B. **In the following table calculate the marginal table that excludes the variable rehabilitation services**.  Then calculate the impact of skilled nursing facility on above average cost, conditional on severity of illness and the surgeon involved.  To do so, create strata for the combination of severity of illness and surgeon and calculate the impact of skilled nursing facility on cost within each strata.  Redo, this time not conditional on surgeon involved.  What does this data tell you about impact of skilled nursing facility on the bundled cost.

|  |  |  |
| --- | --- | --- |
|   | **N**: Skilled Nursing Facility A | Skilled Nursing Facility B |
| **S**: High Severity | Low Severity | High Severity | Low Severity |
| **O**: Orthopedic Surgeon | **R**: Rehab Services | **A**: Above Avg Cost | Below Avg Cost | Above Avg Cost | Below Avg Cost | Above Avg Cost | Below Avg Cost | Above Avg Cost | Below Avg Cost |
| Joe | Yes | 405 | 268 | 453 | 228 | 23 | 23 | 30 | 19 |
| Joe | No | 13 | 218 | 28 | 201 | 2 | 19 | 1 | 18 |
| Jim | Yes | 1 | 17 | 1 | 17 | 0 | 1 | 1 | 8 |
| Jim | No | 1 | 117 | 1 | 133 | 0 | 12 | 0 | 17 |

* **Calculate the marginal table that excludes the variable rehabilitation services**

> ##Create marginal table with frequencies summed over Rehab services

> str(data)

'data.frame': 32 obs. of 6 variables:

 $ O : num 2 2 2 2 2 2 2 2 2 2 ...

 $ R : num 2 2 2 2 2 2 2 2 1 1 ...

 $ N : num 1 1 1 1 2 2 2 2 1 1 ...

 $ S : num 1 1 2 2 1 1 2 2 1 1 ...

 $ Cost : num 1 2 1 2 1 2 1 2 1 2 ...

 $ Observed: int 405 268 453 228 23 23 30 19 13 218 ...

> head(data)

 O R N S Cost Observed

1 2 2 1 1 1 405

2 2 2 1 1 2 268

3 2 2 1 2 1 453

4 2 2 1 2 2 228

5 2 2 2 1 1 23

6 2 2 2 1 2 23

> ###collapse over Rehab services

> data.ONCS.df<-aggregate(Observed~O+N+S+Cost, data=data, FUN=sum) #sums over rehab services.

> data.ONCS.df

 O N S Cost Observed

1 1 1 1 1 2

2 2 1 1 1 418

3 1 2 1 1 0

4 2 2 1 1 25

5 1 1 2 1 2

6 2 1 2 1 481

7 1 2 2 1 1

8 2 2 2 1 31

9 1 1 1 2 134

10 2 1 1 2 486

11 1 2 1 2 13

12 2 2 1 2 42

13 1 1 2 2 150

14 2 1 2 2 429

15 1 2 2 2 25

16 2 2 2 2 37

* **Then calculate the impact of skilled nursing facility on above average cost, conditional on severity of illness and the surgeon involved.  To do so, create strata for the combination of severity of illness and surgeon and calculate the impact of skilled nursing facility on cost within each strata.**

|  |
| --- |
| > #partial counts of Skilled nursing facility on Cost conditional on > #severity and surgeon involved.(Stratified data)> pCounts.df = xtabs(Observed~ N+ Cost + S+O, data.ONCS.df)> pCounts.df, , S = 1, O = 1 CostN 1 2 1 2 134 2 0 13, , S = 2, O = 1 CostN 1 2 1 2 150 2 1 25, , S = 1, O = 2 CostN 1 2 1 418 486 2 25 42, , S = 2, O = 2 CostN 1 2 1 481 429 2 31 37> #convert data frame to a multidimensional array using array function> #array function indicates 2 rows, 2 columns and 4 levels.> pCounts<-array(pCounts.df, c(2,2,4))> pCounts, , 1 [,1] [,2][1,] 2 134[2,] 0 13, , 2 [,1] [,2][1,] 2 150[2,] 1 25, , 3 [,1] [,2][1,] 418 486[2,] 25 42, , 4 [,1] [,2][1,] 481 429[2,] 31 37* **Calculate the impact of skilled nursing facility on cost within each strata.**

*Cochran-Mantel Test to test for conditional independence:***Null Hypothesis of conditional independence**: Skilled nursing facility and Costs are independent given severity of illness and orthopedic surgeon involved. This is equivalent to the statement that all conditional odds ratios given the levels of orthopedic surgeon and severity of illness of the patient is equal to 1.> mantelhaen.test(pCounts) Mantel-Haenszel chi-squared test with continuity correctiondata: pCountsMantel-Haenszel X-squared = 2.7062, df = 1, p-value = 0.09996alternative hypothesis: true common odds ratio is not equal to 195 percent confidence interval: 0.9586451 1.9373978sample estimates:common odds ratio  1.362819 **Inference**: Since the p-value=0.0996 is more than 0.05, it indicates that conditional independence model is a good fit for the data (that is we cannot reject null hypothesis) and that nursing facility and cost are conditionally independent of each other while orthopedic surgeon involved and severity of illness are held constant.***Breslow-Day test to test for homogeneity****:* > breslowday.test(pCounts)Breslow and Day test (with Tarone correction):Breslow-Day X-squared = 1.695941 Breslow-Day-Tarone X-squared = 1.695816 Test for test of a common OR: p-value = 0.6378644 **Inference of Breslow-Day Test**: *Null hypothesis*- The common odds ratio of nursing facility and costs across each level of surgeon involved and severity of illness is identical (which implies nursing facility and costs are independent across each level of severity and surgeon involved). The Breslow-Day statistic is 1.69, p-value = 0.637. Since the p-value=0.637 is more than 0.05, it indicates that homogeneous independence model holds true for the data (that is we cannot reject null hypothesis) and that nursing facility and cost are independent of each other across all levels of surgeon involved and severity of illness.* **Redo, this time not conditional on surgeon involved**

> #partial counts of Skilled nursing facility on Cost conditional on > #severity but not surgeon involved this time.(Stratified data)> pCountsNoS.df = xtabs(Observed~ N+ Cost + S, data.ONCS.df)> pCountsNoS.df, , S = 1 CostN 1 2 1 420 620 2 25 55, , S = 2 CostN 1 2 1 483 579 2 32 62* **What does this data tell you about impact of skilled nursing facility on the bundled cost.**

*Cochran-Mantel Haeszen test:*> mantelhaen.test(pCountsNoS) Mantel-Haenszel chi-squared test with continuity correctiondata: pCountsNoSMantel-Haenszel X-squared = 6.6719, df = 1, p-value = 0.009795alternative hypothesis: true common odds ratio is not equal to 195 percent confidence interval: 1.121729 2.163680sample estimates:common odds ratio  1.557903 **Null Hypothesis of conditional independence**: Skilled nursing facility and Costs are independent given severity of illness. This is equivalent to the statement that all conditional odds ratios given the levels ofseverity of illness of the patient is equal to 1.**Inference**: Since the p-value=0.0097 is less than 0.05, it indicates that conditional independence model is not a good fit for the data (that is we can reject null hypothesis) and that nursing facility and cost are significantly associated with each other while severity of illness is held constant. |