**13. Analýza kvalitativních dat**

V této kapitole se budeme zabývat analýzou kvalitativních dat pomocí tzv. kontingenční tabulky. Tato tabulka nepoužívá ani pořadí a je proto použitelná i pro nominální náhodné veličiny.

S kontingenční tabulkou jsme se setkali již v kapitole o náhodném vektoru.

Mějme dvě kvalitativní náhodné veličiny: X, která má *r* variant a Y, která má *s* variant. Do prvního řádku napíšeme jednotlivé varianty veličiny Y a do prvního sloupce varianty veličiny X. V jádru tabulky pak budou simultánní četnosti příslušné dvojice variant. Poslední řádek obsahuje sloupcové součty a poslední sloupec řádkové součty simultánních četností – tedy marginální četnosti variant veličin X, Y.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***y*[1]** | **. . .** | ***y*[*s*]** | ***nj*.** |
| ***x*[1]** | *n*11 | . . . | *n*1*s* | ***n*1.** |
| **˸** | ˸ |  | ˸ | **˸** |
| ***x*[*r*]** | *nr*1 | . . . | *nrs* | ***nr.*** |
| ***n*.*k*** | ***n*.1** | **. . .** | ***n*.*s*** | ***N*** |

*Kontingenční tabulka absolutních četností*

Simultánní pravděpodobnosti odhadneme pomocí relativních četností a marginální pravděpodobnosti pomocí relativních marginálních četností a .

Mouka

=AVERAGE

Směrodatná odchylka: =STDEV.S

Jednovýběrový t-test:

=T.TEST(B1:B15, 1000, 1, 1)

Array1: vaše data (B1:B15)

Array2: deklarovaná hmotnost (1000)

Tails: 1 (jednostranný test)

Type: 1 (jednovýběrový t-test)

**Test nezávislosti v kontingenční tabulce**

Testujeme nulovou hypotézu

H0: Veličiny X, Y jsou nezávislé proti alternativě H1: Veličiny X, Y nejsou nezávislé.

Myšlenka je v podstatě stejná jako při testu dobré shody: testuje se, jestli jsou pozorované simultánní relativní četnosti shodné s teoretickými relativními četnostmi, které by v případě nezávislých veličin měly být součinem příslušných marginálních relativních četností, tedy

Absolutní simultánní četnost je pak: =

Realizace pivotové statistiky: *r*0 = = … =

(kvantily budeme vybírat z ((*r* – 1)(*s* – 1))

Kritický obor je: W =

Pozn.: Pro tento test musí být splněny podmínky, že minimálně 80% teoretických četností je rovno aspoň 5  
 a zbylých 20% neklesne pod 2.

K měření síly závislosti se používá Cramérův koeficient V = a podle jeho číselné realizace hodnotíme závislost takto: 0 – 0,1 zanedbatelná závislost

0,1 – 0,3 slabá

0,3 – 0,7 střední

0,7 – 1 silná

Za minutu

B1: Statistiky

B2: =AVERAGE(A2:A11) # Průměr

B3: =STDEV.S(A2:A11) # Směrodatná odchylka

B4: =COUNT(A2:A11) # Počet měření

B5: =CONFIDENCE.T(0.05,B3,B4) # Interval spolehlivosti

B6: =(B2-60)/(B3/SQRT(B4)) # T-statistika

B7: =T.DIST.2T(ABS(B6),B4-1) # P-hodnota

**Čtyřpolní kontingenční tabulka**

Mají-li obě veličiny X, Y pouze dvě varianty, příslušná kontingenční tabulka se nazývá čtyřpolní.

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***y*[1]** | ***y*[2]** | ***nj*.** |
| ***x*[1]** | *a* | *b* | ***a+b*** |
| ***x*[2]** | *c* | *d* | ***c+d*** |
| ***n*.*k*** | ***a+c*** | ***b+d*** | ***n*** |

Realizace pivotové statistiky: *r*0 = =

(kvantily budeme vybírat z (1))

Kritický obor je: W =

***Testování pomocí podílu šancí ve čtyřpolní kontingenční tabulce***

V případě, že varianta x[1] představuje „úspěch“, varianta x[2] představuje „neúspěch“  
a varianta y[1] představuje „okolnost 1“, varianta y[2] představuje „okolnost 2“,

pak poměr počtu úspěchů k počtu neúspěchů (tzv. šance) za 1. okolnosti je

a poměr počtu úspěchů k počtu neúspěchů za 2. okolnosti je .  
Výběrový podíl šancí („odds ratio“) je pak OR = =

Jsou-li veličiny nezávislé, pak teoretický podíl šancí je 1, Čím více se výběrový OR bude lišit od 1, tím bude závislost veličin silnější. Protože ale teoretický podíl šancí je z intervalu 0, ) – tedy hodnoty jsou kolem 1 rozmístěny nesymetricky – používá se logaritmus:

Realizace pivotové statistiky: *r*0 = (kvantily budeme vybírat z N(0,1))

Kritický obor je: W =

kapalina do ostřikovačů

C1: Statistiky

C2: =AVERAGE(A2:A9) # Průměr Výrobce 1

C3: =AVERAGE(B2:B9) # Průměr Výrobce 2

C4: =VAR.S(A2:A9) # Rozptyl Výrobce 1

C5: =VAR.S(B2:B9) # Rozptyl Výrobce 2

C6: =F.TEST(A2:A9,B2:B9)# F-test pro rozptyly

C7: =T.TEST(A2:A9,B2:B9,2,2) # T-test pro střední hodnoty

***Testování pomocí Spearmanova koeficientu***

Pro náhodné veličiny ordinálního typu lze použít testování nezávislosti pomocí Spearmanova korelačního koeficientu (viz. předchozí kapitola).