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 Sum of Mean F Value | Pr > F Source DF Squares Square Model 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 Parameter Standard Variable DF **Estimate** Error t Value Pr > |t| Intercept 1 1.08228 1.18426 0.3608 0.91 1 0.60499 0.03985 15.18 <.0001 age

- 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.

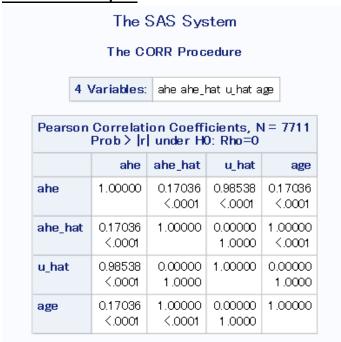
Means Procedure Output:

	The S/	AS 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

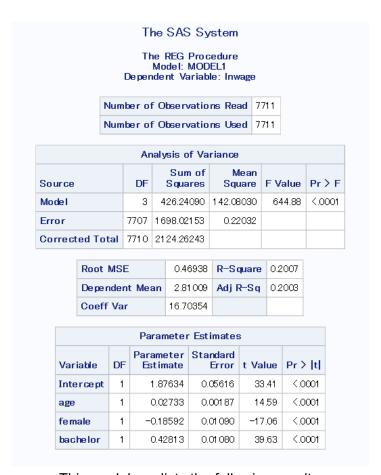
Correlation Output:



- 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.

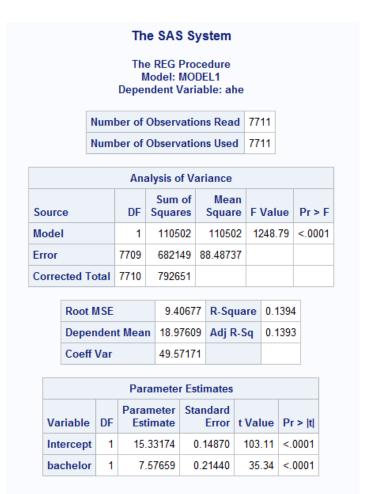
Model: Inwage = age female bachelor



- This model predicts the following results:
 - The R^2 value of 0.2007 suggests that 20% of the variation in Inwage 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)



- 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 SAS System								
	The REG Procedure Model: MODEL1 Dependent Variable: ahe							The REG Procedure Model: MODEL1 Dependent Variable: ahe							
	N	Number of Observations Read					Read 7605					Number of Observations Read			
	N	Number of Observations Used			red 7605			Number of Observations Used				d 77	11		
Analysis of Variance						Analysis of Variance									
Sourc	ce		DF	Sum of		ın re FVa	lue Pr	ource		DF	Sum of Squares	Mea Squar		Value	Pr >
Model			1	32175	321	75 1179	.37 <.00	lodel		1	110502	11050	2 12	48.79	<.000
Error			7603	207418	27.281	13		rror		7709	682149	88.4873	7		
Corre	ected To	tal	7604	239593	3			orrected To	otal	7710	792651				
	Root M	ISE		5.223	13 R-So	uare 0	1343	Root I	MSE		9.406	77 R-Sq	uare	0.139	1
	Depen	Dependent Mean		11.626	37 Adj l	R-Sq 0	1342	Depe	nden	t Mea	n 18.9760	9 Adj F	R-Sq	0.139	3
	Coeff	Var		44.924	.91			Coeff	Var		49.571	71			
		Parameter Estimates								Par	ameter E	stimates			
Va	ariable	DF		meter S	tandard Error	t Value	Pr > t	Variable	DF		meter Stimate	tandard Error	t Val	ue Pi	> t
Int	tercept	pt 1 9.98350 0		0.07665	130.24	<.0001	Intercept	1	15.	.33174	0.14870	103	11 <.	0001	
ha	chelor	1	1 4.21808 0		0.12283	34.34	<.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.