Statistik; 25.10.2022

Regression:

Lineare Regression

* Englisch : Ordinary Least Square (OLS)
* Nur bei kardinalen Skalen
* Gibt Zusammenhang
  + Bei Korrelation will man erst herausfinden ob zusammenhang
  + Weiß auch die Richtung der Korrelation (rechts oben – links unten; links oben – rechts unten)
  + Was beeinflusst was
* Exogene = Variabel, die die andere Variabel beeinflusst; aktiv
  + Wird immer auf x-Achse dargestellt
* endogene = Variabel die von anderer beeinflusst wird; passiv
* Exogene = funktion (exogen)  
  f(exogen)  
  Y = f(exogen) 🡨 aufgebaut wie normale Funktion (y = mx +b; y= bx + a)
  + a = Y-Achsenabschnitt  
    a = Durchschnitt Y – (Steigung x Durchschnitt X); f(0), wenn Funktion gegeben
  + b = Steigung

b = Covarianz/Varianz(X) = f’(x)

* Zwei Punkte in eine lineare Gerade umwandeln : lineare Interpolation (zwischen zwei Polen (hier Punkte) eine Verbindung herstellen
  + Wie in Matheunterricht
* Wie sehen ich ob starke Korrelation:

1. Hüllkurve
2. Gerade einzeichnen (Regressionslinie)
3. Ablesen
   * Statistik

* Nennt man dann „lineare Regressionsschätzung“
* Determinationskoeffizient :
  + zeigt Güte der bestimmten Regressionsgerade an (Wie stark die tatsächlichen Werte mit der theoretischen Gerade übereinstimmen)
  + z.B. wie gut passen die Messwerte zu einem Modell
  + = R2 (berechnung von uns nicht erfordert)
  + 0 < R2 < 1
    - 1 = sehr doll
    - 0 = sehr wenig
* Summe der quadratischen Abweichungen soll möglichst klein sei 🡪 möglichst nah auf der imaginären Gerade liegen
  + Freihandmethode : imaginäre Gerade ohne Berechnungen „freihändig“ einfügen
* Weiterziehen der Geraden:
  + Interpolation : Schätzung im bekannten Bereich
    - z.B.: wenn X (eingesetzter) Wert schonmal vorkam
  + Extrapolation : Schätzung im unbekannten Bereich/weiterziehen der Linie
    - Ungenauer
    - z.B.: wenn X (eingesetzter) Wert noch nie vorkam

nicht lineare Regression:

* wird angewandt, wenn Punkte bei linearer regression zu weit wegliegen
* nicht in linearer Funktion dargestellt
* funktion dann höheren Grades (z.B. quadratisch, also 2.)

multiple Regression

* mehrere exogene Faktoren  
  Y = f(X1, X2, X3, X4)
* exogene Faktoren werden i.d.R. addiert

Skalentransformation

* umwandlung von Zahlen in andere Zahlen, sodass der Graph schöner aussieht (z.B. von normalen Zahlen in logarithmisch)

**Wahrscheinlichkeit**

* immer eine Zahl zwischen 0 und 1  
  0 = unmöglich (0 „oben“, also null richtige Ereignisse; 0/x)  
  1 = sicher/kausal (selbe Anzahl oben wie unten; x/x)
  + <0,5 = unwahrscheinlicher; weniger als 50%  
    >0,5 = wahrscheinlicher, dass es eintritt; mehr als 50% Wahrscheinlichkeit
* Abkürzungen:

p (possibilities) 🡨 haben wir auch schon in Schule verwendet  
w (Wahrscheinlichkeit)

* p = Anzahl „gewollter/richtiger“ Ereignisse / Anzahl aller Ereignisse
  + immer Verhältnis
  + z.B.: Würfeln von einer 4 : 4/6
* mit relativer Häufigkeit Wahrscheinlichkeit berechnen :  
  h = fi / n
  + hier gilt auch : Gesetzt der großen Zahl
* a priori = ohne gegebene Häufigkeiten Wahrscheinlichkeiten berechnen (siehe Würfelspiel)  
  a posteriori = mit gegebenen Häufigkeiten Wahrscheinlichkeiten berechnen
* subjektive Schätzung: Schätzung von Wahrscheinlichkeit ohne Berechnung
  + z.B.: Laie schätzt mit welcher Wahrscheinlichkeit es morgen regnet
* Gegenwahrscheinlichkeit = 1 – p

Odds:

* haben keine Abkürzung
* Quotient, dass Wahrscheinlichkeit oder Gegenwahrscheinlichkeit eintritt
  + z.B.: Würfelt eine 6; Wahrscheinlichkeit : 1/6, Gegenwahrscheinlichkeit 5/6;  
    Odds = (1/6)/(5/6) = 0,2
* „Verhältnis von p und (1-p)
  + Beim Beispiel : Odds = 0,2 oder Odds = 1 : 5
* „Die Chance auf X betragen A zu B“
* Formel :
  + Wahrscheinlichkeit/Gegenwahrscheinlichkeit = Odds  
    p / (1-p) = Odds
* Wird nie sowohl Wahrscheinlichkeit als auch odds verwendet, immer nur 1 von beidem
  + Odds v.a. in den USA

Entscheidungsbaum

* Kenne ich als Baumdiagramme
* Bezeichnungen wie in der Schule
  + A Strich = nicht A  
    A = A
* Unabhängige (disjunkte) Ereignisse = mit zurücklegen
* Abhängige (nichtdisjunkte) Ereignisse = ohne zurücklegen
* Pfadwahrscheinlichkeiten:
  + Wahrscheinlichkeit eines Pfads : “senkrecht”; einzelne Wahrscheinlichkeiten werden **multipliziert**
  + Wahrscheinlichkeiten auf mehreren Pfaden: „waagerecht“; erst einzelne Pfade berechnen + dann addieren 🡪 Aufgabenstellung mit dem logischen oder
  + Ausschnitte aus verschiedene Pfaden : ebenfalls zuerst Pfade und dann addieren

Bedingte Wahrscheinlichkeiten:

* Wahrscheinlichkeit hängt immer vom Ergebnis der vorherigen „Runde“ ab
  + Wird immer kleiner (ohne zurücklegen)
* In Klausur nicht zu berechnen : kamen wir nicht zu
* Ergebnis hängt immer von Frage ab

Predictive Analysis:

* aus den Wahrscheinlichkeiten aus dem Baumdiagramm auf die Zukunft schließen
* einfach selbe Wahrscheinlichkeiten daraus auf neues Szenario anwenden
* z.B.: Goldburger Beispiel

Zufallsprozesse:

Theorie d. Zufall

* zwei verschiedene Möglichkeiten
* Zufall arbeitet immer zentralisiert 🡪 Häufung an Mittelwerten
* Zufallsprozess:
  + Einfach : eher gleichverteilung
  + Komplex/schwer : immer näher an Normalverteilung heran

Erwartungswert:

* Durchschnitt bei unendlich vielen Würfen/was man vor dem Experiment als Ergebnis erwartet
* Nur interessant bei sehr vielen versuchen (?)
* E(X) = Wert 1 x Wahrscheinlichkeit W1 + W2 x Wahrscheinlichkeit W2 + W3 x Wahrscheinlichkeit W3 + ….  
  E(X) = X1 x P(X1 = x) + X2 x P(X=x2) + X3 x P(X=x3) + ….
* z.B.: Würfeln = ausgezahlte €; Erwartungswert : 3,5
  + „Veranstalter“ freut sich über <3,5 : verliert weniger Geld als erwartet
  + „Spieler“ freut sich über >3,5 : gewinnt mehr geld als erwartet
* Standardabweichung und Varianz können auch mit Erwartungswert gebildet werden

Standardabweichung, etc

* Normale Abweichung = Durchschnitt/Erwartungswert +- Standardabweichung
* Alles was außerhalb des normal abweichenden Bereichs liegt : außergewöhnliche Abweichung

Variationskoeffizient: v = s/E(X)

* Gibt an um wie viel Prozent die normale Abweichung vom Erwartungswert abweichen
  + „bis wie viel Prozent abweichung ist im Fall X normal“

Spieltheorie:

* Faires Spiel : E(X) = 0
  + Spieler A und Spieler B haben gleiche Chancen zu gewinnen
* Gleichverteilung, da alle Wahrscheinlichkeiten gleich sind 🡪 Hüllkurve gerade
* Zwei Würfel: eingipflige, symmetrische Verteilung 🡪 Dachverteilung
  + Alle 4 Mittelwerte auf Mitte
* Je komplizierter, desto näher an Normalverteilung (auch Wendepunkte, etc)
* µ = wenn alle Mittelwerte an einem Punkt (laut ihm; internet sagt anders)
* je komplexer, desto kleiner die Normalabweichung, da immer mehr Werte auf Mitte konzentriert (σ wird kleiner)
  + je mehr Werte, desto näher an Normalverteilung