

Logistic Regression Analysis for cross-sectional data

Data: infarc.raw
 1 subject id
 2 oral contraceptive indicator (0=no,1=yes)
 3 age (years)
 4 smoking indicator (0=no,1=yes)
 5 binary response -- whether MI has been suffered
 (0=no,1=yes)

Exploratory Analysis

d

Contains data from C:\My Documents\655LDA\DataLDA\infarc.dta
 obs: 200
 vars: 5 2 Mar 2004 10:21
 size: 4,800 (99.3% of memory free)

variable name	storage type	display format	value label	variable label
id	float	%9.0g		ID
oc	float	%9.0g		Oral contraceptive use(0:NO, 1:Yes)
age	float	%9.0g		Age years
smoke	float	%9.0g		Smoking indicator
mi	float	%9.0g		Miocardial Infarction event

Relation between smoking status and MI.

. tab smoke mi

Smoking indicator	Miocardial Infarction event		Total
	0	1	
0	117	17	134
1	40	26	66
Total	157	43	200

. cci 26 17 40 117

	Exposed	Unexposed	Total	Proportion Exposed
Cases	26	17	43	0.6047
Controls	40	117	157	0.2548
Total	66	134	200	0.3300
Point estimate			[95% Conf. Interval]	
Odds ratio	4.473529		2.077789	9.701643 (exact)
Attr. frac. ex.	.7764629		.5187191	.8969247 (exact)
Attr. frac. pop	.4694892			
chi2(1) = 18.69 Pr>chi2 = 0.0000				

Reject the null hypothesis and conclude that the OR $\neq$ 1.

Relation between OC status and MI.

. tab oc mi

Oral contraceptive use(0:NO, 1:Yes)	Myocardial Infarction event		Total
	0	1	
0	93	9	102
1	64	34	98
Total	157	43	200

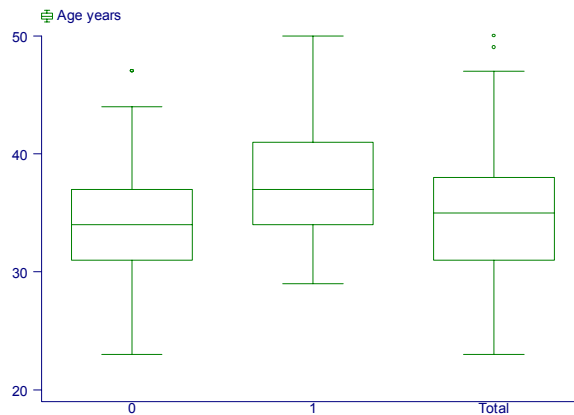
```
. cci 34 9 64 93
```

	Exposed	Unexposed	Total	Proportion Exposed
Cases	34	9	43	0.7907
Controls	64	93	157	0.4076
Total	98	102	200	0.4900
	Point estimate		[95% Conf. Interval]	
Odds ratio	5.489583		2.356981	13.81983 (exact)
Attr. frac. ex.	.8178368		.5757285	.9276402 (exact)
Attr. frac. pop	.6466617			
chi2(1) = 19.82 Pr>chi2 = 0.0000				

Reject the null hypothesis and conclude that the $OR \neq 1$.

Relation between age and MI.

```
. sort mi
. gr age, box by(mi) total s(o) ylab
```



```
. by mi: summ age
```

```
-> mi = 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	157	34.12102	4.682688	23	47

```
-> mi = 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	43	37.90698	5.079435	29	50

Two sample t-test

```
. ttesti 43 37.9 5.08 157 34.1 4.68
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	43	37.9	.7746927	5.08	36.33661	39.46339
y	157	34.1	.3735047	4.68	33.36222	34.83778
combined	200	34.917	.3540173	5.006561	34.21889	35.61511
diff		3.8	.8206078		2.181747	5.418253

Degrees of freedom: 198

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0	Ha: diff ~= 0	Ha: diff > 0
t = 4.6307	t = 4.6307	t = 4.6307
P < t = 1.0000	P > t = 0.0000	P > t = 0.0000

Conclude that the means ages are different for people with vs. people w/out MI.

Logistic regression analysis for independent data

```
. logit mi age smoke oc, nolog
```

Logit estimates

Number of obs	=	200
LR chi2(3)	=	57.83
Prob > chi2	=	0.0000
Pseudo R2	=	0.2777

Log likelihood = -75.187408

mi	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.1625556	.0445365	3.65	0.000	.0752657	.2498455
smoke	1.812237	.4293898	4.22	0.000	.9706485	2.653826
oc	1.979896	.4696893	4.22	0.000	1.059322	2.90047
_cons	-9.114047	1.757076	-5.19	0.000	-12.55785	-5.670242

Same result can be obtained with:

```
. xi:logit mi age i.smoke i.oc, nolog
```

or with:

```
. glm mi age smoke oc, f(bin) l(logit) nolog
```

The estimated increase in log odds of MI for nonsmokers and non-users of OC is .16 with one year increase in age.

SAS PROCEDURE

```
data mi; infile 'infarc.dat';
  input id oral age smoke mi;
run;
/*****
  Fit the logistic regression model using PROC GENMOD.
  We do not use a CLASS statement here, as the covariates are
  either continuous (AGE) or already in "dummy" form (ORAL, SMOKE).
  The model statement with the LINK=LOGIT option results in the
  logistic regression model in equation (10.21). The DIST=BINOMIAL
  specifies the Bernoulli distribution, which is the simplest case
  of a binomial distribution.
*****/
proc genmod data=mi;
  model mi = oral age smoke / dist = binomial link = logit;
run;
```

SAS OUTPUT

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.MI
Distribution	BINOMIAL
Link Function	LOGIT
Dependent Variable	MI
Observations Used	200
Number Of Events	43
Number Of Trials	200

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	196	150.3748	0.7672
Scaled Deviance	196	150.3748	0.7672
Pearson Chi-Square	196	177.5430	0.9058
Scaled Pearson X2	196	177.5430	0.9058
Log Likelihood	.	-75.1874	.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	-9.1140	1.7571	26.9044	0.0001
ORAL	1	1.9799	0.4697	17.7683	0.0001
AGE	1	0.1626	0.0445	13.3216	0.0003
SMOKE	1	1.8122	0.4294	17.8121	0.0001
SCALE	0	1.0000	0.0000	.	.

NOTE: The scale parameter was held fixed.