BIOS 665 Homework 3
Joyce Choe
10/4/2024

1.

a) Stratified analysis: Mantel-Haenszel Test

Assumptions are randomization into independent groups, a binary outcome, a stratification (or confounder) variable to adjust for that is associated with both treatment and outcome, and sufficient sample sizes per cell (assessed by Mantel-Fleiss criterion (> 5). The Mantel-Haenszel test statistic (Q_{MH}) is 5.068 (p-value = 0.0244), which is larger than the chi-squared test statistic under H_0, 3.84 df 1. H_0: Treatment and presence of rash after 24 hours are not associated, after controlling for center.

$$H_0 : OR = 1$$

H_A: Treatment and presence of rash after 24 hours are associated, controlling for center.

$$H_A : OR \neq 1$$

At $\alpha = 0.05$, we reject H_0 and conclude that treatment and presence of rash after 24 hours are associated, controlling for center.

b) i. In Center A, the odds of rash outcome after 24 hours in active treatment was 2.286 times than that of odds in placebo group with 95% CI: (1.189, 4.395).
ii. In Center B, the odds of rash outcome after 24 hours in active treatment was 1.257 times than that of odds in placebo group with 95% CI: (0.670, 2.358).
iii. In both centers, the odds of rash outcome after 24 hours in active treatment was 1.676 times than that of odds in placebo group with 95% CI: (1.068, 2.630), controlling for center.
Assume: same effect of active treatment on outcome in both centers.

c) We use Breslow-Day (BD) test to assess homogeneity of odds ratios between active vs. placebo treatment on rash outcome across centers, assuming sufficiently large sample size per center.
H_0: The odds ratios comparing active treatment to placebo for rash outcome across centers are homogeneous.
Breslow-Day chi-square test statistic 1 df is 1.671 (p-value=0.196). We fail to reject H_0 and conclude that the odds ratios across centers are homogeneous. Thus, the overall OR is a reliable estimate.

2.

a)

$$logit(p_{hi}) = \alpha + \beta_1(centerA) + \beta_2(activetreatment)$$

Variable	Estimate(SE)	Definition/Parameter Interpretation
$logit(p_{hi})$		logit probability of rash response for hth center & ith treatment
α	-0.0925 (0.1974)	reference log odds of center B and placebo treatment
β_1	0.1086 (0.2306)	increment to log odds for center A
β_2	0.5192 (0.2306)	increment to log odds for active treatment

h=center A, center B
i=active, placebo

Assumptions for a dichotomous logistic regression main effects model are a stratified simple random sample, one or more explanatory variables as main effects (categorical or continuous), and a dichotomous response variable that follows a binomial distribution for each factor level combination.

b) The asymptotic odds of rash outcome in active treatment is 1.681 times the odds of rash outcome in placebo treatment, adjusting for center and we are 95% confident that the true odds ratio is between 1.070 and 2.641. Since OR does not contain 1, treatment may be associated with rash outcome, adjusting for center.

c) We use Hosmer and Lemeshow (HL) chi-square to test homogeneity of odds ratios across centers in a logistic regression model, assuming sufficiently large sample size.
H_0: The odds ratios comparing active treatment to placebo for rash outcome across centers are homogeneous because the model fits adequately.
The HL chi-square test statistic is 1.671 with 2 df (p-value=0.4337). At $\alpha = 0.05$, we fail to reject H_0, and conclude that the odds ratios across centers are homogeneous. Thus, overall OR and fitted model are reliable when comparing treatment on response across different centers.

d) The BD test statistic is 1.671 with 1 df, and the HL chi square statistic is 1.671 with 2 df. The HL test is more conservative because $\chi^2_2 = 5.99$, whereas the BD test $\chi^2_1 = 3.84$ under H_0. It is more difficult to reject H_0 using HL test (logistic regression) due to higher degrees of freedom under chi-square than BD test (frequency table), but here the results agree.

3.

a) Assuming sufficiently large sample size per cell (≥ 5), we use the Row Mean Scores Differ Test.
H_0: The mean agreement scores with the new policy are the same compared by health status (2 level dichotomy) adjusted for region. There is no association between agreement outcome and health status level, adjusted for region. The chi square test statistic with 1 df is 8.258 (p-value=0.0041). We reject H_0 at $\alpha = 0.05$ and conclude that there is a significant location shift for degree of agreement score compared by health status, adjusted for region.

b) The favorable respiratory health status is significantly associated with agreement outcome in the West region and tends to show a higher degree of agreement, MH chi-square test statistic 9.939 (p-value=0.0016 at $\alpha = 0.05$).

c) The Cochran-Armitage test will test for a trend in the proportion of agreement with the new policy across 3 ordered levels by respiratory health status dichotomy.
H_0: The proportion of agreement response is the same for favorable vs. unfavorable respiratory health status.
H_a: The proportion of agreement response has an increasing or decreasing trend for favorable to unfavorable respiratory health status.
The Z statistic is 2.855 (2-sided p-value=0.0043) which is larger than standard normal quantile 1.96 under H_0 at $\alpha = 0.05$. We reject H_0 and conclude that there is an increasing proportion of agreement from disagree, neutral, to agree with favorable to unfavorable respiratory health status.

d) Controlling for region, the Cochran-Armitage test will assess trend in proportion of agreement by respiratory health status dichotomy, favorable vs. unfavorable.
H_0: The proportion of agreement response is the same for respiratory health status, controlling for each region.
H_a: The proportion of agreement response has an increasing or decreasing trend for favorable to unfavorable respiratory health status, controlling for each region.
Controlling for West region, the Z statistic is 3.104 (p-value=0.0019), which is larger than standard normal quantile 1.96 under H_0 at $\alpha = 0.05$. We reject H_0 and conclude that there is an increasing trend in proportion of agreement by respiratory health status, controlling for West region.
Controlling for Mid region, the Z statistic is -0.283 (p-value=0.7773). We fail to reject H_0 at $\alpha = 0.05$ and conclude that there is not a significant trend in agreement proportion by health status, favorable to unfavorable, controlling for Mid region.
Controlling for East region, the Z statistic is 1.676 (p-value=0.0938). We fail to reject H_0 at $\alpha = 0.05$ and conclude that there is not a significant trend in agreement proportion by health status, favorable to unfavorable, controlling for East region.

4.

a) Assumptions for a proportional odds model are a stratified simple random sample, an ordinal multilevel response variable, one or more explanatory variables as main effects (categorical or continuous), and sufficient sample size at each factor level combination (> 5).
The null hypothesis for a proportional odds assumption test is that a common beta is expected as an overall estimate for each beta parameter, meeting the proportional fit requirement. For all k beta parameters,

$$H_0 : \beta_k = \beta$$

$$logit(p_{hik}) = \alpha_1 + \alpha_2 + \beta_1(west) + \beta_2(mild) + \beta_3(mod)$$

Variable	Estimate(SE)	Definition/Parameter Interpretation
$logit(p_{hik})$		cumulative logit probability of agreement for hth region & ith health status
α_1	-0.4084 (0.1605)	log odds of agree vs neutral or disagree for east region and favorable status
α_2	0.8360 (0.1640)	log odds of agree or neutral vs disagree for east region and favorable status
β_1	0.2974 (0.1632)	increment for both logodds due to west region
β_2	-0.7779 (0.1846)	increment for both logodds due to mildly unfavorable status
β_2	-0.2200 (0.2235)	increment for both logodds due to moderately unfavorable status

h=east, west
i=favorable, mildly unfavorable, moderately unfavorable
k=agree (1), agree or neutral (2)

b) The proportional odds assumption test shows chi-square test statistic is 10.029 with 3 df (p-value=0.0183). Since not significant, we reject H_0 at $\alpha = 0.05$ in 4.a) and conclude that the assumption is not met.
Assuming sufficient sample size, the GOF test from the Pearson chi squared statistic is 1.8707 with 7 df (p-value=0.0698), and deviance chi squared statistic is 1.9675 with 7 df (p-value=0.0554).
H_0: The proportional odds logistic model fits the data adequately.
We marginally fail to reject H_0 at $\alpha = 0.05$ and cautiously conclude that the model fits the data but not to a widely significant degree.
Since GOF is close to unacceptable, we can test partial proportional odds model for each main effect and retest proportional odds assumption. First, contrast test to determine if main effects are significant. If significant, include both effects and use proportional odds test to check if the equal slopes (proportional) assumption was met for each main effect in partial odds model. Then, re-fit parameter estimates with

unequal slopes adjustment per effect if needed (i.e. effects that were not significantly proportional) before another GOF test to test fit.

c) The adjusted odds ratio of <u>agree</u> vs neutral or disagree and <u>agree or neutral</u> vs disagree of the new policy for West vs. East region is 1.346 with 95% CI (0.978, 1.854).

```
* 1. MH test and measures of association (OR);
data rash;
    input center $ trmt $ response $ count @@;
    num = (response='y'); *numeric code 1=Y, code 0=N;
    cards;
A active y 52 A active n 26
A placebo y 35 A placebo n 40
B active y 45 B active n 34
B placebo y 40 B placebo n 38
;

proc freq data=rash order=data;
    weight count;
    table center*trmt*response /
        nocol nopct chisq cmh(mf) measures; *MH statistic, OR;
    *per center, Mantel-Haenszel (Mantel-Fleiss criterion) test;
run;

* 2. Fit a main effects logistic regression;
* compare homogeneity of odds ratio using GOF H-L chi square test;
proc logistic data=rash order=data;
    freq count;
    class center(ref='B') trmt(ref='placebo') / param=ref; * assign ref groups and reference cell-coding;
    model response(event='y')= center trmt
        /scale=none aggregate lackfit; *residual chi-squared statistic;
    *exact trmt / estimate=both;
run;


**********************************************
* 3. Row means Score test adjusted for stratified variable;
data policy;
    input region $ health $ response $ count @@;
    cards;
west F disagree 26  west F neutral 14  west F agree 50
west U disagree 58  west U neutral 53  west U agree 45
mid F disagree 15  mid F neutral 15  mid F agree 22
mid U disagree 23  mid U neutral 35  mid U agree 40
east F disagree 33  east F neutral 28  east F agree 37
east U disagree 73  east U neutral 58  east U agree 49
;

proc freq data=policy order=data;
    weight count;
    *table region*health*response / nopercent norow nocol cmh2 scores=rank;
    *table region*health*response / nopercent norow nocol cmh2;
    table region*health*response / nopercent norow nocol cmh2 scores=modridit all;
    table health*response
        / cmh chisq measures trend;
    table region*health*response
        / cmh chisq measures trend;
run;

/*
data policy_num;
    input region $ health $ response count @@;
    cards;
west F 0 26  west F 1 14  west F 2 50
west U 0 58  west U 1 53  west U 2 45
mid F 0 15  mid F 1 15  mid F 2 22
mid U 0 23  mid U 1 35  mid U 2 40
east F 0 33  east F 1 28  east F 2 37
east U 0 73  east U 1 58  east U 2 49
;
```

```
* effect size;
proc glm data=policy_num;
    class region health;
    freq count;
    model response = region health;
    estimate 'diff' health -1 1;
run;
*/

* 4. Fit a proportional odds regression model
for ordinal agreement level without Mid region;
data fit;
    input region $ health $ response $ count @@;
    cards;
west F agree 50  west F neutral 14  west F disagree 26
west M agree 30  west M neutral 35  west M disagree 48
west U agree 15  west U neutral 18  west U disagree 10
east F agree 37  east F neutral 28  east F disagree 33
east M agree 28  east M neutral 34  east M disagree 53
east U agree 21  east U neutral 24  east U disagree 20
;

proc logistic data=fit order=data;
    freq count;
    class region(ref='east') health(ref='F') / param=reference;
    model response(ref='agree') = region health
    / scale=none aggregate; * unequalslopes=health;
run;
```