Ben Taylor

ECON 453

04/23/2025

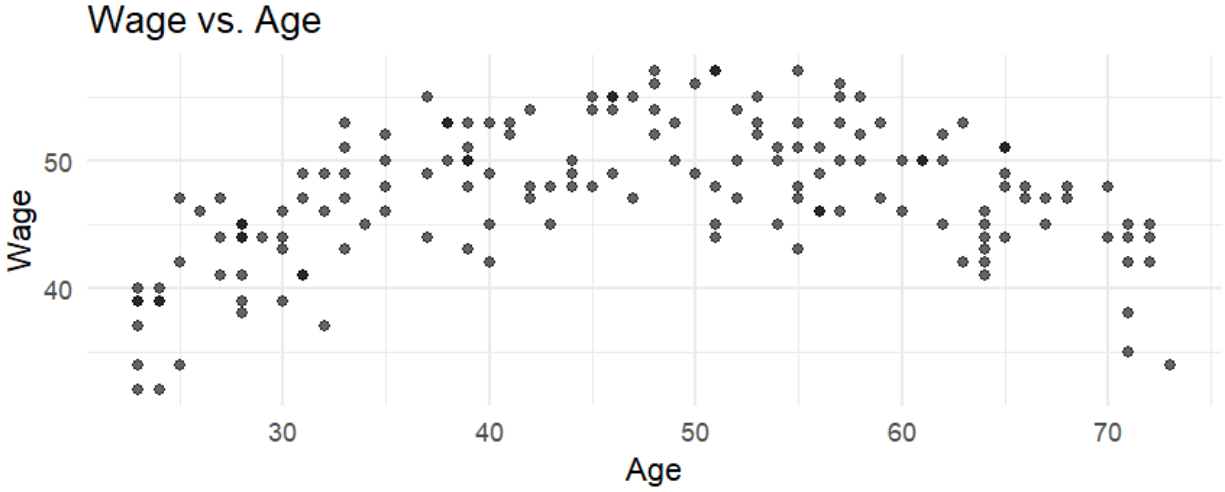
Problem Set 3/4

Question 1:

- N/A – done in R.
- Age and education are both statistically significant at a 1% level on log(wage). The output from my ANOVA test is F-statistic: 52.89 on 5 and 994 DF, p-value: < 0.00000000000000022. This means that the model is jointly significant at the 1% level.
- The log model fits the data better – it has an R^2 of 0.2101 compared to an R^2 of 0.1071 for the model using earnings.
- Comparing the 2 models, adding age^2 improves the fit. The P-value is 0.00000000004434, significantly <0.01.
- Comparing the 2 models, adding education^2 does not improve the fit of the model. The p-value is 0.4638, significantly >0.05.

Question 2:

- From the below chart, we can observe that wages typically highest between 40 and 60, and lowest under 25 and above 70.



	Model 1	Model 2
R2	0.350	0.864

R2 Adj.	0.339	0.861

- Based on the above table, model 2 is significantly better than model 1, as it has significantly higher R^2 and adjusted R^2 values.
- For a 30 year old with a graduate degree, model 1 predicts a hourly wage of $48.17364, while model 2 predicts an hourly wage of $46.82804 .
- According to model 2, predicted wages peak at age 49.01014.

```
     Dataset   Model      MAE       MSE
1      Train Model 1 3.749576 20.723374
2      Train Model 2 1.661867  4.330140
3 Validation Model 1 3.403475 17.310311
4 Validation Model 2 1.765325  4.688266
```

- As you can see in the above table, model 2 has lower MAE and MSE both in and out of sample.

Question 3:

- Income has a slight positive coefficient, which is surprising to me, as I would have expected lower income would typically lead to increased crime rates. My expectations for Poverty were correct, though, as it has an extremely positive coefficient.
- In an area with a poverty rate of 20% and median income of $50,000 , the expected crime rate according to the model would be :
    - $-301.62+53.16\times(20)+4.95\times(50) = 1{,}009.08$ crimes per 100,000 residents.

Question 4:

- The proportion of variation in sales that is explained by advertising is:
    - 199.3/240.92 = R^2 = 0.830
    - 83% is explained by advertising
- Therefore (1-r^2) is unexplained by advertising. This amount is:
    - (1-0.830) = 0.17
    - 17% is unexplained by advertising.