Statistik-Vorlesung 47.KW (17.11.2014) Prof.Dr.L.Paditz
(Dokument als eActivity im ClassPad400 erstellt)

Liebe Studenten,
zuerst gebe ich Ihnen ein Beispiel zum Korrelationstest:
Für n=10 Baustellen ergaben sich folgende Werte der
Jahresgesamtkosten Y und der Jahresproduktion X (in 10^5€):

xliste:={12,15,18,18,20,21,24,25,36,37}

{12,15,18,18,20,21,24,25,36,37}

yliste:={11,12,16,17,18,18,20,21,26,31}

{11,12,16,17,18,18,20,21,26,31}

Mit einem Korrelationstest ist die (lineare) Abhängigkeit von Y
zu X zu untersuchen.

Zwischenschritt: lineare Regression y=a+b·x

LinearReg xliste,yliste,1,y1,On

done

DispStat

done

==========

Lineare Regression

y = a+b·x
a = 2.9409474
b = 0.7105775
r = 0.9770887 (Korr.-koeff.)
r^2 = 0.9547023 (Bestimmtheitsmaß B)
MSe= 1.8458793 (Mean Square Error)

==========

Korrelationstest:

LinRegTTest "≠", xliste, yliste, 1

done

DispStat

done

==========

t-Test für lin. Reg.

$\hat{y}$ = a+b·x (Schätzung für $y=\alpha+\beta \cdot x$)

ß≠0 & ρ≠0 (Art der Alternative)

t = 12.984976 (Testgröße T=t)

prob = 1.1727E-6 (p-Wert)

df = 8 (Freiheitsgrade von T)

a = 2.9409474 (Schätzung für α, Absolutglied)

b = 0.7105775 (Schätzung für ß, Anstieg)

s = 1.3586314 (Standardfehler der MKQ-Schätzung der Regressionsgeraden)

r = 0.9770887 (Korr.-koeff.)

r^2 = 0.9547023 (Bestimmtheitsmaß B)

SEb = 0.0547231 (Standardfehler des Anstiegs bei MKQ-Schätzung)

==========

Hinw.: nicht alle angezeigten Kennzahlen werden für die Testdurchführung benötigt.

Sei $\alpha=0.05$. Wegen p=1.1727E-6<$\alpha=0.05$ ist H_0: ρ=0 abzulehnen.

per Hand:

für K benötigtes t_{n-2}-Quantil: mit $1-\alpha/2$:

invTCDf(0.025,8)

2.306004135

$T:=\sqrt{10-2}*\frac{0.9770887}{\sqrt{1-0.9770887^2}}$

12.98497271

LinearReg xliste,yliste,1,y1,On

done

residual

{-0.4678780013,-1.599610642,0.2686567164,1.26865671▶

$s:=\sqrt{\frac{sum(residual^2)}{10-2}}$

1.358631407

Standardfehler der MKQ-Schätzung der Regressionsgeraden
(Wurzel aus der Reststreuung, mit n-2 normiert)

Berechnung von SEb:

sum((xliste-mean(xliste))^2)⇒SSX

616.4

sum(residual^2)⇒SSD

14.76703439

$SEb:=\sqrt{\frac{1}{10-2}*\frac{SSD}{SSX}}$

0.05472305501

Vorl.-Beisp. 12:

==========

Wahlprognose n=2000, H_n=136 für Partei A.

1. H_0: $p=p_0=0.05$ H_1: $p>p_0$

2. $\alpha = 0.01$ (d.h. $1-\alpha = 0.99=99\%$)

3. T mit N(0,1)-Prüfverteilung (nach ZGWS: Faustregel: $np(1-p) \approx 126,8 > 9$ erfüllt)

$$2000 * \frac{136}{2000} * \left(1 - \frac{136}{2000}\right)$$

126.752

$$t := \frac{\frac{136}{2000} - 0.05}{\sqrt{0.05(1-0.05)}} \sqrt{2000}$$

3.693522068

4. z-Quantil mit $1-\alpha$:

invNormCDf("L",0.99,1,0)

2.326347874

5. Entscheidung: $t > 2.326347874$, d.h. Ablehnung von H_0: Wird bei der Wahl so wie in der Prognose gewählt, wird mit einer statistischen Sicherheit von 99% die 5%-Hürde übersprungen und Partei A kann in den Landtag einziehen.

Mit einem Statistik-Befehl:

OnePropZTest ">",0.05,136,2000

done

DispStat

done

==========

Z-Test(1 Wkt.)

Prop > 0.05 (Art der Alternative)

z = 3.6935221 (t-Wert)

prob = 1.1058E-4 (p-Wert)

$\hat{p}$ = 0.068 (geschätzter Anteilswert 6,8%)

n = 2000

==========

Vorl.-Beisp. 13:

==========

Hersteller behauptet: Ausschußquote max. 4%

Verbraucherschutz zweifelt und führt einen Test durch: n=40 (kleiner Stichprobenumfang, ZGWS nicht anwendbar, Faustregel nicht erfüllt, s.u.)

Stichprobe enthielt 4 Ausschußstücke: Fehlerquote hier $\frac{4}{40}$=10%.

(ZGWS Faustregel: np(1−p)=40∗0.1∗(1−0.1)=3,6<9)

Test für unbekannten Anteilswert p:

1. H_0: $p=p_0=0.04$ H_1: $p>p_0$
2. $\alpha = 0.05$ (d.h. $1-\alpha = 0.95=95\%$)
3. T mit exakter $B(n,p_0)$-Prüfverteilung: $T=H_n$ mit Realisierung t=4
4. Forderung an K: $P(T\in K)\leq\alpha$ und $P(T\notin K)\geq 1-\alpha$
5. Entscheidung:

$B(n,p_0)$-Prüfverteilung tabellieren:

Intervallwktn. für das Intervall [m,40] berechnen:

n:=40

40

p_0:=0.04

0.04

seq(binomialCDf(m,n,n,p_0),m,0,6,1)

{1,0.8046338484,0.4790235959,0.2144652656,0.0748372▶

listToMat(ans)

$$\begin{bmatrix} 1 \\ 0.8046338484 \\ 0.4790235959 \\ 0.2144652656 \\ 0.074837258 \\ 0.02102229673 \\ 4.877808347\text{E-}3 \end{bmatrix}$$

Für m=5 wird das Niveau α=0.05 erstmals unterschritten.
Damit ist K=(4,40]=[5,40], um die Forderung
$P(T \in K) \leq \alpha = 0.05$ einzuhalten.
T ist die zufällige Anzahl der Ausschußstücke in der Stichprobe (Bernoulli-Schema mit p_0=0.04 beachten)
Mit t=4 wird der kritische Bereich K knapp verfehlt, d.h. H_0 kann durch die Verbraucherschützer nicht abgelehnt werden auf Grundlage der erhobenen Daten und dem gewählten Signifikanzniveau α=0.05.

Bem.: B(n, p_0)-Quantil direkt berechnen:
invBinomialCDf(0.95, 40, 0.04)

4

Quantildefinition war: $P(T<\text{Quantil}) \leq 1-\alpha \leq P(T \leq \text{Quantil})$
Kontrolle:
binomialCDf(0, 4, 40, 0.04)

0.9789777033

binomialCDf(0, 3, 40, 0.04)

0.925162742

d.h. für T=5 wird die Nichtablehnungs-Wahrscheinlichkeit $1-\alpha$ erstmals unterschritten und damit $P(T \notin K) \geq 1-\alpha = 0.95$ verletzt.

Diskussion:

im Fall H_0: $p=p_0$ H_1: $p<p_0$ liegt der kritische Bereich links.

im Fall H_0: $p=p_0$ H_1: $p\neq p_0$ ist der kritische Bereich zweiseitig:

$K=[0,m_1]\cup[m_2,n]$ mit $P(T\in K)\leq\alpha$

Das Auffinden der benötigten Quantile wird aufwändiger.