1: Introduction

In this project, we are studying about five variables with a sample size of 250. The variables in consideration are beginning salary, Educational Level, Age of Employee, SEX of Employee Minority classification. Here the Beginning Salary is the dependent variable. Educational Level, Age of Employee, Sex of employee and Minority classification are independent variables. We are going to fit the sample size of 250 in an Estimated Multiple Linear Regression model with the help of SPSS software. Then from the output data of the Linear Regression model we are going to construct a Multiple Linear Equation model. We are going to interpret each of the coefficients of the independent variable and how the change of each unit affects the dependent variable. We are going to find out how strong the Multiple Linear Regression is with the help of "**Coefficient of Multiple Determination**". With the help of our model we are going to predict an outcome. Finally, we are going to determine the confidence intervals for each of the regression coefficients.

2: Estimated Regression Equation

The following data below is obtained from the output of the SPSS Multiple Linear Regression Model Output:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2791.476	1041.513		-2.680	.008
	EDUCATIONAL LEVEL	637.641	54.985	.586	11.597	.000
	AGE OF EMPLOYEE	51.717	13.178	.180	3.925	.000
	SEX OF EMPLOYEE	-1555.560	321.527	-.239	-4.838	.000
	MINORITY CLASSIFICATION	-961.135	380.388	-.118	-2.527	.012

a. Dependent Variable: BEGINNING SALARY

Suppose, we let the dependent variable “Beginning Salary” be y.

We let “Educational Level” be x_1

We let “Age of Employee” be x_2

We let “Sex of Employee” be x_3

We let “Minority Classification” be x_4

$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$

From the table above, a = -2791.476, b_1= 637.641, b_2= 51.717, b_3= -1555.560,

b_4= -961.135

Hence the Estimated Regression Equation is:

$y = -2791.476 + 637.641x_1 + 51.717x_2 - 1555.560x_3 - 961.135x_4$

3: Explaining the meaning of Estimated Regression Coefficients

The following data below is obtained from the output of the SPSS Multiple Linear Regression Model Output:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2791.476	1041.513		-2.680	.008
	EDUCATIONAL LEVEL	637.641	54.985	.586	11.597	.000
	AGE OF EMPLOYEE	51.717	13.178	.180	3.925	.000
	SEX OF EMPLOYEE	-1555.560	321.527	-.239	-4.838	.000
	MINORITY CLASSIFICATION	-961.135	380.388	-.118	-2.527	.012

a. Dependent Variable: BEGINNING SALARY

In the previous question we found our Multiple Linear Regression to be:

$$y = -2791.476 + 637.641x_1 + 51.717x_2 - 1555.560x_3 - 961.135x_4$$

1. **Constant:** The constant of -2791.476 is the Y-intercept, which means, if all the independent variables x_1, x_2, x_3 and x_4 is set to zero then the dependent variable Beginning salary would be $-2791.476. In a real life scenario it does not make sense that the salary would be negative, it also does not makes sense that someone at the age of zero will go for a job. This value only represents the y-intercept for this model, it has no real life meaning.

2. **b_1:** The b_1 is the coefficient of x_1 or Educational Level in the estimated regression model. The value of b_1 is 637.641. This means that if x_1 increases by one unit while the other independent variables are constant, the dependent variable y would increase by 637.641. In real

life, if the Education Level increases by 1 year while the other independent variables are constant, then the beginning salary would increase by $637.641.

3. **b_2:** The b_2 is the coefficient of x_2 or Age of Employee. The value of b_2 is 51.717. This means that if x_2 increases by one unit, then the dependent variable increases by 51.717. In real life, this means that if the Age of an employee increases by one while the other independent variables are constant, then the salary of an employee increases by $51.717.

4. **b_3:** The b_3 is the coefficient of x_3 or Sex of Employee. The sex of employee can either be Male or female. This can be represented by dummy variables 1 or 0. The value of b_3 is -1555.560. Suppose the female gender is represented by 1 and male is represented by 0. So if the x_3 = 1 keeping the other independent variables constant, which represents female, then according to the coefficient b_3, the beginning salary of the employee would decrease by $1555.560.

5. **b_4:** The b_4 is the coefficient of x_4 or Minority Classification. This is also a categorical variable white is the Majority while the rest such as black, Hispanic etc. are minority. Minority is represented by 1 while the majority is represented by 0. The coefficient b_4 is - 961.135. This means suppose if the person is Hispanic, then x_4 would be 1. And the beginning salary of the person would decrease by $961.135.

4: value of coefficient of multiple determination

The following data below is obtained from the output of the SPSS Multiple Linear Regression Model:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.713[a]	.509	.501	2258.299

The value of $\mathbf{R^2}$ is known as **coefficient of multiple determination.** From the table above the value of coefficient of multiple determination is 0.509

Interpret of Coefficient of Multiple Determination:

From this value we can interpret that

0.509 × 100% = 50.9%

50.9% of the dependent variables can be explained by this Estimated Multiple Linear Regression model.

5: predicted beginning salary for the employee who is white male of 40 years and has 16 years of education

From part 2, our estimated multiple linear equation is:

$y = -2791.476 + 637.641x_1 + 51.717x_2 -1555.560x_3 - 961.135x_4$

So here:

Education = 16 years

Age= 40

Sex = Male

Minority = Non-minority or white

Hence, $x_1= 16$, $x_2 = 35$, $x_3= 0$, $x_4 = 0$

$y= -2791.476+ 637.641(16) + 51.717(40) – 1555.560(0) – 961.135(0) = 9479.46$

Hence, according to our Estimated Multiple Linear Regression Model, the salary of the employee would be $9479.46

6: 95% confidence intervals for the regression coefficients

In this sample we are with a sample size of 250 and 4 independent variables.

n = 250 and k= 4

Degree of Freedom = n-k-1 = 250 – 4 -1 = 245

The confidence level = 95%.

Area in each tail of the t distribution = (1 - .95) / 2 = .025

For df= 245 and an area of 0.025 in the right tail, the value of t= 1.970

We are going to obtain the values of s_{bi} corresponding to each coefficient from the table below

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2791.476	1041.513		-2.680	.008
	EDUCATIONAL LEVEL	637.641	54.985	.586	11.597	.000
	AGE OF EMPLOYEE	51.717	13.178	.180	3.925	.000
	SEX OF EMPLOYEE	-1555.560	321.527	-.239	-4.838	.000
	MINORITY CLASSIFICATION	-961.135	380.388	-.118	-2.527	.012

a. Dependent Variable: BEGINNING SALARY

1. **95% confidence level of b_1:**

$b_1 \pm ts_{b1} = 637.641 \pm (1.970 \times 54.985) = 637.641 \pm 100.320$

We can state with 95% confidence that with one year increase of educational level, the beginning salary increases by $537.321 to $737.961

2. 95% confidence level of b_2:

$b_1 \pm ts_{b2} = 51.717 \pm (1.970 \times 13.178) = 51.717 \pm 25.960$

We can state with 95% confidence that with one year increase of age of employee, the beginning salary increases by $25.757 to $77.677

3. 95% confidence level of b_3:

$b_1 \pm ts_{b3} = -1555.560 \pm (1.970 \times 321.527) = -1555.560 \pm 633.408$

We can state with 95% confidence that if the employee is a female ($x_3 = 1$) then the beginning salary decreases by $922.152 to $2188.968

4. 95% confidence level of b_4:

$b_1 \pm ts_{b4} = -961.135 \pm (1.970 \times 380.388) = -961.135 \pm 749.364$

We can state with 95% confidence that if the employee is a Minority or non-white ($x_4 = 1$) then the beginning salary decreases by $211.770 to $1710.499

7: Discussion and conclusion

In this project we have analyzed one dependent variable "Beginning Salary" (y) versus four independent variables: Educational Level (x_1), Age of Employee (x_2), Sex of Employee (x_3) and Minority Classification (x_4). Our sample size was 250. The coefficient of x_1 is 637.641, it has a positive correlation with the dependent variable y, meaning that with one year increase in educational level the Beginning salary increases by $637.641. The coefficient of x_2 is 51.717, it also has a positive correlation with the dependent variable y, and it means if the age of employee increases by one year, then the beginning salary increases by $51.717. The coefficient of x_3 is -1555.560, it has a negative correlation with y, 0 represents male and 1 represents female, if the employee is a female then the salary of the employee decreases by $1555.560. The final independent variable x_4 is the Minority Classification with coefficient
-961.135, if the employee is non-white then x_4= 1, if the employee is white then x_4=0, hence if the employee is nonwhite, then their salary decreases by $961.135.

Let's have a significance level of α = 0.05

The p-value/significant level of the independent variables x_1, x_2, x_3 and x_4 are all 0, which means that they are far less than α = 0.05, hence they are significant. The p-value of Minority classification is 0.012 which is also far less than α, hence it is also important.