Source: https://nacchocommunique.com/tag/social-determinants-of-health/

Summary and Assignment Support

Zi Ye

ENVS225

Exploring the Social World

UNIVERSITY OF LIVERPOOL

Quantitative Block

Week 7: Fundamental Concept + Data Visualisation

Week 8: Multiple Linear Regression

Week 9: Prediction and Model Assessment

Week 10: Dummy Variables

Week 11: Logistic Regression

Week 12: Wrap up

Regression Model

Multiple Linear Regression

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \epsilon$$

Scale/Continuous variables

Scale/Continuous variables

Dummy variables

Y: What is average % of people with long-term illness in the district?

X1: % of male
X2: % of no qualification
X3: % of higher professional

X4: Region

Logistic Regression

$$logit(p) = log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Log odds

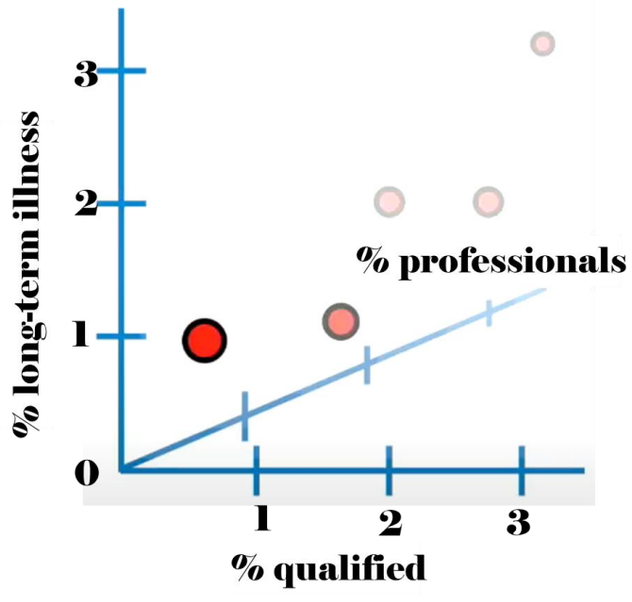
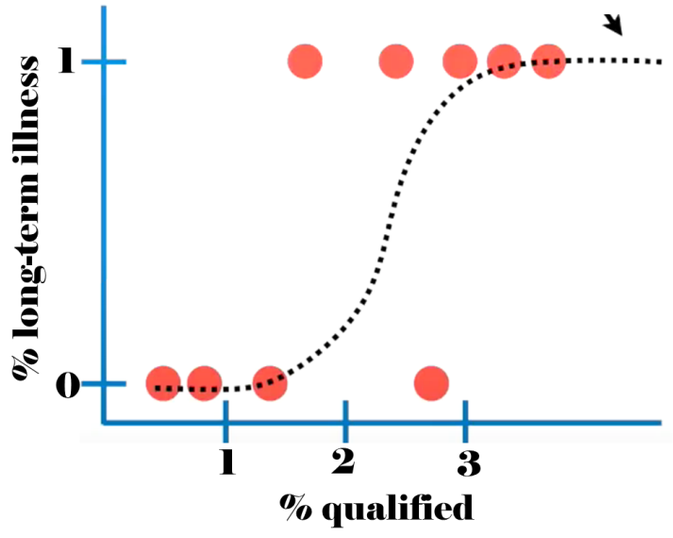
Categorical Variables
*binary

Scale/Continuous variables

Dummy variables

p: Whether the person is willing to commute long distance?

X1: Sex
X2: NSSEC (higher managers, higher professional, routine occupation

	Multiple Linear Regression	Logistic Regression
Output variables (dependent/response)	Continuous/Scales (e.g. Rate, Age, Distance, Height…)	Categorical (e.g. Yes/No, Male/Female, Win/Not win)
Output to predicted…	Y: Mean of the target variable at the given values of the input variable	Log Odds: The probability of the particular levels of the given values of the input variable
Solve problems	Regression	Classification
Practical	What is the average long-term illness rate (%) in Liverpool?	Do you willing to commute long distance?
		

Interpretation

Slope coefficient 1

Slope coefficient 2

Slope coefficient 3

Intercept

Error

Dependent variable

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \epsilon$$

$logit(p)$

Independent variable 1

Independent variable 2

Independent variable 3

Overall Model Fit (R^2)

R-Square / Adjusted R-square: the proportion of variance in the dependent variable (science) which can be predicted from the independent variables

See Week 9 Practical

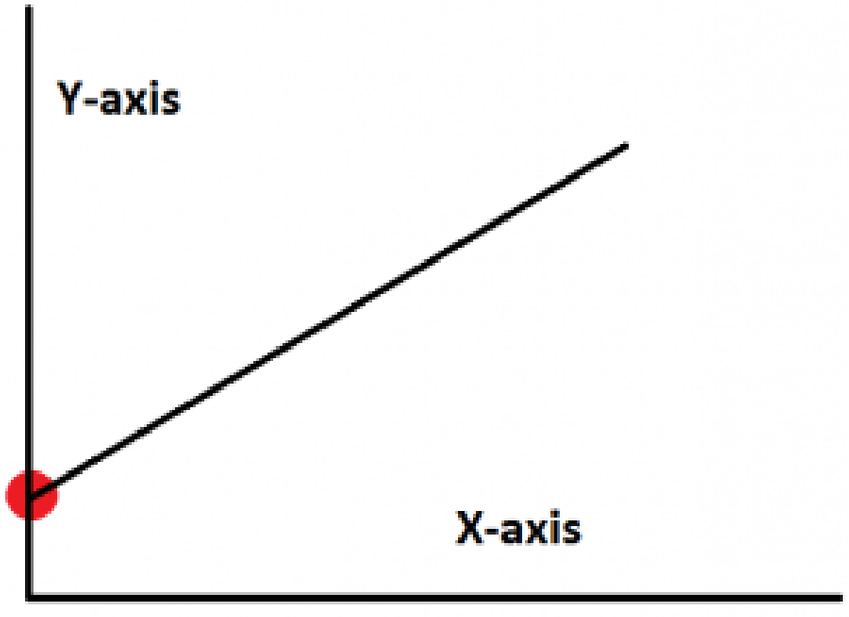
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.836[a]	.699	.697	1.82845

a. Predictors: (Constant), % of Total residents aged 16 to 74 who are in Higher managerial and professional occupations, % of Total resident population who are Males, % of Total residents aged 16 and over with No qualifications

Intercept (Constant)

- The predicted value of Y/Log-odds when all other variables are 0.



See Week 8,9,10,11 Practical

P-value (Sig.)

- help to determine whether the relationships that you observe in your sample also exist in the larger population.

- If the p-value of a coefficient is smaller than 0.05, the coefficient is statistically significant. *You can say that the relationship between this independent variable and the outcome variable is statistically significant.*

- If the p-value of a coefficient is larger than 0.05, the coefficient is not statistically significant. *You can say or conclude that there is no evidence of an association or relationship between this independent variable and the outcome variable.*

See Week 8,9,10,11 Practical

Slope βs

- The estimated change in the Y/Log-odds for one unit change in X_i, holding all other variables constant.

See Week 8,9,10,11 Practical & Week 11 Appendix

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	47.498	6.376		7.450	.000
	% of Total resident population who are Males	-.817	.125	-.181	-6.533	.000
	% of Total residents aged 16 and over with No qualifications	.477	.036	.768	13.404	.000
	% of Total residents aged 16 to 74 who are in Higher managerial and professional occupations	-.018	.046	-.022	-.383	.702

a. Dependent Variable: % of Total resident population who are Ill-health limits activities a lot or a little

F-Statistic/ANOVA

ANOVA: Analysis Of Variance. It decomposes the total variance (or variability/variation) of the outcome to two parts: the variation that can be explained by the included independent variables and the variation that cannot be explained.

See Week 8,9 Practical

ANOVA[a]

Model			Sum of Squares	df	Mean Square	F	Sig.
1	Regression	B	3121.773	3	1040.591	311.252	.000[b]
	Residual	C	1343.983	402	3.343		
	Total	A	4465.755	405			

F-statistic

P value

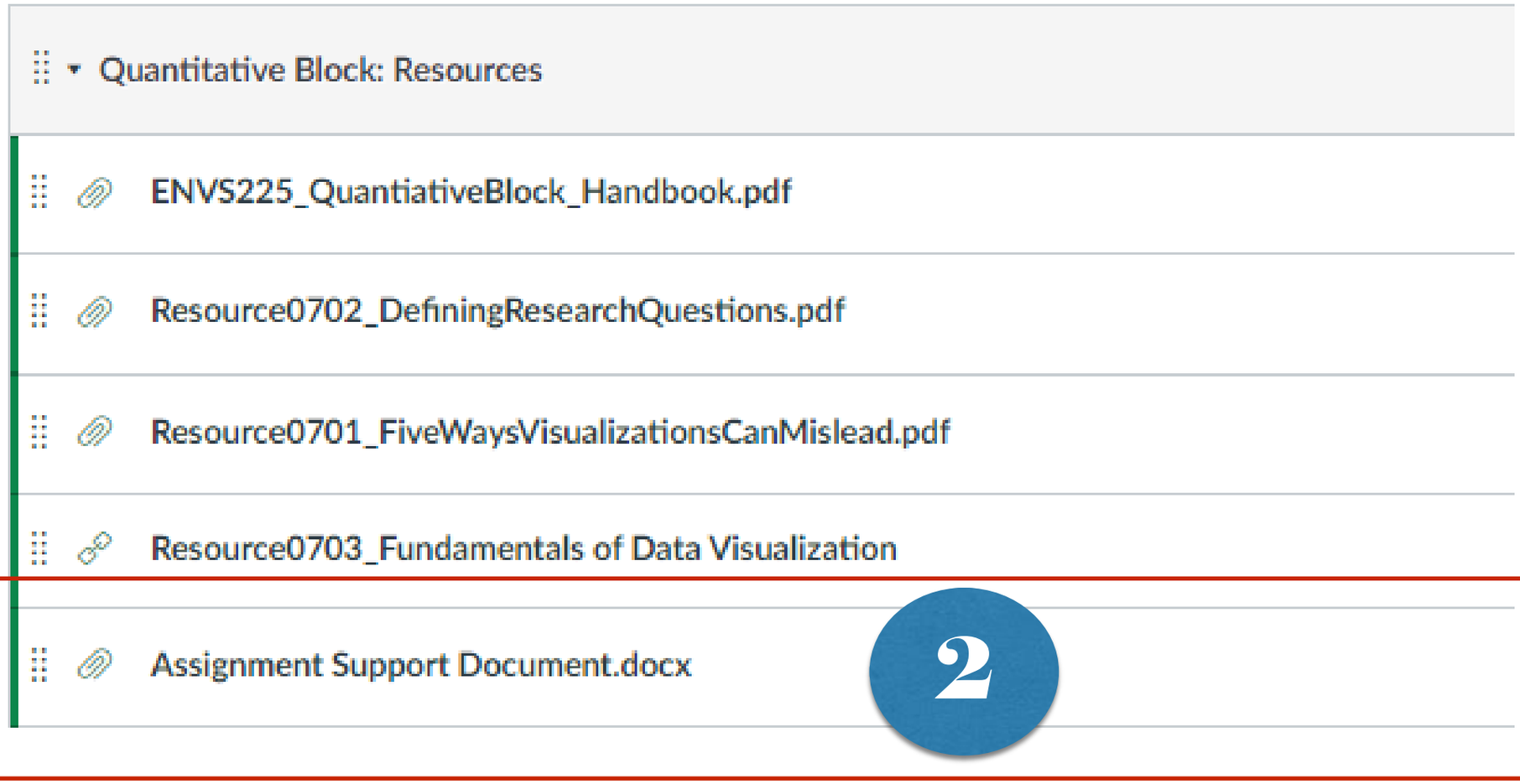
a. Dependent Variable: % of Total resident population who are Ill-health limits activities a lot or a little

b. Predictors: (Constant), % of Total residents aged 16 to 74 who are in Higher managerial and professional occupations, % of Total resident population who are Males, % of Total residents aged 16 and over with No qualifications

Assignment

Submission: before 2pm Tuesday 9th January

1 Handbook



Structure

Or

- 1. Introduction
- 2. Literature Review
- 3. Methodology
- 4. Results and Discussion
- 5. Conclusion
- References

- 1. Introduction
- 2. Literature Review
- 3. Methodology
- 4. Results
- 5. Discussion
- 6. Conclusion
- References

Research Question

- Make sense
- Knowledge gap
- Location: national, regional, local …

Examples from practical

- **How do local factors affect residents' health in England and Wales?**
- **What is the average long-term illness rate in Liverpool?**
- **How does health vary across regions in the UK?**
- **Who is willing to commute long distances?**

Methodology: dataset

- 1. 2021 UK Census Data
- 2. 2021 Annual Population Survey
- 3. Family Resource Survey 2016-17
- 4.2011 Census Sample of Anonymised Records (SAR.sav)

Make sure you talk about the dataset and why you choose those input variables

Methodology: descriptive statistic

For continuous variables

Name of variable	Description of variable	Minimum value	Maximum value	Mean	Standard deviation

For categorical data

Name of variable	Description of variable	Number of unique values	*Frequency of each unique value

Wisely use Practical 07 Data Visualisation to help you describe the dataset you used.

Methodology

For sure

Regression model

Multiple Linear Regression

Or

Logistic Regression

Please make sure you use one regression model

Results and Discussion

- *Overall Model Fit (R^2)
- *P-value (Sig.)
- *Slope βs
- Intercept (Constant)
- Anova

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	47.498 (intercept)	6.376		7.450	.000
	% of Total resident population who are Males	-.817	.125	-.181	-6.533	.000
	% of Total residents aged 16 and over with No qualifications	.477 (Slope βs)	.036	.768	13.404	.000
	% of Total residents aged 16 to 74 who are in Higher managerial and professional occupations	-.018	.046	-.022	-.383	.702

a. Dependent Variable: % of Total resident population who are Ill-health limits activities a lot or a little

P-value

Talk according to your table.

Others

- More than 5 references, uniform style (Chicago, APA, Harvard), ENVS203 literature management

- Use Figure 1, Figure 2, Table 1 … with your Graph/Table. Mention them in the text.

- Interpret your results and discuss with relevance to your literature review – use citations!

- Earn points for illustrations (graphs/maps/charts) in discussion part!

Have a nice Christmas break!

Zi Ye zi.ye@liverpool.ac.uk
Gabriele Filomena gabriele.filomena@liverpool.ac.uk

ENVS225
Exploring the Social World

UNIVERSITY OF LIVERPOOL