SAINT LOUIS UNIVERSITY.

Program-Level Assessment: Annual Report

Program Name (no acronyms): Biostatistics & Health Analytics

Department: Epidemiology & Biostatistics

Degree or Certificate Level: Master of Science

College/School: College for Public Health & Social Justice

Date (Month/Year): May 2023

Assessment Contact: Steve Rigdon

In what year was the data upon which this report is based collected? AY 2021/2022

In what year was the program's assessment plan most recently reviewed/updated? Never. This is a new program.

Is this program accredited by an external program/disciplinary/specialized accrediting organization or subject to state/licensure requirements?

This program is part of the portfolio of programs offered by the CPHSJ which is accredited by CEPH.

If yes, please share how this affects the program's assessment process (e.g., number of learning outcomes assessed, mandated exams or other assessment methods, schedule or timing of assessment, etc.):

Learning outcomes are called "competencies" by CEPH, so we have used that terminology here. At least six competencies are required for CEPH. Beyond these minimal requirements, we have considerable latitude in establishing competencies.

1. Student Learning Outcomes

Which of the program's student learning outcomes were assessed in this annual assessment cycle? (Please provide the complete list of the program's learning outcome statements and **bold** the SLOs assessed in this cycle.)

There are two concentrations to the MS-BSTHA program: (1) traditional biostatistics, and (2) geospatial health data analytics. The competencies for the two programs are given below. The first four competences are identical for the two concentrations.

Traditional Biostatistics	**Assessed In**
T1. FOUNDATIONS: Apply foundational principles of probability and statistics to develop methods for estimation and hypothesis testing.	5025 (I)
T2. ANALYSIS: Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.	5100 (I) 5961 (R)
T3. WRITTEN COMMUNICATION: Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis	5100 (I) 5961 (R)
T4. ORAL COMMUNICATION: Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis	5100 (I) 5961 (R)
T5. DATA AND COMPUTING: Apply the appropriate software to collect, store, manage, clean and analyze data.	5400 (I) 5961 (R)
T6. DESIGN: Design experiments or data collection strategies, including sample size requirements to answer research questions in public health.	5025 (I) 5961 (R)

Geospatial Health Data Analytics	Assessed In
G1. FOUNDATIONS: Apply foundational principles of probability and statistics to develop methods for estimation and hypothesis testing.	5025 (I)
G2. ANALYSIS: Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.	5100 (I) 5961 (R)
G3. WRITTEN COMMUNICATION: Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis.	5100 (I) 5961 (R)
G4. ORAL COMMUNICATION: Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis.	5100 (I) 5961 (R)
G5. DATA MANAGEMENT: Acquire, manage, analyze, and display geospatial health data.	5600 (I)
G6. APPLY SPATIAL & SPATIO-TEMPORAL MODELS: Build and analyze models to assess the health of populations across geographic regions and across both time and geographic regions.	5610 (I)

2. Assessment Methods: Artifacts of Student Learning

Which artifacts of student learning were used to determine if students achieved the outcome(s)? Please describe the artifacts in detail, identify the course(s) in which they were collected, and if they are from program majors/graduates and/or other students. Clarify if any such courses were offered a) online, b) at the Madrid campus, or c) at any other off-campus location.

Competency: T1 and G1: FOUNDATIONS: Apply foundational principles of probability and statistics to develop methods for estimation and hypothesis testing.
BST 5025 Theory of Biostatistics II
Spring 2022
Instructor: Rigdon

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5025	T1 and G1: FOUNDATIONS: Apply foundational principles of probability and statistics to develop methods for estimation and hypothesis testing.	0	4	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.

Midterm Exam

2. (8 Points) Suppose that $X \sim$ BINOMIAL(n, θ), that is, the Binomial distribution with parameters n (the number of trials) and θ (the probability of success on any single trial). [Assesses MS-BSTHA 1]
Using the normal approximation in the theorem below, find an approximate 95% confidence interval for θ if $n = 100$ yielded $x = 25$.

4. (16 Points: 4,4,4,4) Suppose $X_1, X_2, \ldots, X_n \sim$ i.i.d. POISSON(μ). That is, we have a sample of size n from the Poisson distribution with mean μ.

(a) Find and justify a sufficient statistic for μ.
(b) Find the MLE of μ. [Assesses MS-BSTHA 1]
(c) Is the MLE an unbiased estimator for μ? Justify your answer. [Assesses MS-BSTHA 1]
(d) Is the MLE a consistent estimator for μ? Justify your answer. [Assesses MS-BSTHA 1]

Homework 5

2. Suppose that X is the random number of people who show up at a pharmacy for a flu shot on a given day. Suppose further that $X \sim \text{POISSON}(\lambda)$. We have a sample of size n from this Poisson distribution; call these random variables X1,, X2, . . . ,Xn.
(a) What is the MLE of λ?
(b) Show that the probability that there would be no people who show up for the flu vaccine on any given day is $e^{-\lambda}$.
(d) Use the bootstrap method to find a confidence interval for $e^{-\lambda}$.

Competency: T2 and G2: ANALYSIS: Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.
BST 5100 Introduction to General Linear Modeling
Fall 2021
Instructor: Loux

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5100	T2 and G2: ANALYSIS: Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.	0	2	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.

Lab 1

Consider two possible models of sleep time:
Model 1: BMI, sex, race, and age (same as problem 5 above)
Model 2: BMI, sex, race, age, metro status, and exercise

a. For each model, provide: R^2, Deviance, AIC, and BIC.
b. Using each measure of model fit, which model is preferred?
c. Taking all four measures into account, which model best describes difference in sleep time across the sample?

Lab 2

Continuing with the model from question 4, add the interaction between sex and smoking history.

a. Is the interaction statistically significant?
b. Using the results regardless of significance, what is the odds ratio between smoking history and obesity for each sex?

Competency: T2 and G2: ANALYSIS: Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.
BST 5100 Introduction to General Linear Modeling
Spring 2022
Instructor: Loux

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5100	T2 and G2: ANALYSIS: Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.	0	1	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.

Lab 1

Consider two possible models of sleep time:
Model 1: BMI, sex, race, and age (same as problem 5 above)
Model 2: BMI, sex, race, age, metro status, and exercise

d. For each model, provide: R^2, Deviance, AIC, and BIC.
e. Using each measure of model fit, which model is preferred?
f. Taking all four measures into account, which model best describes difference in sleep time across the sample?

Lab 2

Continuing with the model from question 4, add the interaction between sex and smoking history.

c. Is the interaction statistically significant?
d. Using the results regardless of significance, what is the odds ratio between smoking history and obesity for each sex?

Competency: T3 and G3. WRITTEN COMMUNICATION: Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis
BST 5100 Introduction to General Linear Modeling
Fall 2021
Instructor: Loux

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5100	T3 and G3. WRITTEN COMMUNICATION: Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis	0	2	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.

Students perform a semester-long research project in which they develop a research question, identify publicly available data they can use to answer that question, propose and conduct an analysis, and report and interpret their results in an academic research paper format.

Competency: T3 and G3. WRITTEN COMMUNICATION: Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis
BST 5100 Introduction to General Linear Modeling
Spring 2022
Instructor: Loux

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5100	T3 and G3. WRITTEN COMMUNICATION: Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis	0	1	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.
Students perform a semester-long research project in which they develop a research question, identify publicly available data they can use to answer that question, propose and conduct an analysis, and report and interpret their results in an academic research paper format.

Competency: T4 and G4. ORAL COMMUNICATION: Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis
BST 5100 Introduction to General Linear Modeling
Fall 2021
Instructor: Loux

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5100	T4 and G4. ORAL COMMUNICATION: Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis	0	2	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.
Students give an in-class oral presentation of their project, including research question, description of the data set and methodology, and interpretation of their results.

Competency: T4 and G4. ORAL COMMUNICATION: Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis
BST 5100 Introduction to General Linear Modeling
Spring 2022
Instructor: Loux

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5100	T4 and G4. ORAL COMMUNICATION: Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis	0	1	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.
Students give an in-class oral presentation of their project, including research question, description of the data set and methodology, and interpretation of their results.

Competency: T5. DATA AND COMPUTING: Apply the appropriate software to collect, store, manage, clean and analyze data.
BST 5400 Applied Data Management
Fall 2021
Instructor: Elliott

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5400	T5. DATA AND COMPUTING: Apply the appropriate software to collect, store, manage, clean and analyze data.	0	2	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.
This competency was assessed with the following six assignments: **Intro to R assignment** The zipped archive *healthcare.zip* contains multiple inter-related files. Uncompress this archive and extract the files to a folder of your choice. The following are descriptions of the files in the archive: • Patient file: This is demographic information about each patient. There should only be one row per patient (but maybe there is an error, you should check!). • Staff file: This is demographic and salary information about each staff member. There should only be one row per staff (but maybe there is an error, you should check!) • Mortality file: This contains the date of death for each patient (if they have died). • Outpatient visit file: This contains information about each outpatient visit, including the staff and patient for each visit, and the diagnoses associated with the visit (ICD1-3 are diagnosis codes). The clinic code is also included, which identifies the clinic the patient was seen in. • Disease Map file: This provides a map from the diagnosis codes (ICD10) to disease conditions. • Clinic File: This provides a map from the clinic ID to the clinic type. Using the datafiles named above, perform the necessary actions to produce results needed to answer the following questions. Please provide answers to the questions in a question-answer format, and provide the R code needed for producing the needed output in a separate R source file with a filename as follows "firstname-lastname-week4.R" (where you put your first name and last name).

- Import the data into R using the read_ csv() function. (0.5 point)
- How many rows and columns are in each of the dataframes? (0.5 point)
- Look at details of the dataframe's data using describe() (from the package Hmisc). Then, identify the variables in each dataframe that have (a) missing values, (b) outliers (very large/small values, invalid values, etc.), (c) need to be cast to a different class (e.g., numeric should be treated as factor, character should be treated as factor, etc.), (d) if any levels (in a factor variable) need to be merged (1 point)
- Convert columns that have dates (e.g., date of birth) into an appropriate class, employing the appropriate function from the package lubridate. Summarize your changes in a table. (0.5 point)
- Perform the necessary steps to compute the proportions of patients born on each day of the week (e.g., Sunday: 0.12, Monday: 0.08, etc.) and provide the summary in a suitable format. From which file would you obtain the data to answer this question? (0.5 point)
- Create a new column in the appropriate file, label it *Age*, and include in it the age of each patient as of 2010-01-01. Then, using *Age*, compute the mean age, highest and lowest ages found in the data.(0.5 points)
- What proportion of patients in the data are aged over 65 years, as of 2010-01-01? (0.5 point)
- In each of the dataframes (each corresponding to data in each of the files), what are the proportions of observations that are excluded when na.omit() is applied on them? (1 points)
- EXTRA CREDIT: Compute the overall mortality in the cohort of patients on whom data is present in the database (across all of the files). Do this by combining the data from the *patient* and *mortality* files using a join that is appropriate. Which join would you use? Why? Assume that if a patient has an associated date of death recorded in the mortality file that the patient expired, and is alive otherwise. What is the overall mortality rate (in the merged dataframe, compute the proportion of records of patients who died to the total number of patient records). (3 points)

Data modelling assignment

This week's application assignment consists of two parts. For the first part, you will use the logic/causal model that you proposed during week 1 (and for which you developed an entry form during week 3) as the basis for developing a data model. Please address the following questions, using your original logic/causal model as a basis:

1. Provide a diagram of the logic/causal model that you proposed originally (you can copy+paste the diagram you submitted earlier) and provide a written description of the various entities in your logic/causal model. Specifically, identify at least 3 key entities in your model and describe some of their key attributes. Then, explain how these entities are related to each other. Use the material covered in this week's lecture video #1 and in chapters 3-6 of the Watson textbook as a basis for creating descriptions of the entities and their attributes (1 point)
2. Using the descriptions of the three entities and their inter-relationships, create the shell of specific tables that would be needed to capture the entities salient details and their inter-relationships. For each table, indicate which attribute serves as a primary key. In addition, indicate for each table whether there are any attributes (fields) that are related to attributes (fields) of other tables. Identify these as foreign keys and primary keys in the appropriate tables. Here, too, the content from video #1 and from chapters 3-6 will be a basis for your responses. (1 point).

For the second part, you will rely the material presented in the lecture videos #2 and #3, and the associated code and HTML reference materials. And the data to use is contained in the files included in the archive "data.zip" that accompanies this assignment instructions document.

1. Import and examine the data files. You will need to check the "events" file carefully, the data does not start on the first row. Look up the documentation of the function read_csv() and see how you can use the 'skip' argument the read_csv() function for solving this. Create a dataframe that contains the Olympic city for each year ('Edition'). This should have 26 rows, one for each city. (1 points)
2. Using that dataframe, identify any cities that have hosted the Olympics more than once. Write the dplyr code to return a dataframe with the city and number of times Olympics have been hosted, but only for cities that have hosted it more than once. (1 points)
3. How many unique sports, disciplines, and events has there been in the history of the Olympics? Write dplyr code to answer this and return a dataframe. (1 points)
4. EXTRA CREDIT: How many unique sports, disciplines and events were there in each year of the Olympics? (2 points)

Submission guidelines

Please follow the instructions provided above to submit:

- A PDF document with answers to part 1 of the assignment. Name this "assignment-6-part-1-firstname-lastname.PDF" (substitute firstname-lastname with your first and last names, respectively).
- A PDF document with answers to questions of part 2, providing appropriate evidence/results to substantiate your answers. Name this "assignment-6-part-2-firstname-lastname.PDF" (substitute firstname-lastname with your first and last names, respectively).
- An R source file (filename of the form: firstname-lastname-week6.R) with the code needed to produce the required results.

EMR assignment

Please develop the pseudo coding to complete the below task.
NOTE: You do not need to actually write the code (in R or any other language). You just need to describe what steps would eventually need to be coded.
In this exercise, you will generalize the ideas of data management you have learned to date. Instead of writing real code, you will need to think through how to solve a given task using only 'pseudo-code'. This 'pseudo-code' is simply a list of data verbs and the corresponding tables or variables names they act on. You could then translate these ideas into Python code, or SQL code, or R code, or something else!

Potential data verbs
• import
• subset (filter by row or select by column)
• create new variable (same size as original data)
• aggregate (with mean, median, etc)
• aggregate by group
• sort
• merge
• reshape (from long to wide or from wide to long)

Problems to solve
1) Use the medical data to whiteboard a solution to the following problems. For some of the answers, you will need to merge multiple files together.
a. Which staff member makes the most money?
b. Is there a pay disparity across gender amongst staff members? What happens when you consider staff type as well?
c. Which staff member saw the most patients in 2016?
d. Which staff member saw the most patients in primary care settings in 2016?
e. Is there a difference in mortality between men and women?
f. Which disease is most prevalent? Which disease is least prevalent?
g. Are there any diseases that are unevenly distributed across races?
h. Are there any diseases that are unevenly distributed across gender?
i. CHALLENGE: Calculate the mortality rate for every year between 2008 and 2018. Is it generally increasing, or decreasing? Assume patients are only at risk of death as of their first visit (in the outpatient Visit file). Once they have died, they should no longer be at risk in subsequent years…This is a harder question to answer than at first glance. What should the denominator of patients be for every year? How will you calculate it?

Copyright 2016-2019 BH Analytics, LLC used with permission from Evan Carey

SQL assignment

This week's assignment provides you with an opportunity to apply SQL-related concepts we have covered. Using the Synthea database (covered during this week's lectures), address the following questions by completing the necessary activities via SQLiteStudio and SQLite:

1. Create all of the tables necessary to load data spread across all of the files that are part of the Synthea download and load the data into these tables, ensuring that the specific primary key(s) of each table has been set (refer to the data dictionary associated with the CSV files). 1 point.
2. Using appropriate SQL queries, obtain the results to answer the following questions:
    1. Which class of encounters had the highest average base encounter cost? 0.5 points.
    2. Who was the payer for the highest base encounter cost? 0.5 points.
    3. At which healthcare provider was this cost incurred? 0.5 points.
    4. What are the name and location of the provider? 0.5 points.

5. What are the date, description and reason of this encounter? 0.5 points.

3. What are the average values of total claimed costs, for each class of encounters? 0.5 points.
4. This one does not require you to write SQL (you are welcome to, if you want to): in conceptual terms, explain how would you would obtain the data needed to answer the question: Do the total claims made by patients with at least one type of allergy differ from the total claims made by patients who do not have allergies in a statistically significant way? 1 points.

Submission guidelines

Please follow the instructions provided above to submit:

- A SQL script file with your name, assignment-6 in the filename, and .sql as its filename extension.
- Provide appropriate comments to annotate the SQL scripts associated with each question.
- For the final conceptual question, provide your answer via comments included at the end of the .sql file.

Survey data assignment

Find a journal article that uses a national weighted dataset (like BRFSS, NHIS, NHANES, etc.). 2 points
In one paragraph, discuss the degree to which the authors of the article adjusted for the weighting in their analyses. 3 points

SQL and R and web scraping assignment

This week, we have covered data scraping, with a focus on obtaining data from various sources (web-based) that share data via APIs. This assignment provides you an opportunity to reinforce the steps and the overall approach covered during this weeks lecture and demonstration videos. Using that content and the associate source code as a basis, please complete the following tasks, addressing the questions asked.

1. Using the code and data shared, compute the number of deaths in Missouri in the U.S. during the months of June and July 2020. Be sure to use the data from https://coronadatascraper.com as the basis for your analysis (this was covered in the video lecture titled "5: Another Comprehensive Example". 2 points.
2. Next, using the dataset titled "United States COVID-19 Cases and Deaths by State over Time" (referenced on the video titled "6: WHO and CDC data"), obtain a count of deaths in Missouri during June and July 2020. 2 points.
3. Compare the results from the above two steps and comment on whether the results are consistent – that is, are the counts obtained from the two sources of data agreeing with each other? If they are not, speculate on why the disagreement seems to be present. 1 points
4. Using the CDC dataset (step 2) determine the top 5 states that had the most number of confirmed deaths during the months of June and July. 2 points.
5. Using the CDC dataset, determine the death rates – ratio of number of confirmed deaths to the number of admissions – across US states and territories during June and July 2020 (combined, not per individual month). Plot these results using an appropriate graphical layout that highlights the differences. 3 points.

Submission guidelines

Please follow the instructions provided above to submit:

- R file associated with obtaining the data, collating them in a database, and analyzing them. 1 point.
- A PDF where you provide answers to the questions above. The answers do not have be lengthy (please do not include your source code in the PDF) – they should provide the answers sought in a clear and concise form. 1 point.
- Include your name and assignment number in comments at the top of the R sourcefile.

Competency: T6. DESIGN: Apply foundational principles of probability and statistics to develop methods for estimation and hypothesis testing.
BST 5025 Theory of Biostatistics II
Spring 2022
Instructor: Rigdon

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5025	T6. DESIGN: Apply foundational principles of probability and statistics to develop methods for estimation and hypothesis testing.	0	4	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.

Homework 3, Problem 2
Researchers would like to replicate the experiment described in the previous problem, but they would like sample sizes to be sufficiently large to give a strong power. If the true means for the three groups are +100, +100, and 0, respectively, what sample size would be needed to give a power of 0.95? What if we wanted the power to be 0.98? How many volunteers would be needed in total for this study?

Homework 4, Problem 1
Physicians have available two treatments for kidney stones. Treatment A is an invasive surgical procedure, and Treatment B is noninvasive. The choice of treatment was made by the physician. In a study of the effectiveness of the two treatments, researchers studied 700 people who had kidney stones and received one of the two treatments. 350 patients received Treatment A, and 350 received Treatment B. (Coincidentally, both treatments had the same number of patients.) The results are given in the following R chunk.

Final Exam Problem 3. Suppose you run an experiment similar to the one described in the previous problem, except (1) there are three treatments, and (2) the sample size is different. The data are contained in the file GoldbergScores.csv. In this case, there are three treatments stored in the variable Group: A="individual counseling", B="group counseling", C="no counseling". The change in the Goldberg score is in the column called GoldbergChange. (This is the "after" minus the "before" Goldberg scores).
(e) Suppose you are designing a follow-up study. Subject-matter experts tell you that they expect the means for the three groups to be $-3, -3, 0$, respectively. Based on your results from the first study, you believe the standard deviation of the outcome will be about $\sigma = 4$ in each group. How many individuals are needed in each group so that the power is 0.9?

Competency: G5. DATA MANAGEMENT: Acquire, manage, analyze, and display geospatial health data.
BST 5600 R for Spatial Analysis
Spring 2022
Instructor: Rigdon

Course	Competency	Fails to Meet Expectations	Meets Expectations	Exceeds Expectations
5600	G5. DATA MANAGEMENT: Acquire, manage, analyze, and display geospatial health data.	0	2	0

Briefly describe the methods used to assess this competency. Give examples of exam questions, projects, or homework assignments that assess this competency.

Project Description

The project should allow you the chance to apply spatial methods to a problem you have encountered. Most students in this class apply spatial regression (the BYM model), but other methods are acceptable. You are encouraged to talk to me about your topic, especially if it involves something other than the BYM model. Throughout the term, I have discussed the kinds of data that you will have to acquire. Now that we have covered the BYM model along with how to use INLA to get the Bayesian estimates, you should be able to proceed.
Here is an outline about how such a project would look. Models for areal data involve disease counts (or some other count) for cases that occur in predefined regions such as counties. You as the researcher might assume that the disease count depends on the population of the county and on other demographic variables for the county. For example, we will study cases of SIDS (sudden infant death syndrome) in North Carolina counties. Predictor variables might involve average income in the county, percent minority in the county, percent of college graduates in the county, etc. In applying these models you will be able to answer questions such as "What is the effect of variables X1, X2, and X3 on the rate of cases Y when we take into account the spatial correlation?"

You will likely need three sources of data for such a project:

1. the shape files for the state and its counties (Georgia would work great, because we already have the shape files for Georgia, but it's not hard to find the shape files for other states and counties.)
2. the outcome variable (a count variable for each county in the state)
3. demographic characteristics for each county in the state.

There will probably be three or more sources for the data.
[Note: Data management skills are needed to acquire the various data files and manipulate them so that the analysis can be done.]

Competency: APPLY SPATIAL & SPATIO-TEMPORAL MODELS: Build and analyze models to assess the health of populations across geographic regions and across both time and geographic regions.

BST 5620
Course was not offered in 2021/2022

Competencies assessed in the master's project:

T2. ANALYSIS: Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.
T3. WRITTEN COMMUNICATION: Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis.
T4. ORAL COMMUNICATION: Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis .
T5. DATA AND COMPUTING: Apply the appropriate software to collect, store, manage, clean and analyze data.
T6. DESIGN: Design experiments or data collection strategies, including sample size requirements to answer research questions in public health.

G2. ANALYSIS: Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.
G3. WRITTEN COMMUNICATION: Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis.
G4. ORAL COMMUNICATION: Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis.

No students completed the master's project in 2021/2022.

We have separate rubrics to grade the master's project for the traditional biostatistics and the geospatial health data analytics concentrations. They are given on the next two pages.

The competencies for the master's project, done as a requirement for BST 5961, are assessed by a committee of three, determined jointly by the student and the faculty member assigned to teach BST 5961. If a majority of the faculty committee evaluates a competency as meets or exceeds expectations, the student will pass for that competency. The student must pass all competencies.

Assessment and Grading Rubric – Spring 2023
BST 5961 – Geospatial Concentration

Student Name _____

Competency	Fails to meet expectations	Meets expectations	Exceeds expectations
ANALYSIS Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.	Method of analysis is overly simplistic or applied incorrectly. Conclusions may not be justified.	Method of analysis is appropriate and applied correctly. Conclusions are appropriate and follow from the analysis.	Method of analysis is appropriate and applied correctly. Methods used were complex and required considerable outside work to understand and apply. Conclusions are appropriate and follow from the analysis.
Check one box ->			
WRITTEN COMMUNICATION Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis.	The writing, either because of bad organization or spelling/grammar mistakes, makes it difficult to follow.	Writing is clear and mostly free of spelling or grammar mistakes. The report is well organized. Graphics are informative.	Writing is clear and compelling. The report is organized so that it leads inexorably to the conclusions. Explains why the problem is important. Writing is free of spelling or grammar mistakes. Graphics are illuminating.
Check one box ->			
ORAL COMMUNICATION Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis	One or more of the following: speaker cannot be understood, slides are not readable, slides are not organized, unable to answer several questions.	Speaker is understandable. Slides and graphics are informative. Student answers most questions without prompting.	Speaker shows confidence and poise. Engages the listener. Slides follow logically. Presentation was well timed. Graphics are illuminating. Student answered questions with insightful answers.
Check one box ->			
DATA MANAGEMENT Acquire, manage, analyze, and display geospatial health data.	Ambiguity about data sources or the meaning of the data values is unclear.	Data sources and description are clearly explained.	Data sources and description are clearly explained. Description of data collection, storage, processing, and management is done in a document archival system such as R Markdown.
Check one box ->			

Assessment and Grading Rubric – Spring 2023
BST 5961 – Traditional Biostatistics Concentration

Student Name _____

Competency	Fails to meet expectations	Meets expectations	Exceeds expectations
ANALYSIS Apply advanced statistical methods to analyze data and make inferences to answer research questions in public health.	Method of analysis is overly simplistic or applied incorrectly. Conclusions may not be justified.	Method of analysis is appropriate and applied correctly. Conclusions are appropriate and follow from the analysis.	Method of analysis is appropriate and applied correctly. Methods used were complex and required considerable outside work to understand and apply. Conclusions are appropriate and follow from the analysis.
Check one box ->			
WRITTEN COMMUNICATION Describe in writing the process of data collection, the application of statistical methodology, and the results of statistical analysis.	The writing, either because of bad organization or spelling/grammar mistakes, makes it difficult to follow.	Writing is clear and mostly free of spelling or grammar mistakes. The report is well organized. Graphics are informative.	Writing is clear and compelling. The report is organized so that it leads inexorably to the conclusions. Explains why the problem is important. Writing is free of spelling or grammar mistakes. Graphics are illuminating.
Check one box ->			
ORAL COMMUNICATION Orally describe the process of data collection, the application of statistical methodology, and the results of statistical analysis	One or more of the following: speaker cannot be understood, slides are not readable, slides are not organized, unable to answer several questions.	Speaker is understandable. Slides and graphics are informative. Student answers most questions without prompting.	Speaker shows confidence and poise. Engages the listener. Slides follow logically. Presentation was well timed. Graphics are illuminating. Student answered questions with insightful answers.
Check one box ->			

Competency	Fails to meet expectations	Meets expectations	Exceeds expectations
DATA AND COMPUTING Apply the appropriate software to collect, store, manage, clean and analyze data.	Ambiguity about data sources or the meaning of the data values is unclear.	Data sources and description are clearly explained.	Data sources and description are clearly explained. Description of data collection, storage, processing, and management is done in a document archival system such as R Markdown.
Check one box ->			
DESIGN Students should be able to design experiments or data collection strategies to answer research questions in public health.	The design is not able to produce data to answer the research question.	The design for collecting data was suitable for addressing the research question.	The design for collecting data was suitable for addressing the research question. The chosen design was justified and compared to competing designs.
Check one box ->			

3. Assessment Methods: Evaluation Process

What process was used to evaluate the artifacts of student learning, and by whom? Please identify the tools(s) (e.g., a rubric) used in the process and **include them in/with this report document** (please do not just refer to the assessment plan).

Competencies are introduced in various courses and assessed by the instructors according to the descriptions above. Many of the competencies are also addressed ("reinforced" is the term used by CEPH) in the master's project (BST 5961). The competencies for the master's project are assessed by a committee of three, determined jointly by the student and the faculty member assigned to teach BST 5961. If a majority of the faculty committee evaluates a competency as meets or exceeds expectations, the student will pass for that competency. The student must pass all competencies.

4. Data/Results

What were the results of the assessment of the learning outcome(s)? Please be specific. Does achievement differ by teaching modality (e.g., online vs. face-to-face) or on-ground location (e.g., STL campus, Madrid campus, other off-campus site)?

Competencies were assessed for all MS students who took courses in AY 2021/2022 regardless of their cohort. Competencies were evaluated as "Failed to Meet Expectations," "Met Expectations" and "Exceeded Expectations."

Competency	Course/Instructor	Failed to Meet Expectations	Met Expectations	Exceeded Expectations
T1 & G1 Foundations	BST 5025 / Rigdon	0	4	0
T2 & G2 Analysis	BST 5100 / Loux	0	3	0
T3 & G3 Written Comm.	BST 5100 / Loux	0	3	0
T4 & G4 Oral Communication	BST 5100 / Loux	0	3	0
T5 Data and Computing	BST 5400 / Elliott	0	2	0
T6 Design	BST 5025 / Rigdon	0	4	0
G5 Data Management	BST 5600 / Rigdon	0	2	0
G6 Apply Spatio Temporal Models	Not offered in AY 2021/2022			

5. Findings: Interpretations & Conclusions

What have you learned from these results? What does the data tell you? Address both a) learning gaps and possible curricular or pedagogical remedies, and b) strengths of curriculum and pedagogy.

The program is new, and data is scarce, but we believe that students are well prepared when they enter and they are able to meet expectations in all of our competencies.

6. Closing the Loop: Dissemination and Use of Current Assessment Findings

A. When and how did your program faculty share and discuss the results and findings from this cycle of assessment?

I have shared this document with the department chair. We will share the information at a faculty meeting in AY 2023/2024.

B. How specifically have you decided to use these findings to improve teaching and learning in your program? For example, perhaps you've initiated one or more of the following:

Changes to the Curriculum or Pedagogies	• Course content • Teaching techniques • Improvements in technology • Prerequisites	• Course sequence • New courses • Deletion of courses • Changes in frequency or scheduling of course offerings
Changes to the Assessment Plan	• Student learning outcomes • Artifacts of student learning • Evaluation process	• Evaluation tools (e.g., rubrics) • Data collection methods • Frequency of data collection

Please describe the actions you are taking as a result of these findings.

None.

If no changes are being made, please explain why.

With little data, there is not enough information to make proposed changes.

7. Closing the Loop: Review of Previous Assessment Findings and Changes

A. What is at least one change your program has implemented in recent years **as a result of previous assessment data**?

This is the first assessment report for this program.

B. How has the change/have these changes identified in 7A been assessed?

NA

C. What were the findings of the assessment?

NA

D. How do you plan to (continue to) use this information moving forward?

Students are doing well. We had no graduates in the AY 2021/2022, but we will have five graduates in 2022/2023. Next year's assessment report will provide more information that we may use moving forward.