Difference in salary among science Doctorate

Ceesay, Ebrima K

"LA SAPIENZA" UNIVERSITY OF ROME

15 June 2012

LA SAPIENZA UNIVERSITY OF ROMA

Name: Ebrima k. Ceesay
Mat: 1504056

Difference In Salary Among Science Doctorates

Abstract

The paper examined the salary difference among science Doctorates for 2004 and 2006.The variation of certain factors such as experience, gender, marital status, Publication, Presentation, hiring and difference in sciences faculties (like natural sciences, social sciences etc.) leads to differences in salary at the end of the doctorate program. The model found out that male and female have different salary due factors such as early marriage, child bearing and household responsibility reduced female salary than its male counterpart even though they all have the same Doctorates. By using factors influences this, we see despite the entire above determinant, geographic region also play key role for salary difference among sciences doctorate. For example, African pays more economics student than Medical student, but we can see American pays more on Medical students than Economics student. Overall, the result found out that doctorate candidates completed in 2004 there is statistically evidence that there will be salary differences than candidates graduated in 2006.

Key Words: Doctorate programs, salary, gender differences, locations, competency.

Introduction

My investigation is basically forced on salary difference amongst science doctorate. The data analyzing is interested, because it come out with a point that explained why there is a salary difference in difference faculty of sciences with considering factors that leads to either positive

impact on my dependant variable or negative impact on explanatory variable which I basically look at gender, Field of study in difference science and type of work contract and make the model simple and easy to explain. The data dropped some of the variable because of insignificance explanation from them. I also look keenly for the variable simultaneously . I did compared with value with Alpha, but we I cannot reject the evidence that there will be a salary difference between faculty of sciences, but for individuals faculty there will be highly significance or insignificance level due to some are significance and other are insignificance due to different in student t- value. Thus, the intercept and some variables are highly significance. Thus the stand error also explained how deviation there will be in term of salary in different faculties of sciences..In the same vein with p value and t statistic I identified the significance of the data. . I also notice that multiple R-square results is very small to explain what will be remaining for unexplained by the data.

For instance, I did explain my data based on result obtained and not go beyond. I started analysis how independent and dependent are treated by try to explain their relationship one after the other based on Multiple R squared, p-value, adjusted R-square, t value and even the outlier that is determined by residual and so on. Thus I look at the rate or changes of coefficient of beta's related to dependent variable which I used Earnings(Salary). I will now begin my analysis one after the other by explains its coefficient and how it relates to independent variable.

Analysis

 The equation and the table below demonstrated the data:

 Monthly Earnings(Y) = β_0 +$\beta_1$field of science +β_2 Gender +β_3Contract Type +ε------(1)

Medical Science(Y)= β_0 +β_1N+β_2A +β_3E +β_4H +β_5S + β_6P + ε ------------(2)

Where N=Natural Science
A= agricultural Science
E=Engineering and Technology
S=Social Science
H=Humanities
P=political Science and Social Science

University evaluation(Y)= β_0 +$\beta_1$2+$\beta_2$3 +$\beta_3$4+ $\beta_5$5 + ε ============(3)

$H_0 : \beta$=0 or difference from zero(0)

$H_1 : \beta =$ (difference from zero(0))

Regression Results For 2004 --- -

Call:
lm(formula = log(PHD2004$Earnings) ~ factor(PHD2004$FieldScience) +
    factor(PHD2004$Gender) + PHD2004$empl.ty + factor(PHD2004$UnivDegreeEvaluation))

Residuals:
     Min   1Q   Median   3Q   Max
-1.02411 -0.19271 -0.01728 0.14059 1.45954

Coefficients:	Estimate	Std. Error	t-value	Pr(> ItI)
(Intercept)	9.10442	0.36354	25.044	< 2e-16 ***
factor(PHD$FieldScience)1	-0.16533	0.05273	-3.135	0.001826 **
factor(PHD$FieldScience)7	-0.28571	0.07945	3.596	0.000358 ***
factor(PHD$FieldScience)8	-0.25076	0.05721	-4.383	1.45e-05 ***
factor(PHD$FieldScience)10	-0.35680	0.06009	-5.938	5.71e-09 ***
factor(PHD$FieldScience)12	-0.11935	0.06899	-1.730	0.084329 .
factor(PHD$FieldScience)14	-0.22641	0.10019	-2.260	0.024300 *
factor(PHD$Gender)2	-0.13215	0.03444	-3.838	0.000142 ***
factor(PHD2004$empl.ty)r	-0.23078	0.04747	-4.861	1.60e-06 ***
factor(PHD2004$empl.ty)t	-0.32971	0.04648	-7.093	4.99e-12 ***
factor(PHD2004$UnivDegreeEvaluation)2	-1.36941	0.36735	-3.728	0.000217 ***

factor(PHD2004$UnivDegreeEvaluation)3	-1.50897	0.36427	-4.142	4.09e-05 ***
factor(PHD$UnivDegreeEvaluation)4	-1.38303	0.36417	-3.798	0.000166 ***
factor(PHD$UnivDegreeEvaluation)5	-1.41665	0.36139	-3.920	0.000102 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.359 on 459 degrees of freedom
  (100 observations deleted due to missingness)
Multiple R-squared: 0.2525,   Adjusted R-squared: 0.2313
F-statistic: 11.92 on 13 and 459 DF, p-value: < 2.2e-16

> anova(Output.2004)
Analysis of Variance Table

Response: log(PHD$Earnings)	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(PHD$FieldScience)	6	6.333	1.0555	8.1895	<1.975 e-08 ***
factor(PHD$Gender)	1	2.873	2.8733	22.2929	3.114e -06 ***
PHD2004$empl.ty	2	8.269	4.1344	32.0770	9.131e -14 ***
factor(PHD$UnivDegreeEvaluation	4	2.504	0.6260	4.8572	0.0007

			614 ***
Residuals			0.1289
	459	59.160 1	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

I did start my analysis as follows for 2004; we can see directly the result for all the independent variable is significance in the model. We notice for example there is negative correlation between medical science and others field of science on gender, contract type (Which I did group into p, r and t) and university evaluation.   Since it is clearly that those in natural science will receives less than those in other fields of science indicate inversely relationship in terms of salary difference. I did apply causality in the model to see the significance of the variables.

First I controlled contract type and Field of sciences I look clearly the salary difference in **gender on field of science** and contract type. I notice that based on gender there is inversely relationship in salary in different field of science for example the result tell us that there is (16.533%) reduction for natural science meaning female receive at least 13% of salary and male receive almost 87% of the salary in natural science. With very small p value less than α of 0.05% the result is that there is enough evidence that there will be a salary difference in gender on natural science in 2004. In the same vein, The salary gap for medical science is highly significance, because the result tell us that is the reference category and it has higher regression and is positive related to salary on gender of 13% less on female and male 87%, so the huge result may be due to hiring, reproductive for gender and the time it takes for Doctorate students in field of medical to complete their studied and so on. Further, comparing p value with alpha of 0.001%, which is 99%, rejection of null hypothesis that there is salary difference in medical science based on gender. Further, the coefficient of Engineering and Technology tell us that there is 25% less in salary difference in this field of science with equivalent 13% of less is for female and the rest for male and so on. All this sciences indicated that male receives more than female in salary in 2004. I also notice there is a disturbance variable which has negative impacts on my outlier which I will explained below for the boxplot. These outliers tell that there is high means difference in salary based on gender in difference field of sciences if you compare them with medical science. The result for this analysis also indicate type 1 error, because I accept the true when it is obviously false.

Second point is that Salary difference on contract type based on gender. I did divide contract type in to p as(permanent position,National employmet contract, and tempin agency) and t as (Training or apprenticeship contract, Project contract, Contract for occasional work) and r as(Post-doc position, Research grant, Research scholarship) From the result above we see all coefficient for contract Type are negative(indicating reduction on salary based on contra type) and only the reference categories which is permanent position is significance. Despite no promotion and otherwise we see the result is inconsistent with modern life, because high p value with corresponding lower t value also indicate proof that there is not enough evidence to prove based on permanent position for female and male. The coefficient for permanent position is negative meaning there is negative independent variable which showed that there is negative rate of salary difference on permanent position. There is 99% to reject this results. For instance, for temporary position, despite discrimination or otherwise female loss their jobs more than male in others words female changes their jobs frequently than their male counterpart. The intercept indicate positive, if i make the other variable as constant zero to see the rate of change on salary.

Inference or Rejection Region;

I now look how the regression variable affects the result of my analysis positivity or negativity for 2004. Thus, I did obtain low r square which mean that there is small variation in r explained the results and obtained higher value of F test demonstrated the rejection for null hypothesis and ANOVA table also showed that how F test is differed from different field science, gender and contract type and university evaluation. Below will indicate how gender salary gap are explained and unexplained by low r square. I looked gender gap in salary in different field of science and see that the gender gap in social Science is the smallest of all the field of science disciplines considered almost 11 percent. I also notice the gender gap can be decomposed into two components on explaining r square one explained the resulting from gender salary difference in different field of science, and contract type(p,t,r) in observed characteristic and the unexplained gap resulting gender difference in difference field science and contra type in estimated coefficient(β 's).The estimated coefficient indicate that there is a reduction in salary for gender in different doctorate faculties and this mean that there is very low r square that explain there is smallest variation in salary difference on gender in sciences, which is approximately 25 % and the remaining approximately 75 % unexplained. The results given by r square statistically is not well predicts the behaviours of the

model. This low value of r certainly have inconsistent result may be the movement have some positive impact on salary based on gender. I now turn my attention why higher F statistics value, as we know I did   reject $H_0 : \beta$=0 or difference from 0. Because larger F value of almost 11.92 with very low critical value Alpha of 0.1(10%), we reject the null hypothesis that there is significance evidence for salary difference based on gender in science faculties, and in contract type and university evaluation. I notice also because of higher F value I reject null hypothesis even I test on difference on 1 (one) it will   produce a result that there will be enough evidence to reject the null hypothesis. Thus, r square is low some field of sciences are weakly correlated to the others for example medical and social science and there is lots of noise .The r square with corresponding standard error could not interpret well for the salary difference of science doctorate. But we see with highly significant intercept may means the model obtain goodness of fit. The academic performance is highly matter for salary difference in 2004 because the result is highly significance in the data. I point out that if you compare the result in term of decrease in percentage we see that those doctorate received 66-90 will receive higher salary than the others. For example I notice there are 137% reduction of salary for those candidate who score a marked of 91-100 and 150% for 101-105.All the result for academic performance is highly significance and it show positive salary difference in gender on difference field of sciences.

Finally, since small p value we reject null hypothesis. For dropping and added variables helps me to identified significance of my data and indicate there will be salary difference in science doctorate.


Scatter Plot for 2004
The scatter diagram below show positive correlation for most variable in 2004. That is they are directly related to one another.
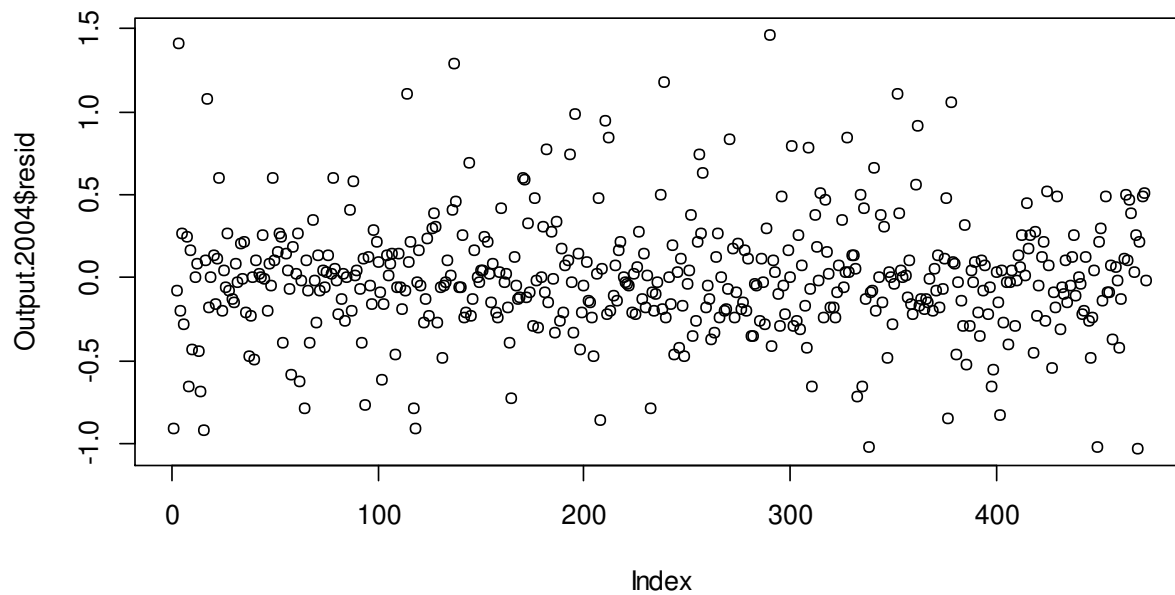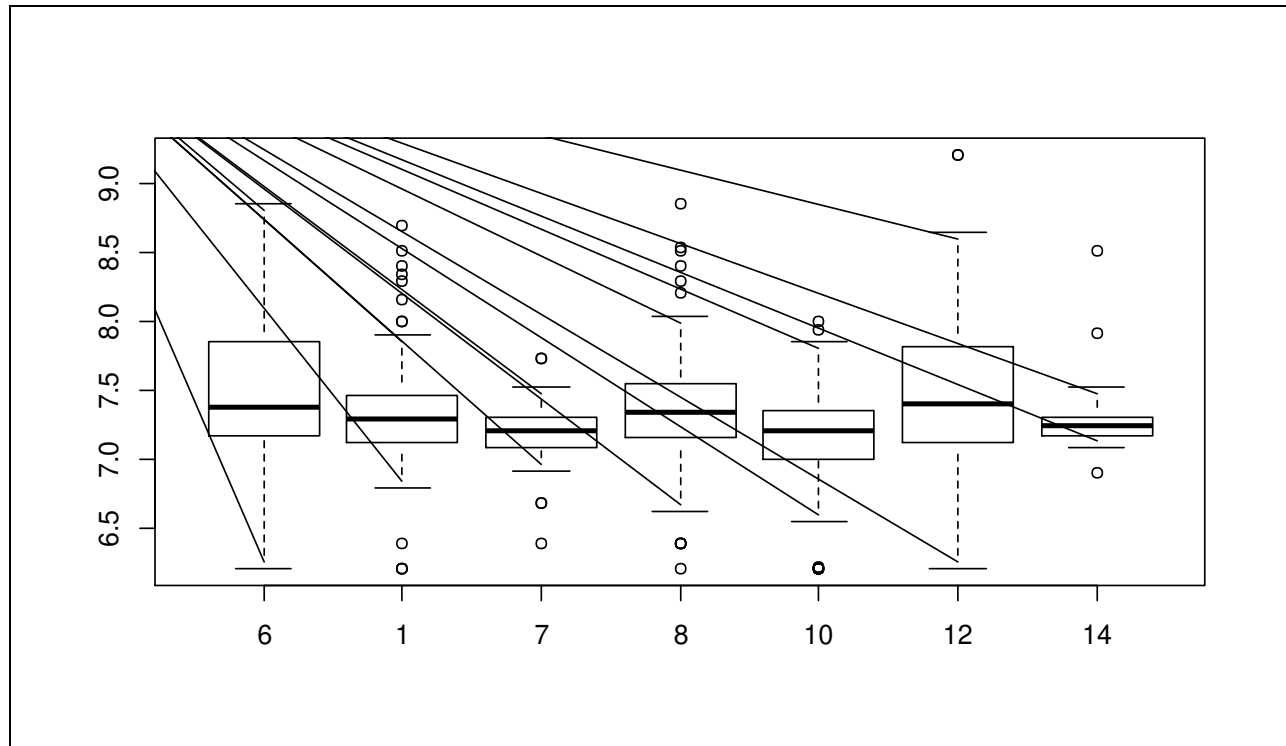
The histogram below indicates normality amongst variables. Meaning there exist normal distribution between earning as dependent variables and all others variables as independent variables on the right. This pinpoint out for variables in 2004 there exist a positive impression and positive result. It is well symmetry.

Histogram of Output.2004$resid

Output.2004$resid
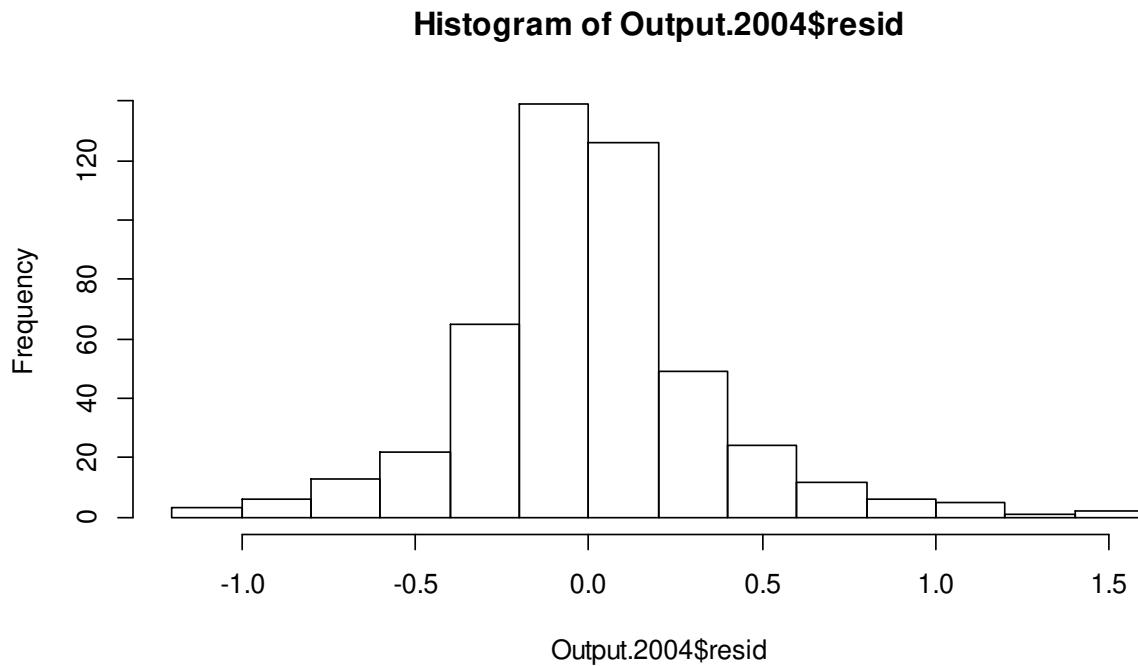
Analysis of Boxplot how Independent Variables influences Earnings (Dependent variables) difference for doctorate sciences salaries. From Table it indicates the sample sizes for categorical variables and Minimum to maximum for Numerical variables Earnings.

|	Earnings	S	Gender	emp1.ty	UnivDegreeEvaluation
min	500	A	1:277	p:329	1: 3
1st Qu	1233	M	2:296	r:73	2: 34
Median or second Qu	1500	P		t:76	3:60
				NA's:95	
Mean	1731	L			4:63
3rd Qu	1850	E			5:413
Max	9999	Sizes			
	NA's:41				

1 Field Sciences:

I see that the boxplot make it clear how some of variables are positive related and others negative related to earnings. I see the variables for medical science is positive and the symmetry is well located and it has a reasonable more more significance than Compared to other field of sciences fore 2004. I notice that the prove for rejection does not means that the alternative hypothesis will be fully accepted for my analysis. It make due to certain variation that after putting we will see that the box plot for the others sciences field are also well symmetry and positive compare to medical sciences in salary difference. This positive test result indicated that medical science doctorate receives more Earnings than the other fields of science. The outlier for humanities and social science are spread far from indicating that there is outlier, which makes measurement errors in this field of sciences.

2. Contract Type

I notice also that contract type p(permanent position)is positive and it indicates that the maximum, mean and quartile value are reasonable for rejection of null hypothesis.   This means that result can used median for salary difference for science doctorate. Thus the median is not affect by outliers and positive for permanent position to be highly significance and it be well documented for the use in this analysis.
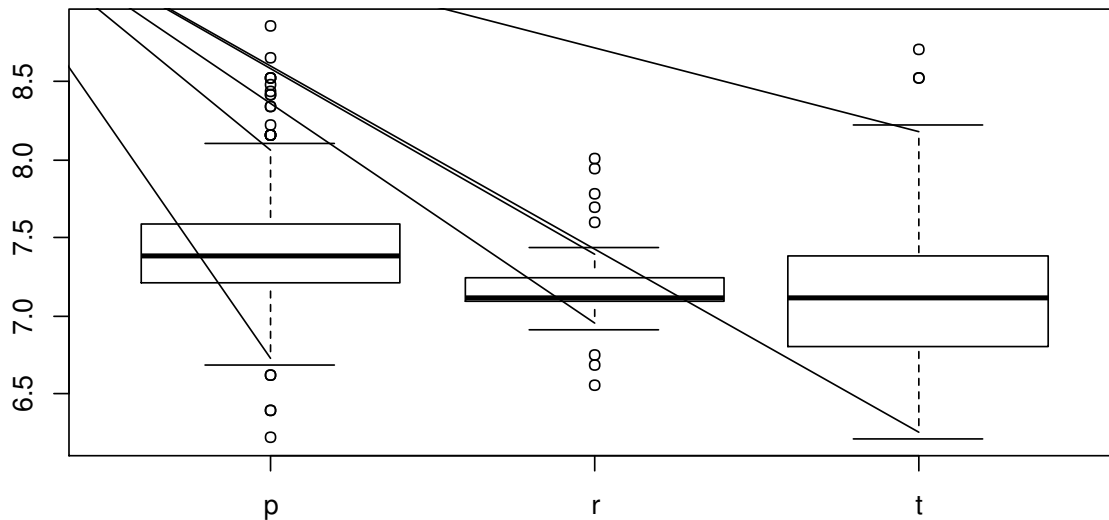
3, Gender
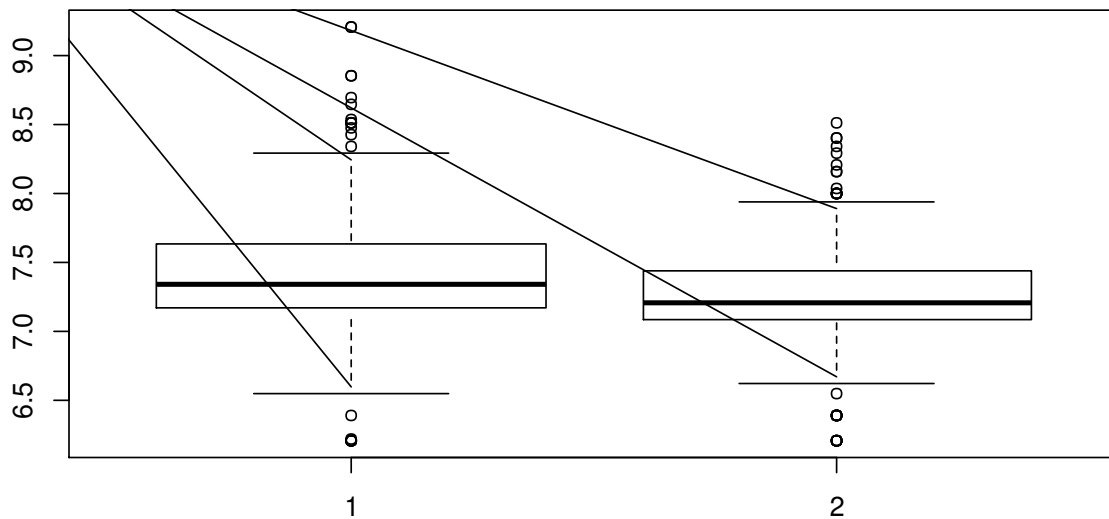
Finally, I see the variation in the two boxplot for male and female respectively for 2004; I notice despite all removing and added value I rejected the null hypothesis that there will be salary difference. But for male has a outlier very far from the median and it show that there may be measurement errors in the data that the rejection region may changes to accept region after identified and detecting the causes of this spread in earnings on gender in difference field of sciences.

Result for 2006

Call:
lm(formula = log(PHD2006$Earnings) ~ factor(PHD2006$FieldScience) +
    factor(PHD2006$Gender) + factor(PHD2006$empl.ty) + factor(PHD2006

$UnivDegreeEvaluation))

Residuals:
    Min   1Q   Median   3Q   Max
-1.12411 -0.17121 -0.00224 0.15348 2.13083

Coefficients:	Estimate	Std. Error	t-value	Pr(> ItI)
(Intercept)	7.85064	0.25427	30.876	< 2e-16 ***
factor(PHD$FieldScience)1	-0.17809	0.04873	-3.654	0.000282 ***
factor(PHD$FieldScience)7	-0.27557	0.07452	-3.698	0.000239 ***
factor(PHD$FieldScience)8	-0.16958	0.05023	-3.376	0.000787 ***
factor(PHD$FieldScience)10	-0.35234	0.05069	-6.951	1.02e-11 ***
factor(PHD$FieldScience)12	-0.15850	0.10321	-2.794	0.005393 **
factor(PHD$FieldScience)14	-0.16586	0.03079	-1.604	0.108635
factor(PHD$Gender)2	-0.08110	0.03689	-2.633	0.008687 **
factor(PHD2006$empl.ty)r	-0.14593	0.04118	-3.956	8.62e-05 ***
factor(PHD2006$empl.ty)t	-0.15141	0.26142	-3.677	0.000259 ***
factor(PHD2006$UnivDegreeEvaluation)2	-0.34234	0.25679	-1.310	0.190888
factor(PHD2006$UnivDegreeEvaluation)3	-0.32629	0.25716	-1.271	0.204374
factor(PHD2006$UnivDegreeEvaluation)4	-0.29612	0.25403	-1.151	0.250024
factor(PHD2006$UnivDegreeEvaluation)5	-0.27297	0.05853	-1.071	0.283044

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.3568 on 556 degrees of freedom
  (144 observations deleted due to missingness)
Multiple R-squared: 0.13,   Adjusted R-squared: 0.1097
F-statistic: 6.391 on 13 and 556 DF, p-value: 2.419e-11

anova(Output.2006)
Analysis of Variance Table

| Response: log(PHD$Earnings)	Df	Sum Sq	Mean Sq	F value
factor(PHD$FieldScience)	6	6.388.	1.05634	17.5787
factor(PHD$Gender)	1	0.737	0.773718	24.6869
factor(PHD$ContractType)	2	3.071	1.53542	18.6293
factor(PHD$UnivDegreeEvaluation	4	0.428	0.10693	2.8208
Residuals	556	70.763	0.12727	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

For 2006 result is similar to 2004 except for academic performance which is not significance at all for 2006. This means that it those not matter academic performance cannot determined the salary difference among science doctorate. For simplicity, the analyses show that there is a reduction of almost 18 percent   in natural science than medical science. The coefficient   is negative explain that earnings for doctorate in medical science increases, the earning in other field of science decrease significantly respectively. The reduction is 28 percents,

Agriculture, 17 percent in Engineering and Technology, 35% in Humanities, 16% in social sciences and 17% in political sciences respectively. For gender is medium significance among to 8%.This reduction tell us that the results for 2006 is biased against female sciences doctorate and male will receive approximately 92% of the salary in field of sciences. Further, for contract type p compared to r and t we see that there exist a significance level of the variables but reduction of approximately 15% and 15% respectively compare to permanent potion ,p.The p value is larger than alpha of 5%, then I accept(do not reject) null hypothesis than there is not enough evidence for salary difference among science doctorate.
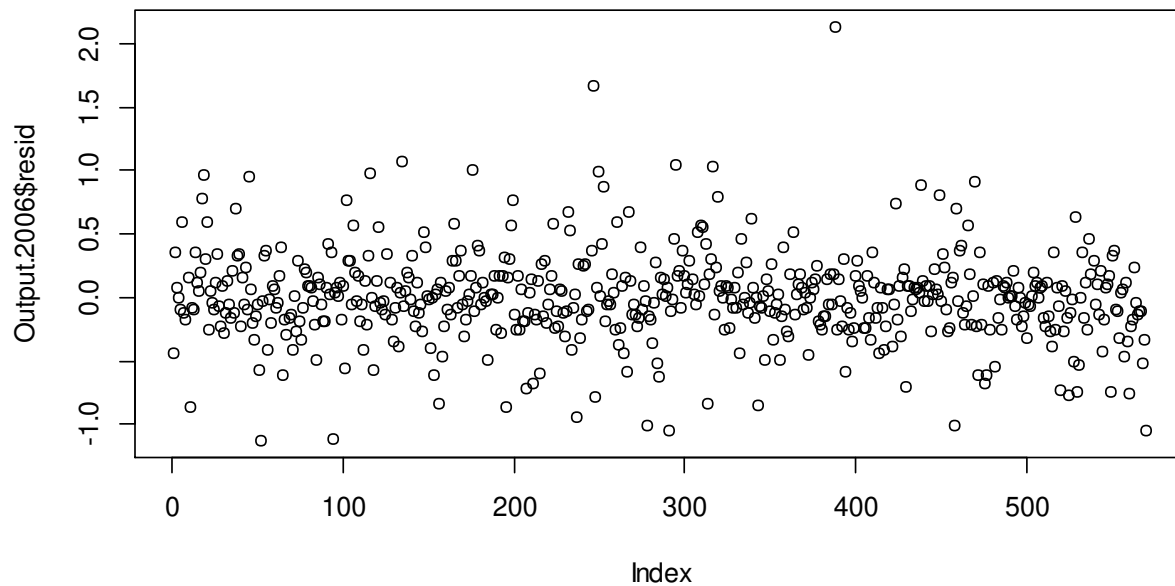

Histogram,
For 2006 is also well shape but may be due observation error we do not reject null hypothesis.

Histogram of Output.2006$resid

Field science

We notice here that for example the boxplot haminities in 2006 has a far outlier compared to 2004 for the same humanities. Tis indicated also wy we accepted the null hypothesis for salary difference in difference field of science
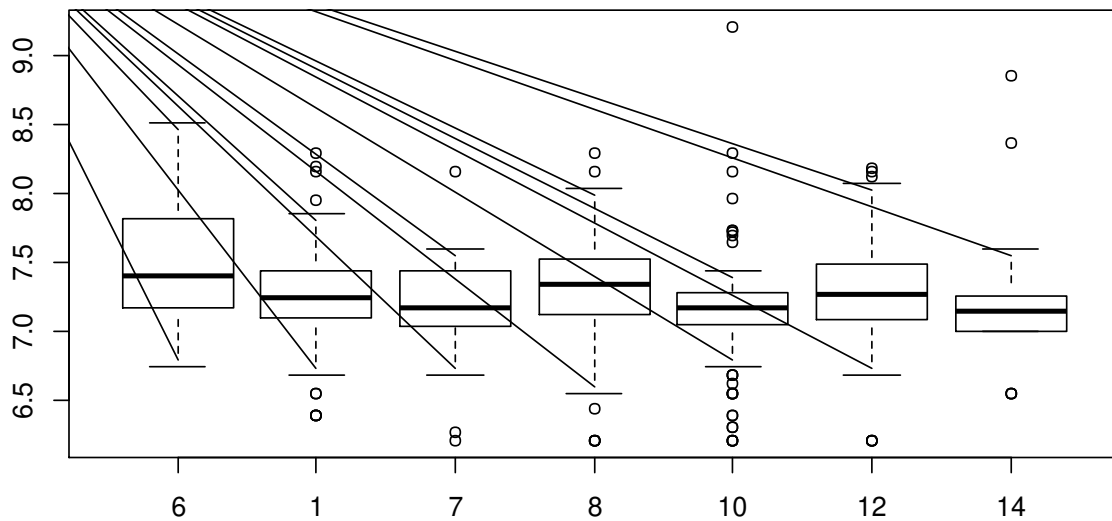
2.
Em pty1
I notice also that contract type p(permanent position)is positive and it indicates that the maximum, mean and quartile value are reasonable for accepting of null hypothesis.We can also see the outlier for r(research position) is very far from the mean and it produced an noise signal.
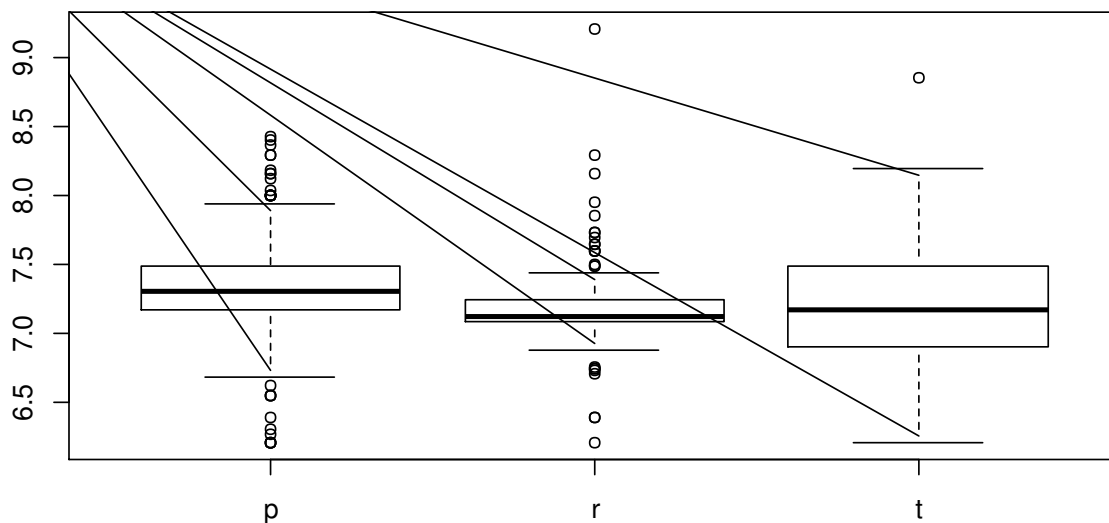
3. Gender
We see also gender one male has larger outlier in 2006 compared tro 2004, which show that it may be due to certain circumstance that male recives far more salary than women due to hiring, discrimination and so on 2006.
Comparing The Result

I notice clearly that the result for 2004 is consistent than 2006 becaose of, first compare Fieldsciences we see that reduction in salary is highly significance than the result for 2006. But we can also ssee the result for Agriculture science remained the same for both years which indicates for salary difference in that fields will be the same for science doctorate.But the result for 2004 is better if we now turned our attention performance, based on the information provided by regression we can notice that it those not matter high or low academic performance will dertermined salary difference. For instance this can be viewed from a very lower result of r square, which only 13%explained, and 87% unexplained the variation in slary difference amongst science doctorate. This may be due to experience, pupolarity, ect.Compareing the rejection and accepting region for 2004 and 2006 respective, we notice that 2004 has lots of evidence for rejection of null hypothesis

that there will be salary difference in science doctorate.This is explained how those variables are significance like university evaluation, fields science, contract type and gender compare to 2006 in which academic performance is not matter for evaluation of salary difference in science doctorate. The comparing result show that those candidate doctorate science candidate completed in 2004 there is a prove that there will be a salary evidence while in 2006 said there may not be salary difference.

Conclusion and Summary

Finally from this paper it explained the salary difference based on gender contract Type and university degree evaluation for doctorate in different field of science like medical science, political and social sciences, social science, Engineering and Technology, Agricultural science, natural science and Humanities. From 2004-2006.The data indicate persistent gender gap, contract and academic performance. Although discrimination, reproductive ect make female to be vulnerable for low salary difference in difference field of science to compared with their male counterpart. Even though the academic rank obviously reduce salary difference it does not entirely explained because of low r square. Salary difference in those doctorate how score a marked of 110 distinction will have 49%less reduction in salary meaning male will gets 51%.. For looking at salary gap for gender in difference science I see that may be discrimination play some role for that because the coefficient for beta tell us that a unit changes in the salary will automatically have impact on categorical reference medical science as explain in equation 3 above..Despite the rejection of these results, raising awareness for different sciences faculty will reduces if not total eradicate gender difference in salary for male and female. I notice that that female in discipline like medical have higher salary than there counterpart male.

Notes: The regression analyses obtained in this study prove problematic in results because of very low r square And corresponding adjusted r especially for 2006.

Reference:

 1. Introduction to statistical Thought by Michael Lavine November 11 2007

 2. Practical Regression and ANOVA using R by Julian J. Faraway, July 2002
3.Gender Difference in salary and promotion in political science by Donna K. Ginther in februart 2004

4. Is MIT an exception? Gender pay Difference Academic Science. By Donna K. Ginther.

5. STAT1401 Probability and Statistical Inference by Dr Julian Stander with contribution from Dr Hilary Sanders September 23, 2010.