How one woman used regression to influence the salaries of many

Amanda Golbeck tells the story of Elizabeth Scott, the Berkeley statistics professor who spent two decades analysing inequities in academic salaries and advocating for change

Elizabeth L. Scott (1917–1988) devoted the last twenty years of her life to shining a spotlight on inequities in academic salaries. At the University of California, Berkeley in the early 1970s, she found significant differences between the salaries of men and women. On average within the same job title, women received over $300 per month less than men; for women directors, the shortfall was about $750 per month. Across the USA, she found that 80% of female faculty members were underpaid on average by at least $1500 per year when compared to men with the same qualifications, and this difference was not explained by career interruptions, it widened with the increase in years since completing a PhD, and it was greater in research universities and the sciences.

Scott discovered all this through a diligent process of data collection, cleaning and analysis. She followed up this work with efforts to convince or cajole university administrators to address the inequalities she uncovered. She had many successes, but would be disappointed to

Amanda L. Golbeck is a professor of biostatistics and associate dean for academic affairs in the Fay W. Boozman College of Public Health at the University of Arkansas for Medical Sciences, and a member of the *Significance* editorial board. She is the author of *Equivalence: Elizabeth L. Scott at Berkeley*, published by Chapman and Hall/CRC.

LEFT Elizabeth Scott addresses a 1970s meeting of the Lockheed Management Association in Sunnyvale, California (exact date unknown).
Courtesy of The Bancroft Library, University of California, Berkeley

learn that, even today, there has been no lasting change. Inequities in pay between men and women remain a problem.

Getting involved

Scott made no conscious plan to study salary inequities in academia. As a renowned statistics professor at Berkeley, she would rather have been working with Jerzy Neyman, one of the fathers of modern statistics and a major collaborator, on the question of whether cloud seeding causes rainfall. She would have preferred researching the clustering of galaxies with Neyman and C. Donald Shane, or partnering with other close colleagues to investigate if ozone depletion causes skin cancer.

However, the call for advocacy, when heard, can be difficult to ignore. For Scott, the call came in 1968, on the day she received a letter from the office of the president of the University of California. The letter asked if she would join a select group of other women leaders on the Berkeley campus to help search for a solution to the then-rampant problems of violence in US cities. Working together, these female academics soon became aware of problems beyond the charge of the group – problems that were specific not to US cities, but to people like them.

In one of many discussions, Scott articulated two important research questions: "Why are there so few women on the faculty?" and "Why are so few working toward and obtaining their PhDs?" At the time, 1200 of the 1245 members of the Berkeley academic senate were men; women comprised less than 4% of the entire group of instructors, assistant professors, associate professors, and professors. Scott spent a year co-leading the development of a trailblazing report to the academic senate that was grounded in hard facts such as these.

As she was working on this report, she learned there was to be a chapter in a forthcoming book that included some salary analyses using linear regression. This gave rise to another research question: to what extent are women earning less than men who have equivalent abilities and performance?

It was a question Scott set about answering using academic payroll data from the Berkeley Chancellor's Office. But there were methodological issues to first contend with.

The limitations of administrative data

Unbiased salary equity analyses depend on accurate and complete data. Administrative databases can have limitations for such analyses because they are set up to satisfy certain administrative needs – to generate pay-cheques, for example. Before Scott began her work on salary equity, payroll-related databases had not been used for assessing inequities. Scott accordingly looked at the databases, and the administrative reports generated from these databases, and concluded that they had many errors.

She worked fervently to improve the quality of these data: to get definitions of variables to be well recorded and consistent over time; and to document how data summaries were computed, including the instructions that programmers were given and the actual computer code that was used. She even checked reports by hand and submitted errors to the administration. She prodded and pressured. The administration was sympathetic. They even pledged to design new reports that would be usable for salary equity analyses, rather than try to adapt the existing payroll reports.

With the data in better shape, they could be useful in a regression model. But which variables could be predictive of a person's salary?

Predictor variables

Regression results depend on the specific models that are built. For salary equity models, like models of other phenomena, decisions had to be made about which predictor variables to include. The first salary equity models for academia were built using large national databases of over 20 000 faculty members. These studies could support the use of large numbers of predictor variables: 20, 30 or more.

In Scott's large studies, she notably incorporated many predictor variables related to academic productivity, including number of books, number of published articles, inclination towards research or towards teaching, hours taught per week, and number of paid consultancies. She believed that using a reasonably large number of variables in university-level studies was the answer to large amounts of spread around the regression lines. However, she was also concerned that too many variables could destroy the reliability of conclusions.

Ultimately, Scott decided that the number of variables should depend on the objective of the study. If the objective was to produce perfect estimates, then many variables should be used, including relevant interaction terms (such as the interaction of date of birth and number of published articles, or the interaction of sex, marital status, and age). However, if the objective was to flag cases of possible salary ▶

How good are the models, and what should be reported?

An important regression measure is R-squared (R^2). This measure has been described in a number of ways: the strength of relationship between variables; how well the predictor variables can predict the outcome; the amount of the variation in the response variable that is explained around its mean; or the goodness of fit of the regression.

Researchers like to see large values of R^2, the closer to 100% the better, as 100% means that all of the variance is explained, as opposed to 0%, which means that none of the variance is explained. For example, a university study that had only three predictors – years since degree, doctorate (yes/no), and a department increment (department-specific number of dollars per year since degree to be added or subtracted) – explained 52% of the variance in men's salaries. Scott's models of Berkeley salaries explained about 64% of the variance.

Scott argued that, even with relatively small models, her model predictions were so good and results so similar from year to year that it was unlikely the observed differences could be due to some factor(s) other than discrimination. She noted that, in a study to flag cases of possible salary discrimination, more women were flagged when there were fewer predictors, but the R^2 did not decrease much with fewer predictors. She doubted that, with more predictors, she could get the R^2 to be much larger than 64%.

The earliest regression modellers of salary equity published only correlations and corresponding significance probabilities. Scott thought the estimated magnitude of differences in salary between men and women should also be reported. In addition to the values of R^2 and correlations, she wanted to see all the regression coefficients reported, because they contained a lot of information that readers could use to evaluate individual salaries.

discrimination to be examined more closely by administrators, Scott reasoned that only a few simple predictors were needed. These predictors could be chosen among those that were easy to obtain: year of birth, highest degree, year of highest degree, sex, and minority status (yes/no).

The problem with rank

Scott worried that there could be discriminations embedded in some of the predictor variables. For one thing, research universities could be biased against hiring women. For another, there could be fewer research opportunities for women than men. Scott worried especially about "rank" as a predictor variable.

Academic positions are organised into ranks. Typical ranks on the tenure ladder are instructor (at the low end of the ladder), assistant professor, associate professor, and professor (at the high end). The problem with rank, Scott reasoned, is that women could be discriminated against in their initial rank assignment and in future promotion decisions. Indeed, when Scott looked at the data for Berkeley, she saw that among people of similar experience and performance, women tended to be appointed at lower salaries (ranks and salary steps within ranks) and advance more slowly than men. This was true year after year, even after the pressures of affirmative action. She found the problem to be worse in some departments than others. What was especially troubling to Scott was that salary deficits at the point of hiring would stay with a person "forever".

As Scott pointed out: "Try looking at the number of years required to go up in rank and the number of years required to go up a given amount in salary roughly corresponding to the change in rank (all for a given field, given rate of publishing, etc.) and you will see that you are looking at two sides of the same coin – women take longer to be promoted (and have smaller probability of being promoted) and women take longer to find their salaries increased by, say, $4000. You will tend to underestimate the underpayment of salary if you use rank as a predictor. Of course, this is true of many other predictors that involve discrimination: Having PhD, getting book published, etc. but rank appears to be worse and more directly tied (after all, [it] tends to be same persons determining both)."[1]

In other words, Scott warned that rank and salary have essentially the same bias; they are "too much the same thing". Discrimination in one, if present, is the same as discrimination in the other. Scott thought there was quite a bit of "monkey business" in studies that used rank as a predictor, and one should be on guard. She ended up feeling strongly that rank should not be included as a predictor variable.

The modelling strategy

For the purpose of analysis, Scott recommended that faculty be grouped into the unit that made the decision about their appointment or promotion. In most cases in academia, this unit would be the department. Accordingly, if the analyses indicated problems with salaries between men and women in a department, then that department should be more carefully watched, because that would be where the problems most likely originated. If a department was small, then it should be grouped together with similar departments.

In her regression modelling of salary, Scott recognised the utility of computing two sets of coefficients, one for men only, and the other for women only. This would allow for specific comparisons of coefficients across genders in each predictor variable (see "How good are the models, and what should be reported?", page 39).

However, Scott's primary modelling strategy was embedded in the question: "Does a person of one sex earn more than what was predicted for the salary of the other sex?" To answer this, she began with a regression equation that was built for male salaries. She then used it to predict salaries for women. For each woman, she compared their predicted salary – based on the male model – with their actual salary. The idea was to flag women whose actual salaries were lower than their estimated salaries, and to bring this to the attention of their administrations.

Scott's strategy was to combine statistical reasoning with common sense, and she challenged the administration to look at the issue of discrimination from multiple viewpoints

ABOVE Elizabeth Scott, pictured in 1972.
Photo republished with permission of the Institute of Mathematical Statistics from: Billard, L. and Ferber, Marianne A. (1991) Elizabeth Scott: Scholar, Teacher, Administrator. *Statistical Science*, **6**(2), 206–216.

Scott thought that the flagged underpayments should be further analysed to determine if there was clustering of underpayments in certain departments, department groupings, or colleges. If clusters were found, all women should be further investigated in those clusters, not just those who were flagged, because of the possibility of systemic discrimination. Furthermore, the reasons for the clustering should be sought so that remedies for the underpayments could be determined.

The flagging was imprecise, and the strategy was not perfect. It used a statistical method (regression) to identify a problem, together with a human method (administrative review) to determine if the problem should be remedied and, if so, what the remedy should be. But it did "serve the purpose of providing a rough estimate of underpayment, which could then be investigated further".

Scott, together with her colleague, the mathematician Mary Gray, realised there were limitations with this approach. If the salaries of flagged women were brought up to the regression line, there was still no guarantee that their discrimination would be completely remedied. Meanwhile, men whose salaries were below the regression line could also claim discrimination. There was also the concern that those remarkable women whose salaries happened to be above the regression line may still be underpaid relative to their academically equivalent male peers. Finally, administrative review would rely on qualitative factors, which could be subjective.

Scott and Gray therefore recommended a statistical remedy over administrative review. This would involve assigning the same salary adjustment to all women in a department or unit, rather than using human reasoning to determine individual salary adjustments to individual women. Since regression identifies class discrimination, they argued that the remedy should be for the entire class. Although the two encouraged institutions to use this approach, it was not widely adopted – probably because of the costs involved.

Powers of persuasion

Scott understood how difficult it would be to satisfy administrators that discrimination was present, especially using statistical modelling methods that depend on assumptions and conditions, and that require choices in specification. Administrators could criticise salary studies at their institutions for using small numbers of predictor variables or the "wrong" predictor variables; they could even point to the deficiencies of their own data to invalidate the studies.

Part of Scott's strategy to convince administrators was to combine statistical reasoning with common sense. She challenged the administration to look at the issues from multiple viewpoints. She would expose as fully as possible the statistical thinking that went into the modelling and the resulting limitations of the results; she would then challenge the responses of administrators with statistical reasoning. She also monitored the progress of affirmative action plans on her campus and politely but firmly wrote letter after letter to her administration, applying pressure for her work to appear in these plans. In so doing, she educated her administration and the Department of Health, Education and Welfare officials on the appropriate use of basic statistics in affirmative action and became a trusted advisor on such matters.

Another strategy used by Scott was to support the work of committees dedicated to advancing academic salary equity. She did

ABOVE Elizabeth Scott (pictured, far left) in conversation at a 1970s meeting of the Lockheed Management Association.
Courtesy of The Bancroft Library, University of California, Berkeley

this at both the local level, with Berkeley's Special Committee on Salary Equity, and the national level, with the American Association of University Professors (AAUP). In 1974 two AAUP committees, W (Women) and Z (Economics), formed a joint subcommittee to develop a do-it-yourself higher education salary evaluation kit. The AAUP was looking for a respected academic to develop the kit and provide technical assistance to users, and it did not take them long to find and commission Scott. The kit was published in 1977 and included a description of what kinds of computer inputs were needed, how packaged least squares regression computer programs could be used for the regression modelling, how computer outputs should be interpreted and evaluated, and what criteria and formulas could be used to correct the salaries.[2]

Scott and the AAUP were optimistic about the potential for the kit to help ensure approximately equal pay for approximately equal work. The kit was released with academic fanfare: there were many announcements in professional newsletters and journals. It was used over many years by universities and individuals across the USA to diagnose salary inequities, and it contributed to many impartial salary determinations. But, ultimately, it was not enough.

Staying involved

In 1978 Scott wrote to her colleague, the statistician William Kruskal at the University of Chicago, about salary equity and broader gender equity issues, saying: "I am indeed distressed by the present use of statistical methods in adversary situations. Partly, my difficulty is that here is an arena in which there is no effort to establish the truth or to get as close to the truth as we can. Rather, the attitude seems to befuddle the judge and to employ any misuse of statistics (such as using rank as a predictor) which might help your own side." In general, she thought there was "entirely too much chance in academic promotion decisions", and that there needed to be less "buttering up" of administrators and more attention to productivity and quality.

It is clear in retrospect that Scott's personal mission was to minimise the role of chance in promotion and salary decisions. In carrying out her mission, she herself admitted that she was depressed by the lopsidedness of the situation.

Despite her work, women continued to enter the academic ranks at lower salaries and they progressed up the ranks more slowly. She saw how university administrations could give assurances about correcting salary and other employment inequities and then drag their feet, and she was frustrated by it. The numbers were startling and did not seem to be getting any better, even with all the efforts toward affirmative action. Furthermore, a tightening job market worked against progress, and there was backlash from the men who were in competition for jobs and salaries. Scott did not expect to see a light at the end of the tunnel, at least not in her lifetime. "Women have a long way to go to attain equity, and many persons need to pay attention to the underlying problems," she wrote.

A few years ago, Mary Gray summed up the progress that had been made since Scott's day. She wrote that: "Women faculty are still paid less on the whole, there are still occasional regression-based studies, there are spot remedies, and often the very best women faculty continue to be underpaid."[1]

What can be done? The best hope, Scott thought, was to keep researching, monitoring, advocating and publishing the findings of these studies. Scott argued that women and minorities need to keep their eyes on the salaries in their units. She would be the first to remind us that, for public universities in the USA, faculty salaries are public information. ■

References

1. Golbeck, A. L. (2017) *Equivalence: Elizabeth L. Scott at Berkeley*. Boca Raton, FL: Chapman and Hall/CRC Press.

2. Scott, E. L. (1977) *Higher Education Salary Evaluation Kit: A Recommended Method for Flagging Women and Minority Persons for Whom There Is Apparent Salary Inequity and a Comparison of Results and Costs of Several Suggested Methods*. Washington, DC: American Association of University Professors.