Chapter 6: Sets of $s \times r$ Tables

6.1 Introduction

- Previous chapters addressed stratified analysis of sets of 2×2 tables, $2 \times r$ tables, and $s \times 2$ tables. These are all special cases of analysis of sets of $s \times r$ tables, which include the following:

  → row and column variables nominally scaled
  → row variable nominally scaled, column variable ordinally scaled
  → row variable and column variable both ordinally scaled

- Mantel-Haenszel procedure extended to handle these situations with the following corresponding alternatives to null hypothesis:

  → general association
  → mean score differences
  → linear correlation

- General idea of stratified analyses: to control for effects of factors that are part of research design,e.g. medical centers in a randomized clinical trial, or explanatory variables that are thought to be related to the response variable. Common strategy for retrospective and observational studies.

- M-H procedure requires **minimal assumptions**. Only assumptions required are randomization of subjects to levels of row variable.

- Minimal assumptions allow hypothesis tests on data that do not meet random sampling or underlying distributional assumptions...however, conclusions then restricted to study population at hand.

- Another advantage of M-H procedure is that sample size requirements are based on total frequencies summed across tables, rather than individual cell sizes.

6.2.4 Summary

- Summary of various types of extended Mantel-Haenszel statistics

MH Statistic	Alternative Hypothesis	SAS Label	Degrees of freedom	Scale	Nonpar. Equiv.
Q_{GMH}	general association	General Association	$(s-1) \times (r-1)$	none	
Q_{SMH}	mean score location shifts	Row Mean Scores Differ	$(s-1)$	column variable ordinal	Kruskal-Wallis
Q_{CSMH}	linear association	Nonzero Correlation	1	row and column variable ordinal	Spearman correlation

6.3 Mantel-Haenszel Applications

- Applications in many different settings:

  Chapter 3: Sets of 2×2 Tables
  Chapter 4: Sets of $2 \times r$ and $s \times 2$ Tables
  Chapter 6: Sets of $s \times r$ Tables

6.3.1 Dumping Syndrome Data

- Randomized clinical trial in four hospitals, where patients were assigned to one of four surgical procedures for treatment of severe duodenal ulcers

v + d:	vagotomy and drainage
v + a:	vagotomy and antrectomy (removal of 25% of gastric tissue)
v + h:	vagotomy and hemigastrectomy (removal of 50% of gastric tissue)
gre:	gastric resection (removal of 75% of gastric tissue)

- Response measured was severity (none, slight, moderate) of dumping syndrome.

- Is type of operation associated with severity of dumping syndrome, after adjusting for hospital?

Dumping Syndrome Data

Hospital	Operation	Severity of Symptoms None	Slight	Moderate	Total
1	v + d	23	7	2	32
1	v + a	23	10	5	38
1	v + h	20	13	5	38
1	gre	24	10	6	40
2	v + d	18	6	1	25
2	v + a	18	6	2	26
2	v + h	13	13	2	28
2	gre	9	15	2	26
3	v + d	8	6	3	17
3	v + a	12	4	4	20
3	v + h	11	6	2	19
3	gre	7	7	4	18
4	v + d	12	9	1	22
4	v + a	15	3	2	20
4	v + h	14	8	3	25
4	gre	13	6	4	23

- Since both row and column variables ordinally scaled, can use correlation statistic Q_{CSMH} to assess the null hypothesis of no association against the alternative that type of operation and severity of response are linearly associated.

```
proc freq order=data;
   weight wt;
   tables hospital*trt*severity / cmh;
   tables hospital*trt*severity / cmh scores=modridit;
run;
```

Table Scores

Summary Statistics for trt by severity
controlling for hospital

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	6.340	0.012
2	Row Mean Scores Differ	3	6.590	0.086
3	General Association	6	10.598	0.102

Standardized Midrank Scores

Summary Statistics for trt by severity
controlling for hospital

Cochran-Mantel-Haenszel Statistics (Modified Ridit Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	6.927	0.008
2	Row Mean Scores Differ	3	7.637	0.054
3	General Association	6	10.598	0.102

7.6 Rank Analysis of Covariance

- Analysis of covariance (ANCOVA) combines features of ANOVA and linear regression to compare some response variable between two or more groups

- ANCOVA increases precision in randomized experiments by using relationship between response and covariates to reduce error variability in comparing treatment groups

- Classical parametric ANCOVA depends on several assumptions. For situations in which these assumptions may not be satisfied, Quade (1967) proposed rank analysis of covariance

- Rank ANCOVA can be combined with extended Mantel-Haenszel statistics to carry out nonparametric comparisons between treatment groups, after adjusting for effects of one or more covariates

- Methodology can be extended to situation with multiple strata

- Example: Data from experiment to evaluate effectiveness of topically applied stannous fluoride (SF) and acid phosphate fluoride (APF) in reducing incidence of dental caries, as compared with placebo treatment (W)

- Compare number of decayed, missing or filled teeth (DMFT) after treatment among the three groups; use number of DMFT before treatment as covariate; stratify analysis by center

- The Mantel-Haenszel mean score statistic for a center × group × before specification yields a *p*-value of 0.0705. Although this provides some suggestion of atypical imbalance of baseline DMFT among the groups, there is no major reason not to proceed with the rank analysis of covariance procedure

- To use Mantel-Haenszel mean score statistic to perform Rank ANCOVA:

1. Compute standardized ranks of response (AFTER) and covariate (BEFORE) in each of three centers:

```
proc rank nplus1 ties=mean out=ranks;
       by center;
       var after before;
run;
```

2. Calculate residuals from regression of ranks:

```
proc reg data=ranks noprint;
        by center;
        model after = before;
        output out=residual r=resid;
run;
```

3. Compare mean values of residuals among groups, stratified by center:

```
proc freq data=residual;
        tables center*group*resid / noprint cmh2;
run;
```

- Results show there is a clearly significant difference among groups after adjusting for baseline number of DMFT and center (row mean score chi-square = 17.593, 2 df, $p < 0.001$)

Results of Stratified Rank Analysis of Covariance

Summary Statistics for group by resid
Controlling for center

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	17.172	<.0001
2	Row Mean Scores Differ	2	17.593	0.0002

Total Sample Size = 69

Example: Study to compare test treatment and placebo for management of a respiratory disorder (Reference: Koch, Carr, Amara, Stokes, and Uryniak [1990])

1. Two centers
2. Four visits
3. Ordered global response with 5 categories (terrible, poor, fair, good, excellent as scores 0,1,2,3,4)
4. Age, gender, baseline are covariables

Partial Listing of Data from a Multicenter, Multivisit Clinical Trial to Compare Two Treatments for Patients with a Respiratory Disorder (Reference: Koch et al [1990])

Center	Patient	Drug	Sex	Age	Base	Visit 1	Visit 2	Visit 3	Visit 4
1	53	A	F	32	1	2	2	4	2
	18	A	F	47	2	2	3	4	4
	54	A	M	11	4	4	4	4	2
	12	A	M	14	2	3	3	3	2
	51	A	M	15	0	2	3	3	3
	20	A	M	20	3	3	2	3	1
	16	A	M	22	1	2	2	2	3
	50	A	M	22	2	1	3	4	4
	03	A	M	23	3	3	4	4	3
	32	A	M	23	2	3	4	4	4
	56	A	M	25	2	3	3	2	3
	35	A	M	26	1	2	2	3	2
	26	A	M	26	2	2	2	2	2
	21	A	M	26	2	4	1	4	2
	08	A	M	28	1	2	2	1	2
	30	A	M	28	0	0	1	2	1
	33	A	M	30	3	3	4	4	2
	11	A	M	30	3	4	4	4	3

Multi-visit example from Koch, Carr et al [1990, Chapter 13]

1. Multi-center, multi-visit clinical trials to compare placebo and test drug for respiratory disorder (111 patients)

    a. two centers
    b. ordered global response with 5 categories (0 = terrible, 1 = poor, 2 = fair, 3 = good, 4 = excellent)
    c. four visits yield four variables
    d. baseline score, age, and gender are covariates

2. Table 5 results from stratified Wilcoxon (van Elteren) tests

Visit 1	Visit 2	Visit 3	Visit 4
0.052	<0.001	0.002	0.020

Visit 3 was specified as primary (and Visit 2 of interest as well)

3. Table 7 results from stratified rank analysis of covariance

Visit 1	Visit 2	Visit 3	Visit 4
0.010	<0.001	< 0.001	0.008

For average rankings over the 4 visits, $p < 0.001$. The covariable was baseline. Also, for the average rankings $p = 0.001$ for each of the two centers.

4. The method in Koch, Tangen et al [1998] for rank measures of association indicated $p = 0.023$ for treatment × visit interaction for the Mann-Whitney estimator for probability of better response with test treatment than control (with ties assigned 0.5 for better, with covariance adjustment for actual values of age, gender and baseline, and centers as strata).

Mantel-Haenszel methods can be formulated prior to data collection

1. They are applicable in some form to all measurement scales

2. They involve no explicit assumption about homogeneity of treatment differences across investigators

3. Adjustment for the association of response with demographic, medical history, baseline, or other explanatory variables can be undertaken through

   A. Refined stratification
   B. Nonparametric covariance extensions

4. Exact *p*-values can be determined by enumeration or simulation

6.4 Advanced Topic: Application to Repeated Measures

6.4.1 Introduction

- M-H strategy also useful for analysis of repeated measures data. Measurements are obtained over time, measured under multiple conditions, or obtained from two or more observers.

- Consider general situation in which *t* measurements of a univariate response variable *Y* are obtained from each of *n* experimental units, e.g.

→ repeated measurements obtained at time points
→ responses measured under multiple conditions
→ responses measured from matched case-control sets
→ responses obtained from members of each family
→ responses obtained from multiple newborns in each litter

- Focus on situation in which Y is categorical, then let c denote number of distinct values of Y and suppose

$$n_{ijk} = \begin{cases} 1 \text{ if subject } i \text{ classified in response } k \text{ at time } j \\ 0 \text{ otherwise} \end{cases}$$

for $i = 1,\ldots, n$; $j = 1,\ldots, t$; and $k = 1,\ldots, c$.

- Then data from subject i can be shown in $t \times c$ contingency table:

Time Point	Response Category			Total
	1	...	c	
1	n_{i11}	...	n_{i1c}	n_{i1+}
⋮	⋮	⋱	⋮	⋮
t	n_{it1}	...	n_{itc}	n_{it+}
Total	n_{i+1}	...	n_{i+c}	n_i

and data can be viewed as a set of n independent two-way contingency tables.

- Mantel-Haenszel statistics can be used to test null hypothesis of no association between rows (time) and columns (response), adjusted for subject.

- Under assumption that marginal totals $\{n_{ij+}\}$ and $\{n_{i+k}\}$ of each table are fixed, null hypothesis is that for each subject, the response Y is distributed at random with respect to t time points.

- The following examples demonstrate use of MH statistics in testing marginal homogeneity for repeated measures.

6.4.2 Dichotomous Response: Two Time Points (McNemar's Test)

- Running shoe company produces new model of running shoe that corrects for overpronation. Company conducted a study on 87 runners who used new shoe for a month. Researchers asked participants whether they experienced occasional heel tenderness before and after they used the new shoe.

- Table 6.7: Heel Tenderness for Runners

Before	After		Total
	No	Yes	
No	48	15	63
Yes	5	19	24

- Can think of measurements for each subject as being one of four 2 × 2 tables, corresponding to the four cells above. Each subject's set of responses can be represented by one of these tables…

(No, No) Configuration Table (48)

Time	Heel Tenderness		Total
	No	Yes	
Before	1	0	1
After	1	0	1

(No, Yes) Configuration Table (15)

Time	Heel Tenderness		Total
	No	Yes	
Before	1	0	1
After	0	1	1

(Yes, No) Configuration Table (5)

Time	Heel Tenderness		Total
	No	Yes	
Before	0	1	1
After	1	0	1

(Yes, Yes) Configuration Table (19)

Time	Heel Tenderness		Total
	No	Yes	
Before	0	1	1
After	0	1	1

- The following analysis demonstrates the Mantel-Haenszel approach to analyzing repeated measurements data:

```
proc freq data=pump;
    tables subject*time*response / noprint cmh;
run;
```

Q_{MH} = 5.00 with p = 0.025

- Another way of obtaining these results is to input original 2 × 2 table (Table 6.7) and specify AGREE option to obtain McNemar's Test:

```
proc freq;
    weight count;
    tables before*after / agree;
run;
```

$Q_M = 5.00$ with $p = 0.025$

6.4.3 Dichotomous Response: Three Repeated Measurements

- The following data are from a study in which 46 patients were each treated with three drugs (A, B, and C). Response to each drug was recorded as favorable (F) or unfavorable (U).

Drug			
A	B	C	Frequency
F	F	F	6
F	F	U	16
F	U	F	2
F	U	U	4
U	F	F	2
U	F	U	4
U	U	F	6
U	U	U	6

- Question: Do the three drugs have similar probabilities for favorable response?

- H_0: Interchangeability (no association between drug and response for each patient)

- Q_{GMH} can test this hypothesis if data are restructured so that there is one 3×2 contingency for each of 46 patients, e.g.:

Drug	Response		Total
	F	U	
A	1	0	1
B	0	1	1
C	1	0	1

- ```
  proc freq data=drug;
     tables patient*drug*response / noprint cmh;
  run;
  ```

$Q_{GMH} = 8.471$ with $p = 0.014$
  ```

6.4.4 Ordinal Response

- Same M-H strategy is appropriate when repeated measurements response variable is ordinally scaled. In this case use the mean score statistic, Q_{SMH}

- Following data is from a study which evaluated efficacy of steam inhalation in treatment of common cold symptoms

- On four successive days, patients self-assessed severity of nasal drainage as no symptoms (0), mild symptoms (1), moderate symptoms (2), or severe symptoms (3)

- Does nasal drainage become less severe following steam inhalation treatment?

Table 6.14 Nasal Drainage Data

Patient ID	Study day				Patient ID	Study day			
	1	2	3	4		1	2	3	4
1	1	1	2	2	16	2	1	1	1
2	0	0	0	0	17	1	1	1	1
3	1	1	1	1	18	2	2	2	2
4	1	1	1	1	19	3	1	1	1
5	0	2	2	0	20	1	1	2	1
6	2	0	0	0	21	2	1	1	2
7	2	2	1	2	22	2	2	2	2
8	1	1	1	0	23	1	1	1	1
9	3	2	1	1	24	2	2	3	1
10	2	2	2	3	25	2	0	0	0
11	1	0	1	1	26	1	1	1	1
12	2	3	2	2	27	0	1	1	0
13	1	3	2	1	28	1	1	1	1
14	2	1	1	1	29	1	1	1	0
15	2	3	3	3	30	3	3	3	3

You can generate the mean scores at days 1-4 using PROC GLM and the LSMEANS statement. You can also use the ESTIMATE statement to determine the direction of the mean difference between days from the 1st day to the 4th day:

```
proc glm;
   class id day;
   model drainage=id day;
   lsmeans day;
   estimate 'direction' day -3 -1 1 3 /divisor=6;
run;
```

Mean Severity Scores

Least Squares Means

day	drainage LSMEAN
1	1.50000000
2	1.36666667
3	1.36666667
4	1.16666667

Average Pairwise Distance Measure Estimate

Dependent Variable: drainage

Parameter	Estimate	Standard Error	t Value	Pr > |t|
direction	-0.16666667	0.07897276	-2.11	0.0377

```
proc freq;
   tables id*day*drainage / cmh noprint;
run;
```

Summary Statistics for day by drainage
Controlling for id

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	4.3548	0.0369
2	Row Mean Scores Differ	3	4.9355	0.1766
3	General Association	9	10.1267	0.3403

Total Sample Size = 120

6.4.5 Ordinal Response with Missing Data

- Following data is from a study which investigated effect of pulse duration on development of acute electrical injury during transesophageal atrial pacing in animals
- Lesion severity classified according to depth of injury by histologic examination using ordinal scale from 0 to 5 (0 = no lesion, 5 = acute inflammation). Missing observations denoted by – in table
- Does increasing pulse duration from 2 to 10 ms tend to increase severity of lesion?

Table 6.15 Lesion Severity Data

ID	Pulse duration (ms)				
	2	4	6	8	10
6	0	0	5	0	3
7	0	3	3	4	5
8	0	3	4	3	2
9	2	2	3	0	4
10	0	0	4	4	3
12	0	0	0	4	4
13	0	4	4	4	0
15	0	4	0	0	0
16	0	3	0	1	1
17	-	-	0	1	0
19	0	0	1	1	0
20	-	0	0	2	2
21	0	0	2	3	3
22	-	0	0	3	0

```
data animals;
   keep id pulse severity;
   input id sev2 sev4 sev6 sev8 sev10;
   pulse=2;  severity=sev2;  output;
   pulse=4;  severity=sev4;  output;
   pulse=6;  severity=sev6;  output;
   pulse=8;  severity=sev8;  output;
   pulse=10; severity=sev10; output;
   datalines;
 6 0 0 5 0 3
 ⋮
22 . 0 0 3 0
;
```

```
proc glm;
   class id pulse;
   model severity = id pulse;
   lsmeans pulse;
   estimate 'direction' pulse -4 -2 0 2 4 / divisor=10;
run;
```

Least Squares Means

pulse	severity LSMEAN
2	-0.13861139
4	1.33516484
6	1.85714286
8	2.14285714
10	1.92857143

Dependent Variable: severity

Parameter	Estimate	Standard Error	t Value	Pr > |t|
direction	0.98841159	0.27649458	3.57	0.0008

```
proc freq;
  tables id*pulse*severity / noprint cmh2 scores=modridit;
run;
```

MH Tests Using Modified Ridit Scores

Cochran-Mantel-Haenszel Statistics (Modified Ridit Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	10.110	0.001
2	Row Mean Scores Differ	4	14.133	0.007

Partial data from the incomplete cases still provide useful information about the association between pulse duration and lesion severity.

6.4.6 Continuous Response

- Table 6.19 shows artificial data collected for purpose of determining if pH level alters action potential characteristics following administration of a drug

- Response variable of interest was measured at up to four pH levels for each of 25 patients

- M-H statistics can still be used to determine if average Vmax differs among four pH values (Q_{SMH}), even though response is a continuous measurement

- M-H methodology accommodates varying numbers of observations per patient (under assumption that missing values are missing completely at random and test statistic is specified with either table scores or ranks)

- ```
  proc freq data=ph_vmax;
     tables subject*ph*vmax / noprint cmh2
         scores=modridit;
  run;
  ```
  ```

Table 6.19 Action Potential Data

Patient	pH Level				Patient	pH Level			
	6.5	6.9	7.4	7.9		6.5	6.9	7.4	7.9
1		284	310	326	14	204	234	268	
2			261	292	15			258	267
3		213	224	240	16		193	224	235
4		222	235	247	17	185	222	252	263
5			270	286	18		238	301	300
6			210	218	19		198	240	
7		216	234	237	20		235	255	
8		236	273	283	21		216	238	
9	220	249	270	281	22		197	212	219
10	166	218	244		23		234	238	
11	227	258	282	286	24			295	281
12	216		284		25			261	272
13			257	284					

MH Mean Score and Correlation Tests: Modified Ridit Scores

Summary Statistics for ph by vmax
Controlling for subject

Cochran-Mantel-Haenszel Statistics (Modified Ridit Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	35.3818	<.0001
2	Row Mean Scores Differ	3	34.7945	<.0001

Effective Sample Size = 66
Frequency Missing = 34

WARNING: 34% of the data are missing.

Mantel-Haenszel Methods can be formulated prior to data collection

1. They are applicable in some form to all measurement scales

2. They involve no explicit assumption about homogeneity of treatment differences across investigators

3. Adjustment for the association of response with demographic, medical history, baseline, or other explanatory variables can be undertaken through:

    a. refined stratification

    b. nonparametric covariance extensions

4. Exact *p*-values can be determined by enumeration or simulation