

# Lazarski, A., “Advantages and limitations of the quantitative data analysis of the coaching performance and results. Is coaching really effective? Are surveys conducted after the coaching sessions representing the best tool to gather data?”

Unpublished paper for newsletter: ‘ODiTK’, 02/10, 2010

## Introduction

A lot of the companies are investing their money into the human resources development. They are hiring trainers to speed up development of their intangible assets. The very simple question could be raised – how to measure results, to convert these intangible into more tangible? How to measure performance of such a coaching sessions? The below presented example is in fact a professionally conducted case study oriented to analyse one aspect of the results of the coaching sessions. The tricky thing here is that coaching sessions have been conducted by two trainers. Therefore it creates also a benchmark possibility of these two colleagues. The whole analysis is presented with the use of SPSS statistical software. The gathered survey data is available in the appendix number two. If you are not willing to go through statistical analysis then please do not hesitate to go straight to the conclusions.

## Data overview

Challenge to analyse data shares similar characteristics to cross sectional design. This is certainly non experimental one since there is no data available on knowledge or competences level before coaching sessions and after it. Also there is no control (being without treatment) group present. Nevertheless cross sectional design rely “on existing differences” (de Vaus, 2001: 170). Therefore the author of this business case analysis will rely on the sample available and will try if necessary to apply statistical control to “remove differences between groups” (de Vaus, 2001: 177). Descriptive statistics were used to shed some more light and present some very first impressions on the data.

		Sex			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	female	16	53.3	53.3	53.3
	male	14	46.7	46.7	100.0
	Total	30	100.0	100.0	

**Table 1** Frequency table - females, males attending coaching sessions

It is clearly visible that slightly more females 53.3% than males 46.7% senior managers attended coaching sessions.

**Coach \* Sex Crosstabulation**

			Sex		Total
			female	male	
Coach Alex	Count	4	10	14	
	% within Coach	28.6%	71.4%	100.0%	
Ruth	Count	12	4	16	
	% within Coach	75.0%	25.0%	100.0%	
Total	Count	16	14	30	
	% within Coach	53.3%	46.7%	100.0%	

**Table 2 Contingency table - coach and female, male senior managers**

The male senior managers visited mainly 71.4% male coach Alex. At the same moment female senior managers preferred to participate 75% in coaching sessions provided by female coach Ruth.

**Coach**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Alex	14	46.7	46.7	46.7
	Ruth	16	53.3	53.3	100.0
Total		30	100.0	100.0	

**Table 3 Frequency table - Alex, Ruth**

Ruth met slightly more managers than Alex. Maybe it was due to the smaller number of sessions with individual managers, so she simply could have more time than Alex.

**Statistics**

		Age	Sessions	Satisfaction
N	Valid	30	30	30
	Missing	0	0	0
Mean		35.2000	7.3667	4.1333
Median		35.5000	7.0000	4.0000
Mode		33.00 <sup>a</sup>	6.00 <sup>a</sup>	4.00 <sup>a</sup>
Std. Deviation		7.41108	2.34128	1.56983
Skewness		-.197	.129	-.235
Std. Error of Skewness		.427	.427	.427
Kurtosis		-.633	-.247	-.281
Std. Error of Kurtosis		.833	.833	.833
Range		28.00	9.00	6.00
Minimum		21.00	3.00	1.00
Maximum		49.00	12.00	7.00

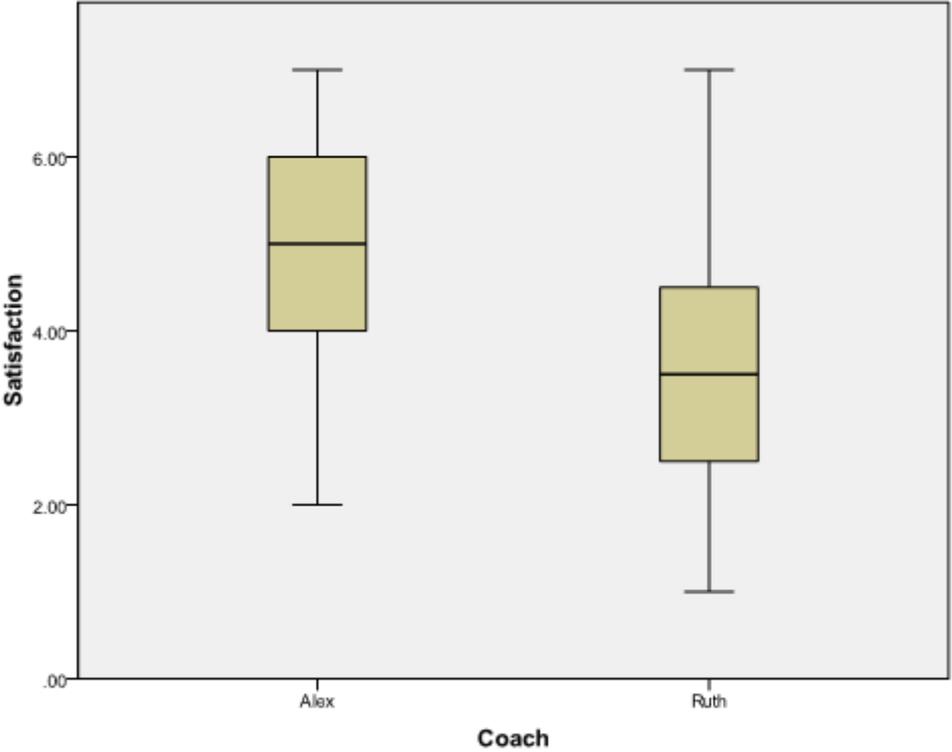
a. Multiple modes exist. The smallest value is shown

**Table 4 Central tendency for age, sessions and satisfaction**

For such a small sample skewness is relatively low therefore observations are spread quite symmetrically on the left and right side of the average values. Also mean is close to the mode especially for age and satisfaction. Kurtosis is also relatively low therefore a lot of observations are located near the average. It implies possible assumption of standard distribution.

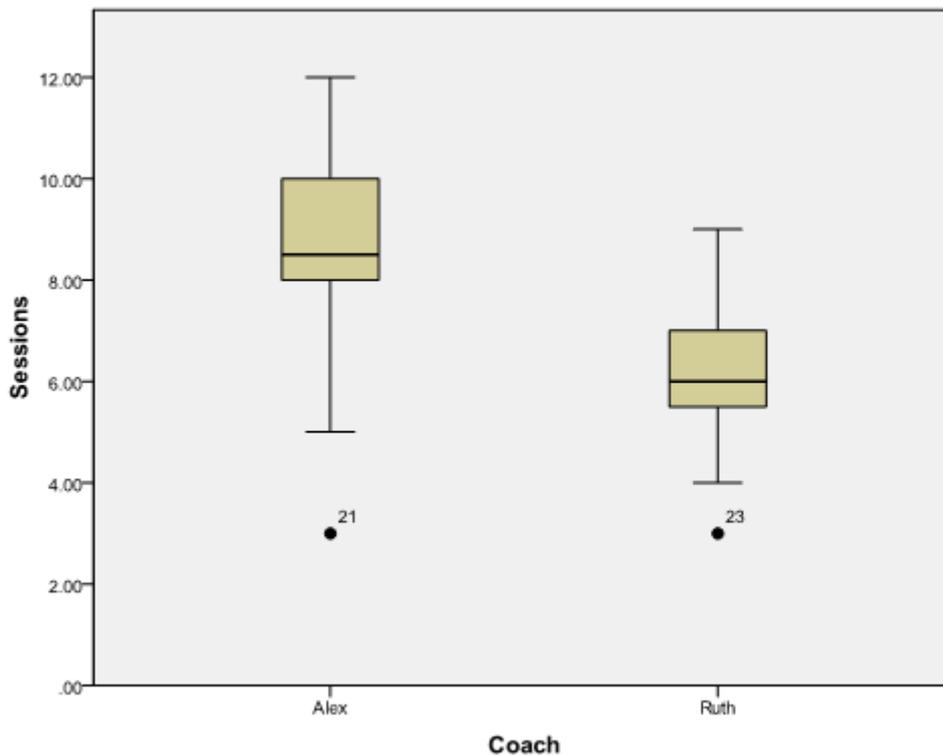
The most common values were 33 years old, 6 sessions, satisfaction level 4. Satisfaction average rating of 4.13 is quite neutral and certainly cannot be considered as an indicator of the success of the whole coaching process.

**Boxplots**



**Graph 1 Boxplots - satisfaction per coach**

In Ruth’s case the range of satisfaction is larger than for Alex. It could be said that Ruth’s satisfaction rating is pulling the average satisfaction “down” - major 50% of her results is located relatively low. Median satisfaction is much higher for Alex than for Ruth.



**Graph 2 Boxplots - number of sessions per coach**

Both Alex and Ruth have outliers (not extreme cases) represented by a managers attending very few (three) sessions. Alex was able to provide more session than Ruth, achieving the highest rank of twelve sessions per senior manager. Median of his sessions is much higher than one in case of Ruth.

### Crosstabulation

To present crosstabulation with coach, sessions were recoded into three equal groups. Number of sessions was categorised as follows: 1 (1-4), 2 (5-8) and 3 (9-12). Categories are in fact defined with the assumption that one session per manager was possible and that the upper edge is defined by the maximum sessions per manager achieved.

**Coach \* sessions categories Crosstabulation**

		sessions categories			Total	
		1.00	2.00	3.00		
Coach	Alex	Count	1	6	7	14
		% within Coach	7.1%	42.9%	50.0%	100.0%
	Ruth	Count	2	12	2	16
		% within Coach	12.5%	75.0%	12.5%	100.0%
Total		Count	3	18	9	30
		% within Coach	10.0%	60.0%	30.0%	100.0%

**Graph 3 Contingency table - coach and sessions categories**

Graph 3 shows that Alex was able in almost 50% of cases to provide 9 or more sessions. Ruth in fact could hardly (12.5% of all cases) cross the “magic” number of 8 sessions per senior manager. To shed some more light onto this division it would be interesting to get to know

what was the minimum and maximum number of sessions that senior manager could participate in.

In the next crosstabulation process satisfaction was categorised into three groups to better polarise negative, neutral and positive feelings. In fact satisfaction even if presented as an interval value tend to be more ordinal one. It is simply hard to say that distances between satisfaction level six and seven is exactly the same as between five and six etc. Therefore data was categorised similarly to the Likert scale into three groups: 1 (1-3) representing rather negative satisfaction, 2 (4) as neutral one and 3 (5-7) as category designating positive satisfaction measurement result.

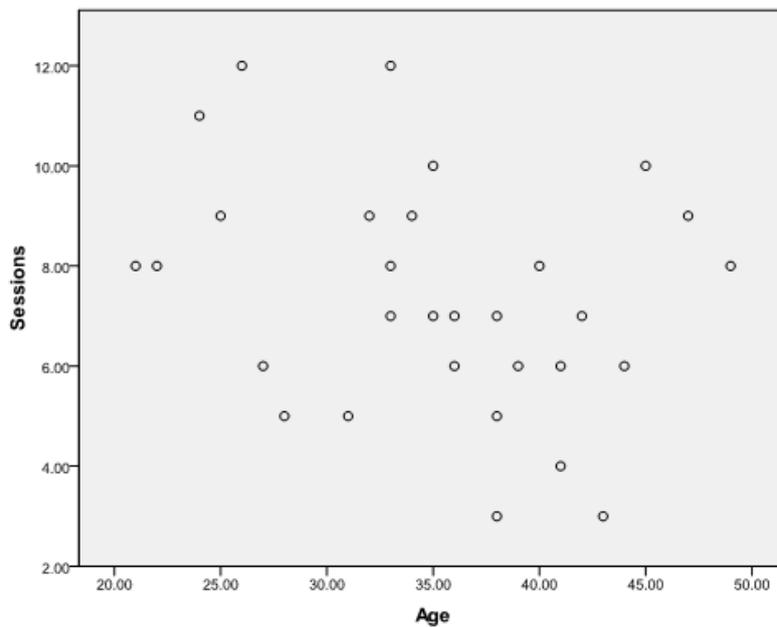
**Coach \* satisfaction categories Crosstabulation**

			satisfaction categories			Total
			1.00	2.00	3.00	
Coach Alex	Count		1	4	9	14
	% within Coach		7.1%	28.6%	64.3%	100.0%
Ruth	Count		8	4	4	16
	% within Coach		50.0%	25.0%	25.0%	100.0%
Total	Count		9	8	13	30
	% within Coach		30.0%	26.7%	43.3%	100.0%

**Graph 4 Contingency table - coach and satisfaction categories**

As previously discussed, Alex was able to coach managers in the higher number of sessions than Ruth. Adding to this, conclusions derived from Graph 4 it could be carefully stressed that Alex was also more effective. He scored in the highest 3 satisfaction category at 64.3% while at the same moment Ruth scored only 25%. Ruth experienced 50% of all opinions in the first category. It designates low satisfaction level of senior managers with the service that she has provided.

## Correlations



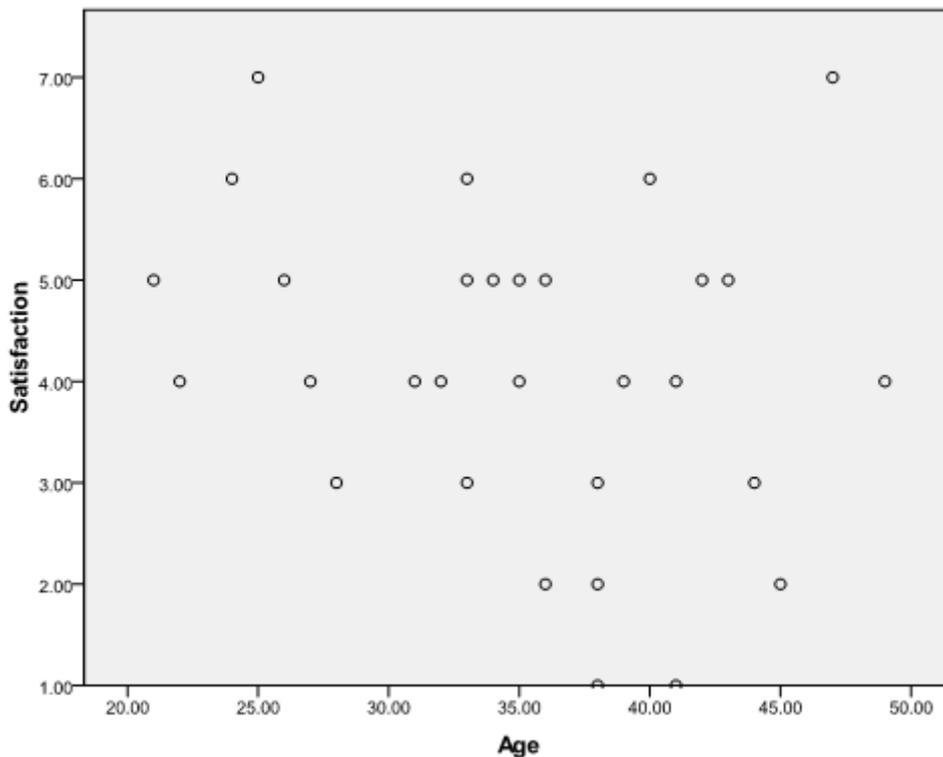
**Graph 5 Scatterplot - age and sessions**

Scatterplot is not representing any “visible” correlation between independent variable age and dependent variable number of sessions. The conclusion is that the age of senior managers did not influence the number of sessions that they have participated in.

		Age	Sessions
Age	Pearson Correlation	1	-.285
	Sig. (2-tailed)		.127
	N	30	30
Sessions	Pearson Correlation	-.285	1
	Sig. (2-tailed)	.127	
	N	30	30

**Table 5 Pearson correlation - age and sessions**

The Pearson correlation is representing a weak negative tendency (Greasley, 2008: 80) of  $-0.285$ . Coefficient of determination shows that only around 8% of variance in sessions number is related to age. Also correlation is not statistically significant since its value of 0.127 is higher than acceptable 0.05.



**Graph 6 Scatterplot - age and satisfaction**

Graph 6 is not representing any “visible” correlation between independent variable age and dependent variable satisfaction. Therefore age of senior managers did not influence their level of satisfaction.

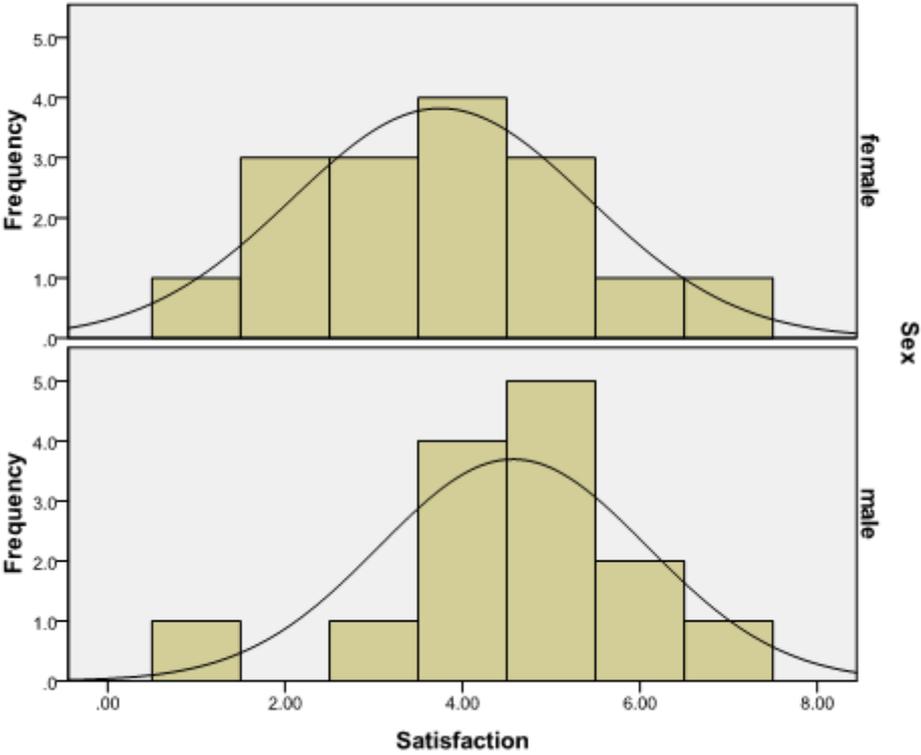
Correlations			
		Age	Satisfaction
Age	Pearson Correlation	1	-.234
	Sig. (2-tailed)		.214
	N	30	30
Satisfaction	Pearson Correlation	-.234	1
	Sig. (2-tailed)	.214	
	N	30	30

**Table 6 Pearson correlation - age and satisfaction**

The Pearson correlation is representing a weak negative tendency of -0.234. Coefficient of determination shows that only around 5% of variance in satisfaction ratings is related to age. Statistical significance of 0.214 represents that there is less than 21.4% chances that the researcher “could have a sample that shows a relationship when there is not one in the population” (Bryman and Bell, 2007: 368). This measured level is not acceptable by professionals since it is above 5%.

### Significant differences

Before conducting tests to analyse if there are any statistically significant differences for male and female satisfaction ratings the major assumptions for application of parametric tests should be checked (Greasley, 2008: 89).



**Graph 7 Histogram - satisfaction of male, female senior managers**

As it is visible assumption of normal distribution could be accepted. Of course it would more beneficial to have a larger sample size what most probably would positively influence the shape of normal distribution. In general “The larger the sample the narrower the band (called confidence interval)” (de Vaus, 2001: 189).

**Tests of Normality**

Sex	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Satisfaction female	.126	16	.200*	.966	16	.766
Satisfaction male	.204	14	.118	.917	14	.198

a. Lilliefors Significance Correction  
 \*. This is a lower bound of the true significance.

**Table 7 Kolmogorov-Smirnov and Shapiro-Wilk tests of normality**

Conducted Kolmogorov-Smirnov and Shapiro-Wilk tests show lack of significance 0.2, 0.118, 0.766, 0.198 are greater than 0.05. Therefore assumption of normal distribution is sustained.

To check for any significant statistical difference independent samples t-test was applied (Easterby-Smith *et al*, 2008: 256).

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Satisfaction	female	16	3.7500	1.61245	.40311
	male	14	4.5714	1.45255	.38821

**Table 8 Group statistics - female, male and satisfaction ratings**

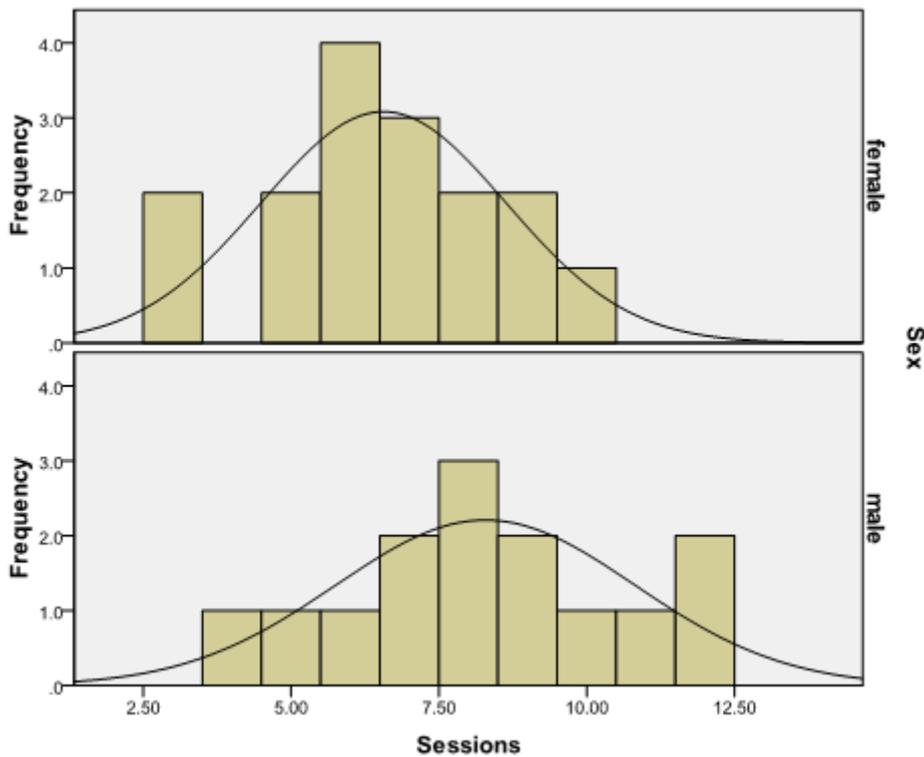
Standard deviation of satisfaction is slightly larger for female than those for male managers. It could designate that it was more problematic for them to judge their satisfaction level. Satisfaction mean of 3.75 for female managers is lower than those for male managers.

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Satisfaction	Equal variances assumed	.415	.525	-1.457	28	.156	-.82143	.56368	-1.97608	.33322
	Equal variances not assumed			-1.468	27.968	.153	-.82143	.55965	-1.96788	.32502

**Table 9 Independent samples t-test - male female satisfaction difference**

Levene's test for equality of variances  $p = 0.525$  points out that its value is not significant. It means that assumption for parametric test of 'homogeneity of variance' is not violated (Greasley, 2008: 89). The results show mean difference  $-0.82$  at  $p = 0.156$ . It implies that since  $p > 0.05$  it could not be said that the difference between male and female senior managers satisfaction is of any statistical significance.

Also in the case of the number of sessions conducted for male/female managers before involving into analysis of if there is statistically significant difference, assumptions for parametric analysis should be fulfilled.



**Graph 8 Histogram - sessions for male, female senior managers**

The normal curve in both histograms as well as presented frequencies suggest that the data are normally distributed. To be certain before entering into parametric independent samples t-test Kolmogorov-Smirnov and Shapiro-Wilk test of normality was applied.

Tests of Normality							
Sex		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Sessions	female	.139	16	.200*	.957	16	.603
	male	.118	14	.200*	.965	14	.811

a. Lilliefors Significance Correction

\*. This is a lower bound of the true significance.

**Table 10 Kolmogorov-Smirnov and Shapiro-Wilk tests of normality**

Conducted Kolmogorov-Smirnov and Shapiro-Wilk test shows lack of statistical significance  $p > 0.05$  and therefore assumption of normal distribution is not violated. To check for any significant statistical differences between female and male number of sessions independent samples t-test was applied.

Group Statistics					
Sex		N	Mean	Std. Deviation	Std. Error Mean
Sessions	female	16	6.5625	1.99896	.49974
	male	14	8.2857	2.43148	.64984

**Table 11 Group statistics - female, male and sessions**

Standard deviation for number of sessions is larger for males than those for females. Sessions mean for female managers is lower than those for male managers.

		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Sessions	Equal variances assumed	4.97	.486	-2.130	28	.042	-1.72321	.80889	-3.38016	-.06627
	Equal variances not assumed			-2.102	25.265	.046	-1.72321	.81978	-3.41068	-.03575

**Table 12 Independent samples t-test - male female sessions difference**

Levene’s test for equality of variances  $p = 0.486$  points out that its value is not significant. It means that assumption for parametric test of ‘homogeneity of variance’ is not violated. The results show mean difference at  $-1.72$  and at  $p = 0.042$ . It implies that since  $p < 0.05$  it could be said that the difference between male and female sessions number is statistically significant. Due to the closeness of  $0.042$  to  $0.05$  this issue should be further analysed. Also attention should be paid to relatively small sample size. Nevertheless it can be concluded that the difference in number of sessions that male and female managers have participated in is statistically significant.

**Conclusions**

It could be said that slightly more female than male senior managers attended coaching sessions. Male managers visited mainly male coach Alex while at the same moment female managers paid a visit mainly to female coach Ruth but it could not be said that the difference between male and female senior managers satisfaction is of any statistical significance.

Unfortunately it is not possible to conclude quantitatively from the data the reasons for this preference. This issue should be investigated in qualitative analysis - for instance during subsequently undertaken unstructured open interview.

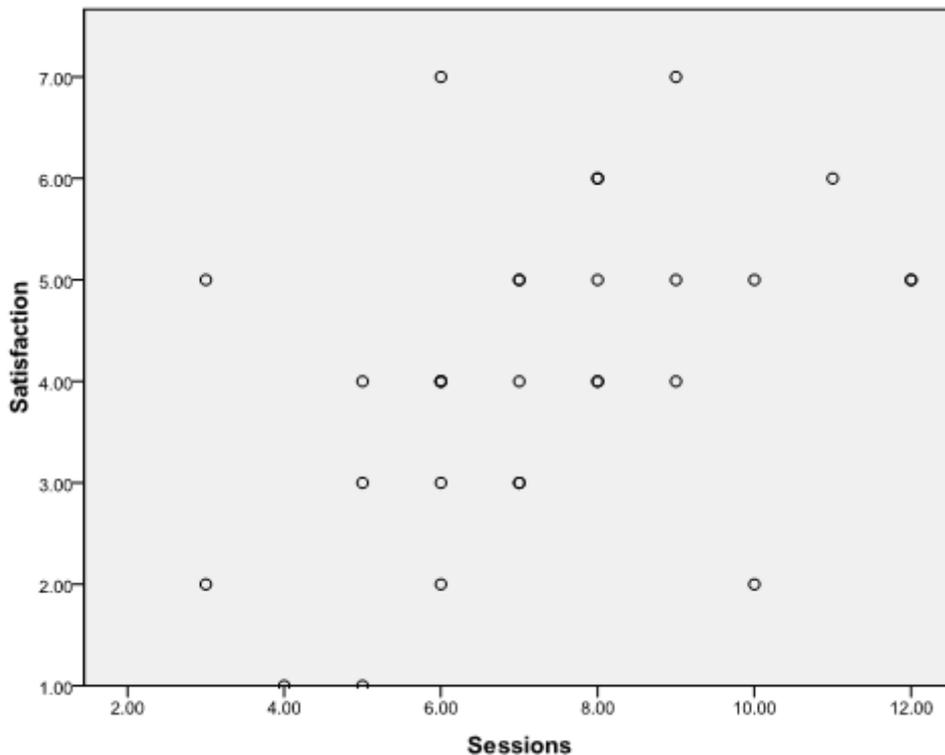
Ruth met slightly more managers than Alex. The question could be raised if it was possible due to the smaller number of sessions maintained by her with individual managers. She could simply spare more time than Alex. Nevertheless this is only an assumption since there is no information on the sessions time length given.

Next issue is that Ruth satisfaction rating is pulling the average satisfaction “down” - almost 50% of her results is located relatively low. Median satisfaction is much higher for Alex than for Ruth. Alex achieved also the highest rank of twelve sessions per senior manager.

Alex was able in almost 50% of cases to service 9 or more sessions. Ruth in fact could hardly cross the number of 8 sessions, which took place only in 12.5% of all cases. Is this long cooperation between senior managers and Alex an indicator of satisfaction and willingness to continue cooperation? Alex was also more effective. He scored in the highest 3 satisfaction category at 64.3% while at the same moment Ruth scored only in the same category at the level of 25%. Ruth experienced 50% of all opinions in the lowest satisfaction

category what designates unfortunately high disappointment with the service that she has provided.

Continuing discussion on the coaching performance there another explanation might also be possible. Maybe satisfaction level is correlated with the number of sessions?



**Graph 9 Scatterplot - sessions and satisfaction for Alex and Ruth**

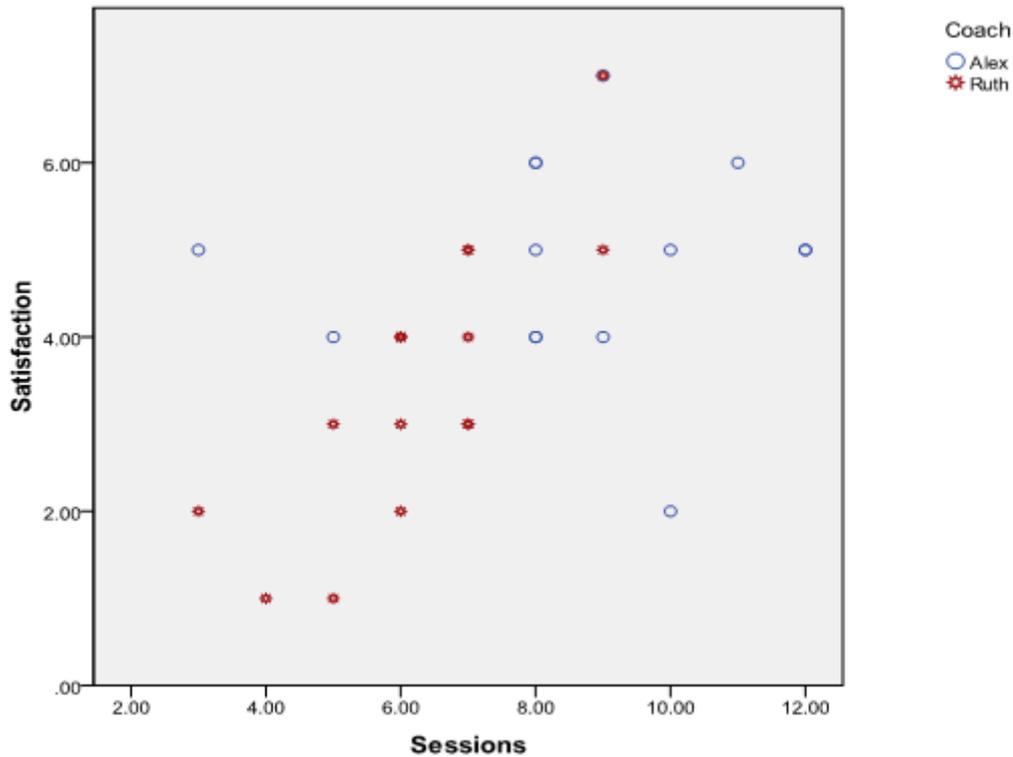
Graph 9 proves that there might be some sort of correlation between number of sessions and the satisfaction.

		Satisfaction	Sessions
Satisfaction	Pearson Correlation	1	.451*
	Sig. (2-tailed)		.012
	N	30	30
Sessions	Pearson Correlation	.451*	1
	Sig. (2-tailed)	.012	
	N	30	30

\*. Correlation is significant at the 0.05 level (2-tailed).

**Table 13 Pearson correlation - sessions and satisfaction for Alex and Ruth**

Due to significance of 0.012 the null hypothesis can be rejected with the low risk of Type I error (Bryman and Bell, 2007: 370). It is not a coincidence that we have noticed during coaching process sessions-satisfaction weak positive correlation. Even if correlation does not prove the cause, “it does mean that a casual explanation is possible” (de Vaus, 2001: 178). This issue was further investigated with the coach being suspect to the statistical control.



**Graph 10 Scatterplot - sessions and satisfaction separated for Alex and Ruth**

Paying attention to the above presented scatterplot it may become apparent that for Ruth such a correlation (sessions-satisfaction) is clearly visible while for Alex it is not. To assure that idea in case of Ruth Pearson correlation has been calculated.

**Correlations**

		Sess	Satis
Sess	Pearson Correlation	1	.800**
	Sig. (2-tailed)		.000
	N	16	16
Satis	Pearson Correlation	.800**	1
	Sig. (2-tailed)	.000	
	N	16	16

\*\* . Correlation is significant at the 0.01 level (2-tailed).

**Table 14 Pearson correlation - sessions and satisfaction for Ruth**

For Ruth this very strong correlation of 0.8 is statistically very significant  $p < 0.05$ . It may be assumed that for her achieved satisfaction level depends on the number of sessions. How did it work in the case of Alex?

		Sess	Satis
Sess	Pearson Correlation	1	.058
	Sig. (2-tailed)		.844
	N	14	14
Satis	Pearson Correlation	.058	1
	Sig. (2-tailed)	.844	
	N	14	14

**Table 15 Pearson correlation - sessions and satisfaction for Alex**

For Alex correlation is very weak 0.58 and it is not in anyway statistically significant. Therefore it designates for him lack of any correlation between number of sessions and achieved satisfaction of senior managers. Probably Alex really possesses some valuable coaching skills that are independent on the number of sessions provided. Regardless on the number of sessions he achieves relatively high satisfaction levels. It could be considered as a good idea to get to know closer his best practises.

To be self-critic about this conclusion it could be said that the sample size is relatively to small to be certain about this result. Therefore such an observation should be confirmed in the next coaching sessions provided. This topic must be clarified. Maybe Ruth could have much higher scores if she would be able to maintain senior manager in the larger number of sessions? The direction of causality between sessions and satisfaction here must be qualitatively discussed with Ruth.

Subsequent findings are that there is no any “visible” correlation between independent variable age and dependent variable number of sessions as well as between independent variable age and dependent variable satisfaction. It can be also concluded that the difference in number of sessions that male and female managers have participated in is statistically significant. Although here statistical significance is not a very strong one.

Some tasks could be suggested to improve our coaching services.

Before starting the next future coaching session qualitative unstructured interview with Alex could be conducted to shed some more light on his best practises. Also as indicated earlier the similar with different focus open interview should be conducted with Ruth. Both results could be considered as a part of the lessons learned. Such qualitative small additional research would certainly enrich gathered quantitative data and could shed some more light on the raised unanswered issues. In fact such an approach could designate more mixed methods oriented approach toward evaluation of coaching programmes.

As regards the survey it would be more interesting to move toward more experimental design. Coaching in fact is about providing a certain treatment (de Vaus, 2001: 48) to the observed senior managers group. Development of the senior managers requires to check their skills before and after that the coaching sessions. Therefore it implies necessity to rebuild survey and add additional variables in order not only to measure satisfaction but also

a certain business skills. Also in the future, questionnaire might be enriched with the open questions. It could help a lot to clarify certain issues raised in the above presented analysis.

## Appendices

### Appendix 1: SPSS variables view

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Senior_manager	Numeric	8	2		None	None	8	Right	Nominal
2	Sex	Numeric	8	2		(1.00, female)...	None	8	Right	Nominal
3	Age	Numeric	8	2		None	None	8	Right	Scale
4	Coach	Numeric	7	2		(1.00, Alex)...	None	8	Right	Nominal
5	Sessions	Numeric	8	2		None	None	8	Right	Scale
6	Satisfaction	Numeric	8	2		None	None	8	Right	Scale

### Appendix 2: Data from the satisfaction questionnaire as in SPSS

Senior manager	Sex	Age	Coach	Sessions	Satisfaction
1.00	1.00	41.00	2.00	6.00	4.00
2.00	1.00	27.00	2.00	6.00	4.00
3.00	2.00	25.00	1.00	9.00	7.00
4.00	2.00	26.00	1.00	12.00	5.00
5.00	2.00	21.00	1.00	8.00	5.00
6.00	2.00	41.00	2.00	4.00	1.00
7.00	2.00	33.00	1.00	8.00	6.00
8.00	1.00	49.00	1.00	8.00	4.00
9.00	2.00	32.00	1.00	9.00	4.00
10.00	1.00	33.00	2.00	7.00	3.00
11.00	1.00	45.00	1.00	10.00	2.00
12.00	2.00	22.00	1.00	8.00	4.00
13.00	2.00	31.00	1.00	5.00	4.00
14.00	1.00	47.00	2.00	9.00	7.00
15.00	1.00	38.00	2.00	5.00	1.00
16.00	2.00	35.00	2.00	7.00	4.00
17.00	2.00	33.00	1.00	12.00	5.00
18.00	2.00	36.00	2.00	7.00	5.00
19.00	2.00	24.00	1.00	11.00	6.00
20.00	1.00	38.00	2.00	7.00	3.00
21.00	1.00	43.00	1.00	3.00	5.00
22.00	1.00	42.00	2.00	7.00	5.00
23.00	1.00	38.00	2.00	3.00	2.00
24.00	1.00	28.00	2.00	5.00	3.00
25.00	1.00	34.00	2.00	9.00	5.00

26.00	1.00	40.00	1.00	8.00	6.00
27.00	2.00	35.00	1.00	10.00	5.00
28.00	2.00	44.00	2.00	6.00	3.00
29.00	1.00	39.00	2.00	6.00	4.00
30.00	1.00	36.00	2.00	6.00	2.00

### References

Bryman, A. and Bell, E. (2007) *Business research methods* Oxford University Press: Oxford, 2<sup>nd</sup> edtn.

De Vaus, David A. (2001) *Research design in social research* Sage Publications Ltd.: London.

Easterby-Smith, M., Thorpe, R., Jackson, Paul R. (2008) *Management research* Sage Publications Ltd.: London, 3<sup>rd</sup> edtn.

Greasley, P. (2008) *Quantitative Data Analysis Using SPSS: An Introduction for Health & Social Science* Open University Press: Maidenhead.