Um tvíkostaaðhvarf

Þessi umfjöllun er byggð á vísindagrein um ofnæmi barna eftir KJ og marga fleiri.

Áhrifaþættir

Hér eru nokkrar þeirra áhættuþátta sem notaðir eru í vísindagreininni:

áhrifaþáttur	breyta	gildi
Fjölskyldusaga um ofnaæmi	fjsaga	0 = engin upp í 5 = mikil
Kyn	kyn	0 = strákur, 1 = stelpa
Leikskóli	leiksk	0 = nei eða sjaldan, 1 = a.m.k. 10 klst. á viku í eitt ár
Reykt á meðgöngu	reykt	0 = nei, 1=já

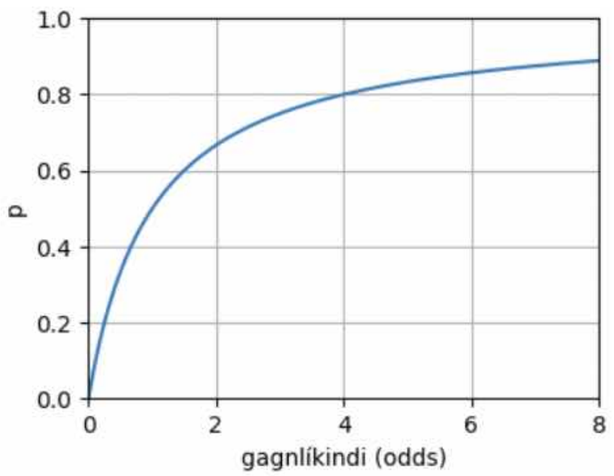
Líkan:

Tvíkostaaðhvarf býr til margfeldislíkan á gagnlíkindi (*odds*). Við látum p tákna líkurnar á að barn á skólaaldri (ca. 8 ára) sé með ofnæmissjúkdóm (astma, exem, frjókornaofnaæmi). Líkanið verður:

$$\begin{aligned}\text{gagnlíkindi} = \frac{p}{1-p} &= \text{grunnlíkur} \times \text{áhrif fjölsk.sögu} \times \text{áhrif kyns} \times \ldots \\ &= a_0 \times a_1^{\text{fjsaga}} \times a_2^{\text{kyn}} \times \ldots\end{aligned}$$

Ástæða þess að gagnlíkindi eru notuð til að mæla áhættuna en ekki bein líkindi er að það er hægt að túlka líkanið þannig að hver áhrifaþáttur hafi bein margfeldisáhrif. Ef við reyndum að gera það sama við líkindi og skoðuðum einstakling með marga jákvæða áhættuþætti mundi áhætta hans hæglega getað orðið meiri en 100%.

Til glöggvunar er hér graf af sambandi gagnlíkinda og líkinda:



Fyrir lítil p er sambandið ekki langt frá því að vera línulegt.

Dæmi:

Gerum ráð fyrir að stikar líkansins séu:

stiki|gildi |:-|:-| grunnlíkur á ofnæmi|0.1 aukning fyrir hvert stig í fjölskyldusögu|1.5 lækkun fyrir stelpur|0.7 lækkun fyrir leikskóla|0.8 aukning ef reykt á meðgöngu|1.5

Líkanið má þá rita:

$$\text{gagnlíkindi} = 0.1 \times 1.5^{\text{fjsaga}} \times 0.7^{\text{kyn}} \times 0.8^{\text{leiksk}} \times 1.5^{\text{reykt}}$$

Skoðum svo stelpu sem fór í leikskóla, fjölskyldusagan hefur gildið 2 og mamman reykti ekki. Samkvæmt líkaninu eru líkurnar á að hún sé með ofnæmi:

$$\begin{aligned} & 0.1 \times 1.5^2 \times 0.7^1 \times 0.8^1 \times 0.9^0 \\ & = 0.1 \times 1.5^2 \times 0.7 \times 0.8 \\ & = 0.126 = 12.6\% \end{aligned}$$

Stikamat

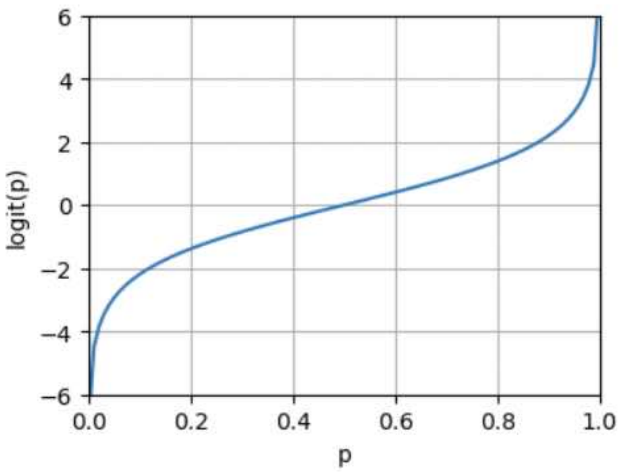
Til að meta stika tvíkostalíkans má taka logra af líkaninu:

$$\text{logit}(p) = \log \frac{p}{1-p} = \log a_0 + (\text{fjsaga}) \log a_1 + (\text{kyn}) \log a_2 + \ldots$$

Hér táknar $\log$ náttúrulegan logra og við höfum notað lograreglurnar:

$$\begin{aligned} \log(AB) &= \log A + \log B \\ \log(A^n) &= n \log A \\ \log 1 &= 0 \end{aligned}$$

Fallið $\log \dfrac{p}{1-p}$ hefur sem sé fengið sérstakt nafn, $\text{logit}(p)$, og hér er graf þess:



Með því að láta $b_i = \log a_i$ verður líkanið:

$$\text{logit}(p) = b_0 + (\text{fjsaga})b_1 + (\text{kyn})b_2 + \ldots$$

Þarna er komið línulegt líkan og stika þess er hægt að meta með (venjulegri) línulegri aðhvarfsgreiningu. Stundum skila forrit sem meta stikana $b_0, b_1, \ldots$ og þá þarf að nota að

$$\text{áhrif} = a_i = e^{b_i}$$

Borgaráhrif

Börnin í stúdíunni sem vísindagreinin lýsir búa í átta borgum í Evrópu og reiknað er með að grunnlíkur á ofnæmi séu ekki þær sömu alls staðar. Þetta er leyst með því að láta a_0 og þar með b_0 vera breytilega á milli borga, sem sé:

$$\text{logit}(p) = B_k + (\text{fjsaga})b_1 + (\text{kyn})b_2 + \ldots$$

þar sem B_k er komið í stað b_0 og k er númer borgar sem viðkomand barn býr í. B_k er logrinn af grunntíðninni í borg k. Til að svona líkan passi fyrir aðhvarfsgreiningarforrit er það stundum ritað með svokölluðum Iverson rithætti:

$$\text{logit}(p) = B_1[k = 1] + \ldots + B_8[k = 8] + (\text{fjsaga})b_1 + (\text{kyn})b_2 + \ldots$$

þar sem hægt er að hugsa sér að allar stærðir séu vigrar og $[P]$ táknar vigur sem er 1 þar sem P er sönn fullyrðing en 0 annarsstaðar.