**Chapter 14**

**Multilevel Modeling: Intercept Regression**

# [H1] Learning Objectives

1. Examine group- and practice-level relationships after controlling for patient-level differences

# [H1] Key Concepts

* Multilevel modeling
* Micro level (or patient level)
* Macro level (or practice level)
* Intercept regression

# [H1] Chapter at a Glance

Multilevel regressions allow the analyst to create two sets of regression models—one for the individual data (patients in a practice) and another for group-level data (e.g., differences between practices). The separation of the modeling efforts allows the analyst to see what factors affect clinical practices independent of patient differences. This chapter introduces the concept of multilevel regression and shows an example of its application.

# [H1] Increasing Use

You might read about multilevel modeling in human resource courses. Multilevel regression was useful in understanding whether educational programs reduce nurse turnover, independent of the characteristics of the nurses involved (Rondeau, Williams, and Wagar 2009). These methods were used to examine job stress as a function of sleep problems regardless of the amount of shift work or of employee characteristics (Lin et al. 2014). You might see multilevel modeling in economic and finance courses. Multilevel regression, also called hierarchical regression, was used to examine variation in long-term care use independent of patient characteristics (Kahn et al. 2012). Hospital managers can use it to decide which nursing home to contract with, which gives them a chance to develop an effective response to reimbursement methods that hold them responsible for outcome in nursing homes. Hierarchical regression has been used heavily in examination of quality of care and in benchmarking. For example, the factors that affect quality of life across patients have been determined through multilevel modeling (Chou, Ma, and Yang 2014). Socioeconomic effects on mortality rate disappear if patients select better-quality hospitals (Gabriel et al. 2018). Hierarchical regression has been used to measure the 30-day morality rate of hospitals caring for pneumonia patients (Bratzler et al. 2011). Multilevel modeling has been used extensively in forecasting. For example, it was used to see whether there is a relationship between impulsivity in preschool children and their body mass index 30 years later (Schlam et al. 2013). The investigators controlled for many individual differences. Despite these differences, longer delay of gratification among four-year-olds was associated with a lower index.

If you are preparing for a career in insurance, you can use multilevel regression to remove individual differences and understand factors that affect satisfaction with managed care plans (Keenan et al. 2009). In courses on environmental risk, one can use multilevel modeling to understand the influence of environment on health outcomes. For example, Svensson and colleagues (2017) used multilevel regression to consider whether city design and neighborhood layouts affect exercise patterns. . In courses on cost of care and accounting, one may run into multilevel regression in studies of factors that can reduce cost. For example, multilevel regression has been used to examine the impact of formulary restrictions on the use of generic drugs (Shen et al. 2017). The point is that multilevel regression is increasingly used to control for individual patient differences while examining factors that affect practice patterns or healthcare organizations.

# [H1] Ideas behind Multilevel Modeling

An accessible and easy to understand review of multilevel modeling is provided in Peugh (2010). In many healthcare studies, we have two sets of variables: macrolevel factors (at the level of the service or practice level) and microlevel characteristics (at the level of the patient). Macrolevel factors focus on organization of care: people who provide care or organizations that do so. They may describe a practice or may refer to a hospital. Patient-level variables describe individuals who received care. In a multilevel modeling, data are modeled in two separate regressions. One focuses on the impact of patient-level data on the outcomes. The other regression examines the impact of practice- or hospital-level features on the variance in outcome left unexplained by the first regression.

 There are two types of multilevel regressions: intercept models and coefficient multilevel models. In this chapter, we focus on intercept regressions. An intercept model can be described with the following two equations. The first focuses on microlevel relationships for the outcome and *r* patient characteristics$ X\_{1},X\_{2},…,X\_{r}$:

$Y\_{ij}=β\_{0j}+β\_{1}X\_{1ij}+β\_{2}X\_{2ij}+…+β\_{r}X\_{rij}+e\_{ij}$.

In this equation, $Y\_{ij} $is the outcome of the *i*th patient in the *j*th practice; $β\_{0j} $is the overall constant (intercept) for each practice; variables $ β\_{1},β\_{2},…,β\_{r} $describe estimated parameters for patient characteristics. Note that $β\_{1}X\_{1ij}+β\_{2}X\_{2ij}+…+β\_{r}X\_{rij} $is the total effect of patient-level characteristics on practice outcomes. $e\_{ij }$ is the random variation in outcome at the individual level. In this regression, all variables are measured across all patients in all practices, hence the indexes *i* and *j* indicating that data are for patient *i* in practice *j*. If the relationship between outcomes and patient features is not linear, a nonlinear model must be fitted.

From this equation, we only use the intercepts, $β\_{0j}$, in subsequent analysis; everything else is ignored. The intercepts do not include any of the patient characteristics and thus describe the average outcome at the practice after removing the effects associated with patient characteristics. This is not the same as saying that the intercept is independent of the predictors. Different sets of predictors result in different intercepts, so intercepts are never independent of predictors; they are a function of predictors used. All we know, each practice has a different average outcome at the end of a microlevel regression.

 The next regression is to examine variations in the calculated average outcome at each practice. We want to explain these variations as a function of *q* practice level features$ Z\_{1},Z\_{2},…,Z\_{q}$ with the equation

$β\_{0j}=γ\_{0}+γ\_{1}Z\_{1j}+γ\_{2}Z\_{2j}+…+γ\_{q}Z\_{qj}+u\_{j}$.

Here, $β\_{0j} $is the average outcome after removing the impact of individual characteristics; γ0 is the average value of the outcome across all practices; $ γ\_{1},γ\_{2},…,γ\_{r} $are the parameters estimated for practice level features. The sum $γ\_{1}Z\_{1j}+γ\_{2}Z\_{2j}+…+γ\_{q}Z\_{qj} $is the total impact of practice level features on outcomes, and $u\_{j}$ is the random error variation at the practice level. If the relationship between practice features and outcome are not linear, then a nonlinear model must be examined.

# [H2] Why Two Separate Models?

An alternative to multilevel modeling is to include all macro- and microlevel variables in one regression. Such a regression will include both the patient characteristics and the practice features. However, studies have shown that putting all the variables into one regression is not reasonable (Dickinson and Basu 2005). When macrolevel variables are correlated with micro level characteristics, patient characteristics may distort relationships among practice-level variables. Practices differ in the patients they care for, and a multilevel analysis is usually more accurate because it removes these differences.

# [H2] Assumptions of Multilevel Analysis

There are a number of ways that multilevel analysis could lead to faulty conclusions. First, each regression equation has its related assumptions. The error term must have a standard normal distribution. The equation must faithfully capture the relationship among patient characteristics and outcome—if the relationship is nonlinear, then the equation must be nonlinear. Similarly, the macrolevel equation must also capture the true relationship among practice features and outcomes.

 A key assumption of the analysis is that all relevant patient-level characteristics are included in the analysis. If there are patient comorbidities that affect the outcome, they should be included in the microlevel equation. Electronic health records contain many predictors, and inclusion of all relevant comorbidities could lead to hundreds and sometimes thousands of predictors. Every effort should be made to remove these patient characteristics.

# [H2] Intraclass Correlation

An important measure that describes the dependencies in the data is the intraclass correlation (ICC) coefficient. This statistic measures the extent to which individuals in the same practice or hospital are more similar to each other than they are to individuals in different groups. It tells us the extent to which variations in patient-level outcomes result from practices. It is calculated as proportion of the total variation in outcome that was explained by the practice-level variables, written as

$$ICC=\frac{Variation explained by practice-level variables}{Variation explained by both practice level and error term}.$$

# [H2] Example: Bypassing Community Hospitals

The need for multilevel analysis arises often when we want to compare hospital performances. For example, suppose we want to understand where ambulances should take trauma victims. When a car accident happens, trauma victims can be taken to the nearby community hospital or to a distant tertiary medical center. The medical center claims, perhaps justifiably, that it can care for the patient better—it has more sophisticated equipment and practitioners who are more experienced with trauma. The community hospital says that it can care for the patient faster, where seconds can save lives.

The decision of when and who should be transferred to the tertiary hospital is complex, with obvious financial impact on the hospitals and real-life differences for the patient. Obviously, it must be based on what is best for the patient. Determining what is best for the patient, however, requires multilevel modeling. First, patient characteristics are used to explain away as much of the variation in mortality rate at different hospitals as possible. Then, hospital features are used to see whether tertiary hospitals are better for the patient than community hospitals.

 For the patient characteristics, we include variables such as severe burn (Yes = 1, No = 0), head trauma (Yes = 1, No = 0), more than 65 years old (Yes = 1, No = 0), and male gender (Yes = 1, No = 0). For each patient, we also report the probability of survival calculated from a cohort of patients with the designated conditions. For hospital characteristics, we include type of hospital (Tertiary medical center =1, Community hospital=1) and availability of burn unit (Yes=1 , No=0).

 Exhibit 14.1 shows the simulated data. Note how hospitals A and B, the two trauma centers, have sicker patients. For example, 50 percent of the patients in tertiary centers have severe burns, versus 37 percent of patients in the community hospitals. Similarly, 55 percent of patients in tertiary centers have head trauma, versus 47 percent of patients in community hospitals. These data suggest that tertiary and community hospitals have different patient case mixes, and a straight regression of probability of survival on both hospital features and patient characteristics may be misleading. Multilevel modeling may be more reasonable.

**Exhibit 14.1** Survival Rate of Trauma Victims in Different Hospitals

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Probability of Survival | Severe Burn | Head Injury | 65+ Years | Male | Hospital | Probability of Survival | Severe Burn | Head Injury | 65+ Years | Male | Hospital |
| 0.695 | 1 | 1 | 1 | 1 | A | 0.753 | 0 | 1 | 0 | 0 | C |
| 0.734 | 1 | 1 | 1 | 0 | A | 0.801 | 0 | 0 | 1 | 1 | C |
| 0.786 | 1 | 1 | 0 | 1 | A | 0.852 | 0 | 0 | 1 | 0 | C |
| 0.819 | 1 | 0 | 1 | 1 | A | 0.952 | 0 | 0 | 0 | 1 | C |
| 0.868 | 1 | 0 | 0 | 1 | A | 0.995 | 0 | 0 | 0 | 0 | C |
| 0.893 | 1 | 0 | 0 | 0 | A | 0.427 | 1 | 1 | 0 | 0 | D |
| 0.853 | 0 | 1 | 1 | 1 | A | 0.483 | 1 | 0 | 1 | 0 | D |
| 0.907 | 0 | 1 | 0 | 1 | A | 0.619 | 1 | 0 | 0 | 0 | D |
| 0.936 | 0 | 0 | 1 | 1 | A | 0.499 | 0 | 1 | 1 | 1 | D |
| 0.990 | 0 | 0 | 0 | 1 | A | 0.667 | 0 | 1 | 0 | 1 | D |
| 0.505 | 1 | 1 | 1 | 1 | B | 0.695 | 0 | 1 | 0 | 0 | D |
| 0.557 | 1 | 1 | 1 | 0 | B | 0.728 | 0 | 0 | 1 | 1 | D |
| 0.613 | 1 | 1 | 0 | 1 | B | 0.807 | 0 | 0 | 1 | 0 | D |
| 0.703 | 1 | 0 | 1 | 0 | B | 0.931 | 0 | 0 | 0 | 1 | D |
| 0.739 | 0 | 1 | 1 | 1 | B | 0.993 | 0 | 0 | 0 | 0 | D |
| 0.745 | 0 | 1 | 1 | 0 | B | 0.261 | 1 | 1 | 1 | 1 | E |
| 0.837 | 0 | 1 | 0 | 1 | B | 0.322 | 1 | 1 | 0 | 1 | E |
| 0.906 | 0 | 0 | 1 | 0 | B | 0.342 | 1 | 1 | 0 | 0 | E |
| 0.978 | 0 | 0 | 0 | 1 | B | 0.395 | 1 | 0 | 1 | 0 | E |
| 0.997 | 0 | 0 | 0 | 0 | B | 0.467 | 0 | 1 | 1 | 0 | E |
| 0.364 | 1 | 1 | 1 | 0 | C | 0.604 | 0 | 1 | 0 | 1 | E |
| 0.431 | 1 | 1 | 0 | 1 | C | 0.676 | 0 | 1 | 0 | 0 | E |
| 0.555 | 1 | 0 | 1 | 0 | C | 0.694 | 0 | 0 | 1 | 1 | E |
| 0.687 | 1 | 0 | 0 | 0 | C | 0.753 | 0 | 0 | 1 | 0 | E |

*Note*: Data are simulated.

The multilevel modeling is done in two steps. In the first step, the variation in odds of survival is explained by the patients’ characteristics. We focus on odds, as opposed to probability of survival, because odds have no upper limit; therefore, regression predictions could be accurate. Probabilities were converted to odds using the formula: $Odds={Probability}/{(1-Probability)}$.

To calculate separate intercepts for each hospital, I defined each hospital as a binary variable an IF statement in Excel. I regressed probability of survival on the hospital binary variables, severe burn, head injury, age more than 65, and male gender variables. The overall intercept was set to zero so that intercept values were calculated for each hospital. Results are shown in exhibit 14.2. There were fifty observations and nine variables used to make the predictions; across the variables, 26 percent of variation in survival odds was explained by patient characteristics. As expected, odds of survival decreased if the trauma victims had severe burns, head injuries, were more than 65 years old, or were male. All hospitals improved survival odds, as indicated by the intercept values for each hospital. Each hospital intercept can be interpreted as the average odds of survival at the hospital, and hospitals differed a great deal in survival odds.

**Exhibit 14.2** Result for Patient Level Regression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Regression Statistics* |  |  |  |  |
| Multiple R | 0.64 |  |  |  |  |
| R-squared | 0.40 |  |  |  |  |
| Adjusted R-squared | 0.26 |  |  |  |  |
| Standard error | 58.14 |  |  |  |  |
| Observations | 50 |  |  |  |  |
|  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  | *Degrees of Freedom* | *Sum of Squares* | *Mean Squares* | *F- Statistic*  | *Significance of F*  |
| Regression | 9 |  93,831.73  |  10,425.75  |  3.08  |  0.01  |
| Residual | 41 |  138,595.89  |  3,380.39  |  |  |
| Total | 50 |  232,427.61  |  |   |   |
|  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t-Statistic* | *p-Value* |  |
| Intercept | 0 |  |  |  |  |
| Hospital A | 102.60 | 28.98 | 3.54 | 0.00 |  |
| Hospital B | 124.78 | 26.22 | 4.76 | 0.00 |  |
| Hospital C | 82.45 | 23.24 | 3.55 | 0.00 |  |
| Hospital D | 73.02 | 22.75 | 3.21 | 0.00 |  |
| Hospital E | 76.39 | 25.51 | 2.99 | 0.00 |  |
| Severe burn | -33.96 | 17.69 | -1.92 | 0.06 |  |
| Head injury | -36.25 | 17.21 | -2.11 | 0.04 |  |
| 65+ years | -44.18 | 16.91 | -2.61 | 0.01 |  |
| Male | -32.29 | 18.10 | -1.78 | 0.08 |   |

 In the next stage, the intercept values for each hospital were regressed on hospital features. Exhibit 14.3 shows the resulting reconstructed data. Note that the dependent variable is the average odds of survival in each hospital. These averages were calculated by removing the effects of patient characteristics in the previous regression. The independent variables were whether the hospital was a tertiary hospital and whether the hospital had a dedicated burn unit. The number of observations is the number of hospitals in the study (five). This number of observations is very small but can nevertheless be used to fit a linear model to the data, if this is the entire population of the hospitals. Were the study conducted at the national level, the number of hospitals would be much larger. The purpose of the analysis is not to generalize to hospitals in other locations but to describe local relationships.

**Exhibit 14.3** Data Organized for Hospital-Level Analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Hospital | Calculated Odds of Survival | Tertiary Center | Has Dedicated Burn Unit |
| A | 102.60 | 1 | 0 |
| B | 124.78 | 1 | 1 |
| C | 82.45 | 0 | 0 |
| D  | 73.02 | 0 | 0 |
| E | 76.39 | 0 | 0 |

 Exhibit 14.4 shows the macrolevel regression. Because all local hospitals are in the study and no inference is being made to hospitals in other locations, the test of statistical significance is irrelevant and should be ignored. Recall that statistics are generated so that we can infer from the sample population differences. Here we are dealing with the entire hospital population and do not need to make any inferences. In this analysis, regression coefficients are the only parameters that matter—all other parameters and statistics can be ignored. These data show that the odds of survival were 25 times greater at tertiary centers; the odds were 22 times greater if the center had a burn unit. These changes were calculated after removing the variation resulting from patient characteristics and therefore apply to the average patient that the ambulance picks up. Now we can go back to our initial question. Should the ambulance bypass a local community hospital to go to one of the two tertiary hospitals? The answer is yes, if during the transfer the odds of mortality do not increase by more than 25 times.

**Exhibit 14.4** Results for Hospital Level Regression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Regression Statistics* |  |  |  |  |
| Multiple R | 0.93 |  |  |  |  |
| R-squared | 0.86 |  |  |  |  |
| Adjusted R-squared | 0.72 |  |  |  |  |
| Standard error | 11.44 |  |  |  |  |
| Observations | 5 |  |  |  |  |
|  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |
|  | Degrees of Freedom | Sum of Squares | Mean Squares | F- Statistic | Significance of F |
| Regression | 2 | 1,835.99 | 917.99 | 40.22 | 0.02 |
| Residual | 2 | 45.65 | 22.82 |  |  |
| Total | 4 | 1,881.64 |  |  |  |
|  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t-Statistic* | *P-Value* |  |
| Intercept | 77.29 | 2.76 | 28.02 | 0.00 |  |
| Tertiary center | 25.31 | 5.52 | 4.59 | 0.04 |  |
| Has burn unit | 22.18 | 6.76 | 3.28 | 0.08 |  |

# [H1] Multilevel Modeling Using SQL

Multilevel modeling is a powerful tool that could help data analysts control for variations in patient characteristic. Unfortunately, this tool relies on regression analysis, a statistical procedure that is difficult to apply within electronic health records. To effectively complete the analysis, data must be organized in matrix format and exported to a statistical package, both steps that are difficult to carryout in today’s massive data within electronic health records. One way to avoid these problems is to use stratification instead of regression. This section shows how SQL can be used to estimate intercept for the microlevel regression models that are subsequently used in the macrolevel analysis.

Instead of regression, data are organized into different strata. A stratum refers to a unique combination of patient characteristics. In SQL, using GROUP BY command produces these strata. In this code, the average outcome or the probability of the outcome event is reported for each strata. Among the strata, the stratum where all patient characteristics are missing, is referred to as a corner case. The reported outcome for that corner case is the estimate for the intercept in the regression model. Like the intercept in the regression model, the outcome for this stratum shows the situation where none of the patient characteristics are present. Unlike, the intercept in the regression, the value of these corner cases can be calculated from the data without regression.

Sometime, there is no corner cases in the data or the number of observations for the corner cases is too low, e.g. less than 30. In these situations, the value of the corner cases must be inferred from other estimates. One way to do so is to impute the value from outcomes in other strata. Separate strata are organized for treated and untreated patients, both sharing the same strata. Then the outcome of treated patients are regressed on untreated patients with same strata. The missing corner cases for treated patients is estimated from the predicted value for the model of the data evaluated at the lowest level for untreated patients.

An example, can demonstrated the steps. Exhibit 14.1 provides data on experiences of seven medical centers with survival from cancer. The first step is to divide the data into various strata, each stratum reflecting the situation for one combination of patient characteristics (severity of burn, presence of head Injury, age older than 65 years, and gender) in one medical center. The following code can be used to obtain the strata for Hospital A:

DROP TABLE #A, #notA

-- Calculations for Hospital A

SELECT Avg([Probability of Survival]) as [Survival Rate A]

 ,[Severe Burn]

 ,[Head Injury]

 ,[65+ Years]

 ,[Male]

INTO #A

FROM [Cancer].[dbo].[Data]

WHERE [Hospital]='A'

GROUP BY [Severe Burn], [Head Injury],[65+ Years],[Male]

-- Calculation for not Hospital A

SELECT Avg([Probability of Survival]) as [Survival Rate Not A]

 ,[Severe Burn]

 ,[Head Injury]

 ,[65+ Years]

 ,[Male]

INTO #NotA

FROM [Cancer].[dbo].[Data]

WHERE [Hospital]<>'A'

GROUP BY [Severe Burn], [Head Injury],[65+ Years],[Male]

-- Merger of the data

Select [Survival Rate A] AS Y, [Survival Rate Not A] AS X, a.[Severe Burn], a.[Head Injury], a.[65+ Years], a.[Male]

INTO #Matched

FROM #A as a inner join #NotA as b

ON a.[Severe Burn]=b.[Severe Burn]

 and a.[Head Injury]=b.[Head Injury]

 and a.[65+ Years]=b.[65+ Years]

 and a.[Male]=b.[Male]

SELECT \* FROM #Matched

-- Estimation of Intercept

Declare @avgX as Float, @avgY as Float

SET @AvgX= (SELECT Avg(X) FROM #Matched)

SET @AvgY= (SELECT Avg(Y) FROM #Matched)

Select Sum((X-@avgX)\*(Y-@AvgY))/SUM((X-@avgX)\*(X-@AvgX)) AS Beta

INTO #Beta FROM #Matched

SELECT Avg(Y-Beta\*X) AS Intercept

FROM #Matched Cross Join #Beta

-- Result = 0.51897

Exhibit 14.5 shows the result of the SQL at point of estimating survival rates at hospital A and other hospital for fixed level of strata. Subsequent steps in SQL estimate beta and alpha paramters of the regression, resulting in an estimate for the intercept of 0.51.

Exhibit 14.5: Strata and outcomes for Hospital A

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Hospital A** | **Other Hospitals** | **Severe Burn** | **Head Injury** | **65+ Years** | **Male** |
| 0.99 | 0.95 | 0 | 0 | 0 | 1 |
| 0.94 | 0.74 | 0 | 0 | 1 | 1 |
| 0.91 | 0.70 | 0 | 1 | 0 | 1 |
| 0.89 | 0.65 | 1 | 0 | 0 | 0 |
| 0.85 | 0.62 | 0 | 1 | 1 | 1 |
| 0.79 | 0.46 | 1 | 1 | 0 | 1 |
| 0.73 | 0.46 | 1 | 1 | 1 | 0 |
| 0.70 | 0.38 | 1 | 1 | 1 | 1 |

Exhibit 14.6 shows the two estimated lines for survival at the hospital A and other hospitals. The X-axis lists the strata. The strata are a nominal scale, composed of various combinations of patient characteristics. For example, one stratum is called “Male”, another is called “Head injury, 65+, and Male”. We have arranged these strata so that later so that later strata have lower survival rates, as if the X axis is a continuous scale that measures the severity of the strata. Notice that the differences in survival rates are due to the hospitals’ performance and not the shared strata. In contrast, the actual survival in each hospital shows both the performance of the hospital and the strata. As the strata include more serious illness, there are lower survival rates. To estimate survival in Hospital A when no patient features are present, one could estimate the intercept of the regression of Hospital A’s survival on other hospital’s survival rates. The predicted values in this regression reflect the combined effect of both the strata and hospital’s performance. The intercept shows the effect of Hospital A but none of the effects of the strata. Thus, the intercept is a measure of survival when all patient features are absent.

In the SQL code, the intercept of this regression needs to be calculated without relying on any statistical package. If X is survival at other hospitals and Y is survival at Hospital A, then the regression line is $Y=α+βX$. The parameters alpha and beta in this equation can be estimated from the formulas:

$$β=\frac{\sum\_{i}^{}\left(\left(X\_{i}-\overbar{X}\right)\left(Y\_{i}-\overbar{Y}\right)\right)}{\sum\_{i}^{}\left(X\_{i}-\overbar{X}\right)^{2}} and $$

$$α=\overbar{Y}-β\overbar{X}.$$

In these two equations, $\overbar{Y}$ is the average survival at Hospital A and $\overbar{X} $is the average survival at other hospitals.. Using these formulas, the intercept in regression of survival at hospital A on other hospitals is estimated to be 0.51. A similar procedure is followed for the other hospitals and the results are then used in the macro-level analysis to see the relationship between survival rate, distance and patient satisfaction. The important point to remember is that it is possible to estimate the survival rates when no patient features are present (the intercept) without using statistical software and purely with SQL.

Exhibit 14.6: Survival Rate at Hospital A and Other Hospitals at Shared Strata

# [H1] Application of Multilevel Modeling to Other Data Types

Longitudinal data is another place where multilevel modeling is useful. Longitudinal data track events for each patient over time. In longitudinal data, a patient’s measures over time are the microlevel regression. The patient-level analysis is the macro level. In the first regression, the influence of time on the outcome is removed. In the second regression, the variation that remains in the outcome is regressed on patient characteristics.

 In the example we used in this chapter, we had a continuous variable: survival probabilities. Often, we only have a dichotomous variable (dead vs. alive). In these situations, both the macro- and microlevel regressions should be done using logistic regression.

# [H1] Measurement Issues

Every micro level datum can be aggregated and made into a macro level datum. For example, one may examine the probability of head trauma as an independent variable in the macrolevel regression. This type of artificial aggregation may hide the true relationship between the patient-level variables and the outcome. Replacing a variable with its average may distort the interaction among the patient-level data. It may also make interpretation of the findings more difficult. It is important for practice-level features to be truly practice level and not an artificial aggregation of patients’ characteristics at the practice.

# [H1] Summary

In this chapter, we learned to use intercept regressions, where the intercept of one regression is used subsequently in another regression. In the initial patient-level regression, the patient characteristics are used to explain the outcome. In the subsequent practice- or hospital-level analysis, practice features are used to explain the variation in outcome that was not explained by patient-level characteristics.

# [H1] Supplemental Resources

A problem set, solutions to problems, multimedia presentations, SQL code, and other related material are on the course website.

**[H1] References**

Bratzler, D. W., S. L. Normand, Y. Wang, W. J. O’Donnell, M. Metersky, L. F. Han, M. T. Rapp, and H. M. Krumholz. 2011. “An Administrative Claims Model for Profiling Hospital 30-Day Mortality Rates for Pneumonia Patients.” *PLoS One* 6 (4): e17401.

Chou, C. Y., M. C. Ma, and T. T. Yang. 2014. “Determinants of Subjective Health-Related Quality of Life (HRQoL) for Patients with Schizophrenia.” *Schizophrenia Research* 154 (1–3): 83–8.

Dickinson, L. M., and A. Basu. 2005. “Multilevel Modeling and Practice-Based Research.” *Annals of Family Medicine* 3 (Suppl 1): S52–60.

Gabriel, E., S. Narayanan, K. Attwood, S. Hochwald, M. Kukar, and S. Nurkin. 2018. “Disparities in Major Surgery for Esophagogastric Cancer Among Hospitals by Case Volume.” *Journal of Gastrointestinal Oncology* 9 (3): 503–16.

Kahn, J. M., R. M. Werner, S. S. Carson, and T. J. Iwashyna. 2012. “Variation in Long-Term Acute Care Hospital Use After Intensive Care.” *Medical Care Research and Review* 69 (3): 339–50.

Keenan, P. S., M. N. Elliott, P. D. Cleary, A. M. Zaslavsky, and B. E. Landon. 2009. “Quality Assessments by Sick and Healthy Beneficiaries in Traditional Medicare and Medicare Managed Care.” *Medical Care* 47 (8): 882–88.

Lin, S. H., W. C. Liao, M. Y. Chen, and J. Y. Fan. 2014. “The Impact of Shift Work on Nurses’ Job Stress, Sleep Quality, and Self-Perceived Health Status.” *Journal of Nursing Management* 22 (5): 604–12.

Peugh, J. L. 2010. “A Practical Guide to Multilevel Modeling”. *Journal of School Psychology.* Feb; 48(1): 85-112.

Rondeau, K. V., E. S. Williams, and T. H. Wagar. “Developing Human Capital: What Is the Impact on Nurse Turnover?” *Journal of Nursing Management* 17 (6): 739–48.

Schlam, T. R., N. L. Wilson, Y. Shoda, W. Mischel, and O. Ayduk. 2013. “Preschoolers’ Delay of Gratification Predicts Their Body Mass 30 Years Later.” *Journal of Pediatrics* 162 (1): 90–3.

Shen, X., B. C. Stuart, C. A. Powers, S. E. Tom, L. S. Magder, and E. M. Perfetto. “Impact of Formulary Restrictions on Medication Use and Costs.” *American Journal of Managed* *Care* 23 (8): e265–74.

Svensson, A. C., J. Stjernschantz Forsberg, D. Seblova, and A. Lager. 2017. “Residential Area and Physical Activity: A Multilevel Study of 68,000 Adults in Stockholm County.” *Scandinavian Journal of Public Health* 45 (1): 25–32