Regression Workshop | Census Data

Current Population Survey

March 4, 2025

Two of the variables in the dataset are *agi* which measures estimated adjusted gross income, and *age* for the Current Population Survey.
To obtain the data for this workshop, download the *cps_panel.dta* from data.ku.edu.

1 Univariate Regression

(a) Use regression to estimate the relationship between household income and age.

(b) What are the null and alternative hypotheses for testing the slope?
(c) What is the slope of the regression line?

(d) What is the standard error of the slope?

(e) What is the formula for and the value of the test statistic for testing the null and alternative hypotheses?

(f) What are the degrees of freedom?

(g) What is the p-value for testing the slope coefficient?

(h) Give the conclusion of the test in context of testing at the 5% significance level.

(i) What is the equation of the regression line?

(j) Interpret the slope of the line in context.

(k) Interpret R^2 in context.

(l) What are the F-statistic and p-value of the ANOVA test?

(m) How does this p-value compare to the two found in parts (d) and (g)?

(n) What is the standard deviation of the error term? Either compute it from the information in the ANOVA table or find it in the output.

(o) State the formula for and find the 95% confidence interval for mean price.

2 Multivariate Regression

Use regression to estimate the relationship between agi, age, $age2$, $forborn$, and $wkslyr$.

(a) What is the regression equation?

(b) Give the coefficient estimates for each explanatory variable in the model.

(c) Interpret the coefficient of age and age squared.

(d) Is age squared a useful variable for predicting agi in this model?

i. State the null and alternative hypotheses.
ii. State the value of the test statistic.
iii. State the p-value.
iv. Sate the generic conclusion.
v. State the conclusion in context.

(e) For the ANOVA test for the regression:

i. State the null and alternative hypotheses.
ii. State the value of the test statistic.
iii. State the p-value.
iv. Sate the generic conclusion.
v. State the conclusion in context.

(f) Interpret the value of R2 in context.

(g) Which predictor variable is most significant in the model?

(h) Which predictor variable is least significant in the model?