KATA LOGO Mathematik – Statistik – Roadmap: Von der Hypothese zum p-Wert

0.

Das eigentliche Forschungsziel ist: Beweis der eigenen Hypothese H_1
Dafür muss Nullhypothese H_0 falsifiziert werden können → **Achtung!** Es gibt einen Trend zur vorschnellen Ablehnung (α-Fehler). → Festlegung für die Wahrscheinlichkeit (WS), dass Forscher genau diesen α-Fehler machen (siehe Punkt 3).

1.

Formulierung der inhaltlichen Hypothese H_0	
Entscheidung: Zweiseitige / einseitige / einfache / zusammengesetzte / spezifische / unspezifische	
Beispiel: Münzwurf zweiseitiger Hypothesentest	
H_0	**H_1**
Nullhypothese	**Gegenhypothese**
Annahme über die Wahrscheinlichkeits-Verteilung der Zufallsvariablen ZV (Normalverteilung ...)	
Dies ist die Behauptung!	
Parametrische Verfahren: H_0 macht Aussage über ≥ 1 Parameter der Verteilung, z. B. p und/oder µ.	
Nichtparametrische Verfahren: H_0 macht keine Aussage über Verteilung.	
Formulierung je nach Interessenlage.	

Gruppe I behauptet: $p > p_0 \Rightarrow H_0 : p \leq p_0$	Gruppe I behauptet also: $H_1 : p > p_0$
Gruppe II behauptet: $p < p_0 \Rightarrow H_0 : p \geq p_0$	Gruppe II behauptet also: $H_1 : p < p_0$
z. B. P (X = Kopf) = 0,5	
Ungerichtet: "Es besteht kein Unterschied"	Ungerichtet: "Es besteht ein Unterschied"
Gerichtet: "Median ist *nicht* < oder sogar > als ..."	Gerichtet: "Median von Gruppe 1 ist < als ...".
Verteilung ist ...	
2 Merkmale haben einen Zusammenhang / keinen Zusammenhang (letzteres oft H_0).	
Gruppen unterscheiden sich durch Merkmale.	
Merkmale haben Parameter ...	
2. Inhaltliche Hypothese in statistische Hypothese übersetzen	
Für parametrische Fragestellungen • einseitige Frage: H0 : $\vartheta \leq \vartheta 0$ gegen H1 : $\vartheta > \vartheta 0$ • einseitige Frage: H0 : $\vartheta \geq \vartheta 0$ gegen H1 : $\vartheta < \vartheta 0$ • zweiseitige Frage: H0 : $\vartheta = \vartheta 0$	
3. Bestimmen des Annahme- und Verwerfungsbereichs durch Festlegen des Signifikanzniveaus	
Annahme- und Verwerfungsbereich sind für die Interessengruppen unterschiedlich	
Signifikanzniveau (5 %) ist komplementär zur Sicherheitswahrscheinlichkeit (95 %)	
Sicherheits- oder Vertrauenswahrscheinlichkeit: 1 - α (bei 0,95 = 1 - 0,05)	

Annahme- und Verwerfungsbereich kann mit σ-Umgebung festgelegt werden
→ Bei 95 %: μ-1,96 σ und μ + 1,96σ

4. **Bestimmen des Testverfahrens**

Zunächst Einteilung der Variablen in Skalenniveaus

1. Nominalskaliert mit > 2 Kategorien
2. Nominalskaliert mit 2 Kategorien
3. Ordinalskaliert
4. Intervallskaliert und nicht normalverteilt
5. Intervallskaliert und normalverteilt

5. **Bestimmung der Güte des gewählten Tests**

Güte 1: Teststärke = Trennschärfe = T = power = 1-β
Mit welcher WS entscheidet Test zugunsten H1 (falls richtig)?

Güte 2: β **=** Wahrscheinlichkeit, dass ein bestehender Unterschied **nicht** erkannt wird.

Güte 3: 1-β **=** Wahrscheinlichkeit, dass ein bestehender Unterschied aufgezeigt wird. Dies ist die eigentliche Teststärke.

Güte Kriterium Nr. 1 eines Tests: Objektivität = Ist Testergebnis unabhängig von jeglichen Einflüssen ?

Güte Kriterium Nr. 2 eines Tests: Reliabilität = Zuverlässigkeit, wie genau misst der Test, was er misst ?

Güte Kriterium Nr. 3 eines Tests: Validität = Grad der Gültigkeit: Misst der Test, was er messen soll ?

Gütefunktion g: Für parametrische Tests und bei festem Stichprobenumfang n (z. B. 36) und SN α (5%).
Ordnet jedem möglichen Parameterwert u die Wahrscheinlichkeit für Ablehnung von H0 zu. Man berechnet die Gütefunktion also nur für den Ablehnbereich. Als Variablen bei z. B. der Hypergeometrischen Verteilung errechnet man nur die Güte von N (alle Werte) und K, aber nicht für n oder k (die beiden letzteren werden zufällig gegeben).
u → g(u) = P (H0 wird abgelehnt | u ist der wahre Wert des unbekannten Parameters).

6. **Ziehen der Stichprobe (z. B. Münzwurf mit p = 0,5 und SN = 5 %)**

	Zeigt sich hier vermuteter Zusammenhang? Wenn nicht → Kein Signifikanztest	
	Kann man annehmen, dass dieser Zusammenhang auch in Ω besteht? Wenn ja → Signifikanztest	
7.	**Klassifikation der Daten: Verhältnis-, Intervall-, Ordinal-, Nominalskala**	
	Voraussetzung für parametrische Tests: Normalverteilung der Daten	
	Voraussetzung vorhanden? Stichworte: Restfehler und Homogenität der Varianzen	Voraussetzung nicht vorhanden? 1. Daten-Transformationen 2. Nicht-parametrische Tests 3. Computerintensive Methoden
8.	**Zusammenfassung der Daten**	
	Graphisch	**In Zahlen**
	Stem-and-Leaf **Histogramm** **Summenkurven** **Box-and-Whiskers**	**Lage- und Streuungsparameter** **Konfidenzintervalle** **Bootstrap**
9.	**Teststatistik berechnen und zu Prüfgröße (optimalen Stichprobenumfang) standardisieren**	
	Art des Tests hängt von Stichprobengröße n und jeweils aufgestellter H_0 ab	
a.	Für das Münzbeispiel: **EW $\mu = n \cdot p$**. Bei 36 Würfen ist $\mu = 36 \cdot 0{,}5 = 18$	
b.	Für das Münzbeispiel: **STA $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{9} = 3$**. (Laplace ist in etwa erfüllt: $\sigma > 3$)	
c.	Für das Münzbeispiel: Für Abdeckung einer 95 % Wahrscheinlichkeit steht in Tabellen für die σ-Umgebung: $z = 1{,}96$	
d.	Für das Münzbeispiel: Umgebung festlegen: $\mu - 1{,}96 \cdot 3$ und $\mu + 1{,}96 \cdot 3$ → $X = 12{,}12$ und $X = 23{,}88$	
e.	Für das Münzbeispiel: Entscheidungsregel aufstellen: Verwirf Annahme, dass Erfolgswahrscheinlichkeit $P = 0{,}5$ ist, wenn die Anzahl der Wappen $X < 13$ oder $X > 23$ ist.	

Liegt Stichprobenergebnis (= Testgröße) also im Annahmebereich, wird nicht die Hypothese bestätigt, sondern man entscheidet sich durch die vorher festgelegte Entscheidungsregel, sie weiter als richtig anzusehen.	
Ist diese WS < 0,05 →	
abzulesen an Verteilung der Testgröße →	
wenn kein Zusammenhang in Ω →	

10.

Testentscheidung	
Man will immer die Nullhypothese ablehnen, also man will verwerfen! Ist sie hier (zu α = 0,05) abzulehnen?	
Bei Ablehnung der Nullhypothese: Ergebnis ist signifikant!	
• Stichprobenergebnis: Ist im Annahmebereich • Wenn H0 nicht zum SN α = 0,05 abgelehnt werden kann, ist H0 korrekt • Ist H0 korrekt, wird aber abgelehnt: → α-Fehler • α-Fehler gibt die WS für H0-Ablehnung an • WS für einen α-Fehler ist höchstens α (0,05) • Kontrolle des α-Fehlers durch Oberschranke α • Wenn WS für α-Fehler geringer, steigt WS für β-Fehler **2.**	• Stichprobenergebnis: Ist im Verwerfungsbereich • H0 ist zum SN α abzulehnen • Wenn H0 abgelehnt wurde, kann nicht zu 100% auf H1 geschlossen werden, da Ablehnung zu 5% • H1 (für z. B. p > 0,5) ist korrekt • Ist H1 korrekt, H0 wurde aber nicht abgelehnt → β-Fehler! • Wenn WS für α-Fehler geringer, steigt WS für β-Fehler **1.**
Bei α = 0,05 ist WS = 0,95 für die richtige Entscheidung H_0 abzulehnen.	β-Fehler ist kein eigentlicher Fehler, da der Test in diesem Fall keine Aussage macht. Trotzdem möchte man ihn vermeiden.
Zu gegebenem α ist eine Nullhypothese genau dann abzulehnen, wenn p ≤ α gilt.	

	abzulehnen, wenn $p \leq \alpha$ gilt.	
11.	**Testen von Hypothesen (Sind Ergebnisse zufälliger Natur oder nicht?)**	
a.	**Assoziationen zwischen ≥ 2 Variablen**	
	Unterteilung in abhängige und unabhängige Variablen ist sinnvoll: Regression	Unterteilung in abhängige und unabhängige Variablen ist nicht sinnvoll: Korrelation
b.	**Vergleich von 2 Durchschnittswerten (x quer und y quer)**	
	1 Stichprobe und 1 bekannter Wert:	One-sample-t-Test und/oder Konfidenzintervall
	2 gepaarte (= abhängige = verbundene) Stichprobenwerte:	Gepaarter t-Test Oft sind derselben Person 2 Werte zugeordnet → Differenzen bilden
	2 unabhängige Stichproben:	Ungepaarter t-Test
c.	**Vergleich von > 2 Durchschnittswerten: Varianzanalyse ANOVA**	
	Unterschiede zwischen festen und zufälligen Effekten	
	Bei 1 Faktor: Einfache ANOVA ? Entspricht das F-Test auf Lokationsunterschied?	Bei ≥ 2 Faktoren: a. 1 fester, 1 zufälliger Effekt: Geblockte ANOVA b. ≥ 2 feste Effekte: Mehrfaktorielle ANOVA
d.	**Vergleich von beobachteten (Stichproben) und erwarteten absoluten Häufigkeiten**	
	χ^2-Test	
	1. Daten sind Zufallsdaten 2. Immer mit Tabelle mit absoluten Häufigkeiten 3. Diskrete Verteilungen 4. Es gibt nur endlich viele mögliche Ausprägungen der zu untersuchenden ZV ($x1, x2, ..., x_n$)	
12.	**Entscheidungsfindung mittels p-Wert (Überschreitungs-Wahrscheinlichkeit)**	

a.

p-Wert: Grad der Unwahrscheinlichkeit. Wie glaubhaft ist ein solcher gemäß Daten extremer Versuchsausgang, wenn H0 wahr ist? p-Wert zeigt WS für solche oder extremere Stichproben-Ergebnisse, wenn H0 gilt. Teststatistik zugrundenehmen: 1 - Sm-1 (TS) = z. B. 1 - S35 (TS) → ausrechnen mit Software R.
H0 ablehnen oder nicht? Hoher p-Wert heißt immer: Nicht signifikant, also Werte sind Ergebnis des Zufalls. Vergleich von p-Wert zu Signifikanzniveau. Je kleiner p-Wert, desto stärker spricht er gegen H0. p-Wert = 0,01: So ein kleiner Wert könnte - falls Ho gilt - durch Zufall auftreten, aber die WS dafür ist ≤ 0,01.
p-Wert im Beispiel: P(T ≥ 2,92) ≈ 1-S5(2,92) ≈ 0,7123 (falls H0 gilt) mit S5 aus chi (χ)-Verteilung mit 5 = 6-1 Freiheitsgraden
Bei (quantitativ) überzufällig starker Abweichung der Daten von H0-Gültigkeit (so dass es keine Zufallsstreuung mehr ist) testet man mit Hilfe des p-Werts.
p-Wert: Gibt die WS an, unter H0 das beobachtete Stichprobenergebnis oder einen in Richtung der Alternativen extremeren Wert zu erhalten.
Je kleiner der p-Wert, desto unwahrscheinlicher ist das erhaltene Ergebnis, wenn H0 wahr ist. Ist p-Wert kleiner oder gleich α, so wird H0 verworfen. Bei p = 0, gilt H0 nicht. Große Werte sagen nur was aus, wenn H0 gilt.
Man nimmt den am wenigsten "plausiblen" Wert, z. B. Trefferzahl 0 bei p > 0,4 bei n = 8. Hier nimmt man k= 0 und setzt es z. B. in die Binomialverteilung ein.
p-Wert = 0,95 bedeutet: Zu 95 % ist sicher, dass Ergebnisse nicht zufällig sind. Zu 95 % sind Ergebnisse der Stichprobe auch Ergebnisse der Variablen.
p-Werte kann man unabhängig vom Signifikanzniveau ausrechnen (kleinstmögliche Oberschranke für WS für diesen oder einen noch unwahrscheinlicheren Wert - bei H0)

b.

Entscheidungsfindung mittels Ablehnungsbereich	
Nur bei manchen Tests. Anwendung der *Testfunktion oder Teststatistik T* .	z. B. t-Test und χ^2-Anpassungstest.
Prüfwert = Realisation einer Teststatistik anhand einer Stichprobe.	
H_0 wird abgelehnt $\Leftrightarrow T(D) \in A$. Mit einer vom Zufall abhängigen Zahl T (D), den Daten D und dem Ablehnungsbereich A.	
Aus den Daten D errechnet sich **p-Wert**	

Signifikanz: Kann das Muster in der Stichprobe erklärt werden durch ... ?		
	Zufall	**System**
Ja	Wenn ja, dann hätte Muster eine Wahrscheinlichkeit von ≤ 5 %	Ein in den Daten sichtbares Muster ist **signifikant,** wenn ein System dahinter steckt und der Zufall zu ≥ 5% ausgeschlossen werden konnte.
Nein	Wenn ein in den Daten sichtbares Muster nur schwer durch Zufall zu erklären ist, ist es **signfikant**	