**Question 1:**The following data provides information on long cancer mortality at different medical centers on a large set of patients.  Medical centers claim that they have higher or lower risk for lung cancer mortality because of patients' comorbidities.  Use multi-level analysis to establish if any particular medical center has a statistically significant excess mortality rate beyond what could be expected from patients' comorbidities:

1. **Identify the intercept for each medical center in a regression of 6-month mortality on patients comorbidities**

table(db$Medical.Center)

 Center 1 Center1 Center2 Center3 Center4 Center5 Center6 Center

 0 2320 2308 2779 4471 2645 2320 2782

sub1 <- subset(db, Medical.Center == "Center1")

sub2 <- subset(db, Medical.Center == "Center2")

sub3 <- subset(db, Medical.Center == "Center3")

sub4 <- subset(db, Medical.Center == "Center4")

sub5 <- subset(db, Medical.Center == "Center5")

Sub6 <- subset(db, Medical.Center == "Center6")

sub7 <- subset(db, Medical.Center == "Center7")

c1 <- glm(Dead.in.6.Months ~ ., data = sub1[, -c(1, 3, 4, 5, 6, 7, 8, 9, 10, 11)], family = "binomial" )

c2 <- glm(Dead.in.6.Months ~ ., data = sub2[, -c(1, 3, 4, 5, 6, 7, 8, 9, 10, 11)], family = "binomial" )

c3 <- glm(Dead.in.6.Months ~ ., data = sub3[, -c(1, 3, 4, 5, 6, 7, 8, 9, 10, 11)], family = "binomial" )

c4 <- glm(Dead.in.6.Months ~ ., data = sub4[, -c(1, 3, 4, 5, 6, 7, 8, 9, 10, 11)], family = "binomial" )

c5 <- glm(Dead.in.6.Months ~ ., data = sub5[, -c(1, 3, 4, 5, 6, 7, 8, 9, 10, 11)], family = "binomial" )

c6 <- glm(Dead.in.6.Months ~ ., data = sub6[, -c(1, 3, 4, 5, 6, 7, 8, 9, 10, 11)], family = "binomial" )

c7 <- glm(Dead.in.6.Months ~ ., data = sub7[, -c(1, 3, 4, 5, 6, 7, 8, 9, 10, 11)], family = "binomial" )

intcp1 <- coef(c1)[1]

intcp1

(Intercept)

 0.5790057

intcp2 <- coef(c2)[1]

intcp2

(Intercept)

 0.5588465

intcp3 <- coef(c3)[1]

intcp3

(Intercept)

 0.5194873

intcp4 <- coef(c4)[1]

intcp4

(Intercept)

 0.6292011

intcp5 <- coef(c5)[1]

intcp5

(Intercept)

 0.5556206

 intcp6 <- coef(c6)[1]

 intcp6

(Intercept)

 0.7727665

 intcp7 <- coef(c7)[1]

 intcp7

(Intercept)

 0.5389053

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

center.incpts

 (Intercept)

intcp1 0.5790057

intcp2 0.5588465

intcp3 0.5194873

intcp4 0.6292011

intcp5 0.5556206

intcp6 0.7727665

intcp7 0.5389053

1. **Regress mortality at the centers (intercepts estimated in previous step) on patients satisfaction with the services and travel distance from referral to 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 |

> intcp1 <- 0.5790057

> intcp2 <- 0.5588465

> intcp3 <- 0.5194873

> intcp4 <- 0.6292011

> intcp5 <- 0.5556206

> intcp6 <- 0.7727665

> intcp7 <- 0.5389053

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

> output2 <- 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

1. **Report the meaning of these two related regressions**

In the first portion of the problem we used the comorbidities to predict if a patient would be dead in 6 months. In doing so our estimation was made at the patient level. In the second regression we took the intercepts to regress on mortality on satisfaction and distance. The output showed that neither of the two were statistically significant therefore not good indicators of dead in 6 months.