**Question 2**: In this problem, we approach multi-level modeling through SQL.  In particular, we address question 1 but this time we use SQL instead of regression.  In particular, for each 2 medical centers follow these steps:

1. Select the data for the medical center.
2. Use SQL for Level 1 model: **SQL code emailed**
   * In temporary table #Cancer, calculate the probability of mortality for patients with cancer at different combination of comorbidities (called strata).  Call the probability "Probability of Strata + Cancer".  You can accomplish this by selecting the probability where lung cancer =1 and grouping by concatenation of all comorbidities.  To concatenate string variables you can use the + sign.
   * In temporary table #NoCancer, calculate the probability of mortality for patients without cancer at different strata.  Call the probability "Probability of Strata + No Cancer".
   * Join the tables #Cancer and #NoCancer on Strata.
3. Use Regression for Level 2 model:
   * Regress the "Probability of Strata + Cancer" on "Probability of Strata + No Cancer".

|  |  |  |
| --- | --- | --- |
|  |  | *Coefficients* |
| **Center 1** | Intercept | **0.393279** |
|  |  |  |
|  |  |  |
|  |  |  |
| **Center 2** |  | *Coefficients* |
|  | Intercept | **0.564593** |
|  |  |  |
|  |  |
|  |  |
| **Center 3** |  | *Coefficients* |
|  | Intercept | **0.495125** |
|  |  |  |
| **Center 4** |  | *Coefficients* |
|  | Intercept | **0.526344** |
|  |  |  |
| **Center 5** |  | *Coefficients* |
|  | Intercept | **0.523612** |
|  |  |  |
|  |  |  |
| **Center 6** |  | *Coefficients* |
|  | Intercept | **0.45297** |
|  |  |  |
| **Center 7** |  | *Coefficients* |
|  | Intercept | **0.429885** |

1. Use the intercept as the impact of cancer in the Medical Center.
2. Examine mortality rates at the center as a function of center's distance to referral source and satisfaction of the patients in the center.

|  |  |  |
| --- | --- | --- |
| **Medical Center** | **Distance** | **% Satisfied** |
| **Center1** | 50 | 79 |
| **Center2** | 80 | 82 |
| **Center3** | 70 | 80 |
| **Center4** | 70 | 79 |
| **Center5** | 80 | 79 |
| **Center6** | 70 | 83 |
| **Center7** | 80 | 81 |

R CODE

##Copied and Pasted each intercept for center from excel doc into a value

intcp1 <- 0.393279012

intcp2 <- 0.564592731

intcp3 <- 0.49512538

intcp4 <- 0.526343725

intcp5 <- 0.523611954

intcp6 <- 0.452969605

intcp7 <- 0.429884674

##Bind all into ONE

center.incpts <- rbind (intcp1, intcp2, intcp3, intcp4, intcp5, intcp6, intcp7)

##Include Data on Distance and %Satisfied

Distance <- rbind(50, 80, 70, 70, 80, 70, 80)

Satisfied <- rbind (79, 82, 80, 79, 79, 83, 81)

#Regress

output <- lm(center.incpts ~ Distance + Satisfied)

summary (output)

**Call:**

**lm(formula = center.incpts ~ Distance + Satisfied)**

**Residuals:**

**intcp1 intcp2 intcp3 intcp4 intcp5 intcp6 intcp7**

**-0.015334 0.058484 0.013908 0.037518 -0.005320 -0.005424 -0.083832**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 0.809108 1.154938 0.701 0.522**

**Distance 0.004011 0.002259 1.775 0.151**

**Satisfied -0.007608 0.014924 -0.510 0.637**

**Residual standard error: 0.05555 on 4 degrees of freedom**

**Multiple R-squared: 0.4416, Adjusted R-squared: 0.1624**

**F-statistic: 1.582 on 2 and 4 DF, p-value: 0.3118**