

Technical Report:

Regression analysis of 2018-19 National Survey for Wales results

This report sets out our basic approach to logistic regression analysis of National Survey results. The first section gives brief details of the approach we have used; the annex contains details of the regression models that we have produced.

Regression analysis

Regression analysis goes beyond descriptive statistics in which the relationship between one independent and one dependent variable is explored. Whilst descriptive statistics are quick and easy to produce and the findings can be useful, they don't take account of the complicated relationships between variables. Regression analysis allows for the relationship between an explanatory variable and the outcome variable to be examined whilst at the same time taking into consideration other explanatory variables that have an effect on the outcome.

The analysis we have used for most survey outcomes is binary logistic regression. Logistic regression is used as it is suitable when looking at categorical outcomes (which is the form taken by most National Survey variables). While it is possible to conduct multinomial logistic regression with multiple categorical outcomes, we usually use logistic regression with binary outcomes (e.g. 'satisfied with visit to GP' vs. 'not satisfied') in order to increase ease of understanding. Outcome variables with more than two outcomes are coded into a binary format prior to regression analysis (for example, merging 'very satisfied' and 'fairly satisfied' together to form a 'satisfied' category). Logistic regression is then used to predict the likelihood of being in a particular category based on the values of the independent variables (predictors).

Procedure

Backwards logistic regression is used in order to create the final models. Firstly, the dependent variable and all other relevant variables are investigated using descriptive statistics. Those that are insignificant predictors at the 0.05 level are dropped. All of the remaining significant variables are placed in the initial model. The contribution of each variable is assessed by looking at the significance value of the t-test for each predictor. The variable with the highest p-value is removed, and the procedure repeated, until only the variables that are significant at the 0.05 level are included. There are multiple different ways in which variables could be entered into the model. We usually use backwards selection as forward approaches often allow for important variables to be missed due to other variables being entered into the model first ("suppressor effects").

Multicollinearity

Many of the variables collected in the National Survey are correlated with one another. Multicollinearity (also known as collinearity) is where one or more explanatory variables in a regression model are highly correlated such that they linearly predict each other with a high

degree of accuracy. However, a key assumption of multivariate regression is that explanatory variables are not too highly correlated with one another. Too high a degree of correlation between predictor variables in a regression model can affect the stability and interpretation of the regression estimates. Therefore, the variables included in the model are tested for multicollinearity.

High multicollinearity can be assessed using the variance inflation factor (VIF) statistic. There is differing advice on what constitutes an acceptable degree of multicollinearity. It is generally suggested that “if the largest VIF is greater than 10 then there is cause for concern”, “if the average VIF is substantially greater than 1 then the regression may be biased” and “tolerance below 0.1 indicates a serious problem” (Field, 2013, p.325)¹.

The VIF is tested firstly with all relevant variables included. Individual variables that fail to meet the assumptions above are removed from the model. The VIF is then tested again on the final model to double-check for multicollinearity. In all of the models the individual variable VIF is no greater than 10 and the mean VIF is no greater than 2.5, suggesting that there is no cause for concern in these models.

Goodness of fit

Goodness of fit describes how well a model fits the data from which it is generated. It can be used to assess how well the data that the model predicts, corresponds to the data that has been collected. It can be measured using the R^2 statistic. The R^2 statistic is the coefficient of determination for multiple regression models and measures how well the data fits the model. It is reported as a percentage of the variation of the outcome explained by the variables included. As with any regression using survey data, we can only consider the variables for which we have data. Associations could be due to some unmeasured factor, and there may also be important unmeasured factors which are simply not captured in the model. Therefore, in social science research R^2 can appear to be low, but this can be due to the complex nature of outcomes being investigated: a model with a low R^2 can still be useful in understanding the relationships between variables.

When calculating R^2 statistics, the data cannot be weighted. The (un-weighted) R^2 statistic for each model is included in the appendices below.

Interaction effects

Interactions can be used to test for the joint effect of two or more predictor variables on an outcome variable. It allows us to explore how the relationships between dependent and independent variables differ by context. For example, when researching the determinants of well-being, prior research showed that the effect of socio-economic conditions on well-being was different for males to females. Therefore, an interaction between household deprivation and gender was added to the model. This found that in deprived households females reported higher well-being than males; however in less deprived households it was the other way round. Looking at this interaction allows a more detailed understanding of how various variables interact to influence the dependent variable.

Causality

Regression analysis can identify relationships between factors; however, it cannot tell us about causality. While for some factors causality is fairly clear based on prior knowledge (e.g.

¹ Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

material deprivation does not cause changes in gender; gender causes changes in material deprivation), for others the relationship between cause and effect is more blurred (e.g. low life satisfaction can cause material deprivation; material deprivation can cause low life satisfaction). Therefore, where prior knowledge does not make the direction of causality clear we have generally noted that causality can operate in either direction (or both).

Weighting

The results of the National Survey are weighted to compensate for unequal selection probabilities and differential non-response (i.e. to ensure that the age and sex distribution of the final dataset matches that of the population of Wales). Our regression models take the weights into account. For details of how the weights are calculated, see [National Survey for Wales 2018-19 Technical Report](#).

Marginal effects

The results are sometimes presented using marginal effects. This differs from the usual regression output, odds ratios. Odds are the probability of an event occurring divided by the probability of the event not occurring. Odds ratios are the ratio between two sets of odds. Odds ratios are somewhat abstract and can often be hard to interpret; indeed, are often misinterpreted². Therefore, we turn the odds ratios in the model into predicted probabilities (risks). Using these we can calculate the probability of an individual in a specified group (e.g. female) meeting the regression criteria (e.g. being in material deprivation) and compare it with the probability for individuals not in the group (e.g. males). This is known as a marginal effect.

Where we have used this approach the results presented are Average Marginal Effects (AMEs). As an example, an AME for the effect of material deprivation on the probability of internet use would be calculated as follows:

1. Generate a logistic regression model for internet use, including material deprivation as one of the predictors.
2. Start at the first person in the dataset.
3. Use the regression model to calculate a predicted probability that this person uses the internet, using their characteristics to set the values for all factors in the model except for material deprivation: set this factor to be 'in material deprivation'. Record the predicted probability generated by the model for that person.
4. Repeat for all the other people in the dataset.
5. Take the mean of the predicted probabilities made for all these people. This is the average adjusted predicted probability of using the internet for living in a deprived household.
6. Do steps 2 to 4 again, except this time setting the material deprivation factor for each person as being 'not in material deprivation'.
7. Take the mean of the predicted probabilities made for all these people. This is the average adjusted predicted probability of using the internet for people not in material deprivation.
8. The difference between the two mean predicted probabilities calculated at steps 5 and 7 is the Average Marginal Effect of material deprivation on internet use.

² See <http://www.bmj.com/content/316/7136/989> for a brief discussion of the issues.

For a more detailed description of Average Marginal Effects please see Williams (2012)³.

Statistically significant differences

Estimates from the National Survey are subject to a margin of uncertainty. Part of the uncertainty comes from the fact that any randomly-selected sample of the population will give slightly different results from the results that would be obtained if the whole population was surveyed. This is known as sampling error. Confidence intervals can be used as a guide to the size of the sampling error. These intervals are calculated around a survey estimate and give a range within which the true value is likely to fall. In 95% of survey samples, the 95% confidence interval will contain the 'true' figure for the whole population (that is, the figure we would get if the survey covered the entire population).

As with any survey, the National Survey is also subject to a range of other sources of error: for example, due to non-response; because respondents may not interpret the questions as intended or may not answer accurately; and because errors may be introduced as the survey data is processed.

Where the text of our reports note a difference between two groups, we have checked to ensure that the confidence intervals for the two groups do not overlap. This suggests that the difference is statistically significant (but as noted above, is not as rigorous as carrying out a formal statistical test), i.e. that there is less than a 5% (1 in 20) chance of obtaining these results if there is no difference between the same two groups in the wider population.

³ Williams, R. (2012). *Using the margins command to estimate and interpret adjusted predictions and marginal effects*. The Stata Journal 12(2):308-331 [Available at: <http://www.stata-journal.com/article.html?article=st0260>].

Annex: Statistical explanations and tables

Internet use and digital skills	6
Climate change and environmental action	7
Community cohesion and safety in the local area	10
Local democracy and access to services.....	14
Childcare.....	20

Internet use and digital skills

Main factors influencing digital skills

Initial list of variables considered: Tenure; Economic status; Has a limiting long-term illness, disability or infirmity; Gender; Age; Highest educational qualification; WIMD overall score (in quintiles); Urban/rural classification; Ability to access information about local authority in the way they would like

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: None

Variables removed as not being significant predictors when part of a regression model:

Tenure; Urban/rural classification; Has a limiting long-term illness, disability or infirmity; Economic status; Gender.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Digital inclusion age group	16-49 (reference group)						
	50-64	0.4732717	0.0418046	-8.47	0	0.3980366	0.5627273
	65-74	0.2853559	0.031555	-11.34	0	0.2297524	0.3544163
	75+	0.1133819	0.0221359	-11.15	0	0.0773323	0.1662366
Highest educational qualification	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	0.7046502	0.0925525	-2.67	0.008	0.5447186	0.9115384
	National Qualification Framework level 2	0.5249024	0.0561408	-6.03	0	0.4256362	0.6473192
	Below National Qualification Framework level 2	0.4545248	0.0939818	-3.81	0	0.3030786	0.6816476
	No qualification	0.2826085	0.0331469	-10.77	0	0.2245683	0.3556492
WIMD overall score (in quintiles)	Most deprived 20% (reference group)						
	Q2	1.486125	0.1888278	3.12	0.002	1.158514	1.90638
	Q3	0.9772703	0.1291978	-0.17	0.862	0.7541949	1.266327
	Q4	1.183438	0.1501617	1.33	0.184	0.9228689	1.517578
	Least deprived 20%	1.749065	0.2316088	4.22	0	1.349245	2.267362
Ability to access information about local authority in the way they would like	Agree (reference group)						
	Neither agree nor disagree	0.5981919	0.0839106	-3.66	0	0.454401	0.7874841
	Disagree	0.7736365	0.0964393	-2.06	0.04	0.6059387	0.9877459
	Constant	1.335895	0.1484338	2.61	0.009	1.074468	1.660928

Statistics

Number of obs = 4,733
 Wald chi²(13) = 423.48
 Prob > chi² = 0.0000

Log pseudolikelihood = -3046.21
 Pseudo R² = 0.1062

Climate change and environmental action

Main factors influencing being very concerned about climate change

Initial list of variables considered: age; gender; whether household contains a child under 16; highest educational qualification; Welsh speaker; tenure; working status of household; material deprivation; WIMD overall score (in quintiles); urban/rural classification; overall satisfaction with life (grouped); participate in sports 3 times a week; religion; ethnicity; sexual orientation; internet use; and bills & credit commitments.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: None

Variables removed as not being significant predictors when part of a regression model:

Sexual orientation; WIMD overall score (in quintiles); internet use; bills & credit commitments; participate in sports 3 times a week; urban/rural classification; whether household contains a child under 16; ethnicity; gender; religion; working status of household; Welsh speaker; tenure; and in material deprivation.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Age (5 groups)	16-24 (reference group)						
	25-44	0.7652516	0.1494243	-1.37	0.171	0.5219119	1.122048
	45-64	1.134917	0.2113232	0.68	0.497	0.7878961	1.634779
	65-74	1.224011	0.2459737	1.01	0.314	0.8255255	1.814848
	75+	1.203432	0.2501369	0.89	0.373	0.8007497	1.808616
Highest educational qualification	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	0.5882027	0.0948783	-3.29	0.001	0.428772	0.8069145
	National Qualification Framework level 2	0.715173	0.0983783	-2.44	0.015	0.5461612	0.936486
	Below National Qualification Framework level 2	0.5939289	0.1313748	-2.36	0.019	0.3849923	0.9162562
	No qualification	0.5878814	0.0747656	-4.18	0	0.4581795	0.7542996
Overall satisfaction with life (grouped)	Low (1-4) (reference group)						
	Medium (5-6)	0.8581665	0.2190554	-0.6	0.549	0.5203482	1.415302
	High (7-8)	0.6477135	0.1555223	-1.81	0.07	0.404577	1.036966
	Very high (9-10)	0.8276289	0.2024683	-0.77	0.439	0.51239	1.336813
	Constant	1.003226	0.304485	0.01	0.992	0.5534188	1.818626

Statistics

Number of obs = 5,334
Wald chi²(11) = 43.64
Prob > chi² = 0.0000
Log pseudolikelihood = -3493.94
Pseudo R² = 0.0165

Main factors influencing purchase of energy efficient appliances

Initial list of variables considered: age; gender; whether household contains a child under 16; highest educational qualification; Welsh speaker; tenure; working status of household; in material deprivation; WIMD overall score (in quintiles); urban/rural classification; overall satisfaction with life (grouped); ethnicity; internet use; bills & credit commitments; and general health.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: None

Variables removed as not being significant predictors when part of a regression model:

Welsh speaker; whether household contains a child under 16; gender; bills & credit commitments; in material deprivation; age group; general health; WIMD overall score (in quintiles); urban/rural classification; and ethnicity.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Highest educational qualification	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	0.6371296	0.1025118	-2.8	0.005	0.4648069	0.8733392
	National Qualification Framework level 2	0.6443393	0.081349	-3.48	0	0.5030937	0.82524
	Below National Qualification Framework level 2	0.6064492	0.1337646	-2.27	0.023	0.3935906	0.9344243
	No qualification	0.4943465	0.0612993	-5.68	0	0.3876875	0.630349
Tenure	Owner-occupier (reference group)						
	Social housing	0.594939	0.0833529	-3.71	0	0.452081	0.7829403
	Private rented	0.5674564	0.0820341	-3.92	0	0.4274437	0.7533315
Working status of household	All working (reference group)						
	Some working	0.833692	0.1006539	-1.51	0.132	0.658018	1.056266
	None working	0.6290742	0.0891155	-3.27	0.001	0.4765621	0.8303942
	No one aged 16-19 (not in FT education) nor aged 19-64 in household	0.6963726	0.0843807	-2.99	0.003	0.5491614	0.8830459
Overall satisfaction with life (grouped)	Low (1-4) (reference group)						
	Medium (5-6)	1.075217	0.2677298	0.29	0.771	0.6600037	1.751644
	High (7-8)	1.383753	0.3261037	1.38	0.168	0.8718848	2.196129
	Very high (9-10)	1.405586	0.3347302	1.43	0.153	0.8813532	2.241636
Internet use	Yes (reference group)						
	No	0.4316038	0.0652253	-5.56	0	0.3209586	0.5803922
	Constant	1.385147	0.3406671	1.32	0.185	0.8553598	2.24307

Statistics

Number of obs = 5,766
Wald chi²(13) = 185.69
Prob > chi² = 0.0000

Log pseudolikelihood = -3759.40
Pseudo R² = 0.0617

Main factors influencing active volunteering to protect the environment

Initial list of variables considered: age; gender; whether household contains a child under 16; highest educational qualification; Welsh speaker; tenure; in material deprivation; built-up areas; overall satisfaction with life (grouped); ethnicity; internet use; bills & credit commitments; general health; sense of community; and satisfaction with ability to access local services and facilities.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None.

Variables removed due to multicollinearity: None

Variables removed as not being significant predictors when part of a regression model: In material deprivation; tenure; sense of community; ethnicity; whether household contains a child under 16; overall satisfaction with life (grouped); Welsh speaker; gender; bills & credit commitments; and age group.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Highest educational qualification	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	0.427184	0.1357947	-2.68	0.007	0.2291038	0.796522
	National Qualification Framework level 2	0.5424099	0.1337525	-2.48	0.013	0.3345264	0.8794777
	Below National Qualification Framework level 2	0.1489811	0.1123575	-2.52	0.012	0.0339765	0.653257
	No qualification	0.4157058	0.0956962	-3.81	0	0.2647512	0.652731
Built-up areas	AS – under 2,000 people is sparsest context (reference group)						
	AL – under 2,000 people in less sparse context	0.9798017	0.2940212	-0.07	0.946	0.5441362	1.764285
	B – 2,000 to 9,999 people	0.4802388	0.1377749	-2.56	0.011	0.2736891	0.8426691
	C – 10,000 to 24,999 people	0.6339249	0.2313536	-1.25	0.212	0.3100189	1.296246
	D – 25,000 to 99,999 people	0.5236613	0.151518	-2.24	0.025	0.2970029	0.9232943
	E – at least 100,000 people	0.5559299	0.1445123	-2.26	0.024	0.3340049	0.92531
Internet use	Yes (reference group)						
	No	0.4279712	0.1417359	-2.56	0.01	0.2236211	0.8190612
General health	Good or very good (reference group)						
	Fair	0.4352029	0.089334	-4.05	0	0.2910488	0.6507554
	Bad or very bad	0.4786772	0.1514582	-2.33	0.02	0.2574626	0.8899616
Satisfied with ability to access local services and facilities	Yes (reference group)						
	No	0.5692737	0.108328	-2.96	0.003	0.3920529	0.8266041
	Constant	0.2175753	0.049086	-6.76	0	0.1398216	0.338567

Statistics

Number of obs = 5,833
Wald chi²(13) = 64.79
Prob > chi² = 0.0000

Log pseudolikelihood = -1296.29
Pseudo R² = 0.0617

Community cohesion and safety in the local area

Main factors influencing dissatisfaction with the local area

Initial list of variables considered: gender; age; whether household contains a child under 16; highest educational qualification; Welsh speaker; people using Welsh in everyday life; tenure; economic status; in material deprivation; WIMD overall score (in quintiles); urban/rural classification; overall satisfaction with life (grouped); participate in sports 3 times a week; religion; ethnicity; sexual orientation; and has a limiting long-term illness, disability or infirmity.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: None

Variables removed as not being significant predictors when part of a regression model: Participate in sports 3 times a week; sexual orientation; highest educational qualification; Welsh speaker; tenure; religion; gender; age; and has a limiting long-term illness, disability or infirmity.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Ethnicity	White (Welsh, English, British, etc.) (reference group)						
	White - Other	0.3551262	0.1213254	-3.03	0.002	0.1817933	0.6937252
	Any other ethnic group	0.7165718	0.1962565	-1.22	0.224	0.4189193	1.225714
Overall satisfaction with life (grouped)	Low (1-4) (reference group)						
	Medium (5-6)	0.60671	0.0903729	-3.35	0.001	0.4530956	0.8124047
	High (7-8)	0.3031111	0.0437951	-8.26	0	0.2283577	0.4023352
	Very high (9-10)	0.2396595	0.0377045	-9.08	0	0.1760674	0.3262198
In material deprivation	In material deprivation (reference group)						
	Not in material deprivation	0.7042219	0.0752122	-3.28	0.001	0.571215	0.8681994
WIMD overall score (in quintiles)	Most deprived 20% (reference group)						
	Q2	0.9324587	0.1127523	-0.58	0.563	0.7357036	1.181834
	Q3	0.677738	0.0869794	-3.03	0.002	0.5270119	0.8715719
	Q4	0.4197185	0.0600502	-6.07	0	0.3170842	0.5555737
	Least deprived 20%	0.2538335	0.0506912	-6.87	0	0.1716176	0.3754361
People using Welsh in everyday life	Yes (reference group)						
	No	1.754613	0.3063242	3.22	0.001	1.246168	2.470506
Urban/rural classification	Urban (reference group)						
	Rural	0.744121	0.0789267	-2.79	0.005	0.6044483	0.9160684
Whether household contains a child under 16	Yes (reference group)						
	No	0.7578075	0.0751341	-2.8	0.005	0.623972	0.9203493
Economic status	Employed (reference group)						
	Unemployed	1.694209	0.3878326	2.3	0.021	1.081715	2.653512
	Economically inactive	1.212444	0.1121443	2.08	0.037	1.011417	1.453426
	Constant	0.3473849	0.0804593	-4.57	0	0.2206279	0.5469673

Statistics

Number of obs = 11,818
 Wald chi²(15) = 447.00
 Prob > chi² = 0.0000

Log pseudolikelihood = -2867.54
 Pseudo R² = 0.0879

Main factors influencing a low sense of community cohesion

Initial list of variables considered: gender; age; whether household contains a child under 16; highest educational qualification; Welsh speaker; people using Welsh in everyday life; tenure; economic status; in material deprivation; WIMD overall score (in quintiles); urban/rural classification; overall satisfaction with life (grouped); participate in sports 3 times a week; religion; ethnicity; sexual orientation; and has a limiting long-term illness, disability or infirmity.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: None

Variables removed as not being significant predictors when part of a regression model:

People using Welsh in everyday life; ethnicity; gender; economic status; in material deprivation; sexual orientation; Welsh speaker; participate in sports 3 times a week; and whether household contains a child under 16.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Age (5 groups)	16-24 (reference group)						
	25-44	0.898516	0.0975812	-0.99	0.324	0.7262449	1.111651
	45-64	0.6781888	0.0737269	-3.57	0	0.5480438	0.8392396
	65-74	0.5553536	0.0653964	-4.99	0	0.4408951	0.6995262
	75+	0.4222328	0.0538626	-6.76	0	0.3288271	0.5421712
Has a limiting long-term illness, disability or infirmity	Yes (reference group)						
	No	0.8317092	0.046636	-3.29	0.001	0.745148	0.9283259
Religion	No religion (reference group)						
	Christian (all denominations)	0.7923498	0.043386	-4.25	0	0.7117188	0.8821155
	Another religion	0.456447	0.07509	-4.77	0	0.3306431	0.6301172
Overall satisfaction with life (grouped)	Low (1-4) (reference group)						
	Medium (5-6)	0.7326359	0.0988225	-2.31	0.021	0.5624356	0.9543411
	High (7-8)	0.5037158	0.0621914	-5.55	0	0.3954502	0.6416221
	Very high (9-10)	0.3915229	0.0495319	-7.41	0	0.305542	0.5016992
WIMD overall score (in quintiles)	Most deprived 20% (reference group)						
	Q2	1.014189	0.0860432	0.17	0.868	0.8588222	1.197662
	Q3	0.7669269	0.0657221	-3.1	0.002	0.6483506	0.9071894
	Q4	0.6752073	0.059534	-4.45	0	0.5680485	0.802581
	Least deprived 20%	0.5753208	0.0513417	-6.19	0	0.4830018	0.6852853
Tenure	Owner-occupier (reference group)						
	Social housing	1.118418	0.0888556	1.41	0.159	0.9571458	1.306862
	Private rented	1.321951	0.1115709	3.31	0.001	1.120405	1.559752

Urban/rural classification	Urban (reference group)						
	Rural	0.8265287	0.0453256	-3.47	0.001	0.7422997	0.9203151
Highest educational qualification	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	0.9311434	0.0823845	-0.81	0.42	0.7828978	1.10746
	National Qualification Framework level 2	0.9641704	0.0702099	-0.5	0.616	0.8359305	1.112083
	Below National Qualification Framework level 2	0.8954677	0.1190737	-0.83	0.406	0.6900212	1.162084
	No qualification	0.7988997	0.0553803	-3.24	0.001	0.6974071	0.9151624
	Constant	4.723279	0.8220092	8.92	0	3.358196	6.643261

Statistics

Number of obs = 9,880
Wald chi²(21) = 510.71
Prob > chi² = 0.0000

Log pseudolikelihood = -6576.37
Pseudo R² = 0.0515

Main factors influencing feeling unsafe walking alone after dark

Initial list of variables considered: gender; age; whether household contains a child under 16; highest educational qualification; Welsh speaker; people using Welsh in everyday life; tenure; economic status; in material deprivation; WIMD overall score (in quintiles); urban/rural classification; overall satisfaction with life (grouped); participate in sports 3 times a week; religion; ethnicity; sexual orientation; having a sense of community; and has a limiting long-term illness, disability or infirmity.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: None

Variables removed as not being significant predictors when part of a regression model:

Ethnicity; sexual orientation; Welsh speaker; participate in sports 3 times a week; people using Welsh in everyday life; economic status; overall satisfaction with life (grouped); and tenure.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Gender	Male (reference group)						
	Female	3.314849	0.2489353	15.96	0	2.861153	3.840488
Age (5 groups)	16-24 (reference group)						
	25-44	0.6967192	0.0983864	-2.56	0.01	0.5282704	0.918881
	45-64	0.7825838	0.1092247	-1.76	0.079	0.5952908	1.028804
	65-74	0.954616	0.1489477	-0.3	0.766	0.7030993	1.296107
	75+	1.181172	0.1961118	1	0.316	0.8530741	1.635458
Has a limiting long-term illness, disability or infirmity	Yes (reference group)						
	No	0.6180328	0.0448307	-6.63	0	0.5361265	0.7124523

Religion	No religion (reference group)						
	Christian (all denominations)	1.233284	0.0947339	2.73	0.006	1.060911	1.433665
	Another religion	1.610927	0.345323	2.22	0.026	1.058303	2.452121
In material deprivation	In material deprivation (reference group)						
	Not in material deprivation	0.6931733	0.0655832	-3.87	0	0.575847	0.8344044
WIMD overall score (in quintiles)	Most deprived 20% (reference group)						
	Q2	0.7969176	0.0803788	-2.25	0.024	0.6539725	0.9711078
	Q3	0.6003248	0.0651755	-4.7	0	0.4852593	0.7426749
	Q4	0.4782216	0.0535091	-6.59	0	0.384049	0.5954863
	Least deprived 20%	0.4228243	0.048092	-7.57	0	0.3383331	0.5284154
Urban/rural classification	Urban (reference group)						
	Rural	0.4917198	0.0400286	-8.72	0	0.4192039	0.5767798
Whether household contains a child under 16	Yes (reference group)						
	No	0.8262377	0.0781304	-2.02	0.044	0.686458	0.99448
Highest educational qualification	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	1.105854	0.1448519	0.77	0.442	0.855464	1.429531
	National Qualification Framework level 2	1.507626	0.145289	4.26	0	1.248142	1.821057
	Below National Qualification Framework level 2	1.20249	0.1889869	1.17	0.241	0.8836984	1.636285
	No qualification	1.748671	0.1586361	6.16	0	1.463823	2.088947
Sense of community	High (reference group)						
	Low	2.379032	0.168283	12.25	0	2.071046	2.732818
	Constant	0.2297133	0.046332	-7.29	0	0.1547046	0.3410902

Statistics

Number of obs = 9,597
 Wald chi²(20) = 862.43
 Prob > chi² = 0.0000
 Log pseudolikelihood = -4009.73
 Pseudo R² = 0.1475

Local democracy and access to services

Main factors influencing knowing how to find what services the local authority provides

Initial list of variables considered:

Gender; age group; local authority; internet use; highest educational qualification; economic status; has a limiting long-term illness, disability or infirmity; WIMD overall score (in quintiles); in material deprivation; mental well-being score; overall satisfaction with life (grouped); people using Welsh in everyday life; tenure; ethnicity; religion; sexual orientation; and Valleys Taskforce area.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: Valleys Taskforce area.

Variables removed as not being significant predictors when part of a regression model:

Sexual orientation; tenure; in material deprivation; ethnicity; has a limiting long-term illness, disability or infirmity; overall satisfaction with life (grouped); WIMD overall score (in quintiles); religion; and economic status.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Gender	Male (reference group)						
	Female	1.112182	0.0580589	2.04	0.042	1.004017	1.232001
Age (3 groups)	16-44 (reference group)						
	45-64	1.466486	0.0910246	6.17	0	1.298506	1.656197
	65+	1.535492	0.1055339	6.24	0	1.341976	1.756913
Local Authority	Isle of Anglesey (reference group)						
	Gwynedd	1.200455	0.2664873	0.82	0.41	0.7769429	1.854825
	Conwy	1.783852	0.3708056	2.78	0.005	1.18692	2.680997
	Denbighshire	1.972567	0.4196654	3.19	0.001	1.299984	2.993131
	Flintshire	2.588532	0.5316127	4.63	0	1.730774	3.871388
	Wrexham	1.979344	0.4235419	3.19	0.001	1.301309	3.010664
	Powys	1.506042	0.2722264	2.27	0.023	1.056762	2.146332
	Ceredigion	1.57738	0.3428801	2.1	0.036	1.030168	2.415265
	Pembrokeshire	1.125156	0.2379188	0.56	0.577	0.7434003	1.702953
	Carmarthenshire	1.204547	0.2371047	0.95	0.344	0.8189773	1.77164
	Swansea	1.270382	0.2526462	1.2	0.229	0.8603049	1.875929
	Neath Port Talbot	3.221118	0.6603966	5.71	0	2.155224	4.814162
	Bridgend	2.170618	0.4591795	3.66	0	1.433896	3.285862
	Vale of Glamorgan	2.205041	0.4414956	3.95	0	1.48932	3.264713
	Cardiff	1.655716	0.3051772	2.74	0.006	1.153704	2.376168
	Rhondda Cynon Taf	1.90133	0.3647622	3.35	0.001	1.305444	2.769217
	Merthyr Tydfil	1.618854	0.391019	1.99	0.046	1.008342	2.599007
	Caerphilly	2.512549	0.5024108	4.61	0	1.697883	3.718102
	Blaenau Gwent	1.19837	0.2724038	0.8	0.426	0.767543	1.871023
	Torfaen	1.550198	0.3378844	2.01	0.044	1.011247	2.376385
Monmouthshire	2.364565	0.4818812	4.22	0	1.585924	3.525493	
Newport	1.752925	0.3585803	2.74	0.006	1.173926	2.617496	
Internet use	Yes (reference group)						
	No	0.5097065	-0.0460302	7.46	0	0.4270222	0.608401

Highest educational qualification	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	0.7609417	-0.0677516	3.07	0.002	0.639092	0.9060234
	National Qualification Framework level 2	0.6951244	-0.0510564	4.95	0	0.6019249	0.8027544
	Below National Qualification Framework level 2	0.6143592	-0.082512	3.63	0	0.4721729	0.7993623
	No qualification	0.5574504	-0.0401101	8.12	0	0.4841276	0.6418782
Mental well-being score	Low well-being (14-44) (reference group)						
	Medium well-being (45-57)	1.333665	0.0925562	4.15	0	1.164055	1.527988
	High well-being (58-70)	2.123402	0.1663277	9.61	0	1.821197	2.475754
People using Welsh in everyday life	Yes (reference group)						
	No	0.7834569	-0.0704751	2.71	0.007	0.6568196	0.9345104
	Constant	0.3460095	-0.0656053	5.6	0	0.2386138	0.5017421

Statistics

Number of obs = 9,480
Wald chi²(32) = 415.15
Prob > chi² = 0.0000

Log pseudolikelihood = -6217.22
Pseudo R² = 0.0464

Main factors influencing people's ability to influence decisions

Initial list of variables considered:

Age; gender; whether household contains a child under 16; highest educational qualification; tenure; in material deprivation; WIMD overall score (in quintiles); overall satisfaction with life (grouped); ethnicity; has a limiting long-term illness, disability or infirmity; satisfaction with local area; economic status; and feeling safe at home and when walking and travelling in local area.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: Valleys Taskforce area.

Variables removed as not being significant predictors when part of a regression model:

In material deprivation; highest educational qualification; tenure; economic status; WIMD overall score (in quintiles); whether household contains a child under 16; gender; ethnicity; has a limiting long-term illness, disability or infirmity; and local authority.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Age (3 groups)	16-44 (reference group)						
	45-64	0.8569316	0.0743926	-1.78	0.075	0.7228547	1.015878
	65+	0.6564388	0.0611905	-4.52	0	0.5468256	0.7880245
Overall satisfaction with life (grouped)	Low (1-4) (reference group)						
	Medium (5-6)	0.8925384	0.2171773	-0.47	0.64	0.5539977	1.437957
	High (7-8)	0.9634197	0.2173875	-0.17	0.869	0.6190833	1.499277
	Very high (9-10)	1.250003	0.2868963	0.97	0.331	0.7971618	1.960089

Feeling safe at home and when walking and travelling in local area	Feel safe (reference group)						
	Don't feel safe	0.7649562	0.0704339	-2.91	0.004	0.6386481	0.9162449
Satisfaction with local area as a place to live	Very satisfied (reference group)						
	Fairly satisfied	0.698649	0.0591136	-4.24	0	0.5918855	0.8246703
	Neither satisfied nor dissatisfied	0.4182849	0.0780974	-4.67	0	0.2900985	0.6031133
	Fairly dissatisfied	0.2619999	0.0592242	-5.93	0	0.1682248	0.408049
	Very dissatisfied	0.4067837	0.1152303	-3.18	0.001	0.233476	0.7087366
	Constant	0.3843175	0.0911052	-4.03	0	0.2414942	0.6116086

Statistics

Number of obs	=	6,688	Log pseudolikelihood	=	-3575.96
Wald chi ² (10)	=	118.68	Pseudo R ²	=	0.0276
Prob > chi ²	=	0.0000			

Main factors influencing satisfaction with access to services and facilities

Initial list of variables considered: age; local authority; access to a car; has public transport links in the local area; urban/rural morphology; has a limiting long-term illness, disability or infirmity; general health; in material deprivation; WIMD overall score (in quintiles); overall satisfaction with life (grouped); internet use; highest educational qualification; economic status; tenure; gender; ethnicity; religion; sexual orientation; and Valleys Taskforce area.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: Valleys Taskforce area.

Variables removed as not being significant predictors when part of a regression model:

Access to a car; has a limiting long-term illness, disability or infirmity; ethnicity; overall satisfaction with life (grouped); internet use; tenure; in material deprivation; economic status; religion; highest educational qualification; WIMD overall score (in quintiles); sexual orientation; age; and gender.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Local Authority	Isle of Anglesey (reference group)						
	Gwynedd	0.9017448	0.174881	-0.53	0.594	0.6166018	1.31875
	Conwy	1.756698	0.3741297	2.65	0.008	1.157214	2.666737
	Denbighshire	2.394334	0.5293498	3.95	0	1.552378	3.692939
	Flintshire	1.623896	0.3198743	2.46	0.014	1.103796	2.389062
	Wrexham	1.675267	0.3454323	2.5	0.012	1.118331	2.509562
	Powys	1.145417	0.1742813	0.89	0.372	0.8500581	1.543399
	Ceredigion	1.421706	0.2735931	1.83	0.067	0.975	2.073075
	Pembrokeshire	1.883576	0.3644295	3.27	0.001	1.289125	2.752146
	Carmarthenshire	1.422368	0.2602436	1.93	0.054	0.9937376	2.035879
	Swansea	1.989476	0.3651182	3.75	0	1.388424	2.850725
	Neath Port Talbot	1.233915	0.2328561	1.11	0.265	0.8524169	1.786151
Bridgend	1.724849	0.3504653	2.68	0.007	1.15824	2.568641	

	Vale of Glamorgan	1.710953	0.3363834	2.73	0.006	1.163823	2.515297
	Cardiff	3.755807	0.7470719	6.65	0	2.543254	5.546471
	Rhondda Cynon Taf	1.6322	0.3042073	2.63	0.009	1.132733	2.351902
	Merthyr Tydfil	1.290784	0.2696022	1.22	0.222	0.8571671	1.943755
	Caerphilly	1.435954	0.2740289	1.9	0.058	0.9878758	2.087271
	Blaenau Gwent	0.6773014	0.1308465	-2.02	0.044	0.4638099	0.989063
	Torfaen	1.213231	0.2583359	0.91	0.364	0.7992733	1.841585
	Monmouthshire	1.70909	0.3157246	2.9	0.004	1.189927	2.454763
	Newport	1.784386	0.3629442	2.85	0.004	1.197717	2.658418
Has public transport links in the local area	Not selected (reference group)						
	Public transport links (eg. train station or bus route)	1.37004	0.1074797	4.01	0	1.17478	1.597755
Urban/rural morphology	Urban >10k (reference group)						
	Town & fringe	0.5167231	0.0474955	-7.18	0	0.4315371	0.6187249
	Village	0.7654958	0.0648108	-3.16	0.002	0.6484488	0.9036702
	Hamlet	0.506623	0.0553722	-6.22	0	0.4089323	0.6276512
General health	Good or very good (reference group)						
	Fair	1.168325	0.0811397	2.24	0.025	1.019643	1.338687
	Bad or very bad	1.02492	0.0910643	0.28	0.782	0.8611134	1.219886
	Constant	2.198293	0.375453	4.61	0	1.572919	3.072309

Statistics

Number of obs = 11,839
Wald chi²(27) = 402.00
Prob > chi² = 0.0000

Log pseudolikelihood = -5715.76
Pseudo R² = 0.0449

Main factors influencing understanding of local councillors' role

Initial list of variables considered: age; gender; religion; ethnicity; local authority; mental well-being score; overall satisfaction with life (grouped); internet use; highest educational qualification; economic status; tenure; in material deprivation; WIMD overall score (in quintiles); whether household contains a child under 16; has a limiting long-term illness, disability or infirmity; and Valleys Taskforce area.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: Valleys Taskforce area.

Variables removed as not being significant predictors when part of a regression model:

In material deprivation; economic status; gender; has a limiting long-term illness, disability or infirmity; internet use; overall satisfaction with life (grouped); and ethnicity.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Age (3 groups)	16-44 (reference group)						
	45-64	1.507947	0.1136418	5.45	0	1.300882	1.747971
	65+	2.154723	0.1860912	8.89	0	1.81919	2.552142
Religion	No religion (reference group)						
	Christian (all denominations)	1.145874	0.0673715	2.32	0.021	1.021153	1.285829
	Another religion	1.481859	0.2617629	2.23	0.026	1.048206	2.09492
Local Authority	Isle of Anglesey (reference group)						
	Gwynedd	0.9620037	0.2185905	-0.17	0.865	0.6162588	1.501725
	Conwy	0.7678505	0.1618083	-1.25	0.21	0.5080469	1.160512
	Denbighshire	1.362393	0.2943511	1.43	0.152	0.892065	2.080694
	Flintshire	1.296608	0.2679478	1.26	0.209	0.86478	1.944071
	Wrexham	1.207919	0.2571686	0.89	0.375	0.7958206	1.833414
	Powys	0.9301579	0.1687634	-0.4	0.69	0.6518066	1.327378
	Ceredigion	0.9172879	0.2025858	-0.39	0.696	0.5949976	1.414152
	Pembrokeshire	0.8666345	0.1824305	-0.68	0.497	0.5736595	1.309235
	Carmarthenshire	0.7397153	0.1484679	-1.5	0.133	0.4991377	1.096248
	Swansea	0.6239832	0.1215752	-2.42	0.015	0.4259191	0.9141528
	Neath Port Talbot	0.970556	0.2000781	-0.14	0.885	0.6479586	1.453764
	Bridgend	0.8904663	0.1934749	-0.53	0.593	0.5816662	1.363205
	Vale of Glamorgan	0.5104847	0.1076791	-3.19	0.001	0.3376249	0.7718466
	Cardiff	0.9217818	0.171857	-0.44	0.662	0.6396319	1.328392
	Rhondda Cynon Taf	0.8537006	0.164313	-0.82	0.411	0.5854282	1.244909
	Merthyr Tydfil	0.9945374	0.2464848	-0.02	0.982	0.6118718	1.616522
	Caerphilly	0.7317409	0.150223	-1.52	0.128	0.4893387	1.094221
	Blaenau Gwent	0.6440497	0.1482471	-1.91	0.056	0.4101948	1.011227
	Torfaen	0.566357	0.1245774	-2.58	0.01	0.3680084	0.8716112
	Monmouthshire	0.775381	0.1558809	-1.27	0.206	0.5228672	1.149844
Newport	0.6424996	0.1338518	-2.12	0.034	0.4271125	0.9665036	
Mental well-being score	Low well-being (14-44) (reference group)						
	Medium well-being (45-57)	1.22873	0.0882932	2.87	0.004	1.067312	1.41456
	High well-being (58-70)	1.686622	0.1370753	6.43	0	1.438265	1.977866
Highest educational qualification	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	0.8733727	0.0815619	-1.45	0.147	0.727291	1.048796
	National Qualification Framework level 2	0.8586859	0.0663101	-1.97	0.049	0.738078	0.9990022
	Below National Qualification Framework level 2	0.6947625	0.1039366	-2.43	0.015	0.5181988	0.931486
	No qualification	0.8286302	0.0618593	-2.52	0.012	0.7158408	0.9591908
Tenure	Owner-occupier (reference group)						
	Social housing	1.315341	0.1172308	3.08	0.002	1.104522	1.566399
	Private rented	0.7771267	0.0727425	-2.69	0.007	0.646868	0.9336154

WIMD overall score (in quintiles)	Most deprived 20% (reference group)						
	Q2	1.204007	0.1150916	1.94	0.052	0.9983027	1.452097
	Q3	1.176867	0.1129937	1.7	0.09	0.9749932	1.420539
	Q4	1.119108	0.1113801	1.13	0.258	0.9207797	1.360156
	Least deprived 20%	1.371834	0.1339748	3.24	0.001	1.13285	1.661234
Whether household contains a child under 16	Yes (reference group)						
	No						
		0.8304731	0.0622411	-2.48	0.013	0.7170194	0.9618785
	Constant	0.3666457	0.0737051	-4.99	0	0.2472484	0.5437005

Statistics

Number of obs = 8,811
Wald chi²(38) = 338.61
Prob > chi² = 0.0000

Log pseudolikelihood = -5602.17
Pseudo R² = 0.0408

Childcare

Main factors influencing use of formal childcare

Initial list of variables considered: gender of parent; gender of child; type of household; number of children in household; whether parents speak Welsh; economic status of the parents; age of parent; age of child; whether parent has a long term limiting illness or infirmity; in material deprivation; urban/rural classification; tenure; highest qualification of the parent; and working status of household.

Variables removed due to missing values: None

Variables removed as not being significant predictors on their own: None

Variables removed due to multicollinearity: None

Variables removed as not being significant predictors when part of a regression model:

Gender of parent; tenure; number of children in household; urban/rural classification; gender of child; whether parent has a long term limiting illness or infirmity; economic status of the parents; age of parent; type of household; and in material deprivation.

Final model

Variable	Categories	Odds ratio	Standard error	z	P>z	95% Confidence interval	
						Lower	Upper
Welsh speaking ability of parent	Yes (reference group)						
	No	0.6878468	0.1064239	-2.42	0.016	0.5079179	0.9315151
	No, but have some Welsh speaking ability (Spontaneous only)	1.207957	0.2310187	0.99	0.323	0.83035	1.757282
Age of child	0-4 (reference group)						
	5-10	0.573452	0.0727015	-4.39	0	0.4472835	0.7352097
	11-14	0.0791002	0.017509	-11.46	0	0.0512581	0.1220654
Highest educational qualification of parent	National Qualification Framework levels 4-8 (reference group)						
	National Qualification Framework level 3	0.7918403	0.1513928	-1.22	0.222	0.5443717	1.151807
	National Qualification Framework level 2	0.5708059	0.0960808	-3.33	0.001	0.4104027	0.7939018
	Below National Qualification Framework level 2	0.876087	0.2418975	-0.48	0.632	0.5099422	1.505128
	No qualification	0.3418855	0.0782659	-4.69	0	0.218283	0.5354778
Working status of household	All working (reference group)						
	Some working	0.2541639	0.0469253	-7.42	0	0.1769945	0.364979
	None working	0.1396033	0.0355413	-7.73	0	0.0847599	0.2299328
	No one aged 16-19 (not in FT education) nor aged 19-64 in household	1	(empty)				
Constant		1.481646	0.2376505	2.45	0.014	1.081969	2.028963

Statistics

Number of obs = 2,258
 Wald chi²(10) = 270.74
 Prob > chi² = 0.0000

Log pseudolikelihood = -1190.38
 Pseudo R² = 0.1972

For further information please contact:

Scott Armstrong
Social Research and Information Division
Knowledge and Analytical Services
Welsh Government, Cathays Park
Cardiff, CF10 3NQ
Email: surveys@gov.wales
Telephone: 03000 250056

Mae'r ddogfen yma hefyd ar gael yn Gymraeg.
This document is also available in Welsh.

OGI © Crown Copyright Digital ISBN: 978-1-78964-916-1