EPI 202 Lab 5 Practice Problem Solutions

1) The MI onset Study was a multicenter epidemiologic study of the determinants of myocardial infarction (MI) onset and prognosis. In this study, patients with acute myocardial infarction were recruited from 64 medical centers across the US and interviewed about lifestyle and health characteristics within an average of 3 days of hospitalization. You are interested in whether physical activity is associated with the 10-year cumulative incidence of death following MI among 3845 individuals. The variable pa_cat takes the value 1 for individuals who exercise at least once a week (physically active) and the value 0 for individuals who exercise less than once a week (physically inactive). There was no loss to follow-up in this study. The following data were observed:

	pa_cat=1	pa_cat=0
Dead	76	1066
Survived	453	2250

a) Calculate the cumulative incidence of death within 10 years among the physically active and physically inactive individuals separately.

Physically active CI = 76/(76+453) = 0.144 by the end of 10 years

Physically inactive CI = 1066/(1066+2250) = 0.321 by the end of 10 years

b) Calculate the odds of death within 10 years among the physically active and physically inactive individuals separately.

Physically active odds = 76/453 = 0.168 within 10 years

Physically inactive odds = 1066/2250 = 0.474 within 10 years

c) Calculate the CIR and OR for the association between physical activity and death within 10 years. How do these measures compare?

$$CIR = \frac{76/(76+453)}{1066/(1066+2250)} = 0.45 \; by \; the \; end \; of \; 10 \; years$$

$$OR = \frac{76/453}{1066/2250} = 0.35 \; by \; the \; end \; of \; 10 \; years$$

The OR is farther from the null than the CIR.

You decide to fit the following logistic regression model to the above data:

$\text{logit}(p) = \beta_0 + \beta_1 * \text{pa\_cat}$ where p=probability of death within 10 years

d) What values will you find for β_0 and β_1? Interpret these values.

$\beta_0 = \ln(1066/2250) = -0.747$

The log odds of death within 10 years is -0.747 among physically inactive individuals, assuming no selection bias or information bias.

$\beta_1 = \ln(76/453) - \ln(1066/2250) = \ln((76/453) / (1066/2250)) = -1.038$

The log odds of death within 10 years is 1.038 lower among physically active individuals than among physically inactive individuals, assuming no confounding, selection bias, or information bias. β_1 can also be interpreted as the log OR comparing the odds of death within 10 years among physically active individuals to among physical inactive individuals.

You fit the regression model and receive the following output:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.74702    0.03718 -20.091  < 2e-16 ***
pa_cat      -1.03814    0.12941  -8.022 1.04e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e) Use the model results to calculate the OR and CIR for the association between physical activity and death within 10 years. How do these measures compare to those you calculated in question 3c?

$$OR = e^{\beta 1} = e^{-1.038} = 0.35 \text{ by the end of } 10 \text{ years}$$

$$CIR = \frac{e^{\beta 0+\beta 1}/(1+e^{\beta 0+\beta 1})}{e^{\beta 0}/(1+e^{\beta 0})} = \frac{e^{-0.747-1.038}/(1+e^{-0.747-1.038})}{e^{-0.747}/(1+e^{-0.747})} = 0.45 \text{ by the end of } 10 \text{ years}$$

The estimated OR and CIR from the model are identical to those calculated in Question 3c.

You are concerned about confounding by age and decide to stratify on age in 3 categories (age_cat=1: <50 years, age_cat=2: 50-64 years, age_cat=3: 65+ years).

age_cat=1

	pa_cat=1	pa_cat=0
Dead	10	70
Survive	177	502
Total	187	572

age_cat=2

	pa_cat=1	pa_cat=0
Dead	28	288
Survive	198	924
Total	226	1212

age_cat=3

	pa_cat=1	pa_cat=0
Dead	38	708
Survive	78	824
Total	116	1532

f) Calculate the Mantel-Haenszel OR for the association between physical activity and death within 10 years, adjusting for age category. Interpret this OR.

Note: this estimate can be obtained using the case-control tab in the EPI 202 calculator

$$OR_{MH} = \frac{\sum_{i=1}^{I} \frac{a_i(N_{0i} - b_i)}{T_i}}{\sum_{i=1}^{I} \frac{b_i(N_{1i} - a_i)}{T_i}} = \frac{\frac{10 * 502}{759} + \frac{28 * 924}{1438} + \frac{38 * 824}{1648}}{\frac{70 * 177}{759} + \frac{288 * 198}{1438} + \frac{708 * 78}{1648}} = \frac{43.603}{89.489}$$
$$= 0.49 \; by \; the \; end \; of \; 10 \; years$$

After adjusting for age, the odds of death within 10 years among physically active individuals were 0.49 times the odds among physically inactive individuals, assuming no residual confounding by age, no confounding by other variables, no selection bias, and no information bias.

You decide to fit the following logistic regression model to the above data to adjust for age:

logit(p) = β_0 + β_1*pa_cat + β_2*(age_cat=2) + β_3*(age_cat=3)

You fit the regression model and receive the following output:

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          -1.9974     0.1204 -16.594  < 2e-16 ***
pa_cat               -0.7130     0.1346  -5.297 1.18e-07 ***
as.factor(age_cat)2   0.8237     0.1349   6.107 1.02e-09 ***
as.factor(age_cat)3   1.8546     0.1292  14.357  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

g) Use the model results to estimate the age-adjusted OR for the association between physical activity and death within 10 years. Interpret this OR. Compare this OR to the one you calculated in the prior question.

$$OR = e^{\beta 1} = e^{-0.713} = 0.49 \; by \; the \; end \; of \; 10 \; years$$

After adjusting for age, the odds of death within 10 years among physically active individuals were 0.49 times the odds among physically inactive individuals, assuming no residual confounding by age, no confounding by other variables, no selection bias, and no information bias.

The OR is identical to the OR computed in Question 3f.

h) You are interested in testing whether the association between physical activity and death within 10 years differs for females and males on the ratio scale, after adjusting for age. Write out the model you would fit to address this question. Provide interpretations of the beta coefficients pertaining to physical activity and sex.

$\text{logit}(p) = \beta_0 + \beta_1 * \text{pa\_cat} + \beta_2 * \text{female} + \beta_3 * \text{female} * \text{pa\_cat} + \beta_4 * (\text{age\_cat}=2) + \beta_5 * (\text{age\_cat}=3)$

β_1 is the log OR of death within 10 years comparing physically active males to physically inactive males, holding age constant

β_2 is the log OR of death within 10 years comparing physically inactive females to physically inactive males, holding age constant

β_3 is the difference between females and males in log ORs for physical activity and death within 10 years, holding age constant

i) You fit this model and obtain the following output. Is there evidence of effect measure modification by sex on the multiplicative scale? State the null and alternative hypotheses of the test of homogeneity of the odds ratio. Report and interpret the p-value for this test.

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         -2.00014    0.12206 -16.386  < 2e-16 ***
pa_cat              -0.72529    0.15095  -4.805 1.55e-06 ***
female               0.01193    0.08218   0.145    0.885
as.factor(age_cat)2  0.82363    0.13490   6.106 1.02e-09 ***
as.factor(age_cat)3  1.85157    0.13042  14.197  < 2e-16 ***
pa_cat:female        0.07537    0.33618   0.224    0.823
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

H_0: There is no effect measure modification by sex on the multiplicative scale for the association between physical activity and death within 10 years, after adjusting for age ($OR_{female}=OR_{male} \Leftrightarrow \beta_3=0$)

H_A: There is effect measure modification by sex on the multiplicative scale for the association between physical activity and death within 10 years, after adjusting for age ($OR_{female} \neq OR_{male} \Leftrightarrow \beta_3 \neq 0$)

P-homogeneity=0.823 from Wald test

The probability of observing odds ratios as or more different than the observed odds ratios for females and males is 0.823, if the null were true, assuming no residual confounding by age, no confounding by other variables, no selection bias, and no information bias.

We fail to reject the null hypothesis at a 0.05 level of significance and conclude that there is no evidence of effect measure modification by sex on the multiplicative scale for the association between physical activity and death within 10 years, assuming no residual confounding by age, no confounding by other variables, no selection bias, and no information bias.