SAINT LOUIS UNIVERSITY.

Program-Level Assessment: Annual Report

Program: Data Science, B.S.	Department:
Degree or Certificate Level:	College/School:
Date (Month/Year):	Primary Assessment Contact:
In what year was the data upon which this report is based collected?	
In what year was the program's assessment plan most recently reviewed/updated?	

1. Student Learning Outcomes

Which of the program's student learning outcomes were assessed in this annual assessment cycle?

Graduates will apply statistics to analyze data sets.

2. Assessment Methods: Artifacts of Student Learning

Which artifacts of student learning were used to determine if students achieved the outcome(s)? Please identify the course(s) in which these artifacts were collected. Clarify if any such courses were offered a) online, b) at the Madrid campus, or c) at any other off-campus location.

1. Answers on the final exam in STAT 4880 (Bayesian Statistics) to a question regarding building an appropriate model in order to analyze a data set were evaluated.
2. Examples of specific topics taught in CSCI 1080 were solicited for each learning outcome to determine whether the course indeed covers all program level outcomes at a basic level.

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.

1. Darrin Speegle used a rubric (see below) to assign one of 3 levels to each students' response.
2. Abby Stylianou provided course specific examples for each of the five program level learning outcomes from her course.

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)?

1. The students achieved a very high level of proficiency overall in the measure. Of the 11 students observed, 8 scored in the ``meets goals" level, 2 scored in the ``progressing" level, and 1 scored in the ``does not meet goals" level.
2. All program level outcomes are indeed covered in the CSCI 1080 course at an elementary level.

5. Findings: Interpretations & Conclusions

What have you learned from these results? What does the data tell you?

1. To a large extent, students in STAT 4880 seem to know how to build and interpret models at a level appropriate for graduates of the program. While this is only one aspect of the learning objective “Graduates will apply statistics to analyze data sets,” it is a good start to understanding how well our graduates will be able to apply statistics to data.
2. CSCI 1080 is a crucial introductory course for the data science major. We will, in the future, provide objective assessment on the degree to which the program level learning outcomes are covered.

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

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

This is our first year as a program, and we have not yet discussed the findings.

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.

1. The result of the assessment in STAT 4880 was that the students are doing very well in the area measured. There seems to be no action necessary at this point.
2. We will be implementing assessment of learning outcomes in CSCI 1080 in Fall, 2021.

If no changes are being made, please explain why.

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 assessment data?

This is the program’s first year.

B. How has this change/have these changes been assessed?

This is the program's first year.

C. What were the findings of the assessment?

This is the program's first year.

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

This is the program's first year.

IMPORTANT: Please submit any assessment tools and/or revised/updated assessment plans along with this report.

Assessment, Fall 2020

The purpose of this assessment is to determine the extent to which students are meeting learning objective 2 of the assessment plan, "Graduates will apply statistics to analyze data sets."

According to our curriculum mapping, students in this class will be expected to reach level 3, which includes

1. Synthesis: Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.
2. Evaluation: Make judgments about the value of ideas or materials.

The following problem was given to students on the final exam.

Consider the seed germination data that is available at https://mathstat.slu.edu/~speegle/Fall2020/4880/data/germination/small_germ.csv. Researchers planted seeds of 15 different types of plants and counted the number of seeds that successfully germinated and the number that did not germinate. Create a hierarchical model with vague priors for θ_i $(1 \leq i \leq 15)$, the proportion of times a seed of type i will germinate and use it to answer the questions below. (Specification and checking of model is worth 15 points).

1. (5 points) Find a 95 percent credible interval for the mode ω of the proportion of seeds of all types that will germinate.
2. (5 points) Find a 95 percent credible interval for the difference in proportion of seeds of type ABL and of type ASJ that will germinate.
3. (5 points) Interpret your answer in (b) with respect to a ROPE of $(-.05, .05)$.
4. (5 points) The actual data is given in https://mathstat.slu.edu/~speegle/Fall2020/4880/data/germination/med_germ.csv, where seeds were planted in batches of 20 inside of paper cups. Create a new model which has two hierarchies; one corresponding to the cups inside of the plant species, and one corresponding to the plant species. Find a 95 percent credible interval for the difference of the modes of the proportion of seeds of type ABL and type ASJ that will germinate.

There are three levels of achievement for this problem:

1. Meets Goals of Program in This Area. The student answers all questions factually correctly. The student makes appropriate choices of models for parts (a-b) and (d) of the problem. The student interprets the answers to the problem in the context described in (c). The students' work is clear without any logical flaws.
2. Progressing Towards Goals of Program in This Area. The student answered at least one of the questions correctly. The student made appropriate choice of model for either part (a-b) or part (d). The student interprets the answer to the problem in the context described in (c), perhaps with some minor flaws. The students' work lacks some clarity and may contain minor logical flaws in addition to mis-modeling.

3. Not Meeting Goals of Program in This Area. The student did not answer any of the questions correctly. The student did not make an appropriate model choice for either part (a-b) nor part (d). The students' work had major logical flaws and/or clarity issues.

1. Consider the `adipose` data set that is in `fosdata`. You will need to re-install `fosdata` and restart R in order to access the data set. Note the ordering of levels for `sex`.

   (a) Remove all observations which correspond to `vat` 5 or less.

   (b) Create a new variable called `waist_height` that is the waist measurement divided by the stature measurement.

   (c) Create a Bayesian model of the log of visceral adipose tissue described by waist_height. Allow different slopes and intercepts for males and females.

   (d) Create a 95 percent credible interval for the difference in slope of log of visceral tissue described by waist_height between males and females. Interpret relative to a ROPE of $(-1, 1)$.

   (e) Find a 95 percent credible interval for the log of the visceral adipose mass of a female patient with waist_height ratio 0.5.

1. Graduates will use programming and other computer science skills to analyze and interact with data.

Students in 1070 learn to use Python (and appropriate libraries like Numpy, pandas, and sklearn) for loading datasets and performing data analysis, visualization, and basic machine learning.

2. Graduates will apply statistics to analyze data sets.

Students learn basics about statistical significance, correlation and multicollinearities when learning about linear, non-linear and multiple regression. In previous semesters (although not the most recent one), we have spent a significant portion of the semester on probability, conditional probability, and hypothesis testing (definition of the null hypothesis, definition and computation of p-value, etc.).

3. Graduates will be able to acquire and manage complex data sets.

Throughout the semester students work with relatively complex datasets that have been pre-processed for them. At the end of the semester, students perform a group "Exploratory Data Analysis" project, that requires them to define a question that they wish to answer, find and curate an appropriate dataset to support investigating that question, and do visualization/analysis on that dataset. In general, these datasets have been what I'd call *moderately* complex. Most students are working with data at the scale of hundreds of examples, with tens of attributes. One goal I have for future semesters is to get students working with more complex data.

4. Graduates will be able to use technical skills in predictive modeling.

Students learn about linear, non-linear and multiple regression. We also spend a fair amount of time on machine learning algorithms for supervised and unsupervised classification, which perhaps doesn't fit as directly into predictive modeling (this past semester we covered k-Nearest Neighbors and k-Means Clustering in depth; in previous semesters we've covered more algorithms, such as the Naive Bayes classifier and introductions to deep learning).

5. Graduates will be able to visualize data to facilitate the effective presentation of data-driven insights.

A significant portion of the early semester is spent on learning concepts in data visualization (different types of plots, when they're appropriate to use, and how to use them). During the semester, quizzes and exams included open ended reflections where students were expected to be able to convey clearly insights gleaned from data, or from data visualizations or summaries. Students are also responsible for an end of the semester group presentation of their own exploratory data analysis, where they must present their question/hypothesis, visualizations of the data they worked with, and visualizations/summaries of the conclusions they made based on their data analysis/machine learning deployment.