

The Basis of Comprehensive Statistical Analysis of the Sydney'2000 Olympic Games Basketball and Volleyball Tournaments

Wiesław Wagner

Institute of Statistics, University School of Physical Education

Rybaki Street 19, 61-884 Poznań, Poland

wwag@poczta.onet.pl

The analysis of the sport tournament in team sport games is conducted in terms of various criteria. The scale and scope of the analysis depends on the statistical data about competing teams, their players and played games. With regard to the basketball and volleyball tournaments played during the Olympic Games in Sydney in 2000, during qualification matches were played in a round robin system, and then the teams competed in a knockout system for positions from 1 to 8. It is suggested that the comprehensive analysis of the competition in team sport games should be discussed in the following order:

- | | |
|---|---|
| 1 ^o Kind and type of the sports tournament, | 5 ^o statistical analysis of matches played in the qualifying round, |
| 2 ^o Database with teams and players, | 6 ^o statistical analysis of matches played by the teams competing for positions 1-8. |
| 3 ^o Division into groups and system of play, including further stages of knockout games, | |
| 4 ^o Database with played matches, | |

Basketball matches were played in the old system – two halves of 20 minutes each with a possibility of overtime in case of a draw after 40 minutes of play.

In this paper the following symbols:

- I–H, II–H, EM – game states of I and II half and end of the game,
- q – number of teams participating at the tournament and we assume that it is even number,
- $Z = \{Z_1, Z_2, \dots, Z_q\}$ – collection of basketball teams,
- $m = m_e + m_p$ – number of played games at the tournament, when m_e – number of played game in the elimination round, m_p – in the cup round,
- (Z_i^k, Z_j^k) - pair defining k -game between teams $Z_i, Z_j; i, j=1, 2, \dots, q; i \neq j$ for $k=1, 2, \dots, m$,
- $(X_{1i}^k, X_{1j}^k), (X_{2i}^k, X_{2j}^k), (X_{3i}^k, X_{3j}^k)$ - pairs results k -games between teams (Z_i, Z_j) for game states I-H, II-H, EM,
- $M_k = \{(Z_i^k, Z_j^k) : (X_{1i}^k, X_{1j}^k), (X_{3i}^k, X_{3j}^k)\}$ – statistical data about k -game,
- M_1, M_2, \dots, M_m – sequence of played games at the tournament.

Volleyball matches were played under the new system, best of five sets, with each set up to 25 points with the exception of the fifth set which is played to 15 points. The following abbreviations are used:

- S-I, S-II, S-III, S-IV, S-V – set I, set II, set III, set IV and set V,
- $q, Z = \{Z_1, Z_2, \dots, Z_q\}, m = m_e + m_p, (Z_i^k, Z_j^k)$ - have the same symbols, like 1,
- s_k – number of played sets in k -games,
- $(X_{gi}^k, X_{gj}^k); g = 1, 2, \dots, s_h$ - pairs s_h results k -game between teams (Z_i, Z_j)
- $M_k = \{(Z_i^k, Z_j^k) : s_h, (X_{gi}^k, X_{gj}^k); g = 1, \dots, s_h\}$ – statistical data about k -volleyball game.

For the analysis of basketball and volleyball games the following index are prepared (in %):

- ◆ index of partial efficiency of game - $WSS = 100 W / (W + P)$, where W and P express number of
- ◆ won and lost played games,

- ◆ index of points intensity - $WS = 100 U / S$, where U and S express number of won and lost played games,
- ◆ index of points efficiency - $WUS = 100 U / (U + S)$, U and S have the same meaning like WS .

For the statistical analysis of both sports, various statistical methods have been proposed. They may be divided into three main groups:

- methods of statistical description of the sample,
- methods of statistical inference based on the F-Fisher's and Student's t tests,
- methods of analysis of variance for single and cross classification.

In method in group a) sample statistics are used: minimum, maximum, range, average, median, standard deviation, coefficient of variation and skewness coefficient. Except for these methods grouping in structural rows and cross-classifications are used.

Numerical data are: (1) number of points scored by the team in each half of the basketball match or in a set in volleyball, (2) point match differences (PMD) or point set differences (PSD), (3) number of points scored by teams winning a match (set), (4) number of points conceded by the team in a match (set). Except for the data concerning matches directly, also methods from group a) analyse the data about players (age, body height, body mass, spike, block). Methods from group b) are used to compare averages between group A and B. They are based on F-Fisher's test to compare homogeneity of variance or t-Student's test to compare averages. These comparisons are made between relevant features of groups A and B. The method of analysis of variance is used for simultaneous comparison of averages for teams in groups and teams competing for positions from 1 to 8. It is used for the cross data *Countries x Parts of match* (5), *Countries x Parts of match* (8), *Countries x Sets of match* (5) and *Countries x Sets of match* (8). A case with multiplication factor (5) relates to qualifications in groups and one with multiplication factor (8) to teams competing for positions 1-4.

Generally, A means factor relating to countries, B factor relating to the parts of game or sets, with adequately a and b levels A_1, A_2, \dots, A_a and B_1, B_2, \dots, B_b levels. The combinations of levels of these factors $A \times B$ have credited combinations $(A_i B_j)$, which include $r=5$ or $r=8$ observations for the two-way cross classification the following null hypothesis are verified:

H_0^A : there is no differentiation between averages for levels A_1, A_2, \dots, A_a A factors,

H_0^B : there is no differentiation between averages for levels B_1, B_2, \dots, B_b B factors,

H_0^{AB} : there is no differentiation between averages subclass $(A_i B_j)$; $i=1, \dots, a$; $j=1, \dots, b$, or there is no interaction between levels of A and B factors.

The results of hypothesis testing are given in table (see **Rosner** 1988). Hypothesis H_0^A , H_0^B , H_0^{AB} are rejected, where $F_{cal} > F_{tab}$, and it means the significant of given factor. The verification of hypothesis starts with H_0^{AB} hypothesis about interaction. In case of reject of its, the multiple comparisons of subclass are done, if not, the conclusions about H_0^A and H_0^B hypothesis is conducted

Table Two-way analysis of variance

Source of Variation	SS	df	MS	F_{cal}	F_{tab}
Factor A	SSA	$f1=a-1$	$MSA=SSA/f1$	$FA=MSA/MSE$	$F_{\alpha;f1,f4}$
Factor B	SSB	$f2=b-1$	$MSB=SSB/f2$	$FB=MSB/MSE$	$F_{\alpha;f2,f4}$
Interaction AB	SSAB	$f3=(a-1)(b-1)$	$MSAB=SSAB/f3$	$FAB=MSAB/MSE$	$F_{\alpha;f3,f4}$
Error	SSE	$f4=ab(r-1)$	$MSE=SSE/f4$	-	-
Total	SSG	-	-	-	-

The mentioned methods are used following statistical analysis: (a) game played in group A, (b) game played in group B, (c) comparison of games played in group A and B and (d) games played by teams 1-8 places.

REFERENCES

Rosner B., (1988): *Fundamentals of biostatistics*. Duxbury Press, Harvard University.