Dr. Quapp: Statistik für Mathematiker mit SPSS

## Hinweise zur 15. Übung No 1 - Poissonverteilung -

**1.] In einem Saal mit 10 Maschinen werde die Anzahl ausgefallener Maschinen registriert. Bei 200 Kontrollen ergaben sich folgende Werte:**

| **Anzahl Ausfall** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Absolute Häufigkeit $h_m$** | **41** | **62** | **45** | **22** | **16** | **8** | **4** | **2** | **0** | **0** | **0** |

**Es wird vermutet, daß die Anzahl eine Poisson-verteilte Zufallsgröße ist, da unter normalen Produktionsbedingungen ein Maschinenschaden ein seltenes Ereignis ist, das auch vom Ausfall einer anderen Maschine unabhängig ist. Geben Sie eine Schätzung für den Parameter der Poisson-Verteilung an. Überprüfen Sie, ob die Anzahl der ausgefallenen Maschinen als eine mit dem geschätzten Parameter Poisson-verteilte Zufallsgröße angesehen werden kann.**
**Hinweis: $\chi^2$ - Test**

Man kann zwei Versionen der Lösung der Aufgabe angehen: einmal die Formeln des $\chi^2$ - Tests direkt berechnen, oder diesen $\chi^2$ - Test in SPSS aufrufen. Bei beiden Wegen braucht man den Vergleich mit der angenommenen theoretischen Poisson-Verteilung

$$H_0: \quad P(X=m) = \frac{\lambda^m}{m!}\, e^{-\lambda}, \quad m = 0, 1, 2, ...$$

für den Ausfall einer Maschine. Bei 200 Maschinen also die Werte $200\,P(X=m)$. Dazu muß als Schätzung der Parameter $\lambda$ bestimmt werden.

```
GET FILE='D:\spss151.sav'.
EXECUTE .
WEIGHT BY h_m .
FREQUENCIES VARIABLES=ausfall
  /STATISTICS=MEAN .
```

Die 11 Zeilen der Datentabelle werden durch die Wichtung der Variablen "Ausfall" mit deren absoluter Häufigkeit sozusagen auf die wirklichen 200 Fälle aufgebläht, in →Daten, → Fälle wichten. Mit einem einfachen Aufruf der beschreibenden Statistik kann dann der Mittelwert von Ausfall berechnet werden: 1,8.

```
AUSFALL                                              Valid     Cum
Value Label              Value  Frequency  Percent  Percent  Percent
                             0         41     20,5     20,5     20,5
                             1         62     31,0     31,0     51,5
                             2         45     22,5     22,5     74,0
                             3         22     11,0     11,0     85,0
                             4         16      8,0      8,0     93,0
                             5          8      4,0      4,0     97,0
                             6          4      2,0      2,0     99,0
                             7          2      1,0      1,0    100,0
                                  -------  -------  -------
                         Total        200    100,0    100,0
Mean           1,800  Valid cases     200       Missing cases      0
```

Der Mittelwert dient als Schätzung des Erwartungswertes der Poisson-Verteilung, der gleich $\lambda$ ist. Mit VerP=CDF.POISSON(ausfall, 1.8) ergibt sich die zugehörige theoretische Verteilungsfunktion, und in →Berechnen, →Zeitreihen kann man diese Variable differenzieren, um die Dichte "pm" zu erhalten. Mit pm200=pm*200 sind die theoretischen Häufigkeiten einer Poissonverteilung zu $\lambda = 1.8$ bereitgestellt. Die Test-Statistik für den $\chi^2$ - Test ist

$$\frac{(pm200 - h\_m)^2}{pm200}$$

welche noch über alle Zeilen kumulativ aufzusummieren ist: wieder in →Berechnen, →Zeitreihen mit der Funktion Kumulative Summe. Dabei ergibt sich eine Spalte dieser kumulativen Werte. Wird nun noch eine Spalte dagegengesetzt, die den kritischen Wert des $\chi^2$ - Tests zu $1 - \alpha = 0.95$ und 10 Freiheitsgraden berücksichtigt, was mit der inversen Verteilungsfunktion des $\chi^2$ - Tests machbar ist: $k_c$=IDF.CHI(0.95, 10), dann zeigt sich, daß in Zeile 8 dieser Wert 18,3 von der kumulatinen $\chi^2$-Summe überschritten wird. Also kann die Nullhypothese, daß die Ausfälle Poisson-verteilt zu $\lambda = 1.8$ sind, nicht aufrecht erhalten werden.

In der **zweiten Version** der Lösung wird der $\chi^2$ - Test direkt in SPSS aufgerufen, in →Nichtparametrische Tests. Dort ist die Variable *ausfall* zu verwenden, aber im Gegensatz zur Würfelaufgabe von Serie 12, No 3, haben wir hier nun in den einzelnen Zeilen von *ausfall* verschiedene Wahrscheinlichkeiten zu erwarten! Dies muß direkt Zeile für Zeile eingetragen werden. Die Werte für die Häufigkeiten der Poissonverteilung zu $\lambda = 1.8$ sind schon in der Datentabelle mit $pm200$ berechnet. Vorher sind in der Datentabelle diejenigen Zeilen zu löschen, die Nullhäufigkeit haben, das betrifft die Werte 8, 9 und 10 für *ausfall*.

```
WEIGHT   BY h_m .
NPAR TEST
  /CHISQUARE=ausfall
  /EXPECTED=33.06 59.51 53.56 32.13 14.46 5.21 1.56 0.4
  /MISSING ANALYSIS.
```

Es erfolgt eine Fehlermeldung: 2 Zellen haben erwartete Häufigkeiten kleiner als 5. Diese Voraussetzung war im $\chi^2$ - Test gefordert: Man muß somit noch in der Datentabelle weiter zusammenfassen, etwa die Zeilen 5, 6, und 7 zu einer Zeile mit dem durchschnittlichen *ausfall* 6, und der beobachteten absoluten Häufigkeit 14, und der theoretischen Häufigkeit pm200(5+6+7)=7.17.

```
NPAR TEST
  /CHISQUARE=ausfall
  /EXPECTED=33.06 59.51 53.56 32.13 14.46 7.17
  /MISSING ANALYSIS.
```

Dieser Test liefert letztendlich das erwartete Resultat

```
- - - - - Chi-Square Test AUSFALL
                 Cases
     Category  Observed  Expected  Residual
            0        41     33,08      7,92
            1        62     59,54      2,46
            2        45     53,59     -8,59
            3        22     32,15    -10,15
            4        16     14,47      1,53
            6        14      7,17      6,83
                    ---
        Total       200
          Chi-Square           D.F.         Significance
            13,2358              5              ,0213
```

da die " Significance" 0,02 kleiner als die geforderte 0,05 ist.

Lädt man die gelöschten Zeilen neu, kann man sich auch noch ein Bild der Verhältnisse verschaffen: In →LinienPlot ist → mehrfach anzuklicken, und • Werte einzelner Fälle.

```
GRAPH
 /LINE(MULTIPLE)= VALUE( h_m pm200 ) BY ausfall .
```

Obwohl die Kurven sich durchaus ähnlich sehen, kann man für *h_m* nicht die Poissonverteilung annehmen.
Bemerkung: Würde man den KS-Test auf Poissonverteilung verwenden, bekäme man eine positive Aussage für *h_m*, und diese mit sehr hoher Signifikanz. Dies zeigt aber nur, daß der KS-Test sehr grob ist. (Und wenn er denn zu einer Ablehnung führt, dann müssen die Daten in der Tat sehr "schlecht" gewesen sein.)