**Econ 422 Practice 2**

**Due Date: March 29**

Use the given dataset, investigate the relationship between one’s ***physical attractiveness*** and ***economic success***.

**1. what** is the research question? What is the null-hypothesis?

**2. Data** Description.

a. what are the key variables in the dataset? Why do you think they are helpful in our research?

b. more detailed information on the key variables.

**3. Empirical Method**

**Caveat: theory guide model, but not the other way around.**

a. what is the fundamental theory?

b. what is the benchmark regression model?

c. what model specifications we will consider? Why?

**4. Conclusion**

a. what can you conclude based on your result?

b. what are the implications of your empirical research?

c. what are the limitations and future directions of your research

**5. In a one-page** slide, present your result to the audience who knows NOTHING about your research.

Reference:

**One example: Step-by-Step guide**

**(this is only an example. Always be creative!):**

**Research question**: examine the impact of look on earnings.

* Null hypothesis: individual’s physical attractiveness is independent from his or her economic success.
* Alternative hypothesis:

**Data description.**

data source and collection method:

19xx quality of American Life survey (QAL) contains interviews of 1260 respondents. The dataset contains usual labor-market and demographic variables of interest to economists. Most importantly, it contains information on the respondents’ look. In this survey, the interviewer, who visited the respondent in his or her abode, had to rate the respondent’s physical appearance on a five-point scale, along which looks rang from homely to strikingly handsome or beautiful. Because we are interested in investigating the relationship between one’s look and income. It is essential to have a dataset that have information on both productivity-enhancing characteristics (such as education level, working experiences, etc.) and one’s physical attractiveness level. However, in the dataset, we do not have detailed information on the respondent’s occupation. The only information we know is that if the respondent is in service industry.

* More detail on the key variable: ***looks***

(distribution, gender difference, use both graph and summarize table)

Do you think the distribution makes sense?

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Men | Women | Pooled |
| 1. homely |  |  |  |
| 2. below average |  |  |  |
| 3. average |  |  |  |
| 4. above average |  |  |  |
| 5. strikingly beautiful or handsome |  |  |  |
| # of Observation |  |  |  |

**Empirical Method**

* **Explain why do you think one’s look might correlated with income**. In your own research, you will spend considerable amount of time in this part. Read past studies, write literature reviews, find existing evidence, etc. Here, I will skip this part and jump to the theoretical model

Where denotes a set of productivity-enhancing characteristics.

* Empirically, let’s start with a quick look about the relationship between income and look without any controls. (report correlation coefficient and statistical test result beside the graph).
* Benchmark regression

Consider:

1. what do you wanted to include in the ?

2. Is your model adequate?

3. Do you have any concerns about heteroskedasticity problem in ***looks***?

4. Do you have any other concerns about the model?

* Other model specification

**Example**: Service industry might be more sensitive to looks. Therefore, I am interested in if there is any interaction effect between one’s look and occupation (service industry).

Regression (1):

Regression (2):

Regression (3):

Report regression results

|  |  |  |  |
| --- | --- | --- | --- |
| ***Variables*** | (1) | (2) | (3) |
|  | log(wage) | log(wage) | log(wage) |
| ***looks (From 1 to 5)*** | coefficient (s.e) | coefficient (s.e) | coefficient (s.e) |
| ***Beauty Sensitive Occupations*** | coefficient (s.e) | coefficient (s.e) | coefficient (s.e) |
| ***Looks x  Beauty Sensitive Occupations*** | coefficient (s.e) | coefficient (s.e) | coefficient (s.e) |
| ***var 1*** | coefficient (s.e) | coefficient (s.e) | coefficient (s.e) |
| ***var 2*** |  |  |  |
| ***…*** |  |  |  |
| ***Constant*** |  |  |  |
| ***Adjust R-square*** |  |  |  |
| # of observations |  |  |  |

\*significant at 10% level;

\*\* significant at 5% level;

\*\*\*significant at 1% level

* Oftentimes, you don’t have to report ALL of the independent variables in the table. Most of the regressors are control variables and do not directly answer the research question. A better way to report your result is to only report important key terms in the regression. And explain what other variables we are controlling.
* Since very few people were categorized as strikingly handsome or homely, the value on looks were highly concentrated in the middle (i.e., average). We have very few observations in extreme cases (strikingly handsome/beautiful or homely). Therefore, we create another two dummy variables: if one’s look score equal or above 4, then we say this person is “above average”; if one’s look score equal or smaller than 2, then we say this person is “below average”. Modify your regression with above average and below average, and then report your regression results.
* What other mechanism/regression/analysis might be interesting? For example, do you think it is possible that good looking people often self-select themselves into occupations that have a higher return on their pretty faces? How to examine this idea?