Statistik-Ü 2.KW (WS 2014/2015) Prof.Dr.L.Paditz
(Dokument als eActivity im ClassPad400 erstellt)

Liebe Studenten,
hier finden Sie noch einmal den vollständigen Lösungsweg für ausgewählte Ü-Aufg. mit dem ClassPad:

STAT D 10:
===========
Nutzung des Befehls **TwoPropZTest:**
(Syntax des Befehls beachten)

Faustregel $n*p*(1-p)\geq 9$ erfüllt:
$n_1*\hat{p}_1*(1-\hat{p}_1)=$

$500*\frac{152}{500}*(1-\frac{152}{500})$

105.792

$n_2*\hat{p}_2*(1-\hat{p}_2)=$

$200*\frac{78}{200}*(1-\frac{78}{200})$

47.58

TwoPropZTest mit N(0,1)-Prüfverteilung kann angewendet werden:
TwoPropZTest "<",152,500,78,200

done

DispStat

done

================================
Z-Test (2 Wktn.)

$p_1<p_2$

$z = -2.18844$ (t-Wert)

prob = 0.0143188 (p-Wert < α=0.05)

$\hat{p}_1 = 0.304$ (=152/500)

$\hat{p}_2 = 0.39$ (= 78/200)

$\hat{p} = 0.3285714$ (=(152+78)/700)

$n_1 = 500$

$n_2 = 200$

==============================

STAT E 1:

==========

Bem.: nach DIN53804(Teil1) werden $\sqrt{n}=\sqrt{100}=10$ Klassen empfohlen.
Da die Urdatenliste nicht vorliegt, wird mit der vorgeg. Klasseneinteilung gearbeitet.

Die ML-Schätzung nach B10 ergibt

$\mu=\bar{x}=1485.1$ und $\sigma^2=\frac{n-1}{n}s^2=\frac{99}{100}869.8^2$, d.h.

$\mu:=1485.1$

1485.1

$\sigma:=\sqrt{\frac{99}{100}869.8^2}$

865.4400728

n*pj-Werte (Faustregel $n*p_j \geq 5$ erfüllt):

100*normCDf(-∞, 500, σ, μ)⇒np1

12.75049569

100*normCDf(500, 1000, σ, μ)⇒np2

16.00560531

100*normCDf(1000,2000,σ,μ)⇒np3

43.65030259

100*normCDf(2000,∞,σ,μ)⇒np4

27.59359641

{16,18,28,38}⇒listh

{16,18,28,38}

{np1,np2,np3,np4}⇒listnp

{12.75049569,16.00560531,43.65030259,27.59359641}

Nutzung des Befehls **ChiGOFTest:**

(Syntax des Befehls beachten)

ChiGOFTest listh,listnp,1

done

DispStat

done

==============================

χ^2 GOF Test

χ^2 = 10.612471 (t-Wert)

prob = 1.1233E-3 (p-Wert < α=0.05)

df = 1

==============================

Quantil für krit. Bereich:

invChiCDf(0.05,1)

3.841458821

Rechnung "per Hand" über die N(0,1)-Verteilung:

z.B. für 3.Klasse [1000,2000):

$100*normCDf(\frac{1000-\mu}{\sigma},\frac{2000-\mu}{\sigma},1,0) \Rightarrow np3$

43.65030259

$\frac{1000-\mu}{\sigma}$

-0.5605240793

$\frac{2000-\mu}{\sigma}$

0.5949574282

Tabellenwerte der N(0,1)-Verteilung:

normCDf(-∞,-0.5605240793,1,0)

0.28756101

normCDf(-∞,0.5949574282,1,0)

0.7240640359

Intervallwkt.:

0.7240640359-0.28756101

0.4365030259

ans*100

43.65030259

usw.

STAT E 4:

==========

Poisson-Vert. mit Parameter λ, geschätzt durch $\bar{x}$

1. Hypothesen:

   Ho: $P(X=i)=e^{-\bar{x}}*\frac{\bar{x}^i}{i!}$ $\forall i \in N$, H_1: $\exists i$: $P(X=i) \neq e^{-\bar{x}}*\frac{\bar{x}^i}{i!}$

2. $\alpha=0.05$

3. $T=\sum_{i=0}^{5}\left(\frac{(h_i-n*p_i)^2}{n*p_i}\right)$ ist unter Ho näherungsweise χ^2-vert.

mit 6-1-1=4 FG

Klassenzusammenfassung (rechts) wegen $n*p_i \geq 5$

4. $K=(\chi^2_{4,0.95}\ ,\ \infty)$

5. Entsch. wie üblich

Quantil:

invChiCDf(0.05,4)

9.487729037

{211,267,160,65,21,4,2}⇒listh

{211,267,160,65,21,4,2}

seq(i,i,0,6,1)⇒listx

{0,1,2,3,4,5,6}

mean(listx,listh)⇒λ

1.230136986

sum(listh)

730

$n*p_j$-Werte (Faustregel $n*p_j \geq 5$ erfüllt?):

730*poissonPDf(0,λ)⇒np0

213.3443545

730*poissonPDf(1,λ)⇒np1

262.4427812

730*poissonPDf(2,λ)⇒np2

161.420286

730*poissonPDf(3,λ)⇒np3

66.18968804

```
730*poissonPDf(4,λ)→np4
                                              20.35559584
730*poissonPDf(5,λ)→np5
                                              5.008034265
730*(1-poissonCDf(0,5,λ))→np6
                                              1.239260177
```

rechte Randklasse zu schwach besetzt!

Klassenzusammenfassung:

```
730*(1-poissonCDf(0,4,λ))→np5
                                              6.247294442
{np0,np1,np2,np3,np4,np5}→listnp
    {213.3443545,262.4427812,161.420286,66.18968804,20.▶
{211,267,160,65,21,6}→listh
                                    {211,267,160,65,21,6}
ChiGOFTest listh,listnp,4
                                                     done
DispStat
                                                     done
```

==============================

χ^2 GOF Test

χ^2 = 0.1689646 (t-Wert)

prob = 0.9966261 (p-Wert > α=0.05)

df = 4

==============================

Quantil für krit. Bereich:

```
invChiCDf(0.05,4)
                                              9.487729037
```

Ergebnis:

kein Einwand gegen Ho, d.h. mit einer Irrtumswkt. von 5% besteht kein Einwand gegen die Vermutung einer Poisson-Vert. mit dem Parameter λ=1,230.
(Mit einer Sicherheitswkt. von 95% konnte nicht nachgewiesen werden, dass vermutlich keine Poisson-Vert. mit λ=1,230 vorliegt)

STAT E 5:
==========
Kolmogorov-Test
µ:=1500

1500

σ:=5

5

1. Ho: $F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)\ \forall x\in\mathbf{R}$, H_1: $\exists x$: $F(x)\neq\Phi\left(\frac{x-\mu}{\sigma}\right)$
2. α=0.01
3. $T=\sup|F_n(x)-\Phi\left(\frac{x-\mu}{\sigma}\right)|$ ist unter Ho Kolmogorov-verteilt
4. K=(0,369 , ∞)
5. Entsch. wie üblich: t=0.288∉K, d.h. kein Einwand gegen Ho.

Treppenfunktion $F_n(x)$ (empir. Vert.-fkt.)
{1501,1496,1487,1506,1503,1490,1495,1502,1499,1495}▶
{1501,1496,1487,1506,1503,1490,1495,1502,1499,1495}
sortA(listx)→listx
{1487,1490,1495,1495,1496,1499,1501,1502,1503,1506}

$$\text{Define } y1(x)=\begin{cases} 0, & x<1487 \\ .1, & 1487\le x<1490 \\ .2, & 1490\le x<1495 \\ .4, & 1495\le x<1496 \\ .5, & 1496\le x<1499 \\ .6, & 1499\le x<1501 \\ .7, & 1501\le x<1502 \\ .8, & 1502\le x<1503 \\ .9, & 1503\le x<1506 \\ 1, & x\ge 1506 \end{cases}$$

done

Define $y2(x)=\text{normCDf}(-\infty, x, \sigma, \mu)$

done

Define $y3(x)=|y1(x)-y2(x)|$

done

y3(x)=abs(y1(x)-y2(x))

Lösung aus der Grafik ablesbar:

y3(1496)

0.2881446014

Einzelwerte in den Sprungstellen (von rechts):

$$\begin{bmatrix} y3(1487) \\ y3(1490) \\ y3(1495) \\ y3(1496) \\ y3(1499) \\ y3(1501) \\ y3(1502) \\ y3(1503) \\ y3(1506) \end{bmatrix}$$

$$\begin{bmatrix} 0.09533881198 \\ 0.1772498681 \\ 0.2413447461 \\ 0.2881446014 \\ 0.1792597094 \\ 0.1207402906 \\ 0.1445782584 \\ 0.1742531178 \\ 0.1150696702 \end{bmatrix}$$

Einzelwerte in den Sprungstellen (von links):

$\varepsilon := 0.000001$

$1E{-6}$

$$\begin{bmatrix} y3(1487-\varepsilon) \\ y3(1490-\varepsilon) \\ y3(1495-\varepsilon) \\ y3(1496-\varepsilon) \\ y3(1499-\varepsilon) \\ y3(1501-\varepsilon) \\ y3(1502-\varepsilon) \\ y3(1503-\varepsilon) \\ y3(1506-\varepsilon) \end{bmatrix}$$

$$\begin{bmatrix} 4.661185307E{-3} \\ 0.07724987885 \\ 0.04134479446 \\ 0.1881446594 \\ 0.07925978765 \\ 0.02074036877 \\ 0.04457833204 \\ 0.0742531844 \\ 0.01506970906 \end{bmatrix}$$

$t=0.288 \notin K$, d.h. kein Einwand gegen Ho.

(Mit einer Sicherheitswkt. von 90% konnte nicht nachgewiesen werden, dass vermutlich keine Normalvert. mit μ=1500 und σ=5 vorliegt.)

stop

STAT E 6:

==========

Kolmogorov–Test

λ:=1/200

5E-3

1•Ho

2. α=0.05

3. $T=\sup|F_n(x)-1-e^{-\lambda x}|$ ist unter Ho Kolmogorov–verteilt

4. K=(0,454 , ∞)

5. Entsch. wie üblich: t=0.4701$\in$K, d.h. Ablehn. von Ho.

Treppenfunktion $F_n(x)$ (empir. Vert.–fkt.)

{301,127,191,241,135,170,224,205}⇒listx

{301,127,191,241,135,170,224,205}

sortA(listx)⇒listx

{127,135,170,191,205,224,241,301}

dim(listx)⇒n

8

$$\text{Define } y1(x)=\begin{cases} 0, & x<127 \\ 1/n, & 127\le x<135 \\ 2/n, & 135\le x<170 \\ 3/n, & 170\le x<191 \\ 4/n, & 191\le x<205 \\ 5/n, & 205\le x<224 \\ 6/n, & 224\le x<241 \\ 7/n, & 241\le x<301 \\ 1, & 301\le x \end{cases}$$

done

$$\text{Define } y2(x)=\begin{cases} 1-e^{-\lambda x}, & x\ge 0 \\ 0, & x<0 \end{cases}$$

done

$$\text{Define } y3(x)=|y1(x)-y2(x)|$$

done

y3(x)=abs(y1(x)-y2(x))

Lösung aus der Grafik ablesbar:

y3(126.999999)

0.470064509

Einzelwerte in den Sprungstellen (von rechts):

$$\begin{bmatrix} y3(127) \\ y3(135) \\ y3(170) \\ y3(191) \\ y3(205) \\ y3(224) \\ y3(241) \\ y3(301) \end{bmatrix}$$

$$\begin{bmatrix} 0.3450645117 \\ 0.2408435794 \\ 0.1975850681 \\ 0.1151878554 \\ 0.01620353459 \\ 0.07627979462 \\ 0.1746919995 \\ 0.2220172938 \end{bmatrix}$$

Einzelwerte in den Sprungstellen (von links):

ε:=0.000001

1E-6

$$\begin{bmatrix} y3(127-\varepsilon) \\ y3(135-\varepsilon) \\ y3(170-\varepsilon) \\ y3(191-\varepsilon) \\ y3(205-\varepsilon) \\ y3(224-\varepsilon) \\ y3(241-\varepsilon) \\ y3(301-\varepsilon) \end{bmatrix}$$

$$\begin{bmatrix} 0.470064509 \\ 0.3658435768 \\ 0.3225850659 \\ 0.2401878535 \\ 0.1412035328 \\ 0.04872020375 \\ 0.04969200101 \\ 0.09701729494 \end{bmatrix}$$

stop

STAT E 8:

==========

χ^2–Unabhängigkeitstest:

sei $A=\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$

1. Ho: Geschlecht und Rauchergewohnheit sind (stochast.) unabhängig

   H_1: Rauchergewohnheit ist geschlechtsspezifisch
2. $\alpha=0.05$
3. $T=n*\frac{(\det(A))^2}{H_{1.}H_{2.}H_{.1}H_{.2}}$ ist unter Ho χ^2-verteilt mit 1 FG
4. $K=\{t \mid t>\chi^2_{1,0.95}\}$
5. Entsch. wie üblich: $t \notin K$, d.h. kein Einwand gegen Ho (Geschlechtsspezifik mit 95% Sicherheit nicht nachweisbar anhand der ausgewerteten Daten)

$A:=\begin{bmatrix} 32 & 76 \\ 34 & 58 \end{bmatrix}$

$$\begin{bmatrix} 32 & 76 \\ 34 & 58 \end{bmatrix}$$

Nutzung des Befehls **ChiTest:**

ChiTest A

done

DispStat

done

====================

χ^2 Test

χ^2 = 1.2062364 (t-Wert)

prob = 0.2720788 (p-Wert>α=0.05)

df = 1

====================

Matrix der erwarteten Häufigkeiten im Fall der Stoch. Unabhängigkeit:

$H_{ij} = H_{i}.H_{.j} \quad \forall\ i,j$

Systemvariable "Expected" nutzen:

Expected

$$\begin{bmatrix} 35.64 & 72.36 \\ 30.36 & 61.64 \end{bmatrix}$$

per Hand:

$$200*\frac{\det(A)^2}{108*92*66*134}$$

1.206236421

Quantil:

invChiCDf(0.05,1)

3.841458821

STAT E 9:

==========

χ^2–Unabhängigkeitstest:

$$A:=\begin{bmatrix} 25 & 41 & 23 \\ 36 & 91 & 47 \\ 37 & 64 & 33 \\ 30 & 36 & 37 \end{bmatrix}$$

$$\begin{bmatrix} 25 & 41 & 23 \\ 36 & 91 & 47 \\ 37 & 64 & 33 \\ 30 & 36 & 37 \end{bmatrix}$$

ChiTest A

done

DispStat

done

=====================

χ^2 Test

χ^2 = 9.8984943 (t-Wert)

prob = 0.1289922 (p-Wert>α=0.05)

df = 6

=====================

Expected

$$\begin{bmatrix} 22.784 & 41.296 & 24.92 \\ 44.544 & 80.736 & 48.72 \\ 34.304 & 62.176 & 37.52 \\ 26.368 & 47.792 & 28.84 \end{bmatrix}$$

fRound(ans, 1)

$$\begin{bmatrix} 22.8 & 41.3 & 24.9 \\ 44.5 & 80.7 & 48.7 \\ 34.3 & 62.2 & 37.5 \\ 26.4 & 47.8 & 28.8 \end{bmatrix}$$

Ergebnis:

kein Einwand gegen Ho, d.h. mit 95% Sicherheit ist eine Anhängigkeit zwischen den betrachteten zwei Merkmalen nicht nachweisbar (anhand der ausgewerteten Daten)