Chapter 1: Introduction

1.1 Overview

- Contingency tables
- Categorical response measures
- Hypothesis testing
- Modeling strategies

1.2 Scale of Measurement

• Dichotomous

Treatment	Fav	Unfav	Total
Placebo	16	48	64
Test	40	20	60

• Ordinal

Improvement

Sex	Trt	Marked	Some	None	Total
Female	Active	16	5	6	27
Female	Placebo	6	7	19	32
Male	Active	5	2	7	14
Male	Placebo	1	0	10	11

• Ordinal continued

RHEUMATOID ARTHRITIS DATA

TREATMENT	SEX	AGE	IMPROVEMENT	TREATMENT	SEX	AGE	IMPROVEMENT
Active	М	27	1	Placebo	М	37	0
Active	M	29	0	Placebo	M	44	0
Active	M	30	0	Placebo	M	50	0
Active	M	32	2	Placebo	M	51	0
Active	M	46	2	Placebo	M	52	0
Active	M	58	2	Placebo	M	53	0
Active	M	59	0	Placebo	M	59	0
Active	M	59	2	Placebo	M	59	0
Active	M	63	0	Placebo	M	62	0
Active	M	63	0	Placebo	M	62	0
Active	M	64	0	Placebo	M	63	2
Active	M	64	1	Placebo	F	23	0
Active	M	69	0	Placebo	F	30	0
Active	M	70	2	Placebo	F	30	0
Active	F	23	0	Placebo	F	31	1
Active	F	32	0	Placebo	F	32	0
Active	F	37	1	Placebo	F	33	2
Active	F	41	0	Placebo	F	37	0
Active	F	41	2	Placebo	F	44	0
Active	F	48	0	Placebo	F	45	0
Active	F	48	2	Placebo	F	46	0
Active	F	55	2	Placebo	F	48	0
Active	F	55	2	Placebo	F	49	0
Active	F	56	2	Placebo	F	51	0
Active	F	57	2	Placebo	F	53	0
Active	F	57	2	Placebo	F	54	0
Active	F	57	2	Placebo	F	54	0

Active	F	58	0	Placebo	F	54	2
Active	F	59	2	Placebo	F	55	2
Active	F	59	2	Placebo	F	57	0
Active	F	60	2	Placebo	F	57	1
Active	F	61	2	Placebo	F	58	1
Active	F	62	1	Placebo	F	59	1
Active	F	62	2	Placebo	F	59	2
Active	F	66	2	Placebo	F	61	0
Active	F	67	2	Placebo	F	63	1
Active	F	68	1	Placebo	F	64	0
Active	F	68	2	Placebo	F	65	2
Active	F	69	0	Placebo	F	66	0
Active	F	69	1	Placebo	F	66	0
Active	F	70	1	Placebo	F	66	1
				Placebo	F	68	1
				Placebo	F	74	2

Improvement: 0 = NONE; 1 = SOME; 2 = MARKED

Nominal

Neighborhood

Party	Bayside	Highland	Longview	Sheffeld
Democrat	221	160	360	140
Independent	200	291	160	311
Republican	208	106	316	97

• Discrete Counts

Periods with Colds

Sex	Res	0	1	2	Total
Female	Rural	45	64	71	180
Female	Urban	80	104	116	300
Male	Rural	84	124	82	290
Male	Urban	106	117	87	310

Grouped Survival Times

Controls

Interval	No Recur	Recur	Withdraw	At Risk
0-1 Years	50	15	9	74
1-2 Years	30	13	7	50
2-3 Years	17	7	6	30

Active

Interval	No Recur	Recur	Withdraw	At Risk
0-1 Years	69	12	9	90
1-2 Years	59	7	3	69
2-3 Years	45	10	4	59

1.3 Sampling Frameworks

- Determine appropriate assumptions
- Determine type of inference that is possible
- Issue: what applies by virtue of study design versus rational assumption to support a likelihood function

- 1) Historical data
 - Observational data
 - Geographic or circumstantial definition
- 2) Experimental data
 - Random allocation of subjects
- 3) Sample survey data
 - Random sample of subjects from a larger study population

Major difference in the three frameworks is the use of randomization to obtain them.

- Randomization unit: Single subject vs. Clusters
- More complex designs may be used for selection

1.4 Overview of Analysis Strategies

- Hypothesis testing with randomization methods to test for an association
- Example: Multi-center study, hypothesis is whether there is an association between treatment and outcome, controlling for any effect of center.
- Null hypothesis implies hypergeometric distribution
 - basis for exact tests or expected values and covariance structure for approximate tests
- Statistical modeling techniques using maximum likelihood estimation or weighted least squares to describe the nature of the association

• Most common response function modeled for categorical data is the "logit"

$$logit(p) = log\left(\frac{p}{1-p}\right),\,$$

where *p* is the proportion of subjects with an event outcome.

• With sampling assumed to produce independent binomial distributions, maximum likelihood provides parameter estimates for the relationship of response to explanatory variables

Most often used methods

Dichotomous

- 1. Fisher's test
- 2. Mantel-Haenszel
- 3. Logistic Regression

Ordinal

- 1. Wilcoxon test for tied data in categories
- 2. Extended Mantel-Haenszel
- 3. Proportional odds model for logistic regression on ordered data

Grouped survival (time to event data)

- 1. Mantel-Haenszel test on life table (Mantel-Cox test)
- 2. Proportional hazards model via Poisson regression

Multivisit and Crossover Studies

- 1. Use of Extended Mantel-Haenszel
- 2. Gart test for crossover studies
- 3. Survey data logistic regression
- 4. Generalized Estimating Equations Methods (GEE)
- 5. Conditional logistic regression for crossover studies