Joey Angel

Clinical Trials- Homework 5

Question 1:

```
1  data FactorX;
2     input Factor $ Treatment $ Response $ Count @@;
3     datalines;
4  + A 1 40 + B 1 32
5  - A 1 60 - B 1 48
6  + A 0 10 + B 0 18
7  - A 0 90 - B 0 102
8  ;
9
10 ods graphics on;
11 proc freq data= FactorX;
12    tables Factor*Treatment*Response /
13            relrisk plots(only)=relriskplot(stats) cmh;
14    weight Count;
15 run;
16 ods graphics off;
17
```

The FREQ Procedure

**Summary Statistics for Treatment by Response
Controlling for Factor**

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	4.4586	0.0347
2	Row Mean Scores Differ	1	4.4586	0.0347
3	General Association	1	4.4586	0.0347

With p-value .0347 < .05 we reject the null hypothesis and conclude there is a statistically significant difference between the response rate for the A and B treatments when adjusting for the effect of factors + and -.

```
18 proc freq data=DATA;
19   table Treatment*Response /chisq nocol nopercent;
20   weight Count;
21   format Treatment $txfmt.;
22 run;
```

Statistic	DF	Value	Prob
Chi-Square	1	4.0404	0.0444

Similarly, we would also reject the null hypothesis and conclude there is a significant difference in the response rate for the A and B treatments without adjusting for the effect of factors + and -. However, the p-value did increase to .0444 from .0347 but both are significant at the 5% level.

Question 2:

```
26  /*Question 2*/
27  data DATA2;
28     input Factor $ Treatment $ Response $ Count @@;
29     datalines;
30  + A 1 47 + B 1 26
31  - A 1 56 - B 1 51
32  + A 0 12 + B 0 15
33  - A 0 85 - B 0 108
34  ;
35
36  ods graphics on;
37  proc freq data= DATA2;
38     tables Factor*Treatment*Response /
39             relrisk plots(only)=relriskplot(stats) cmh;
40     weight Count;
41  run;
42  ods graphics off;
```

The FREQ Procedure

**Summary Statistics for Treatment by Response
Controlling for Factor**

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	4.2201	0.0399
2	Row Mean Scores Differ	1	4.2201	0.0399
3	General Association	1	4.2201	0.0399

With p-value .0399 < .05 we reject the null hypothesis and conclude there is a statistically significant difference between the response rate for the A and B treatments when adjusting for the effect of factors + and -.

```
44  proc freq data=DATA2;
45    table Treatment*Response /chisq nocol nopercent;
46    weight Count;
47    format Treatment $txfmt.;
48  run;
```

Statistics for Table of Treatment by Response

Statistic	DF	Value	Prob
Chi-Square	1	6.8283	0.0090

Again, we would also reject the null hypothesis and conclude there is a significant difference in the response rate for the A and B treatments without adjusting for the effect of factors + and -. However, the p-value did decrease from .0399 to .009. This indicates the factor is likely not an effective confounding variable we need to include in the analysis.

Question 3:

```
51  /*Question 3*/
52  data DATA3;
53     input Factor $ Treatment $ Response $ Count @@;
54     datalines;
55  + A 1 33 + B 1 38
56  - A 1 64 - B 1 45
57  + A 0 8 + B 0 21
58  - A 0 95 - B 0 96
59  ;
60
61  ods graphics on;
62  proc freq data= DATA3;
63     tables Factor*Treatment*Response /
64            relrisk plots(only)=relriskplot(stats) cmh;
65     weight Count;
66  run;
67  ods graphics off;
```

The FREQ Procedure

**Summary Statistics for Treatment by Response
Controlling for Factor**

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

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	4.5771	0.0324
2	Row Mean Scores Differ	1	4.5771	0.0324
3	General Association	1	4.5771	0.0324

With p-value .0324 < .05 we reject the null hypothesis and conclude there is a statistically significant difference between the response rate for the A and B treatments when adjusting for the effect of factors + and -.

```
69  proc freq data=DATA3;
70    table Treatment*Response /chisq nocol nopercent;
71    weight Count;
72    format Treatment $txfmt.;
73  run;
```

Statistics for Table of Treatment by Response

Statistic	DF	Value	Prob
Chi-Square	1	1.9798	0.1594

This time we would fail to reject the null hypothesis and find that there is not a significant difference in the response rate for the A and B treatments without adjusting for the effect of factors + and -. The p-value increased from .0324 to .1594 and is therefore no longer significant at the 5% level. This indicates the factor is an effective confounding variable we should include in the analysis.

Question 4:

First we set up the data as in example 11 but kept the age variable and defined logage = log(age). Now we consider whether the relationships between logcnt and logbase and between logcnt and logage are linear. logage clearly has a linear relationship but logbase could also be used with a high order term but we will stick with it being a linear relationship since it is still reasonable.

We then estimate the dispersion parameter phi using the model cnt = trt|visit logbase|logage|trt|visit and store it to select an error structure in the next step. The SAS output tells us the estimate is 2.1243.

We will now choose a covariance structure. Using the unstructured covariance we the error: "ERROR: Error in computing the variance function," so we will not use the unstructured covariance as it is not a good fit for our model. We get three very similar values but the AR(1) gave us a slightly smaller value than the CS so we will continue with the AR(1) variance structure.

```
Indep.:    -1302.3390
AR(1)  :   -1302.4522
CS     :   -1302.4424
```

Now we will create models and choose the one with the lowest QICu as our model. We reduce the mean model with backward selection and by eliminating higher-order interactions. We also estimate phi using the working models to have a better estimate of the standard error. Of the 14 models we tested, the model with the lowest QICu is number 7 which has the following score statistics and QICu:

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
trt	1	6.35	0.0117
visit	3	7.08	0.0695
trt*visit	3	6.68	0.0828
logbase	0	.	.
logbase*trt	1	4.08	0.0435
logbase*visit	3	7.94	0.0473
logbase*trt*visit	3	6.10	0.1067
logage	1	7.39	0.0065
logage*visit	3	6.14	0.1048

```
mean1:     -1278.0590
mean2:     -1293.4990
mean3:     -1307.0317
mean4:     -1266.2944
mean5:     -1179.3371
mean6:     -1307.0731
mean7:     -1331.1432
mean8:     -1297.7000
mean9:     -1308.9752
mean10:    -1201.4321
mean11:    -1308.5561
mean12:    -1193.1612
mean13:    -1221.3559
mean14:    -1229.1075
```

In our model, logage does not have an interaction term with trt so it is unnecessary and not possible to make a contrast statement including logage*trt. The results show the point estimate of the mean ratio at the mean of logbase is 1.4201 although this would fail to be statistically significant from 0 at the 5% significance level.

The GENMOD Procedure

		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
trt-wald	3.8500	1.7109	8.6636	1.3481	0.4138	0.05	0.5370	2.1591	10.61	0.0011
trt-wald at mean	1.4201	0.9986	2.0196	0.3507	0.1797	0.05	-0.0014	0.7029	3.81	0.0509

Contrast Results for GEE Analysis

Contrast	DF	Chi-Square	Pr > ChiSq	Type
trt-score	1	6.35	0.0117	Score
trt-wald	1	10.61	0.0011	Wald
trt-score at mean	1	3.84	0.0502	Score
trt-wald at mean	1	3.81	0.0509	Wald