# Main Instructions

1. Open 'faculty.sav'
2. Run frequencies, graphs and correlations to get an overview of the data file
3. The aim of the analysis will be to understand gender differences in salary and to see the extent to which they persist after controlling for various variables
4. Run a t-test comparing mean salary for males and females; calculate cohen's d;
5. Determine and describe the relationship between each of the other predictors with salary (i.e., FTE, Rank, Articles, Experience)
6. Run an ANCOVA (Analyze - GLM - Univariate) with Sex as IV, but include covariates that you think are legitimate bases for salary discrimination (i.e., the aim is get an estimate of wage differences controlling for legitimate discriminators); Get an estimate of the marginal mean salary for males and females.
7. (a) Why might you want to re-run excluding "Adjuncts"? (b) Re-run ANOVA excluding adjuncts and interpret the results; examine marginal means. (to filter, see "Data - select cases")
8. Try to create an overall index out of articles and experience that could be used to determine salary. You might want to transform one or more of the two variables. You may want to make them into z-scores. You may or may not want to weight the two variables equally in any composite.
9. With the composite variable you created in the previous step and the filter excluding adjunct: (a) create a scatter plot with x=composite; y=salary; and different colours for sex; fit a line of best fit to each gender. (b) run another ANCOVA with sex as fixed factor and your composite as covariate; examine homogeneity of regression slopes assumption; examine marginal means for sex.
10. Provide an overall summary of what can be said about sex differences in salary in this sample and whether any differences appear justified or not. Note any limitations or complexities.

# Answers

## 2. Run frequencies, graphs and correlations to get an overview of the data file

Analyze - descriptives - frequencies is a good place to get frequencies (you might also ask for some descriptive statistics such as means, sd, min, max; you might also ask for histograms, which would be useful for the continuous variables)

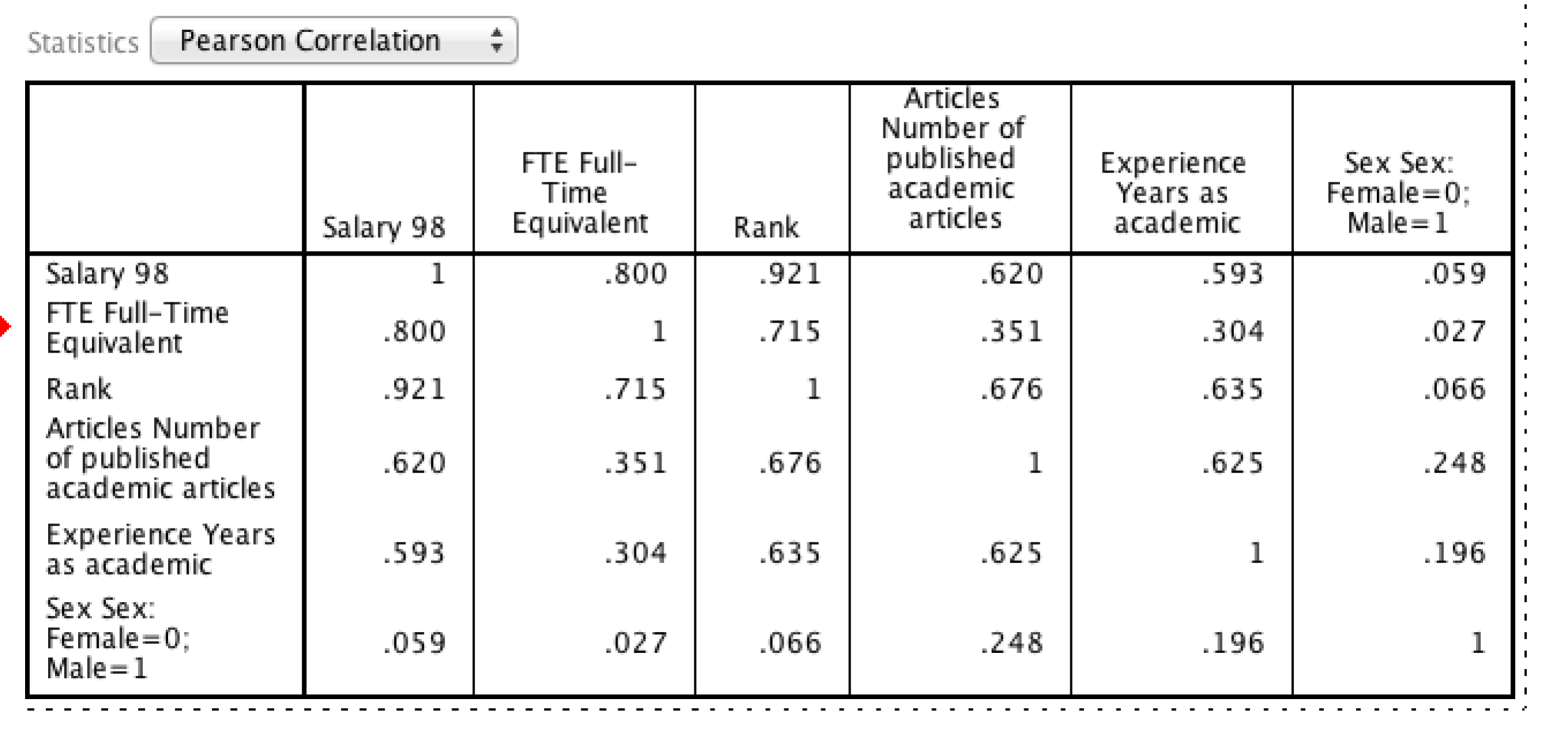
Analyze - correlate - bivariate is a good place to get correlations

The aim is to understand each of the variables and where appropriate get a feel for the relationship between the variables.

The frequency analysis should highlight things like:

* Most of the dataset are full time
* There is a scale from 1 to 5 for rank, however its unclear whether it is an ordered scale.
* Number of published articles is quite positively skewed as is typical for such count based achievement measures
* Years of academics is also positively skewed with more people with fewer years of service. This may be a property of the sampling, but it also makes sense statistically that if you sample tenure that it would typically have such a distribution, because some people don't stay very long and others stay a long time.
* There's roughly equal numbers of males and females in the sample.

The correlation analysis is show below



A few observations:

* There are very strong correlations between salary, FTE, and rank. This makes sense given that salary is often directly tied to rank. Salary is also often proportional to the FTE.
* There are also relatively strong correlations between publication count, experience, and salary. This also makes sense because over time academics get promotions and salaries tend to go up. Similarly those who are good at their job may be more likely to stay as academics. There may also be a link between publication count and time to get publications. Thus, it is somewhat unclear the degree to which publications are a proxy for time on the job or are directly linked to promotions and salary. It is quite likely that it is a bit of both.
* Sex appears relatively unrelated to salary, but there are some small correlations with publication count and years of experience, with males having more publications and more years of experience. The relationship is weak, but may reflect either a history of more men in academia or men staying in the workforce for longer. That said the correlations are small, so not much should be read into it given the small sample size.

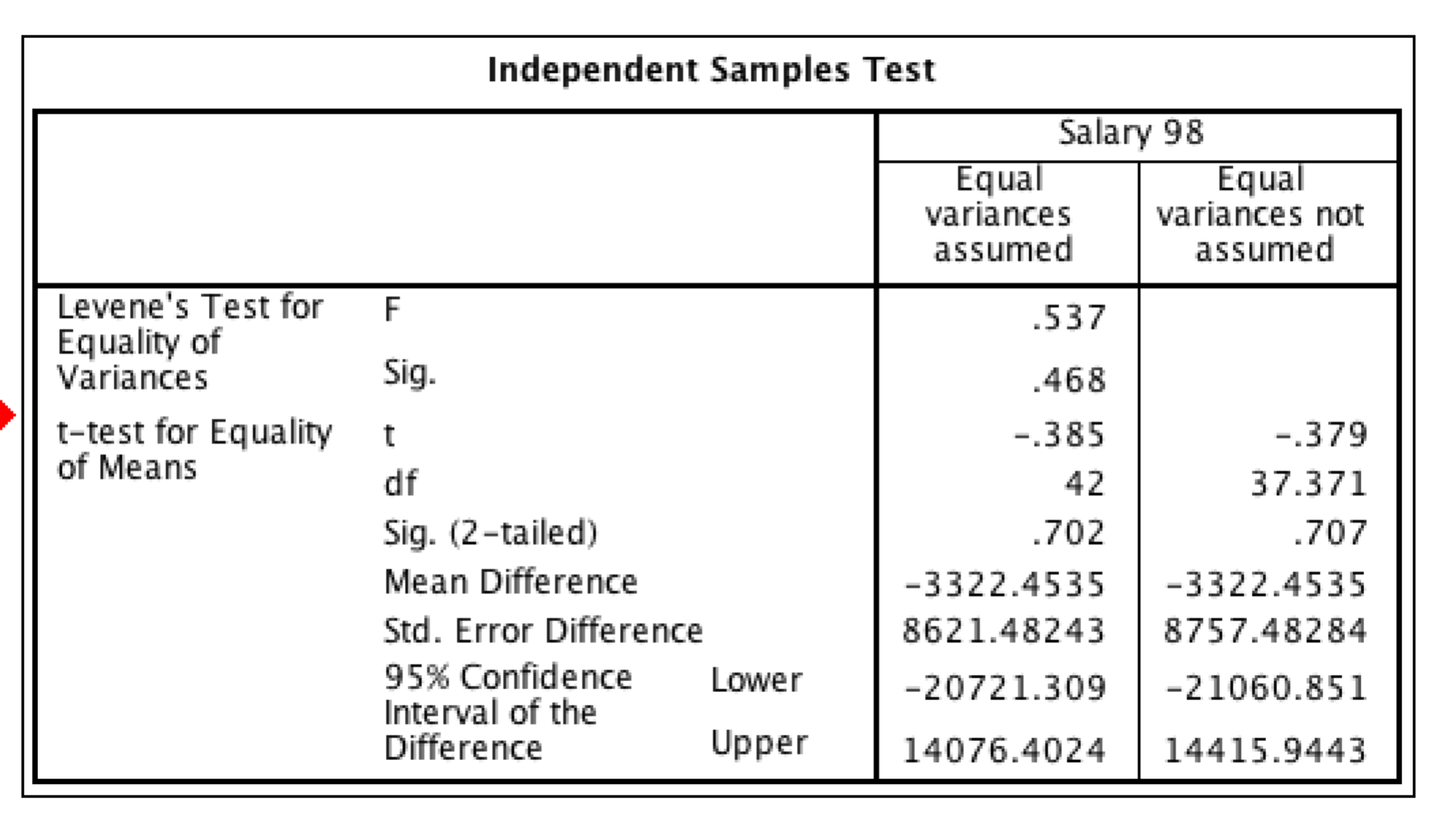
## 3. The aim of the analysis will be to understand gender differences in salary and to see the extent to which they persist after controlling for various variables

## 4. Run a t-test comparing mean salary for males and females; calculate cohen's d;

"Analyze - compare means - independent samples t-test"

Grouping variable: sex

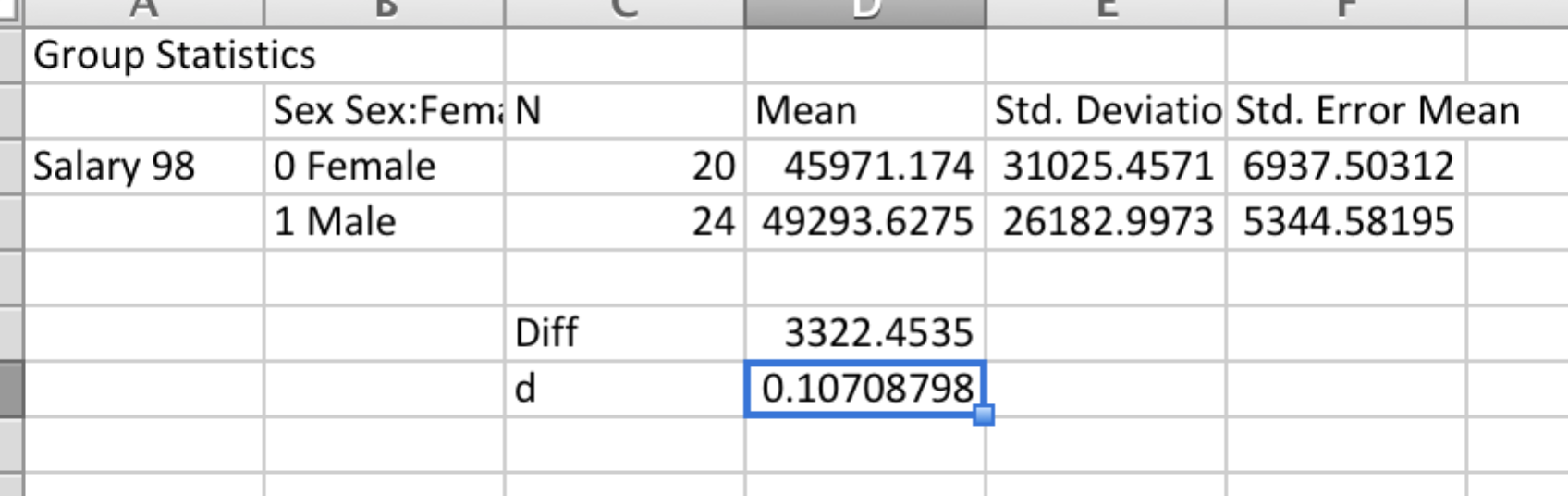
Test variable: Salary



The significance test suggests that there are no significant differences between male and female salaries (p=.702).

To get Cohen's D, we need to do some hand calculations by dividing the difference in means by the standard deviation. There is some choice about which standard deviation to use. You could use one of the online calculators: http://www.uccs.edu/~lbecker/

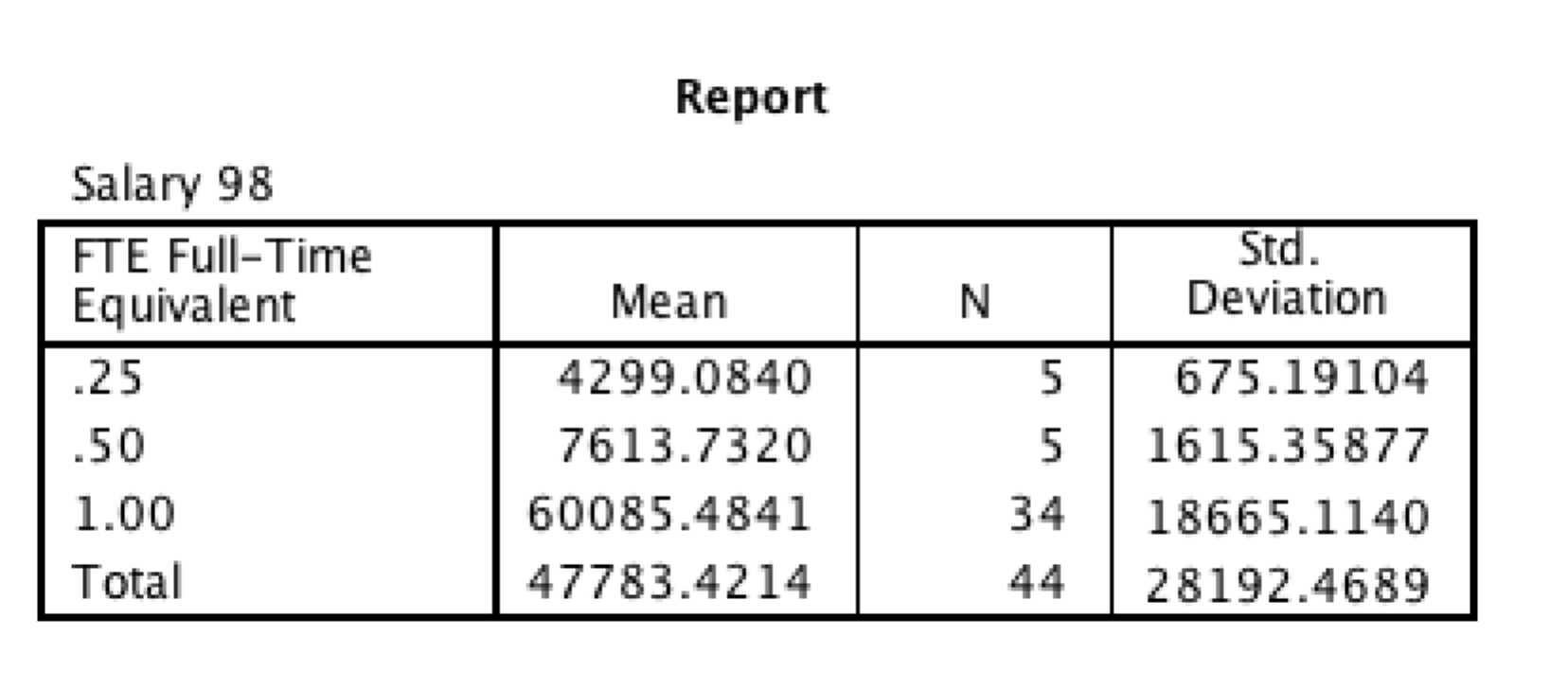
For simplicity I just used the Female standard deviation and used a formula in excel after pasting the table of descriptive statistics from SPSS into Excel.



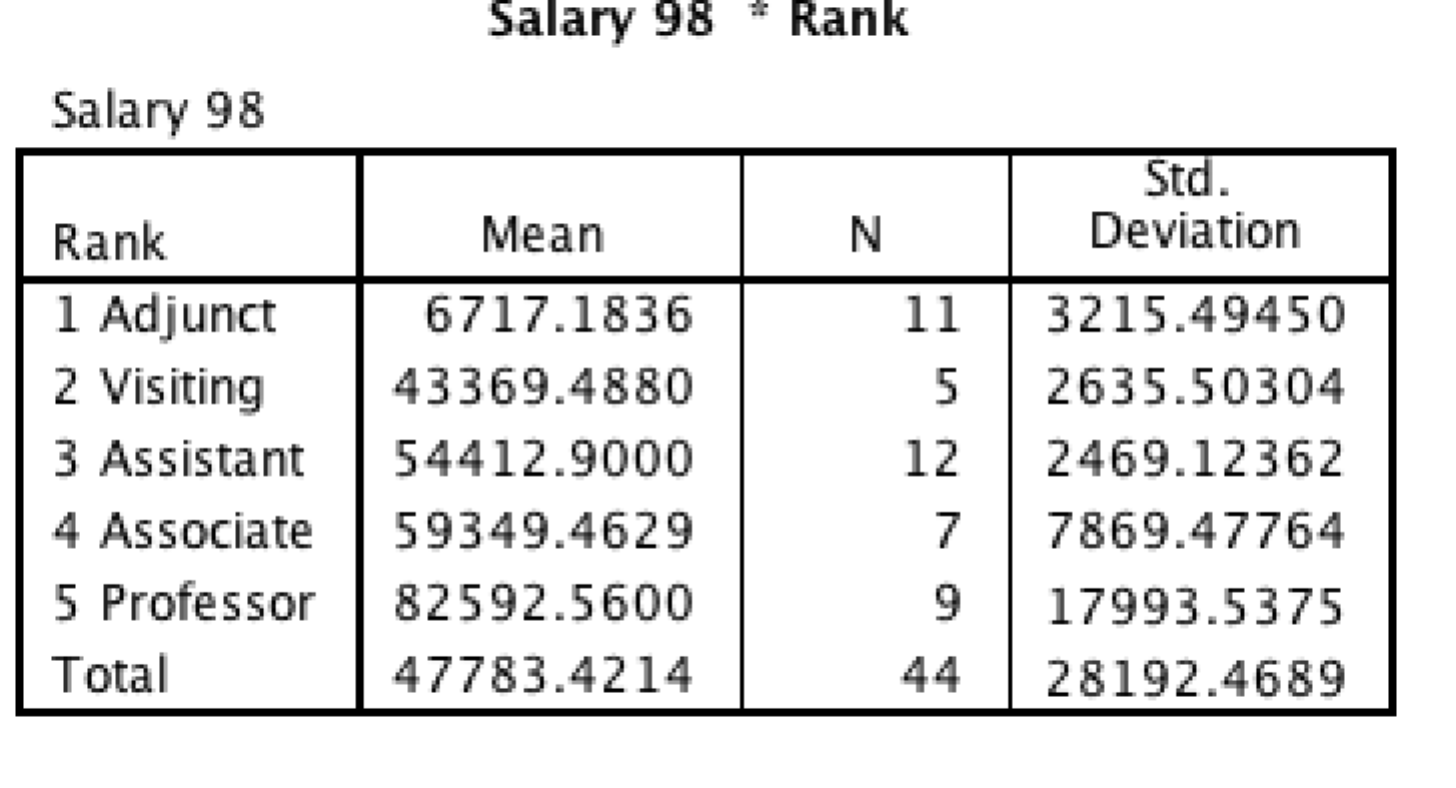
Thus, the difference was around .1 of a standard deviation.

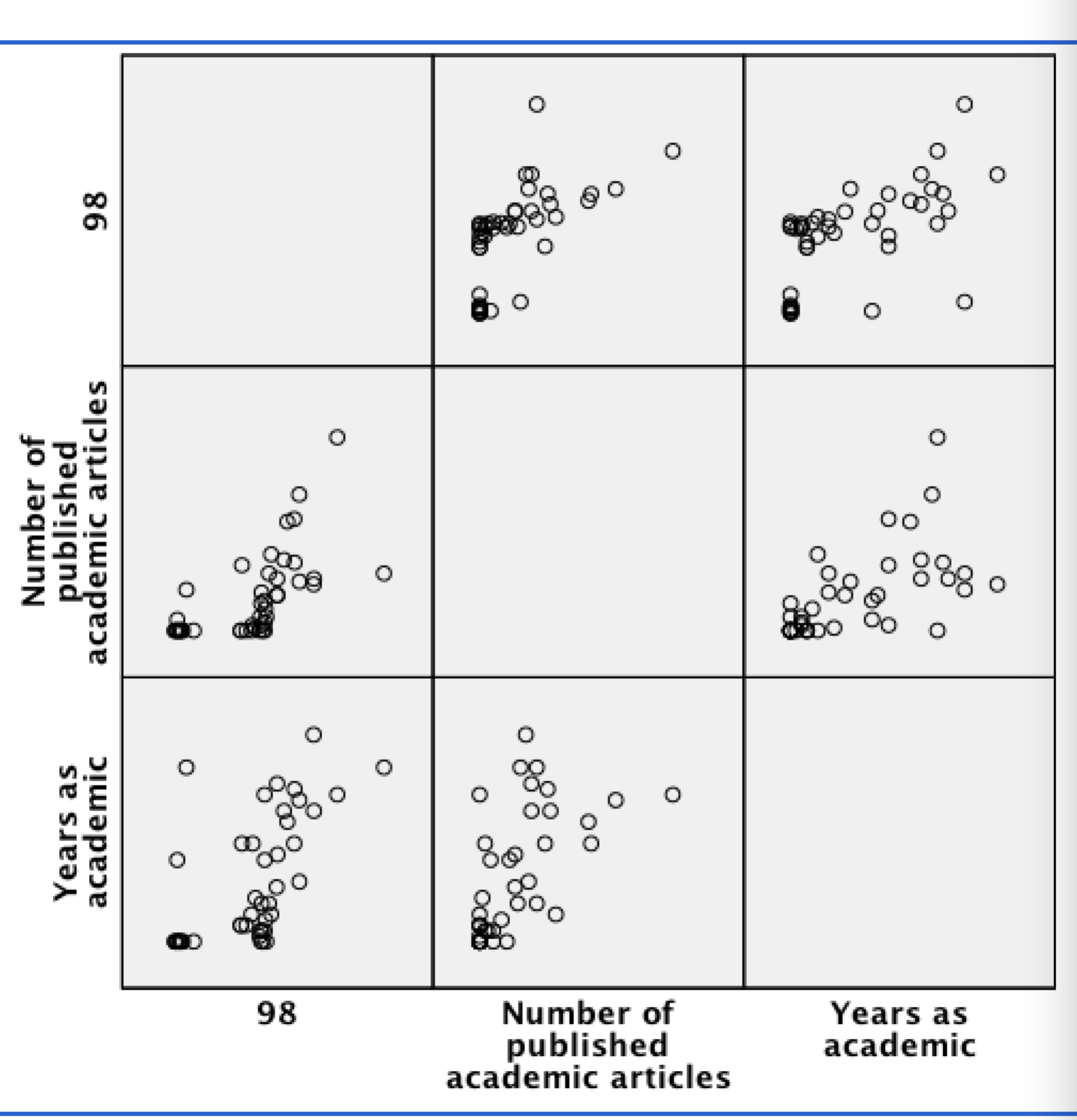
## 5. Determine and describe the relationship between each of the other predictors with salary (i.e., FTE, Rank, Articles, Experience)

* FTE: The correlation with FTE is very strong. However, a table makes it clear that there is a massive difference between those on 1.0 and those not.



* Rank: The adjunct rank also appears to be a bit of an outlier.



* A matrix scatter plot might be a good way to explore the relationship between salary, publication counts, and years as academic. The relationships look roughly linear, although the lack of data for cases with many publications makes it a little difficult to assess, and the relationships are far from perfect.
* 

## 6. Run an ANCOVA (Analyze - GLM - Univariate) with Sex as IV, but include covariates that you think are legitimate bases for salary discrimination (i.e., the aim is get an estimate of wage differences controlling for legitimate discriminators); Get an estimate of the marginal mean salary for males and females.

There are several options for this. I excluded rank because if males or females are systematically being promoted then that might be a consequence of bias rather than a factor to be corrected for.

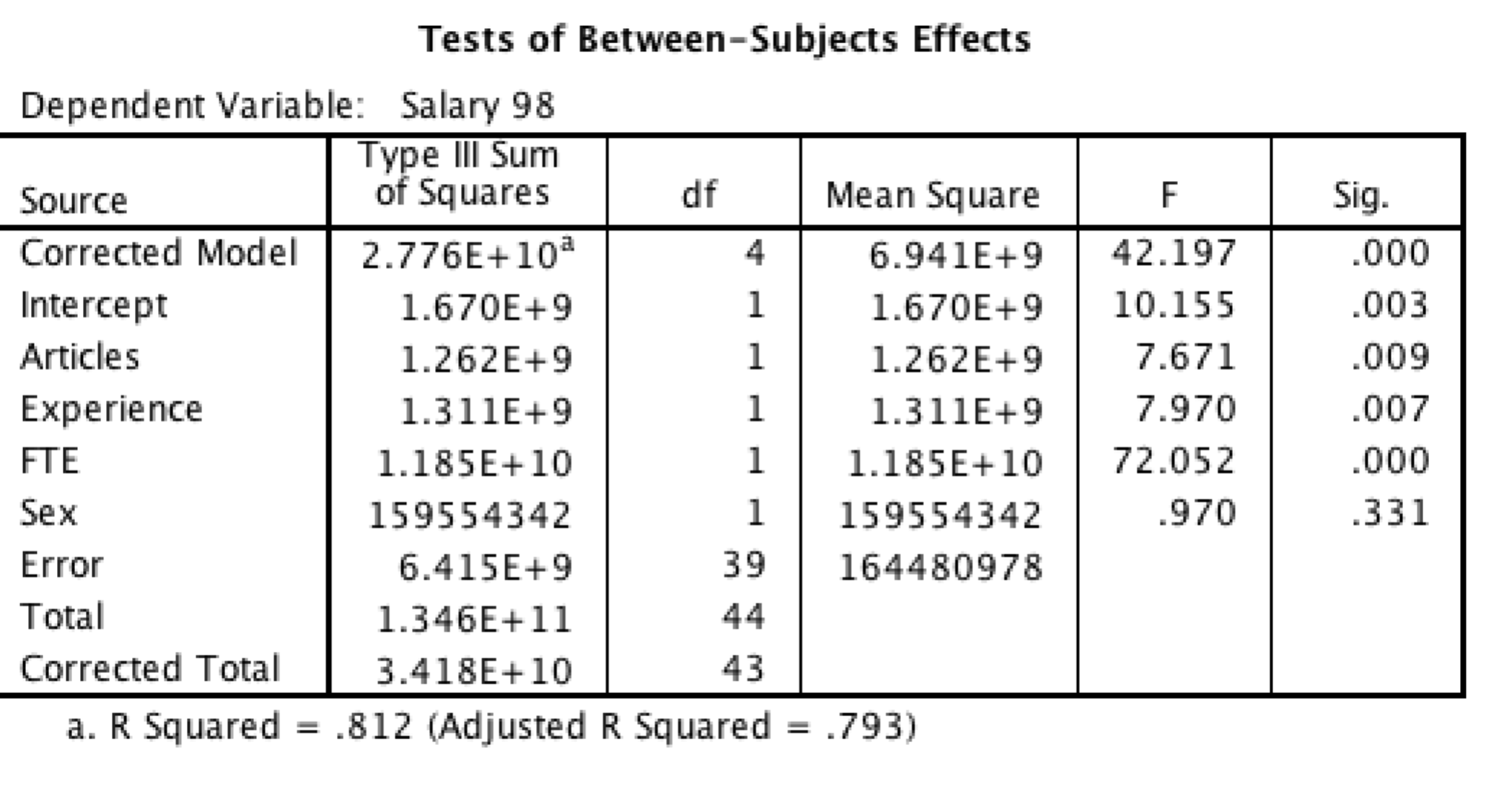
Analyze - GLM - Univariate

DV: salary

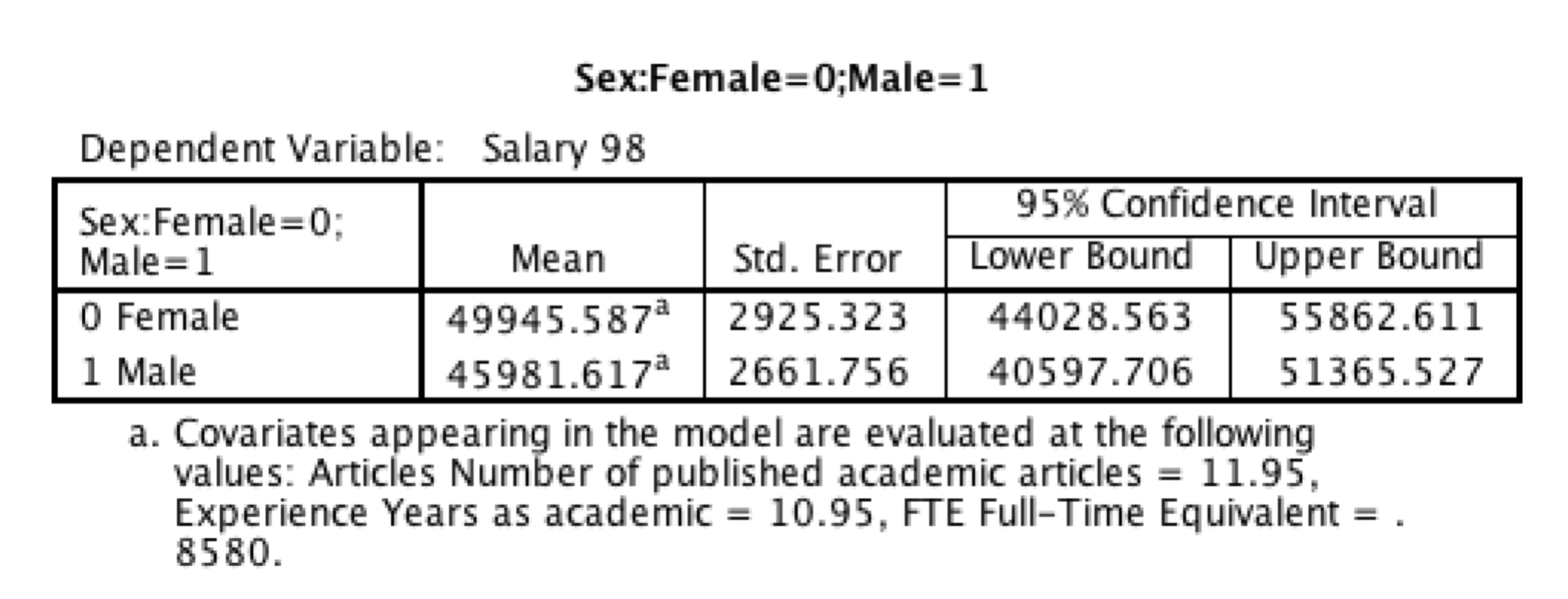
Fixed factors: Sex;

Covariates: FTE (you could put this in covariates, given it's non-linear effect, but we'll keep it simple); Years as academic; publication count

Options: Display means for sex



This table shows the significant effect or articles, experience, and FTE but not sex.



This table shows the marginal means. Interestingly, after adjusting for these factors the sample mean for females is about $4,000 more than males, whereas on the raw means, females were earning around $4,000 less.

## 7. (a) Why might you want to re-run excluding "Adjuncts"?

Adjuncts had a massively different income to the other ranks. Thus, they seem to be paid completely different amounts.

## (b) Re-run ANOVA excluding adjuncts and interpret the results; examine marginal means. (to filter, see "Data - select cases")

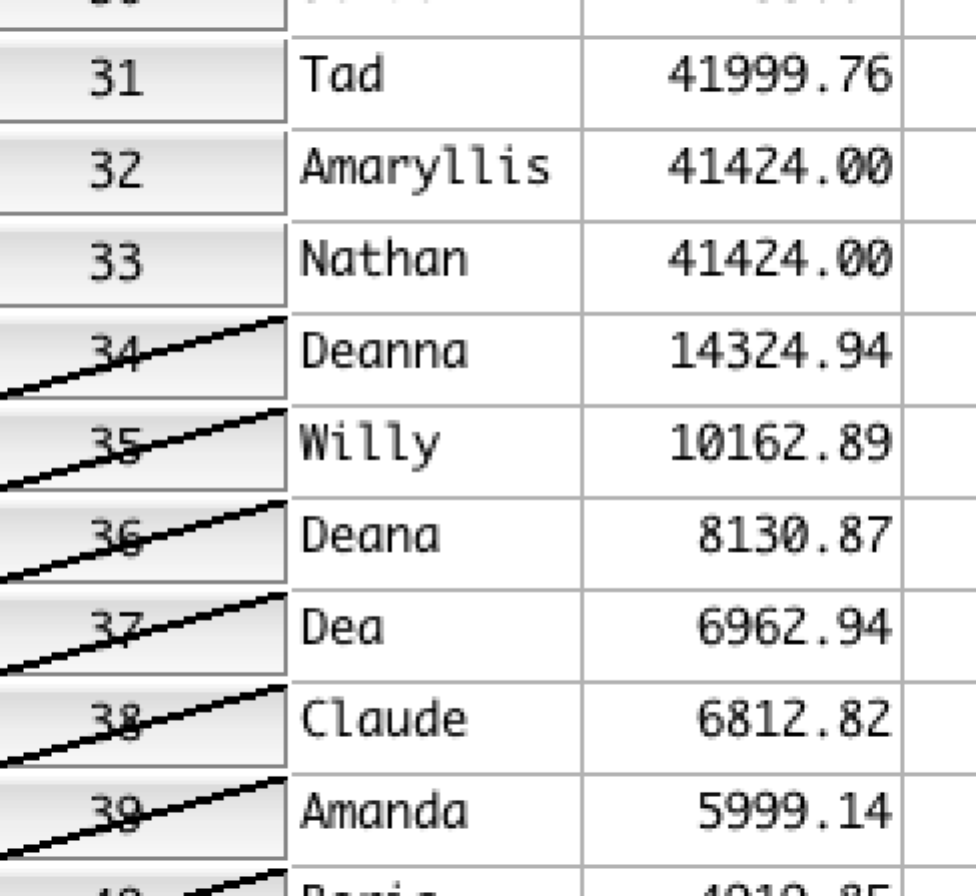
To exclude cases, go to "Data - select cases"

If condition is satisfied; click "if..."

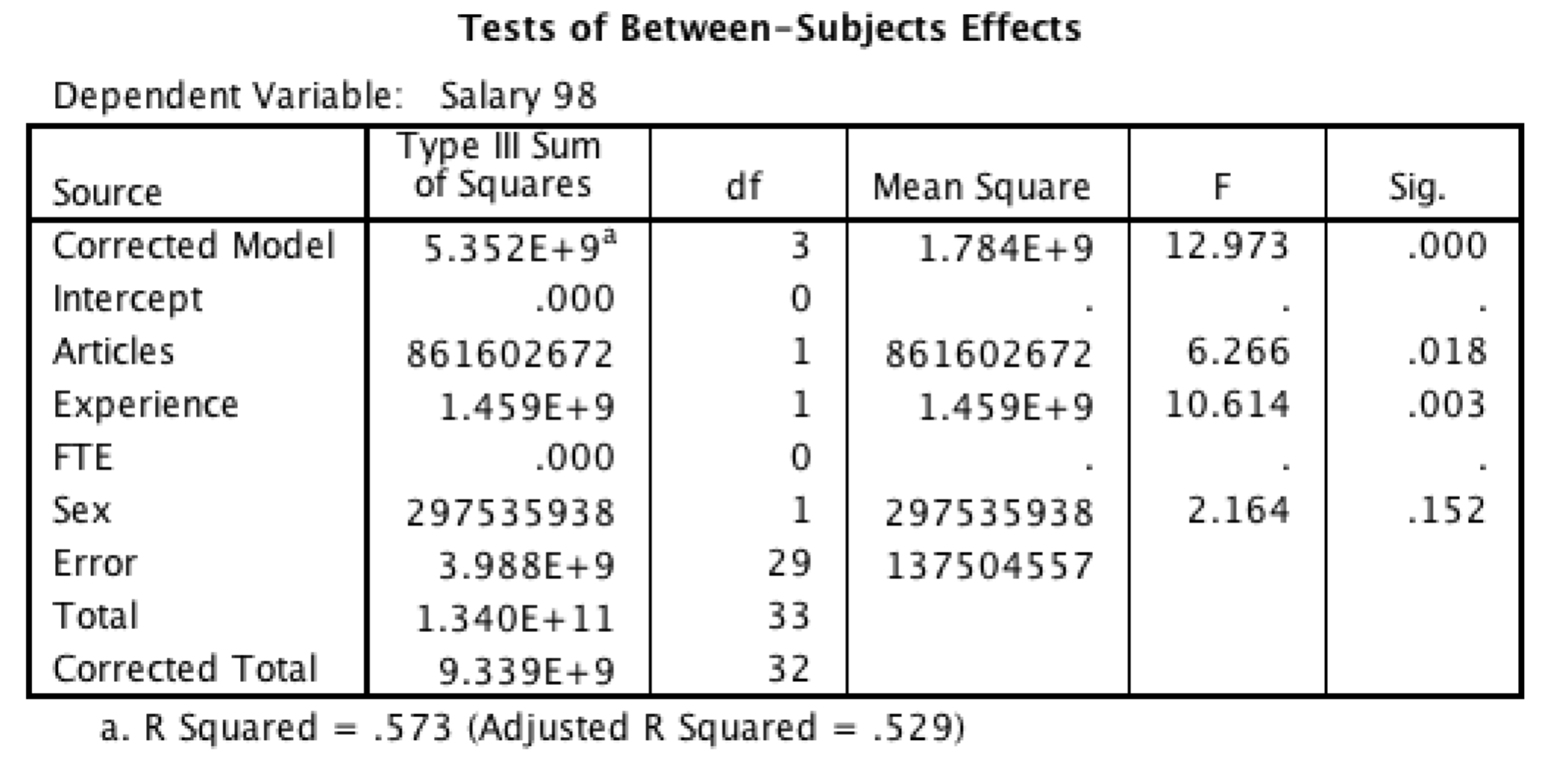
rank ~= 1

The "~=" means "not equal" and 1 is the value adjuncts.

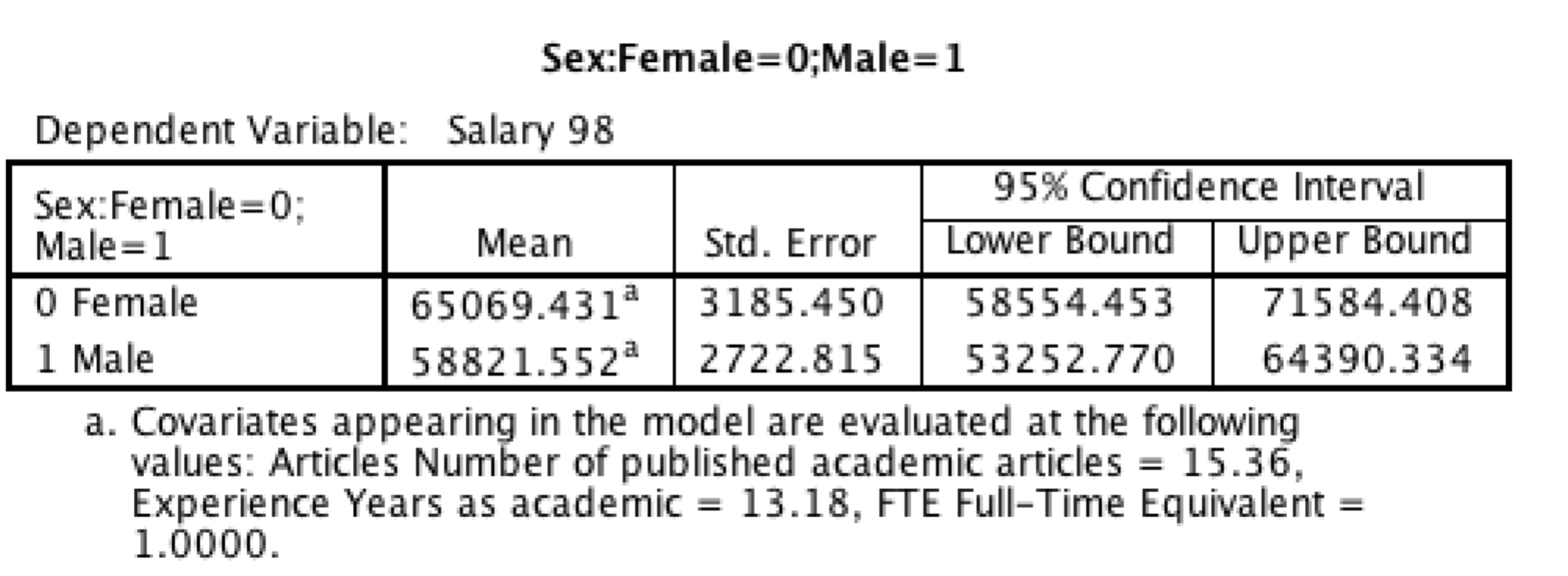
After you run this command, you will see that these cases have a vertical line through them.



Then run "Analyze - GLM - Univariate" as before.



The new results exclude FTE presumably because it is only the adjuncts that are not 1.00. Experience and articles are still statistically significant.

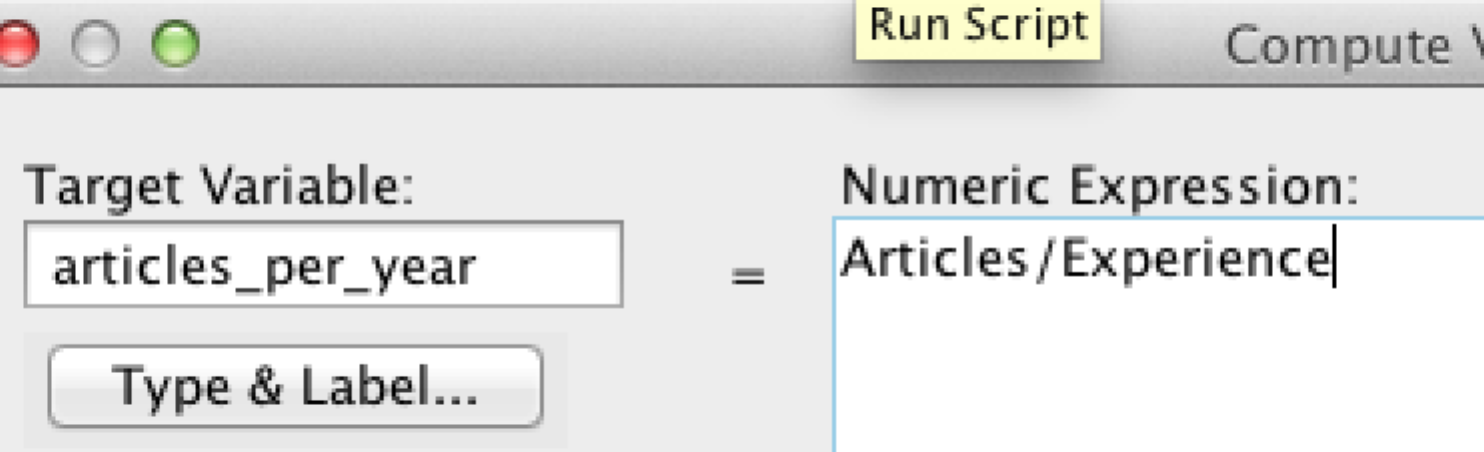


All the marginal means have increased because we have excluded the adjuncts. Now adjusted, females earn a bit over $6,000 more than males but it is a small sample size and it is not statistically significant.

## 8. Try to create an overall index out of articles and experience that could be used to determine salary. You might want to transform one or more of the two variables. You may want to make them into z-scores. You may or may not want to weight the two variables equally in any composite.

There is no right answer here. But one approach that initially appeals to is as follows: publication count and experience seem to be double dipping on experience. With more time there is more time to produce publications. So, I'm going to divide publications by experience to get a publications per year variable. I'll then predict salary from these two variables to get a regression equation for salary.

Transform - compute



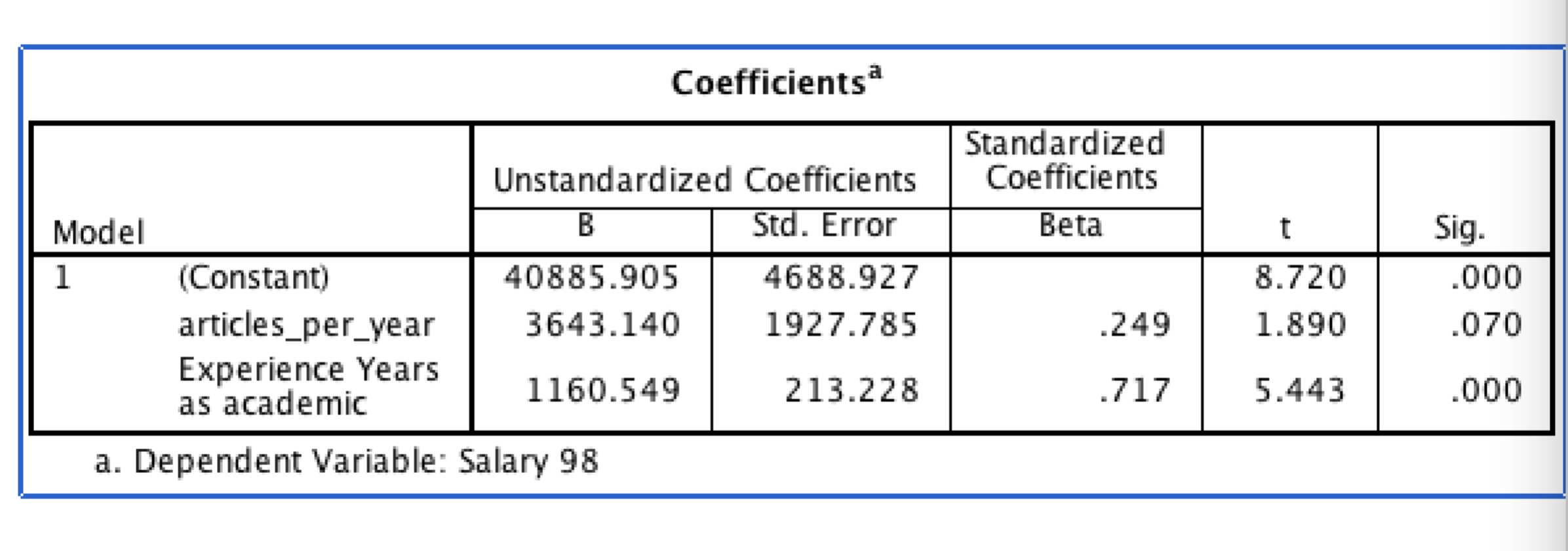
Analyze - Regression - linear

DV: salary

IV: articles per year and experience

Adjuncts are still filtered out

This model explained 53.7% of variance in salary.



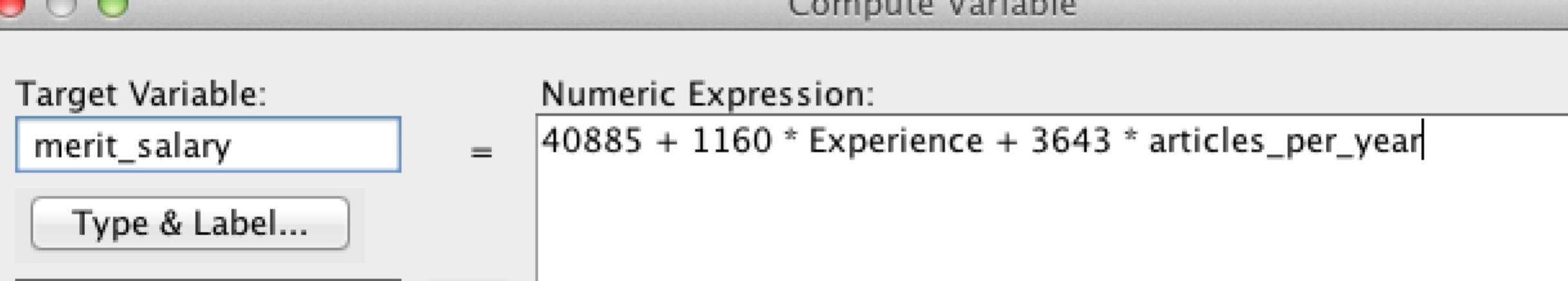
The regression equation is wage is:

$40,885 + $1,160 \* (years as academic) + $3,643 (articles\_per\_year)

Caution: Of course this is just an exercise, and there is a vast literature on structuring pay and benefits.

## 9. With the composite variable you created in the previous step and the filter excluding adjunct: (a) create a scatter plot with x=composite; y=salary; and different colours for sex; fit a line of best fit to each gender.

Use transform - compute to create the composite



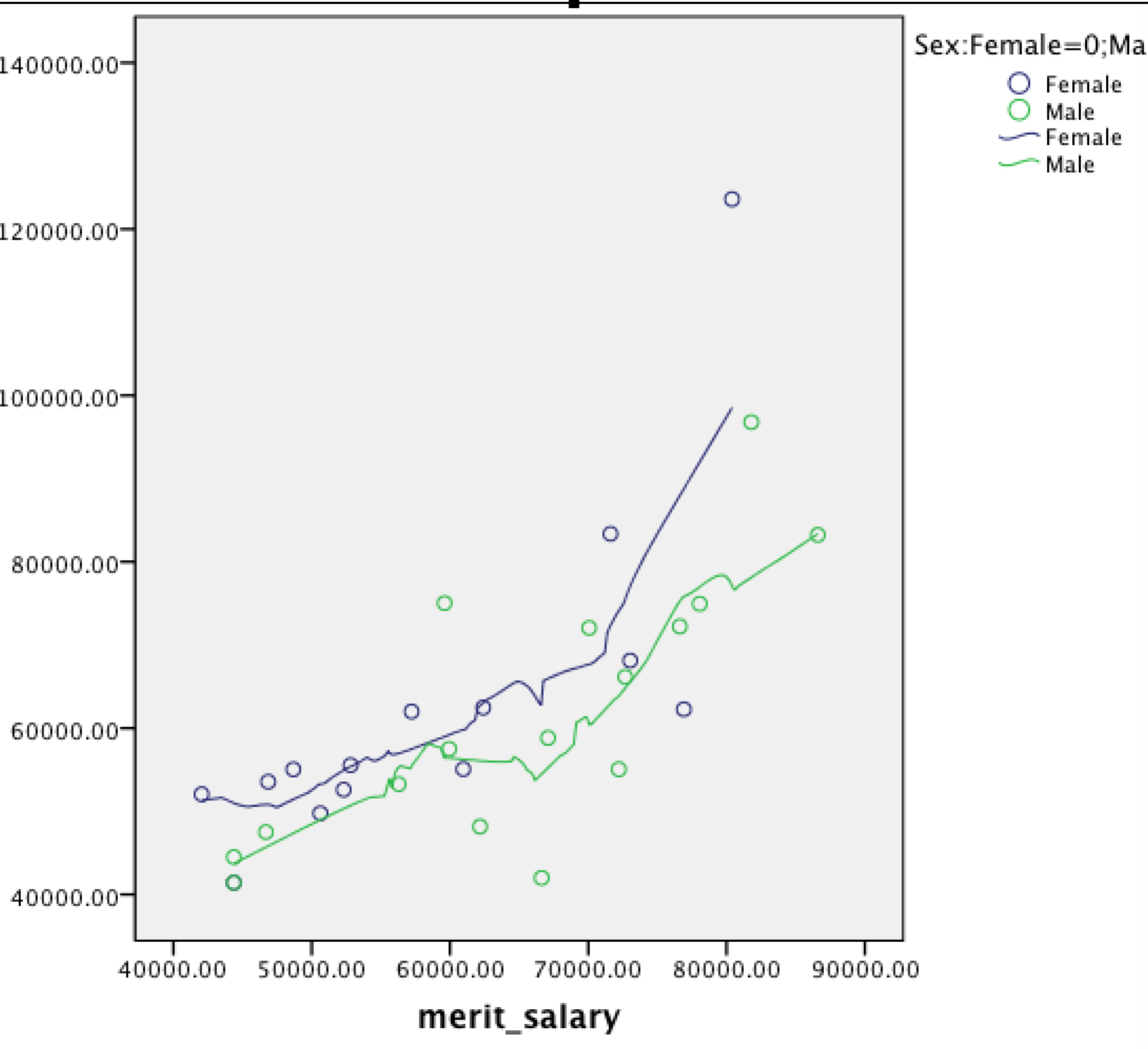
Graphs - legacy - scatter - simple scatter

Y: salary

X:merit\_salary

set markers by: sex

Double click on graph and select fit line at subgroups (either linear or loess)



Ultimately it a small sample size so probably not too much should be read into it.

## (b) run another ANCOVA with sex as fixed factor and your composite as covariate; examine homogeneity of regression slopes assumption; examine marginal means for sex.

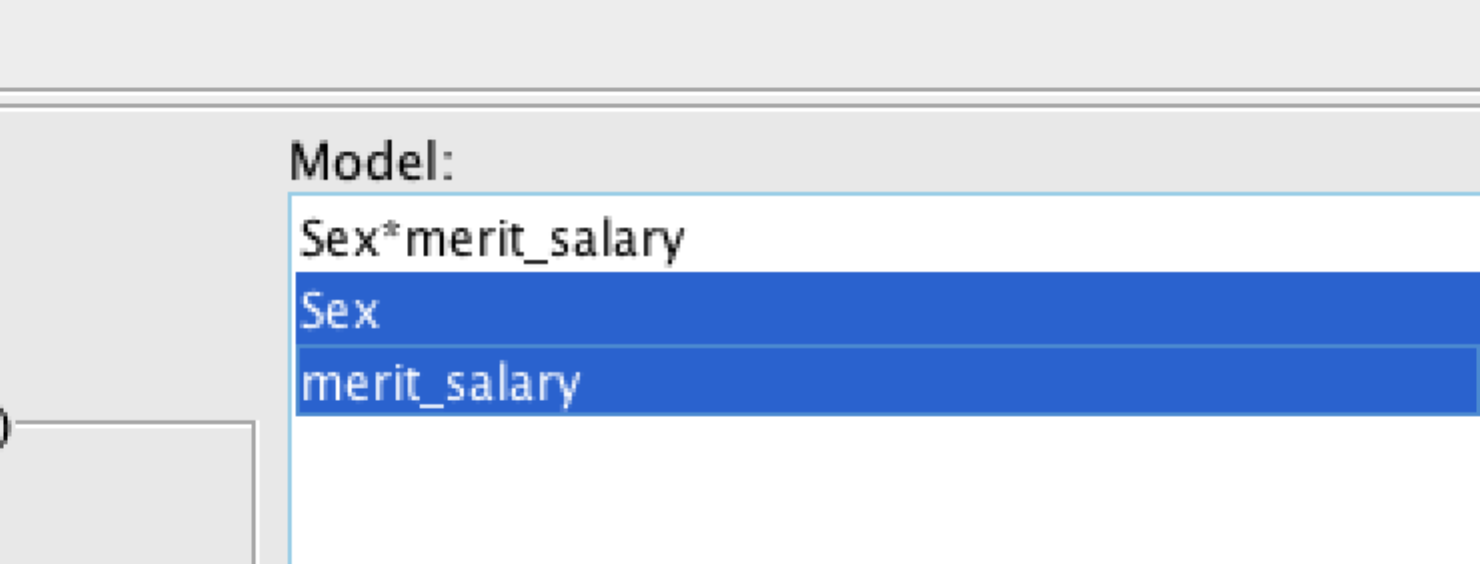
Analyze - GLM - Univariate

DV: salary

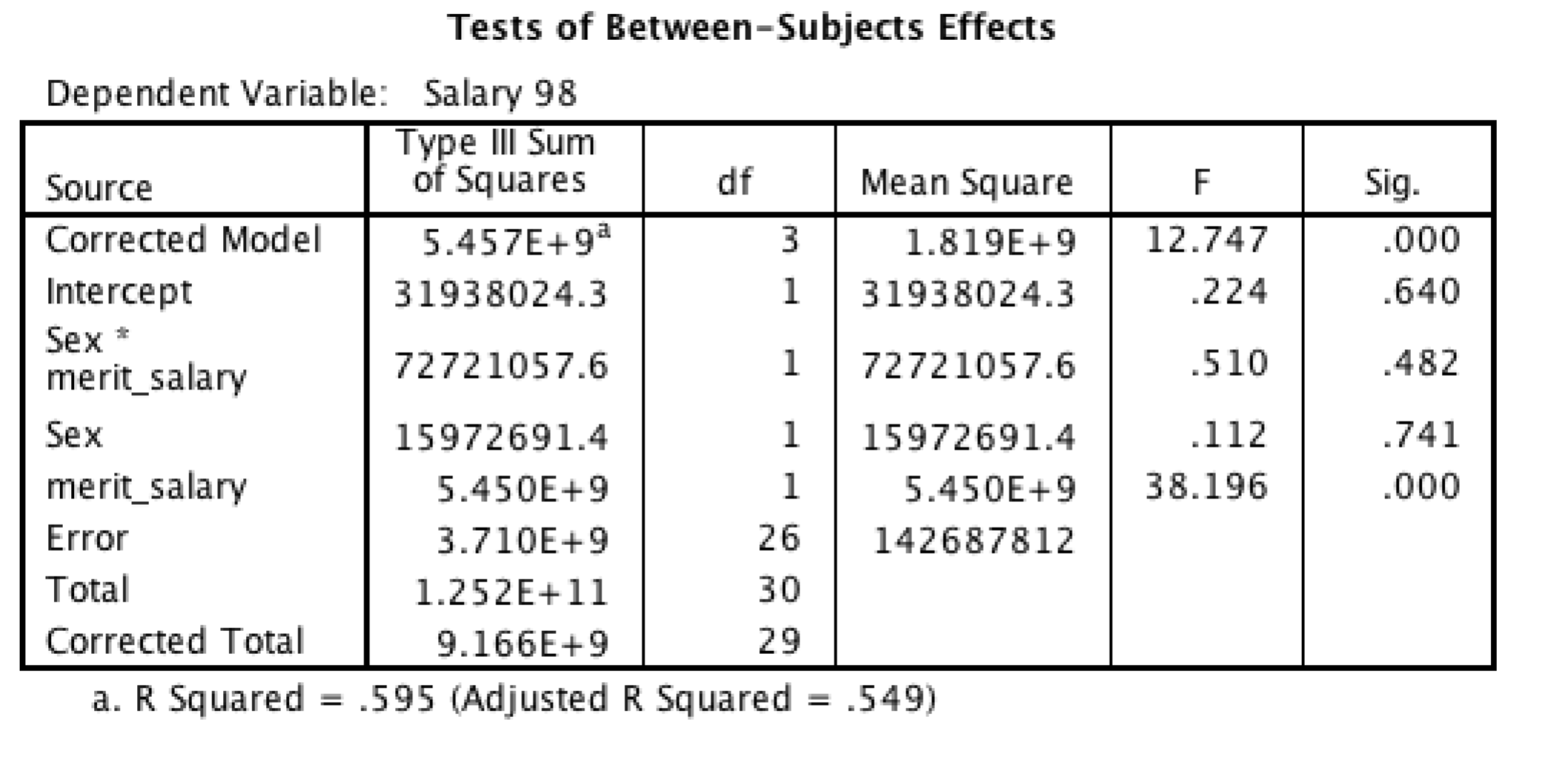
Fixed factor: sex

Covariate: merit\_salary

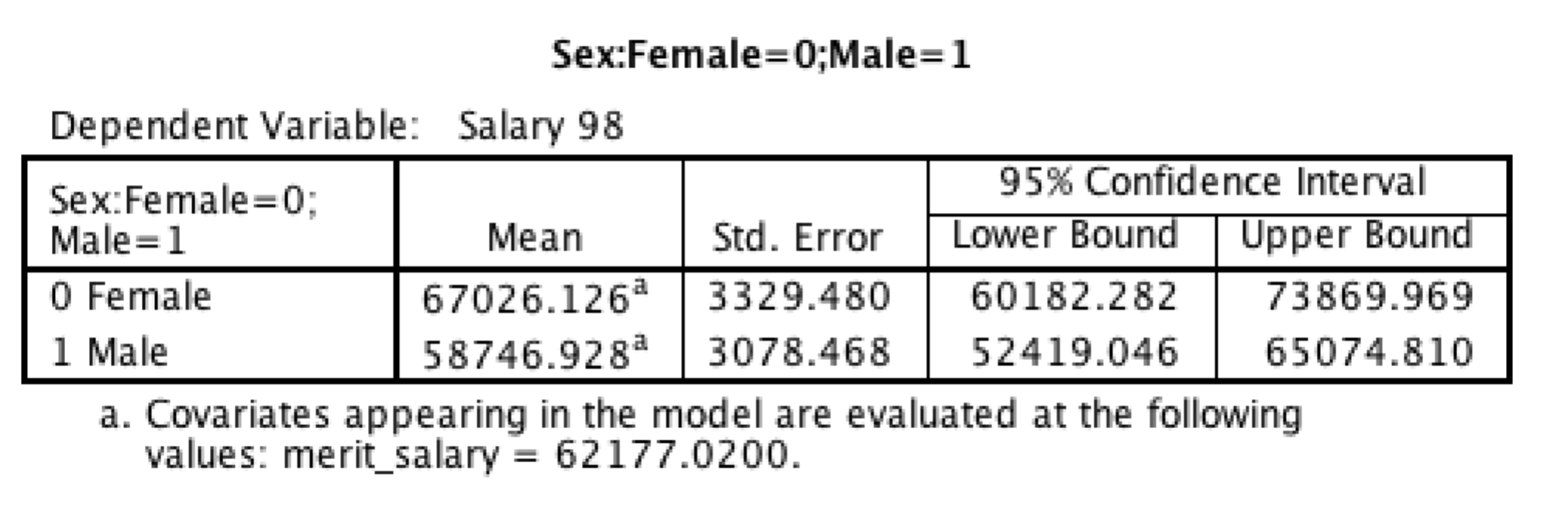
Model: custom: add main effects and interaction ( this is required because by default interactions between covariates and fixed factors are not supplied)



Options: get marginal means for sex



Ultimately there is no significant interaction between sex and merit salary so, you might retain the homogeneity of regression slopes relationship. I.e., the relationship between "operationalised merit "and pay is the same for males and females.



## 10. Provide an overall summary of what can be said about sex differences in salary in this sample and whether any differences appear justified or not. Note any limitations or complexities.

Overall, there were no significant differences between males and female salaries before or after controlling for covariates. There were a number of complexities related to adjuncts that meant that they really needed to be excluded or modelled separately. The sample size was also inadequate for complete tests.

Removing the adjuncts and controlling for the covariate seemed to flip the tendency in the data from males being paid more to females being paid more. That said, there were some general intercorrelations in the data to suggest that women in the sample tended to have spent less time in academia. Thus, at the risk of over analysing the small sample size, the data was more consistent with a model suggesting that possibly females are slightly more likely to drop out of academia and that the shift to the greater inclusion of women in academia is a historical trend, which will require time to work through the system. That said, the sample size is small and probably shouldn't be over-analysed.