Datenanalyse

May 27, 2023

1 Basics

Mittelwert (diskret/stetig)

$$\bar{x} = \sum_{i=1}^{n}(x_i)$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty}(f(x) * x)dx$$

Varianz (diskret/stetig)

$$\sigma^2 = \frac{1}{n} * \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\sigma^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Geschätzte Varianz

$$s^2 = \frac{1}{n-1} * \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Pearson Korrelation

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x s_y}$$

2 Faltung

2.1 Diskret

$$f_Z(z) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} f_X(x_i) * f_Y(x_j) * \mathbb{I}_z(x_i + y_j)$$

Wenn beide Träger teil der natürlichen Zahlen sind, dann darf folgende Formel benutzt werden. ($T_1 = T_2 = \mathbb{N}_0$)

$$f_Z(z) = \sum_{i=0}^{z} f_X(i) * f_Y(z-i)$$

2.2 Stetig

Wenn X und Y stetig verteilte Zufallsvariablen sind und $Z = X + Y$. Dann gilt

$$f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) dy$$

2.3 Bivariate

$$(f * g)(x) := \int f(T) * g(x - T) dT$$

http://www.biancahoegel.de/mathe/analysis/faltung.html

2.4 Dichtetransformationssatz

Es ist eine Funktion "f" gegeben und eine Funktion "g"

$$h = g^{-1}$$

$$f = |h'| * f(h) * \mathbb{I}_\Omega$$

3 Schätzer

To do...

- MSE
- Konsistenz von Schätzern

3.1 Maximum-Likelihood-Methode

Likelihood Funktion

$$L = f(x1) * ... * f(xn)$$

Einsetzten der Likelihood Funktion in den Logarithmus.

$$l(p) = log(L) = log(f_{x1}) + ... + log(f_{xn})$$

Erste Ableitung bilden

$$l(p)' = \frac{\delta}{\delta p} l(p)$$

Zweite Ableitung bilden

$$l(p)'' = \frac{\delta}{\delta p} l'(p)$$

Notwendige Bedingung

$$l(p)' = 0$$

Hinreichende Bedingung

$$l(p)' \neq 0$$

3.2 Momentenmethode

Erwartungswert berechnen und mit dem Empirischen Wert gleichsetzten

$$\mathbb{E}[X] = \bar{x}$$

3.3 Gütekriterien für Punktschätzer

Erwartungstreue ist, wenn für einen Schätzer gilt, dass der Erwartungswert ($E_\theta(T)$) gleich dem geschätzten Wert (θ) ist. Also:

$$E_\theta(T) = \theta$$

Der Bias beschreibt, wie viel der Schätzer vom Erwartungswert abweicht.

$$Bias_\theta(T) = E_\theta - \theta$$

Zudem gibt es noch asymptotische Erwartungstreue. Das bedeutet, dass bei unendlich viele Beobachtungen, der Schätzer erwartungstreu wird ($n- > \infty$)

$$\lim_{n->\infty} E_\theta(T) = \theta$$

4 Konfidenz Intervall

Erwartungswert

$$[\bar{x} - z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{1-\alpha/2} * \frac{\sigma}{\sqrt{n}}]$$

Varianz

$$[\frac{(n - 1 * S^2)}{q_{1-a/2}}; \frac{(n - 1 * S^2)}{q_{a/2}}]$$

5 Hypothesen Tests

To do...

- ab 6.4 unvollständig

$$v = \sqrt{n} * \frac{\bar{x} - h_0}{\sigma}$$

Einstichprobentest (Bernoulli Experiment)

$$v = \sqrt{n} * \frac{\bar{x} - p_0}{\sqrt{q * (1 - q)}}$$

$$p = 2 * (1 - \phi(|v|))$$

phi ist eine Verteilungsfunktion. Entscheidungsregeln (bekanntes σ):

- $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$: H_0 wird abgelehnt, falls $|v| > z_{1-\alpha/2}$
- $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$: H_0 wird abgelehnt, falls $v < z_\alpha$
- $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$: H_0 wird abgelehnt, falls $v > z_{1-\alpha}$

Entscheidungsregeln (unbekanntes σ):

- $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$: H_0 wird abgelehnt, falls $|v| > t_{1-\alpha/2}$
- $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$: H_0 wird abgelehnt, falls $v < t_\alpha$
- $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$: H_0 wird abgelehnt, falls $v > t_{1-\alpha}$

6 Regression

6.1 Einfache lineare Regression

$$f(x) = a + b * x$$

$$b = \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}}{\sum_{i=1}^{n}(y^2) - n\bar{y}^2}$$

$$a = \bar{y} - b * \bar{x}$$

Es ist hilfreich eine Tabelle mit folgenden Spalten zu erstellen:

x	y	x^2	y^2	xy
$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$

Zudem sollten auch $\bar{x}$ und $\bar{y}$ ausgerechnet werden

6.2 Multiple Lineare Regression

$$\beta = (X^T * X)^{-1} * X^T * Y$$

falls $(X^T * X)^{-1}$ existiert