1. Co-Contributors:
2. This Proposal is Part of a Package:  Yes
3. Course/Programs related to this Package:
   | Linguistics Major: This course is Preparation for the Linguistics Major
4. Course Status:  Launched
5. Subject/Abbrev:  LING - Linguistics
6. Department:  450 - LINGUISTICS - *Active*
7. School:  62 - ARTS & LETTERS - *Active*
8. Number:  0270
9. Suffix:
10. Catalog Number:  0
11. Full Title:  Elementary Statistics for Language Studies
12. Abbrev. Title:  STATFORLANG
13. Variable Title:  No
14. General Education:
   a. Proposed for General Education:  Yes
   b. GE Fulfill:  IIA4 - Foundations of Learning - Natural Sciences & Quantitative Reasoning - Mathematics/Quantitative Reasoning
   c. If Not Approved GE, Still New Course:  No
15. Special Course Designator:  -
16. Cross Listed Courses:  No
17. Campus:  R - San Diego Campus
18. Description:  Introductory statistics and probability using language data to illustrate statistical ideas and emphasizing concepts that facilitate the analysis of quantitative language data. Probabilistic explanation in linguistics. Measures of central tendency and dispersion, significance, correlation, regression, mixed models. Students with credit or concurrent registration in the following lower division statistics courses will be awarded a total of four units for the two (or more) courses: Linguistics 270; Administration, Rehabilitation and Postsecondary Education 201; Biology 215; Civil Engineering 160; Economics 201; Political Science 201; Psychology 270; Sociology 201; Statistics 119, 250.
19. General Text:
20. Course Hours Description:
21. Course Statement:
22. Proposed Start Year:  2019  Term:  Fall
23. Variable Units:  No
24. Units:  3.00
25. Mode of Instruction:
<table>
<thead>
<tr>
<th>Units</th>
<th>Staffing Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Lecture:</td>
<td>3.00</td>
</tr>
<tr>
<td>b. Lecture 2:</td>
<td>0</td>
</tr>
<tr>
<td>c. Seminar:</td>
<td>0</td>
</tr>
<tr>
<td>d. Laboratory:</td>
<td>0</td>
</tr>
<tr>
<td>e. Activity:</td>
<td>0</td>
</tr>
<tr>
<td>f. Supervision:</td>
<td>0</td>
</tr>
<tr>
<td>g. Discussion/Clinical Prog:</td>
<td>0</td>
</tr>
<tr>
<td>h. ROTC:</td>
<td>0</td>
</tr>
<tr>
<td>i. NTI:</td>
<td>0</td>
</tr>
</tbody>
</table>
26. Repeatable for Credit:  No
27. Grading Methods:  +LETTER (C/N OK)
28. Prerequisites:  
N/A
29. Resources:  
   a. Faculty Resources
      ■ Describe Available Resources:
         There are several TT faculty in the department capable of teaching this class, including Gawron, Malouf, Doyle, and Csomay.
      ■ Describe Needed Resources:
30. Relationship of this Course to Total Educational Program of University:
   a. What other course or courses (in your department and others) cover subject matter similar to proposed content?
      ■ BIOL 0215 - Biostatistics *Active*
      ■ CIV E 0160 - Statistical Methods for the Built Environment *Active*
      ■ ECON 0201 - Statistical Methods *Active*
      ■ POL S 0201 - Elementary Statistics for Political Science *Active*
      ■ PSY 0280 - Statistical Methods in Psychology *Active*
      ■ SOC 0201 - Elementary Social Statistics *Active*
      ■ STAT 0250 - Statistical Principles and Practices *Active*
   b. Will this course replace a course now offered?  No
      i. If yes, which course(s)?
   c. Has this course been offered as a topics course?  No
      i. If yes, which semester and year?
   d. Is course being proposed in response to academic review or accreditation recommendations?  No
      i. If yes, explain fully.
   e. Does this course affect the STAR ACT?  No
      i. If yes, explain fully.
31. Justification:
   a. Need for course:
      Understanding language variation along various dimensions (within an individual, across individuals and communities, across linguistic contexts) offers a unique perspective from which to understand the role of statistical ideas in the social sciences. Facts about the frequency of language types (sounds, words, phrases) play a key role in explaining language variation across communities (dialects), the emergence of new forms (language change), and language processing. Additionally, because sounds, words, and phrases are not normally distributed, interpreting quantitative linguistic data requires special care. This class covers introductory statistics and probability, using language data to illustrate statistical ideas and emphasizing those concepts that facilitate linguistic explanation and the analysis of quantitative data.
   b. Justify level of course (a justification for all 500-level courses is required by Graduate Council):
      Like other introductory statistics classes specialized for psychological and social science (Political Science 201, Sociology 201, Economics 201, Psychology 280), this is a 200-level course, emphasizing methods and quantitative relations in a particular scientific or social science setting rather than mathematical rigor.
   c. Justify Cr/NC only grading (if applicable):
   d. Justify if no prerequisite(s) for 300 through 700 level courses:
32. Needs Met by Course:
   a. Satisfies:
      i. Degree/Major
   b. Does this course affect a program?  Yes
   c. List courses for which this course will be required as a prerequisite:
   d. Have Course Change proposals been submitted to make prerequisite changes?  No
   e. List which other departments or programs will use this course:

33. **Required Student Course Materials:**
   a. **Textbooks:**

   b. **Manuals:**

   c. **Periodicals:**

   d. **Software:**
      - RStudio w R 5.3.2. https://www.rstudio.com/, (5.3.2/e).

   e. **Other:**

   f. **Accessibility Materials:** Yes

   g. **Accessibility**
      
      | Textbooks, manuals, and software will be made available 6 weeks in advance of the start of the course; accessible electronic copies of the texts will be available. The syllabus includes a statement inviting students with disabilities to see the instructor to discuss accommodations in advance. The course will use Blackboard to post class information, course materials, and pointers to and help with the software. |

34. **Writing Component (300 level courses or above):**
   Assignments and exams alike will call for short answers, of at most a paragraph. Students will nevertheless be asked to explain their answers, to construct arguments for them, and to describe the data and their models using appropriate terminology.

35. **Grading Standards:**
   - Assignments: 40%
   - Quizzes: 20%
   - Midterm 20%
   - Final 20%

36. **Graduate Student Requirements (500 level courses):**
   a. **For 500 numbered courses, specify any special assignments for graduate students:**
   b. **For 500 numbered courses, if grading standards and weights are identical for graduate and undergraduate students, please justify:**

37. **Student Learning Outcomes:**
   1. **Outcome 1:**
      i. Students will be able to compute counts, contingency tables, and basic descriptive statistics from linguistics data, as well as plot data distributions, and explore data graphically in other ways. Since these skills will be exercised in R, a statistical software package, this targets GE Goal One (Quantitative Reasoning).
      ii. **Course Activity:**
         
         | Readings, the software manual, and inclass software demonstrations will guide students through an introduction to the statistics package R. |
      iii. **Assessment Strategy:**
         
         | The first quiz will test knowledge of the core language concepts of R. The first assignment results will assess competence in using R for basic descriptive statistics. |

   2. **Outcome 2:**
      i. Students will be able to conduct basic significance tests and explain the meaning of the results;
      ii. **Course Activity:**
         
         | Readings, discussion of implications of example test, and an assignment testing a linguistic hypothesis. |
      iii. **Assessment Strategy:**
         
         | Midterm and performance on the assignment. |

   3. **Outcome 3:**
      i. Students will calculate a linear regression model for a dependent linguistic variable, and describe what the model
shows about each independent variable; this targets Quantitative Reasoning Goal Two (real-world problems).

ii. **Course Activity:**
   
   Readings, in class demonstrations using R, an assignment predicting a continuous linguistic variable such as vowel frontness using an independent variable such as community membership.

iii. **Assessment Strategy:**
   
   Quiz and performance on the assignment.

4. **Outcome 4:**
   
   i. Discriminate between different types of probability distribution and test goodness of fit of data to distributions; this targets Quantitative Reasoning Goal Two (real-world problems).

   ii. **Course Activity:**
   
   Readings, discussion of implications of goodness of fit plots, R demonstrations of relevant tools, an assignment testing goodness of fit of various word distributions to a Poisson distribution. This targets Quantitative Reasoning Goal Two (real-world problems).

   iii. **Assessment Strategy:**
   
   Final and performance on the word distribution assignment.

5. **Outcome 5:**
   
   i. Identify cases where individual variation (such as variation across speakers or variation across words) is an integral part of community language patterns, and explain how this can be built into a model as a random effect.

   ii. **Course Activity:**
   
   Readings and lecture. Discussion of various example modeling tasks distinguishing mixed and random effects, A mixed effects assignment using data about sentence acceptability, treating the verbs as random effects.

   iii. **Assessment Strategy:**
   
   Quiz, final exam, and assignment results.

38. **Design and Conduct**
   
   a. **Include topics to be covered:**

   i. Probability and counting, Binomial distribution.

   ii. Conditional probability, Bayes’ Theorem

   iii. Mean, Variance, Normal distributions

   iv. Other types of probability distribution

   v. Hypothesis testing: Significance and confidence

   vi. Word frequencies, Zipf’s Law

   vii. Word distributions, burstiness of language, words that do not follow a Poisson distribution

   viii. Information theory: entropy and compression

   ix. Correlation and regression

   x. Mixed models

   b. **Approximate time in number of weeks:**

   i. 2

   ii. 1

   iii. 1

   iv. 1

   v. 2

   vi. 1

   vii. 1

   viii. 2

   ix. 2

   x. 2

39. **General Education:**
   
   a. **Section A:**

   i. **Outline Selection:**

   1. Communication and Critical Thinking
II. Foundations of Learning

☐ A. Natural Sciences and Quantitative Reasoning
☐ 4. Mathematics/Quantitative Reasoning

III. Lifelong Learning and Self-Development (Area E)

IV. Explorations

ii. Justify Classification:

This is a Foundations of Learning course because it teaches the study of language from a scientific perspective, emphasizing data-driven methods and reproducible results. It falls in the category of quantitative reasoning because the course content focuses on statistical methods for understanding and explaining language data. With respect to Goal I (computational skills and mathematical concepts), the course includes extensive assignments using R (a statistical software package) to do statistical hypothesis-testing, distribution visualization, distribution-fitting, and regression. With respect to Goal II (using quantitative methods to solve and communicate real-world problems), the course includes assignments guiding students through the process of building multi-factor models of speaker judgments of the acceptability of sentences, as well as models explaining contrasting distributional properties of content words and function words. Students also explore the issue of how to quantitatively characterize the variation that underlies dialect boundaries.

b. Section B:

Foundations of Learning - Natural Sciences and Quantitative Reasoning - Mathematics/Quantitative Reasoning

Goal 1: Apply appropriate computational skills and use basic mathematical concepts to analyze problems in natural and social sciences.

Activity:

Use the statistic software package R to provide counts from linguistics data, construct contingency tables, compute basic descriptive statistics, plot probability distributions, and explore data graphically in other ways.

Assessment:

These preliminary computational skills will be assessed in a graded assignment in the first two weeks, and computational skills will be developed in all assignments, which will all be graded and all involve the use of R.

Goal 2: Use methods of quantitative reasoning to solve and communicate answers to real-world problems.

Activity:

Assignment 1. Apply a type of probability distribution (the Poisson distribution) to describe lexical dispersion. Assignment 2. Use inter-community variation in vowel formants to identify dialect boundaries. Assignment 3. Using a regression model to account for speaker grammaticality judgments.

Assessment:

Assignments 1, 2, and 3 will be graded. Concepts involving the Poisson distribution, inter- and intra-speaker variation in vowel production, and regression will also be evaluated on exams.

c. Section C:

i. 1. Construct, analyze, and communicate arguments.

Course Work:

Students will review research showing that words divide into those whose distributions fit a Poisson distribution fairly well and those that do not. Function words generally turn out to be among the words best fitting a Poisson distribution. Natural constraints on communication and coherence explain this.

Assessment:

The assignment that leads students through the process of fitting word distributions against Poisson distributions also guides them to make predictions about which words will be the best fits. Finally they will be asked to construct an argument for how natural constraints on communication can account for the distributional facts they observe.
ii. ☑ 2. Apply theoretical models to the real world.

**Course Work:**

1. Applying a type of probability distribution (the Poisson distribution) to describe lexical dispersion. 2. Using inter-community variation in vowel formants to identify dialect boundaries. 3. Using a multifactor regression model to account for speaker grammaticality judgments.

**Assessment:**

1. An assignment using distribution-fitting techniques to test the fit of lexical dispersion facts to a Poisson distribution. 2. An assignment exploring community variation in vowel formants to describe dialects. 3. An assignment exploring a regression model to account for speaker grammaticality judgments.

iii. ☑ 3. Contextualize phenomena.

**Course Work:**

1. Categorical speaker judgments of sentence acceptability can be explained as the combined effect of an assortment of mild preferences that add up to make a sentence highly improbable. In such cases, it is whether a preference violation occurs in conjunction with other violations that determines speaker judgments of acceptability. 2. Students will review research showing that words divide into those whose distributions fit a Poisson distribution fairly well and those that do not. The research claims that the words that best fit a Poisson distribution are semantically bleached words or grammar words like "the" and "is", while those that do not are semantically specific.

**Assessment:**

1. In several assignments, students will explore the predictive power of a multi factor account of speaker judgments of sentence acceptability. 2. Students will construct frequency distributions for a variety of words, and will be asked to test the hypothesis that the empirical frequency distributions fit a Poisson distribution. They will also be asked to determine what prediction the literature makes for each word.

iv. ☑ 4. Negotiate differences.

**Course Work:**

**Assessment:**

v. ☑ 5. Integrate global and local perspectives.

**Course Work:**

**Assessment:**

vi. ☑ 6. Illustrate relevance of concepts across boundaries.

**Course Work:**

**Assessment:**

vii. ☑ 7. Evaluate consequences of actions.

**Course Work:**

**Assessment:**

d. Section D:

i. **Forms of Communication:**

Students will need to master a technical vocabulary for data distributions and properties of models (e.g. linear versus non-linear). They will be asked to describe the kinds of models they are using and
to describe their shortcomings. They will also have to be able to choose appropriate visualizations to answer the questions they are asking, and to exhibit those answers to others.

d. Section E:

<table>
<thead>
<tr>
<th>i. Required For Majors: Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>It serves Linguistics majors by showing how quantitative methods can reveal fundamental properties of language linked to its communicative purpose and its role in community cohesion and identity. It serves GE students because it satisfies the GE goals of a quantitative methods class by applying statistical methods to real world problems involving language and linguistic communities.</td>
</tr>
</tbody>
</table>

e. Section F:

| i. GE Cultural Diversity Upper Division Only: No |
| ii. Consider if Rejected: No |