

# Kontingenční tabulky - SAS

Zjistěte, zda očkování proti určité chorobě mělo vliv na onemocnění touto nemocí.

Pokud existuje závislost, určete její sílu

| Očkování | Výskyt choroby |     | Celkem |
|----------|----------------|-----|--------|
|          | Ano            | Ne  |        |
| Ano      | 6              | 72  | 78     |
| Ne       | 65             | 57  | 122    |
| Celkem   | 71             | 129 | 200    |

$H_0$  - mezi uvedenými znaky neexistuje závislost.

```
data choroba;  
input ockovani $ onemocneni $ count @@;  
datalines;  
NO NO 57 NO YES 65  
YES NO 72 YES YES 6  
;  
proc freq data=choroba;  
tables ockovani*onemocneni/expected chisq measures;  
weight count;  
Title " Vliv očkování na onemocnění určitou  
chorobou";  
run;
```

The FREQ Procedure

|           | ockovani | onemocneni |        |
|-----------|----------|------------|--------|
| Frequency |          |            |        |
| Expected  |          |            |        |
| Percent   |          |            |        |
| Row Pct   |          |            |        |
| Col Pct   | NO       | YES        | Total  |
| NO        | 57       | 65         | 122    |
|           | 78.69    | 43.31      |        |
|           | 28.50    | 32.50      | 61.00  |
|           | 46.72    | 53.28      |        |
|           | 44.19    | 91.55      |        |
| YES       | 72       | 6          | 78     |
|           | 50.31    | 27.69      |        |
|           | 36.00    | 3.00       | 39.00  |
|           | 92.31    | 7.69       |        |
|           | 55.81    | 8.45       |        |
| Total     | 129      | 71         | 200    |
|           | 64.50    | 35.50      | 100.00 |

Statistics for Table of ockovani by onemocneni

| Statistic                   | DF | Value   | Prob   |
|-----------------------------|----|---------|--------|
| Chi-Square                  | 1  | 43.1824 | <.0001 |
| Likelihood Ratio Chi-Square | 1  | 49.2863 | <.0001 |
| Continuity Adj. Chi-Square  | 1  | 41.2144 | <.0001 |
| Mantel-Haenszel Chi-Square  | 1  | 42.9665 | <.0001 |
| Phi Coefficient             |    | -0.4647 |        |
| Contingency Coefficient     |    | 0.4214  |        |
| Cramer's V                  |    | -0.4647 |        |

The FREQ Procedure

Statistics for Table of ockovani by onemocneni

| Statistic                         | Value   | ASE    |
|-----------------------------------|---------|--------|
| Gamma                             | -0.8638 | 0.0586 |
| Kendall's Tau-b                   | -0.4647 | 0.0525 |
| Stuart's Tau-c                    | -0.4338 | 0.0535 |
| Somers' D C R                     | -0.4559 | 0.0543 |
| Somers' D R C                     | -0.4736 | 0.0548 |
| Pearson Correlation               | -0.4647 | 0.0525 |
| Spearman Correlation              | -0.4647 | 0.0525 |
| Lambda Asymmetric C R             | 0.1127  | 0.1465 |
| Lambda Asymmetric R C             | 0.1923  | 0.1309 |
| Lambda Symmetric                  | 0.1544  | 0.1213 |
| Uncertainty Coefficient C R       | 0.1894  | 0.0458 |
| Uncertainty Coefficient R C       | 0.1842  | 0.0449 |
| Uncertainty Coefficient Symmetric | 0.1868  | 0.0452 |

Z hodnoty  $\chi^2$  (chi square = 43,18) a z p-hodnoty k ní se vztahující ( $p < .0001$ ) je patrné, že nulovou hypotézu lze zamítnout a závislost je statisticky významná. Kromě této charakteristiky můžeme použít i věrohodnostní poměr (*Likelihood Ratio*), kde se v testování postupuje stejným způsobem. Jako míru těsnosti závislosti můžeme využít Phi coefficient, Pearsonův koeficient kontingence (*Contingency Coefficient*) či Cramérovo V (*Cramer's V*). Všechny tyto hodnoty značí střední závislost. K hodnocení můžeme dále použít příslušnou míru typu PRE. V našem případě je vhodnou měrou koeficient lambda, který vyjadřuje míru redukce chyby pro predikci, jestliže známe hodnotu vysvětlující proměnné.

V souboru 27 dotázaných bylo zjišťováno, zda existuje vztah mezi spokojeností s bydlením a polohou bytu. Proveďte příslušný test.

| Spokojenost s bydlením | Poloha bytu |                 | Celkem |
|------------------------|-------------|-----------------|--------|
|                        | V centru    | Na okraji města |        |
| Ano                    | 9           | 8               | 17     |
| Ne                     | 3           | 7               | 10     |
| Celkem                 | 12          | 15              | 27     |

$H_0$  - mezi uvedenými znaky neexistuje závislost.



The FREQ Procedure

Table of spokojenost by poloha

| spokojenost | poloha  |        |        |
|-------------|---------|--------|--------|
| Frequency   |         |        |        |
| Expected    |         |        |        |
| Percent     |         |        |        |
| Row Pct     |         |        |        |
| Col Pct     | CENTRUM | OKRAJ  | Total  |
| ANO         | 9       | 8      | 17     |
|             | 7.5556  | 9.4444 |        |
|             | 33.33   | 29.63  | 62.96  |
|             | 52.94   | 47.06  |        |
|             | 75.00   | 53.33  |        |
| NE          | 3       | 7      | 10     |
|             | 4.4444  | 5.5556 |        |
|             | 11.11   | 25.93  | 37.04  |
|             | 30.00   | 70.00  |        |
|             | 25.00   | 46.67  |        |
| Total       | 12      | 15     | 27     |
|             | 44.44   | 55.56  | 100.00 |

Statistics for Table of spokojenost by poloha

| Statistic                   | DF | Value  | Prob   |
|-----------------------------|----|--------|--------|
| Chi-Square                  | 1  | 1.3421 | 0.2467 |
| Likelihood Ratio Chi-Square | 1  | 1.3705 | 0.2417 |
| Continuity Adj. Chi-Square  | 1  | 0.5738 | 0.4488 |
| Mantel-Haenszel Chi-Square  | 1  | 1.2924 | 0.2556 |
| Phi Coefficient             |    | 0.2229 |        |
| Contingency Coefficient     |    | 0.2176 |        |
| Cramer's V                  |    | 0.2229 |        |

WARNING: 25% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test

|                               |               |
|-------------------------------|---------------|
| Cell (1,1) Frequency (F)      | 9             |
| Left-sided Pr <= F            | 0.9422        |
| <b>Right-sided Pr &gt;= F</b> | <b>0.2256</b> |
| Table Probability (P)         | 0.1678        |
| Two-sided Pr <= P             | 0.4244        |

Výstupní tabulka je stejná jako v předchozím příkladu, co je zde důležité, je informace označená jako **WARNING**. Jde o to, že více než 20% teoretických četností je menších než 5 a proto  $\chi^2$  test není použitelný a je užít Fisherův exaktní test, který je také pod tímto upozorněním proveden. Z řádku “*Right-sided Pr >= F=0.2256*“ je patrné, že pravděpodobnost chyby překračuje obvykle volenou  $\alpha=0,05$  a tedy nulovou hypotézu o nezávislosti nelze zamítnout.

Byl sledován výskyt alergií u dětí v souvislosti s bydlením v silně znečištěném ovzduší. Kolikrát je vyšší pravděpodobnost výskytu alergie u dítěte žijícího ve znečištěném ovzduší a o kolik se zvýší pravděpodobnost výskytu alergie ?

| Alergie | Bydlení v silně znečištěném ovzduší |    |
|---------|-------------------------------------|----|
|         | Ano                                 | Ne |
| Ano     | 28                                  | 11 |
| Ne      | 16                                  | 35 |

Statistics for Table of alergie by znečisteni

| Statistic  | DF | Value   | Prob   |
|------------|----|---------|--------|
| Chi-Square | 1  | 14.4514 | 0.0001 |

Z velikosti p-value je patrné, že závislost je významná. Nás však zajímá ohrožení.

Estimates of the Relative Risk (Row1/Row2)

| Type of Study             | Value  | 95% Confidence Limits |         |
|---------------------------|--------|-----------------------|---------|
| Case-Control (Odds Ratio) | 5.5682 | 2.2313                | 13.8952 |
| Cohort (Col1 Risk)        | 2.2885 | 1.4576                | 3.5929  |
| Cohort (Col2 Risk)        | 0.4110 | 0.2409                | 0.7010  |

Hodnocení: Relativní riziko (*Cohort Col1Risk*) = 2,2885 a tedy pravděpodobnost výskytu alergie u dětí žijících v silně znečištěném ovzduší je 2,88-krát vyšší než u dětí žijících ve zdravém prostředí.

Křížový poměr (*Odds ratio*) – šance na onemocnění = 5,5682, což znamená, že děti žijící v silně znečištěném ovzduší mají 5,56-krát vyšší šanci na onemocnění než děti žijící ve zdravém prostředí.

Atributivní riziko (*Cohort Col2Risk*) = 0,411 udává, že pravděpodobnost výskytu alergií u dětí žijících ve znečištěném ovzduší je o 0,41 vyšší než u dětí žijících ve zdravém prostředí.