Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Research Methods

M5X01282

Student: Georgiana Florina Cune

Appraisal and analysis of 'Visits to the local area' questionnaire

1. Introduction

A questionnaire is the main instrument used for collecting data for a survey research. The survey is presented to individuals who have been selected from a population of interest. For a survey to be successful and useful for getting information, it must be conducted properly and well-thought before it's given out to respondents.
In order to avoid and minimize bias, the sample of individuals must be selected randomly with every member of the target population to have an equal chance of being included in the sample (Rumsey, 2010).
The analysis carried out in this paper is obtained from a questionnaire carried out by a 2nd undergraduate student who was employed by the Welsh Tourist Board to survey the visitors to various beauty sports around the coast of South West Wales during the summer vacation. The content of the questionnaire will be critically appraised and analysed in IBM's Statistical Package for Social Science (SPSS) after a data cleansing will be conducted to correct any anomalies.

2. Appraisal of the questionnaire and improvement suggestion for future surveys

According to (Rumsey, 2010), the planning and process of a survey can be broken down into ten elements that should be checked:

1. Target population is well defined.
2. Sample matches the target population.
3. Sample is randomly selected.
4. Sample size is large enough.
5. Nonresponse is minimized.
6. Type of survey is appropriate.
7. Questions are well worded.
8. Survey is properly timed.
9. Personnel are well trained.
10. Proper conclusions are made.

The questionnaire is showing data from the survey carried by the student. The survey contains questions related to different variables.
The questionnaire looks is rather busy and can confuse the respondents. In completing the survey, most respondents will lose interest because the questions to answer are too long. In order for a survey to be answered by all the participants and obtain the answers needed to perform the analysis, short time for completion and ease for answering the questions are two of the key elements that enable a high questionnaire response.
The data from questions 4.1 to 4.21 and 4a.0 to 4a.5 is richer, but much more complex to analyse.

In the student's questionnaire, the target population is the visitors to South West Wales who visit various beauty sports in the summer. The sample of 464 individuals was randomly selected. The sample size is large enough, with a margin of error of 0.046 or 4.6% ($\pm 1/\sqrt{464}$).

3. Clearing the data

All the variables in SPSS Variable View were treated as 'Nominal', but the variables from 4.1. to 4a.5. need to be changed to 'Ordinal' because they include a degree of satisfaction among the visitors surveyed, where values range from 1 = 'Very dissatisfied' to 5 = 'Very satisfied' and 6 = 'Don't know' (Chetty, 2015). Other variables changed to 'Ordinal' are the values from question 11., 12., 13., 14., 15., 20., and 22, because the values represent categories with some intrinsic ranking (Chetty, 2015). The variables from questions 10 and 23 are changed and treated as a 'Scale' because their values represent ordered categories.

Looking on the data it can be seen that some participants did not complete all the questions in the survey, or the data set is already completed and has missing values, therefore the data has to be cleansed in SPSS to correct any anomalies.
There are participants who did not answer at all the first question *1. How often do you visit the coastal area?* It is not a bad idea to check if the participants, who did not answer the first question, did not answer the rest of the questions as well. It may be an evidence of a response set since for these participants was entered the score '99' for every variable even if that did not make sense.

Statistics

		1.1: I live Here	1.2: 2nd Home	1.3: 1st Visit	1.4: Monthly Visits	1.5: Visit 2+ times a Year	1.6: Visit Annually	1.7: Visit Every other Year	1.8: Visit couple times in 5 years	1.9: Visit once in 5 years
N	Valid	4	4	4	4	4	4	4	4	4
	Missing	0	0	0	0	0	0	0	0	0

Table 1. Number of respondents who have not answered question 1.

From the data analysis in SPSS, 4 respondents did not answer question 1 at all *(Table 1).*

Questionnaire Identifier

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	94	1	25.0	25.0	25.0
	154	1	25.0	25.0	50.0
	168	1	25.0	25.0	75.0
	372	1	25.0	25.0	100.0
	Total	4	100.0	100.0	

Table 2. Questionnaire Identifier for participants who did not answer question 1.

In this case, based on the fact that there is 'No reply' on the first question and '99' value everywhere on the Data View in SPSS, these candidates (ID: 94,154, 168, 372 in the SPSS Data View) as can be seen in Table 2, will be worthwhile for deleting them.
Question *7. In which locality did you spend MOST OF YOUR TIME during your most recent visit to this local area?* there are values which have not been recorded in the variable data set in SPSS. The *MostTime* variable was recorded by the student in SPSS as type string and calculations cannot be performed on this type of variable, therefore the variable *MostTime* has been automatic recoded into a new numeric variable *MostTimeNumeric,* where calculation can be performed when running a SPSS analysis (Table 3). SPSS has created new variables that no longer have alphanumeric characters. The new variable has numeric value labels showing the original values.

MostTime into MostTimeNumeric (7 MostTime)	
Old Value	New Value - Value Label
01	1 Aberystwyth
02	2 Borth
03	3 Borth & Ynys Las
04	4 Ceredigion
05	5 Clarach
06	6 Machynlleth Area
08	7 Ynys Las
09	8 Ynys Las - Aberystwyth
11	9 Pembrokeshire
12	10 Wales
13	11 Swansea/Gower
14	12 Carmarthenshire
15	13 Gwynedd
16	14 Monmouthshire
17	15 Powys
99	16 No Reply
Aberaeron	17 Aberaeron
Aberafan	18 Aberafan
Aberavon	19 Aberavon
Aberdovey	20 Aberdovey
Aberdyfi	21 Aberdyfi
Aberdyfi & Snowdonia	22 Aberdyfi & Snowdonia
Aberystwyth - Barmouth	23 Aberystwyth - Barmouth
Aberystwyth - Borth	24 Aberystwyth - Borth

Aberystwyth - Borth - Ynys Las	25 Aberystwyth - Borth - Ynys Las
Aberystwyth - Llandudno	26 Aberystwyth - Llandudno
Aberystwyth & Ynys Las	27 Aberystwyth & Ynys Las
Aberystwyth Borth	28 Aberystwyth Borth
Aberystwyth Clarach Bay	29 Aberystwyth Clarach Bay
Aberystwyth-Corris-NantYrArian	30 Aberystwyth-Corris-NantYrArian
Aberystwyth-DevilsBridge-Pembr	31 Aberystwyth-DevilsBridge-Pembr
Anglesey	32 Anglesey
Arthog	33 Arthog
Barmouth - Tywyn	34 Barmouth - Tywyn
Beach	35 Beach
Blackpill & Clyne	36 Blackpill & Clyne
Blaenau Gwent	37 Blaenau Gwent
Blaenavon	38 Blaenavon
Borth - Aberaeron	39 Borth - Aberaeron
Borth - Aberystwyth - Rhayder	40 Borth - Aberystwyth - Rhayder
Borth - Tywyn	41 Borth - Tywyn
Borth & area	42 Borth & area
Borth-YnysLas-Ceredigion	43 Borth-YnysLas-Ceredigion
Bryn Uchel	44 Bryn Uchel
Builth Wells	45 Builth Wells
Cardigan	46 Cardigan
Carneddau Mountains	47 Carneddau Mountains
Ceredigion - Carmarthenshire	48 Ceredigion - Carmarthenshire
Clarach-Borth-YnysLas	49 Clarach-Borth-YnysLas
Coast	50 Coast
Conwy	51 Conwy
Cornwall	52 Cornwall
Criccieth	53 Criccieth
Darowen	54 Darowen
Dartmoor	55 Dartmoor
Dinas Mawddwy	56 Dinas Mawddwy
Dolgellau	57 Dolgellau
Dollgellau - Borth	58 Dollgellau - Borth
Dovey Estuary	59 Dovey Estuary
Dyfi Valley - Machynlleth	60 Dyfi Valley - Machynlleth
Eglwys Fach - Machynlleth	61 Eglwys Fach - Machynlleth
Furnace	62 Furnace
Furnace - Ynys Las	63 Furnace - Ynys Las
Glas Pwll	64 Glas Pwll
Glyncorrwg	65 Glyncorrwg
Gower	66 Gower
Gower beach	67 Gower beach
Gower beaches	68 Gower beaches
Gower Coast	69 Gower Coast
Kenfig	70 Kenfig
Llandre	71 Llandre
Llandrindod	72 Llandrindod
Llandystrad	73 Llandystrad
Llanelli	74 Llanelli
Llangennith	75 Llangennith
Llangurig	76 Llangurig
Llanmadoc	77 Llanmadoc
Llanmadoc/Fairwood/Mumbles	78 Llanmadoc/Fairwood/Mumbles

Llanon - Aberaeron	79 Llanon - Aberaeron
Llanrhystud	80 Llanrhystud
Llantwit Major	81 Llantwit Major
Local Park	82 Local Park
Mid Wales	83 Mid Wales
Mid Wales Coast	84 Mid Wales Coast
Mumbles	85 Mumbles
Mumbles/Gower	86 Mumbles/Gower
Mumbles/Southgate/Llanmadoc	87 Mumbles/Southgate/Llanmadoc
Newton, Powys	88 Newton, Powys
No Reply	89 No Reply
Northumbria	90 Northumbria
Ogmore	91 Ogmore
Oxwich	92 Oxwich
Pembrey	93 Pembrey
Pembrokeshire	94 Pembrokeshire
Pembrokeshre	95 Pembrokeshre
Pendine	96 Pendine
Port Eynon	97 Port Eynon
Powys	98 Powys
Rhayder	99 Rhayder
Rheidiol Valley	100 Rheidiol Valley
Rhossili	101 Rhossili
Seaside town	102 Seaside town
Snowdonia	103 Snowdonia
South Wales	104 South Wales
Southerndown	105 Southerndown
Southgate	106 Southgate
St Davids - Devils Bridge	107 St Davids - Devils Bridge
Swansea	108 Swansea
Swansea/Mumbles	109 Swansea/Mumbles
Trefenter near Aberystwyth	110 Trefenter near Aberystwyth
Vale of Glamorgan	111 Vale of Glamorgan
Village 3 miles inland	112 Village 3 miles inland
Ynys Hir - Ynys Las -Aberdovey	113 Ynys Hir - Ynys Las -Aberdovey
Ynys Las - Borth	114 Ynys Las - Borth

Table 3. Automatic recode of MostTime variable into MostTimeNumeric (7 MostTime) variable

The case is the same for variable *19. Home*, where string type variable had to be be automatic recoded into new variable *HomeNumeric.* In this table there are user-missing values coded -1 and -2, which must be changed into missing values for SPSS to recognize and transform them into system-missing values.
The student has created new variables ‘Walking’ where they could have been all integrated in the Scale category Walking: Under 3 miles, 3-5 miles, 5-9 miles, 10+ miles. They have to be recoded into different variables.
The output in Table 5 is the resulting frequency table, showing that out of a total of 464 respondents 168 are male, 261 are female, 31 did not reply and 4 have the value ‘Joint’. Because the data on gender refers to male and female only, the ‘Joint’ value will be considerate missing. The valid percentage, calculated after the missing values are ignored, shows that 36.5% of respondents are male, 56.7% are female and 6.7% didn’t answer the question.

21: Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	168	36.2	36.5	36.5
	Female	261	56.3	56.7	93.3
	No Reply	31	6.7	6.7	100.0
	Total	460	99.1	100.0	
Missing	Joint	4	.9		
Total		464	100.0		

Table 5. Gender count of respondents

4. Data analysis

The primary data was analysed by the student using IBM's Statistical Package for Social Science (SPSS).

a. *Chi-square test for annually visits to the coast (Q1) and the belief that the visits are in the summer season (Q2)*

H_1 – There is no difference between the annually visits to the coast and the belief that the visits are in the summer.
H_a – There is a difference between the annually visits to the coast and the belief that the visits are in the summer.

The two variables used are the frequency of how often people visit the coastal area from Q1: frequently – annually and whether these visits are in the summer or not. The Chi-square test function will be used to test this hypothesis. Before using the Chi-square test analysis, it is important to select only the respondents who answered the question, therefore using only Yes '1' / No '0' variables and eliminating the No answer variable '99'.

The Chi-square test function suggests that the assumption for the expected count is not less than 5 or 20% of the cells have expected count greater than 5.

Table 2.1.1. Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
1.6: Visit Annually * 2.2: Visited in Summer	433	100.0%	0	0.0%	433	100.0%

Table 2.1.2. - 1.6: Visit Annually * 2.2: Visited in Summer Crosstabulation

			2.2: Visited in Summer		
			No	Yes	Total
1.6: Visit Annually	No	Count	51	330	381
		Expected Count	53.7	327.3	381.0
	Yes	Count	10	42	52
		Expected Count	7.3	44.7	52.0

Total	Count	61	372	433
	Expected Count	61.0	372.0	433.0

Table 2.1.3 - Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.292[a]	1	.256		
Continuity Correction[b]	.854	1	.355		
Likelihood Ratio	1.196	1	.274		
Fisher's Exact Test				.286	.176
Linear-by-Linear Association	1.289	1	.256		
N of Valid Cases	433				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.33.

b. Computed only for a 2x2 table

The Chi-square test results *(Table 2.1.3):*

The assumption for a 2x2 table that all the expected counts are at least 10 has been violated because one of the expected counts (7.3) is less than 10. Because of this the Fisher's Exact Test is used. The significance value (2- sided) .286 is greater than α=0.05, which means that the result is not statistically significant and accept the null hypothesis (H_1), there is no difference between the annually visits to the coast and the belief that the visits are in the summer, and reject the alternative hypothesis (H_a).

b. Chi-square test for gender (Q21) and who attends the sport events (Q3)

H_1 – There is no difference between gender and that males attend more sport events
H_a – There is a difference between and that males attend more sport events

Table 2.2.1. Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
21: Gender * 3.27: Attend Sport Event	404	93.3%	29	6.7%	433	100.0%

Table 2.2.2. - 21: Gender * 3.27: Attend Sport Event Crosstabulation

			3.27: Attend Sport Event		
			No	Yes	Total
21: Gender	Male	Count	152	7	159
		Expected Count	151.5	7.5	159.0
	Female	Count	233	12	245
		Expected Count	233.5	11.5	245.0
Total		Count	385	19	404
		Expected Count	385.0	19.0	404.0

Table 2.2.3. Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.053[a]	1	.818		
Continuity Correction[b]	.000	1	1.000		
Likelihood Ratio	.053	1	.818		
Fisher's Exact Test				1.000	.511
Linear-by-Linear Association	.053	1	.818		
N of Valid Cases	404				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.48.

b. Computed only for a 2x2 table

Chi-square= .053a, degrees of freedom = 1 and significance value p= .818.

With p value =.818 and greater than α=0.05, which means that the result is not statistically significant and accept the null hypothesis (H_1), there is no difference between gender and that males attend more sport events and reject the alternative hypothesis (H_a).

The crosstabulation function in SPSS showed that out of 404 respondents, 153 males and 233 females do not attend sport events, whereas only 7 males and 12 females attend sport events.

c. *Chi-square test for reason of visit pleasure (Q5) and the month to most recent trip to the area for recreation (Q6)*

H_1 – There is no difference between visiting the coast for pleasure and month in which most recent trip was taken.
H_a – There is a difference between visiting the coast for pleasure and month in which most recent trip was taken.

2.3.1. Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
5.2: Here on Pleasure * 6: Which Month was your most recent visit	464	100.0%	0	0.0%	464	100.0%

2.3.2. - 5.2: Here on Pleasure * 6: Which Month was your most recent visit Crosstabulation

			6: Which Month was your most recent visit											
			Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	No Reply	Total
5.2: Here on Pleasure	No	Count	0	0	0	0	0	0	4	1	1	0	4	10
		Expected Count	.0	.1	.1	1.4	1.4	.9	3.3	1.4	.3	.0	.9	10.0
	Yes	Count	1	3	5	59	58	36	136	63	12	2	26	401
		Expected Count	.9	2.6	5.2	55.3	56.2	38.0	133.1	57.0	13.8	1.7	37.2	401.0
	No Reply	Count	0	0	1	5	7	8	14	2	3	0	13	53
		Expected Count	.1	.3	.7	7.3	7.4	5.0	17.6	7.5	1.8	.2	4.9	53.0
Total		Count	1	3	6	64	65	44	154	66	16	2	43	464
		Expected Count	1.0	3.0	6.0	64.0	65.0	44.0	154.0	66.0	16.0	2.0	43.0	464.0

2.3.3. Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	42.736[a]	20	.002
Likelihood Ratio	39.974	20	.005
Linear-by-Linear Association	16.178	1	.000
N of Valid Cases	464		

a. 20 cells (60.6%) have expected count less than 5. The minimum expected count is .02.

2.3.4. Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.303	.002
	Cramer's V	.215	.002
N of Valid Cases		464	

Chi-square = 42.736a, degrees of freedom df= 20 and significance value p=.002
With p value =.002 and less than α=0.05, means that the result is statistically significant and reject the null hypothesis (H_1), and accept the alternative hypothesis (H_a), there is a is a difference between visiting the coast for pleasure and month in which most recent trip was taken. Cramer's V (3x11), shows there is a positive relationship between visiting the coast for pleasure and the month in which recent trip was taken.

5. Conclusion

Questionnaires are an important tool when collecting data for a survey research. It is important to have a target population well defined with a randomly selected population. In order for a survey to be answered by all the participants and obtain the answers needed to perform the analysis, short time for completion and ease for answering the questions are two of the key elements that enable a high questionnaire response.