Zusammenfassung Methoden
VL2

Einheit	Objekte, über die man eine Aussage macht.
Population	Alle Einheiten
Stichprobe	Teilmenge der Population/der vorhandenen Einheiten
Merkmal	Eigenschaft einer Einheit; Man möchte Aussagen über Merkmale machen
Merkmalsträger	Einheiten, die auf ein Merkmal hin untersucht werden
Variable	Merkmal mit verschiedenen Ausprägungen
Konstante	Merkmal mit nur einer Ausprägung
Abhängige Variable	Merkmal, das man erklären will
Unabhängige Variable	Merkmal, das man zur Erklärung anwendet
Messung	numerische Darstellung von Werten einer Variable → ist an Regeln gebunden, Werte sollen repräsentativ sein
zulässige Transformation	wenn gleiche empirische Ergebnisse durch verschiedene numerische Zuteilungen ausgedrückt werden können, ohne dass die Aussage falsifiziert wird

Messebene	Verhältnis	Transformation	„Mittelwert"
Nominal	Ähnlichkeit	1:1	Modus
Ordinal	Gröser/kleiner	Rangordnung muss bewahrt werden	Median
Intervall	Zahlenmässige Distanz; Nullpunkt nicht fix	Abstände müssen bewahrt werden	Arithmetisches Mittel
Verhältnis	Absoluter Nullpunkt; Verhältnisse zwischen Objekten	Verhältnisse müssen bewahrt werden	Geometrischer Mittelwert
Absolut	Nur ein Wert	Keine Änderung	Bsp. Zählung aller Männer

VL3: Univariate deskriptive Statistik und Datenvisualisierung

Deskriptive Statistik	Stichprobendaten zusammenfassen und leicht verständlich präsentieren -> keine Rückschlüsse auf Population!
Häufigkeitstabelle	- zeigt verschiedene Ausprägungen einer Variable und ihre Häufigkeiten (h) im Datensatz - gültig für alle Messebenen
Gruppierung von Daten	bei sehr vielen Ausprägungen der Variable
Binbreite	Breite einer Gruppe: k = Anzahl Gruppen

$$b \approx \frac{x_{Max} - x_{Min}}{k}$$

Proporz	= Relative Häufigkeit eines Wertes j bei einer Stichprobengrösse n: f(j) = h(j)/n
Kumulative Häufigkeit	gibt an, wie oft eine Variable den Wert j oder tiefer annimmt (Geht nicht für Nominalskala) → relative Kumulative Häufigkeit
Limit	Zahlen, die oben und unten an Summenzeichen geschrieben werden
Stabdiagramm	= Säulendiagramm Höhe der Säule = Häufigkeit der Kategorie Nominal- und Ordinalskalen Abstand zwischen den Säulen
Histogramm	für Intervall- und Verhältnisskalen Häufigkeit = Grösse einer rechteckigen Fläche Keine Zwischenräume, evt. Gruppierung der Daten
Kerndichtefunktion	Häufigkeiten werden durch stetige Funktion angezeigt ➔ optimaler Einblick in Daten verschaffen (Häufigkeiten an jedem einzelnen Punkt) wichtiger Parameter: Bandbreite je kleiner die Bandbreite, desto genauer wird die Funktion Vorgehen: Für jeden Punkt eine Dichtefunktion erstellen und diese aneinanderreihen
Lagemasse	Zentrale Tendenz einer Häufigkeitsverteilung Beantwortung Frage: Wie sieht typische Einheit aus? Modus, Median, Mittelwert
Modus	Wert, der am häufigsten vorkommt Für alle Messebenen anwendbar Nicht zwingend eindeutig
Median	x mit Wellenlinie Wert, der genau an mittlerer Stelle steht Ab Ordinalebene anwendbar
Quantil	p. Quantil ist ein Wert Qp, bei dem p % der gemessenen Werte links von (oder auf) Qp liegen Median = 50. Quantil
Spezielle Quantile	Terzil Quartil Quintil Dezil
Arithmetisches Mittel	R: mean Mittelwert, Durchschnitt

	x mit Strich Alle Werte addiert, geteilt durch Stichprobengrösse Gilt ab Intervall-Ebene Weniger robust als Median, da stärker von Ausreisser beeinflusst
Ausreisser	atypische Werte: entsprechen nicht den Erwartungen
Streuungsmasse	befassen sich mit Variation der Werte, Unterschiede der Einheiten bezüglich einer Variable Interquartilabstand, Spannweite, Varianz, Standardabweichung
Spannweite	Unterschied vom höchsten zum tiefsten Wert R = x(max) – x(min) Ab Intervallebene anwendbar Nachteil: nur Extremwerte
Interquartilabstand	Differenz zwischen 25. Und 75. Quantil (1. Und 3. Quartil) Ab Ordinalebene anwendbar IQR = gross -> grosse Variation zwischen Einheiten, grosse Streuung Vorteil: robust
Boxplot	Darstellung von Verteilung Mittelstrich: Median Box Länge: IQR Whisker: 1.5x IQR in beide Richtungen Punkte ausserhalb = Ausreisser
Varianz	eine Art Mittelwert Für jeden Wert wird seine Abweichung zum Mittelwert ausgerechnet und quadriert ➔ Summe aller dieser Werte werden durch (n-1) geteilt $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Standardabweichung	Quadratwurzel der Varianz
Schiefe	misst, ob Verteilung der Werte symmetrisch oder asymmetrisch ist $v = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$ V = 0 – Verteilung ist Symmetrisch (Normalverteilung) V < 0 – Verteilung asymmetrisch, linksschief: wenige niedrige Werte

V > 0 – Verteilung asymmetrisch, rechtsschief: wenig hohe Werte

Schiefe & Lagemasse	Mittelwert = Modus: symmetrische Verteilung Mittelwert < Modus: linksschiefe Verteilung Mittelwert > Modus: rechtsschiefe Verteilung

Wölbung

= (exzessive) Kurtosis
misst Steilheit der Verteilung

$$w = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - 3$$

w = 0 – Verteilung normalgipflig (Normalverteilung)
w < 0 – Verteilung flachgipflig (platykurtisch)
w > 0 – Verteilung steilgipflig (leptokurtisch)

VL4: Multivariate deskriptive Statistik für diskrete Variablen
Multivariate deskriptive Statistik: Zusammenhänge von verschiedenen Variablen herausfinden

Diskrete Variablen

Variablen, die eine endliche Anzahl Werte annehmen können
v.a. Nominal- und Ordinalskalen, aber auch Intervall- und Verhältnisskalen in Gruppierungen

Häufigkeitstabelle

bildet die gemeinsame Häufigkeit zweier Variablen ab
2x2 Tabelle

	X		
Y	0	1	
0	h_{00}	h_{01}	$h_{0.} = h_{00} + h_{01}$
1	h_{10}	h_{11}	$h_{1.} = h_{10} + h_{11}$
	$h_{.0} = h_{00} + h_{10}$	$h_{.1} = h_{01} + h_{11}$	$n = h_{0.} + h_{1.} = h_{.0} + h_{.1}$

Randverteilung h0. : Gesamte Anzahl Fälle, wo Variable Y = 0
Randverteilung h.0: Gesamte Anzahl Fälle, wo X = 0

R x C – Tabelle

Variable X hat C verschiedene Ausprägungen, Variable Y hat R verschiedene Ausprägungen
Tabelle hat C x R Zellen

h(ij)

Häufigkeit der Werte Y = i und X = j (verallgemeinert)
Randverteilung

$$h_{i.} = \sum_{j=1}^{C} h_{ij}$$

relative Häufigkeit

n als Grundlage: f(ij) = hij/n
→ Addition aller Zellen = 1
→ kein Unterschied zw. abhängiger und unabhängiger Var.

Bedingte rel. Häufigkeit	Wenn X unabhängige Variable Randverteilung von X als Grundlage: f(ij) = h(ij)/h(.j) Frage: von allen Fällen mit Wert j für X, welchen Anteil hat der Wert i für Y? → Addition aller Zellen innerhalb einer Spalte = 1
Relatives Risiko	Wahrscheinlichkeit eines Ergebnisses einer Variable, verglichen zwischen verschiedenen Gruppen der anderen Variable (2x2 Tabelle) Bsp. Frage: Ist das Risiko, innerhalb eines Jahres abzustürzen, grösser für Minderheitsregierungen oder für Mehrheitsregierungen? → Bedingte relative Häufigkeit Minderheitsregierungen / bedingte relative Häufigkeit Mehrheitsregierungen = RR

Oddsverhältnis

Verhältnis zwischen Wahrscheinlichkeit, dass ein Ereignis stattfindet und Wahrscheinlichkeit, dass es nicht stattfindet

	Ereignis	
Gruppe	Tritt ein	Tritt nicht ein
A	*a*	*b*
B	*c*	*d*

Ω = (a/b) / (c/d)

keinen Bezug auf Randverteilung, sondern auf absolute Häufigkeit

Assoziazionsmass

Messung, wie stark die Werte zweier Variablen zusammenhängen

Mass	Messebene	Nur symmetrisch?
Cramér's V	nominal	ja
Goodman & Kruskal's λ	nominal	nein
Spearman's Rangkorrelation	ordinal	ja

Cramer's V

Phi-Koeffizient

Für 2x2 Tabellen:

$$\phi = \frac{h_{11} \cdot h_{00} - h_{10} \cdot h_{01}}{\sqrt{h_{0.} \cdot h_{1.} \cdot h_{.0} \cdot h_{.1}}}$$

für RxC Tabellen:

$$V = \sqrt{\frac{\phi^2}{min(R, C) - 1}}$$

Wert zwischen 0 (=keine Assoziation zwischen Variablen) und 1 (=perfekte Assoziation, kann nur erreicht werden bei gleichen Ausprägungen der Variablen)

Goodman & Kruskal's

Unterschied zwischen abh. Und unabh. Variable

Proportionale Fehlerverringerung durch Einbezug der unabhängigen Variable

	PRE = (E1-E2)/E1 E1 = Fehler bei Ignorieren der unabhängigen Variable E2 = Fehler bei Berücksichtigen der unabh. Varb. Wert zwischen 0 (unabh. Variable hat keine Aussagekraft) und 1 (unabh. Variable kann abh. Variable perfekt erklären)
Spearman's Rangkorr.	Ab Ordinalebene Keine Unterscheidung abh. und unabh. Begrenzt zwischen -1 und 1 (Zusammenhang positiv oder negativ) Gemessene Werte werden der Reihe nach in Ränge überführt Positive Korrelation: niedrige Ränge von X gehen mit niedrigen Rängen von Y zusammen Negative Korrelation: hohe Ränge von X -> niedrige Ränge von Y Hoch = höher als Mittelwert der Ränge. Je näher an 1/-1, desto stärker die Beziehung $\phi = \frac{h_{11} \cdot h_{00} - h_{10} \cdot h_{01}}{\sqrt{h_{0.} \cdot h_{1.} \cdot h_{.0} \cdot h_{.1}}}$

VL5: Multivariate deskriptive Statistik für stetige Variablen

Stetige Variablen	können unendlich viele Ausprägungen annehmen innerhalb eines Intervalls v.a. Intervall- und Verhältnisskalen
Streudiagramm	Kartesische Koordination x-Achse = werte von X, y-Achse = Werte von Y ➔ gemeinsame Werte darstellen
Jitter	Wenn sich bei Streudiagramm viele Punkte überlagern, kann man die Datenpunkte ein wenig zerstreuen (jittern)
Lineare Assoziation	Annäherung der Punkte in Streudiagramm an eine gerade Linie: gemessen durch Kovarianz und Korrelation
Kovarianz s	misst das Muster der Daten und dessen Ausmass s(xy) > 0: Positive **lineare** Assoziation: hohe Werte von X = hohe Werte von Y s(xy) < 0: Negative lineare Assoziation: hohe Werte von X = tiefe Werte von Y s(xy) = 0: kein linearer Zusammenhang „hoch“ = höher als Mittelwert jeweils Differenz der X-Werts vom X-Mittelwert * Differenz des Y-Werts vom Y-Mittelwert $s_{XY} = \frac{1}{n-1} \sum_i (x_i - \bar{x}) \cdot (y_i - \bar{y})$

Nachteil Kovarianz

hat keine Ober- und Untergrenze -> man sieht nicht, wie stark der Zusammenhang ist
Ändert sich durch Transformation der Werte

Korrelation r

= Pearsonscher Korrelationskoeffizient
= Produktmomentkorrelation
Bereich von -1 bis 1

$$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y}$$

Kovarianz / Produkt der Standardabweichungen
Vorteil: ändert sich nicht durch Transformation der Werte
Wichtig: Korrelation nur wenn Zusammenhang linear!
r(xy) = 0.1: klein
r(xy) = 0.3: mittel
r(xy) = 0.5: gross

Einfache Regressionsanalyse: Unterscheiden zwischen unabhängiger und abhängiger Variable

Regressionslinie

Linie, die sich am ehesten den Punkten im Streudiagramm annähert wird berechnet
Y wird vorhergesagt anhand von X

$$\hat{y} = a + b \cdot x$$

a und b müssen geschätzt werden, sodass die bestmögliche Annäherung an die tatsächlichen Daten geschehen kann
Bei perfekter linearer Korrelation: Regressionslinie exakt richtig
a = erwarteter Wert für y, wenn X=0
b = erwartete Änderung in Y, wenn X um eine Messeinheit erhöht wird

Residuum e

Differenz zwischen eigentlichem y und vorhergesagtem auf Regressionslinie

Scheinkorrelation

Korrelation zwischen zwei Variablen, die nur aufgrund einer Drittvariablen besteht

Statistische Kontrolle

Einbezug der Drittvariable in die Schätzung des Zusammenhangs von X und Y
→ partieller Korrelationskoeffizient

$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ} \cdot r_{YZ}}{\sqrt{1 - r_{XZ}^2} \cdot \sqrt{1 - r_{YZ}^2}}$$

Partieller Korrelationskoeffizient erster Ordnung: Einbezug einer Drittvariable
→ wenn r(xy) oder r(yz) = 0, dann hat Drittvariable keinen Einfluss auf den normalen Korrelationskoeffizent

VL6: Einführung in die Wahrscheinlichkeitstheorie

Zufälligkeit	Statistischer Begriff für Unsicherheit bezüglich Variablen
Wahrscheinlichkeit	numerische Masse für Zufälligkeit
Frequentistische Def.	Relative Häufigkeit des Ergebnisses, wenn die Anzahl der Versuche (n) unendlich gross (oder genügend gross) ist $$Pr_j = \lim_{n\to\infty} \frac{h_j}{n}$$
Bayessche Def.	Wahrscheinlichkeit = Überzeugungsgrad bezgl. Einer Aussage (kann auf empirische Evidenz und wiederholbares Ereignis bezogen sein, muss aber nicht)
Versuch	Prozess, der Sammlung verschiedener Ergebnisse generiert
Stichprobenpunkte	Sammlung verschiedener Ergebnisse Ereignisraum, unmögliches Ereignis, elementares Ereignis
Ereignisraum S	Menge, die alle Stichprobenpunkte umfasst; Menge aller Ereignisse; irgendein Ereignis aus S wird sich immer ergeben Pr(S) = 1
Unmögliches Ereignis Ø	Menge ohne Ergebnisse
Elementares Ereignis	Menge mit nur einem Stichprobenpunkt
Komplementärereignis	Alle Elemente, die nicht zu Ereignis A gehören Pr(A') = 1 – Pr(A)
Vereinigung	Alle Elemente, die zu A **oder** zu B gehören: A ∪ B Pr(A ∪ B) = Pr(A) + Pr(B) – Pr(A ∩ B) → Addieren aller einzelnen Wahrscheinlichkeit minus die Bereiche, die doppelt gezählt wurden (Durchschnitt)
Durchschnitt	Alle Elemente, die zu A **und** B gehören A ∩ B
Disjunkte Ereignisse	schliessen sich gegenseitig aus: A ∩ B = Ø Pr(A1 ∪ A2 ∪ A3 …) = $\sum_{j=1}^{J} \Pr(A_j)$
Axiome von Kolmogoroff	1. Für Jedes Ereignis A gilt Pr(A) >= 0 2. Sicheres Ereignis: Pr(S) = 1 3. disjunkte Ereignisse: Pr(A ∪ B) = Pr(A) + Pr(B)
Durchschnitt berechnen	1. Ereignisraum auf die Stichprobenpunkte limitieren, die zu einem der Ereignisse gehören

	2. Innerhalb dieses beschränkten Ereignisraumes die Wahrscheinlichkeit des anderen Ereignisses berechnen 3. Korrektur für Einschränkung des Ereignisraumes
Formel Durchschnitt	bei Abhängigen Variablen $\Pr(A \cap B) = \Pr(B\|A) \cdot \Pr(A)$ $= \Pr(A\|B) \cdot \Pr(B)$
Gemeinsame Wahrscheinl.	Pr(A ∩ B)
Bedingte Wahrscheinlichk.	Pr(A|B) und Pr(B|A) $\Pr(A\|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$ $\Pr(B\|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$
Randwahrscheinlichkeit	Pr(A), Pr(B)
Statistische Unab.	Bedingte Wahrscheinlichkeit ist gleich der Randwahrscheinlichkeit Pr(A|B) = Pr(A) Durchschnitt: Pr(A ∩ B) = Pr(A) * Pr(B)
Bayesscher Satz	$\Pr(A\|B) = \frac{\Pr(B\|A) \cdot \Pr(A)}{\Pr(B)}$
Beispiel Bayesscher Satz	Überdenken einer Hypothese nach Berücksichtigung der Daten

VL7: Zufallsvariablen und Verteilungen

Zufallsvariable	Funktion über Stichprobenraum Jed--em Stichprobenpunkt wird ein reeller Wert zugewiesen „Zufall“ weil von Unsicherheit geprägt
Kennzeichnung	Zufallsvariable: Grossbuchstaben Deren Ausprägungen: Kleinbuchstaben
Univariate Verteilung	für jede Zufallsvariable kann eine Verteilung dargestellt werden Konsistent mit Axiomen von Kolmogoroff Aussage über die Wahrscheinlichkeiten der Ausprägungen, je nach Art der Zufallsvariable
diskrete Zufallsvariable	Wahrscheinlichkeitsmasseverteilung $f(x) = \Pr(X = x)$

stetige Zufallsvariable

Wahrscheinlichkeitsdichtefunktion

$$\int_a^b f(x)dx = \Pr(a \leq X \leq B)$$

Summe (Integral) aller Wahrscheinlichkeiten der Werte zwischen A und B. Wenn A und B = - unendlich / unendlich, dann ist das Ergebnis = 1.

Träger

Werte von X, die mit einer gewissen Wahrscheinlichkeit (nicht 0) auftreten

Parameter

Charakterisieren die Verteilung der Zufallsvariable
- Lageparameter: Lage der Verteilung
- Skalenparameter: Ausbreitung der Verteilung
- Gestaltsparameter: alle anderen

Kumulative Verteilung

Wahrscheinlichkeit, dass X den Wert x oder kleiner annimmt
Funktion: F(x) = Pr(X>=x) -> Grosses F!

-> Formel Diskret

$$F(x) = \sum_{X \leq x} f(x)$$

-> Formel Stetig

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Eigenschaften F(x)

Werte sind immer zwischen 0 und 1.
Wenn X zunimmt, kann F(x) nicht abnehmen.

Multivariate Verteilung

Gemeinsame Verteilung

Wahrscheinlichkeit des Auftretens zweier Ereignisse von zwei verschiedenen Variablen

-> diskret:

$$f(x,y) = \Pr(X = x \cap Y = y)$$

-> stetig:

Doppelintegral:

$$\int_a^b \int_c^d f(x,y)dxdy = \Pr(a \leq x \leq b \cap c \leq y \leq d)$$

Randverteilung

Addition aller Verteilungen der verschiedenen Ausprägungen einer Variable zusammen mit einer Ausprägung der anderen Variable
→ Kombination der zwei Variablen nicht von Bedeutung

-> diskret

$$f(x) = \sum_y f(x,y)$$

-> Stetig

$$f(x) = \int_y f(x,y)dy$$

Bedingte Verteilung — Wahrscheinlichkeit einer Variable bei gegebener anderer Variable

$$f(x|y) = \frac{f(x,y)}{f(y)}$$

Statistische Unabhängigkeit

$$f(x|y) = f(x)$$
$$f(y|x) = f(y)$$

$$f(x,y) = f(x) \cdot f(y)$$

VL8 Übliche Verteilungen

Woher kommt Verteilung? - theoretische Überlegung
- Empirische Beobachtung

Merkmale
- diskrete Zufallsvariablen
- beschreibt Anzahl der Erfolge von jeweils gleichartigen (gleiche Erfolgsneigung π) und unabhängigen Versuchen
- immer nur jeweils zwei Möglichkeiten: Erfolg oder Misserfolg

Kennzeichnung

$$X \sim \mathcal{BN}(n, \pi)$$

„Zufallsvariable X ist verteilt als Binomialvariable (Anzahl Versuche, Zustimmungswahrscheinlichkeit)“

Massefunktion

$$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

Binomialkoeffizient — Vektor zeigt die Anzahl Möglichkeiten an, genau x Erfolge zu erzielen

Fakultät

$$\binom{2}{x} = \frac{2!}{x!(2-x)!}$$

Berechnung von Vektoren: 2! = 2 * 1

Probleme Binomialv.
- Leute handeln nicht unabhängig voneinander
- Pi ist nicht immer gleich

Beta-Binomialverteilung — Generalisierung der Binomialverteilung; Erfolgswahrscheinlichkeit nicht immer gleich

$$\binom{n}{x} \frac{1}{B(\alpha,\beta)} B(y+\alpha, n-y+\beta)$$

Normalverteilung am meisten angewandte Verteilung

Wichtige Merkmale Hauptkonzepte: Mittelwert/Erwartungswert μ und Varianz σ^2

Symmetrie der Wahrscheinlichkeitsdichte um den Erwartungswert

Konzentration der Wahrscheinlichkeitsmasse in der Mitte (Gipfel)

Bezeichnung X ist normalverteilt, Mittelwert ist 2 und Varianz ist 4 (Standardabweichung also Wurzel(4))

$$X \sim N(2,4)$$

Standardnormalverteilung Erwartungswert = 0

Standardabweichung/Varianz = 1

VL9: Merkmale von Verteilungen

Mittelwert von Zufallsvariablen

Berechnung Alle Werte einer Variable mit Verteilungsfunktion multiplizieren und aufsummieren

-> Diskret

$$\mu = \sum_j x_j \cdot f(x_j)$$

-> stetig

$$= \int x \cdot f(x) dx$$

Varianz Berechnung Alle Abweichungen zum Mittelwert mit Verteilungsfunktion multiplizieren und aufsummieren

-> diskret

$$\sigma^2 = \sum_j (x_j - \mu)^2 \cdot f(x_j)$$

-> stetig

$$\sigma^2 = \int (x - \mu)^2 \cdot f(x) dx$$

Erwartungswerte nach Wahrscheinlichkeit gewichtetes Mittel der Werte, die die Zufallsvariable annimmt

Theoretischer Mittelwert E[X], basiert nicht auf Daten

Was man erwartet, welche Ausprägung eine Variable annehmen wird → Mittelwert

Varianz als Erwartungsw.

$$\sigma^2 = E\left[(X - E[X])^2\right] = \begin{cases} \sum_j (x_j - \mu)^2 \cdot f(x_j) & \text{if } X \text{ diskret} \\ \int (x - \mu)^2 \cdot f(x) dx & \text{if } X \text{ stetig} \end{cases}$$

Rechenregeln zu E	- Wenn X eine Konstante: E[X] = X - wenn k eine Konstante und X eine Zufallsvariable: E[k*X] = k * E[X] E einer Summe ist gleich Summe aller E
Bedingter Erwartungswert	Wert von Y, den man erwarten kann, wenn X eine bestimmte Ausprägung hat E[Y|X]
-> diskret:	$\sum_j y_j \cdot f(y_j \| x)$
-> stetig:	$\int y \cdot f(y \| x) dy$
Gesetz der iterierten E	E[E[Y|X]] = E[Y] Der Erwartungswert vom Erwartungswert von Y gegeben X ist gleich dem Erwartungswert von Y
Momente	Möglichkeit Definition Wölbung/Schiefe oder Kovarianz Quantitatives Mass für die Form einer Punktemenge
erstes (Original)Moment	= Mittelwert, Erwartungswert: $\mu_1 = E[X^1] = \mu$
erstes zentrales Moment	immer 0
zweites zentrales Moment	Varianz: $\mu_2' = E\left[(X - E[X])^2\right] = \sigma^2$
drittes zentrales Moment	Schiefe: $\mu_3^* = \frac{E\left[(X - E[X])^3\right]}{\sigma^3}$
viertes zentrales Moment	Wölbung: $\mu_4^* = \frac{E\left[(X - E[X])^4\right]}{\sigma^4}$

Funktionen von Zufallsvariablen
Lineare Funktionen von Zufallsvariablen produziert neue Zufallsvariable
→ Y wird in Funktion abhängig von X dargestellt

Funktion	$Y = \sum_i (a_i + b_i X_i)$
Erwartungswert	$E[Y] = \sum_i a_i + \sum_i b_i E[X_i]$
Varianz	$Var[Y] = \sum_i b_i^2 Var[X_i] + \sum_i \sum_j b_i b_j Cov[X_i, X_j]$

Erwartungswerte von unabhängigen Zufallsvariablen

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Vorlesung 10

Stichprobenfluktuation	Mit jeder Zusammensetzung der Stichprobe können sich die Schätzungen zu Parametern (z.B. Mittelwert einer Verteilung) verändern.
Schätzer	Mittels erhobener Werte der Stichprobe werden Parameter der Population geschätzt, die danach die Wahrscheinlichkeitsverteilung der Population bestimmen: y(Strich) ist Schätzer von μ

Stichprobentheorie

Zufallsstichprobe	Jede Einheit der Population hat eine positive Wahrscheinlichkeit, in die Stichprobe gewählt zu werden. -> notwendig, um mit statistischem Verfahren, ohne Verzerrung, Rückschlüsse auf die Population zu machen.
Gegenteil Zufallsstichprobe	Entweder lässt sich Selektionswahrscheinlichkeit nicht genau bestimmen oder sie ist 0.
Einfache Stichprobe	Zufallsstichprobe; - jede Einheit hat die gleiche Chance, selektiert zu werden - Jeder kann nur einmal in Stichprobe vorkommen - Jede Stichprobenzusammensetzung mit n Einheiten hat die gleiche Chance, ausgewählt zu werden -> wird selten wirklich praktiziert
Annahme einfache Stichp.	Population = unendlich gross
Bezeichnung Population	N Einheiten
Bezeichnung Stichprobe	n Einheiten, n < N
Stichprobenumfang	n
Anzahl Stichproben?	So viele verschiedene Zusammensetzungen an Stichproben gibt es $\binom{N}{n}$

Schätzer	Regel, die aussagt, wie man aufgrund der beobachteten Daten einen Parameter schätzen soll -> Regel kann auf jeden Datensatz angewendet werden
Bezeichnung Parameter	θ
Bezeichnung Schätzer	$\hat{\theta}$
Schätzung	Spezifischer Wert des Schätzers, den man aufgrund der Daten berechnet -> Wert für einen bestimmten Datensatz
Goldberger-Manski	Definition des Schätzers: Man soll für den geschätzten Wert denjenigen nehmen, den man bei Stichprobe herausgefunden hat. – μ mit $\bar{y}$ – σ^2 mit s^2 – σ_{XY} mit s_{XY}

Stichprobenverteilung
Schätzer ist auch eine Zufallsvariable, da nicht jede Stichprobe dieselben Schätzer Produziert

Stichprobenverteilung	Wahrscheinlichkeitsverteilung des Schätzers, die die Wahrscheinlichkeitsdichte zu allen möglichen Werten des Schätzers zeigt für die Stichprobe mit Umfang n. → Wert basiert auf allen möglichen gemachten Stichproben
n.i.d.	Verschiedene Werte von X (xi) sind unabhängig und kommen aus der gleichen, normalen Population → wenn xi voneinander unabhängig und normalverteilt sind, ist auch deren Mittelwert normalverteilt
Merkmale Stichprob.vert.	- Mittelwert - Standardabweichung/ Varianz → die zwei machen Aussagen darüber, ob man gute Schlussfolgerungen zur Population ziehen kann - mittleres Fehlerquadrat
Mittelwert Stichpr.vert.	= Erwartungswert eines Schätzers $E[\hat{\theta}]$
Verzerrung/Bias	Wenn Erwartungswert nicht gleich ist wie Populationswert, gibt es Verzerrung $B = E[\hat{\theta}] - \theta$
Erwartungstreuer Schätz.	Wenn Bias/Verzerrung = 0 ist. → Erwartungswert von Mittelwert ist = Mü $E[\bar{x}] = \mu$

Annahmen erwartungstreuer Schätzer
- alle x werden aus gleicher Population gezogen
- keine systematischen Messfehler bei X vorhanden
- keine fehlenden Daten, und wenn, dann vollständig zufällig fehlend

Standardfehler
Standardabweichung einer Stichprobenverteilung
Streuung des Schätzers über die Stichproben und die Genauigkeit des Schätzers
→ Streuung des Schätzers gibt einen Eindruck über die Genauigkeit, mit der man die Parameter der Population schätzen kann)
= s.e.

Varianz
Standardfehler^2 -> Standardfehler = Standardabweichung

Eigenschaften Standardf.
Abhängig von Varianz der Zufallsvariable und Stichprobenumfang
- je weniger Zufallsvariable variiert, desto genauer die Schätzung
- je grösser die Stichprobe, desto genauer die Schätzung
  → Vervierfachung Stichprobenumfang = Halbierung Standardfehler

s.e./var des Mittelwerts

$$Var[\bar{x}] = \frac{\sigma^2}{n}$$
$$s.e.[\bar{x}] = \frac{\sigma}{\sqrt{n}}$$

Was wenn X = Konstante?
Varianz & Standardfehler = 0, keine Stichprobenfluktuation

Mittleres Fehlerquadrat
MSE
(Abweichungen der Schätzungen zu den Realwerten)^2
Kombination von Verzerrung und Varianz der Schätzer

$$MSE = B^2 + Varianz$$
$$MSE = E\left[\left(\hat{\theta} - \theta\right)^2\right]$$

MSE von Mittelwert
- ist erwartungstreuer Schätzer, also B = 0

$$MSE[\bar{x}] = \frac{\sigma^2}{n}$$

Zentraler Grenzwertsatz
bei nicht normalverteilten Verteilungen
- wenn n genügend gross ist (man genügend viele Stichproben Elemente in einer Stichprobe erhebt), nähert sich die Verteilung einer Normalverteilung an

Z-Transformation	„Umformung“ irgend einer Verteilung in eine Standardnormalverteilung $Z = \frac{X-\mu}{\sigma},\ Z \sim N(0,1)$

Vorlesung 11: Einführung in das Testen von Hypothesen (Inferenz)

Hypothese	Aussage über Verteilung oder deren Parameter
Hypothesentest	statistisches Verfahren, um die Konsistenz einer Hypothese mit empirischen Daten zu prüfen
Einfache Hypothese	Vollständige Umschreibung einer Verteilung, z.B. präziser Wert eines Parameters
Zusammengesetzte H.	unvollständige Umschreibung (Bsp. Mittelwert ist mindestens 3)
Nullhypothese	Ho, widerlegt unsere Theorie ➔ Ziel: Widerlegen der Nullhypothese
Alternative Hypothese	Ha oder H1, entspricht unserer Theorie
Zweiseitiger Test	Ho ist einfache und H1 zusammengesetzte Hypothese Zweiseitige Fragestellung: Keine Aussage über die Richtung des Unterschieds zwischen H0 und H1 → ungerichtet
Einseitiger Test	H0 und H1 beides zusammengesetzte Hypothesen → gerichtet

Testverfahren
Klassische Testverfahren

Nullhypothese	muss im Vorhinein definiert werden Meistens Verneinung eines Effekts
Teststatistik	Zufallsvariable, hat Wahrscheinlichkeitsverteilung Misst Diskrepanz zwischen Empirie und Nullhypothese z-Test
Verfahren nach Fisher	1. Konstruiere eine einfache Null Hypothese 2. Wahl eine Test Statistik mit eine bekannte Verteilung 3. Ziehe eine Zufallsstichprobe und erhebe Daten 4. Berechne die Test Statistik auf Grund der erhobenen Daten 5. Berechne die p-Wert: Die Wahrscheinlichkeit unsere Wert der Test Statistik, oder ein extremere Wert, wenn H_0 gültig ist 6. Die Null Hypothese wird zurückgewiesen wenn p zu niedrig ist
p-Wert nach Fisher	bedingte Wahrscheinlichkeit; gegeben der Richtigkeit der Nullhypothese, zu welcher Wahrscheinlichkeit erhält man tatsächlich diesen Wert (oder einen extremeren)?

je nach Signifikanzniveau führt der P-Wert zur Verifizierung oder Falsifizierung der Nullhypothese

Verfahren Neyma/Pear

1. Konstruiere H_0 und H_A Unterschied zu Fisher: Ha wird explizit mit einbezogen
2. Wahl eine Typus-I Fehler-Wahrscheinlichkeit: sie gibt die langfristige Wahrscheinlichkeit H_0 unrecht zurück zu weisen. wird im Vorhinein festgelegt (Signifikanzniveau)
3. Wahl eine Test Statistik mit eine bekannte Verteilung
4. Ziehe eine Zufallsstichprobe und erhebe Daten
5. Berechne die Test Statistik auf Grund der erhobenen Daten, sowie ihre p-Wert
6. Falls $p < \alpha$, dann wird H_0 zurück gewiesen. Dass heisst, der Evidenz ist eher konsistent mit H_A.
7. Falls $p \geq \alpha$ wird das Ergebnis als nicht-signifikant bezeichnet. Dass heisst, das am diesem Moment H_0 nicht zurückgewiesen worden kann.

Signifikanzniveau

Alpha; Irrtumswahrscheinlichkeit: p-Werte unter diesem Niveau führen zur Zurückweisung der Nullhypothese
Wird im Vorhinein festgelegt
-> Typus-I Fehler

Hypothesentest

Entscheidung über H_0	Realität über H_0: Richtig	Realität über H_0: Falsch
Zurückgewiesen	Typus-I Fehler	☺
Nicht Zurückgewiesen	☺	Typus-II Fehler

Entscheidung über H_0	Realität über H_0: Richtig	Realität über H_0: Falsch
Zurückgewiesen	α	$1-\beta =$ Schärfe
Nicht Zurückgewiesen	$1-\alpha$	β

Ziel: Alpha (Signifikanzniveau, Typus-I Fehler) minimieren!

Test-Statistik

$$T = \frac{\sqrt{n}(\hat{\pi} - \pi_0)}{\sqrt{\pi_0(1-\pi_0)}} \sim \mathcal{N}(0,1)$$

- ➔ wie Z-Transformation: über dem Bruchstrich: Pi(Dach) = X, Pi(null) = Mü; unter dem Bruchstrich: Wurzel der Varianz = Standardabweichung

→ Wert von T = Teststatistik: kumulierte Wahrscheinlichkeitsdichte dieser Statistik in Standardnormalverteilung: p-Wert
→ wenn zweiseitiger Test (ungerichtet): Absolutwert von T
→ wenn einseitiger Wert (Gerichtet): nicht Absolutwert
→ wenn Einseitig: je nach Alternativhypothese wird positive oder negative Abweichung angeschaut

Kritische Werte

Kritischer Wert	Alternative zu Signifikanzniveau: Schwellenwert für Test-Statistik, der Annahme- und Ablehnungsbereiche einer Hypothese aufzeigt Für Standardnormalverteilungen (z-Test bei Hypothesen zu Anteilen) T(c) = ± 1.96 für **alpha = 0.05** bei zweiseitigem Test T(c) = ± 1.64, je nach Alternativhypothese Bei Alpha = 0.1: T(c) = 1.282
Annahmebereich	Menge der Werte einer Teststatistik, die nicht zur Ablehnung der Nullhypothese führen
Ablehnungsbereich	Menge der Werte einer Teststatistik, die zur Ablehnung der Nullhypothese führen (Wahrscheinlichkeitsdichte dieser Werte (= p-Wert) ist kleiner als Signifikanzniveau) (T(c) > 1.96, < -1.96) → je nachdem ob einseitiger oder zweiseitiger Test, absoluter Wert oder nicht

Testschärfe

Testschärfe/Trennschärfe	Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn sie falsch ist 1 – Beta Ziel: Testschärfe maximieren (0.8 oder höher)
Kritischer Wert Schätzer $\hat{\pi}_c$	Benötigt dazu: Kritischer Wert von T (± 1.96 oder ± 1.64) $T_c = \frac{\sqrt{n} \cdot (\hat{\pi}_C - \pi_0)}{\sqrt{\pi_0 \cdot (1 - \pi_0)}} = 1.64$
Berechnung Testschärfe	benötigt dazu: Kritischer Wert des Schätzers $\hat{\pi}_c$ $1 - \beta = \Pr(\hat{\pi} \geq .91 \mid H_A) = \Pr\left(T \geq \frac{\sqrt{n} \cdot (.91 - .80)}{\sqrt{\pi_1 \cdot (1 - \pi_1)}}\right)$ Annahme hier: Ha ist korrekt 0.91: Kritischer Wert des Schätzers 0.8: Wert aus Ha → je grösser, desto besser
Eigenschaften Testschärfe	- Je grösser n, desto grösser die Testschärfe Effektgrösse: Unterschied zwischen H0 und Ha - je grösser der Unterschied zwischen Ho und Ha, desto grösser die Testschärfe

	- Je grösser Alpha (Typus-I Fehler), desto höher die Testschärfe
Teststatistiken	- T-Verteilung: für Mittelwerte; Wenn man keine Informationen über Population (weder Mittelwert noch Varianz) hat; Parameter: Freiheitsgrade (n-1) - Z-Test: für Verteilungen (p) -> Standardnormalverteilung