Date: 5/5/2008
To: AugCore Subcommittee
From: Milo Schield
Topic: Request Approval of GST 200 as "Quantitative Applications" (QA) course

Background: GST 200, **Quantitative Reasoning / Statistical Literacy**, is designed for students in majors that do not require a quantitative course. Such majors include English, History and Philosophy. GST 200 has already been approved as a Quantitative Foundations (QF) course. This course has been reviewed and/or taught by faculty from different disciplines: Business, Economics and Physics. It would be open to faculty from other departments to teach just as Critical Thinking (GST 100) is open to all faculty to teach. Typically, three sections are offered each year. More can be offered if needed.

Request: Please approve GST 200 as a Quantitative Application (QA) course.

Course: Attached is a GST 200 syllabus and a packet of the right-wrong web-based assignments.

- Appendix 1 includes the rationale for designating GST 200 as a QF course.
- Appendix 2 includes the criteria for a course to qualify as a QA course.

In depth studies in GST 200 include mathematical relationships, statistical relationships and logical analysis. This can be seen from the documentation of the assignments.

Details of the final learning project are now presented. This is a major student-generated project. Each student has a unique set of data that involves variables of their own choosing – just as in BUS 379 and in BUS 264. But instead of focusing on the manipulation of the data using Excel, GST 200 students focus on taking into account the influence of a confounder and on interpreting the influence of differing definitions on the percentages and related associations. Thus, GST 200 maintains focus on the influence of context – as appropriate for students in the liberal arts.

Terms: Outcome, predictor and confounder are chosen by and unique to each student.

- Outcome: The result of interest in an association. Must be two-valued.
- Predictor: A factor that is associated with the outcome. Must be continuous or ordinal.
- Confounder: A 3rd factor that is mixed up with the predictor and the outcome.
- Assembly: The process of defining a variable that can influence the count or percent.
  Example: Define a high GPA student.
  Assembly #1: Any GPA above 3.9. Assembly #2: Any GPA above 3.0.
  Assembly #1 gives a smaller percentage of students that assembly #2.
- Standardizing: "Taking into account" the influence of a related factor.
  Normally this involves multivariate regression. But when the variables are all binary, this can be done using simple algebra or a simple graphical technique. All students in GST 200 are well-versed in standardizing.
  See www.StatLit.org/pdf/2006SchieldSTATS.pdf

Assignment (short version):
1. Create questions for the data analysis project?
2. Obtain count data under two different scenarios.
3. Analyze summary statistics before and after taking standardizing for Assembly #1
4. Analyze summary statistics before and after taking standardizing for Assembly #2
5. Comment on the differences before and after standardizing. Explain.
6. Comment on the differences between Assembly#1 and Assembly#2. Explain.

DATA ANALYSIS PROJECT (GST 200) Long Version.

Goal: To use descriptive statistics to describe and compare quantitative associations between factors, to determine the influence of assembly in defining binary factors given either ordinal or continuous data, to determine the influence of a binary confounder, and to determine the influence of sample size and chance on student-generated data.

Computer Inputs from each student:
Specify outcome (binary) plus predictor and confounder (ordinal or continuous).
Specify distribution of each factor: Minimum, expected value and maximum.
Assembly1: Specify cut-point to convert predictor and confounder from continuous to binary.
Assembly2: Select a different cut point for either the predictor or the confounder.
Note: All this data is presented in the grayed cells in the top part of the Excel page.

Computer Output: Sample is randomly generated and counts are summarized into tables.
Note: This data is presented just below the top part of the computer output page (next page).

STUDENT ACTIVITY: Translate summary counts into percentages.
See example on next page.

Data 1.1: Describe percentages under definitions in Assembly 1 and in Assembly 2.

Data 1.2: "Adjust" for confounder influence (standardize) in Assembly 1 and Assembly 2.

Data 1.3: Calculate the **margin of error** for the various percentages. (Future version)

REPORT:

- **Compare*** common-part percentages *within* Assembly 1 and within Assembly 2.
- **Compare*** similar percentages *between* Assembly 1 and Assembly 2.
- **Describe** the influence of assembly*.
  * Use percentage & likely grammar for comparisons.

Within Assembly 1,

- **compare*** common-part percentages *between* un-standardized and standardized.
- **describe** the influence of taking into account the influence of a confounder*.
- **calculate** what percentage of original difference is explained by confounder
  * Use percentage & likely grammar for comparisons.

For both Assembly 1 and Assembly 2,

- **determine** if associations are statistically significant both before and after adjust.
- **Summarize*** influence of confounding on predictor-outcome association on Assembly 1 and 2 in terms of strength of association and statistical significance.
  * Use percentage & likely grammar for comparisons.

FORM: Write up your results in a report. Identify any relevant plausible source of bias.

	Student Name:	**Leah Dodds**		Group	**College students**		GST 200 Spring Day 2007-8		Ver 3b
	Outcome (binary)		Min	Max		0 =	1=		
Q1.	**Do you attend musical events?**		0%	100%	**50%**	don't attend	attend musical events	Two groups	
Claim	**High-GPA students are more likely to attend musical events than low-GPA students**							ASSEMBLY	
						Smaller values	Bigger values	Min & above	Second
	Predictor (Multi-value)		Min	Max	Expected	0 =	1=	Cut-point for 1's	Cut-point
Q2.	**What is your GPA?**		**0**	**40**	**30**	**low GPA**	**high GPA**	**35**	**30**
Q3.	**How many years have you taken music lessons?**		**0**	**16**	**4**	**little music training**	**lots music training**	**6**	**6**
	Confounder (multi-value)								

	SCENARIO #1: COUNT TABLES (ALL and OUTCOME1)				OUTPUT		SCENARIO #2: COUNT TABLES (ALL and OUTCOME1)				
	B	C	D	E	F	G	H	I	J	K	
13	**Count of ALL**	**Predictor0**	**Predictor1**			13	**Count of ALL**	**Predictor0**	**Predictor1**		13
14	Row Labels	low GPA	high GPA	Grand Total		14	Row Labels	low GPA	high GPA	Grand Total	14
15	lots music training	15	8	23	**CNFDR1**	15	lots music training	7	16	23	15
16	little music training	62	15	77	**CNFDR0**	16	little music training	42	35	77	16
17	Grand Total	77	23	100		17	Grand Total	49	51	100	17
18						18					18
19	**attend musical events**	**Predictor0**	**Predictor1**			19	**attend musical events**	**Predictor0**	**Predictor1**		19
20	Row Labels	low GPA	high GPA	Grand Total		20	Row Labels	low GPA	high GPA	Grand Total	20
21	lots music training	9	6	15	**CNFDR1**	21	lots music training	5	10	15	21
22	little music training	17	8	25	**CNFDR0**	22	little music training	13	12	25	22
23	Grand Total	26	14	40		23	Grand Total	18	22	40	23

ANALYSIS

			SCENARIO #1:		SCENARIO2	
Q1	What percentage of HiGPA students attend music events?	P(O1|P1)	+D23 / +D17	**60.9%**	+J23 / +J17	**43.1%**
Q2	What percentage of LowGPA students attend music events?	P(O1|P0)	+C23 / +C17	**33.8%**	+I23 / +I17	**36.7%**
Q3	What is the difference in percentage points?	P(O1|P1) - P(O1|P0)	Q1-Q2	**27.1 PctPt**		**6.4 PctPt**
Q4	Is assoc. between GPA and attend ME positive?	P(O1|P1) > P(O1|P0)	Yes/No	**yes**		**yes**
	READINESS					
R1	What % of HiGPA,HiMusicTrain students attend ME?	P(O1|P1C1)	+D21 / +D15	**75.0%**	+J21 / +J15	**62.5%**
R2	What % of HiGPA,LowMusicTrain students attend ME?	P(O1|P1C0)	+D22 / +D16	**53.3%**	+J22 / +J16	**34.3%**
R3	What % of LowGPA,HiMusicTrain students attend ME?	P(O1|P0C1)	+C21 / +C15	**60.0%**	+I21 / +I15	**71.4%**
R4	What % of LowGPA,LowMusicTrain students attend ME?	P(O1|P0C0)	+C22 / +C16	**27.4%**	+I22 / +I16	**31.0%**
R5	What percentage of all students are HiGPA?	P(P1)	+D17 / +E17	**23.0%**	+J17 / +K17	**51.0%**
R6	What percentage of all students have lots of music training?	P(C1)	+E15 / +E17	**23.0%**	+K15 / +K17	**23.0%**
R7	What percentage of HiGPAs have lots of music training?	P(C1|P1)	+D15 / +D17	**34.8%**	+J15 / +J17	**31.4%**
R8	Is assoc. between GPA and music training positive?	Yes if R7 > R6	Yes/No	**Yes**	Yes/No	**Yes**
	STANDARDIZATION					
S1	What is standardized percentage of HiGPAs who attend ME?	R2 + [R6*(R1-R2)]		**58.3%**		**40.8%**
S2	What is standardized prcntg of LowGPAs who attend ME?	R4 + [R6(R3-R4)]		**34.9%**		**40.3%**
S3	What is the difference in the standardized percentages?	S1-S2	+S1-S2	**23.4 PctPt**	+S1-S2	**0.5 PctPt**
S4	Is standardized assoc. between GPA and attend ME positive?	Yes, If S1>S2	Yes/No	**Yes**	Yes/No	**Yes**
S5	Is GPA-AttendME association same after std as before?	Yes, if S4 = Q4	Yes/No	**Yes**	Yes/No	**Yes**
S6	What is the influence of standardization on Association?		in/decrease, reduce	**Reduce**	in/decrease, reduce	**Reduce**
S7	**What % of original difference is explained by MT?**		X% or N/A	**13.6%**	X% or N/A	**92.0%**

APPENDIX 1: Rationale for designating GST 200 as a QF course:

As a QF course, GST 200 addresses three of the four "q" skill areas. These are:

- Mathematical relationships.
  - Graphical representations include charts of counts, rates and proportions, graphs of association and confounding.
  - Symbolic representations include formulas for percentage change, percentage of difference explained by a confounder, percentage of cases attributed to a predictor, number of cases attributed to a predictor, margin of error, sample size and confidence intervals (See attached list of such formulas).
  - Numerical representations focus on arithmetic comparisons (difference, times ratio and percentage change) and on various models of a relationship: linear model and weighted averages.
  - Ordinary English representations of arithmetic relationships. Even though algebra is both concise and precise most people don't speak algebra. Students learn to express arithmetic comparisons in ordinary English. They learn that 8 is 4 times 2, but is only 3 times (300%) more than 2. They learn that 6% is not 2% more than 4%. They learn that one cannot compare temperatures as a ratio or a percentage change (e.g., it is 20% hotter today than last week). At least a fifth of the course focuses on the description and comparisons of ratios (proportions, percentages and rates) using ordinary English to describe and compare conditional probabilities. Students use an on-line program to practice decoding written statements (reading) and another cutting-edge program to practice writing such statements in ordinary English. The latter program reads ordinary English and gives students immediate feedback on the accuracy of their answer.

- Statistical relationships.
  - Data analysis is a substantial component of the course, but the data analyzed tends to be summary data presented in tables, graphs and statements (Blacks are less likely to commit suicide than whites).
  - Elementary probability is a larger part of this course than it is in most introductory statistics courses but again the focus is on ordinary English instead of algebra. Students must distinguish between the probability of A given B and the probability of B given A. E.g., P(A|B) versus P(B|A). But they must do so in ordinary English. They must see that "the percentage of men who are smokers" is the inverse of "the percentage of men among smokers." They must see that "Widows are more likely among suicides than widowers" is the inverse of "Widows are more likely to suicide than widowers." Students must be able to identify which of these two statements has the larger percentage. (a) the percentage of male smokers who are runners (b) the percentage of smokers who are male runners. [In this case if one is larger, it must be the first.].
  - Association vs. causation. Students spend considerable time distinguishing association from causation. They learn that statements such as "Women who eat nuts have a lower risk of breast cancer" do not imply causation. They realize that getting breast cancer is not repeatable, the study cannot be rerun using the same subjects so there is no way to switch a given subject and see what would have happened if they had changed between not eating nuts and eating nuts.

- Logical analysis.
  - Deductive reasoning. Students study some of the basic deductive arguments in statistics. These are not the basic forms of Aristotelian syllogisms. They learn that if two groups have different rates or percentages of a given outcome, then the excess in the larger can be attributed to the associated group and that when that excess is divided by the larger

rate, the resulting proportion is the percentage of those outcomes attributable to the associated group. For example, if 1 reader in a 1,000 of car and sport magazines gets pregnant while 20 readers in a 1,000 of home and fashion magazines gets pregnant, then 95% of the pregnancies among readers of home and fashion magazines can be attributed to reading home and fashion magazines. ☺

- Arguments: About a fifth of the course involves logical analysis involving arguments. Students are given news stories, press releases and statistical studies. They must identify the point of the article and evaluate the strengths and weaknesses of the statistics used as evidence for the truth of the conclusion. In each case, they examine how the statistics used might be influenced context (confounding), assembly (choice of definitions and presentation), randomness (chance) and error or bias.
- Fallacies. Identifying statistical fallacies is a key element of this course. Students learn that a medical test with 99% accuracy may be less than 50% accurate in predicting disease when someone tests positive. They learn that just because most car accidents occur with 25 miles from home" does not mean "they are safer when driving more than 25 miles from home."
- Counterexamples. Counterexamples are perhaps the central part of the course. Students evaluate the strength of inductive arguments using alternative explanations – hypothetical thinking. For example, suppose a study finds that 69% of kids in middle school are involved in 'bullying.' A statistically-literate reader must realize that 'bullying' is a soft term – there is no generally accepted definition. Such a reader must think hypothetically about how this term could have been defined to change the size of the percentage.

APPENDIX 2: QA COURSE REQUIREMENTS

To qualify as a QA course it must include instruction in at least three of the following "q" skill areas (with at least two in depth):

- Mathematical relationships – graphical, symbolic, and numerical representations; proportions, percents, estimation.
- Statistical relationships – data analysis (including graphical analysis), elementary probability.
- Algebraic Relationships – modeling, functions, algebraic representations.
- Logical Analysis – deductive reasoning, fallacies, arguments, counterexamples.

And

- A required significant quantitative reasoning project.

Student learning objectives include:

- Prerequisite arithmetic and algebra skills at the level of Math Placement Group 3;
- Foundational mastery of the "q" skills; and
- Demonstration through the final project that the student can
  - Pose quantitative questions;
  - Make and communicate reasoned choices as to applicable quantitative methods;
  - Demonstrate the ability to apply quantitative methods; and
  - Use the results to reason and articulate answers/conclusions.