THEMA UND GRUNDLAGEN

Ausgehend von einer Stichprobe mit (n) Einzelergebnissen ist der arithmetische Mittelwert bei vermuteter Normalverteilung der Gesamtpopulation nur mit einer bestimmten Bandbreite und mit einer definierten, bzw. wählbaren Vertrauens - würdigkeit (Vertrauensniveau VN ≤ 100%) prognostizierbar. Diese Bandbreite wird als Konfidenzintervall des Mittelwertes bezeichnet.

Der generelle Ansatz zur Berechnung der Begrenzung lautet wie folgt:

Obergrenze, resp. Untergrenze = $x_{quer} \pm s \cdot t_s \cdot n^{-0.5}$

x_{quer} = Mittelwert der Stichprobe
s = empirische Standardabweichung der Stichprobe
t_s = Koeffizient (Verzerrungsfaktor) je nach VN und Grösse (n) der Stichprobe bei unbekannter Streuung der Gesamtheit (σ)
n = Grösse der Stichprobe

Die massgebenden Koeffizienten (t_s) je nach VN und Stichprobengrösse (n) sind in der statistischen Literatur tabelliert, so beispielsweise für VN90%, 95%, 98%, 99%. Dazwischen liegende Werte (z.B. für VN 92.5%) dürfen aber nicht linear interpoliert werden, sondern folgen gegen den Endwert VN 100% hin einer logistischen Funktion (Sättigungskurve). Der rechnerische Ansatz zur Ermittlung von beliebigen Zwischen - werten (t_s) – gültig zwischen VN 90% und VN 100%, ab einer Grösse (n) ≥ 10 – ist nachstehend aufgeführt. Die individuelle Berechnung von entsprechenden Konfidenzintervallen kann, darauf basierend – mittels eines separaten xls – files erfolgen.

ZUSAMMENSTELLUNG DER REGRESSIONSFOMELN

In **„horizontaler Richtung“**, d. h. zwischen (tabellierten) VN 90% und VN 99% folgen die t_s – Werte pro „Stückzahl (n)“ der folgenden logistischen Summenkurve:

$\mathbf{t_{s\,(n)\,schätz}}$ = e^ [ln ($t_{s\,(n)\,schätz\infty}$) – $A_{(n)}$ · e^($B_{(n)}$·x)],
mit $x_{Eingabe}$ = ln [1/(1-VN% · 0.01)]

Aus den Statistiktabellen für (t_s) je nach „Stückzahl“ (n), folgen nach obiger Formel iterativ die Koeffizienten $A_{(n)}$, $B_{(n)}$, $t_{s((n)\,schätz\infty)}$ sowie der Korrelationskoeffizient r ~……:

Anzahl n	f = (n-1)	Konstantentherm $A_{(n)}$	Regressionskoeff. $B_{(n)}$	Max. Wert $t_{S\,((n)\,schätz\,\infty)}$	Korrel. Koeffizient r
6	(5)	2.947278	-0.205834	12.609475	> - 0.9999
7	(6)	2.687370	-0.221645	9.7488539	> - 0.9999
11	(10)	2.249962	-0.262717	6.1944443	> - 0.9999
16	(15)	2.075976	-0.283763	5.1659436	> - 0.9999
21	(20)	1.996442	-0.298681	4.7093285	> - 0.9999
26	(25)	1.953683	-0.307916	4.4709054	> - 0.9999
31	(30)	1.926896	-0.311156	4.3529750	> - 0.9999
41	(40)	1.892214	-0.318344	4.1830880	> - 0.9999
51	(50)	1.873256	-0.323513	4.0827449	> - 0.9999
∞	∞	1.800567	-0.337439	3.7676576	> - 0.9999

Tabelle 1

In „**vertikaler Richtung**“, d. h. zwischen n = 6 und n ≥ 51 ergeben sich aus einer analogen logistischen Summenkurve aus <u>iterativem</u> Prozess, wie oben

→ für die **Regression aller (A)**, mit $x_{Eingabe}$ = ln(n) und mit y = $[(1 / A_{(n)}) \cdot 100]$
$A_{(n)\ schätz} = e\hat{}[\ln(A_{(n),\ schätz\ \infty}) - a_{(A)} \cdot e\hat{}(b_{(A)} \cdot x)]$

→ für die **Regression aller (B)**, mit $x_{\ Eingabe}$ = ln (n) und mit y = $- B_{(n)} \cdot 100]$
$B_{(n)\ schätz} = e\hat{}[\ln(B_{(n),\ schätz\ \infty}) - a_{(B)} \cdot e\hat{}(b_{(B)} \cdot x)]$

→ für die **Regression aller < $t_{s\ schätz\ \infty}$ >**, mit $x_{\ Eingabe}$ = ln (n) und mit y = $[(1 / t_{s\ (n)}) \cdot 100]$
$t_{s((n)\ schätz\infty)\ schätz} = e\hat{}[\ln(t_{s((n)\ schätz\infty),\ schätz\ \infty}) - a_{(ts\ (n)\ schätz\ \infty)} \cdot e\hat{}(b_{(ts\ (n)\ schätz\ \infty)} \cdot x)]$

Aus den Kenngrössen $A_{(n)}$, $B_{(n)}$, sowie $t_{s\ (n)\ schätz\ \infty}$ je nach „Stückzahl“ (n) laut Tabelle 1, folgen aus <u>iterativem</u> Prozess die „Koeffizienten für Zwischenwerte“ $a_{(A)}$ und $b_{(A)}$ zur Ermittlung von $A_{(n)\ schätz}$, resp. $a_{(B)}$ und $b_{(B)}$ zur Ermittlung von $B_{(n)\ schätz}$ sowie von $a_{\ (ts\ (n)\ schätz\ \infty)}$ und $b_{\ (ts\ (n)\ schätz\ \infty)}$ zur Bestimmung der benötigten Hilfsgrösse <$t_{s(n)schätz\infty),\ schätz}$>.

Hilfs – Koeff.	Zur Ermittlung des Konstanten - Therms $A_{(n)\ schätz}$	Zur Ermittlung des Regressions - Koeff. $B_{(n)\ schätz}$	Zur Ermittlung des massg. Max -Wertes $t_{s((n)\ schätz\ \infty),\ schätz}$
$a_{(A)}$	4.888525		
$b_{(A)}$	-1.299752		
$A_{(n),\ schätz\ \infty}$	1.819313 (54.965) umgeformt)		
$r_{\ (A)}$	> -0.9999		
$a_{(B)}$		3.410940	
$b_{(B)}$		-1.069559	
$B_{(n),\ schätz\ \infty}$		-0.34012 (34.0125 umgeformt)	
$r_{\ (B)}$		> -0.9996	
$a_{\ (ts\ (n)\ schätz\ \infty)}$			13.993261
$b_{\ (ts\ (n)\ schätz\ \infty)}$			-1.394169
$t_{s((n)\ schätz\ \infty),\ schätz\ \infty}$			3.8574 (25.9237 umgeformt)
$r_{(ts\ (n)\ schätz\ \infty)}$			> -0.9998

Tabelle 2

ABFRAGE – FILE (xls)

Mit dem separat konzipierten Excel – Rechenprogramm – basierend auf den hier gezeigten Grundlagen – kann aus einer Stichprobe von n – Einzelwerten (10 ≤ n ≤ ∞) für das wünschbare Vertrauensniveau (90% ≤ VN ≤ 100%) der Ober – und der Unterwert des Konfidenzintervalls zum Mittelwert abgefragt werden. Desgleichen die daraus resultierende (±) Abweichung. Die punktuell mögliche, minimale Fehlerkote gegenüber direkt vergleichbaren Werten (aus t_s – Statistiktabellen mit festgelegtem VN, geteilt durch $n^{0.5}$) ist in der Regel belanglos, sofern die Mindest – Stückzahl (n ≥ 10) eingehalten ist.

Das eingetragene „Norm“beispiel (Mittelwert x_{quer} = 0, empirische Standard - abweichung s = 1, Stückzahl n = 50 und VN 95%) → gelbe Eingabefelder, sind schreibgeschützt und können mit eigenen Festlegungen überschrieben werden. Die grau untermalten Zahlen sind Hilfswerte und für den Nutzer ohne Interesse.

21. November 2011 / Ba.