**Teste de ajustamento do qui-quadrado**

Ex1, pág 189

Feito à mão na aula

Comparar valores com os do spss

**Nº rapazes em família com 5 filhos**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 0 | 8 | 2,5 | 2,5 | 2,5 |
| 1 | 40 | 12,5 | 12,5 | 15,0 |
| 2 | 88 | 27,5 | 27,5 | 42,5 |
| 3 | 110 | 34,4 | 34,4 | 76,9 |
| 4 | 56 | 17,5 | 17,5 | 94,4 |
| 5 | 18 | 5,6 | 5,6 | 100,0 |
| Total | 320 | 100,0 | 100,0 |  |

**Nº rapazes em família com 5 filhos**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed N | Expected N | Residual |
| 0 | 8 | 10,0 | -2,0 |
| 1 | 40 | 50,0 | -10,0 |
| 2 | 88 | 100,0 | -12,0 |
| 3 | 110 | 100,0 | 10,0 |
| 4 | 56 | 50,0 | 6,0 |
| 5 | 18 | 10,0 | 8,0 |
| Total | 320 |  |  |

**Test Statistics**

|  |  |
| --- | --- |
|  | Nº rapazes em família com 5 filhos |
| Chi-Square(a) | 12,000 |
| df | 5 |
| Asymp. Sig. | ,035 |
| Exact Sig. | ,035 |
| Point Probability | ,000 |

a 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 10,0.

Ex4, pag.190

X- modo de pagamento

Distribuição de X:

P(X=cheque)=0,35

P(X=cartão)=0,48

P(X=dinheiro)=0,17

N=200 vendas

H0: o modo de pagamento na época natalicia é concordante com a informação que a loja tem

H1: o modo de pagamento na época natalicia não é concordante com a informação que a loja tem

Alfa=0,05

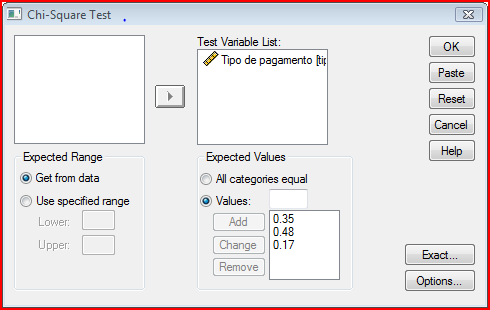
**Tipo de pagamento**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Cheque | 47 | 23,5 | 23,5 | 23,5 |
| Cartão | 116 | 58,0 | 58,0 | 81,5 |
| Dinheiro | 37 | 18,5 | 18,5 | 100,0 |
| Total | 200 | 100,0 | 100,0 |  |

|  |  |  |
| --- | --- | --- |
|  | | Valores esperados |
| Valid | Cheque | 200\*0,35=70 |
|  | Cartão | 200\*0,48=96 |
|  | Dinheiro | 200\*0,17=34 |
|  | Total | 200 |

Teste de ajustamento do qui-quadrado no SPSS

Analyze->Nonparametric tests->chi-square



Na caixa dos valores esperados deverão ser colocadas as frequências observadas por ordem das categorias. Também se podem colocar as probabilidades (o SPSS calcula depois as freq esperadas).

No exact escolher o teste exacto.

Outputs

**Tipo de pagamento**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed N | Expected N | Residual |
| Cheque | 47 | 70,0 | -23,0 |
| Cartão | 116 | 96,0 | 20,0 |
| Dinheiro | 37 | 34,0 | 3,0 |
| Total | 200 |  |  |

**Test Statistics**

|  |  |
| --- | --- |
|  | Tipo de pagamento |
| Chi-Square(a) | 11,989 |
| df | 2 |
| Asymp. Sig. | ,002 |
| Exact Sig. | ,002 |
| Point Probability | ,000 |

a 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 34,0.

(condições de aplicabilidade verificadas)

Como sig=0,002 < alfa=0,05 rejeitamos a hipótese de o modo de pagamento ser concordante com a informação que a loja tem.

Ex2 pag189

X- nº de ampolas defeituosas por caixa

H0: X segue distribuição de Poisson

H1: X não segue dist de Poisson

O paraâmetro lambda não é conhecido. Terá de ser estimado através da média amostral.

**Nº ampolas defeituosas**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 0 | 66 | 33,0 | 33,0 | 33,0 |
| 1 | 84 | 42,0 | 42,0 | 75,0 |
| 2 | 38 | 19,0 | 19,0 | 94,0 |
| 3 | 9 | 4,5 | 4,5 | 98,5 |
| 4 | 2 | 1,0 | 1,0 | 99,5 |
| 5 ou + | 1 | ,5 | ,5 | 100,0 |
| Total | 200 | 100,0 | 100,0 |  |

**Statistics**

Nº ampolas defeituosas

|  |  |  |
| --- | --- | --- |
| N | Valid | 200 |
| Missing | 0 |
| Mean | | 1,00 |

Média amostral =1

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Prob (lambda chapeu = xbarra =1) | Freq esperadas |
| Valid | 0 | 0,3679 | 200\*0,3679=73,58 |
|  | 1 | 0,3679 | 73,58 |
|  | 2 | 0,1839 | 36,78 |
|  | 3 | 0,0613 | 12,26 |
|  | 4 | 0,0153 | 3,06 |
|  | 5 ou + | 0,0037 | 0,74 |
|  | Total | 1 | 200 |

Reparem que existe uma categoria da variável que tem valor esqperado inferior a 1 e 2 em 6 (o que corresponde a 33%) com valor esperado inferior a 5. Isto viola as condições de aplicabilidade pelo que vamos ter que juntar categorias.

Se não dessem conta disto, o spss alertavamos da seguinte forma:

**No quadro dos valores de teste possuem um rodapé com essa informação:**

**Test Statistics**

|  |  |
| --- | --- |
|  | Nº ampolas defeituosas |
| Chi-Square(a) | 3,622 |
| df | 5 |
| Asymp. Sig. | ,605 |
| Exact Sig. | ,591 |
| Point Probability | ,000 |

a 2 cells (33,3%) have expected frequencies less than 5. The minimum expected cell frequency is ,7.

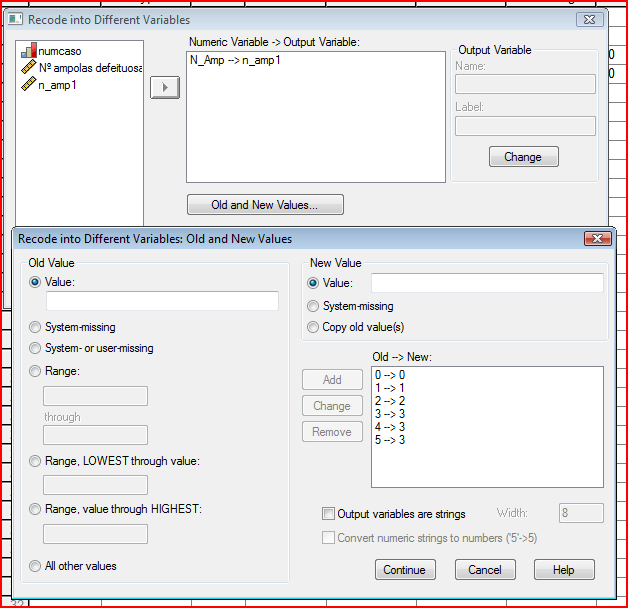
Como juntar categorias:

Uma hipótese será juntar as duas últimas categorias na categoria “4ou +”. A sua freq esperada seria 3,8 (que é inferior a 5). Então teríamos 1 em 5 categorias (20%) com valor esperado inferior a 5. Viola as condições de aplicabilidade. Logo vamos ter que agregar as 3 últimas categorias.

|  |  |  |
| --- | --- | --- |
|  | | Freq esperadas |
| Valid | 0 | 200\*0,3679=73,58 |
|  | 1 | 73,58 |
|  | 2 | 36,78 |
|  | 3 ou + | 16,06 |
|  | Total | 200 |

Agora já temos as condições de aplicabilidade verificvadas logo podemos utilizar o teste do qui-quadrado.

Não esquecer que temos que criar no spss uma nova variável em que as 3 últimaas categorias estão juntas. (Tranform -> Recode into Different Variables)



Fazendo o teste do qui-quadrado à nova variável, temos:

**n\_amp1**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed N | Expected N | Residual |
| 0 | 66 | 73,6 | -7,6 |
| 1 | 84 | 73,6 | 10,4 |
| 2 | 38 | 36,8 | 1,2 |
| 3 ou + | 12 | 16,1 | -4,1 |
| Total | 200 |  |  |

**Test Statistics**

|  |  |
| --- | --- |
|  | n\_amp1 |
| Chi-Square(a) | 3,323 |
| df | 3 |
| Asymp. Sig. | ,344 |
| Exact Sig. | ,346 |
| Point Probability | ,000 |

a 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 16,1.

Sig=0,346 > alfa=0,05 logo não se rejeita H0.

Portanto o responsável pelo controlo de qualidade tem razão. (com alfa de 0,05)

**Teste de independência do Qui-quadrado**

Ex9 pág.193

N=1000

|  |  |  |  |
| --- | --- | --- | --- |
| Sexo | expresso | semanário | independente |
| F | 150 | 50 | 150 |
| M | 350 | 200 | 100 |

Testar se a preferência pelos vários semanários é influenciada pelo sexo dos leitores

Alfa =0.05

Teste de independência do qui-quadrado

H0: A preferência pelos semanários é independente do sexo dos leitores (A HIPÓTESE NULA É SEMPRE A HIPÓTESE DA INDEPENDÊNCIA)

H1: A preferência pelos semanários dependente do sexo dos leitores

Cálculo dos valores esperados

Analyse->Descriptive Statistics->Crosstabs

| **Género \* Semanário preferido Crosstabulation** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | Semanário preferido | | | Total |
| Expresso | Semanário | Independente |
| Género | Feminino | Count | 150 | 50 | 150 | 350 |
| Expected Count | 175,0 | 87,5 | 87,5 | 350,0 |
| % within Género | 42,9% | 14,3% | 42,9% | 100,0% |
| % within Semanário preferido | 30,0% | 20,0% | 60,0% | 35,0% |
| Masculino | Count | 350 | 200 | 100 | 650 |
| Expected Count | 325,0 | 162,5 | 162,5 | 650,0 |
| % within Género | 53,8% | 30,8% | 15,4% | 100,0% |
| % within Semanário preferido | 70,0% | 80,0% | 40,0% | 65,0% |
| Total | | Count | 500 | 250 | 250 | 1000 |
| Expected Count | 500,0 | 250,0 | 250,0 | 1000,0 |
| % within Género | 50,0% | 25,0% | 25,0% | 100,0% |
| % within Semanário preferido | 100,0% | 100,0% | 100,0% | 100,0% |

Ao analisar as freq. esperadas vemos que são todas superiores a 5, logo as condições de aplicabilidade verificam-se.

O SPSS também nos dá informação sobre as freq. Esperadas para podermos concluir sobre as condições de aplicabilidade. Essa informação vem no rodapé do quadro do teste.

| **Chi-Square Tests** | | | |
| --- | --- | --- | --- |
|  | Value | df | Asymp. Sig. (2-sided) |
| Pearson Chi-Square | 98,901a | 2 | ,000 |
| Likelihood Ratio | 97,322 | 2 | ,000 |
| Linear-by-Linear Association | 48,902 | 1 | ,000 |
| N of Valid Cases | 1000 |  |  |
| a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 87,50. | | | |

Sig=0,000<alfa =0,05 rejeita-se H0

Logo a preferência pelos semanários depende do sexo dos leitores.

Na tabela verificamos que os homens preferem o expresso e o semanário e as mulheres preferem o independente (ver as percentagens em linha e em coluna).

Ex11, pág. 194

Pretende-se verificar se o nível de absentismo está relacionado com o facto do trabalhador ser do sexo masc ou fem.

Teste de independência do qui-quadrado

H0: O nível de absentismo é independente do sexo

H1: O nível de absentismo depende do sexo

Quadro com as frequências observadas:

| **Nível de absentismo \* Género Crosstabulation** | | | | |
| --- | --- | --- | --- | --- |
| Count | | | | |
|  | | Género | | Total |
| Masculino | Mulheres |
| Nível de absentismo | Mto baixo | 5 | 20 | 25 |
| Baixo | 10 | 14 | 24 |
| Médio | 8 | 22 | 30 |
| Elevado | 3 | 10 | 13 |
| Mto elevado | 4 | 4 | 8 |
| Total | | 30 | 70 | 100 |

Quadro com as frequências observadas e esperadas:

| **Nível de absentismo \* Género Crosstabulation** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | | Género | | Total |
| Masculino | Mulheres |
| Nível de absentismo | Mto baixo | Count | 5 | 20 | 25 |
| Expected Count | 7,5 | 17,5 | 25,0 |
| Baixo | Count | 10 | 14 | 24 |
| Expected Count | 7,2 | 16,8 | 24,0 |
| Médio | Count | 8 | 22 | 30 |
| Expected Count | 9,0 | 21,0 | 30,0 |
| Elevado | Count | 3 | 10 | 13 |
| Expected Count | 3,9 | 9,1 | 13,0 |
| Mto elevado | Count | 4 | 4 | 8 |
| Expected Count | 2,4 | 5,6 | 8,0 |
| Total | | Count | 30 | 70 | 100 |
| Expected Count | 30,0 | 70,0 | 100,0 |

2 células (em 10, o que corresponde a 20%) com freq esperada inferior a 5. Atenção que as condições de aplicabilidade não se verificam.

Para resolver o problema vamos ter que juntar categorias. Neste caso juntamos a categoria “elevado” com a categoria “mto elevado”.

Criar no spss uma nova variável (através do recode into different variable) que tem estas duas últimas categorias juntas numa só.

| **Nível de absentismo \* Género Crosstabulation** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | | Género | | Total |
| Masculino | Mulheres |
| Nível de absentismo | Mto baixo | Count | 5 | 20 | 25 |
| Expected Count | 7,5 | 17,5 | 25,0 |
| Baixo | Count | 10 | 14 | 24 |
| Expected Count | 7,2 | 16,8 | 24,0 |
| Médio | Count | 8 | 22 | 30 |
| Expected Count | 9,0 | 21,0 | 30,0 |
| Elevado e Mto Elevado | Count | 7 | 14 | 21 |
| Expected Count | 6,3 | 14,7 | 21,0 |
| Total | | Count | 30 | 70 | 100 |
| Expected Count | 30,0 | 70,0 | 100,0 |

Todas as células têm valor esperado acima de 5, pelo que as condições de aplicabilidade agora estão verificadas.

| **Chi-Square Tests** | | | |
| --- | --- | --- | --- |
|  | Value | df | Asymp. Sig. (2-sided) |
| Pearson Chi-Square | 3,016a | 3 | ,389 |
| Likelihood Ratio | 3,023 | 3 | ,388 |
| Linear-by-Linear Association | ,339 | 1 | ,560 |
| N of Valid Cases | 100 |  |  |
| a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 6,30. | | | |

Sig=0,389> alfa=0,05 não se rejeita a hipótese de independência

Logo o nível de absentismo não está relacionado com o facto do trabalhador ser do sexo masc ou fem, é independente do sexo.