Strokes and the 'Friday afternoon problem':
An audit using logistic regression.

Mario Hair: University of the West of Scotland

UNIVERSITY OF THE WEST *of* SCOTLAND UWS

Stroke

↓

Scan (within 48hrs)

Ischemic ↓ **antiplatelets good**

Hemorrhagic ↓ **antiplatelets bad**

1

Problems with the data

- Duplicate cases.
- Numbers entered as text.
- Dates not in a format that could be read by SPSS.
- The key value - hours between admission and medication was embedded in a sentence.
  e.g."Time between first admission and administration: 65 hours"

By Friday afternoon SPSS file of 712 cases with no duplicates created (original file had 1356 cases).

2

First quick analysis

- **Outcome by days between admittance and medication**

		Days to medication				Total
		1-2	3-4	5-6	7+	
Outcome	Alive	258	229	69	52	608
		88.7%	90.2%	79.3%	65.0%	85.4%
	Dead	33	25	18	28	104
		11.3%	9.8%	20.7%	35.0%	14.6%
Total		291	254	87	80	712

		Days to med		Total
		1-4	5+	
Outcome	Alive	487	121	608
		89.4%	72.5%	85.4%
	Dead	58	46	104
		10.6%	**27.5%**	**14.6%**
Total		545	167	712

Odds ratio = 3.19
95% CI (2.07 , 4.93)

Big difference after 4 days.
Those who get meds after 4 days are 3.19 times more likely to die.

Research question: Is 'hours to med' a significant predictor of survival (status) once covariates are taken into account?

Variables

Status	0 = Alive 1 = Dead. (Survival at 30 days after stroke) 104 deaths out of 712 cases (14.6%).
Hours_to_med	Hours between admit and the antiplatelet drug

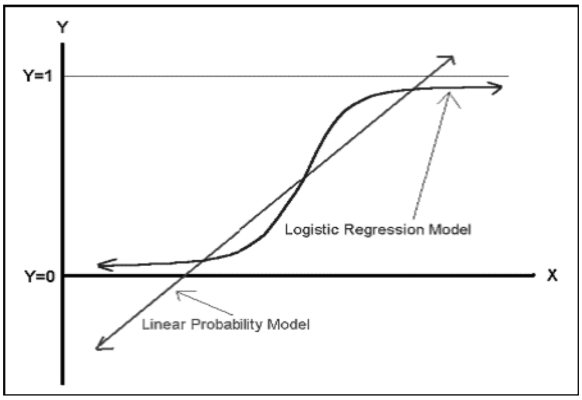
Covariates

Barthel_score	Ranges from 0 to 100 with lower score indicating greater severity
Type_stroke	1 = LACS, 2 = PACS/POCS 3 = TACS. TACS is the most serious, LACS least.
Age	
Sex	Males = 0, Females = 1

When dependent variable is dichotomous can use logistic regression

Logistic regression

Linear probability model : $p(\text{outcome}) = \sum_{i=1}^{n} b_i x_i$



Logistic regression model : $\text{Logit}(p) = \ln\left(\frac{p}{1-p}\right) = \sum_{i=1}^{n} b_i x_i$

5

Maximum likelihood estimation

- Logistic regression uses maximum likelihood to fit the parameters
- It seeks parameters so as to maximise the log likelihood function (LL)
- $LL = \sum [\, Y_i \ln(\hat{Y}_i) + (1 - Y_i) \ln(1 - \hat{Y}_i)]$

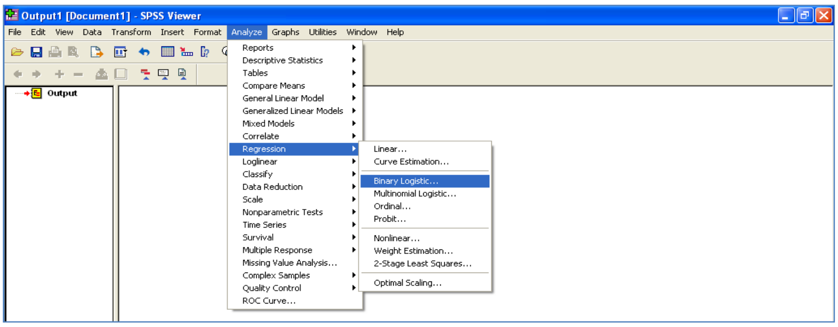
Where Y_i is the value of the dependent variable for case i

And $\hat{Y}_i$ is the predicted probability of 'success' for case I

- In fact SPSS seeks to minimise -2 log likelihood (-2LL)
- Changes to -2LL have a chi-square distribution
- It is simple to calculate -2LL only when there are no covariates in the model. After that, there is no closed form solution so iterative methods are used. Stops when successive iterations converge.

6

Logistic regression in SPSS

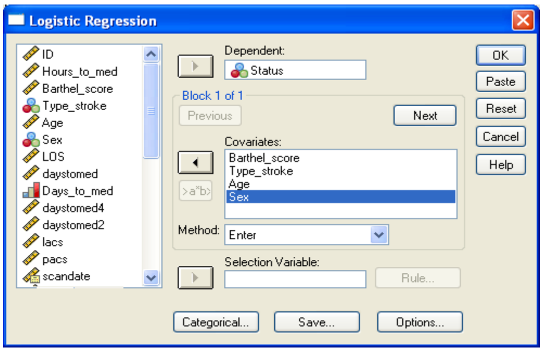


Hierarchical logistic regression: covariates entered in blocks

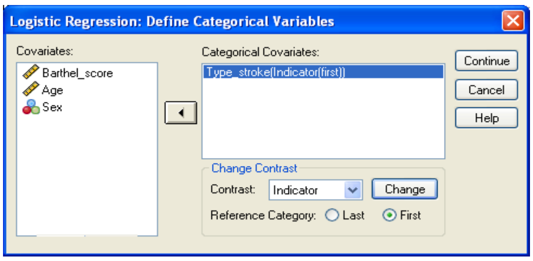
- **Dependent variable : status**
- **First block : all the covariates**
- **Second block : hours to med**

Logistic regression in SPSS

In the first block all the covariates are entered

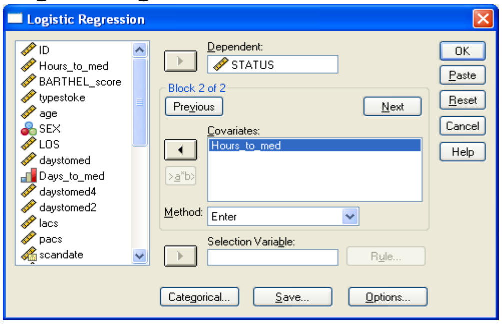


Dummy variables can be created easily



Logistic regression in SPSS

- In the second block only hours to med is added



Unweighted Cases(a)		N	Percent
Selected Cases	Included in Analysis	673	94.5
	Missing Cases	39	5.5
	Total	712	100.0
Unselected Cases		0	.0
Total		712	100.0

			Parameter coding	
		Frequency	(1)	(2)
type of stroke	LACS	214	.000	.000
	PACS/POCS	359	1.000	.000
	TACS	100	.000	1.000

The output (model fit block 1)

- **Block 1: Method = Enter**

  Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	224.14	5	.000
	Block	224.14	5	.000
	Model	224.14	5	.000

Fall in -2LL

Sum is the initial -2LL
224.14 + 345.00 = 569.14

- **Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	345.00(a)	.283	.496

Final -2ll

a Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

- **McFadden's R^2 is the proportional reduction in -2LL fall in -2LL value divided by the initial -2LL value.**

  $$R_L^2 = \frac{224.14}{569.14} = 0.39$$

- **so the covariates account for a 39% fall in the -2LL value.**

The output (model fit block 2)

- **Block 2: Method = Enter**

  Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	9.63	1	.002
	Block	9.63	1	.002
	Model	233.77	6	.000

- **Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	335.377(a)	.293	.514

- McFadden's R^2 $\quad R_L^2 = \dfrac{233.77}{569.14} = 0.41$

- Hours to med has contributed a significant 2% fall in -2LL over and above that caused by the covariates

The output (interpreting coefficients)

- **Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
Barthel_score	**-.068**	.010	44.995	1	**.000**	**.935**
Type_stroke			18.945	2	.000	
Type_stroke(1)	**.410**	.362	1.279	1	.258	1.507
Type_stroke(2)	**1.541**	.387	15.851	1	**.000**	4.670
Age	**.060**	.015	16.337	1	**.000**	1.062
Sex	**-.097**	.294	.108	1	.743	.908
Hours_to_med	**.005**	.002	8.241	1	**.004**	**1.005**
Constant	**-6.085**	1.243	23.955	1	.000	.002

- **The significant variables are Barthel score, TACS stroke type , age and hours to med.**

- **The B column gives the coefficients of the regression equation**

$$\text{Logit(p)} = \ln\left(\frac{p}{1-p}\right) = -6.085 - 0.068\text{Barthel\_score} + \ldots\ldots + 0.005\text{Hours\_to\_med}$$

- **Exponential of beta (EXP(B)): Change in the odds of dying for a unit increase.**

  An increase of one hour before getting the medication increases the odds of death by a factor of 1.005. An increase of 24 hours increases the odds by 1.005^{24} = 1.127and so on.

Conclusion

- **The time taken to administer the antiplatelets is a significant predictor even when other factors are taken into account.**
- **Overall every extra day increases odds of death by 12.7%**
- **But survival rates seem to reduce dramatically after 4 days.**

13

Except……

It's all a load of b*ll*cks !

The cardinal rule for a consultant statistician:

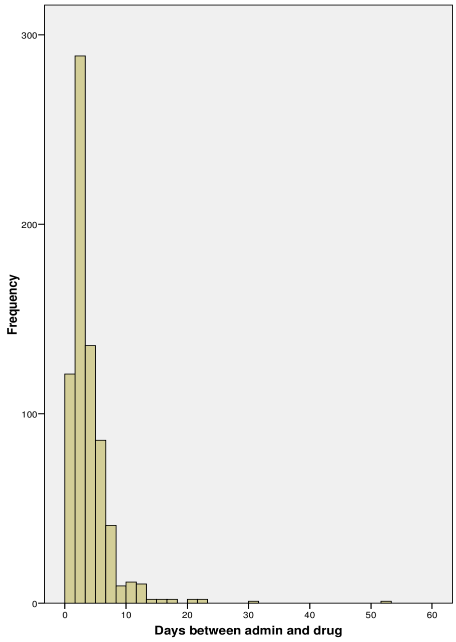
NEVER EVER EVER believe the client

ALWAYS check the data first

14

Checking the data: Distribution of days to medication

Days to med	Frequency	Cumulative %
1	121	16.9
2	167	40.3
3	122	57.4
4	136	76.4
5	52	83.7
6	34	88.3
7	25	91.9
8	16	94.1
9	9	95.4
10	6	96.2
11	5	96.9
12	7	97.9
13	3	98.3
14	2	98.6
15	**1**	**98.7**
16	**1**	**98.9**
17	**2**	**99.2**
20	**1**	**99.3**
21	**1**	**99.4**
22	**1**	**99.6**
23	**1**	**99.7**
30	**1**	**99.9**
53	**1**	**100.0**



The effect of extreme cases:

If the 10 patients with over 14 days to med are excluded, time to med not significant.

- **Block 2: Omnibus Tests of Model Coefficients**

		Chi-square	df	Sig.
Step 1	**Step**	2.852	1	.091
	Block	2.852	1	.091
	Model	212.239	6	.000

- **Variables in the Equation**

	B	S.E.	Wald	df	Sig.	Exp(B)
Barthel_score	-.071	.011	44.030	1	.000	.931
Type_stroke			14.334	2	.001	
Type_stroke(1)	.343	.363	.892	1	.345	1.409
Type_stroke(2)	1.357	.394	11.873	1	.001	3.884
Age	.057	.015	14.772	1	.000	1.059
Sex	-.050	.296	.029	1	.865	.951
hours_to_med	**.004**	**.002**	**2.834**	**1**	**.092**	**1.004**
Constant	-5.661	1.247	20.619	1	.000	.003

Conclusions

- Always Always check the data first (no matter how rushed you are).
- Never believe a client about data (no matter how nice s(he) seems).
- Logistic regression is a great technique but maybe not so robust?

  (diagnostics are not as good as for linear regression)

Any questions!

17