Anmerkungen zur statistischen Auswertung der Abwitterung von CDF/CIF-Tests

Die Norm 12390-9 fordert zusätzlich zur Angabe der Prüfergebnisse die Angabe der **Mittelwerte** und der **Standardabweichungen**.

Die Norm beschreibt kein Abnahmekriterium. Dies wird z. B. für den CDF-Test in der Literatur mit 1500 g/m² angegeben, den der Mittelwert der Serie nicht überschreiten darf. In der Regel wird das Abnahmekriterium vom Auftraggeber oder vom Prüfinstitut unter Bezugnahme auf die Literatur oder anderen Prüf- und Qualitätsvorschriften festgelegt.

Im Merkblatt "Frostprüfung von Beton" der BAW wird ein Abnahmekriterium definiert, wobei der Mittelwert der Prüfserie 1500 g/m² und die 95%-**Quantile** der Prüfserie 1800 g/m² nicht überschreiten darf

Es gibt verschiedene Methoden, um die 95%-Quantile zu bestimmen. Diese unterscheiden sich besonders bei kleinen Stichproben erheblich. Oft wird die Auswertung auch mit Tabellenkalkulationsprogrammen wie MS-Excel oder OpenOffice-Calc gemacht. Diese Programme bieten bereits fertige Funktionen, um die statistischen Werte zu berechnen. Nachdem beim CDF/CIF-Test meist nur fünf Werte zu Grunde liegen, ist es besonders wichtig, die dazu richtige Funktion auszuwählen.

Mittelwert

Es ist sicherlich das arithmetische Mittel gemeint. Dies ist definiert als die Summe aller Werte geteilt durch die Anzahl aller Werte.

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Die entsprechende Calc-Funktion ist "MITTELWERT"

Standardabweichung

Die Standardabweichung ist definiert als die Quadratwurzel der **Varianz**.

$$\sigma = \sqrt{Var}$$

Es gilt:

- 68,3 % aller Messwerte liegen im Bereich des Mittelwertes ± Standardabweichung.
- 95 % aller Messwerte liegen im Bereich des Mittelwertes ± 1,96·Standardabweichung

Die **Varianz** ist das arithmetische Mittel der quadrierten Abweichungen:

$$Var_n = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n}$$

Die entsprechende Calc-Funktion für die Standardabweichung ist "STABWN"

Empirische Standardabweichung

Es gibt noch den Begriff der korrigierten oder **empirischen Varianz**:

$$Var_{n-1} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n-1}$$

Der Unterschied ist $n-1$ im Nenner. Man begründet dies damit, dass die Standardabweichung einen Bereich um einen Wert x beschreibt. Nachdem dieser Wert x schon einer von n Werten ist, sind nur noch $n-1$ Stichprobenwerte frei, die Abweichungen beinhalten können.

Die empirische Standardabweichung ist wegen des Teilers n-1 immer etwas größer als die Standardabweichung. Mit größeren Stichproben nähern sich beide Standardabweichungen an.

Die empirische Standardabweichung ist demnach:

$$S = \sqrt{Var_{n-1}}$$

Die entsprechende Calc-Funktion für die empirische Standardabweichung ist "STABWA".

95%-Quantil

Als Quantil Qp wird in der Statistik ein Merkmalswert bezeichnet, unterhalb dessen ein vorgegebener Anteil aller Fälle der Verteilung liegt. Der Anteil ist der Wert p (liegt zwischen 0 und 1), der dem Prozentanteil entspricht. Wir suchen in unserem Fall das Quantil $Q_{95\%}$.

Es werden in der Literatur mehrere Möglichkeiten angegeben ein Quantil zu bestimmen. In dem Wikipedia-Artikel über Quantile wird das empirische Quantil beschrieben. Es wird der Wert in der gegebenen Verteilung gesucht, für den die entsprechende Prozentangabe, in unserem Beispiel 95%, zutrifft. Ist $n{\cdot}p$ ganzzahlig, wird der entsprechende Wert genommen. Ist $n{\cdot}p$ nicht ganzzahlig, wird der nächst höhere Wert herangezogen. Dies ist bei kleinen Stichproben nicht aussagekräftig. Bei nur 5 Werten, wie beim CDF-Test, würde immer der höchste Wert die 95%-Quantile bestimmen.

Die in den meisten Kalkulationsprogrammen implementierte Methode ist folgende: Wie vorher beschrieben wird der Wert in der gegebenen Verteilung gesucht, für den die entsprechende Prozentangabe, in unserem Beispiel 95%, zutrifft. Liegt der gesuchte Prozentwert zwischen zwei Werten, wird interpoliert.

Die entsprechende Calc-Funktion ist "QUANTIL"

Auf Nachfrage beim BAW (tel. H. Reschke, 09/2012) wird allerdings zur Bewertung des CDF-Tests für die Bestimmung der 95% Quantile die Student-t-Verteilung zu Grunde gelegt. Diese soll für kleine Stichproben besser geeignet sein.

$$x = \bar{x} + t \cdot S$$

Die folgende Tabelle zeigt t-Werte für die Freiheitsgrade df bei bei 95% aller möglichen t-verteilten Zufallsvariablen.

$$df = n - 1$$

	Wahrscheinlichkeit 95%
df	**t-Wert**
1	6,314
2	2,920
3	2,353
4	2,132
5	2,015
6	1,943
7	1,895
8	1,860
9	1,833

Beispiel für die Unterschiede der Methoden

gegeben sind fünf Abwitterungswerte:

1370, 1220, 1290 1250, 1240 (g/m²)

Sortiert nach der Größe und den Prozentwerten zugeordnet ergibt:

1220	1240	1250	1290	1370	(g/m²)
0%	25%	50%	75%	100%	Quantile

Auswertung der 95%-Quantile empirisch nach Wikipedia:

$n \cdot p = 5 \cdot 0{,}95 = 4{,}75$ ergibt aufgerundet 5

Wir suchen den 5. Messwert in der Reihe. Der entsprechende Wert ist **1370 g/m²**

Auswertung der 95%-Quantile mit der "QUANTIL" Funktion eines Tabellenkalkulationsprogrammes:

der 4. Messwert ist 1290 g/m² ≙ 75%

der 5. Messwert ist 1370 g/m² ≙ 100%

interpoliert ergibt:

1354 g/m² ≙ 95%

Die 95%-Quantile ist **1354 g/m².**

Auswertung der 95%-Quantile mit der Student-t-Verteilung:

gesucht ist der t-Wert für 95%:

mit:
$n = 5$

und
$$df = 5 - 1 = 4$$

folgt
$t = 2{,}132$

mit
$x = 1274$ g/m² und
$S = 59{,}4$ g/m²

folgt:
$$x = 1274 g/m^2 + 2{,}132 \cdot 59{,}4\, g/m^2 = 1401 g/m^2$$

Die 95%-Quantile ist **1401 g/m².**