Low Birth Weight

Jack Tubbs

September 2021

Contents

Discussion of the Problem

The data contain information about infant mortality in 2003 and were obtained from the US National Center for Health Statistics. A random sample of 2,500+ observations is used in this example. This data are observational, in which case, meaningful inference is limited. The description below is for a causal inference example, which is beyond the scope of this course, given in SAS.

Our approach is to investigate this problem using the material given in the first part of Chapter 3 in the methods lecture notes.

The main variables in the analysis are as follows:

- The treatment variable is **Smoking**. It is an indicator of maternal smoking behavior, with values Yes and No.

- The outcome variable is **Death**. It is an indicator of infant death within one year of birth, with values Yes and No.

- The mediator variable is **LowBirthWgt**. It is an indicator of low birth weight (less than 2,500 grams), with values Yes and No.

The analysis also includes five confounding covariates:

- **AgeGroup** represents maternal ages of less than 20, between 20 and 35, and greater than 35, with values 1, 2, and 3, respectively.

- **Drinking** is an indicator of maternal drinking during pregnancy, with values Yes and No.

- **Married** is an indicator of marital status, with values Yes and No.

- **Race** is an indicator of race, with values Asian, Black, Hispanic, Native (native American), and White.

- **SomeCollege** is an indicator of whether the mother has 12 or more years of education, with values Yes and No.

R

Needed Packages

```r
if(!require(FSA)){install.packages("FSA")}
if(!require(ggplot2)){install.packages("ggplot2")}
if (!require("mosaic")) install.packages("mosaic", dep=FALSE)
if (!require("nortest")) install.packages("nortest", dep=TRUE)
if (!require("epitools")) install.packages("epitools", dep=TRUE)
if (!require("prettyR")) install.packages("prettyR", dep=TRUE)
if (!require("rms")) install.packages("rms", dep=TRUE)
# add other as needed
```

Problem – Infant Birth Weight Data

Read data from SAS input file

```r
# this data came from SASHELP.BWEIGHT
bw = read.csv('bwgt.csv', header = TRUE)
bw = data.frame(bw)
#summary(bw)
bw = transform(bw, AgeGroup.f = as.factor(AgeGroup))
bw = transform(bw, Race.f = as.factor(Race))
bw = transform(bw, Drinking.f = as.factor(Drinking))
bw = transform(bw, Death.f = as.factor(Death))
bw = transform(bw, Smoking.f = as.factor(Smoking))
bw = transform(bw, SomeCollege.f = as.factor(SomeCollege))
bw = transform(bw, LowBirthWgt.f = as.factor(LowBirthWgt))
```

```r
tally(~ AgeGroup + Race.f, data=bw)
```

```
##          Race.f
## AgeGroup Asian Black Hispanic Native White
##        1     8    91       83      6   169
##        2   101   375      475     22  1337
##        3    36    52       66      4   264
```

```r
tally(~ Race.f | AgeGroup.f, data=bw)
```

```
##          AgeGroup.f
## Race.f        1    2    3
##    Asian      8  101   36
##    Black     91  375   52
##    Hispanic  83  475   66
##    Native     6   22    4
##    White    169 1337  264
```

```r
library(mosaic)
mytab = tally(~ Race.f | AgeGroup.f, data=bw)
addmargins(mytab)
```

```
##          AgeGroup.f
## Race.f        1    2    3  Sum
```

2

```
##    Asian       8  101   36  145
##    Black      91  375   52  518
##    Hispanic   83  475   66  624
##    Native      6   22    4   32
##    White     169 1337  264 1770
##    Sum       357 2310  422 3089
```

```r
prop.table(mytab, 1)
```

```
##          AgeGroup.f
## Race.f             1          2          3
##    Asian    0.05517241 0.69655172 0.24827586
##    Black    0.17567568 0.72393822 0.10038610
##    Hispanic 0.13301282 0.76121795 0.10576923
##    Native   0.18750000 0.68750000 0.12500000
##    White    0.09548023 0.75536723 0.14915254
```

```r
library(epitools)
attach(bw)
mytab = tally(~ LowBirthWgt.f | Death.f, data=bw)
addmargins(mytab)
```

```
##             Death.f
## LowBirthWgt.f   No  Yes  Sum
##           No  2278  198 2476
##           Yes  205  408  613
##           Sum 2483  606 3089
```

```r
prop.table(mytab, 1)
```

```
##             Death.f
## LowBirthWgt.f        No        Yes
##           No  0.92003231 0.07996769
##           Yes 0.33442088 0.66557912
```

```r
riskratio(x=Smoking.f, y=Death.f)
```

```
## $data
##          Outcome
## Predictor   No Yes Total
##            156  41   197
##     No    1786 405  2191
##     Yes    541 160   701
##     Total 2483 606  3089
##
## $measure
##          risk ratio with 95% C.I.
## Predictor  estimate     lower     upper
##           1.0000000        NA        NA
##      No   0.8881678 0.6670955 1.182502
##      Yes  1.0966911 0.8087967 1.487063
##
## $p.value
##          two-sided
## Predictor midp.exact fisher.exact chi.square
##                  NA           NA         NA
##      No    0.4206715    0.4449380  0.4220298
```

```
##      Yes  0.5556210    0.6286442  0.5493629
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

SAS

Code 1

```
/*
The Sashelp.BirthWgt data set contains 100,000 random observations about
infant mortality in 2003 from the US National Center for Health Statistics.
Each observation records infant death  within one year of birth, birth weight,
maternal smoking and drinking behavior, and other background characteristics
of the mother.
*/

title "Sashelp.bweight --- Infant Birth Weight";
data birthwgt; set sashelp.birthwgt;
run;

proc contents data=birthwgt varnum;
ods select position;
run;

title "The First Five Observations Out of 100,000";
proc print data=birthwgt(obs=10);
run;
```

Sashelp.bweight — Infant Birth Weight

The CONTENTS Procedure

Variables in Creation Order			
#	Variable	Type	Len
1	LowBirthWgt	Char	3
2	Married	Char	3
3	AgeGroup	Num	8
4	Race	Char	9
5	Drinking	Char	3
6	Death	Char	3
7	Smoking	Char	3
8	SomeCollege	Char	3

The First Five Observations Out of 100,000

Obs	LowBirthWgt	Married	AgeGroup	Race	Drinking	Death	Smoking	SomeCollege
1	No	No	3	Asian	No	No	No	Yes
2	No	No	2	White	No	No	No	No
3	Yes	Yes	2	Native	No	Yes	No	No
4	No	No	2	White	No	No	No	No
5	No	No	2	White	No	No	No	Yes
6	No	No	2	White	No	No	No	
7	No	No	2	Asian	No	No	No	Yes
8	No	No	3	White	No	No	No	Yes
9	No	Yes	1	Black	No	No	No	No
10	No	No	2	Native	No	No	No	Yes

Code 2

I have changed 'Yes' responses to 'Affirm' as SAS orders the variables in the tables using an alphabetical ordering. This new order allows one to have a better interpretation of results.

```
*Create a new smaller data set;
title 'New Sample of Size 2,500';
proc surveyselect data=birthwgt out=new2 method=srs n=2500
                  seed=2021;
run;

/* I needed more death records than the srs gave me */
data new; set birthwgt; if death = 'Yes';
run;

/*merge the two files into one */
data new_bwgt; set new new2;
run;

data new_bwgt; set new_bwgt;
if LowBirthWgt = 'Yes' then LowBirthWgt = 'Affirm';
if Death = 'Yes' then Death = 'Affirm';
if Smoking = 'Yes' then Smoking = 'Affirm';
if Drinking = 'Yes' then Drinking = 'Affirm';


title 'Test for Association between Low Birth Weight and Smoking';
proc freq data=new_bwgt;* order=freq;
tables smoking*LowBirthWgt/norow nopercent chisq relrisk riskdiff;
run;

title 'Test for Association between Low Birth Weight and drinking';
proc freq data=new_bwgt;* order=freq;
tables drinking*LowBirthWgt/norow nopercent chisq relrisk riskdiff;
run;
```

New Sample of Size 2,500

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	BIRTHWGT
Random Number Seed	2021
Sample Size	2500
Selection Probability	0.025
Sampling Weight	40
Output Data Set	NEW2

Test for Association between Low Birth Weight and Smoking

The FREQ Procedure

Table of Smoking by LowBirthWgt			
Smoking	LowBirthWgt		
	Aff	No	Total
Aff	155 26.96	546 23.56	701
No	420 73.04	1771 76.44	2191
Total	575	2317	2892
Frequency Missing = 197			

Note	Statistics for Table of Smoking by LowBirthWgt

In the following table there is not a significant association at the .05 level between Low Birth Weight and Smoking. This is seen in the chi-square statistic and the relative risk and odds ratio.

Statistic	DF	Value	Prob
Chi-Square	1	2.8856	0.0894
Likelihood Ratio Chi-Square	1	2.8341	0.0923
Continuity Adj. Chi-Square	1	2.7038	0.1001
Mantel-Haenszel Chi-Square	1	2.8846	0.0894
Phi Coefficient		0.0316	
Contingency Coefficient		0.0316	
Cramer's V		0.0316	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	155
Left-sided Pr <= F	0.9593
Right-sided Pr >= F	0.0510
Table Probability (P)	0.0102
Two-sided Pr <= P	0.0921

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.2211	0.0157	0.1904	0.2518	0.1909	0.2537
Row 2	0.1917	0.0084	0.1752	0.2082	0.1754	0.2088
Total	0.1988	0.0074	0.1843	0.2134	0.1844	0.2138
Difference	0.0294	0.0178	−0.0054	0.0643		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.7789	0.0157	0.7482	0.8096	0.7463	0.8091
Row 2	0.8083	0.0084	0.7918	0.8248	0.7912	0.8246
Total	0.8012	0.0074	0.7866	0.8157	0.7862	0.8156
Difference	−0.0294	0.0178	−0.0643	0.0054		
Difference is (Row 1 - Row 2)						

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.1970	0.9725	1.4734
Relative Risk (Column 1)	1.1535	0.9796	1.3582
Relative Risk (Column 2)	0.9636	0.9218	1.0074

Note	Sample Size = 2892 Frequency Missing = 197

Test for Association between Low Birth Weight and drinking

In the following table there is not a significant association at the .05 level between Low Birth Weight and Drinking. This is seen in the chi-square statistic and the relative risk and odds ratio.

The FREQ Procedure

Table of Drinking by LowBirthWgt			
Drinking	**LowBirthWgt**		
	Aff	*No*	*Total*
Aff	74 12.87	325 14.03	399
No	501 87.13	1992 85.97	2493
Total	575	2317	2892
Frequency Missing = 197			

Note	Statistics for Table of Drinking by LowBirthWgt

Statistic	DF	Value	Prob
Chi-Square	1	0.5187	0.4714
Likelihood Ratio Chi-Square	1	0.5263	0.4682
Continuity Adj. Chi-Square	1	0.4260	0.5140
Mantel-Haenszel Chi-Square	1	0.5185	0.4715
Phi Coefficient		−0.0134	
Contingency Coefficient		0.0134	
Cramer's V		−0.0134	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	74
Left-sided Pr <= F	0.2588
Right-sided Pr >= F	0.7835
Table Probability (P)	0.0423
Two-sided Pr <= P	0.4998

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.1855	0.0195	0.1473	0.2236	0.1485	0.2271
Row 2	0.2010	0.0080	0.1852	0.2167	0.1854	0.2172
Total	0.1988	0.0074	0.1843	0.2134	0.1844	0.2138
Difference	−0.0155	0.0210	−0.0568	0.0258		
Difference is (Row 1 - Row 2)						

8

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.8145	0.0195	0.7764	0.8527	0.7729	0.8515
Row 2	0.7990	0.0080	0.7833	0.8148	0.7828	0.8146
Total	0.8012	0.0074	0.7866	0.8157	0.7862	0.8156
Difference	0.0155	0.0210	−0.0258	0.0568		
Difference is (Row 1 - Row 2)						

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.9053	0.6906	1.1869
Relative Risk (Column 1)	0.9229	0.7406	1.1500
Relative Risk (Column 2)	1.0194	0.9689	1.0725

Note	Sample Size = 2892 Frequency Missing = 197

Code 3

```
title 'Test for Association between Low Birth Weight and Smoking';
title2 'Controlling for Death';
proc freq data=new_bwgt;* order=freq;
tables death*smoking*LowBirthWgt /nopercent norow chisq cmh;
run;

title 'Test for Association between Low Birth Weight and Drinking';
title2 'Controlling for Death';
proc freq data=new_bwgt;* order=freq;
tables death*drinking*LowBirthWgt /nopercent norow chisq cmh;
run;
```

Test for Association between Low Birth Weight and Smoking

Controlling for Death

The FREQ Procedure

Table 1 of Smoking by LowBirthWgt			
Controlling for Death=Aff			
Smoking	LowBirthWgt		
	Aff	No	Total
Aff	102 26.63	58 31.87	160
No	281 73.37	124 68.13	405
Total	383	182	565
Frequency Missing = 41			

Note	Statistics for Table 1 of Smoking by LowBirthWgt Controlling for Death=Aff

Statistic	DF	Value	Prob
Chi-Square	1	1.6664	0.1967
Likelihood Ratio Chi-Square	1	1.6467	0.1994
Continuity Adj. Chi-Square	1	1.4185	0.2337
Mantel-Haenszel Chi-Square	1	1.6635	0.1971
Phi Coefficient		−0.0543	
Contingency Coefficient		0.0542	
Cramer's V		−0.0543	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	102
Left-sided Pr <= F	0.1172
Right-sided Pr >= F	0.9172

10

Fisher's Exact Test	
Table Probability (P)	0.0344
Two-sided Pr <= P	0.2304

Note	Sample Size = 565 Frequency Missing = 41

Table 2 of Smoking by LowBirthWgt			
Controlling for Death=No			
Smoking	LowBirthWgt		
	Aff	No	Total
Aff	53 27.60	488 22.86	541
No	139 72.40	1647 77.14	1786
Total	192	2135	2327
Frequency Missing = 156			

Note	Statistics for Table 2 of Smoking by LowBirthWgt Controlling for Death=No

Statistic	DF	Value	Prob
Chi-Square	1	2.2246	0.1358
Likelihood Ratio Chi-Square	1	2.1450	0.1430
Continuity Adj. Chi-Square	1	1.9666	0.1608
Mantel-Haenszel Chi-Square	1	2.2237	0.1359
Phi Coefficient		0.0309	
Contingency Coefficient		0.0309	
Cramer's V		0.0309	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	53
Left-sided Pr <= F	0.9410
Right-sided Pr >= F	0.0821
Table Probability (P)	0.0231
Two-sided Pr <= P	0.1532

Note	Sample Size = 2327 Frequency Missing = 156

Note	Summary Statistics for Smoking by LowBirthWgt Controlling for Death

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0640	0.8003
2	Row Mean Scores Differ	1	0.0640	0.8003
3	General Association	1	0.0640	0.8003

Common Odds Ratio and Relative Risks				
Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel−Haenszel	1.0328	0.8004	1.3327
	Logit	1.0374	0.8066	1.3341
Relative Risk (Column 1)	Mantel−Haenszel	1.0170	0.8909	1.1610
	Logit	0.9675	0.8563	1.0932
Relative Risk (Column 2)	Mantel−Haenszel	0.9954	0.9599	1.0323
	Logit	0.9809	0.9513	1.0114

Breslow-Day Test for Homogeneity of Odds Ratios	
Chi-Square	3.8091
DF	1
Pr > ChiSq	0.0510

Note	Sample Size = 2892 Frequency Missing = 197

Test for Association between Low Birth Weight and Drinking

Controlling for Death

The FREQ Procedure

Table 1 of Drinking by LowBirthWgt			
Controlling for Death=Aff			
Drinking	LowBirthWgt		
	Aff	No	Total
Aff	45 11.75	25 13.74	70
No	338 88.25	157 86.26	495
Total	383	182	565
Frequency Missing = 41			

Note	Statistics for Table 1 of Drinking by LowBirthWgt Controlling for Death=Aff

Statistic	DF	Value	Prob
Chi-Square	1	0.4487	0.5029
Likelihood Ratio Chi-Square	1	0.4419	0.5062
Continuity Adj. Chi-Square	1	0.2843	0.5939
Mantel-Haenszel Chi-Square	1	0.4479	0.5033
Phi Coefficient		−0.0282	
Contingency Coefficient		0.0282	
Cramer's V		−0.0282	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	45
Left-sided Pr <= F	0.2940
Right-sided Pr >= F	0.7912
Table Probability (P)	0.0852
Two-sided Pr <= P	0.4976

Note	Sample Size = 565 Frequency Missing = 41

13

Table 2 of Drinking by LowBirthWgt			
Controlling for Death=No			
Drinking	*LowBirthWgt*		
	Aff	*No*	*Total*
Aff	29 15.10	300 14.05	329
No	163 84.90	1835 85.95	1998
Total	192	2135	2327
Frequency Missing = 156			

Note	Statistics for Table 2 of Drinking by LowBirthWgt Controlling for Death=No

Statistic	DF	Value	Prob
Chi-Square	1	0.1608	0.6884
Likelihood Ratio Chi-Square	1	0.1581	0.6909
Continuity Adj. Chi-Square	1	0.0858	0.7696
Mantel-Haenszel Chi-Square	1	0.1607	0.6885
Phi Coefficient		0.0083	
Contingency Coefficient		0.0083	
Cramer's V		0.0083	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	29
Left-sided Pr <= F	0.7002
Right-sided Pr >= F	0.3773
Table Probability (P)	0.0775
Two-sided Pr <= P	0.6660

Note	Sample Size = 2327 Frequency Missing = 156

Note	Summary Statistics for Drinking by LowBirthWgt Controlling for Death

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.0102	0.9194
2	Row Mean Scores Differ	1	0.0102	0.9194
3	General Association	1	0.0102	0.9194

Common Odds Ratio and Relative Risks				
Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel−Haenszel	0.9834	0.7080	1.3659
	Logit	0.9836	0.7110	1.3609
Relative Risk (Column 1)	Mantel−Haenszel	0.9908	0.8261	1.1883
	Logit	0.9668	0.8191	1.1413
Relative Risk (Column 2)	Mantel−Haenszel	1.0021	0.9611	1.0449
	Logit	0.9942	0.9592	1.0305

Breslow-Day Test for Homogeneity of Odds Ratios	
Chi-Square	0.5994
DF	1
Pr > ChiSq	0.4388

Note	Sample Size = 2892 Frequency Missing = 197

Code 4

```
title 'Test for Association between Low Birth Weight and Death';
title2 '';
proc freq data=new_bwgt;* order=freq;
tables LowBirthWgt*death/norow nopercent chisq relrisk riskdiff;
run;

ods latex close;
```

Test for Association between Low Birth Weight and Death

In the following table there is a significant association at the .05 level between Death and Low Birth Weight. This is seen in the chi-square statistic and the relative risk and odds ratio.

The FREQ Procedure

Table of LowBirthWgt by Death			
LowBirthWgt	Death		
	Aff	No	Total
Aff	408 67.33	205 8.26	613
No	198 32.67	2278 91.74	2476
Total	606	2483	3089

Note	Statistics for Table of LowBirthWgt by Death

Statistic	DF	Value	Prob
Chi-Square	1	1068.5596	<.0001
Likelihood Ratio Chi-Square	1	897.1241	<.0001
Continuity Adj. Chi-Square	1	1064.8493	<.0001
Mantel-Haenszel Chi-Square	1	1068.2137	<.0001
Phi Coefficient		0.5882	
Contingency Coefficient		0.5070	
Cramer's V		0.5882	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	408
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

16

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.6656	0.0191	0.6282	0.7029	0.6267	0.7029
Row 2	0.0800	0.0055	0.0693	0.0907	0.0696	0.0914
Total	0.1962	0.0071	0.1822	0.2102	0.1823	0.2106
Difference	0.5856	0.0198	0.5468	0.6245		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.3344	0.0191	0.2971	0.3718	0.2971	0.3733
Row 2	0.9200	0.0055	0.9093	0.9307	0.9086	0.9304
Total	0.8038	0.0071	0.7898	0.8178	0.7894	0.8177
Difference	−0.5856	0.0198	−0.6245	−0.5468		
Difference is (Row 1 - Row 2)						

Note all the confidence intervals do not contain one, indicating a strong association between infant birth weight and survival.

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	22.8979	18.3410	28.5869
Relative Risk (Column 1)	8.3231	7.2003	9.6210
Relative Risk (Column 2)	0.3635	0.3249	0.4067

Note	Sample Size = 3089