|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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:

Gdzie:

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

|  |  |  |  |
| --- | --- | --- | --- |
| *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

77,88377

Z tablic 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

Prosta regresji :

Czyli mamy:

I stąd prosta regresji