

## PROC REPORT - Decathlon

Manuela Wern  
SCHUFA Holding AG  
Massenbergstr. 13a  
44787 Bochum  
manuela.wern@schufa.de

Stefanie Daum  
SCHUFA Holding AG  
Massenbergstr. 13a  
44787 Bochum  
stefanie.daum@schufa.de

### Zusammenfassung

Die Darstellung von Analyseergebnissen in Form von Berichten ist ein wichtiger Bestandteil des Arbeitsalltags für viele SAS-Anwender. Für die Erstellung von Berichten direkt aus SAS heraus ist die Prozedur PROC REPORT eine gute Wahl. Jedoch wird oft auf andere Möglichkeiten zugegriffen, da die Hürden dieser Prozedur nicht für alle SAS-Anwender leicht zu nehmen sind. Die verschiedenen Disziplinen von PROC REPORT sollen deshalb anhand von Beispielen aus dem Zehnkampf trainiert werden. Nach dem abschließenden 1500-Meter-Lauf haben die Leser einen Überblick über die Möglichkeiten der Prozedur PROC REPORT, z.B. Listenberichte erstellen, verdichtete Berichte erstellen, Grafiken einbinden.

**Schlüsselwörter:** BASE SAS, PROC REPORT

## 1 Einführung

Mit der Prozedur **PROC REPORT** ist es möglich, komplexe Darstellungen von Analyseergebnisse aus SAS heraus zu erzeugen. Dazu werden u.a. Teile aus anderen SAS Prozeduren wie z.B. **PROC MEANS** verwendet.

```
PROC REPORT <options>;
  COLUMN variable list and column specifications;
  DEFINE column/column usage and attributes;
  COMPUTE column; compute block statements;
  ENDCOMP;
RUN;
```

Die Prozedur REPORT kann sowohl interaktiv gesteuert werden, als auch durch Befehle innerhalb eines SAS Programms – eine Verwendung ist also im Batchmodus möglich. Im Folgenden wird nur die zweite Variante vorgestellt. Die Standard-Einstellung bei PROC REPORT ist die über verschiedene Fenster interaktiv gesteuerte erste Variante. Um das automatische Öffnen der interaktiven Fenster zu umgehen, wird bei den Optionen NOWINDOWS bzw. die verkürzte Schreibweise NOWD gesetzt.

Ebenfalls bei den Optionen wird wie üblich der darzustellende bzw. auszuwertende Datensatz mit einem DATA= angegeben.

Die am häufigsten verwendeten Statements von PROC REPORT sind

COLUMN	Hier werden alle Variablen aufgezählt, die für die Erstellung der Tabelle verwendet werden.
DEFINE	Für jede Variable des COLUMN-Statements wird ein DEFINE-Statement verwendet, um anzugeben, wie die Variable verwendet werden soll.
COMPUTE	Mit diesem Statement können neue Spalten erzeugt bzw. Berechnungen durchgeführt werden.

Im DEFINE-Statement wird festgelegt, wie PROC REPORT die im COLUMN-Statement angegebenen Variablen verwenden soll. Unter anderem sind folgende Typen möglich:

DISPLAY	Hiermit werden die Werte der Variablen angezeigt. Dies ist die Standardeinstellung für alphanumerische Variablen.
ANALYSIS	Wird für Variablen verwendet, für die eine Statistik berechnet werden soll. Dies ist die Standardeinstellung für numerische Variablen.
GROUP	Die so gekennzeichnete Variable wird für eine Gruppierung verwendet, um z.B. verdichtete Berichte zu erstellen.
ORDER	Nach der Variable mit diesem Attribut wird die Ausgabe sortiert.
ACROSS	Für jeden Wert der Variable wird eine Spalte im Bericht ausgegeben.
COMPUTED	Kennzeichnet eine Variable, die nicht in der zu Grunde liegenden Datei steht, sondern erst noch in einem COMPUTE-Block berechnet wird.

Der Aufbau eines COMPUTE-Blocks ist dabei immer wie folgt:

```
COMPUTE <location> <report_item> </options>;  
  ... one ore more SAS language elements ...  
ENDCOMP;
```

Alle folgenden Beispiele sind für die ODS-Ausgabe optimiert.

Verschiedene Anwendungsbeispiele der **PROC REPORT** Prozedur werden mit Hilfe von Zehnkampf-Daten vorgestellt. Folgende Daten standen für die Auswertungen zur Verfügung:

- Zehnkampfergebnisse der Olympischen Spiele von 1912 bis 2012
- Zehnkampfergebnisse der Leichtathletik-Weltmeisterschaften von 1983 bis 2013

Quellen:

- (1) <http://www.decathlon2000.com/eng/championships>
- (2) <http://www.iaaf.org/results>

## 2 Der 100m-Lauf oder einfacher Bericht mit sortierter Anzeige

Die erste und auch kürzeste Disziplin beim Zehnkampf ist der 100m-Lauf. Mit einem ebenfalls kurzen PROC REPORT Prozeduraufruf soll der Startschuss für den PROC REPORT-Decathlon gegeben werden. Für die Betrachtung der Fragestellung, wer die schnellste jemals beim Zehnkampf bei Olympischen Spielen gelaufene Zeit erreicht hat, ist folgender PROC REPORT Aufruf hilfreich:

```
PROC REPORT DATA=PROJEKT.OLYMPIA NOWD;
  COLUMN JAHR ATHLET METER_100;
  DEFINE JAHR           /DISPLAY;
  DEFINE ATHLET         /DISPLAY;
  DEFINE METER_100      /ORDER ORDER=FORMATTED FORMAT=6.2;
RUN;
```

Hinter dem COLUMN-Statement werden die für den gewünschten Bericht relevanten Variablen angegeben. Im DEFINE-Statement wird für die Variablen „Jahr“ und „Athlet“ festgelegt, dass diese angezeigt werden sollen. Die Variable „Meter\_100“ wird ebenfalls angezeigt, allerdings wird mit ORDER eine Sortierung nach dieser Variablen erzwungen. Die Art der Sortierung wird mit dem nachfolgenden ORDER= spezifiziert, wobei FORMATTED der Standardeinstellung entspricht, also nicht angegeben werden müsste. Bei anderen SAS-Prozeduren ist die Standardeinstellung INTERNAL, so dass empfohlen wird ORDER=FORMATTED mit aufzuführen, da möglicherweise die Standardeinstellung in zukünftigen Versionen angepasst werden könnte.

Die Angabe des Formats ist in diesem Fall wichtig, denn ohne diese Angabe würden die Werte nach dem Format Best12. sortiert. Dies hätte zur Folge, dass zunächst die ganzen Zahlen 11, 12, 13, dann die Zahlen mit einer Stelle hinter dem Komma und zuletzt die Zahlen mit zwei Stellen hinter dem Komma aufgeführt würden.

Das Ergebnis mit der angegebenen Formatierung sieht dann folgendermaßen aus (Auszug):

**Tabelle 1:** Sortierte Zeiten beim 100m-Lauf

Jahr	Athlet	Meter_100
2012	Ashton Eaton	10.35
1968	William Toomey	10.41
2012	Trey Hardee	10.42
1984	Daley Thompson	10.44
2004	Bryan Clay	
2008	Bryan Clay	
1996	Chris Huffins	10.47

Jahr	Athlet	Meter_100
2000	Chris Huffins	10.48
2012	Damian Warner	

Die schnellste bei einem Zehnkampf bei Olympischen Spielen gelaufene Zeit hält Ashton Eaton mit 10,35 Sekunden.

Es fällt auf, dass einige Zellen leer sind. Das liegt nicht an Missings, sondern daran, dass z.B. Bryan Clay 2004 und 2008 die gleiche Zeit gelaufen ist wie Daley Thompson 1984. Gibt man die Option SPANROWS an, werden die Zellen mit gleichen Werten verbunden, so dass deutlicher wird, dass es sich in diesem Fall um die gleichen Zeiten handelt.

Bei Angabe von ORDER gibt es folgende Auswahlmöglichkeiten:

DATA	Sortiert die Daten nach ihrem Auftreten im Datensatz.
FREQ	Sortiert die Daten nach ihrer aufsteigenden Häufigkeit.
FORMATTED	Sortiert die Daten nach ihren formatierten Werten.
INTERNAL	Sortiert die Daten nach ihren unformatierten Werten analog der Prozedur SORT.

Eine Sortierung mit ORDER=INTERNAL hätte ebenfalls zum obigen Ergebnis geführt.

### 3 Der Weitsprung oder verdichteter Bericht und Berechnung einer Statistik

Um zu erfahren, wie sich die Weiten beim Weitsprung, der 2. Disziplin, in den letzten einhundert Jahren bei den Olympischen Spielen verändert haben, wird mit folgendem Prozeduraufruf pro Ausprägung der Variable „Jahr“ ein Mittelwert berechnet:

```
PROC REPORT DATA=PROJEKT.OLYMPIA NOWD;  
  COLUMN JAHR WEITSPRUNG;  
  DEFINE JAHR          /GROUP;  
  DEFINE WEITSPRUNG    /ANALYSIS MEAN FORMAT=COMMAX5.2;  
RUN;
```

Hinter der Variablen „Jahr“ steht GROUP, damit diese Variable als Gruppierungsvariable herangezogen wird und pro Ausprägung dieser Variablen eine Statistik berechnet werden kann. Hinter der Variablen „Weitsprung“ steht ANALYSIS. Standardmäßig wird hier die Statistik SUM, also die Summe aller Werte pro Ausprägung der Gruppierungsvariablen berechnet. Um andere Statistiken zu erhalten, werden diese einfach angegeben, in diesem Fall die Statistik MEAN. Die Angabe einer Statistik im DEFINE-Statement impliziert die ANALYSIS-Option, so dass diese eigentlich nicht angegeben

werden müsste. Um den Code leichter lesbar zu machen, bietet es sich aber an, auf diese Angabe nicht zu verzichten. Nachfolgend das Ergebnis dieses Prozeduraufrufs in verkürzter Darstellung:

**Tabelle 2:** Mittlere Weitsprungweiten pro Olympische Spiele

Jahr	Weitsprung
1912	6,31
1920	6,05
1924	6,31
1928	6,21
1932	6,74
1936	6,69
⋮	
1992	7,26
1996	7,40
2000	7,09
2004	7,23
2008	7,13
2012	7,21

Die Steigerung im Laufe der Zeit ist zu erkennen: Im Unterschied zum Mittelwert der Weite im Jahr 1912 liegt dieser Wert im Jahr 2012 90cm weiter.

## 4 Das Kugelstoßen oder Unterbrechung des Berichts für eine Zusammenfassung mit **BREAK**

Für die Absolvierung der 3. Disziplin ist ein kleiner Kraftakt von Nöten. Es sollen nicht nur die Kugelstoßergebnisse pro Jahr ausgegeben werden, sondern auch das beste Ergebnis des Jahres nochmals explizit angezeigt werden. Dafür wird der folgende Prozeduraufruf gestartet:

```
PROC REPORT DATA=PROJEKT.OLYMPIA NOWD;
  COLUMN JAHR ATHLET KUGEL;
  DEFINE JAHR          /GROUP;
  DEFINE ATHLET        /DISPLAY;
  DEFINE KUGEL         /ANALYSIS 'Kugelstoßen' MAX FORMAT=COMMAX6.2;
  BREAK AFTER JAHR/SUMMARIZE;
RUN;
```

Hier liegt mit „Jahr“ eine Gruppierungsvariable und mit „Athlet“ eine Display-Variable vor. Der schöneren Form halber soll die Variable „Kugel“ im Bericht umbenannt werden in „Kugelstoßen“. Dies geschieht über die DEFINE-Anweisung.

Mit der Anweisung `BREAK AFTER` wird nach jeder neuen Ausprägung der Variablen „Jahr“ das beste Kugelstoßergebnis (maximaler Wert pro Jahr) ausgegeben und somit eine „Unterbrechung“ des Berichts erzeugt. Hier ein Auszug aus dem Ergebnis:

**Tabelle 3:** Kugelstoßergebnisse nach Jahr und Ausgabe des besten Ergebnisses pro Jahr

Jahr	Athlet	Kugelstoßen
1912	James Thorpe	12,89
	Hugo Wieslander	12,14
	Charles Lomberg	11,67
	Gösta Holmer	10,98
	James Joseph Donahue	9,67
	Eugene LeRoy Mercer	9,76
	Waldemar Wickholm	11,09
	Erik Kugelberg	9,98
	Karl von Halt	11,12
	Josef Schäffer	11,50
	Aleksander Schulz	10,08
	Alfreds Alslebens	8,48
1912		12,89
1920	Helge Lovland	11,19
	Brutus Kerr Hamilton	11,64
	⋮	
	Eduard Mikhan	14,75
	Dmitriy Karpov	16,47
	Jan Felix Knobel	15,29
	Daniel Awde	
	Kurt Felix	13,28
	Roman Šebrle	
	Mihail Dudaš	13,76
2012		16,47

Nach 100 Jahren schaffte es der Sieger dieser Disziplin, die 7,257 kg schweren Kugeln bei Olympischen Spielen um mehr als 3,50 m weiter zu stoßen.

## 5 Der Hochsprung oder Berechnung einer Kennzahl mit Hilfe von COMPUTE

Die nächste Disziplin ist der Hochsprung. Dieser ist für Zehnkämpfer durch den vertikalen Absprung nicht leicht zu meistern. Mit Hilfe einer COMPUTE-Anweisung sollen die für die jeweilige Höhe erreichten Punkte berechnet werden und in einem Bericht dargestellt werden.

Für den hier gewünschten Bericht sieht dann der Prozeduraufruf folgendermaßen aus:

```
PROC REPORT DATA=PROJEKT.OLYMPIA NOWD;
  COLUMN VERANSTALTUNG ATHLET HOCHSPRUNG PUNKTE_HOCHSPRUNG;
  DEFINE VERANSTALTUNG      /DISPLAY;
  DEFINE ATHLET             /DISPLAY;
  DEFINE HOCHSPRUNG         /DISPLAY FORMAT=COMMAX6.2;
  DEFINE PUNKTE_HOCHSPRUNG /COMPUTED  FORMAT=8. 'Hochsprungpunkte';
  COMPUTE PUNKTE_HOCHSPRUNG;
      PUNKTE_HOCHSPRUNG
          = FLOOR(0.84658 * (HOCHSPRUNG*100 - 75)**1.42);
  ENDCOMP;
RUN;
```

Schon in der COLUMN-Anweisung wird die zu berechnende Variable „Punkte\_Hochsprung“ aufgeführt. In der DEFINE-Anweisung zu dieser Variablen wird mit COMPUTED angegeben, dass diese Variable noch zu berechnen ist, die Berechnungsformel wird im entsprechenden COMPUTE-Block hinterlegt.

Ein Auszug aus dem Ergebnis ist in der folgenden Tabelle zu sehen:

**Tabelle 4:** Hochsprungergebnisse (Höhe und Punkte)

Veranstaltung	Athlet	Hochsprung	Hochsprungpunkte
05. Olympische Spiele	James Thorpe	1,87	687
05. Olympische Spiele	Hugo Wieslander	1,75	585
05. Olympische Spiele	Charles Lomberg	1,80	627
05. Olympische Spiele	Gösta Holmer	1,70	544
05. Olympische Spiele	James Joseph Donahue	1,65	504
05. Olympische Spiele	Eugene LeRoy Mercer	1,65	504
05. Olympische Spiele	Waldemar Wickholm	1,60	464
05. Olympische Spiele	Erik Kugelberg	1,65	504
05. Olympische Spiele	Karl von Halt	1,70	544
05. Olympische Spiele	Josef Schäffer	1,55	426
05. Olympische Spiele	Aleksander Schulz	1,55	426
05. Olympische Spiele	Alfreds Alslebens	1,70	544

Veranstaltung	Athlet	Hochsprung	Hochsprungpunkte
07. Olympische Spiele	Helge Lovland	1,65	504
07. Olympische Spiele	Brutus Kerr Hamilton	1,60	464
07. Olympische Spiele	Bertil Ohlson	1,65	504
07. Olympische Spiele	Gösta Holmer	1,70	544
07. Olympische Spiele	Evert Nilsson	1,75	585
07. Olympische Spiele	Waldemar Wickholm	1,60	464

Da die entsprechende Formel zur Berechnung der Punkte nicht linear ist, sind Aussagen zur Anzahl der Punkte pro Zentimeter nicht genau möglich, aber bei einer Sprunghöhe um 1,70m gibt es für je fünf Zentimeter Höhe ca. 40 Punkte.

## 6 Der 400m-Lauf oder Steuerung der Farben einer Zelle mit CALL DEFINE

Die letzte Disziplin des ersten Tages ist der 400m-Lauf, die längste Sprintstrecke in der Leichtathletik. Als sehr gute Leistung im Zehnkampf gelten Zeiten unter 47 Sekunden, daher sollen die Ergebnisse unter 47 Sekunden farblich hervorgehoben werden. Ein nützliches Werkzeug ist dabei CALL DEFINE. Mit CALL DEFINE kann der Wert eines Attributs einer Zelle verändert werden.

Mit folgendem Programmschritt werden die Farben der Spalte „Meter\_400“ in Abhängigkeit vom Wert der Variable geändert.

```

PROC REPORT DATA=PROJEKT.OLYMPIA NOWD;
COLUMN VERANSTALTUNG METER_400;
DEFINE VERANSTALTUNG /GROUP;
DEFINE METER_400 /ANALYSIS MIN FORMAT=6.2 '400m-Lauf';
COMPUTE METER_400;
    IF METER_400.MIN LT 47 THEN
        CALL DEFINE (_COL_, "STYLE",
                    "STYLE=[BACKGROUNDCOLOR=DARKGREY COLOR=WHITE]");
ENDCOMP;
RUN;

```

Das CALL DEFINE-Statement benötigt drei Parameter: die Spalte bzw. Zelle, die zu ändern ist, das zu ändernde Attribut und den Wert des zu ändernden Attributs. Im obigen Prozedurschritt wird mit \_COL\_ die aktuelle Spalte (hier „Meter\_400“) angesprochen, für die das STYLE-Attribut geändert werden soll. Im dritten Argument wird mit BACKGROUNDCOLOR die Hintergrundfarbe geändert und mit COLOR die Schriftfarbe.



In Kombination mit einer IF-Abfrage werden die Zellen also je nach Wert des Minimums pro Veranstaltung beim 400m-Lauf unterschiedlich eingefärbt: Für einen Wert kleiner als 47 mit dunkelgrauem Hintergrund und weißer Schrift. Das führt zu folgendem Ergebnis:

**Tabelle 5:** Beste Zeit beim 400m-Lauf pro Veranstaltung und Markierung der Zeiten unter 47 Sekunden

Veranstaltung	400m-Lauf
05. Olympische Spiele	49.90
07. Olympische Spiele	52.80
08. Olympische Spiele	51.40
09. Olympische Spiele	51.40
10. Olympische Spiele	49.00
11. Olympische Spiele	49.40
14. Olympische Spiele	50.50
15. Olympische Spiele	50.33
16. Olympische Spiele	48.35
17. Olympische Spiele	48.00
18. Olympische Spiele	48.20
19. Olympische Spiele	45.68
20. Olympische Spiele	46.25
21. Olympische Spiele	47.12
22. Olympische Spiele	48.01
23. Olympische Spiele	46.97
24. Olympische Spiele	47.44
25. Olympische Spiele	47.73
26. Olympische Spiele	46.82
27. Olympische Spiele	46.41
28. Olympische Spiele	46.81
29. Olympische Spiele	47.16
30. Olympische Spiele	46.90

Bei den 19. Olympischen Spielen 1968 in Mexiko-Stadt wurden zum ersten Mal bei Olympia die 400m im Zehnkampf unter 47 Sekunden gelaufen und das mit 45,68 Sekunden sogar deutlich.

## 7 Der Hürdenlauf oder die Berechnung von verschiedenen Statistiken zu einer Variable

Beim Weitsprung wurde die Berechnung des Mittelwerts erreicht, indem im DEFINE-Statement das Schlüsselwort MEAN angegeben wurde. Es ist aber ebenso möglich, die gewünschten Statistiken im COLUMN-Statement anzugeben. Dies ist auch eine einfache Möglichkeit, gleich mehrere Statistiken berechnen zu lassen. Folgender Prozedurschritt berechnet für die Variable „Huerden“ gleich drei Werte, nämlich das Minimum, den Mittelwert und das Maximum.

```
PROC REPORT DATA=PROJEKT.OLYMPIA NOWD;  
  COLUMN VERANSTALTUNG JAHR HUERDEN, (MIN MEAN MAX) BESTWERT;  
  DEFINE VERANSTALTUNG      /GROUP;  
  DEFINE JAHR                /GROUP;  
  DEFINE HUERDEN             /ANALYSIS "110 Meter Hürdenlauf"  
                             FORMAT=5.2;  
  DEFINE BESTWERT            /COMPUTED  FORMAT=5.2  
                             "Zehnkampfrekord 1996";  
  COMPUTE BESTWERT;  
      BESTWERT = 13.47;  
  ENDCOMP;  
  WHERE 1992 <= JAHR <= 2012;  
RUN;
```

Im COLUMN-Statement werden die gewünschten Statistiken aufgezählt, mit einem Komma von der Variable „Huerden“ getrennt, zu der diese Statistiken berechnet werden sollen.

Es ist auch möglich zunächst die Statistiken und dann, mit einem Komma getrennt, die Variable aufzuführen. Das führt dazu, dass die Überschriften vertauscht werden, also MIN, MEAN und MAX über „110 Meter Hürdenlauf“ stehen.

Die Variablen „Veranstaltung“ und „Jahr“ dienen als Gruppierungsvariablen. Da pro Ausprägung der Variablen „Veranstaltung“ nur eine Ausprägung der Variablen „Jahr“ vorliegt, führt dies dazu, dass die Variable „Jahr“ im Grunde nur zusätzlich angezeigt wird.

Die Variable „Huerden“ ist als Analysevariable angegeben. Die gewünschten Statistiken sind bereits im COLUMN-Statement aufgezählt. Zusätzlich zu den statistischen Kennzahlen soll noch die gültige Bestleistung für diese Disziplin beim Zehnkampf in der Tabelle mit aufgeführt werden. Hierfür wird die Variable „Bestwert“ definiert, für die in einem COMPUTE-Block der Wert des Rekords eingetragen wird.

Die Ausgabe wurde mit dem WHERE-Statement auf die letzten fünf Olympischen Spiele eingegrenzt.

**Tabelle 6:** Verschiedene Statistiken zum 110m-Hürdenlauf pro Veranstaltung

		110 Meter Hürdenlauf			Zehnkampfrekord 1996
Veranstaltung	Jahr	min	mean	max	
25. Olympische Spiele	1992	13.95	14.85	15.42	13.47
26. Olympische Spiele	1996	13.47	14.47	15.20	13.47
27. Olympische Spiele	2000	13.87	14.56	15.27	13.47
28. Olympische Spiele	2004	13.97	14.61	15.80	13.47
29. Olympische Spiele	2008	13.90	14.60	16.25	13.47
30. Olympische Spiele	2012	13.54	14.55	15.59	13.47

Wie in der Tabelle zu sehen ist, wurde der Zehnkampfrekord 1996 bei den Olympischen Spielen aufgestellt. Diese Bestleistung wurde von Frank Busemann erzielt, der bei den Olympischen Spielen in Atlanta die Silbermedaille gewann.

## 8 Der Diskuswurf und der Stabhochsprung oder Zusammenfassung der Ergebnistabelle mit RBREAK

Beim Kugelstoßen wurde mit dem BREAK-Statement nach jeder Ausprägung der Variable „Veranstaltung“ der Bericht für eine Zusammenfassung unterbrochen, in diesem Fall wurde das Maximum ausgegeben. Es ist ebenso möglich, mit RBREAK eine Zusammenfassung des ganzen Berichts zu erzeugen. Syntax und Optionen sind analog zum BREAK-Statement, bei dem allerdings eine GROUP-Variable angegeben werden muss.

Es sollen nun die erzielten Punkte beim Diskuswurf und beim Stabhochsprung miteinander verglichen werden und eine Zusammenfassung am Ende des Berichts erzeugt werden. Um den Vergleich zu erleichtern, soll zusätzlich eine Differenz zwischen den Punkten beim Diskuswurf und den Punkten beim Stabhochsprung gebildet werden.

```
PROC REPORT DATA=PROJEKT.OLYMPIA_PUNKTE NOWD;
  COLUMN JAHR DISKUS_P, (N MEAN) STABHOCHSPRUNG_P, (N MEAN)
    DIFFERENZ;
  DEFINE JAHR / GROUP;
  DEFINE DISKUS_P / ANALYSIS FORMAT=5.0;
  DEFINE STABHOCHSPRUNG_P / ANALYSIS FORMAT=5.0;
  DEFINE DIFFERENZ / COMPUTED FORMAT=5.0 "Punkte Differenz";
  COMPUTE DIFFERENZ;
    DIFFERENZ = DISKUS_P.MEAN - STABHOCHSPRUNG_P.MEAN;
  ENDCOMP;
  RBREAK AFTER / SUMMARIZE;
  COMPUTE AFTER;
```

```

        JAHR      = 'GESAMT';
    ENDCOMP;
RUN;

```

Für die Variablen „Diskus\_p“ und „Stabhochsprung\_p“ werden schon in der COLUMN-Anweisung die gewünschten Statistiken N (Anzahl) und MEAN (Mittelwert) festgelegt. Die Variable „Differenz“ wird berechnet als Differenz des Mittels der Punkte beim Diskuswurf und des Mittels der Punkte beim Stabhochsprung. Diese Differenz wird für jede Ausprägung der GROUP-Variable „Jahr“ gebildet. Mit RBREAK AFTER / SUMMARIZE wird der Bericht mit einer Zusammenfassung abgeschlossen. Hierbei wird durch SUMMARIZE für jede Spalte ein Wert abhängig vom Typ der dazugehörigen Variable bestimmt. So wird für eine Analysevariable der Wert der angegebenen Statistik über alle Beobachtungen berechnet (hier Anzahl und Mittelwert), für eine COMPUTED-Variable wird der Wert durch den COMPUTE-Block dieser Variable berechnet. Der zweite COMPUTE-Block dient dazu, den Wert der Variable „Jahr“ für die abschließende Zeile auf „GESAMT“ zu setzen.

**Tabelle 7:** Vergleich der Punkte beim Diskuswurf und beim Stabhochsprung pro Jahr

	Diskus_p		Stabhochsprung_p		
Jahr	N	MEAN	N	MEAN	Punkte Differenz
1912	11	515	11	396	119
1920	12	530	12	380	150
1924	23	489	23	381	109
1928	26	551	26	413	139
1932	11	646	11	465	181
1936	17	619	17	500	118
1948	28	608	28	415	193
1952	21	623	21	505	118
1956	12	642	12	477	165
1960	22	683	22	530	153
1964	18	682	18	629	53
1968	20	721	20	674	47
1972	22	733	22	695	38
1976	21	713	21	728	-15
1980	15	735	15	783	-49
1984	23	737	23	757	-20
1988	33	713	33	835	-123
1992	24	744	24	826	-82

	Diskus_p		Stabhochsprung_p		
Jahr	N	MEAN	N	MEAN	Punkte Differenz
1996	28	764	28	877	-113
2000	25	721	25	865	-144
2004	30	750	30	832	-83
2008	24	765	24	810	-45
2012	27	759	27	823	-64
<i>GESAMT</i>	<i>493</i>	<i>682</i>	<i>493</i>	<i>662</i>	<i>20</i>

Die so erzeugte Tabelle bietet einen guten Überblick über die bei diesen beiden Disziplinen erreichten mittleren Punktzahlen bei Olympischen Spielen. Seit 1976 wurden im Mittel beim Stabhochsprung immer mehr Punkte erzielt als beim Diskuswurf.

## 9 Der Speerwurf oder Einbinden von Grafiken in einen Bericht

Mit PROC REPORT ist es möglich, vorhandene Grafiken direkt in einen Bericht einzubinden. Dies kann z.B. in der Überschrift sein oder direkt in einer Zelle. Mit dem folgenden Programmschritt soll ein Bericht erzeugt werden, der die bei Olympischen Spielen erzielten Speerwurfweiten mit denen bei einer Weltmeisterschaft vergleicht. Um dies anschaulich zu machen, sollen sowohl die Olympischen Ringe als auch das WM-Logo in den Bericht eingebunden werden. Hierzu wird die Variable „Wettkampf“ benötigt, die angibt, ob es sich bei dem Wert um ein Ergebnis bei Olympischen Spielen handelt oder bei einer Weltmeisterschaft. Diese Variable dient als Gruppierungsvariable und beinhaltet die Namen der zwei verwendeten png-Grafiken. Betrachtet werden in diesem Beispiel nur die Jahre ab 1983, da erst seit diesem Jahr die IAAF-Leichtathletik-Weltmeisterschaften veranstaltet werden.

```
PROC REPORT DATA=PROJEKT.OLYMPIA_WM NOWD;
  COLUMN WETTKAMPF LOGO SPEER;
  DEFINE WETTKAMPF /GROUP;
  DEFINE LOGO      /COMPUTED;
  DEFINE SPEER     /ANALYSIS MEAN FORMAT=COMMAX5.2;
  WHERE JAHR > 1982;
  COMPUTE LOGO / CHAR;
  CALL DEFINE('LOGO','STYLE',"STYLE=
    {PREIMAGE='d:\Decathlon\" || TRIM(WETTKAMPF) || ".png' }");
  ENDCOMP;
RUN;
```

Neben der Variablen „Wettkampf“ als Gruppierungsvariable und der zu analysierenden Variable „Speer“, für die der Mittelwert berechnet werden soll, wird im COLUMN-

Statement die Variable „Logo“ aufgeführt. Diese Spalte soll die gewünschte Grafik enthalten und wird im COMPUTE-Block als alphanumerische Variable mit der Länge 10 angelegt.

Mit der CALL DEFINE-Anweisung wird für die Variable „Logo“ der Wert des STYLE-Attributs verändert. Im dritten Argument wird mit PREIMAGE die Grafik vor dem Inhalt der Zelle, der in diesem Fall leer ist, eingebunden. Mit POSTIMAGE wird die Grafik nach dem Zellinhalt eingefügt.

**Tabelle 8:** Mittlere erzielte Speerwurfweiten bei Olympischen Spielen und Weltmeisterschaften

Veranstaltung	LOGO	Speer
OLYMPIA		59,54
WM		60,64

Im Mittel ist der Speer bei den Weltmeisterschaften der Jahre 1983-2013 weiter geflogen als bei den Olympischen Spielen des vergleichbaren Zeitraums.

## 10 Der 1.500m-Lauf oder verdichteter Bericht mit ACROSS

Der 1.500m-Lauf ist die letzte Disziplin des Zehnkampfs und schließt den zweiten Tag ab. Nun kann analysiert werden, ob an beiden Tagen des Zehnkampfs im Schnitt gleich viele Punkte erzielt werden, oder ob die Leistung erkennbar nachlässt. Es wurden nur Athleten berücksichtigt, die den Zehnkampf auch beendet haben. Um zum einen eine Gruppierung nach Tag 1 und 2 zu erreichen und zum anderen pro Jahr auszuwerten, ist das DEFINE-Attribut ACROSS hilfreich. Mit ACROSS wird für jede Ausprägung der Variable eine eigene Spalte erzeugt:

```

PROC REPORT DATA=PROJEKT.OLYMPIA NOWD;
  COLUMN TAG JAHR, PUNKTE, (MEAN);
  DEFINE TAG /GROUP;
  DEFINE JAHR /ACROSS;
  DEFINE PUNKTE /ANALYSIS FORMAT=5.;
RUN;

```

Dieser Prozeduraufruf erzeugt pro Ausprägung der Variable „Tag“ eine Zeile und durch ACROSS pro Ausprägung der Variable „Jahr“ eine Spalte. Hier ein Auszug aus dem Ergebnis:

**Tabelle 9:** Vergleich der Mittelwerte der erzielten Punkte an Tag 1 und 2 pro Jahr

	Jahr							
	1912	1920	1924	1928	...	2004	2008	2012
	Punkte	Punkte	Punkte	Punkte		Punkte	Punkte	Punkte
TAG	MEAN	MEAN	MEAN	MEAN		MEAN	MEAN	MEAN
1	3048	2847	3073	3105		4100	4114	4136
2	2542	2510	2403	2577		3872	3798	3897

Bei sämtlichen olympischen Zehnkämpfen wurden am zweiten Tag weniger Punkte erzielt als am ersten Tag.

## Literatur

- [1] Art Carpenter: Carpenter's Complete Guide to the SAS REPORT Procedure, SAS Institute Inc., Cary, NC, USA - 2nd printing, 2013
- [2] Ben Cochran: Thirty Three Tricks with PROC REPORT: A Quick Introduction to the Powerful REPORT Procedure,  
<http://www.mwsug.org/proceedings/2010/tutorials/MWSUG-2010-109.pdf>
- [3] SAS 9.3 Help and Documentation