2017 Qual - Problem 2

Elaine Kaye Kearney

7/6/2020

2017 Qual - Problem 2

a)

Table 1: Descriptive Statistics for Baseline Variables

	Overall Sample ($n = 204$)	Control ($n = 103$)	Intervention ($n = 101$)
Age (years)	9.05 (0.971)	9.02 (0.918)	9.08 (1.026)
Baseline Obesity Status			
Not Obese	71 (34.8%)	39 (37.9%)	32 (31.7%)
Obese	133 (65.2%)	64 (62.1%)	69 (68.3%)
Sex			
Male	91 (44.6%)	50 (48.5%)	41 (40.6%)
Female	113 (55.4%)	53 (51.5%)	60 (59.4%)
Race			
African-American	68 (33.3%)	33 (32.0%)	35 (34.65%)
Caucasian	136 (66.7%)	70 (68.0%)	66 (65.35%)
Parent BMI			
$\geq 30\ kg/m^2$	167 (81.9%)	92 (89.3%)	75 (74.3%)
$< 30\ kg/m^2$	37 (18.1%)	11 (10.7%)	26 (25.7%)

Continuous variables are reported as mean (SD). Categorical variables are reported as N(%).

Table 2: Descriptive Statistics for Follow-up Visit Obesity Status

	Control ($n = 103$)	Intervention ($n = 101$)
Follow-up Obesity Status		
Missing	24 (23.3%)	14 (13.9%)
Not Obese	40 (38.8%)	46 (45.5%)
Obese	39 (37.9%)	41 (40.6%)

Categorical variables are reported as N(%).

b) Extent of missingness

Based on Table 2, we see that about 23% of the outcome variable $Y1$ is missing in the control group and 14% of the outcome variable $Y1$ is missing in the control group. Moreover, 63% of the missing values were from patients in control group. Therefore, we do see that missingness varies by intervention group. Thus, missingness do seem to have a relationship with intervention group, which must be accounted for in the analysis.

Chi-square test of 2x2 table (marginal result)

c)

Let θ_{ijklmn} be the probability of being obese at follow-up for the ith intervention group ($i = 1$ for control, 2 for intervention), jth baseline obesity status ($j = 1$ for not obese, 2 for obese), kth sex ($k = 1$ for male, 2 for

female), lth race ($l = 1$ for Caucasian, 2 for African-American), mth parent weight ($m = 1$ for parent BMI $< 30kg/m^2$, 2 o.w.), and nth age in years.

$$logit(\theta_{ijklmn}) = \beta_0 + \sum_{u=1}^{6} \beta_u X_{ijklmnu}$$

where

$$X_{ijklmn1} = 1 \text{ if intervention, 0 o.w.}$$

$$X_{ijklmn2} = 1 \text{ if obese at baseline visit, 0 o.w.}$$

$$X_{ijklmn3} = 1 \text{ if female, 0 o.w.}$$

$$X_{ijklmn4} = 1 \text{ if African-American, 0 o.w.}$$

$$X_{ijklmn5} = 1 \text{ if parent BMI} \geq 30kg/m^2, \text{ 0 o.w.}$$

$$X_{ijklmn6} = \text{age (years)}$$

and

β_0 = log odds of being obese at follow-up for children who are in the control group, not obese at baseline, male, Caucasian, have parents BMI $< 30kg/m^2$, and with age of 0 years.

β_1 = increment in log odds of being obese at follow-up for children in the intervention group.

β_2 = increment in log odds of being obese at follow-up for children who are obese at baseline.

β_3 = increment in log odds of being obese at follow-up for children who are female.

β_4 = increment in log odds of being obese at follow-up for children who are African-American.

β_5 = increment in log odds of being obese at follow-up for children who have parents with BMI $\geq 30kg/m^2$.

β_6 = increment in log odds of being obese at follow-up for each one year in age.

Table 3: Parameter Estimates and Standard Error Estimates for Logistic Regression Model

Parameter	Parameter Estimate	Standard Error Estimate
β_0	-1.24	2.240
β_1	-0.92	0.488
β_2	4.09	0.582
β_3	-0.80	0.484
β_4	-0.23	0.464
β_5	-0.34	0.546
β_6	-0.01	0.235

Children in the intervention group had 0.4 times the odds of being obese at the follow-up visit (in the complete case analysis) than for children in the control group. This odds ratio has a corresponding 95% confidence interval of (0.152, 1.033), which includes the null value of 1. Therefore, at the alpha = 0.05 significance level, we do not reject the null hypothesis that that the risk of being obese at follow-up is equal in the groups. However, based on the magnitude of this odds ratio, we do have evidence that the intervention group has lower risk of being obese at follow-up, compared to the control group. Perhaps analyzing more of the data will provide more information.

d)

Table 4: Parameter Estimates and Standard Error Estimates for Logistic Regression Model after performing Multiple Imputation

Parameter	Parameter Estimate	Standard Error Estimate
β_0	-1.18	2.197
β_1	-0.95	0.491
β_2	4.15	0.601
β_3	-0.80	0.531
β_4	-0.23	0.475
β_5	-0.32	0.559
β_6	-0.02	0.230

Children in the intervention group had $exp(-0.95) = 0.39$ times the odds of being obese at the follow-up visit (in the multiple imputation analysis) than for children in the control group. This odds ratio has a corresponding 95% confidence interval of (0.147, 1.013), which includes the null value of 1. Therefore, at the alpha = 0.05 significance level, we do not reject the null hypothesis that that the risk of being obese at follow-up is equal in the groups. However, based on the magnitude of this odds ratio, we do have evidence that the intervention group has lower risk of being obese at follow-up, compared to the control group. The results after performing multiple imputation are very similar to those using the complete case analysis.

MI analysis steps:

1. Missing data are filled in m times, which I set to be 25 in this case, to create m completed data sets.

Let Z_{i0} denote the vector composed of Y_{i0}, which is $Y0$ for individual i and the baseline covariates for individual i, X_i. The model used to impute Y_{i1}, which is $Y1$ for individual i is

$$logit\{Pr(Y_{i1} = 1|Y_{i0}, X_i)\} = Z'_{i0}\gamma$$

where γ is the vector of regression parameters relating Y_{i1} to Y_{i0} and the baseline covariates.

2. m completed data sets are analyzed using the appropriate statistical method for the scientific question of interest. Here, I fit a logistic model like I fit in part c for each of the m completed data sets.

3. Results from m analyses are combined to get final estimates. Let $\hat{\beta}^{(k)}$ be the parameter estimates based on fitting the model for the kth completed data set. The final multiple imputation estimate of the parameter estimates is

$$\hat{\beta} = \frac{1}{m}\sum_{k=1}^{m}\hat{\beta}^{(k)}$$

and the estimated covariance of $\hat{\beta}$ is

$$\hat{Cov}(\hat{\beta}) = W + \big(1 + \frac{1}{m}\big)B$$
$$W = \frac{1}{m}\sum_{k=1}^{m}\hat{Cov}(\hat{\beta}^{(k)})$$
$$B = \frac{1}{m-1}\sum_{k=1}^{m}\big(\hat{\beta}^{(k)} - \hat{\beta}\big)\big(\hat{\beta}^{(k)} - \hat{\beta}\big)'$$

and the confidence intervals follow from point estimates and estimated covariance.

e)

The analyses in parts C are valid for MCAR. The analyses in parts D are valid for MCAR, MAR. Seems that intervention group is strongly correlated with missingness.