



1

# Part 4

## Lecture 2 Logistic Regression



# Who we are...

**Pascal Tyrrell, PhD**      *Associate Professor*  
Department of Medical Imaging , Faculty of Medicine  
Department of Statistical Sciences , Faculty of Arts and Science

**Paul Corey, PhD**      *Professor Emeritus*  
Biostatistics Program, Dalla Lana Faculty of Public Health  
Institute of Medical Science, Faculty of Medicine  
Department of Statistical Sciences, Faculty of Arts and Science



The POPULATION

DISEASE

EXPOSURE	YES	NO	TOTAL
YES	A	B	A + B
NO	C	D	C + D

$$\text{RELATIVE RISK} = \frac{A / (A + B)}{C / (C + D)}$$

$$\text{ODDS RATIO} = \frac{A / B}{C / D}$$

# THE POPULATION

# DISEASE

EXPOSURE	YES	NO	TOTAL
YES	A	B	A + B
NO	C	D	C + D

$$\text{RELATIVE RISK (RR)} = \frac{A / (A + B)}{C / (C + D)} = \frac{A \times (C + D)}{C \times (A + B)} = \frac{A \times C + A \times D}{A \times C + B \times C}$$

$$\text{ODDS RATIO (OR)} = \frac{A / B}{C / D} = \frac{A \times D}{B \times C}$$

NOTE: For a rare disease **A** and **A x C** are very small and  
 RR = OR (approximately)

# THE POPULATION

# DISEASE

EXPOSURE	YES	NO	TOTAL	PERCENT	RR
YES	10	990	1000	1.0	<b>5.0</b>
NO	20	9980	10000	0.2	

RELATIVE RISK ( RR ) =  $\frac{10 / (10 + 990)}{20 / (20 + 9980)} = \frac{10 \times (20 + 9980)}{20 \times (10 + 990)} = \frac{100,000}{20,000} = \mathbf{5.00}$

ODDS RATIO ( OR ) =  $\frac{10 / 990}{20 / 9980} = \frac{10 \times 9980}{990 \times 20} = \frac{99800}{19800} = \mathbf{5.04}$

# THE POPULATION

EXPOSURE	DISEASE		TOTAL	PROPORTION
	YES	NO		
YES	100	900	1,000	0.10
NO	100	9,900	10,000	0.01

$$\text{RELATIVE RISK} = \frac{100 / (1,000)}{100 / (10,000)} = 10.0$$

$$\text{ODDS RATIO} = \frac{100 / 900}{100 / 9900} = 11.0$$

```

DATA T ;
INPUT  GROUP $  SUCCESS $  N  @@  ;
DATALINES ;
DRUG      NO  3  DRUG      YES  7
PLACEBO   NO  8  PLACEBO  YES  2

RUN ;

PROC PRINT DATA = T; RUN ;

PROC FREQ DATA = T ;

TABLES GROUP * SUCCESS/ NOPERCENT NOCOL
NOROW

CHISQ FISHER EXPECTED ;

WEIGHT N ;

RUN ;

```

**The SAS System**

Obs	GROUP	SUCCESS	N
1	DRUG	NO	3
2	DRUG	YES	7
3	PLACEBO	NO	8
4	PLACEBO	YES	2

**The FREQ Procedure**

Frequency Expected	Table of GROUP by SUCCESS		
	GROUP	SUCCESS	
	NO	YES	Total
DRUG	3 5.5	7 4.5	10
PLACEBO	8 5.5	2 4.5	10
Total	11	9	20



```
DATA PERSONS ; INPUT GROUP $ SUCCESS $ @@;
```

```
DATALINES ;
```

```
DRUG NO DRUG NO DRUG NO DRUG YES
```

```
DRUG YES DRUG YES DRUG YES DRUG YES
```

```
DRUG YES DRUG YES
```

```
PLACEBO NO PLACEBO NO PLACEBO YES PLACEBO YES
```

```
PLACEBO YES PLACEBO YES PLACEBO YES PLACEBO YES
```

```
PLACEBO YES PLACEBO YES
```

```
RUN ;
```

```
PROC FREQ DATA = T ;
```

```
TABLES GROUP * SUCCESS/ NOPERCENT NOCOL NOROW
```

```
CHISQ FISHER EXPECTED ;
```

```
RUN ;
```

<<< My Note: No WEIGHT statement needed



### The FREQ Procedure

Frequency Expected	Table of GROUP by SUCCESS			
	GROUP	SUCCESS		Total
		NO	YES	
	DRUG	3 5.5	7 4.5	10
	PLACEBO	8 5.5	2 4.5	10
	Total	11	9	20

### Statistics for Table of GROUP by SUCCESS

Statistic	DF	Value	Prob
Chi-Square	1	5.0505	0.0246
Likelihood Ratio Chi-Square	1	5.3002	0.0213
Continuity Adj. Chi-Square	1	3.2323	0.0722
Mantel-Haenszel Chi-Square	1	4.7980	0.0285
Phi Coefficient		-0.5025	
Contingency Coefficient		0.4490	
Cramer's V		-0.5025	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test	
Cell (1,1) Frequency (F)	3
Left-sided Pr <= F	0.0349
Right-sided Pr >= F	0.9973
Table Probability (P)	0.0322
Two-sided Pr <= P	0.0698

Sample Size = 20

```

DATA TTT ;
LIKELIHOOD0 = ((11/20)11) × ((9/20)9) ;
LIKELIHOOD1 = ((3/10)3) × ((7/10)7) ;
LIKELIHOOD2 = ((8/10)8) × ((2/10)2) ;
LR = LIKELIHOOD0 / (LIKELIHOOD1 × LIKELIHOOD2 ) ;
MINUS2LOGL = -2 × LOG(LR) ;
RUN ; PROC PRINT ; RUN ;

```

## The SAS System

Obs	LIKELIHOOD0	LIKELIHOOD1	LIKELIHOOD2	LR	MINUS2LOGL
1	.000001054	.002223566	.006710886	0.070644	5.30022

Statistics for Table of GROUP by SUCCESS

Statistic	DF	Value	Prob
Chi-Square	1	5.0505	0.0246
Likelihood Ratio Chi-Square	1	5.3002	0.0213
Continuity Adj. Chi-Square	1	3.2323	0.0722
Mantel-Haenszel Chi-Square	1	4.7980	0.0285
Phi Coefficient		-0.5025	
Contingency Coefficient		0.4490	
Cramer's V		-0.5025	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

### Fisher's Exact Test

Cell (1,1) Frequency (F)	3
Left-sided Pr <= F	0.0349
Right-sided Pr >= F	0.9973
Table Probability (P)	0.0322
Two-sided Pr <= P	0.0698

Sample Size = 20

```

TITLE1 "    LIKELIHOOD RATIO TEST    ";
DATA TT ;
P1  = (0.30)3 x (0.70)7 ;
P2  = (0.80)8 x (0.20)2 ;
PH0 = (0.55)11 x (0.45)9 ;
LR  = PH0 / (P1 x P2); MINUS2LOGLR = - 2 x LOG(LR) ;
Z   = SQRT(MINUS2LOGLR) ;
PVALUE1 = SDF('CHISQUARE' , MINUS2LOGLR, 1) ;
PVALUE2 = 2 x SDF('NORMAL' , Z, 0, 1); RUN ;
PROC PRINT ; RUN ;

```

### LIKELIHOOD RATIO TEST

Obs	P1	P2	PH0	LR	MINUS2LOGLR	Z	PVALUE1	PVALUE2
1	.002223566	.006710886	.000001054	0.070644	5.30022	2.30222	0.021323	0.021323

**SCORE TEST**    CHI SQUARE = 5.0505    P = 0.0246

# THE SCORE TEST – ASSUME NULL HYPOTHESIS $P_1 = P_0$

$$Z = \frac{p_1 - p_0}{\sqrt{\frac{P_m \times Q_m}{n_1} + \frac{P_m \times Q_m}{n}}} = \frac{\frac{8}{10} - \frac{3}{10}}{\sqrt{\frac{11}{20} \times \frac{9}{20} \times \left(\frac{1}{10} + \frac{1}{10}\right)}}$$

$$= \frac{0.5}{\sqrt{0.0495}} = \frac{0.5}{0.222486} = 2.24733$$

$$Z^2 = 5.0505$$

$$Z = 2.247$$

$$p = 0.0246$$

## CORRECTION FACTOR FOR CONTINUITY

$$Z = \frac{(p_1 - p_0) - \left(\frac{1}{2n_0} + \frac{1}{2n_1}\right)}{\sqrt{\frac{P_m \times Q_m}{n_1} + \frac{P_m \times Q_m}{n}}} = \frac{\frac{8}{10} - \frac{3}{10} - \left(\frac{1}{20} + \frac{1}{20}\right)}{\sqrt{\frac{11}{20} \times \frac{9}{20} \times \left(\frac{1}{10} + \frac{1}{10}\right)}}$$

$$= \frac{0.5 - 0.10}{\sqrt{0.0495}} = \frac{0.40}{0.22246} = 1.79807$$

$$Z^2 = 3.2331 \quad p = 0.0712$$

Fisher's Exact Test	
Cell (1,1) Frequency (F)	3
Left-sided Pr <= F	0.0349
Right-sided Pr >= F	0.9973
Table Probability (P)	0.0322
Two-sided Pr <= P	0.0698

P VALUE BASED ON NUMBER OF POSSIBLE COMBINATIONS AS OR MORE EXTREME THAN THE OBSERVED DIFFERENCE ( )

Number of Ways of Getting Each Table

$$3 \quad 7 \quad 10 \quad 10C3 = 120$$

$$() \quad 8 \quad 2 \quad 10 \quad 10C8 = 45 \quad \text{Sum} = 45 \times 120 = 5400$$

$$2 \quad 8 \quad 10 \quad 10C2 = 45$$

$$9 \quad 1 \quad 10 \quad 10C9 = 10 \quad \text{Sum} = 10 \times 45 = 450$$

$$11 \quad 9 \quad 10 \quad 10C1 = 10$$

$$0 \quad 0 \quad 10 \quad 10C10 = 1 \quad \text{Sum} = 1 \times 10 = \underline{10}$$

$$\text{TOTAL} = 5860$$

The FREQ Procedure

Frequency Expected	Table of GROUP by SUCCESS			
	GROUP	SUCCESS		Total
		NO	YES	
	DRUG	3 5.5	7 4.5	10
	PLACEBO	8 5.5	2 4.5	10
	Total	11	9	20

P VALUE IS THE PROPORTION OF POSSIBLE COMBINATIONS AS EXTREME OR MORE EXTREME THAN THE OBSERVED DIFFERENCE

Total Number of 2 by 2 tables =  ${}_{20}C_{11} = 167,960$

P Value =  $2 \times 5860 / 167,960 = 0.0698$

Fisher's Exact Test	
Cell (1,1) Frequency (F)	3
Left-sided Pr <= F	0.0349
Right-sided Pr >= F	0.9973
Table Probability (P)	0.0322
Two-sided Pr <= P	0.0698

### The FREQ Procedure

Frequency Expected	Table of GROUP by SUCCESS			
	GROUP	SUCCESS		Total
		NO	YES	
	DRUG	3 5.5	7 4.5	10
	PLACEBO	8 5.5	2 4.5	10
	Total	11	9	20

### Statistics for Table of GROUP by SUCCESS

Statistic	DF	Value	Prob
Chi-Square	1	5.0505	0.0246
Likelihood Ratio Chi-Square	1	5.3002	0.0213
Continuity Adj. Chi-Square	1	3.2323	0.0722
Mantel-Haenszel Chi-Square	1	4.7980	0.0285
Phi Coefficient		-0.5025	
Contingency Coefficient		0.4490	
Cramer's V		-0.5025	
<b>WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.</b>			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	3
Left-sided Pr <= F	0.0349
Right-sided Pr >= F	0.9973
Table Probability (P)	0.0322
Two-sided Pr <= P	0.0698

Sample Size = 20



		SUCCESS		
		NO	YES	
DRUG	A	B	$p1 = B / A$	
GROUP CONTROL	C	D	$p2 = D / C$	

$$p0 = (B + D) / (A + B + C + D) \quad q0 = 1 - p0$$

If we assume the difference between the observed **proportions**  $p1$  and  $p2$  is due to chance then we replace them by  $p0$  and  $q0$ . We can test the hypothesis that the **probabilities**  $P1$  and  $P2$  are equal, that is, that  $P1 = P2$ .

**Probability** tells you the likelihood of something happening, while **proportion** is just the comparison of (measurable) quantities. If you have a standard deck, the **proportion** of diamonds there is  $1/4$ , and so is the **probability** to draw one.

# Table of GROUP by SUCCESS

GROUP	SUCCESS		Total
	NO	YES	
DRUG	3	7	10
PLAC	8	2	10
Total	11	9	20

# Table of GROUP by SUCCESS

GROUP	SUCCESS		Total	Probability
	NO	YES		
DRUG	3	7	10	$\left(\frac{3}{10}\right)^3 \times \left(\frac{7}{10}\right)^7$
PLACEBO	8	2	10	$\left(\frac{8}{10}\right)^8 \times \left(\frac{2}{10}\right)^2$
Total	11	9	20	$\left(\frac{11}{20}\right)^{11} \times \left(\frac{9}{20}\right)^9$

## Likelihood Ratio

$$LR = \frac{{}_N C_X \times P_1^X \times Q_1^{N-X} \times {}_M C_Y \times P_2^Y \times Q_2^{M-Y}}{{}_N C_X \times P_0^X \times Q_0^{N-X} \times {}_M C_Y \times P_0^Y \times Q_0^{M-Y}}$$

$$LR = \frac{\left(\frac{3}{10}\right)^3 \times \left(\frac{7}{10}\right)^7 \times \left(\frac{8}{10}\right)^8 \times \left(\frac{2}{10}\right)^2}{\left(\frac{11}{20}\right)^{11} \times \left(\frac{9}{20}\right)^9}$$

*Under the Null Hypothesis  $H_0: P_1 = P_2$   
and the combinatorial constants cancel out.*

# *LIKELIHOOD RATIO TEST*

$$LR = \frac{\left(\frac{3}{10}\right)^3 \times \left(\frac{7}{10}\right)^7 \times \left(\frac{8}{10}\right)^8 \times \left(\frac{2}{10}\right)^2}{\left(\frac{11}{20}\right)^{11} \times \left(\frac{9}{20}\right)^9}$$

*- 2 × ln(LR) has a Chi Square*

*Probability Distribution with 1 Degree of Freedom*

```

DATA TTT ;
LIKELIHOOD0 = ((11/20) 11) ((9/20) 9) ;
LIKELIHOOD1 = (( 3/10) 3) ((7/10) 7) ;
LIKELIHOOD2 = (( 8/10) 8) ((2/10) 2) ;
LR = LIKELIHOOD0 / (LIKELIHOOD1 LIKELIHOOD2 ) ;
MINUS2LOGL = -2 LOG(LR) ;
RUN ; PROC PRINT ; RUN ;

```

## The SAS System

Obs	LIKELIHOOD0	LIKELIHOOD1	LIKELIHOOD2	LR	MINUS2LOGL
1	.000001054	.002223566	.006710886	0.070644	5.30022

## Statistics for Table of GROUP by SUCCESS

Statistic	DF	Value	Prob
Chi-Square	1	5.0505	0.0246
Likelihood Ratio Chi-Square	1	5.3002	0.0213
Continuity Adj. Chi-Square	1	3.2323	0.0722
Mantel-Haenszel Chi-Square	1	4.7980	0.0285
Phi Coefficient		-0.5025	
Contingency Coefficient		0.4490	
Cramer's V		-0.5025	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

## Fisher's Exact Test

Cell (1,1) Frequency (F)	3
Left-sided Pr <= F	0.0349
Right-sided Pr >= F	0.9973
Table Probability (P)	0.0322
Two-sided Pr <= P	0.0698

Sample Size = 20

# LOGISTIC MODEL

```

DATA T ;
INPUT DRUG $ SUCCESS $ N @@ ;
DATALINES ;
ASPIRIN NO 3 ASPIRIN YES 7
PLACEBO NO 8 PLACEBO YES 2
RUN ;
PROC LOGISTIC DATA = T ;
CLASS DRUG ; WEIGHT N ;
MODEL SUCCESS = DRUG ; RUN ;
    
```

## The LOGISTIC Procedure

Model Information	
Data Set	WORK.T
Response Variable	SUCCESS
Number of Response Levels	2
Weight Variable	N
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	4
Number of Observations Used	4
Sum of Weights Read	20
Sum of Weights Used	20

Response Profile			
Ordered Value	SUCCESS	Total Frequency	Total Weight
1	NO	2	11.000000
2	YES	2	9.000000

Probability modeled is SUCCESS='NO'.

Class Level Information		
Class	Value	Design Variables
DRUG	ASPIRIN	1
	PLACEBO	-1

### Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

### Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	29.526	26.225
SC	28.912	24.998
-2 Log L	27.526	22.225

### Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5.3002	1	0.0213
Score	5.0505	1	0.0246
Wald	4.5304	1	0.0333

### Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
DRUG	1	4.5304	0.0333

- << Derived from the ratio of likelihoods.
- << Derived from the standard Gaussian Z value.
- << Derived from the logistic regression analysis.



CURE	NO	YES	TOTAL	PROPORTION	ODDS
DRUG	3	7	10	0.7	$7/3 = 2.3333$
PLACEBO	8	2	10	0.2	$2/8 = 0.25$

RELATIVE RISK 1 =  $0.7 / 0.2 = 3.50$

RELATIVE RISK 2 =  $0.2 / 0.7 = 0.29$

ODDS RATIO 1 =  $2.3333 / 0.25 = 9.333$

ODDS RATIO 2 =  $0.25 / 2.33 = 0.107$

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.2695	0.5247	0.2638	0.6075
DRUG	ASPIRIN	1	-1.1168	0.5247	4.5304	0.0333

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
DRUG ASPIRIN vs PLACEBO	0.107	0.014	0.838

```

PROC LOGISTIC DATA = T ;
CLASS DRUG (ref="ASPIRIN") / param = ref ;
WEIGHT N ;
MODEL SUCCESS = DRUG ; RUN ;

```

Note:  $\text{Exp}(2.2336) = 9.333$

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	29.526	26.225
SC	28.912	24.998
-2 Log L	27.526	22.225

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5.3002	1	0.0213
Score	5.0505	1	0.0246
Wald	4.5304	1	0.0333

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
DRUG	1	4.5304	0.0333

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.8473	0.6901	1.5076	0.2195
DRUG	PLACEBO	1	2.2336	1.0494	4.5304	0.0333

Odds Ratio Estimates			
Effect		Point Estimate	95% Wald Confidence Limits
DRUG PLACEBO vs ASPIRIN		9.333	1.193 72.989

```

PROC LOGISTIC DATA = T ;
CLASS DRUG (ref="ASPIRIN") / param = ref ;
WEIGHT N ;
MODEL SUCCESS = DRUG ; RUN ;

```

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	29.526	26.225
SC	28.912	24.998
-2 Log L	27.526	22.225

My note:  
 $27.526 - 22.225 = 5.3002$

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	5.3002	1	0.0213
Score	5.0505	1	0.0246
Wald	4.5304	1	0.0333

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
DRUG	1	4.5304	0.0333

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.8473	0.6901	1.5076	0.2195
DRUG	PLACEBO	1	2.2336	1.0494	4.5304	0.0333

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
DRUG PLACEBO vs ASPIRIN	9.333	1.193	72.989

# So what if we had more than 1 predictor?

```
PROC LOGISTIC DATA = T ;  
CLASS DRUG ; WEIGHT N ;  
MODEL SUCCESS = DRUG SEX AGE ; RUN ;
```



# End of Lecture 2

*Next up in Part 4 Lecture 3: Diagnostic Accuracy*

