Model 1 Output:

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: ahe

Number of Observations Read	7711
Number of Observations Used	7711

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	23006	23006	230.43	<.0001
Error	7709	769646	99.83730		
Corrected Total	7710	792651			

Root MSE	9.99186	R-Square	0.0290
Dependent Mean	18.97609	Adj R-Sq	0.0289
Coeff Var	52.65500		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.08228	1.18426	0.91	0.3608
age	1	0.60499	0.03985	15.18	<.0001

- Our coefficient estimate can be interpreted as: As workers age increases by one year, average hourly earnings increase $0.6499
- Our R^2 value of 0.0289, which means that only 2.89% of the variation in earnings is described by age helps hint that age does not account for a large fraction of the variance in earnings across individuals. But, such a lowR^2 value could also be due to omitted variables in our model, since we see that age is statistically significant at a 1% level, meaning that we should include other dependent variables in our model or control for these omitted variables.

Means Procedure Output:

The SAS System

The MEANS Procedure

Variable	Mean	Corrected SS	Variance
ahe	18.9760924	792651.46	102.8082303
ahe_hat	18.9760924	23005.74	2.9838829
u_hat	2.456952E-14	769645.72	99.8243474
age	29.5772273	62855.76	8.1524982

From our means procedure, you can tell that ahe and ahe_hat are not identical. Although the mean is the same for both ahe and ahe_hat, ahe has significantly more variance than ahe_hat. This proves that although the average value across them may be the same, ahe's data points deviate significantly more from the mean than ahe_hat

- Mean:
  - Ahe = 18.9760924
  - Ahe_hat = 18.9760924
- Standard Deviations:
  - Ahe = 10.13944
  - Ahe_hat = 1.727392

Correlation Output:

The SAS System

The CORR Procedure

4 Variables:	ahe ahe_hat u_hat age

Pearson Correlation Coefficients, N = 7711
Prob > |r| under H0: Rho=0

	ahe	ahe_hat	u_hat	age
ahe	1.00000	0.17036 <.0001	0.98538 <.0001	0.17036 <.0001
ahe_hat	0.17036 <.0001	1.00000	0.00000 1.0000	1.00000 <.0001
u_hat	0.98538 <.0001	0.00000 1.0000	1.00000	0.00000 1.0000
age	0.17036 <.0001	1.00000 <.0001	0.00000 1.0000	1.00000

- U_hat & ahe_hat => 0

- This correlation makes sense because U_hat represents the residual of predicted value ahe_hat. When deriving the least squares formula, it is found that the dependent variables have zero correlation with the error terms.

- Age & ahe_hat => 1
  - This correlation also makes sense. When predicting a_hat using OLS, it produces a sum of the residual values which equals 0. Since (X_i, Y_i(hat)) all lie on a line, Y(hat) is defined by X.
- Age & u_hat => 0
  - This correlation also makes sense. This is because residuals will always have zero correlation with the independent variables (age in this case), as seen in the OLS derivation.

Model: lnwage = age female bachelor

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: lnwage

Number of Observations Read	7711
Number of Observations Used	7711

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	426.24090	142.08030	644.88	<.0001
Error	7707	1698.02153	0.22032		
Corrected Total	7710	2124.26243			

Root MSE	0.46938	R-Square	0.2007
Dependent Mean	2.81009	Adj R-Sq	0.2003
Coeff Var	16.70354		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.87634	0.05616	33.41	<.0001
age	1	0.02733	0.00187	14.59	<.0001
female	1	-0.18592	0.01090	-17.06	<.0001
bachelor	1	0.42813	0.01080	39.63	<.0001

- This model predicts the following results:
  - The R^2 value of 0.2007 suggests that 20% of the variation in lnwage is explained by the model. This figure is very low and suggests that we are leaving out many important determinants of wage.

- The very small p-value of <.0001 suggests that you would reject the null hypothesis. P-value represents the probability of obtaining an effect at least as extreme as the one in our sample data.
- For every year old that a person is, they will earn 2.73% higher average hourly salary, holding other variables constant
- If someone is a female, they will earn 18.59% less earnings per hour than a male, holding all other variables constant

- The model predicts that if one pursued a bachelors degree, their average hourly earnings would be 42.8% higher than if they chose not to.

Model: ahe = bachelor (only for 2008)

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: ahe

Number of Observations Read	7711
Number of Observations Used	7711

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	110502	110502	1248.79	<.0001
Error	7709	682149	88.48737		
Corrected Total	7710	792651			

Root MSE	9.40677	R-Square	0.1394
Dependent Mean	18.97609	Adj R-Sq	0.1393
Coeff Var	49.57171		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	15.33174	0.14870	103.11	<.0001
bachelor	1	7.57659	0.21440	35.34	<.0001

- You can see here that although 13.94% of the deviation in average hourly earnings is explained by bachelors degree, it is still not statistically significant with a p-value of <.0001. Although R^2 here is much more than we saw in the regression between age and age, it is still a very small amount and not statistically significant.

Comparing the effect of having a bachelors degrees in 2004 and 1992:

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: ahe

Number of Observations Read	7605
Number of Observations Used	7605

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	32175	32175	1179.37	<.0001
Error	7603	207418	27.28113		
Corrected Total	7604	239593			

Root MSE	5.22313	R-Square	0.1343
Dependent Mean	11.62637	Adj R-Sq	0.1342
Coeff Var	44.92491		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.98350	0.07665	130.24	<.0001
bachelor	1	4.21808	0.12283	34.34	<.0001

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: ahe

Number of Observations Read	7711
Number of Observations Used	7711

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	110502	110502	1248.79	<.0001
Error	7709	682149	88.48737		
Corrected Total	7710	792651			

Root MSE	9.40677	R-Square	0.1394
Dependent Mean	18.97609	Adj R-Sq	0.1393
Coeff Var	49.57171		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	15.33174	0.14870	103.11	<.0001
bachelor	1	7.57659	0.21440	35.34	<.0001

- Seen above are the regression outputs for 2004 and 1992, respectively. You can see that bachelors degree had a higher R^2 value the later the year, and it has trended downwards. There is not a large difference between 2004 and 2008 values, but this is expected due to the short time change. The larger difference between 1992 and 2004 can be explained by less people having a bachelors degree in 1992, so it makes more intuitive sense that it had a bigger impact on wages back then.