|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Test niezależności** |  |  |  |  |  |  |  |  |  |  |  |  |
| Dla zbadania wpływu palenia (zmienna *X*) na raka (zmienna *Y*)wylosowano n=874 osoby z populacji brytyjskiej  |  | Prosta regresji 2-go rodzaju (korelacja) |
| i uzyskano wyniki: |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | *X \ Y* | ma raka | nie ma raka |  |  |  |  |  |  |  |  |
|  |  | pali  | 412 | 299 |  |  |  |  |  |  |  |  |  |
|  |  | nie pali | 32 | 131 |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Na poziomie istotności *α*=0.02 |  zweryfikować hipotezę o niezależności palenia i zachorowań na raka. |  |  |  |  |

**Rozwiązanie:**

Aby zbadać niezależność cech użyjemy testu χ2 wyrażonego wzorem:

$$χ\_{e}^{2}=\sum\_{i=1}^{k}\sum\_{j=1}^{m}\frac{(n\_{ij}-np\_{ij})^{2}}{np\_{ij}}$$

Gdzie:

 *n=k\*m* – ilość wszystkich obserwacji

$n\_{ij}- liczebność próbki i,j; np\_{ij}=\frac{n\_{.j}\*n\_{i.}}{n}- liczebność teoretyczna$

|  |  |  |  |
| --- | --- | --- | --- |
| *X \ Y* | ma raka | nie ma raka | Y brzegowy |
| pali  | 412 | 299 | 711 |
| nie pali | 32 | 131 | 163 |
| X brzegowy | 444 | 350 |  |

Tabela *npij*

|  |  |  |
| --- | --- | --- |
| X \ Y | ma raka | nie ma raka |
| pali  | 361,1945 | 349,8055 |
| nie pali | 82,8055 | 80,1945 |

Liczymy empiryczne χ2

$χ\_{e}^{2}=\frac{(412-361,1945)^{2}}{361,1945}+\frac{(299-349,8055)^{2}}{349,8055}+\frac{(32-82,8055)^{2}}{82,8055}+\frac{(131-80,1945)^{2}}{80,1945}=$77,88377

Z tablic $χ\_{0,02}^{2}=$5,411895 (z (2-1)(2-1)=1 stopni swobody.

77,88377>5,411895 – hipotezę o niezależności odrzucamy.

**Korelacja i regresja**:

Dwuwymiarowa zmienna losowa (*X, Y*) ma rozkład dany tabelą:

|  |  |  |  |
| --- | --- | --- | --- |
| *X \ Y* | 1 | 2 | 3 |
| 0 | 0,3 | 0,2 | 0,1 |
| 1 | 0,2 | 0,1 | 0,1 |

Znaleźć rozkłady brzegowe *X* i *Y*, znaleźć ich współczynnik korelacji i prostą regresji 2go rodzaju.

Rozwiązanie:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *X \ Y* | 1 | 2 | 3 | *X* brzegowy |
| 0 | 0,3 | 0,2 | 0,1 | 0,6 |
| 1 | 0,2 | 0,1 | 0,1 | 0,4 |
| *Y* brzegowy | 0,5 | 0,3 | 0,2 |  |

Celem policzenia współczynnika korelacji musimy znaleźć rozkład zmiennej *XY*:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *XY* | 0 | 1 | 2 | 3 |
| *p* | 0,6 | 0,2 | 0,1 | 0,1 |

*EX* = 0\*0,6 + 1\*0,4 = 0,4

*EY* = 1\*0,5 + 2\*0,3 + 3\*0,2 = 1,7

*EXY* = 0\*0,6 + 1\*0,2 + 2\*0,1 + 3\*0,1 = 0,9

*EX*2 = 0\*0,6 + 1\*0,4 = 0,4 więc *D*2*X* = 0,4 – 0,42 = 0,24

*EY*2 = 1\*0,5 + 4\*0,3 + 9\*0,2 = 3,5 więc *D*2*Y* = 3,5 – 1,72 = 0,61

$$ϱ=\frac{EXY-EX EY}{\sqrt{D^{2}X D^{2}Y}}=\frac{0,9-0,4\*1,7}{\sqrt{0,24\*0,61}}=\frac{0,22}{0,383}=0,5744$$

Prosta regresji :

$$y=αx+β $$

$$gdzie: α=ρ\frac{σ\_{2}}{σ\_{1}} oraz β=EY-\frac{σ\_{2}}{σ\_{1}}EX$$

Czyli mamy:

$$α=0,5744\*\frac{\sqrt{0,24}}{\sqrt{0,61}}=0,3603$$

$$β=1,7-0,5744\*\frac{\sqrt{0,24}}{\sqrt{0,61}}\*0,4=1,5559$$

I stąd prosta regresji

$$y=0,3603x+1,5559 $$