

BIOSTATS 640 - Intermediate Biostatistics Spring 2021

Unit 4. Categorical Data Analysis Summary

Part I: Single 2x2 Table

1. When is single 2x2 table analysis performed?	To describe overall pattern of exposure and event occurrence, to assess overall association, to obtain information for later use in assessing confounding and effect modification.
2. What is the research question?	Overall is there an association between exposure and event?
3. Probability models	Product Poisson, Product Binomial
4. Data Descriptions - numerical	See below for a suggested format.
5. Data Descriptions - graphical	See Part II: Stratified Analysis of 2x2 Tables
6. Test of General Association	Fisher Exact (always valid) or Chi Square (approximation)
7. Confidence Interval Estimation	See Notes 4: pp 20-23

Data Example -

Cups Coffee per day	MI	Control
≥ 5	7	18
< 5	20	112

Suggested Data Description (Numerical)

Crude Association Heart Attack (MI) with Coffee Consumption, n=157 (# Events = 27, 17.2%)

	n	# events (%)	OR	(95% CI)	Significance ^a
<u>Coffee Consumption</u>					
≥ 5 cups/day	25	7 (28.0%)	2.18	0.80, 5.96	.12
< 5 cups/day	132	20 (15.2%)	ref	-	

^a Fisher Exact Test or Chi Square Test (Null: No Association)

Part II: Stratified Analysis: K 2x2 Tables

1. When is a stratified analysis of K 2x2 tables performed?	This is performed when we suspect that a third variable (this will be the stratification variable) influences the exposure-event association of interest, either via confounding or modification, or intermediary.
2. What is the research question?	There is more than one. Taking them in order: _1. Is the exposure-event relationship different, depending on the level of the third variable? _2. If there is no modification, is there confounding? _3. What are the patterns of association?
3. Probability models	Product Poisson, Product Binomial
4. Data Descriptions - numerical	See below for a suggested format.
5. Data Descriptions - graphical	See below for suggested graphs
6. Test of Homogeneity of Association across strata	See Notes 4 pp 26-35
7. Given no effect modification, Test of Association	See Notes 4 pp 37-39

Data Example -

Stratum 1: FORMER SMOKER

Cups Coffee per day	MI	Control
≥ 5	7	18
< 5	20	112

Stratum 2: 1-14 CIGARETTES/DAY

Cups Coffee per day	MI	Control
≥ 5	7	24
< 5	33	11

Stratum 3: 35-44 CIGARETTES/DAY

Cups Coffee per day	MI	Control
≥ 5	27	24
< 5	55	58

Stratum 4: 45+ CIGARETTES/DAY

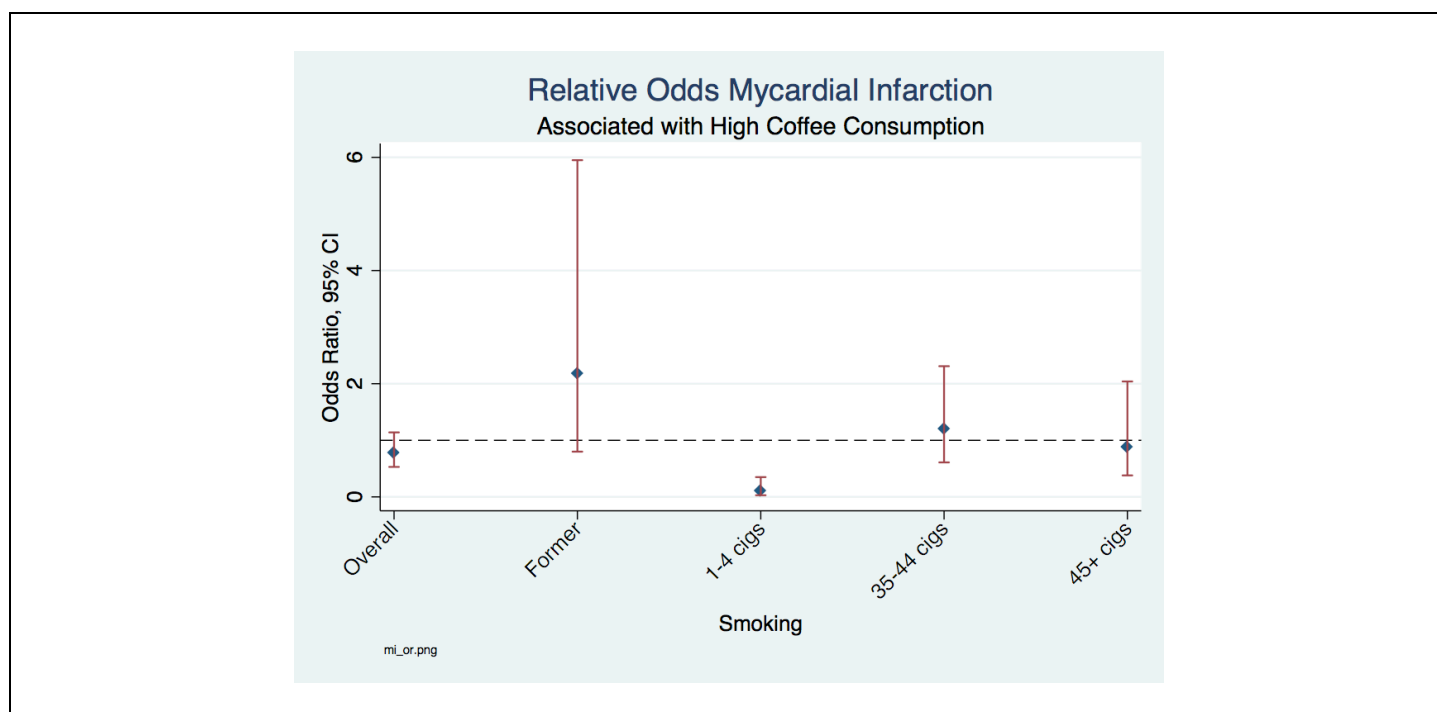
Cups Coffee per day	MI	Control
≥ 5	30	17
< 5	34	17

Suggested Data Description (Numerical)
Adjusted Associations Heart Attack (MI) with Coffee Consumption, by Smoking
n=157 (# Events = 27, 17.2%)

	n	# events (%)	OR	(95% CI)	Significance ^a
Former Smokers					
<u>Coffee Consumption</u>					
≥ 5 cups/day	25	7 (28.0%)	2.18	0.80, 5.96	.15
< 5 cups/day	132	20 (15.2%)	ref	-	
1-4 cigarettes/day					
<u>Coffee Consumption</u>					
≥ 5 cups/day	31	7 (22.6%)	0.10	0.03, 0.35	< .0001
< 5 cups/day	44	20 (75.0%)	ref	-	
35-44 cigarettes/day					
<u>Coffee Consumption</u>					
≥ 5 cups/day	51	27 (52.9%)	1.19	0.61, 2.31	.74
< 5 cups/day	113	55 (48.7%)	ref	-	
45+ cigarettes/day					
<u>Coffee Consumption</u>					
≥ 5 cups/day	47	30 (63.8%)	0.88	0.38, 2.04	.83
< 5 cups/day	51	34 (66.7%)	ref	-	
Mantel-Haenszel Test of Homogeneity of OR					0.0002

^a Fisher Exact Test (Null: No Association)

Suggested Graphical Description (Numerical)
Adjusted Associations Heart Attack (MI) with Coffee Consumption, by Smoking
n=157 (# Events = 27, 17.2%)



Part III: 2xC Table Analysis of Trend

1. When is a 2xC Analysis of Trend performed?	Okay, this is pretty self evident. We perform this analysis when we want to learn if higher levels of exposure are associated with more occurrences of event.
2. What is the research question?	Is there a "dose-response"?
3. Probability models	Product Multinomial
4. Data Descriptions - numerical	See below for a suggested format.
5. Data Descriptions - graphical	See below for suggested graphs
6. Test of Homogeneity of Odds w Dose	See Notes 4 pp 10-15
7. Test of Trend of Odds w Dose	See Notes 4 pp 40-45

Data Example (Event outcome is Esophageal Cancer) -

	Alcohol Consumption (g/day)				Total
	0-39	40-79	80-119	120+	
Cases	29	75	51	45	200
Controls	386	280	87	22	775
Total	415	355	138	67	975

Suggested Data Description (Numerical)

Distribution of Events of Esophageal Cancer with Alcohol Consumption

n=975 (# Events = 200, 20.5%)

	n	# events (%)	OR	(95% CI)	Significance ^b
Total	975	200 (20.5%)			
Alcohol Consumption (g/day)					
0-39	414	29	1.0 (ref)	Ref	
40-79	355	75	3.56	2.23, 5.67	
80-119	139	51	7.69	4.44, 13.33	
120+	67	45	27.15	12.48, 59.10	
Test of Homogeneity (Null: Equal Odds)					< .0001
Score Test of Trend of Odds					< .0001

Suggested Graph

Note – In the spirit of full disclosure...this graph was made with a smaller version of the data set so #'s don't quite match)

