

Strategie der Inferenzstatistik* (8)

- Beispiel IV -

Da die Münzwürfe voneinander unabhängig sind, kann die Multiplikationsregel der Wahrscheinlichkeit für unabhängige Ereignisse, mit der $P(A \cap B)$ berechnet werden kann, auf das Münzwurfbeispiel angewandt werden. In dem Beispiel ist bei 6 Würfeln 6 mal Kopf erschienen, und wir würden gerne wissen, ob die Münze fair ist. Die Wahrscheinlichkeit, bei 6 Würfeln 6 mal Kopf zu werfen beträgt:

$$P(6 \text{ mal Kopf}) = P(\text{Kopf}) * P(\text{Kopf}) * P(\text{Kopf}) * P(\text{Kopf}) * P(\text{Kopf}) * P(\text{Kopf})$$

Allerdings erfordert die Formel, dass $P(\text{Kopf})$ bekannt ist. Um die Wahrscheinlichkeit zu bestimmen, **bei 6 Würfeln** 6 mal Kopf zu erhalten, muss also bekannt sein, mit welcher Wahrscheinlichkeit die Münze bei **einem** Wurf Kopf zeigt. Diese Wahrscheinlichkeit ist **nur dann 0.5, wenn die Münze fair ist**. Dies ist das einzige, was wir sicher wissen. Dieses Wissen kann für die Entscheidung, ob wir die Münze als fair annehmen (und weiterspielen) oder die Freundin als betrügerisch betrachten (und das Spiel beenden) sollen, wie folgt genutzt werden:

*) Angelehnt an: Welkowitz, J., Ewen, R.B. & Cohen, J. (1982). Introductory statistics for the behavioral sciences. New York (3rd ed.): Academic Press.

Strategie der Inferenzstatistik (9)

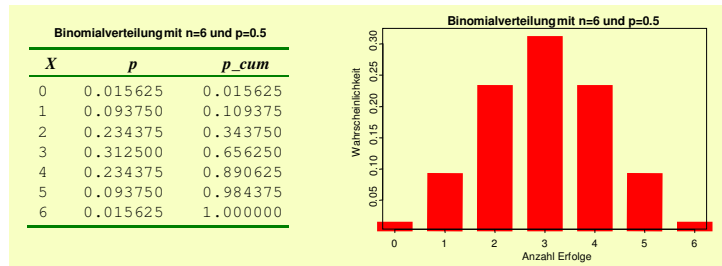
- Beispiel V -

- **Schritt 1:** Annahme, die Münze sei fair (d.h. $P(\text{Kopf}) = 0.5$)
- **Schritt 2:** Bestimmung der Wahrscheinlichkeit von 6 mal Kopf (bei 6 Würfeln) ausgehend von der Annahme in Schritt 1.
- **Schritt 3:** Schritt 2 ergibt die Wahrscheinlichkeit von 6 mal Kopf (bei 6 Würfeln) **wenn** die Münze fair ist. Gestützt auf diese Wahrscheinlichkeit gibt es nun die Wahl zwischen zwei Entscheidungen:
 - **A)** Entscheidung, die Annahme von Schritt 1, dass die Münze fair sei, nicht zurückzuweisen;
 - **B)** Die in Schritt 2 berechnete Wahrscheinlichkeit erscheint derart klein, dass die Annahme in Schritt 1 als unplausibel betrachtet wird. Obwohl 6 mal Kopf möglich ist, ist es vielleicht besser, die Annahme $P(\text{Kopf}) = .5$ zurückzuweisen und anzunehmen, dass die Münze **nicht** fair ist.

Strategie der Inferenzstatistik (10)

- Beispiel VI -

Wenn die Münze fair ist, ist $P(6 \text{ mal Kopf}) = 0.5^6 = 0.016$. Da es sich bei den sechs Münzwürfen um Bernoulli-Experimente handelt, kann hier auch die Binomialverteilung für die Wahrscheinlichkeiten, bei 6 Würfeln mit $p = 0.5$ X mal Kopf zu erhalten, berechnet werden:



Man kann zwischen zwei Alternativen wählen: Entweder die Münze ist fair und ein Ereignis mit einer Wahrscheinlichkeit von 0.016 ist eingetreten (**Hypothese 0**) **oder** die Münze ist unfair, **$P(\text{Kopf})$ ist nicht 0.5 und deshalb ist $P(6 \text{ mal Kopf})$ nicht gleich 0.016 (Hypothese 1).**

Strategie der Inferenzstatistik (11)

- Beispiel VII -

Sozialwissenschaftler folgen häufig der willkürlichen (aber vernünftigen) Konvention, Ereignisse mit einer Wahrscheinlichkeit von weniger als 0.05 als „hinreichend unwahrscheinlich zur Rechtfertigung der Zurückweisung der Nullhypothese“ zu betrachten (und Ereignisse größer als 0.05 als „nicht hinreichend unwahrscheinlich“ zur Rechtfertigung einer Zurückweisung).

Dieser Konvention folgend würde man die Annahme zurückweisen, dass die Münze fair ist, und das Spiel beenden.

Achtung: Die statistische Analyse kann eine korrekte Entscheidung nicht garantieren und es ist immer noch möglich, dass die Entscheidung fehlerhaft ist (hier: unfair gegenüber einer ehrlichen Freundin zu sein).

Im Folgenden werden das Münzbeispiel und das Beispiel einer Evaluationsstudie, in der in einer Stichprobe von 100 Inhaftierten nach Durchführung eines neuen Aggressionstrainings ein Rückgang der beobachteten Aggressivität um 5 Punkte registriert wurde, hinsichtlich ihrer Gemeinsamkeiten bei der Anwendung der Strategie der Inferenzstatistik verglichen.

Strategie der Inferenzstatistik (12)

- Vergleich Münzenexperiment und Evaluationsstudie (I) -

	Münzenexperiment	Evaluationsstudie (Aggressionstraining)
Population	unendliche Anzahl von Münzwürfen	große Anzahl von Personen (Inhaftierte)
Stichprobe	kleinere, machbare Anzahl von Münzwürfen	kleinere, untersuchbare Anzahl von Personen
Hypothese 0	Annahme: Münze ist fair, jede Abweichung vom 50:50 Kopf-Zahl Verhältnis in Stichprobe ist zufällig	Annahme: Aggressionstraining ist unwirksam , jede Abweichung von einer exakter Null-Verbesserung in Stichprobe hängt von Einzelfällen ab, d.h. ist zufällig
Hypothese 1	Ergebnisse sind nicht zufällig (Münze ist gezinkt)	Ergebnisse sind nicht zufällig (Aggressionstraining ist wirksam)

Strategie der Inferenzstatistik (13)

- Vergleich Münzenexperiment und Evaluationsstudie (II) -

	Münzenexperiment	Evaluationsstudie (Aggressionstraining)
Experiment	Daten beobachten (Münze sechs mal werfen, Ergebnis: 6 mal Kopf)	Daten beobachten (Stichprobe messen, Ergebnis: Mittelwert 5 Punkte niedriger)
statistische Analyse	Berechnung von P (sechs mal Kopf)	Berechnung von P (mindestens fünf Punkte Rückgang)
Ergebnis der statistischen Analyse	P (sechs mal Kopf) wenn die Münze fair ist (Antwort: $p = 0.016$)	P (mindestens fünf Punkte Rückgang) wenn Aggressionstraining nicht wirksam ist (Antwort: $p = 0.240$)

Strategie der Inferenzstatistik (14)

- Vergleich Münzenexperiment und Evaluationsstudie (III) -

	Münzenexperiment	Evaluationsstudie (Aggressionstraining)
Entscheidung wenn P „nicht hinreichend klein“ ist nach „.05-Regel“ (P=[.050-1.00])	Hypothese 0 nicht zurückweisen. Die Annahme, die Münze sei fair, ist plausibel. Weiterspielen. Risiko: Fortgesetzte finanzielle Unterstützung einer unehrlichen Freundin (Risiko \leftrightarrow β -Fehler)	Hypothese 0 nicht zurückweisen. Die Annahme, das neue Aggressionstraining sei unwirksam, ist plausibel. Bisherige Methoden beibehalten. Risiko: Tatsächlich bessere Methode wird nicht eingeführt (Risiko \leftrightarrow β -Fehler)
Entscheidung wenn P „hinreichend klein“ ist nach „.05-Regel“ P=[.00-.050]	Hypothese 0 zurückweisen, Hypothese 1 annehmen. Die Annahme, die Münze sei fair, ist nicht plausibel. Spiel beenden. Risiko: Unfaire Beschuldigung einer ehrlichen Freundin (Größe des Risikos $\leq P$)	Hypothese 0 zurückweisen, Hypothese 1 annehmen. Die Annahme, das neue Aggressionstraining sei unwirksam, ist nicht plausibel. Neue Methode einführen. Risiko: Einführung einer Methode, die nicht besser oder schlechter als die Standardmethode ist (Größe des Risikos $\leq P$)

Strategie der Inferenzstatistik (15)

- Vergleich Münzenexperiment und Evaluationsstudie (IV) -

	Münzenexperiment	Evaluationsstudie (Aggressionstraining)
tatsächliche Entscheidung	($p = 0.016$) Hypothese 0 zurückweisen. Spiel beenden. Da die Nullhypothese zurückgewiesen wurde, müssen wir nicht länger annehmen, dass ein Ereignis mit einer Wahrscheinlichkeit von 0.016 eingetreten ist.	($p = 0.240$) Hypothese 0 nicht zurückweisen. Man kann nicht annehmen, dass das Aggressionstraining unter den gegebenen Bedingungen wirksam ist.

Strategie der Inferenzstatistik (16)

- Statistische Modelle I -

Die Binomialverteilung erlaubt es, die Wahrscheinlichkeit für eine beliebige Zahl von Erfolgen bei einer beliebigen Anzahl von Versuchen (Bernoulli-Experimenten) unter der Annahme einer bestimmten Erfolgswahrscheinlichkeit zu berechnen.

Die Binomialverteilung ist ein **statistisches Modell** für die Wahrscheinlichkeit von Ereignissen unter bestimmten Bedingungen.

Beispiele:

Wäre bei den 6 Münzwürfen 5 mal Kopf geworfen worden, wäre das Ergebnis nach der „05-Regel“ „nicht hinreichend unwahrscheinlich“ gewesen, die Nullhypothese $P(\text{Kopf}) = .5$ zurückzuweisen; die Wahrscheinlichkeit von 5 mal Kopf oder häufiger ist $p = .094 + .016 = .109$.

Auf die Frage: „Wie groß ist die Wahrscheinlichkeit, bei sechs Würfeln mit einer fairen Münze 5 mal oder häufiger Kopf **oder** 5 mal oder häufiger Zahl zu erhalten“ wäre die Antwort $p = .109 + .109 = .219$.

Das statistische Modell liefert Informationen über beliebige mögliche Ergebnisse. In ähnlicher Weise lassen sich statistische Modelle für andere Anzahlen von Bernoulli-Experimenten formulieren.

Strategie der Inferenzstatistik (17)

- Statistische Modelle II -

Im Gegensatz zum Münzbeispiel ist es im Fall der Evaluationsstudie schwierig, ein statistisches Modell basierend auf allen möglichen und allen günstigen Ereignissen zu berechnen. Ein (nicht praktikable) Möglichkeit wäre, eine sehr große Zahl von gleich großen Zufallsstichproben aus der Population zu ziehen, die **empirische Stichprobenverteilung** des Mittelwertes zu bestimmen und in Analogie zur Binomialverteilung diese Verteilung als statistisches Modell für die Bestimmung des Anteils der **empirisch bestimmten** günstigen Ereignisse an den **empirisch bestimmten** totalen Ereignissen zu benutzen (an diese Idee knüpfen moderne, rechenintensive **Bootstrap-Verfahren** an).

In der Inferenzstatistik werden statt dessen **theoretische Stichprobenkennwerteverteilungen** benutzt. So zeigt der **zentrale Grenzwertsatz**, dass eine Summe stochastisch unabhängiger Größen, deren Wahrscheinlichkeitsverteilungen gleiche Varianz haben, unabhängig von der Verteilung der Größen in der Population **approximativ** (d.h. mit wachsender Stichprobengröße) **normalverteilt** sind.

Z.B. sind die Mittelwerte beliebiger Merkmale von gleich großen Stichproben selbst normalverteilt um den Populationsmittelwert μ , weshalb die **Normalverteilung als statistisches Modell** zur Prüfung der Wahrscheinlichkeit eines Stichprobenmittelwerts benutzt werden kann.

Strategie der Inferenzstatistik (18)

- Zusammenfassung -

- Ziel der Inferenzstatistik sind **Schlüsse über eine nicht beobachtbare Population** gestützt auf **Informationen aus Zufallsstichproben** aus dieser Population.
- Eine vorläufige Hypothese 0 (Nullhypothese) wird als gültig angenommen, um statistische Berechnungen durchzuführen, die die Wahrscheinlichkeit der beobachteten Ergebnisse liefern, **wenn** die Nullhypothese gültig ist.
- Sind die Ergebnisse „hinreichend unwahrscheinlich“ wird basierend auf dieser Wahrscheinlichkeit die Nullhypothese zu Gunsten der Alternativhypothese **zurückgewiesen**; die Nullhypothese wird **beibehalten**, wenn die Ergebnisse unter der Annahme der Gültigkeit der Nullhypothese nicht hinreichend unwahrscheinlich sind.
- Ein Zusammenfassung aller möglichen Ergebnisse einer bestimmten experimentellen Situation wird als **statistisches Modell** bezeichnet. Weil es unpraktikabel ist, derartige Modelle empirisch zu konstruieren, werden **theoretische Modelle von Stichprobenkennwerteverteilungen** benutzt, um zu entscheiden, ob eine Nullhypothese beibehalten oder zurückgewiesen werden soll.