

Statistiek voor A.I.

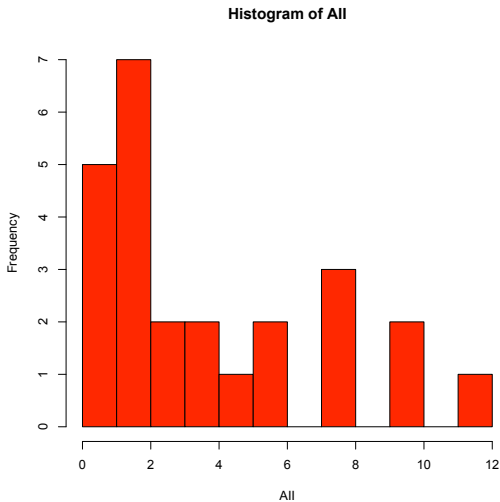
College 8

Dinsdag 9 Oktober

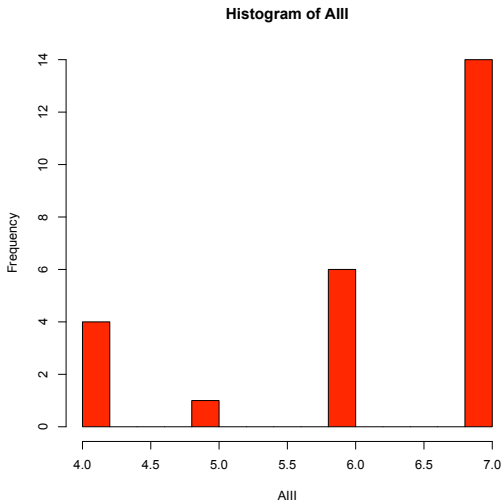
2 Deductieve statistiek

Kansrekening

Het aantal uren dat jullie per week sporten.



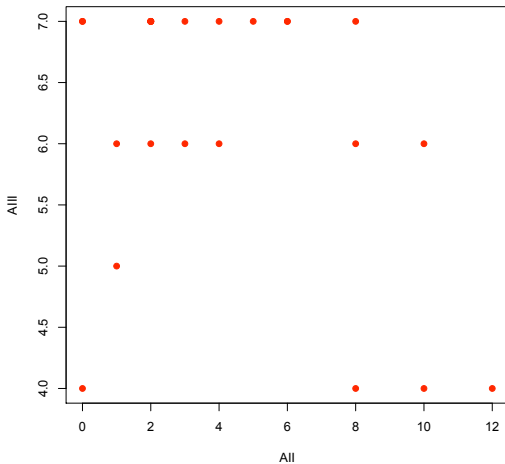
Het aantal zelfgemaakte maaltijden dat jullie per week eten.



Jullie - onderzoek Floor, Minke, Annemarie

Horizontaal: aantal uren dat jullie per week sporten.

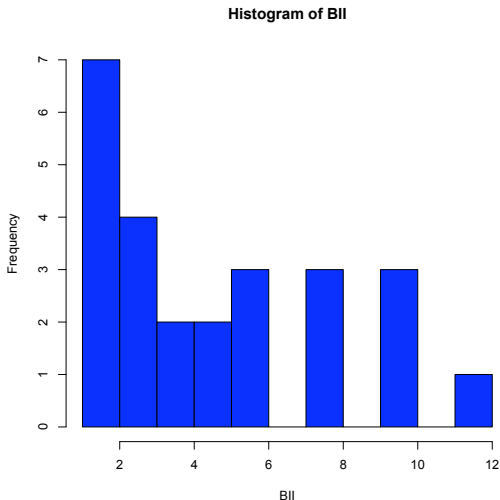
Verticaal: aantal zelfgemaakte maaltijden dat jullie per week eten.



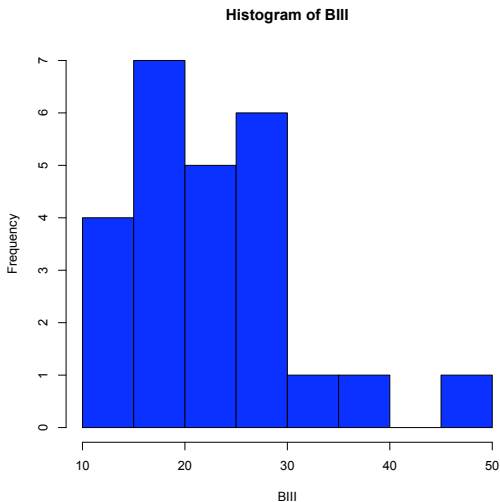
Correlatiecoëfficiënt: -0.4.

Jullie - onderzoek Renate en Marleen

Gemiddeld aantal uren per week dat jullie gesport hebben in het eerste jaar (incl.fietsen).



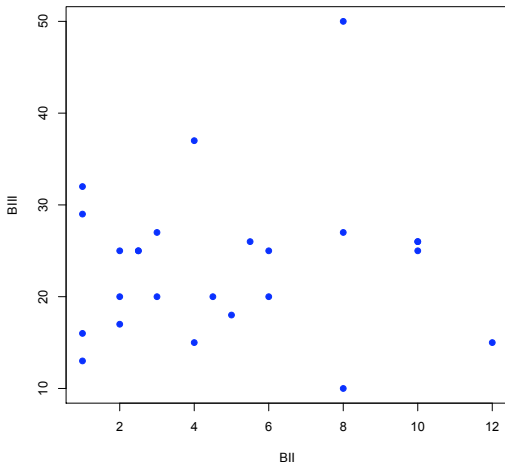
Gemiddeld aantal uren studie per week in het eerste jaar (incl.colleges).



Jullie - onderzoek Renate en Marleen

Horizontaal: Gemiddeld aantal uren per week sporten in het eerste jaar.

Verticaal: Gemiddeld aantal uren studie per week in het eerste jaar.

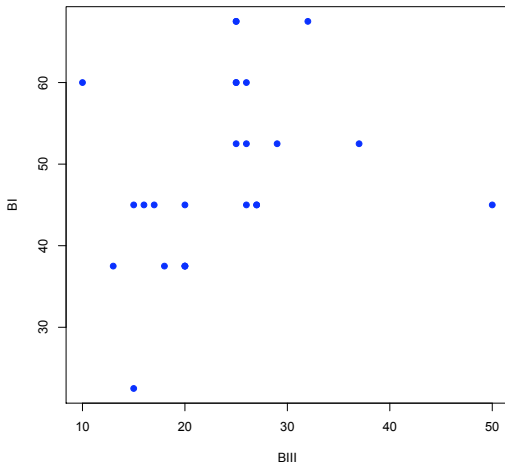


Correlatiecoëfficiënt: 0.07.

Jullie - onderzoek Renate en Marleen

Horizontaal: Gemiddeld aantal uren studie per week in het eerste jaar.

Verticaal: Aantal studiepunten behaald in het eerste jaar.



Correlatiecoëfficiënt: 0.31.

Vandaag :

- Stelling van Bayes
- Bayesiaans leren
- Geschiedenis

Stelling van Bayes

Stelling van Bayes

Stelling Als $H_1, \dots, H_n$ een partitie van de uitkomstenruimte is en E een gebeurtenis, dan geldt voor elke $i \leq n$:

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{j=1}^n P(E|H_j)P(H_j)}.$$

Bewijs Merk op dat voor elke $j \leq n$ geldt:

$$P(E|H_j) = \frac{P(E \cap H_j)}{P(H_j)} \quad P(E \cap H_j) = P(E|H_j)P(H_j).$$

Omdat $H_1, \dots, H_n$ een partitie van de uitkomstenruimte is geldt

$$P(E) = \sum_{j=1}^n P(E \cap H_j) = \sum_{j=1}^n P(E|H_j)P(H_j).$$

Daarom

$$P(H_i|E) = \frac{P(E \cap H_i)}{P(E)} = \frac{P(E|H_i)P(H_i)}{\sum_{j=1}^n P(E \cap H_j)} = \frac{P(E|H_i)P(H_i)}{\sum_{j=1}^n P(E|H_j)P(H_j)}.$$

Stelling van Bayes

Voorbeeld De chips in de iPhone5 komen van drie Samsung fabrieken in Azië. 60% van fabriek I, 30% van II en 10% van III. De kans dat een chip defect is is in de fabrieken respectievelijk 0.01, 0.04 en 0.2.

Een chip in een iPhone5 is defect. Wat is de kans dat die van fabriek I komt?

H_1 : chip komt van I, H_2 : chip komt van II, H_3 : chip komt van III. E : de chip is defect.

De kans dat de chip van fabriek I komt:

$$P(H_1 | E) = \frac{P(E | H_1)P(H_1)}{\sum_{j=1}^3 P(E | H_j)P(H_j)} = \frac{0.01 \times 0.6}{0.01 \times 0.6 + 0.04 \times 0.3 + 0.2 \times 0.1} = 0.1578947.$$

De kans dat de chip van fabriek II komt:

$$P(H_2 | E) = \frac{P(E | H_2)P(H_2)}{\sum_{j=1}^3 P(E | H_j)P(H_j)} = \frac{0.04 \times 0.3}{0.01 \times 0.6 + 0.04 \times 0.3 + 0.2 \times 0.1} = 0.3157895.$$

De kans dat de chip van fabriek III komt:

$$P(H_3 | E) = \frac{P(E | H_3)P(H_3)}{\sum_{j=1}^3 P(E | H_j)P(H_j)} = \frac{0.2 \times 0.1}{0.01 \times 0.6 + 0.04 \times 0.3 + 0.2 \times 0.1} = 0.5263158.$$

Merk op: $P(H_1 | E) + P(H_2 | E) + P(H_3 | E) = 1.$

Stelling Als H_1 , H_2 en E twee gebeurtenissen zijn dan geldt dan geldt:

$$\frac{P(H_1 | E)}{P(H_2 | E)} = \frac{P(E | H_1) P(H_1)}{P(E | H_2) P(H_2)}.$$

Bewijs Merk op dat voor alle A en B geldt:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B | A)P(A)}{P(B)}.$$

Voor E en de H_i leidt dit tot

$$\frac{P(H_1 | E)}{P(H_2 | E)} = \frac{P(E | H_1)P(H_1)P(E)}{P(E)P(E | H_2)P(H_2)} = \frac{P(E | H_1) P(H_1)}{P(E | H_2) P(H_2)}.$$

Bayesiaans leren

Definitie *Bayesiaans leren* heeft (in essentie) de volgende vorm:

Er zijn een aantal *hypotheses* $H_1, \dots, H_n$ die samen een partitie van de uitkomstenruimte vormen.

De hypotheses zijn meer of minder waarschijnlijk: de (initiële) bijbehorende verdeling is de *a-priori* verdeling, de kansen $P(H_i)$ zijn de *a-priori* kansen.

Na het verkrijgen van nieuwe informatie/data/gebeurtenis E worden de kansen van de hypotheses aangepast volgens de stelling van Bayes:

$$P(H_i | E) = \frac{P(E | H_i)P(H_i)}{\sum_{j=1}^n P(E | H_j)P(H_j)}.$$

De kansen $P(H_i | E)$ zijn de *a-posteriori* kansen. De kansen $P(E | H_i)$ zijn de *likelihoods* van E .

Om de a-posteriori kansen $P(H_i | E)$ te berekenen moeten de a-priori kansen $P(H_i)$ en de likelihoods $P(E | H_i)$ bekend zijn.

Definitie De a-priori kansen geven de *bias* bij aanvang weer:

Bij $P(H_i) > P(H_j)$ wordt H_i waarschijnlijker geacht dan H_j .

Bij $P(H_i) = P(H_j)$ worden beide hypotheses even waarschijnlijk geacht.

Leren:

Op grond van telkens nieuwe data $E_1, E_2, \dots$ wordt de verdeling van de hypotheses voortdurend aangepast, $P_0, P_1, P_2, \dots$:

- P_0 is de a-priori verdeling, waarbij $P_0(H_i) = P(H_i)$.
- Na het verkrijgen van data E_1 wordt de nieuwe verdeling P_1 , waarbij $P_1(H_i) = P_0(H_i | E_1)$.
- Na het verkrijgen van data E_2 wordt de nieuwe verdeling P_2 , waarbij $P_2(H_i) = P_1(H_i | E_2)$.
- Etc.

Definitie Het *Brute-Force MAP Learning Algoritme* output na een *update* op grond van informatie E een hypothese h_{MAP} die de hoogste waarschijnlijkheid heeft met betrekking tot E , dat wil zeggen een hypothese H_i waarvoor $P(H_i | E)$ maximaal is.

Voorbeeld Stel dat een spamfilter aanneemt dat het percentage (c) van alle emails die het woord VIAGRA bevatten dat spam is, 80% of 90% is.

Aanvankelijk worden beide waarden voor even waarschijnlijk gehouden:

$P_0(c = 80\%) = P_0(c = 90\%) = 0.5$. Er wordt aangenomen dat de verdeling van spam/niet-spam binomiaal is.

Jij bent het spamfilter aan het trainen en van de 100 emails die het woord VIAGRA bevatten geef je aan dat er 98 spam zijn. Noem deze gebeurtenis E .

Dan zal op grond van deze data het filter voor $c = 90\%$ kiezen:

$$P_1(c = 90\%) = P_0(c = 90\% | E) = \frac{P_0(E | c = 90\%)}{P_0(E | c = 90\%) + P_0(E | c = 80\%)} =$$
$$\frac{\binom{100}{98}(0.9)^{98}(0.1)^2}{\binom{100}{98}(0.8)^{98}(0.2)^2 + \binom{100}{98}(0.9)^{98}(0.1)^2} = 0.9999612.$$

Daarmee is $P_1(c = 80\%) = 1 - 0.9999612 < P_1(c = 90\%)$, en zoals verwacht is $c = 90\%$ de meest waarschijnlijke hypothese.

Voorbeeld Gegeven is een valse munt waarvan bekend is dat de kans p op K 0.42 of 0.43 is. Beide mogelijkheden worden even waarschijnlijk geacht:

$$P(p = 0.42) = P(p = 0.43) = 0.5.$$

De munt wordt 7 maal geworpen, X is het aantal maal K . Stel dat $X = 3$. Welke hypothese is op grond van deze informatie het waarschijnlijkste?

$$\begin{aligned} P(p = 0.42 | X = 3) &= \frac{P(X = 3 | p = 0.42)}{P(X = 3 | p = 0.42) + P(X = 3 | p = 0.43)} = \\ &= \frac{\binom{7}{3}(0.42)^3(0.58)^4}{\binom{7}{3}(0.42)^3(0.58)^4 + \binom{7}{3}(0.43)^3(0.57)^4} = 0.4997. \end{aligned}$$

Dus $P(p = 0.43 | X = 3) = 1 - 0.4997 = 0.5003 > 0.4997$, en daarmee is $p = 0.43$ de meest waarschijnlijke hypothese op grond van de data "3 maal K bij 7 worpen".

Voorbeeld Gegeven is een valse munt waarvan bekend is dat de kans p op K 0.42 of 0.43 is. Op grond van de verkregen informatie dat er bij het 7 maal werpen van de munt 3 maal K is gegooid (zie vorige slide), is de (nieuwe) verdeling van p : $P(p = 0.42) = 0.4997$ en $P(p = 0.43) = 0.5003$.

De munt wordt nog eens 11 maal gegooid, Y is het aantal maal K . Stel dat $Y = 4$. Welke hypothese is op grond van deze informatie het waarschijnlijkste?

$$\begin{aligned}
 P(p = 0.42 \mid Y = 4) &= \\
 &= \frac{P(Y = 4 \mid p = 0.42)P(p = 0.42)}{P(Y = 4 \mid p = 0.42)P(p = 0.42) + P(Y = 4 \mid p = 0.43)P(p = 0.43)} = \\
 &= \frac{\binom{11}{4}(0.42)^4(0.58)^7(0.4997)}{\binom{11}{4}(0.42)^4(0.58)^7(0.4997) + \binom{11}{4}(0.43)^4(0.57)^7(0.5003)} = 0.51.
 \end{aligned}$$

Daarmee is $p = 0.42$ de meest waarschijnlijke hypothese op grond van de laatste data "4 maal K bij 11 worpen".

Merk op: Hoewel de bias voor $p = 0.42$ ($P(p = 0.42) = 0.4997$) lager is dan die voor $p = 0.43$ ($P(p = 0.43) = 0.5003$) is de nieuwe informatie "4 maal K bij 11 worpen" zodanig dat daarna $p = 0.42$ toch het meest waarschijnlijk is.

Geschiedenis

Girolamo Cardano (1501-1576)



Liber de Ludo Aleae (Boek van het Spel met de Dobbelstenen)

Antoine Gombaud, M. le Chevalier de Méré (1607-1684)

Blaise Pascal (1623-1663) en Pierre de Fermat (1601-1665)



Pascal in 1654: Ik zie dat waarheid hetzelfde is in Toulouse als in Parijs.

Finis