BIOS 665 Homework 2
Joyce Choe
9/26/2024

1.

a)

$$logit(p_{hi}) = \alpha + \beta_1(ruraltype) + \beta_2(highstress) + \beta_3(mediumstress)$$

Variable	Estimate(SE)	Definition/Parameter Interpretation
$logit(p_{hi})$	n/a	logit probability of favorable response for hth residence type & ith stress
α	0.9071 (0.2122)	reference log odds of urban type and low stress
β_1	-0.0436 (0.1482)	log odds ratio of favorable response comparing rural type to urban type
β_2	-0.6473 (0.2284)	increment to log odds for high stress
β_3	-0.7109 (0.2321)	increment to log odds for medium stress

h=urban, rural
i=low, medium, or high

The 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 housing type*stress combination.

b) The odds ratio estimate for medium stress to low stress on favorable response controlling for residence type is 0.491 with 95% CI: (0.312, 0.774).

c) To calculate the odds ratio estimate:

$$e^{\hat{\beta_3}} = e^{-0.7109} = 0.491 \tag{1}$$

To calculate 95% CI OR, $\alpha = 0.05$:

$$e^{\hat{\beta_3} \pm z_{1-\alpha/2} \cdot SE(\hat{\beta_3})} = e^{-0.7109 \pm 1.96 \cdot (0.2321)} \approx (0.312, 0.774) \tag{2}$$

d) Test whether residence type has no effect on opinion response at 2-sided $\alpha = 0.05$ level.

$$H_0 : \beta_1 = 0$$

Assumptions for the Wald test are a sufficiently large sample with approximately normal distribution and a categorical variable as the explanatory variable. The null hypothesis is that residence type is not associated with opinion response, controlling for stress. The Wald chi-square statistic(Q_W) is 0.0868 (p-value=0.7683). At $\alpha = 0.05$, we fail to reject H_0 and conclude that residence type does not have a significant effect on opinion response based on the model.

e) The predicted probability for unfavorable responses for
a. Urban, low stress individual $\hat{p} = 0.288$
b. Rural, high stress individual probability $\hat{p} = 0.446$

f) To test how close the observed and fitted values from the model are, we can use goodness-of-fit (GOF) tests: Pearson chi-square and Likelihood (deviance) chi-square. The assumptions are sufficiently large samples ($>$ 10 samples per group, 80% of predicted counts at least 5, others at least 2, and no 0 counts.)
H_0: The model fits adequately or the residuals are significantly close to 0.
The degrees of freedom for the Pearson and deviance chi-square test statistic under H_0 is 6-4 = 2. The Pearson chi-square (Q_P) is 1.6623 (p-value = 0.1897) and deviance chi-square statistic (Q_L) is 1.6634 (p-value=0.1895). We fail to reject H_0 at $\alpha = 0.05$ and conclude the model fits the data adequately. Note: If Pearson and deviance test statistics are similar, then the sample size was sufficiently large.

2.

a)

$$logit(p_{hij}) = \alpha + \beta_1(center2) + \beta_2(baselinegood) + \beta_3(treatmenttest)$$

Variable	Definition/Parameter Interpretation
$logit(p_{hij})$	logit probability of favorable outcome for hth center, ith baseline, & jth treatment
α	reference log odds of center 2, poor baseline, & placebo treatment
β_1	increment to log odds for center 1
β_2	increment to log odds for good baseline
β_3	increment to log odds for test treatment

h=1, 2
i=good, poor
j=placebo, test

b)

Parameter	Estimate	Standard Error	Wald Chi Square
α Intercept	-1.056	0.255	17.15
β_1 Center	-0.102	0.103	0.98
β_2 Baseline	0.826	0.324	6.50
β_3 Treatment	3.132	1.020	9.43

c) 95% CI for odds ratio of treatment parameter:

$$e^{\hat{\beta_3} \pm z_{1-\alpha/2} \cdot SE(\hat{\beta_3})} = e^{3.132 \pm 1.96 \cdot (1.020)} \approx (3.10, 169.22) \tag{3}$$

We are 95% confident that the odds of favorable outcome for test treatment is (3.10, 169.22) times higher than the odds of favorable outcome for placebo treatment, after adjusting for center and baseline effects.

d) 95% CI for odds ratio of baseline parameter:

$$e^{\hat{\beta_2} \pm z_{1-\alpha/2} \cdot SE(\hat{\beta_2})} = e^{0.826 \pm 1.96 \cdot (0.324)} \approx (1.21, 4.31) \tag{4}$$

We are 95% confident that the odds of favorable outcome for good baseline is (1.21, 4.31) times higher than the odds of favorable outcome for poor baseline, after adjusting for center and treatment effects.

e) Predicted probabilities for favorable outcome using model for:
i. Center 1, poor baseline, placebo treatment

$$\hat{p} = \frac{e^{\hat{\alpha} + \hat{\beta_1}}}{1 + e^{\hat{\alpha} + \hat{\beta_1}}} \approx 0.239 \tag{5}$$

ii. Center 2, good baseline, test treatment

$$\hat{p} = \frac{e^{\hat{\alpha + \beta_1} + \hat{\beta_2} + \hat{\beta_3}}}{1 + e^{\hat{\alpha + \beta_1} + \hat{\beta_2} + \hat{\beta_3}}} \approx 0.943 \tag{6}$$

3.

a)

$$logit(p_i) = \alpha + \beta_1(dose10mg) + \beta_2(dose100mg)$$

Variable	Definition/Parameter Interpretation
$logit(p_i)$	logit probability of favorable response for ith dose
α	reference log odds of dose 1 mg
β_1	increment to log odds for dose 10 mg
β_2	increment to log odds for dose 100 mg

i=1 mg, 10 mg, 100 mg

Parameter	Estimate	Standard Error	Wald Chi Square
α Intercept	-0.5878	0.2494	5.55
β_1 Dose 10 mg vs. 1 mg	0.5878	0.3455	2.89
β_2 Dose 100 mg vs. 1 mg	1.9741	0.3892	25.72

b)

$$logit(p_i) = \alpha + \beta_1 log_{10}(dose)$$

Variable	Definition/Parameter Interpretation
$logit(p_i)$	logit probability of favorable response for continuous $log_{10}(dose)$
α	log odds of favorable outcome when dose=$log_{10}(1) = 0$ mg
β_1	log odds of favorable outcome for every $log_{10}(dose)$=10-fold mg

dose=[1, 100]

Parameter	Estimate	Standard Error	Wald Chi Square
α Intercept	-0.722	0.2318	9.70
β_1 Dose	0.9646	0.1903	25.70

c) The 99% CI for $\hat{\beta_1}$ at α=0.01 is

$$\hat{\beta_1} \pm z_{1-\alpha/2} \cdot SE(\hat{\beta_1}) = 0.9646 \pm 2.5758 \cdot 0.1903 \approx (0.474, 1.455) \tag{7}$$

We are 99% confident that the log-odds of favorable outcome for every 10-fold mg increment is (0.474, 1.455), which does not contain the null value 0.

To test GOF for a logistic model with a continuous explanatory variable, we use the residual chi-square test via forward selection (intercept vs. 1 parameter fitted model). Assumptions for this test are sufficient sample size.

H_0: The model fits adequately.

The chi-square residual score statistic 1 d.f. (Q_{RS}) = 27.76 measures the model residuals linear association with the explanatory variable. The p-value is $<$0.0001, so we reject H_0 and conclude the model does not fit adequately. Note: test may not be reliable with small sample sizes due to biased high power (type I error) and result in falsely poor model fit conclusion.