Statistical Methods II
Statistics 211

General Comments on the Faculty Salary Data

The following are some general comments about issues related to the analysis of the project data. Note that, as always, I believe the strongest generalization of analysis results comes the statistical question is decided before looking at the data. If the statistical model used is based on data driven decisions, we are subject to problems due to multiple comparisons. In particular:

1. If we use the data to decide how to model the predictor of interest, we can easily inflate the type I error. For instance, in one data set that I sometimes use in this class, I have determined (by randomly permuting the data across dose groups– so the null hypothesis would be true) that if one considers dummy variables, quadratic, linear, dichotomous at dose greater than 0, or a combination of linear and the threshold term, the true type I error is about .15 rather than .05. Clearly this is not valid practice. Thus we have to be careful about data driven modeling of the predictor of interest.
2. If we use the data to decide which confounders/precision variables to model, there is a tendency to overfit your data. That is, some of those variables might just be modeling random noise in this sample, and thus the results might not generalize in the sense that the modeled variables are surrogate for the random noise rather than measuring the scientific quantity you think they are.
3. If we use the data to decide how to model the confounders/precision variables, we may also be fitting random noise, thereby overfitting our model.
4. If your model is overfit, you may have estimated the standard error lower than it really is. This will tend to inflate your type I error, though it is probably not as grievous an error as the problem with the predictor of interest.

In light of the above, I tend to spend a good deal of time trying to decide the best models to use before looking at the data.

1 Thinking About the Problem

The overarching goal of this project is to explore whether there is any evidence suggestive of discrimination by sex in the employment of the faculty at a single university (University of Washington). The specific aim was to explore whether we could find evidence of discrimination in the salaries paid to the faculty. Most of the problems that arise in the analysis of this data can be anticipated prior to looking at the data. That is, the more time spent thinking about the problem tended to mean that less time was required to analyze the data.

1.1 Sampling Scheme

The first thing that should be considered is the sampling scheme. First, we must consider whether we have a sample which is to be used to make inference about a population, or whether our data represent the population of interest. Even if these data represent information on all faculty in the nonmedical fields employed in 1995, I still believe that we are interested in making inference about the university's employment practices in general. That is, the 1995 cross sectional picture undoubtedly represents fluctuations around some general tendency. In some years the university may have been more lucky or less lucky in attracting top candidates from each sex. Because of such fluctuations, we do not think that inequality of salaries can be interpreted deterministically. Instead, we should interpret that inequality in terms of the variability of salaries within groups that are homogeneous with respect to other predictors. We are then generalizing to this university's employment practices in other similar years.

Secondly, we need to realize that in sampling only those faculty employed in 1995, we have only a snapshot of the faculty in that year. We do not have information on other faculty who were hired in previous years, but who left the university in the intervening years.

1.2 Factors Affecting Salary

I first find it useful to consider the factors that affect level of salary. The following are the general categories that I could think of.

1. Factors relating to the faculty member.
    - Skill: People with more innate talent for their job are generally more valued, and thus would be expected to receive higher salaries. We have no variable that measures this directly, though it might be reflected in rank.
    - Training: People who have invested more time in getting training would be of more value, and thus might receive higher salaries. *deg* (the degree obtained by a faculty member) might be indicative of this, though the meaning of a Ph.D. relative to some other degree might vary substantially across fields. Also, fifty years ago it was more frequently the case that faculty might not have Ph.D.s
    - Experience: The length of time a person has been working in a field may add to his/her value to the university. *yrdeg* will be an indication of this, providing it can be assumed that the time between receiving the degree and starting at the university was spent gaining experience. It should be noted that if women are more likely to take parental leave, then *yrdeg* might not be as indicative of the the time spent working in their field for them. *startyr* might be important if there were any aspects of experience specific to this university (as opposed to working in the field in general) that were important. *rank* will also be indicative of experience, as faculty in higher rank will generally have been in the field longer, and, at least in the case of tenured faculty, faculty who do not receive promotion are not retained.
    - Productivity: The number of publications, grants, copyrights, and patents (or productions in Fine Arts, etc.) would be indicative of a faculty member's productivity in non-educational endeavors, and evaluations of quantity and quality of teaching would indicate the faculty member's productivity in educational endeavors at the university. While it would seem clear that such productivity is likely to influence salaries, we have no direct measures of such activity. This productivity might be reflected in *rank*.
2. Factors relating to job description
    - Faculty whose jobs involve some amount of administrative duties will typically be paid more to reflect the higher demands on their time.
3. Factors relating to market forces
    - The amount that individual faculty members are paid will undoubtedly reflect the supply and demand of competent people to perform the work. This then suggests that certain fields will tend to be better compensated than others.

(a) Demand: Faculty members in fields that are judged highly useful by society will tend to get paid more than faculty members in fields that are judged to be less useful.
(b) Supply: If the number of applicants for positions in a field is much higher than the number of positions in the field, it is likely that salaries in that field will be lower.

4. Factors relating to the economic environment

   Patterns of inflation will affect salaries, and there is a tendency for such inflationary pressures to affect starting salaries more than raises, at least for government employees. The end result is salary compression in which the difference between salaries for starting employees and salaries for more senior employees is diminished.

   Over the time period covered by this study (about 50 years if we consider the time from the earliest start year until 1995), the pattern of inflation has been nonlinear: In the 50s and early 60s, inflation was very low. In the 70s, inflation was very high. Recently, inflation has been relatively low.

5. Patterns of raises

   There is a general tendency to award pay raises as percentages of current salary. Thus, as years go by, the variance of salaries increases.

1.3 Association between Sex and Factors Affecting Salary

The degree to which any of the above factors affecting salary will confound our ability to detect sex discrimination at the university will be determined by the nature of associations between sex and those factors.

1. Factors relating to the faculty member.

   - Skill: The basis for investigation of sexual discrimination is the presumption that women are equal to men in their capabilities. If this is not true, we can stop the analysis right here, because we have no data to assess the innate cappabilities of the individual faculty members. Hence, I proceed by assuming that there are no differences between men and women in their innate skill, and I will leave later studies (having more appropriate data to answer that question) to determine whether any differences in salaries that I find in this analysis might be justified based on unmeasured differences between the male and female faculty in this study.
   - Training: Due to historical and social trends, women tended not to enroll in educational programs leading to higher degrees until the recent past. Hence, there might be some tendency for fewer women to have Ph.D.s. However, it is not clear to me that women would necessarily be hired in fields where Ph.D.s were requisite for the job. Thus, I would think the past tendency not to allow women to get degrees would be perhaps responsible for there being fewer women hired at the university in the 1950s, for instance, but that it would not explain why those women received lower salaries. Instead, I would think any association between sex and degree is more likely reflective of the field not requiring Ph.D.s, rather than that women were hired in the absence of a Ph.D.
   - Experience: I do not think it would be surprising to find that more women tend to have been hired in later years than in the 1950s and 1960s. There was until recent times clear bias against hiring women in a great many jobs, and as noted above, the number of women pursuing higher education in many fields tended to be smaller than that for men.
   - Productivity: As with skill, I believe the presumption for this analysis should be that women and men tend to be equally productive in terms of teaching, research, and service. If, under this assumption, we decide that women's salaries lag men's, then more detailed comparisons can be made on a case by case basis to consider productivity, etc.

2. Factors relating to job description

- To the extent that administrators are generally more experienced than the faculty at large, it would not be surprising to find that women are less commonly administrators: The historical patterns in entering higher education would cause there to be fewer women with higher levels of experience. However, there is also the possibility that sex discrimination has caused the representation of women among administrators to be even lower than the historical reasons would predict.

3. Factors relating to market forces
    - Demand: It does not seem likely to me that society's demand for a particular field is a direct result of there being more women than men in that field, or vice versa. (I note that it is not that unusual for third graders to decide that they don't want something because the opposite sex had something to do with it, but I suspect that this is not a major factor after elementary school.)
    - Supply: To the extent that women have only been allowed to enter certain fields, it is reasonable to expect that the supply of applicants might be high enough to result in an oversupply of labor. In that case, the salaries for the fields that are more accepting of women historically might have lower salaries. Under that scenario, lower salaries for those fields would be a result of historical discrimination against women.
4. Factors relating to the economic environment

   One would expect that inflation would affect men and women equally. However, given the different patterns of hiring men and women over time, there may be a differential effect of inflation on the sexes.
5. Patterns of raises

   In the absence of discrimination, and under the assumptions of general equality between the sexes with respect to skill and productivity, we would expect to see similar patterns of raises for the two sexes. However, given the historical discrimination against women and the possibility that attempts have been made to achieve parity in salary, there may be a tendency for women to receive increasingly better raises over time.

1.4 Discrimination by Sex

There are a number of ways that discrimination against women has been alleged to affect the salaries women receive, and not all of these can be addressed by the data at hand. For instance, our data represent only those people who are employed at the university in 1995. Hence we cannot address the following types of discrimination:

1. Women might be less likely to be hired than a man with comparable qualifications.
2. Women might be less likely to be retained than a man with comparable qualifications.

In particular, it should be noted that we have no way of knowing whether any trend toward having more women hired more recently reflects a new trend, or whether it has always been the practice of the university to hire women at higher rates into the lower ranks and then fire them after a few years. The latter practice could result in the pattern that we see in the data, and it could well represent discrimination.

With respect to the questions that these data can answer, I find it useful to consider each of the potential confounding variables with respect to the extent to which they might reflect discrimination against women.

1. *deg*: I will take this variable to represent training as discussed above. I suspect that this variable may well mean very different things in different fields and at different times. I might consider an interaction term with *field* and *yrdeg* when both variables are included in the model. I note that I am not very

concerned that this variable is truly a confounder, nor do I suspect that it is a particularly strong predictor of salary after adjusting for other confounders.

2. *yrdeg*: This is the best measure of experience, providing there is not a large trend toward women having taken more parental leave. Its association with sex probably represents historical discrimination against women or discrimination against women at the level of admission to educational programs. Neither of these forms of discrimination are of particular interest in deciding whether the university discriminates in awarding salaries to women in 1995, so I am perfectly comfortable in adjusting for this variable.

3. *startyr*: After adjusting for *yrdeg*, the interpretation of this variable may indicate those faculty who are recruited away from other institutions. It may reflect discrimination if there is a greater tendency for the university to try to hire men in such a situation than women. However, there may be a myriad of reasons for a faculty member to be hired several years after obtaining his/her degree, and thus I do not feel that any discrimination effect would necessarily dominate. It should also be noted that *yrdeg* and *startyr* would be expected to be highly collinear, and their effects on *salary* are probably nonlinear due to inflation patterns, historical discrimination, and attempts to remedy historical discrimination. Hence we might need to be careful that the modeling of these two variables simultaneously not just allow two degrees of freedom to model a nonlinear effect of a single strong association, rather than independent associations. (In any case, we should be cautious in overinterpreting either effect when both variables are modeled.)

4. *field*: As noted above, academic field almost certainly represents historical discrimination against women. However, I think there may be some possibility that salary disparities among fields represents continued discrimination against women: There may be little motivation to re-evaluate whether a given field is deserving of higher pay if that field is dominated by women. (I note that historically men often decreed that there was no need to pay women as well as men, because it was presumed that men would be the primary support of their families.)

5. *admin*: This variable probably represents lack of supply of women with long experience in the field and possibly any discrimination in awarding administrative positions to women. My own belief is that the historical issues predominate, but I note that when we compare two individuals getting their degrees in the same year, we should suspect discrimination if the man is markedly more likely to be an administrator than the woman. Hence, after adjustment for year of degree, I am more inclined to regard any residual confounding of the sex effect by *admin* as evidence of discrimination. If such exists, I would be very interested in the model which did not adjust for *admin*.

6. *rank*: After adjusting for year of degree, I think that in the absence of discrimination and under the presumption that men and women are equally capable, the distribution of men and women ought to be the same for each rank. Hence, any remaining confounding I will interpret to be discrimination, until a more detailed case by case study could demonstrate otherwise. Under this scenario, in the absence of discrimination, *rank* is a precision variable (because no differences exist by sex). If differences exist by sex, then I will be taking that to mean that there is some possibility of discrimination in awarding promotions, and I would be regarding *rank* a surrogate for the response (or a variable in the causal pathway). Failure to adjust for the precision is not such a big deal, while adjusting for a variable in the causal pathway will ruin your ability to answer the scientific question of greatest interest.

7. *sex*: This is of course the predictor of interest, and thus should be included in all models

1.5 Methods of the analysis

The response variable is clearly *salary*, and linear regression would be the logical choice of method. I would argue that a logarithmic transformation of salary is logical given the typical pattern of granting raises as a percentage of current salary. Such a pattern will tend to cause greater variation in those subpopulations having a higher mean salary. Of course, the actual form of modeling *salary* will have to satisfy the key assumptions of linear regression.

Based on the above discussion, there is a hierarchy of models that I will consider in order to address the questions.

10. I would first consider a model in which only *sex* is modeled. In the absence of any discrimination (current or historical), and under the assumption that men and women are equally qualified, there ought to be no difference between the sexes in their salary in 1995. As noted above, however, the existence of historical discrimination is widely accepted. Hence, my purpose in presenting this analysis is just to tell a story: "Women currently tend to receive $x\%$ less pay than men. Some of this may be due to a smaller supply of properly trained and experienced women, which is in turn possibly due to historical controls. When we adjust for possible effects of historical discrimination, we find that the difference is $y\%$ ..."

11. My 'headline' model (that is, the model that I would base a headline saying "UW Women Faculty Paid $X\%$ Less Than Men Faculty") would be based on adjustment for all the possible confounders but *rank*. ($X\%$ turned out to be about 7.0%.)

12. My next model would be to adjust for *rank* in order to separate the amount of the decrease in salary that reflects lower pay at the same rank from the amount of the decrease is due to women having a lower salary. So the results might be summarized as "UW women faculty tend to be paid about 7.0% lower than comparable men faculty having the same training, experience and administrative duties. This lower salary is caused by women's salaries tending to be 4.7% lower than men's salaries for the same rank, and the remaining 2.3% decrease in salary reflects the fact that women are less likely to be in the higher ranks than men of comparable training and experience."

1.6 Modeling of Variables

Having decided what general models to use, we must also consider the way those variables should be modeled. Some part of this is done after looking at the data, but the more that I can anticipate it beforehand, the happier I am with my analysis.

As noted above, I would reasonably expect some heteroscedasticity in *salary*, and a logarithmic transformation would seem reasonable (see comments below).

As our predictor of interest is dichotomous, there is nothing to consider here.

When modeling our other predictors, we need to consider whether they are confounders or precision variables, and whether we want to spend much time trying to figure out the best representation. As noted above, it probably doesn't matter too much how you model precision variables, but failure to model gross departures from linearity in confounders may mean that some residual confounding (sometimes looking like interactions) may exist.

I would model the other variables as dummy variables, although I have some sympathy for modeling rank continuously (coded 0, 1, 2) if you want. But again, we have lots of data, and I don't really care about the effect of rank on salary.

There are many other interactions with *sex* that could be considered. In particular, interactions between *sex* and *field* and between *sex* and *rank* could be included whenever modeling those other predictors. If that is done, tests for differences between the sexes must also consider the interactions, and estimates of the effect must be made within each rank or field category. I would probably make such a decision before looking at the data. Even if an interaction exists, it might be of interest to estimate a common effect across the university, especially if the first attempt at remediation would be to do an across the board pay increase for women (probably an unlikely event). In exploratory analyses, I would certainly look at and report these interactions.

2 Problems Arising During Analysis of the Data

As most will note, some transformation of the outcome is likely to occur. This is not because the outcome is skewed, but because of the presence of heteroscedasticity. (I note that when the data are skewed, it is often the case that heteroscedasticity exists, but it is possible that the response could have a skewed distribution and that all the necessary assumptions for linear regression are satisfied. This can be the case when the predictors have skewed distributions as well, thereby making the distribution of response look skewed.)

We have to make sure that the assumptions of linear regression are satisfied. A residual plot clearly shows heteroscedasticity when *salary* is modeled untransformed. A logarithmic transformation takes care of most the problem, but not all of it. Possible causes for the remaining heteroscedasticity include

1. Some confounder has not been modeled. This is moot if you have examined the residuals after adjusting for all of your variables (including *rank*).
2. Some interaction has not been modeled. You could consider looking for interactions, though you must be careful of the multiple comparison problem.
3. Some other transformation is necessary. This will help if the heteroscedasticity represents a mean-variance relationship. Such a relationship would show up if there were a pattern of increasing variability on the plot of residuals versus fitted values. But a couple notes are in order:
   (a) you might not be able to find the variance stabilizing transformation. (there are some methods whereby you can estimate it if the correct transformation is in a power family of transformations, e.g., the Box-Cox family).
   (b) you might not be able to interpret the parameters easily (this can be overcome by giving predicted values for each sex using interesting values for all the other predictors, but that is still suboptimal to me).

It is pretty much a judgement call whether the heteroscedasticity is something you have to deal with.

1. If the heteroscedasticity is truly just a mean-variance relationship, then under the null hypothesis, there would be no heteroscedasticity and hypothesis tests are valid (you may lose a little power, and confidence intervals will likely have incorrect coverage probabilities).
2. We are truly only concerned about a variance relationship with the predictor of interest. It is only in that case that we could truly increase the type I error.

3 Presentation of results

Below are selected tables and comments that I would likely present in a report of this analysis. In writing the report to the client, I consider it important that you discuss the limitations of the data up front (as we have done), and then revisit some of those issues in the discussion when it comes to discussing whether sex discrimination exists. Similarly, you should discuss the difference between the interpretations of the various models as they relate to the practical problem of deciding how much discrimination exists.

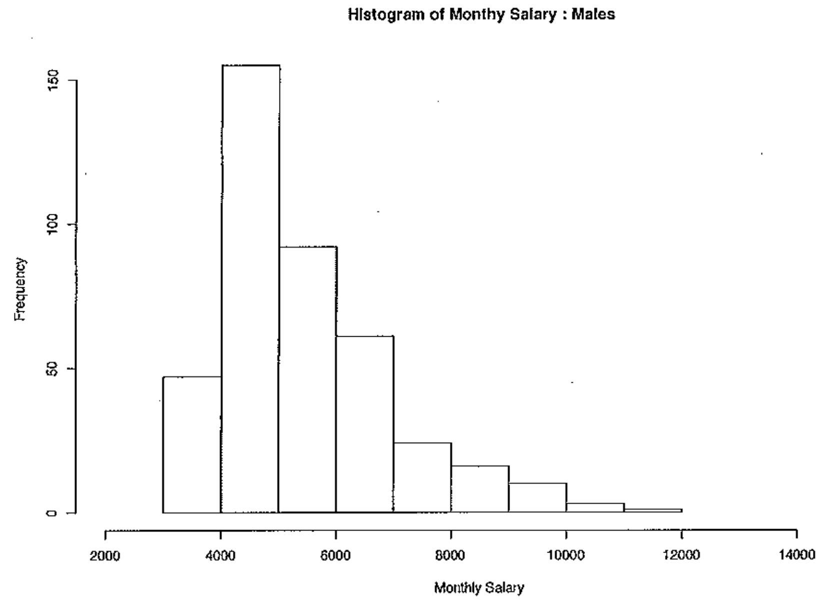
I also think it would be useful to suggest other data that might have been useful (number of publications, citations, courses taught).

1. First I would begin with simple descriptive statistics by gender as presented in table 1.

Table 1: Faculty characteristics stratified by gender. Summary statistics represent mean (sd) for continuous characteristics and frequency (%) for categorical characteristics.

Characteristic	Female (N=409)	Male (N=1188)
Year of Degree	81.11 (8.7)	74.37 (9.64)
Highest Degree		
- Other	56 (13.7%)	88 (7.4%)
- PhD	334 (81.7%)	1016 (85.5%)
- Prof	19 (4.6%)	84 (7.1%)
Field		
- Arts	80 (19.6%)	140 (11.8%)
- Other	287 (70.2%)	780 (65.7%)
- Prof	42 (10.3%)	268 (22.6%)
Starting Year at UW	85.47 (8.02)	79.62 (10.17)
Faculty Rank		
- Assist	145 (35.5%)	170 (14.3%)
- Assoc	138 (33.7%)	299 (25.2%)
- Full	126 (30.8%)	719 (60.5%)
Administrative Duties		
- No	377 (92.2%)	1051 (88.5%)
- Yes	32 (7.8%)	137 (11.5%)

2. Next, I would consider the unconditional distribution of salary by gender as in Figure 1.



(a) Females

(b) Males

Figure 1: Distribution of monthly salary by gender.

3. Next, I would summarize salary differences by gender and each of the potential confounders as given in Table 2 and Figure 2.

Table 2: Average monthly salary (sd) stratified by gender and faculty characteristics.

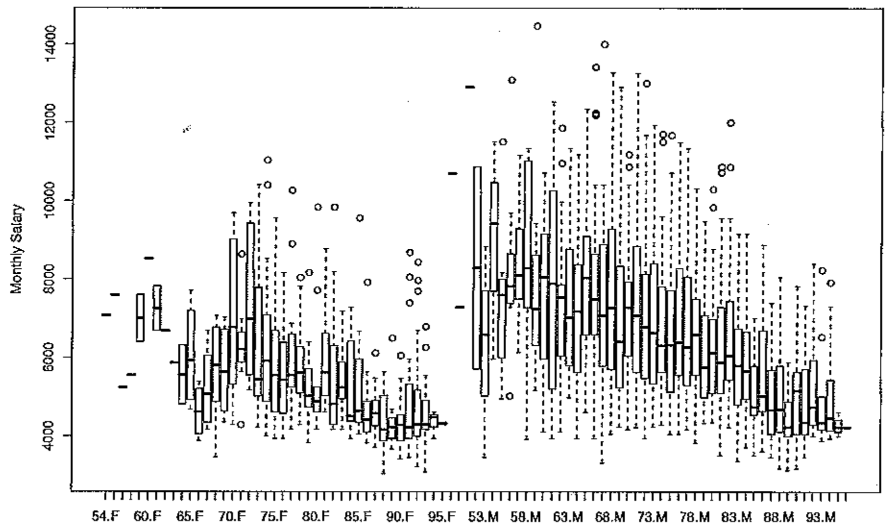
Characteristic	Female (N=409)	Male (N=1188)
Highest Degree		
- Other	5134.5 (1206.3)	5759.0 (1587.4)
- PhD	5390.1 (1486.8)	6731.6 (2084.5)
- Prof	6289.1 (1825.7)	7751.5 (2149.5)
Field		
- Arts	4910.5 (1158.1)	5488.1 (1280.2)
- Other	5340.2 (1411.6)	6641.7 (2062.1)
- Prof	6710.8 (1759.7)	7643.0 (2118.4)
Faculty Rank		
- Assist	4502.9 (981.5)	4773.8 (1116.7)
- Assoc	5018.9 (858.1)	5480.5 (1224.2)
- Full	6839.7 (1435.7)	7714.8 (1943.4)
Administrative Duties		
-No	5280.4 (1410.1)	6506.6 (2002.7)
-Yes	6769.8 (1627.2)	8458.0 (1938.3)

4. When getting to the main comparison of interest, first I would discuss the unadjusted difference in salaries: The median salary for men is estimated to be approximately 23.2% higher than that of women (95% CI for ratio: 1.192 1.272; p-Value<.001).

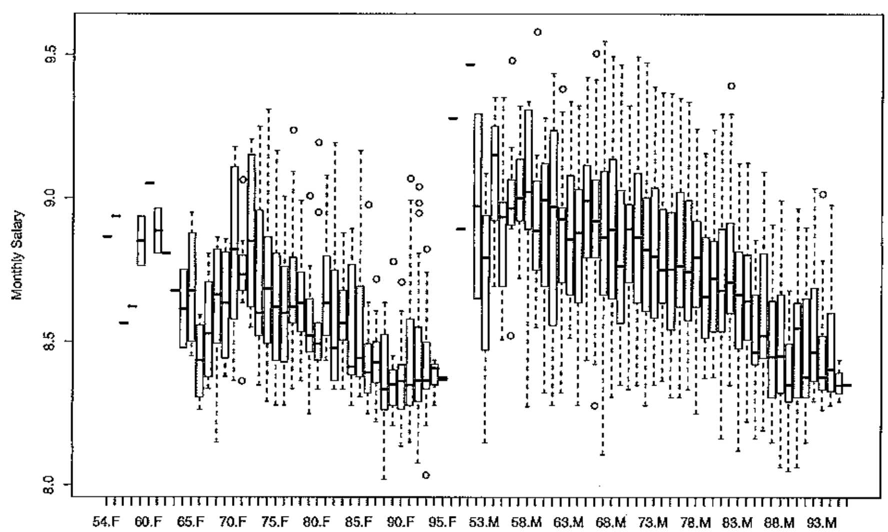
Next, I would present the adjusted results (without *rank*) as given in Table 3.

Table 3: Linear regression model based estimates of the relative difference in median salary.

Covariates	Ratio of Median Salary (95% CI)	P-value
Gender (M vs. F)	1.070 (1.041, 1.099)	<.001
Year of degree (per 5 yrs)	0.923 (0.914, 0.932)	<.001
Highest Degree		
PhD	1.0	
Other	0.894 (0.857, 0.931)	<.001
Prof	1.040 (0.993, 1.089)	0.100
Field		
Arts	1.0	
Other	1.124 (1.086, 1.163)	<.001
Prof	1.342 (1.288, 1.399)	<.001
Starting Year (per 5 yrs)	1.012 (1.002, 1.021)	0.013
Administrative Duties (Y vs. N)	1.235 (1.190, 1.281)	<.001



(a) Monthly Salary



(b) Log-transformed Salary

Figure 2: Monthly salary (untransformed and log-transformed) stratified by gender and year of degree.

Next I would discuss the difference in salaries if we condition on rank: With adjustment for rank and those covariates listed in Table 3, the median salary for men is estimated to be approximately 4.7% higher than that of women (95% CI for ratio: 1.023 1.072; p-Value<.001)

5. Finally I would discuss exploratory analysis results investigating effect modification by rank and field (I found none, but I would still discuss it as it is of interest).