Tyler Maus

ADVANCED QUALITATIVE DATA ANALYSIS
Homework No. 3

Problems: #2.3, 2.8, 2.12, 2.19, 2.39

2.3 The response variable is injury type: fatal or nonfatal.

Letting I = injury is fatal & S = safety equipment in use...

a) Difference: $P(I=y \mid S=N) - P(I=y \mid S=y) = \frac{\pi_{11}}{\pi_{1+}} - \frac{\pi_{21}}{\pi_{2+}}$

$$= \frac{1601}{164128} - \frac{510}{412878} = 0.0085193501$$

b) Relative Risk $RR = \frac{P(I=y \mid S=N)}{P(I=y \mid S=y)} = \left(\frac{\pi_{11}}{\pi_{1+}}\right) \Big/ \left(\frac{\pi_{21}}{\pi_{2+}}\right)$

$$= 7.896965138$$

c) Odds Ratio $OR = \frac{\text{odds}(I=y \text{ vs. } N \mid S=N)}{\text{Odds}(I=y \text{ vs. } N \mid S=y)} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$

$$= \frac{(1601)(412{,}368)}{(510)(162{,}527)} = 7.964904872$$

- The differences of proportions is 0.0085193501 indicating that the probability of a fatality is 0.0085 higher when no seat belt is used compared to when a seatbelt is used.
- The relative risk is 7.896965138, indicating that the probability of fatality is 7.90 times greater when not using a seat belt relative to when a seat belt is used
- The odds ratio is 7.964904872, indicating that the odds of a fatality when not wearing a seatbelt is 7.96 times that of odds of a fatality when wearing a seatbelt.
- Since the probability π_{11} of a fatality is relatively small (< 0.1%) the odds ratio and relative risk are approximately equal.

2.8 a) This interpretation reflects a risk ratio rather than an odds ratio, as the interpretation references probability. The correct interpretation would be that the <u>odds</u> of survival for femal[illegible] was 11.4 that of the odds of survival for men.
The quoted interpretation would be approximately correct whe[illegible] the odds ratio and risk ratio are approximately equal; this occurs when the outcome of interest (survival) is close t[illegible] zero for both groups (male & female) — this makes sense in the context of the Titanic, as there was a famously lo[illegible] survival rate.

b) odds of survival for females = $\pi_{11}/\pi_{12} = 2.9$

odds ratio = $\dfrac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = 11.4$

thus odds of survival for males = $\pi_{21}/\pi_{22} = 3.931034483$

$\left(\dfrac{3.9}{2.9}\right)^{-1} = 0.7435897436$ is the proportion of females who surviv[illegible]

$\left(\dfrac{4.93...}{3.93...}\right)^{-1} = 0.7972027972$ is the proportion of males who surviv[illegible]

Admitted?

	Y	N	
M	μ_{11}	μ_{12}	Gender
F	μ_{21}	μ_{22}	

2.12 A = whether admitted
G = gender
D = department

conditional AG odds ratio: $\theta_{AG(d)} = \dfrac{\mu_{11d}\,\mu_{22d}}{\mu_{12d}\,\mu_{21d}}$

AG marginal odds ratio: $\theta_{AG} = \dfrac{\mu_{11+}\,\mu_{22+}}{\mu_{12+}\,\mu_{21+}}$

conditional $\theta_{AGA} = \mu_{11A}\mu_{22A} / \mu_{12A}\mu_{21A} = \dfrac{(512)(19)}{(313)(89)} = 0.3492$
-047

$\theta_{AGB} = \mu_{11B}\mu_{22B} / \mu_{12B}\mu_{21B} = \dfrac{(353)(8)}{(207)(17)} = 0.80250071$

$\theta_{AGC} = (120)(391)/(205)(202) = 1.133059047$

$\theta_{AGD} = (138)(244)/(279)(131) = 0.9212837561$

$\theta_{AGE} = (53)(299)/(138)(94) = 1.221631206$

$\theta_{AGF} = (22)(317)/(351)(24) = 0.8278727445$

marginal $\theta_{AG} = (1198)(1278)/(1493)(557) = 1.841080037$

- The odds of males being admitted to department A is 0.349212017 that of the odds of females
- " " department B is 0.8025007101 that of the odds of females
- " " department C is 1.1330596247 that of " "
- " " department D is 0.921283756 that of " "
- " " department E is 1.221631206 that of " "
- " " department F is 0.827872744 that of the odds of females
- The odds of males being admitted to departments A through F is 1.841080037 that of the odds of females

The different indications of the AG association are due to Simpson's paradox — some departments have an outsized influence on the overall odds ratio.

2.19 In SAS I find that Cohen's Kappa (simple Kappa) is estimated at 0.1293 w/ a standard error of 0.0686. We are 95% confident that the true value of Kappa falls between -0.0051 and 0.2638. As a kappa value of 1 indicates perfect agreement, in this case we would reject a null hypothesis of perfect agreement at $\alpha = 0.05$

In addition $\gamma = 0.3604$; we are 95% confident that γ lies between 0.1401 and 0.5806. The value $\gamma \neq 0$, which would imply independence. It is not close to -1 or 1 either, which would imply perfect discordance or concordance respectively.

```
options ls=72 ps=max nocenter;
data fun;
do husband=1 to 4;
    do wife=1 to 4;
        input count @@;
        output;
    end;
end;
datalines;
7 7 2 3
2 8 3 7
1 5 4 9
2 8 9 14
;
run;
proc freq data=fun;
weight count;
table husband*wife/measures agree cl;
run;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of husband by wife					
	wife				
husband	1	2	3	4	Total
1	7 7.69 36.84 58.33	7 7.69 36.84 25.00	2 2.20 10.53 11.11	3 3.30 15.79 9.09	19 20.88
2	2 2.20 10.00 16.67	8 8.79 40.00 28.57	3 3.30 15.00 16.67	7 7.69 35.00 21.21	20 21.98
3	1 1.10 5.26 8.33	5 5.49 26.32 17.86	4 4.40 21.05 22.22	9 9.89 47.37 27.27	19 20.88
4	2 2.20 6.06 16.67	8 8.79 24.24 28.57	9 9.89 27.27 50.00	14 15.38 42.42 42.42	33 36.26
Total	12 13.19	28 30.77	18 19.78	33 36.26	91 100.00

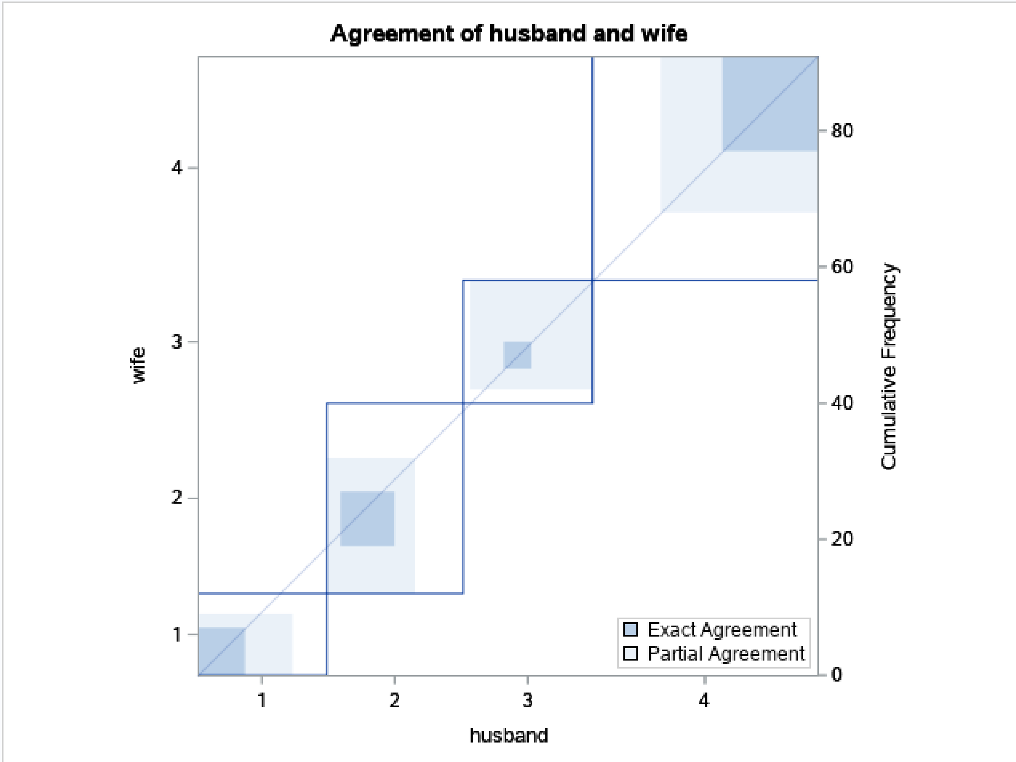
Statistics for Table of husband by wife

Statistic	Value	ASE	95% Confidence Limits	
Gamma	0.3604	0.1124	0.1401	0.5806
Kendall's Tau-b	0.2661	0.0859	0.0977	0.4345
Stuart's Tau-c	0.2573	0.0840	0.0927	0.4219
Somers' D C|R	0.2633	0.0853	0.0961	0.4304
Somers' D R|C	0.2690	0.0868	0.0990	0.4391
Pearson Correlation	0.3336	0.0977	0.1420	0.5251
Spearman Correlation	0.3143	0.0995	0.1192	0.5094
Lambda Asymmetric C|R	0.0862	0.0824	0.0000	0.2477
Lambda Asymmetric R|C	0.0862	0.0824	0.0000	0.2477
Lambda Symmetric	0.0862	0.0744	0.0000	0.2321
Uncertainty Coefficient C|R	0.0645	0.0323	0.0012	0.1279
Uncertainty Coefficient R|C	0.0628	0.0317	0.0007	0.1249
Uncertainty Coefficient Symmetric	0.0637	0.0320	0.0010	0.1264

Symmetry Test		
Chi-Square	DF	Pr > ChiSq
3.8778	6	0.6932

Kappa Statistics				
Statistic	Estimate	Standard Error	95% Confidence Limits	
Simple Kappa	0.1293	0.0686	-0.0051	0.2638
Weighted Kappa	0.2374	0.0783	0.0839	0.3909

Sample Size = 91


Agreement of husband and wife
wife
4
3
2
1
1
2
3
4
husband
Cumulative Frequency
80
60
40
20
0
Exact Agreement
Partial Agreement

2.39 $V[y] = 1 - \max_j\{\pi_{+j}\}$ and $V[y|i] = 1 - \max_j\{\pi_{j|i}\}$

Show that independence implies $\lambda = 0$ but that the converse is not true.

Proportional reduction in variance measure has form:

$$\left(V[y] - E[V[y|x]]\right) / V[y]$$

We know that under independence $\pi_{j|i} = \pi_{+j}$, as the choice of i has no impact on the probability/frequency of a given level of

Therefore $V[y] = V[y|i] = 1 - \max\{\pi_{+j}\}$ so

$$\lambda = \left(V[y] - E[V[y|x]]\right) / V[y] = \left(V[y] - V[y]\right) / V[y] = 0$$

under independence.

However we can choose π_{i+} and π_{j+} and π_{ij} s.t.

$V[y] = E[V[y|x]]$ without independence

As a corner case consider $\pi_{11} = 4e14$, $\pi_{ij} = 1$ $\forall i \neq 1$ or $j \neq 1$

4e14	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

Agreement is near 100% yet $V[y] \approx 1 - 1 = 0$ and $E[V[y|i]] \approx 1 - 1 = 0$ so $\lambda = 0$.

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of husband by wife					
	wife				
husband	1	2	3	4	Total
1	4E14 100.00 100.00 100.00	1 0.00 0.00 25.00	1 0.00 0.00 25.00	1 0.00 0.00 25.00	4E14 100.00
2	1 0.00 25.00 0.00	1 0.00 25.00 25.00	1 0.00 25.00 25.00	1 0.00 25.00 25.00	4 0.00
3	1 0.00 25.00 0.00	1 0.00 25.00 25.00	1 0.00 25.00 25.00	1 0.00 25.00 25.00	4 0.00
4	1 0.00 25.00 0.00	1 0.00 25.00 25.00	1 0.00 25.00 25.00	1 0.00 25.00 25.00	4 0.00
Total	4E14 100.00	4 0.00	4 0.00	4 0.00	4E14 100.00

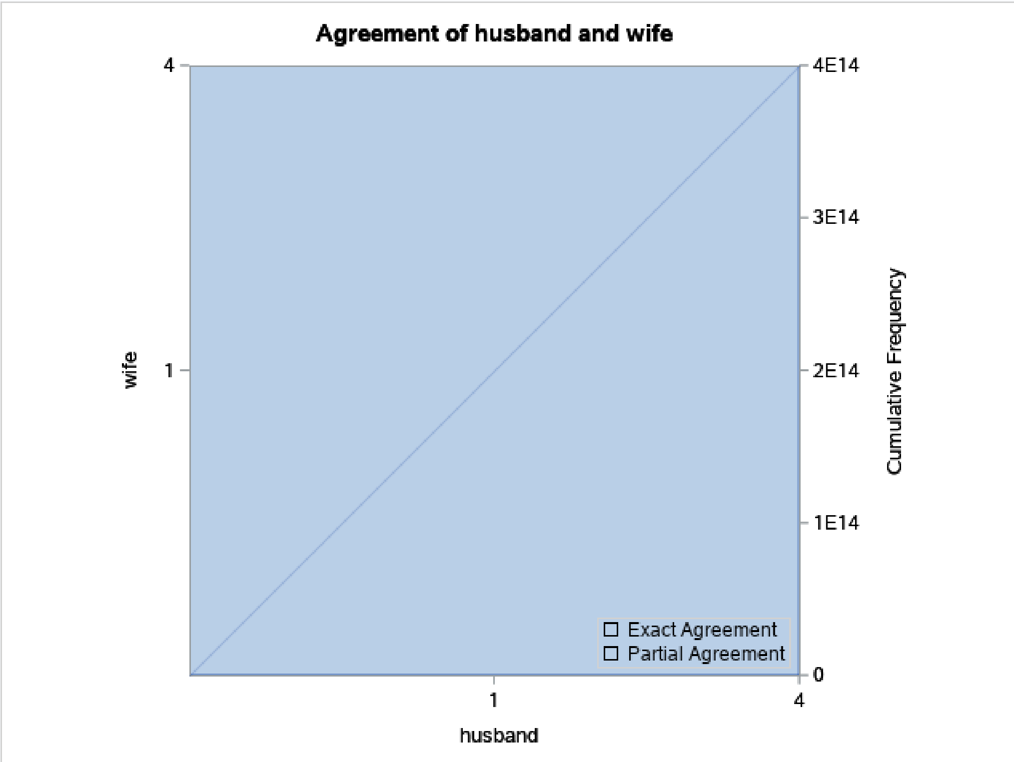
Statistics for Table of husband by wife

Statistic	Value	ASE	95% Confidence Limits	
Gamma	1.0000	0.0000	1.0000	1.0000
Kendall's Tau-b	0.7496	0.1014	0.5508	0.9484
Stuart's Tau-c	0.0000	0.0000	0.0000	0.0000
Somers' D C|R	0.7496	0.1251	0.5045	0.9947
Somers' D R|C	0.7496	0.1251	0.5045	0.9947
Pearson Correlation	0.6433	0.1125	0.4228	0.8637
Spearman Correlation	0.7508	0.0932	0.5682	0.9334
Lambda Asymmetric C|R	0.0000	0.2041	0.0000	0.4001
Lambda Asymmetric R|C	0.0000	0.2041	0.0000	0.4001
Lambda Symmetric	0.0000	0.1768	0.0000	0.3465
Uncertainty Coefficient C|R	0.6978	0.1214	0.4598	0.9358
Uncertainty Coefficient R|C	0.6978	0.1214	0.4598	0.9358
Uncertainty Coefficient Symmetric	0.6978	0.0999	0.5021	0.8935

Symmetry Test		
Chi-Square	DF	Pr > ChiSq
0.0000	6	1.0000

Kappa Statistics				
Statistic	Estimate	Standard Error	95% Confidence Limits	
Simple Kappa	.	.	.	.
Weighted Kappa	.	.	.	.

Sample Size = 4E14


Agreement of husband and wife
4
1
wife
4E14
3E14
2E14
1E14
0
Cumulative Frequency
Exact Agreement
Partial Agreement
1
4
husband