Alexander Coward <alexander.coward@gmail.com>

Plans going forward

Arthur Ogus <ogus@math.berkeley.edu> Tue, Dec 3, 2013 at 4:00 PM
To: Alexander Coward <alexander.coward@gmail.com>
Cc: "F. Michael CHRIST" <mchrist@berkeley.edu>, Arthur Ogus <ogus@berkeley.edu>, Thomas Scanlon <scanlon@math.berkeley.edu>, Ole Hald <hald@berkeley.edu>

Dear Alexander,

Thank you for the frank and enlightening discussion today. I'm attempting to write a summary of what I think we should/need to do going forward. Still open to adjustments.

1. I'll try to find an office for you which is more centrally located. Kathy says that 987 Evans may be available as of January. It's a small office, I think, but with a better location. Please take a look and see if you find it satisfactory; otherwise we will have to wait until next Fall.

2. Please do write your own response to the two assessments, including in particular any inaccuracies or follow-up materials you think relevant. You may want to speak or write to Mike or Ani directly as well.

3. Please consult very closely with Ole Hald in the course of making up your final examination for Math 1A, and also with regard to the grading scale when the exam is finished.

4. Please consult with Tom Scanlon about Math 16B which you will be teaching this Spring. In particular we would like to see the following:

   a. A weekly selection of homework problems that students are assigned and encouraged somehow to do, and which the GSI's will know to be prepared to help students with during section. Encouragement could take the form of having students be expected to work these problems on the board during section, or to work them (or variants therefore) on quizzes. Or it could be that there is an online system of help for this set of problems so students can self-assess their understanding. I believe Paul Vojta may have some technology to help with this. It may be difficult to set up but I expect it would be of considerable educational value.

   b. Pay careful attention to the weekly schedule that has been worked out by the faculty for Math 16B, making sure that the required topics are covered in approximate accord wight his schedule. This schedule is typically made available to students in advance; if you have strong educational objection to that and can convince Tom, it could be made available afterwards. Ideally there should be a website.

   c. Put together a body of teaching material (teaching portfolio) which will allow other faculty to understand, in so far as possible, what you covered and how. In particular, lecture notes, emails you send out (wouldn't web dissemination be better?), examinations, your own comments on what worked and didn't.… Here I don't want to create an extra burden beyond what is generally expected of faculty, but this will be important in helping us understand exactly what you taught and also in helping us adapt some of your ideas.

5. We will try to figure out how students from your sections perform in Math 1B next semester if we can.

Does this sound feasible?

Alexander Coward <alexander.coward@gmail.com>

Follow-up

2 messages

Alexander Coward <alexander.coward@gmail.com> Mon, Oct 13, 2014 at 11:01 PM
To: Arthur Ogus <ogus@math.berkeley.edu>

Dear Arthur,

Thank you for the phone call today. Here is a follow-up note relating to our conversation

1) You mentioned that statistics had been collected for how my students from 1A in Fall 2013 performed in Math 1B compared to students who had not taken Math 1A with me, and that my students had performed better. I was not aware that this had been done. Please could you let me know a summary of the statistics that were collected and the methodology that was used.

2) You asked me for my reaction to the evaluation plan. I mentioned that I thought that along with the other four items (faculty observations, course materials, student evaluations, and `other') I thought it was important to include a section for continuing professional development. You said that you didn't want to change the document at this stage. I'm afraid I must protest about this. You indicated to me that the document was in draft form and that I would have the opportunity to provide input. Since you indicated that you agree that professional development is indeed a legitimate thing to look at, and that it would be taken into consideration, it seems to me important that it be included as an item to consider, and that this be recorded in the evaluation plan. Otherwise the evaluation plan is not accurate.

3) You mentioned that unless I insist on sticking to the timeline you outlined in your memo of April 18th, in which my re-appointment decision is scheduled to be made in October 2014, the department would now like to make this decision in February of 2015, after the department has made its decision about whether to continue with long-term lecturers. I don't think it would be sensible for me to insist upon a decision being made in October since you indicated that the department is not ready to make a well informed decision. In fact, while it is disconcerting for me to adjust my expectations, and I would have appreciated it if this information could have been shared sooner, I do in fact agree with you that it makes more sense to decide about whether to re-appoint individual lecturers after the broader decision about whether to continue with long-term lecturers in general.

I have some further questions I forgot to mention:

1) I have recently been observed by Hald, Evans, and Rezakhanlou. Have reports of these observations been made? If so, will I be sent them?

2) On page 2, item 3, of the evaluation plan you sent me, it states that materials from the Fall 2013, Spring 2014, and Fall 2014 semesters will be looked at in connection with my re-appointment decision. Since you said that in fact the department would like to consider my performance and materials relating to Math 1B in Spring 2015, will the evaluation plan be updated to correct this inaccuracy?

There is one further thing that I must mention. You were keen for me to state what my preferences are regarding the timeline for my re-appointment decision. It is really not for me to be making these kinds of decisions, because I am not well informed about how far the department has progressed along the evaluation plan that it intends to follow. However it is very important to me that information shared with me, both in writing and in person, is accurate. I'd really appreciate it if care could be taken to make sure that I am not given inaccurate information in the future. On the basis of our conversation, my understanding is that a decision on whether to continue with long-term lecturers will be made first, and then a decision regarding my re-appointment will be made afterwards, with this process scheduled to be completed by the end of February 2015. If this is not accurate, then please let me know.

Thanks again for the call. I really appreciate the hard work you're putting in to make the lecture experiment a success.

All my best,

Alexander

Arthur Ogus <ogus@math.berkeley.edu> Tue, Oct 14, 2014 at 10:14 AM
To: Alexander Coward <alexander.coward@gmail.com>
Cc: "F. Michael CHRIST" <mchrist@berkeley.edu>, Craig Evans <evans@math.berkeley.edu>

Dear Alexander,

Thank you for the summary below. I will attempt to clarify and adjust some of your points.

On Oct 13, 2014, at 11:01 PM, Alexander Coward <alexander.coward@gmail.com> wrote:

> Dear Arthur,
>
> Thank you for the phone call today. Here is a follow-up note relating to our conversation
>
> 1) You mentioned that statistics had been collected for how my students from 1A in Fall 2013 performed in Math 1B compared to students who had not taken Math 1A with me, and that my students had performed better. I was not aware that this had been done. Please could you let me know a summary of the statistics that were collected and the methodology that was used.

I am only authorized to send you the summary report, which I will attach below.

>
> 2) You asked me for my reaction to the evaluation plan. I mentioned that I thought that along with the other four items (faculty observations, course materials, student evaluations, and `other') I thought it was important to include a section for continuing professional development. You said that you didn't want to change the document at this stage. I'm afraid I must protest about this. You indicated to me that the document was in draft form and that I would have the opportunity to provide input. Since you indicated that you agree that professional development is indeed a legitimate thing to look at, and that it would be taken into consideration, it seems to me important that it be included as an item to consider, and that this be recorded in the evaluation plan. Otherwise the evaluation plan is not accurate.

Although I did express my personal opinion during that phone conversation, that opinion is not official. The evaluation plan was prepared in consultation with several faculty.
I am willing, based on standard department practices, to modify the written plan in a different manner, by encouraging the candidates to submit their own self-statement, which
could include a description of professional development, as well as other information of their choice.

>
> 3) You mentioned that unless I insist on sticking to the timeline you outlined in your memo of April 18th, in which my re-appointment decision is scheduled to be made in October 2014, the department would now like to make this decision in February of 2015, after the department has made its decision about whether to continue with long-term lecturers. I don't think it would be sensible for me to insist upon a decision being made in October since you indicated that the department is not ready to make a well informed decision. In fact, while it is disconcerting for me to adjust my expectations, and I would have appreciated it if this information could have been shared sooner, I do in fact agree with you that it makes more sense to decide about whether to re-appoint individual lecturers after the broader decision about whether to continue with long-term lecturers in general.

Let me clarify further. The decision about whether to reappoint you is not, strictly speaking, dependent on the decision about the department's long-term use of lecturers.
The department could decide to reappoint you for one more year even if decides to discontinue the use of lower-

divisionlecturers, or it could decide not to reappoint you even it does decide to continue the use of lecturers, (and there are two more obvious possibilities). Thus in a strict logical sense, the issue of your reappointment is independent of the issue of the long-term use of lecturers. Of course, in a strategic sense, the issues are in fact linked.

Contract Lecturers are typically appointed year-by-year, depending on the performance of the individual, the needs of the department, and its budget. Typically such decisions are made in the spring. However, since we were aware that you might be interested in applying for a position with a fall deadline, we thought it important to give you information about our decision on your appointment this fall. We are willing to make the decision now if you so desire, for your convenience. On the other hand, in spring, the department will have more information (about your performance, about its needs and budget, and also its long-term plans with regard to lecturers). But the same logical independence will still apply then, and, depending on how the discussion goes, it may be necessary to decide on your reappointment before the long-term issue is settled.

Thus what I need from you now is a statement of preference, based on your needs (not the department's) of when you would like a decision on your reappointment for next year.

>
> I have some further questions I forgot to mention:
>
> 1) I have recently been observed by Hald, Evans, and Rezakhanlou. Have reports of these observations been made? If so, will I be sent them?

No written reports have yet been submitted.
>
> 2) On page 2, item 3, of the evaluation plan you sent me, it states that materials from the Fall 2013, Spring 2014, and Fall 2014 semesters will be looked at in connection with my re-appointment decision. Since you said that in fact the department would like to consider my performance and materials relating to Math 1B in Spring 2015, will the evaluation plan be updated to correct this inaccuracy?

Is this something I said on the phone?
>
> There is one further thing that I must mention. You were keen for me to state what my preferences are regarding the timeline for my re-appointment decision. It is really not for me to be making these kinds of decisions, because I am not well informed about how far the department has progressed along the evaluation plan that it intends to follow. However it is very important to me that information shared with me, both in writing and in person, is accurate. I'd really appreciate it if care could be taken to make sure that I am not given inaccurate information in the future. On the basis of our conversation, my understanding is that a decision on whether to continue with long-term lecturers will be made first, and then a decision regarding my re-appointment will be made afterwards, with this process scheduled to be completed by the end of February 2015. If this is not accurate, then please let me know.

Please see the discussion above which attempts to clarify and correct this point.
>
> Thanks again for the call. I really appreciate the hard work you're putting in to make the lecture experiment a success.
>
> All my best,
>
> Alexander

I have been informed that this analysis was done by Prof. Denis Auroux, former Co-Chair for Strategic Planning of the UC Berkeley Mathematics Department.

Here is the result of the grade comparison:

"We have reviewed the performance in Math 1B in Spring 2014 of students who had taken Math 1A in Fall 2013 with Alex Coward vs. those who had taken it with another instructor. The average final grade obtained in Math 1B by those students who had taken 1A with Coward was slightly higher than for the other students. However it is not clear that the difference is statistically significant or can be attributed to teaching effectiveness, as it could simply reflect a different pool of students. In any case it appears that the students who took 1A with Alex Coward were, on average, at least as well prepared for Math 1B as those who took it with someone else."

Best regards,

Arthur

Alexander Coward <alexander.coward@gmail.com>

evaluation plan

1 message

Arthur Ogus <ogus@math.berkeley.edu> Wed, Oct 29, 2014 at 10:37 AM
To: Alexander Coward <coward@math.berkeley.edu>

Dear Alexander,

I've just returned from a meeting with administrative experts, who gave me the go ahead to forward to our evaluation plan
(which I may already have done), as well as Philip Stark's analysis of the data on the grades of students in Math 1B.
I regret the administrative delay.

It is now our responsibility to decide whether or not to reappointment you for the academic year 2015-16, as we had agreed last April. How this will be done is explained in the attached plan. As I told you earlier, you have the right to submit your own self-evaluation, which the committee will then consider in formulating its report. Please let me know if you need more time to submit such a document and if you are willing to allow us some more time to take it into consideration.

Here are the promised documents.

Best regards,

Arthur Ogus
Professor of Mathematics
Chair, Department of Mathematics
University of California
Berkeley, California
ogus@math.berkeley.edu
www.math.berkeley.edu/~ogus

2 attachments

evaluationplan.pdf
47K

from Stark.rtf
3K

Dear Arthur--

I performed a stratified permutation test of the null hypothesis that the three sections of 1A performed comparably in 1B. More specifically, the null hypothesis is that in each section of 1B, the labeling of students as having come from each section of 1A is as if at random, holding the number of students labeled with each of those 1A sections fixed.

The test statistic was the following:
For each section of 1B, calculate the three average grades corresponding to each of the three 1A sections. Take the standard deviation of those three numbers. Add those numbers for the three sections of 1B.

If students from different sections performed quite differently, we would expect this sum to be large. If they perform similarly, we would expect this sum to be small. How big is "big?"

To assess whether the observed sum was surprisingly large, I compare it to the distribution of values that result from randomly re-labeling the students in each section of 1B as having come from the three sections of 1A (keeping the number in each section fixed). In 100,000 random relabelings, the sum was at least as large as the sum for the actual data 13.3% of the time. That is, the p-value of the null hypothesis is about 13.3%. This would not ordinarily be considered "statistically significant." It is evidence that the sections of 1A are not quite alike, but not strong evidence.

Since we do not think that students selected which section of 1A to take "as if at random," differences other than the amount that students learned in 1A could account for part or all of the observed difference in their performance in 1B. There is no _statistical_ basis for concluding that any differences in the 1B performance--to the extent that such differences exist--are attributable to differences in what they learned in 1A.

Regards,
Philip

Students from my Math 1A class performed an average of 0.17 grade points better in Math 1B, compared to students who took Math 1A with another instructor. The overall sample size is approximately 1000 students.

Note that the test statistic used by Stark implicitly compares my students to more of my students, since I taught two of the sections of Math 1A being referred to here.

Report on A. Coward

A. Classroom performance. We each (except Ogus) attended one of Alexander Coward's classes this semester, and in addition M. Christ, D. Auroux and A. Adhikari visited his classes during the 2013-14 academic year. Write-ups on these visits are in the appendices. We have in addition interviewed 5 GSIs from AC's Math 16B last spring and 16A this semester, and have reviewed his course evaluations from 2013-14. The answers following are culled from these various sources.

Rigor and correctness of material presented. AC seemed to be careful and correct in the lectures we and others observed.

Appropriateness of level of material presented. This is hard to assess from one-time visits, but AC's classroom discussions seemed acceptable, if not at a very high level, for the Math 16A,B classes. However, there were some questions raised about the level of the presentation in the Math 1A lectures.

Interaction with students, level of attention paid, reaction to questions. Extremely outstanding in all the classes visited. In each of the classes, AC developed a really great rapport with the students. Attendance in his classes is very good.

Projection of voice, clarity and visibility of writing. Very good in all the lectures visited.

Special teaching techniques or technology. AC makes good use of the document camera in class, to project pages from the textbook for class discussion and to project students' answers to in class math questions.

Organization, clarity of statement and review of main goals. Some of the visiting faculty commented that the overall goals of the day's lecture should have been made clearer. Several others we have interviewed have stated that AC often seems unorganized in the class room.

Clarity of presentation. Good, but see the more detailed comments above.

B. Course materials. We have also gathered various information about grade distributions for AC's classes during 2013-14, materials provided on the course websites, etc. The earlier reports of M. Christ and D. Auroux also address some of these issues.

Syllabus. AC has not posted a day-by-day syllabus for any of his classes, but the sections to be covered each week are indicated in the weekly assignments on the course webpages.

Homework assignments. This is a major issue. AC did not collect written homework assignments for his Math 1A classes during Fall, 2013. He was advised to do so, and during the Spring, 2014 semester assigned all of the problems in the various sections. During the Fall, 2014 he has assigned selected problems from each section. However students are not required to hand in their written solutions to homework problems; instead they show their work to GSIs, who sign off on a cover sheet.

Level of examinations, including reviews of final examinations as available. We have reviewed various midterms and final exams, the levels of difficulty of which seem appropriate. We attach to this report copies of some of his exams. Apparently AC does not have the GSIs give quizzes in sections.

Value of lecture notes or other materials. We are not aware that any such materials were provided, except for some hand-witten notes he emailed to students during Math 1A.

Grading. AC's stated grading policies seem fine, and the class grades he gave during the 2013-14 were neither too generous nor too stingy.

Interaction with the GSIs. These GSIs report that AC has provided very little detailed guidance about running their sections, and many expressed concern about the practice of not collecting homework.

Office hours. His office hours are not stated on the course website for this semester. So we asked and found out that they are 4-5 on Mondays, 10:15-12 on Wednesdays and Fridays. In our view these are an inadequate number of office hours, inadequately published.

Course website, online materials, and online communication with students. AC's course websites seem fine, posting the grading policies, homework assignments, etc. It would be very good to post also a day-by-day syllabus for each class.

Intellectual content. It has proved difficult for us to judge the content in the various courses really accurately, but what evidence we have found is incorporated into the other remarks in this document.

Student Evaluations Coward's numerical scores were extremely high, with scores of 6.4, 6.6 in 16B, for example, and very few low

scores. Nevertheless, many students complained about poor organization of the structure of the course and especially about lack of clarity in homework assignments; some students specifically wished that homework had been in some sense required.

Learning Outcomes An attempt was made to measure the effectivity of Coward's teaching by comparing the performance in Math 1B of the students who took Math 1A from Coward with the students who took Math 1A from [redacted] the previous semester. The data indicate that Coward's students did on the average at least as well as [redacted] and perhaps slightly better. Although the positive difference is not judged to be statistically significant, the fact that his students fared well can be viewed as a positive sign. There is also informal evidence that his attention to careful writing of arguments has had a beneficial effect on his students.

C. Summary.

1. AC is extremely dynamic and interactive in the big lectures, and the students really like him and his courses.

2. There are however many reports that AC is sometimes disorganized in class, and so we are somewhat concerned that he is able to effectively cover all the material at an appropriately high level.

3. AC's policy of not systematically collecting homework nor having quizzes in sections is at variance with usual practices in the department.

4. AC has also had trouble providing effective guidance to his GSIs, both in guidance about how they should conduct their courses and in grading examinations. This has been a potentially serious problem.

5. AC has not yet made an attempt to share his teaching methods or materials with other instructors of lower-division courses. We had hoped that this would be a useful contribution to the department.

Craig Evans, Ole Hald, Fraydoun Rezakhanlou, Arthur Ogus (Hald points out that he did not participate in, and therefore does not endorse, some aspects of this report.)

Best,

Arthur

Alexander Coward <alexander.coward@gmail.com>

Some questions

5 messages

Alexander Coward <alexander.coward@gmail.com> Wed, Jul 1, 2015 at 2:05 PM
To: stark@stat.berkeley.edu
Cc: Craig Evans <evans@math.berkeley.edu>

Dear Chair Stark,

In the Fall of 2014 you were asked, in connection with my re-appointment evaluation as a Lecturer in the Mathematics department, to provide your expert commentary on whether the score gain my students exhibited the following semester, which averaged 0.17 grade points, was statistically significant.

I have attached an analysis that I received on October 29th 2014 from Arthur Ogus, then chair of the Mathematics department, apparently signed by you.

In this analysis, you state the following:

"The test statistic was the following: For each section of 1B, calculate the three average grades corresponding to each of the three 1A sections. Take the standard deviation of those three numbers. Add those numbers for the three sections of 1B."

Thus, you divided my students into two groups corresponding to my 8am and 11am classes, and then used a test statistic that penalized consistency in performance between these two groups. You reported that this resulted in a p-value of 0.13 and concluded that the difference was not statistically significant. This seems to be a very strange way to evaluate whether my students' score gain of 0.17 grade points was statistically significant.

I am afraid to say that I have since discovered there was foul play in my Fall 2014 re-appointment process, including, but not limited to, the suppression of student evaluation data from my personnel file, a topic I know you take an interest in. Given this, I am concerned that the attached statistical analysis may have been part of a boarder deliberate distortion of the truth.

Therefore, I am afraid I must ask you the following questions:

1) Did you indeed write the attached document?

2) If you did write the document, what instructions did you receive?

3) If you did write the document, do you stand by it as being an appropriate use of statistics for its intended purpose?

I am sorry to have to ask you these questions is such a direct way. Perhaps you were not informed of the purpose of the analysis and this explains the strange choice of test statistic.

My reason for asking you these questions is that the matter of my students' performance in the following semester became an important question in my re-appointment decision, and I am absolutely resolved that any and all wrongdoing in that process must be either corrected or brought into the light for all to see.

All best wishes,

Alexander Coward

from Stark.rtf
3K

Philip B. Stark <stark@stat.berkeley.edu> Thu, Jul 2, 2015 at 10:04 AM
Reply-To: stark@stat.berkeley.edu
To: Alexander Coward <alexander.coward@gmail.com>
Cc: Philip Stark <stark@stat.berkeley.edu>, Craig Evans <evans@math.berkeley.edu>

Hi Alexander--

I did write that document (although not in that format).

I don't understand why you think the result is negative. As I point out in the note, there really isn't any basis for a statistical test, because students are not assigned to sections of the course at random.

I felt the test was appropriate for a variety of reasons.

I don't know what test you are proposing instead, but nothing can get around the fact that the allocation isn't random.

The (notional) test I used was designed to see whether there was evidence of any difference at all among the three sections. Differences, to the extent that they exist, might be due to teaching, to differences between the students who chose to take different sections, to time of day, or to other factors. There is weak evidence of difference among the sections. Whether any differences can be attributed to your teaching is another matter, one that statistics cannot speak to, especially given the absence of random assignment.

Cheers,
Philip
[Quoted text hidden]
--
Philip B. Stark | Professor | Department of Statistics | University of California
Berkeley, CA 94720-3860 | 510-394-5077 | statistics.berkeley.edu/~stark | @philipbstark

---

Alexander Coward <alexander.coward@gmail.com> Thu, Jul 2, 2015 at 4:17 PM
To: stark@stat.berkeley.edu
Cc: Craig Evans <evans@math.berkeley.edu>

Dear Chair Stark,

Thank you for the reply, and for confirming that you did write the document at hand.

I'm afraid your response does not make sense. On the one hand you say "As I point out in the note, there really isn't any basis for a statistical test," and yet just one sentence later you say "I felt the test was appropriate for a variety of reasons."

These sentiments plainly contradict each other.

The fact is that you did perform a test, you did obtain a p-value of 0.13, and you did say that the p-value was not small enough that it would ordinarily be considered "statistically significant." It is important that there is clarity on whether you believe that the test you performed was appropriate for its intended purpose of evaluating my teaching or not. If, on reflection, you believe that the test you performed was not appropriate, then you should say so.

I will ask you again:

Question 1: Do you stand by the test you performed as an appropriate use of statistics for its intended purpose of evaluating my teaching?

This is a very simple question, and given the importance this question had on the evaluation of my teaching that took place, I am entitled to a singular `yes' or `no' answer.

You also pose the question of what test I would suggest instead, and the answer to this question is informed by the purpose the test was used for. The analysis you completed was referred to, though not directly, in a report on my teaching. See attached. The report states: "The data indicate that Coward's students did on the average at least as well as Steel's and perhaps slightly better. [...] the positive difference is not judged to be statistically significant."

Thus the question at hand is whether the positive difference in my students scores' is statistically significant, not whether the difference in subsequent performance of students from the three 1A classes is statistically significant. For this reason an appropriate test statistic is one that measures the positive difference, as referred to in the report on my teaching, if that is the purpose for which the test is to be used.

This leads to the following question:

Question 2: If the test statistic had been the average score my students from 1A received in 1B the next semester, minus the average score of those students in 1B who took 1A with another instructor, using the same type of stratified permutation test as you performed originally, what would be the resulting p-value, that is the probability of observing a score gain at least that great by chance?

All best wishes,

Alexander Coward

[Quoted text hidden]

Report on A Coward.pdf
49K

Philip B. Stark <stark@stat.berkeley.edu> Fri, Jul 3, 2015 at 1:01 AM
Reply-To: stark@stat.berkeley.edu
To: Alexander Coward <alexander.coward@gmail.com>
Cc: Craig Evans <evans@math.berkeley.edu>

Alexander--

Without randomization, no causal inference.

Without randomization, the p-value at best "notional."

With the caveat that the p-value was notional and not a genuine p-value, I think the test was appropriate.

Even if there had been randomization, the classes differ in ways other than who taught them, so there's no way to attribute differences in performance--statistically significant or not--to differences in instruction.

I don't understand what you hope to accomplish with this communication. The report says your students performed better, on average, than those in the other sections. No amount data analysis now will make up for the fact that there was no randomization.

Regards,
Philip

Alexander Coward <alexander.coward@gmail.com>

Statistics

2 messages

Alexander Coward <alexander.coward@gmail.com> Mon, Jul 6, 2015 at 5:35 PM
To: Craig Evans <evans@math.berkeley.edu>
Cc: Philip Stark <stark@stat.berkeley.edu>

Dear Interim Chair Evans,

You will have seen my correspondence with Philip Stark. I am writing to formally ask you the following question:

Do you think the statistical analysis conducted by Philip Stark was appropriate for its intended purpose of evaluating my teaching?

Thank you for your attention in this matter.

Yours sincerely,

Alexander Coward

From Stark.pdf
431K

Craig Evans <evans@math.berkeley.edu> Tue, Jul 7, 2015 at 9:14 AM
To: Alexander Coward <alexander.coward@gmail.com>

Dear Alexander,

I am not a statistician and consequently am not at all qualified to have an opinion on this.

Please let me know if you'd like to meet with me again.

Best wishes, Craig
[Quoted text hidden]

Craig Evans is Interim Chair of the UC Berkeley Mathematics Department for 2015-2016. He is a Fellow of the American Mathematical Society and a Member of the National Academy of Sciences.

The question at hand, of whether one population of students performed statistically significantly better than another, is something that freshman undergraduates learn about at Berkeley in courses like Math 10, which Professor Evans helped develop.

Fall 2013 MATH 1A 001 LEC

Calculus

Schedule:

Section	Days/Time	Location	Instructor	CCN
001 LEC	TuTh 2-330P	155 DWINELLE	Non-Coward	53403

Units/Credit	Final Exam Group	Enrollment
4	5: TUESDAY, DECEMBER 17, 2013 8-11A	Limit:432 Enrolled:384 Waitlist:0 Avail Seats:48 [on 10/27/13]

Fall 2013 MATH 1A 002 LEC

Calculus

Schedule:

Section	Days/Time	Location	Instructor	CCN
002 LEC	MWF 8-9A	155 DWINELLE	COWARD, A	53448

Units/Credit	Final Exam Group	Enrollment
4	4: MONDAY, DECEMBER 16, 2013 7-10P	Limit:405 Enrolled:391 Waitlist:0 Avail Seats:14 [on 10/27/13]

Fall 2013 MATH 1A 003 LEC

Calculus

Schedule:

Section	Days/Time	Location	Instructor	CCN
003 LEC	MWF 11-12P	155 DWINELLE	COWARD, A	53490

Units/Credit	Final Exam Group	Enrollment
4	2: MONDAY, DECEMBER 16, 2013 1130-230P	Limit:430 Enrolled:413 Waitlist:0 Avail Seats:17 [on 10/27/13]

Total Coward Students: 391+413 = 804
Total Non-Coward Students: 384

Spring 2014 MATH 1B 001 LEC

Calculus

Schedule:

Section	Days/Time	Location	Instructor	CCN
001 LEC	MWF 9-10A	155 DWINELLE	RESHETIKHIN, N Y	53436

Units/Credit	Final Exam Group	Enrollment
4	4: MONDAY, MAY 12, 2014 7-10P	Limit:500 Enrolled:476 Waitlist:0 Avail Seats:24 [on 05/28/14]

Spring 2014 MATH 1B 002 LEC

Calculus

Schedule:

Section	Days/Time	Location	Instructor	CCN
002 LEC	TuTh 2-330P	100 LEWIS	AGANAGIC, M	53481

Units/Credit	Final Exam Group	Enrollment
4	2: MONDAY, MAY 12, 2014 1130-230P	Limit:287 Enrolled:285 Waitlist:12 Avail Seats:2 [on 05/28/14]

Spring 2014 MATH 1B 003 LEC

Calculus

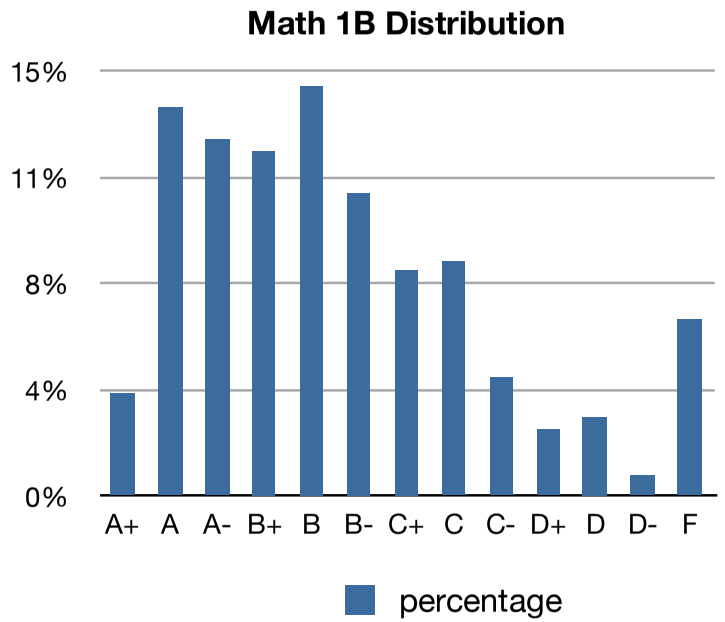
Schedule:

Section	Days/Time	Location	Instructor	CCN
003 LEC	MWF 12-1P	155 DWINELLE	HARRINGTON, L A	53526

Units/Credit	Final Exam Group	Enrollment
4	11: WEDNESDAY, MAY 14, 2014 3-6P	Limit:400 Enrolled:301 Waitlist:0 Avail Seats:99 [on 05/28/14]

Total Math 1B Sample Size: 476+285+301 = 1062
Pro-rated number of Coward Students: 1062*804/(384+804) = 719
Pro-rated number of Non-Coward Students: 1062-719 = 343

Grade	number	percentage	
A+	296	4%	
A	1119	14%	30%
A-	1026	13%	
B+	994	12%	
B	1181	14%	37%
B-	873	11%	
C+	651	8%	
C	675	8%	20%
C-	342	4%	
D+	193	2%	
D	228	3%	6%
D-	61	1%	
F	509	6%	6%
Total	8148	100%	



This program labels 1062 students with the department average grade distribution for Math 1B, as detailed on the previous page. It also randomly labels 719 of them as “Coward” and the remaining 343 “Non-Coward”. The improvement between the “Coward” and the “Non-Coward” students is then calculated, and this is done 100,000 times. The proportion of the time that this is greater than 0.17 is then displayed, which is found to equal 0.00869. Thus the probability, or p-value, of seeing a score gain of at least 0.17 grade points by chance was in fact less than 1%, contradicting the claim that my students’ score gain was not statistically significant. I encourage others to audit this code and this finding.

```
1  import random
2
3  def printlinebyline(l):
4      for i in range(len(l)):
5          print l[i]
6
7  def makestudentlist():
8      studentlist = []
9      for i in range(1062):
10         studentlist.append(0)
11     for i in range(1062):
12         studentlist[i] = ["NC","b"]
13     listsub = random.sample(range(1062),719)
14     for i in listsub:
15         studentlist[i][0]="C"
16     for i in range(191):
17         studentlist[i][1]=4.0
18     for i in range(191,329):
19         studentlist[i][1]=3.7
20     for i in range(329,456):
21         studentlist[i][1]=3.3
22     for i in range(456,605):
23         studentlist[i][1]=3.0
24     for i in range(605,722):
25         studentlist[i][1]=2.7
26     for i in range(722,807):
27         studentlist[i][1]=2.3
28     for i in range(807,892):
29         studentlist[i][1]=2.0
30     for i in range(892,934):
31         studentlist[i][1]=1.7
32     for i in range(934,955):
33         studentlist[i][1]=1.3
34     for i in range(955,987):
35         studentlist[i][1]=1.0
36     for i in range(987,998):
37         studentlist[i][1]=0.7
38     for i in range(998,1062):
39         studentlist[i][1]=0
40     return studentlist
41
42
43
44
45 def cowardaverage(l):
```

```
    total = 0
    for i in range(len(l)):
        if l[i][0] == "C":
            total = total + l[i][1]
    return total/719.0

def noncowardaverage(l):
    total = 0
    for i in range(len(l)):
        if l[i][0] == "NC":
            total = total + l[i][1]
    return total/343.0




bigger = 0.0
smaller = 0.0
for i in range(100000):
    studentlist1 = makestudentlist()
    if cowardaverage(studentlist1) - noncowardaverage(studentlist1) > 0.17:
        bigger = bigger + 1
    else:
        smaller = smaller + 1
p = bigger/(bigger+smaller)
print p
```