The Good, the Bad and the Ugly:
The Good and Bad Analysis of Quantitative Data
Loury: Gender Earnings Gap
Qualitative Data

Tufte

- Group 1: What did John Snow do in his analysis of the cholera epidemic that allowed him to support his theory (causal relation) about water and cholera?
- Group 2: What ended up obfuscating the relationship between temperature and O-ring failure in the analysis pre and post Challenger Disaster?
Classic Approach to Research

- How to write a research paper, an article and what to look for in statistical research
  - Find an interesting or pressing question
  - Review literature and develop a theoretic model
    - What is a theoretic model?
    - b. Specify the empirical model: selection of independent variables and functional form.
    - What do we mean by an empirical model? How is this different from a theoretic model?
    - c. Hypothesize the expected signs of the coefficients
    - d. Collect the data
    - e. Estimate and evaluate the equation
    - f. Document the results

Opening of the Gender Earnings Gap

- What do we learn in the opening two paragraphs?
  - i. The issue addressed by the paper
  - ii. What has gone before
  - iii. How this paper contributes to the literature (knowledge)
  - iv. Why is this important?

Review Literature and Develop a Theoretic Model

- Relevance of the Research Question
  - College educated women made up a large and growing segment of the population
  - While most studies of the decline in the gender earnings gap present aggregate results across different years of schooling, factors that are responsible for narrowing the gender gap may differ by skill level.
  - An analysis of the effects of college selectivity and academic performance makes it possible to determine whether gains have been uniform for all college educated women or whether some individuals have benefitted relative to others.
Explanations of Changes in Gender Differences in Earnings

- relative gain in the number of years worked by women compared to men.
- changed in perceptions about appropriate role of women in the labor market
- changes in industrial structure & the desired skill mix of workers within industry
- Note that each reason is developed and often supported by other information

Data and Empirical results

- Description of data used in the study
  - NLS data sets
    - longitudinal (follow individuals over time)
    - when and how collected
    - types of data collected
    - easy to find the data and obtain additional information (standard data sets)
  - Data is limited to white employees, why?
    - How does this limit what we can learn from the study?
    - What would be the natural next step?
Table information:
- name in form readily understood
- mean and standard deviation by gender
- difference in means (what is missing: see fn10)

Focused discussion of factors most relevant to this study:
- male female contrasts between 1979 and 1986:
  - raw weekly earnings
  - changing gap in number of business and education majors by gender
  - other german factors:
Display of Regression Results

- Lay-Out of Table 2
  - name, coefficient and SE of coefficient by gender and year.
  - size of sample and $r^2$

Implicit Discussion of Specification

- ground on wage equations is very well plowed.
  - education variables:
    - college graduate
    - years of college 2 – 3
    - College GPA
    - Business Major
    - Engineering Major
    - Other Major
  - College Characteristics:
    - Median College SAT Score
  - Other Characteristics:
    - Married
    - Union Member
    - Tenure on Current Job

- How complete is this specification?

The Results: 1979

- start by showing that the results conform with prior work (conventional wisdom) on factors which we are not immediately concerned with

- Cut to the chase:
  - returns to getting a degree are higher for women than for men:
    - base is one year of college:
    - estimate return for 2 – 3 years with no degree
    - estimate return to degree
    - compare return to degree to: those with one year of college
Return to College Education

- Coefficient on College grad
  - men: 0.1214
  - women: 0.3173
- compare return to degree to those with 2 - 3 years of college (calculate significance of the difference:
  - men: 0.1214 - 0.1004 = 0.02
  - women: 0.3173 - 0.1115 = 0.21
- show that this difference is statistically significant

Return to College Quality

- Women’s earnings are more sensitive to the quality of the school than men’s earnings:
  - Median SAT:
    - men: 0.0187 (0.0214)
    - women: 0.0565 (0.0282)

Repeat for 1986 estimates and compare to 1979

- Return to graduating fell for women both absolutely and relative to men
- Return to women’s GPA rose substantially
- Larger relative returns to Business major, larger returns to engineering major.
- Change in the structure of returns to college to women. In 1979 the return was to graduating from college, now it has more to do with performance in college and the major.
- Suggests that the changes in returns reflect within occupation changes in worker demand which favored women.
Conclusion

- Brief summary of results and major points, doesn’t beat a dead horse.

---

What You Need To Ask In Analyzing Statistical Research

- What is the question being asked by the researchers?
- How does this fit into prior research? Does the current research come out of prior research, or is it de novo with little base in previous work?
  - De novo work requires closer examination than work which is firmly based in prior, established research.
- What is the ‘theoretic’ model is used by the researchers and what are the implications of this model? What hypotheses are derived about the relationship of interest.

---

Questions to Ask: The Empirical Model

- What is the dependent variable? How similar is it to the dependent variable in the theoretic model and how well measured in the variable (is the definition similar to that in the theoretic model)?
- What are the variables of interest? Are they measured in a fashion which is compatible with the theoretic model?
- What controls (other variables) are included in the model? Are they derived from prior literature? Are they included because of the implications of the theoretic model?
- Is the model an extension of prior research? If so, does it follow previously models reasonably closely?
- Is the specification reasonable given the problem (are we using a linear form to estimate a non-linear model)?
Questions to Ask: The Sample

- How was the sample obtained?
  - Was it obtained from a dependable source or with a dependable methodology?
  - If the data was assembled by the researcher, was the sample random and did it have a reasonable response rate?
  - Is there evidence that the sample is appropriate and accurate?
- What limitations does the sample impose on the interpretation of the results? (E.g.: cannot generalize to all workers from a male sample).

Empirical Methods

- We will consider this in our next lecture

Questions: The Results

- Are the results presented in a fashion which provides sufficient information?
  - Coefficients, standard errors or t-statistics
  - R-square and R-square adjusted
  - Number of observations
  - Access to all coefficients, either in tables or readily available from the authors.
- Are the results with controls consistent with prior research or are those results unusual? If so, is there a reasonable explanation?
Results: Continued

- Variables of interest:
  - Do the hypotheses tests support the views of the authors? How strong is the support?
  - Can additional interesting implications be derived from the estimates?

Questions: Conclusions

- Do the conclusions reflect the results?
- Are the explanations provided in the conclusion plausible?

Working with Qualitative Data

How can we incorporate qualitative data into our models?
How do we interpret the results?
Regression Analysis: weekearn versus age, years ed

The regression equation is

\[ \text{weekearn} = -706.63 + 6.8717 \times \text{age} + 83.463 \times \text{years ed} \]

47576 cases used 7582 cases contain missing values

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-706.63</td>
<td>19.24</td>
<td>-36.73</td>
<td>0.000</td>
</tr>
<tr>
<td>age</td>
<td>6.8717</td>
<td>0.2118</td>
<td>32.45</td>
<td>0.000</td>
</tr>
<tr>
<td>years ed</td>
<td>83.463</td>
<td>1.137</td>
<td>73.38</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ S = 524.7 \quad R-Sq = 12.9\% \quad R-Sq(adj) = 12.9\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2</td>
<td>1938910706</td>
<td>969455353</td>
<td>3520.80</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>47573</td>
<td>13099281161</td>
<td>275351</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47575</td>
<td>15038191866</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

~ Dummy (Indicator) Variables ~

Thinking about this data. We believe that gender and region have an independent affect on earnings. We would like to separate the effect of gender and region from that of education.

~ Dummy (Indicator) Variables ~

Gender and region are qualitative variables. Let's take a look at them...

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>1.0000</td>
<td>2.0000</td>
<td>1.0000</td>
<td>2.0000</td>
</tr>
<tr>
<td>region</td>
<td>1.0000</td>
<td>4.0000</td>
<td>2.0000</td>
<td>3.0000</td>
</tr>
</tbody>
</table>
Dummy (Indicator) Variables

- Work with Gender first, need to create a variable to capture the effect of gender:
  - Female: recode 1 (male) to 0, recode 2 (female) to 1
  - Label this variable female
  - Add female to the regression

Regression Analysis: weekearn versus age, years ed, Female

The regression equation is

weekearn = - 402 + 6.29 age + 76.4 years ed - 319 Female

47576 cases used 7582 cases contain missing values

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-401.76</td>
<td>18.87</td>
<td>-21.29</td>
<td>0.000</td>
</tr>
<tr>
<td>age</td>
<td>6.2874</td>
<td>0.2021</td>
<td>31.11</td>
<td>0.000</td>
</tr>
<tr>
<td>years ed</td>
<td>76.432</td>
<td>1.089</td>
<td>70.16</td>
<td>0.000</td>
</tr>
<tr>
<td>Female</td>
<td>-318.522</td>
<td>4.625</td>
<td>-68.87</td>
<td>0.000</td>
</tr>
</tbody>
</table>

S = 500.4       R-Sq = 20.8%     R-Sq(adj) = 20.8%

Analysis of Variance
<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>3</td>
<td>3126586576</td>
<td>1042195525</td>
<td>4162.27</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>47572</td>
<td>11911605290</td>
<td>250391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47575</td>
<td>15038191866</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What happens to the regression when we add female to our model?
- Interpret the coefficient on female:
  - What is the effect if the respondent is female? -318.52*1
  - What is the effect if the respondent is male? -318.52*0
- What happens to the other coefficients?
- What happens to the constant?
- What happens to $r^2$?
We could just as easily defined a variable *male* as:
- Recode 1 (male) → 1
- Recode 2 (female) → 0

How would this have affected the estimates?

Regression Analysis: weekearn versus age, years ed, Male

The regression equation is

\[
\text{weekearn} = -720.28 + 6.29 \text{ age} + 76.43 \text{ years ed} + 318.52 \text{ Male}
\]

47576 cases used 7582 cases contain missing values

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-720.28</td>
<td>18.35</td>
<td>-39.25</td>
<td>0.000</td>
</tr>
<tr>
<td>age</td>
<td>6.2874</td>
<td>0.2021</td>
<td>31.11</td>
<td>0.000</td>
</tr>
<tr>
<td>years ed</td>
<td>76.432</td>
<td>1.089</td>
<td>70.16</td>
<td>0.000</td>
</tr>
<tr>
<td>Male</td>
<td>318.522</td>
<td>4.625</td>
<td>68.87</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ S = 500.4 \quad R^2 = 20.8\% \quad R^2(\text{adj}) = 20.8\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>3</td>
<td>3126586576</td>
<td>1042195525</td>
<td>4162.27</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>47572</td>
<td>11911605290</td>
<td>250391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47575</td>
<td>15038191866</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>3</td>
<td>3126586576</td>
<td>1042195525</td>
<td>4162.27</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>47572</td>
<td>11911605290</td>
<td>250391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>47575</td>
<td>15038191866</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dummy (Indicator) Variables

Regression Analysis: weekearn versus age, years ed, Female

The regression equation is

\[
\text{weekearn} = -402 + 6.29 \text{ age} + 76.4 \text{ years ed} - 319 \text{ Female}
\]

47576 cases used 7582 cases contain missing values

Predictor        Coef     SE Coef          T        P

Constant      -401.76       18.87     -21.29    0.000
age            6.2874      0.2021      31.11    0.000
years ed       76.432       1.089      70.16    0.000
Female       -318.522       4.625     -68.87    0.000

S = 500.4       R-Sq = 20.8%     R-Sq(adj) = 20.8%

Analysis of Variance

Source            DF          SS          MS         F        P

Regression         3  3126586576  1042195525   4162.27    0.000
Residual Error 47572 11911605290      250391
Total          47575 15038191866

Dummy (Indicator) Variables

- How has the coefficient on the gender variable changed? What is the implication of this?
- What has happened to the intercept between the two definitions of gender?
- What has happened to the other coefficients and to r-squared?

A small thought experiment:

- Compare the results for the two equations for a 32 year old male with 18 years of education.
  - Female Equation: weekearn = -402 + 6.29*32 + 76.4*18 - 319*0
  - Male Equation: weekearn = -720 + 6.29*32 + 76.4*18 + 319*1
- So which version you use is arbitrary.
- Implicitly, the “female” equation takes men as the base of measurement (hence the higher intercept), while the “male” equation takes women as the base of measure (hence the lower intercept).
Dummy (Indicator) Variables

- A More Complex Case: Controlling for the Effects of Region
  - Why might we want to have a measure of where people live in our earnings equation?
  - The Region Variable Takes the Form:
    - 1 = NE
    - 2 = MW
    - 3 = S
    - 4 = W
  - What happens if we enter the region variable just as it is?

Regression Analysis: earnings versus age, years ed, Female

The regression equation is

earnings = - 360 + 76.1 years ed + 6.29 age - 319 Female - 14.2 region

47576 cases used 7582 cases contain missing values

Predictor        Coef     SE Coef    T    P
Constant      -359.67       19.88     -18.09  0.000
years ed       76.073       1.090      69.78  0.000
age            6.2875      0.2020      31.12  0.000
Female       -318.796       4.623     -68.96  0.000
region        -14.245       2.125      -6.71  0.000
S = 500.2       R-Sq = 20.9%     R-Sq(adj) = 20.9%
Dummy (Indicator) Variables

- How do we interpret the region variable?
  - Does this make sense? Explain.
- Example of nonsense aspect: We can recode NE to 4 and West to 1 without losing any information. What happens now when we regress this new variable?

Regression Analysis: weekearn versus years ed, age, Female, New Region

The regression equation is:

\[
\text{weekearn} = -428.82 + 76.137 \text{ years ed} + 6.2775 \text{ age} - 318.810 \text{ Female} + 12.899 \text{ New Region}
\]

47576 cases used; 7582 cases contain missing values.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-428.82</td>
<td>19.39</td>
<td>-22.12</td>
<td>0.000</td>
</tr>
<tr>
<td>years ed</td>
<td>76.137</td>
<td>1.090</td>
<td>69.85</td>
<td>0.000</td>
</tr>
<tr>
<td>age</td>
<td>6.2775</td>
<td>0.2020</td>
<td>31.07</td>
<td>0.000</td>
</tr>
<tr>
<td>Female</td>
<td>-318.810</td>
<td>4.623</td>
<td>-68.96</td>
<td>0.000</td>
</tr>
<tr>
<td>New Regi</td>
<td>12.899</td>
<td>2.122</td>
<td>6.08</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[
S = 500.2 \quad R^2 = 20.9\% \quad R^2(\text{adj}) = 20.8\%
\]

- We cannot meaningfully enter REGION as a variable. Why?

Instead let's create 4 indicator variables, one for each region, using the code command:

<table>
<thead>
<tr>
<th>region</th>
<th>NE</th>
<th>NW</th>
<th>South</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Dummy (Indicator) Variables

Regression analysis outcomes versus years of age:

- West is highly correlated with other X variables
- West has been removed from the equation

The regression equation is:

\[
\text{weekearn} = -392.10 + 75.89 \text{ years ed} + 6.25 \text{ age} - 318.40 \text{ Female} + 65.81 \text{ NE} - 2.17 \text{ South} + \text{West}
\]

47,576 cases used 7582 cases contain missing values

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-410.25</td>
<td>19.26</td>
<td>-21.31</td>
<td>0.000</td>
</tr>
<tr>
<td>years ed</td>
<td>75.895</td>
<td>1.089</td>
<td>69.67</td>
<td>0.000</td>
</tr>
<tr>
<td>age</td>
<td>6.2532</td>
<td>0.2019</td>
<td>30.98</td>
<td>0.000</td>
</tr>
<tr>
<td>Female</td>
<td>-318.406</td>
<td>4.619</td>
<td>-68.93</td>
<td>0.000</td>
</tr>
<tr>
<td>NE</td>
<td>65.814</td>
<td>6.714</td>
<td>9.80</td>
<td>0.000</td>
</tr>
<tr>
<td>South</td>
<td>-2.168</td>
<td>6.262</td>
<td>-0.35</td>
<td>0.729</td>
</tr>
<tr>
<td>West</td>
<td>18.155</td>
<td>6.594</td>
<td>2.75</td>
<td>0.006</td>
</tr>
</tbody>
</table>

S = 499.7 R-Sq = 21.0% R-Sq(adj) = 21.0%

- Why are there only three region controls rather than 4?
- What is the implicit base group for the three region variables?
- What do we learn from this?

Can arbitrarily set a different base group:

The regression equation is:

\[
\text{weekearn} = -410 + 75.9 \text{ years ed} + 6.25 \text{ age} - 318 \text{ Female} + 65.8 \text{ NE} - 2.17 \text{ South} + 18.2 \text{ West}
\]

47,576 cases used 7582 cases contain missing values

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-410.25</td>
<td>19.26</td>
<td>-21.31</td>
<td>0.000</td>
</tr>
<tr>
<td>years ed</td>
<td>75.895</td>
<td>1.089</td>
<td>69.67</td>
<td>0.000</td>
</tr>
<tr>
<td>age</td>
<td>6.2532</td>
<td>0.2019</td>
<td>30.98</td>
<td>0.000</td>
</tr>
<tr>
<td>Female</td>
<td>-318.406</td>
<td>4.619</td>
<td>-68.93</td>
<td>0.000</td>
</tr>
<tr>
<td>NE</td>
<td>65.814</td>
<td>6.714</td>
<td>9.80</td>
<td>0.000</td>
</tr>
<tr>
<td>South</td>
<td>-2.168</td>
<td>6.262</td>
<td>-0.35</td>
<td>0.729</td>
</tr>
<tr>
<td>West</td>
<td>18.155</td>
<td>6.594</td>
<td>2.75</td>
<td>0.006</td>
</tr>
</tbody>
</table>

S = 499.7 R-Sq = 21.0% R-Sq(adj) = 21.0%
### Dummy (Indicator) Variables

- A Few Important Points About Dummy Variables:
  - Indicators are always 0/1 variables. They show whether a qualitative condition is or is not present.
    - Where you have more than two categories, you need to have more than one indicator variable. Consider the following variable for region.
  - You will always need one less indicator than you have categories.
  - You are implicitly measuring your effect against the omitted group.

### Interactions

- Doing something a little fancier with indicator variables:
  - To this point, we have assumed that the returns to education were not affected by gender. Is this a reasonable position?
  - It could be that gender affects weekly earnings both by (1) directly moving the intercept and (2) by altering the returns to education.

### Interactions

- Think about the following model:
  \[
  \text{weeklyearnings} = \beta_0 + \beta_1 \cdot \text{education} + \beta_2 \cdot \text{female} + \beta_3 \cdot (\text{female} \cdot \text{education}) + e
  \]
  - by interacting (multiplying together) the female indicator with education, we are able to "pick out" the specific effect of gender on return to education.
  - Use the calculator to multiply female by continuous education.
Interactions

Regression Analysis: weekearn versus years ed, age, Female

The regression equation is

weekearn = - 402 + 76.4 years ed + 6.29 age - 319 Female

47576 cases used 7582 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant      -401.76       18.87     -21.29    0.000
years ed       76.432       1.089      70.16    0.000
age            6.2874      0.2021      31.11    0.000
Female       -318.522       4.625     -68.87    0.000
S = 500.4       R-Sq = 20.8%     R-Sq(adj) = 20.8%

Interactions

Regression Analysis: weekearn versus years ed, age, Female, fem-ed

The regression equation is

weekearn = - 453 + 79.7 years ed + 6.26 age - 212 Female - 6.74 fem-ed

47576 cases used 7582 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant      -453.36       25.14     -18.03    0.000
years ed       79.730       1.521      52.41    0.000
age            6.2604      0.2023      30.95    0.000
Female        -212.28       34.52      -6.15    0.000
fem-ed         -6.738       2.169      -3.11    0.002
S = 500.3       R-Sq = 20.8%     R-Sq(adj) = 20.8%

What do we learn about the effect of gender on weekly earnings in this model?

What happens to general returns to education?

Do women realize a separate effect education on earnings?

What does this do to the magnitude of the coefficient on the female indicator variable?
Interactions

How to think about this:

- Base equation: \( \text{weekearn} = -453 + 79.7 \times \text{years ed} + 6.26 \times \text{age} - 212 \times \text{Female} - 6.74 \times \text{fem-ed} \)
- Male equation: (Female = 0)
  \( \text{weekearn} = -453 + 79.7 \times \text{years ed} + 6.26 \times \text{age} - 212 \times 0 - 6.74 \times 0 \times \text{education} \)
- Female Equation: (female = 1)
  \( \text{weekearn} = -453 + 79.7 \times \text{years ed} + 6.26 \times \text{age} - 212 \times 1 - 6.74 \times 1 \times \text{education} \)
  \( \text{weekearn} = -453 + 212 + 79.7 \times \text{years ed} - 6.74 \times 1 \times \text{education} + 6.26 \times \text{age} \)
  \( \text{weekearn} = -665 + (79.7 - 6.26) \times \text{years ed} + 6.26 \times \text{age} \)
  \( \text{weekearn} = -665 + 73.44 \times \text{years ed} + 6.26 \times \text{age} \)