Zusammenschrift zur Normalverteilung

8. April 2015

1 Normalverteilung

1.1 Bestimmte Wahrscheinlichkeit

$$dnorm(x, \mu, \sigma) = \text{Wahrscheinlichkeit für einen bestimmten Wert} \tag{1}$$

1.2 Kumulative Wahrscheinlichkeit

$$pnorm(x, \mu, \sigma) = \text{Wahrscheinlichkeit für höchstens einen Wert} \tag{2}$$

1.3 Inverse kumulative Wahrscheinlichkeit

$$qnorm(p, 0, 1) = x = \text{Wert für eine bestimmte Wahrscheinlichkeit} \tag{3}$$

2 Verschiebung zur Standardnormalverteilung

$$u = \frac{x - \mu}{\sigma} = qnorm(p, 0, 1) \tag{4}$$

3 Normalverteilung und Stichproben

Wenn eine Stichprobe mit n Werten von einer Grundgesamtheit genommen wird, dann verändern sich μ und σ, die eigentlich für die Grundgesamtheit gelten, folgendermaßen:

$$\overline{\mu} = \mu \tag{5}$$

$$\overline{\sigma} = \frac{\sigma}{\sqrt{n}} \tag{6}$$

4 Annäherung der Binomialverteilung zur Normalverteilung

Es gilt:

$$\mu = n * p \tag{7}$$

$$\sigma = \sqrt{n * p * (1 - p)} \tag{8}$$

Es muss erkannt werden, dass die gegebenen Daten binomialverteilt sind! Dies ist gegeben, wenn die Daten immer nur Ja/Nein Ergebnisse sind und bei einer Wiederholung sich die Wahrscheinlichkeit nicht ändert.

5 Suchen nach Variablen anhand einer Binomialverteilung

5.1 Grundlegendes

Mit den im Kapitel section 1 - Normalverteilung beschriebenen Funktionen können nun folgende Werte gesucht werden:

x Der bestimmte Wert von einer Grundgesamtheit

n Die Grundgesamtheit

p Die Wahrscheinlichkeit, wieviel von der Grundgesamtheit y eintritt ($y = n * p$)

p_2 Die Wahrscheinlichkeit einen Wert x von der Grundgesamtheit n zu erhalten

Wichtig ist, dass die folgenden Vorgehensweisen sich nicht auf die Binomialverteilung beschränkt. Es muss nur μ und σ gegeben sein! Für all die nachstehenden Suchen sind diese beide definiert durch:

- $\mu = p * n$
- $\sigma = \sqrt{n * p * (1 - p)}$

5.2 Suche nach x

Gegebene Variablen:

- n
- p
- p_2

$$qnorm(p_2, 1, 0) = \frac{x - \mu}{\sigma} \text{solve, x} \rightarrow \tag{9}$$

5.3 Suche nach n

Gegebene Variablen:

- x
- p
- p_2

$$qnorm(p_2, 0, 1) = \frac{x - \mu}{\sigma} \text{solve, n} \rightarrow \tag{10}$$

5.4 Suche nach p

Gegebene Variablen:

- x
- n

Wenn es sich um eine Binomialverteilung handelt, dann gilt: $\sigma = \sigma_0$, da eine Wurzel gezogen wurde.

$$pnorm(x, \mu, \sigma) = p_2 \tag{11}$$

6 Zweiseitiger Zufallsstreubereich

6.1 Grundlegendes

Beim zweiseitigen Zufallsstreubereich wird ermittelt, welche Unter- und Obergrenze die Werte in einem bestimmten Bereich symmetrisch verteilt um den Erwartungswert besitzt.
Gegebene Variablen:

μ

σ

p Die Wahrscheinlichkeit (Größe) des Streubereichs symmetrisch um den Erwartungswert μ

6.2 Berechnung von $x_{u,o}$

$$\alpha = 1 - p \tag{12}$$

$$x_u = \mu - qnorm(1 - \frac{\alpha}{2}, 0, 1) * \sigma \tag{13}$$

$$x_o = \mu + qnorm(1 - \frac{\alpha}{2}, 0, 1) * \sigma \tag{14}$$

6.3 Berechnung von $x_{u,o}$ bei Stichproben

Es kommt nun zustätzlich die gegebene Variable der Stichprobenanzahl **n** hinzu.

$$\alpha = 1 - p \tag{15}$$

$$\overline{\sigma} = \frac{\sigma}{\sqrt{n}} \tag{16}$$

$$x_u = \mu - qnorm(1 - \frac{\alpha}{2}, 0, 1) * \overline{\sigma} \tag{17}$$

$$x_o = \mu + qnorm(1 - \frac{\alpha}{2}, 0, 1) * \overline{\sigma} \tag{18}$$

7 Konfidenzintervall

7.1 Grundlegendes

Das Konfidenzintervall gibt an, in welchen Intervall die Standardabweichung am ehestens bei gegebenen Stichproben liegt. Das bedeutet, dass σ selbst nicht ermittelt werden kann, jedoch der Bereich, in der sich σ bewegt.

Gegeben Variablen müssen sein:

m Ein einzeiliger Stichprobenvektor

s Die Standardabweichung s berechnet aus m errechnet mit $Stdev()$ bzw. $stdev()$

p Das Konfidenznevau (meistens 95 %)

7.2 Suche nach dem Konfidenzintervall

Es werden folgende Variablen gesucht:

σ_u Die untere Grenze des Vertrauensbereichs

σ_o Die obere Grenze des Vertrauensbereichs

Berechnung:

$$\alpha = 1 - p \tag{19}$$

$$p_u = 1 - \frac{\alpha}{2} \tag{20}$$

$$p_o = \frac{\alpha}{2} \tag{21}$$

$$\sigma_u = s * \sqrt{\frac{f}{qchisq(p_u, f)}} \tag{22}$$

$$\sigma_o = s * \sqrt{\frac{f}{qchisq(p_o, f)}} \tag{23}$$

Der p %ige Vertrauensbereich für σ liegt dann bei: $[\sigma_u; \sigma_o]$

8 Ermittlung von μ

8.1 Ermittlung basierend auf Stichproben

Es wird, ähnlich wie bei section 7 - Konfidenzintervall, die Erwartungswert μ zu Stichproben in einem Intervall errechnet. Es ist dabei σ gegeben!

Gegebene Variablen:

m Ein einzeiliger Stichprobenvektor

n Die Länge von m

σ Die Standardabweichung

α Die Signifikanz in Prozent

8.1.1 Berechnung von μ

$$\overline{x} = mean(m) \tag{24}$$

$$u = qnorm(1 - \frac{\alpha}{2}, 0, 1) \tag{25}$$

$$\mu_u = \overline{x} - u * \frac{\sigma}{\sqrt{n}} \tag{26}$$

$$\mu_o = \overline{x} + u * \frac{\sigma}{\sqrt{n}} \tag{27}$$

8.2 Studentsche Verteilung

8.2.1 Grundlegendes

Es wird der Erwartungswert μ auf Grund einer Stichprobe ermittelt. Hier ist jedoch σ nicht bekannt und muss selbst errechnet werden. Der Rest deckt sich mit Kapitel subsection 8.1 - Ermittlung basierend auf Stichproben.
Gegebene Variablen:

m Ein einzeiliger Stichprobenvektor

n Die Länge von m

s Die Standardabweichung s berechnet aus m errechnet mit $Stdev()$ bzw. $stdev()$

α Die Signifikanz in Prozent

8.2.2 Berechnung von μ

$$\overline{x} = mean(m) \tag{28}$$

$$s = Stdev(m) bzw. stdev(m) \tag{29}$$

$$f = n - 1 \tag{30}$$

$$\mu_u = \overline{x} - qt(1 - \frac{\alpha}{2}, f) * \frac{s}{\sqrt{n}} \tag{31}$$

$$\mu_o = \overline{x} + qt(1 - \frac{\alpha}{2}, f) * \frac{s}{\sqrt{n}} \tag{32}$$

9 χ^2-Verteilung

9.1 Grundlegendes

Die χ^2-Verteilung wird verwendet um zu Überpüfen:

- Wie wahrscheinlich ist eine bestimmte Standardabweichung s? (Kapitel subsubsection 9.6.2 - Suche nach p)
- Zu einer Wahrscheinlichkeit p wird welche Standardabweichung s erwartet? (Kapitel subsubsection 9.6.1 - Suche nach s)
- Eine alternierende Standardabweichung s ist zur Standardabweichung σ gesucht.

Wichtig dabei ist, dass es sich bei der Standardabweichung s um eine Standardabweichung gegeben von bestimmten Werten handelt. Somit wird s entweder durch einen externen Rechner gegeben oder selbst mit den folgenden Funktionen ermittelt:

$Stdev(m)$ m ist ein einzeiliger Vektor mit der Grundgesamtheit

$stdev(m)$ m ist ein einzeiliger Vektor mit Stichproben

9.2 Freiheitsgrade

Die Werte n, die freigewählt werden können. Die Definition der Freiheitsgrade:

$$f = n - 1 \tag{33}$$

9.3 Bestimmte Wahrscheinlichkeitsdichte

$$dchisq(x, f) = \text{Wahrscheinlichkeitsdichte für einen bestimmten Wert} \tag{34}$$

9.4 Kumulative Wahrscheinlichkeitsdichte

$$x_{\text{prüf}} = f * \frac{s^2}{\sigma^2} \tag{35}$$

$$pchisq(x_{\text{prüf}}, f) = \text{Wahrscheinlichkeitsdichte für höchstens einen Wert} \tag{36}$$

9.5 Inverse kumulative Wahrscheinlichkeitsdichte

$$\chi^2_{f,p} = qchisq(p, f) = x_{\text{prüf}} \tag{37}$$

9.6 Suche nach Werten

Es müssen wie im Kapitel subsection 5.1 - Grundlegendes μ und σ gegeben sein. Außerdem müssen die Freiheitsgrade f bekannt sein ($\rightarrow$ also die Anzahl der Werte!).

9.6.1 Suche nach s

Gegebene Variablen:

- p
- Es ist bekannt ob eine maximale (Kapitel subsection 9.4 - Kumulative Wahrscheinlichkeitsdichte) oder eine bestimmte (Kapitel subsection 9.3 - Bestimmte Wahrscheinlichkeitsdichte) Standardabweichung gesucht ist. (Matchad beherrscht nur die inverse kumulative Wahrscheinlichkeitsdichte!)

$$x_{\text{prüf}} = qchisq(p, f) \tag{38}$$

$$x_{\text{prüf}} = f * \frac{s^2}{\sigma^2} \text{solve, s} \rightarrow \tag{39}$$

9.6.2 Suche nach p

Gegebene Variablen:

- s bzw. m für $Stdev(m)$ oder $stdev(m)$

  Es wird somit die Berechnung des Kapitels subsection 9.4 - Kumulative Wahrscheinlichkeitsdichte verwendet.