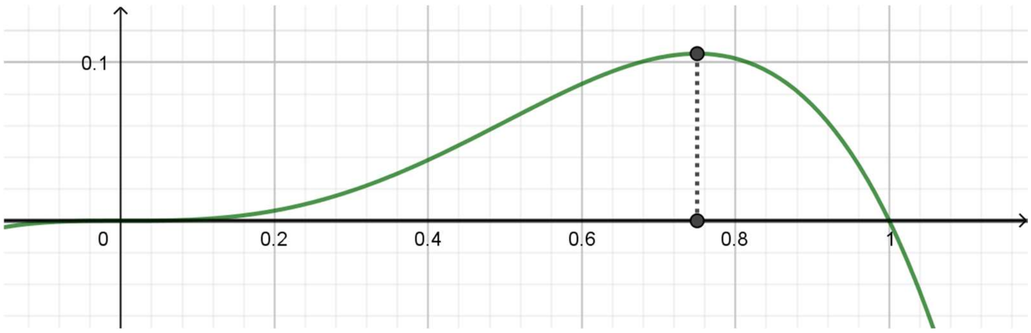
Ein zweites Verfahren ist die sogenannte

Maximum-Likelihood Methode (ML-Methode)

Dazu ein einfaches Beispiel:
Ein Versuch, der mit Wahrscheinlichkeit p misslingt, wird so oft wiederholt, bis er das erste Mal gelingt (sogenannte geometrische Verteilung).
Die Anzahl X der Misserfolge bis zum ersten Erfolg hat dann die Verteilung:
$P(X = x) = (1-p)p^x$
Wie ist p zu schätzen, wenn angenommen wird, dass der Versuch nach 3 Misserfolgen gelungen ist?

Die Idee besteht darin, die Funktion
$P(X = 3) = (1-p) \cdot p^3 = p^3 - p^4$

zu betrachten, und p so zu bestimmen, dass $P(X = 3)$ möglichst gross und damit plausibler wird (Likelihood bedeutet Plausibilität). Es ist damit angebracht, diesen plausibelsten Wert als Schätzwert $\hat{p}$ zu verwenden.



Notwendige Bedingung dafür ist:
$P'(X = 3) = 3p^2 - 4p^3 = p^2 \cdot (3 - 4p) = 0 \qquad p \neq 0, 1$
hat neben der Zahl 0 die positive Lösung $\hat{p} = \frac{3}{4}$.

Der Schätzwert für die Erfolgswahrscheinlichkeit der **Binomialverteilung** ergibt sich auch mit dem MLP:
Der Binomialkoeffizient ist eine Konstante und kann weggelassen werden.
Als Ableitung von $p^x \cdot (1-p)^{n-x}$ nach p erhält man nach einiger Rechnung

$$(p^x \cdot (1-p)^{n-x})' =\cdot \frac{p^x \cdot (1-p)^{n-x} \cdot (np - x)}{p \cdot (1-p)} = 0 \qquad p \neq 0, 1$$

Mit dieser notwendigen Bedingung für ein Maximum erhält man als Lösung $\hat{p} = \frac{x}{n}$

Das in vielen Fällen anwendbare ML-Methode kann verallgemeinert werden. Eine Vereinfachung ergibt sich, wenn man statt der Likelihood den Logarithmus der Likelihood betrachtet:
$(\ln(p^x \cdot (1-p)^{n-x}))' = x \cdot \ln p + (n - x) \cdot \ln(1-p) \qquad p \neq 0, 1$
Als Ableitung nach π erhält man
$x \cdot \frac{1}{p} - \frac{n-x}{1-p} = 0$
mit der Lösung $\hat{p} = \frac{x}{n}$.

Eigenschaften von Schätzungen:

1. Erwartungstreue

Ein Schätzer heisst erwartungstreu, wenn der Erwartungswert des Schätzers mit dem gesuchten Parameter übereinstimmt.

Beispiel:

Wählt man als Schätzer für den Erwartungswert μ einer normalverteilten Zufallsvariablen X den arithmetischen Mittelwert

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

dann ist der Schätzer erwartungstreu (unbiased) d.h. es gilt: $E(\hat{\mu}) = \mu$. Ein solcher Schätzer heisst unbiased.

Das Ergebnis der Stichprobe kann nämlich als Realisation der Zufallsvariablen X aufgefasst werden. Da ihre Wahrscheinlichkeitsverteilung (die Normalverteilung) und damit auch μ und σ^2 bekannt sind, gilt:

$$E(\bar{X}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i\right) = \frac{1}{n} \cdot E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(x_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} \mu = \frac{1}{n} \cdot n \cdot \mu = \mu$$

Für die Varianz ist der folgende Schätzer erwartungstreu:

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-1} \cdot \left(\sum_i X_i^2 - n \cdot \bar{X}^2\right)$$

Bemerkenswert ist, dass im Nenner $n-1$ steht (statt wie erwartet n).
Es gilt nämlich mit den Gesetzen für E(X) und V(X):

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \cdot \left(E(X_i^2) - n \cdot E(\bar{X}^2)\right) = \frac{1}{n-1}\left(n \cdot \sigma^2 - n \cdot \frac{1}{n} \cdot \sigma^2\right) = \frac{n-1}{n-1} \cdot \sigma^2 = \sigma^2$$

Der Beweis wird vereinfacht unter der Annahme, dass $E(X_i) = 0$ also $E(X) = 0$ ist
Damit vereinfacht sich die Berechnung der Varianz:

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

reduziert sich auf

$$V(X) = E(X^2) = \sigma^2$$

Für den **empirischen Standardfehler des Mittelwerts** gilt damit

$$\hat{\sigma}_{\bar{x}}^2 = \frac{1}{n(n-1)} \cdot \left(\sum_i X_i^2 - n \cdot \bar{X}^2 \right)$$

Beispiel:
Aus einer Population von normalverteilten Stichprobengrössen mit μ = **178 cm** und σ = **7 cm** wurden durch Simulation 150 Stichproben vom Umfang n = 100 gezogen und für jede der Mittelwert und die Standardabweichung erhoben.
Die Mittelwerte schwankten zwischen 175.39 und 180 19.
Der Mittelwert der 150 Stichprobenmittelwerte war $\boldsymbol{\hat{\mu} = 178.03\ cm}$
Die Standardabweichung schwankten zwischen 5.86 und 8.15
Der Mittelwert der 180 Stichproben-Standardabweichungen war $\boldsymbol{\hat{\sigma} = 7.00\ cm}$
Die Standardabweichung der Stichprobenmittelwerte, der sogenannte Standardfehler des Mittelwerts beträgt nur **0.68 cm**

Zusammenfassung der Ergebnisse**:**

	Population	Schätzwert aus der Stichprobe
Erwartungswert	**μ = 178 cm**	$\boldsymbol{\hat{\mu} = 178.03\ cm}$
Standardabweichung	**σ = 7 cm**	$\boldsymbol{\hat{\sigma} = 7.00\ cm}$
Standardfehler	$\boldsymbol{\frac{\sigma}{n} = \frac{7}{\sqrt{100}} = 0.7}$ **cm**	$\boldsymbol{\frac{\hat{\sigma}}{n} = \frac{7}{\sqrt{100}} = 0.68}$ **cm**

Fazit:

Betrachtet man statt der einzelnen Stichprobenwerte den Mittelwert, dann wird die Schätzung genauer. Wird also der Stichprobenumfang vervierfacht, so halbiert sich der Standardfehler.

2. Konsistenz

Ist ein Schätzwert konsistent, dann bedeutet dies, dass der Schätzwert sich mit wachsendem Umfang der Stichprobe dem geschätzten Parameter beliebig genau annähert, die Differenz der beiden Werte (bias genannt) hat den Grenzwert 0.

Für die Varianz des Schätzers für den Erwartungswert μ gilt:

$$V(\hat{\mu}) = V(\bar{X}) = V\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Mit wachsendem Stichprobenumfang n nähert sich also die Varianz des Schätzers 0.

3. Effizienz

Sie bezeichnet die Genauigkeit, mit der ein Parameter geschätzt wird. Je kleiner die Varianz ist, desto grösser ist die Genauigkeit bzw. die Effizienz.

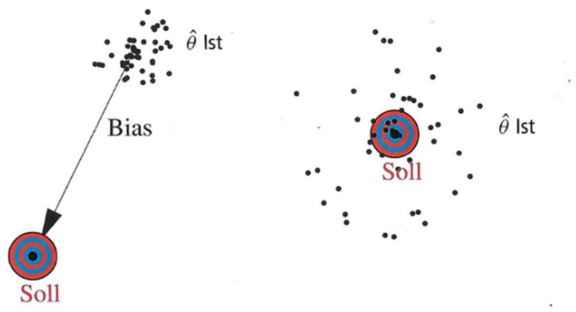
Beispiel:
Schätzt man den Erwartungswert von $\bar{X}$ von X mit $\frac{1}{2}(X_1 + X_2)$,dann ist der Schätzer zwar erwartungstreu, aber nicht konsistent denn

$$Var\left(\frac{1}{2}(X_1 + X_n)\right) = \left(\frac{1}{2}\right)^2 \cdot \left(Var(X_1) + Var(X_n)\right) = \frac{1}{4} \cdot (\sigma^2 + \sigma^2) = \frac{1}{2} \cdot \sigma^2$$

4. Suffizienz

Erschöpfend ist ein Schätzwert, wenn er alle in den Daten der Stichprobe enthaltenden Informationen berücksichtigt, d.h., wenn durch die Berechnung eines weiteren statistischen Kennwerts keine zusätzliche Information über den Parameter gewonnen werden kann.

In der Abbildung zwei Beispiele zum Thema Bias und Varianz von Schätzungen



Linke Zielscheibe:
grosser Bias und kleine Varianz
Rechte Zielscheibe:
unbiased aber grosse Varianz