**a)**

The regression model is as follows:

*bpsystol =*β0 + β1*female* + β2*race +*  β3*age +*  β4*rural +* β5*bmi +* β6*diabetes* + ε

**b)**

Ordinary least squares multiple linear regression. The dependent variable is continuous and there are more than one independent variables.

**c)**

Table 1

*Results for Multiple Linear Regression Analysis of Systolic Blood Pressure and Socioeconomic Factors*

|  |  |  |
| --- | --- | --- |
|  | Unstandardized regression coefficients | 95% CI |
| Intercept | 72.35 (1.10) \* | (70.19, 74.50) |
| *Female* (ref.a male) | -4.11 (.38) \* | (-4.86, -3.37) |
| *race (ref. White)* |  |  |
| *Black* | 2.6(.63) \* | (1.22, 3.71) |
| *Other* | 1.94 (1.38) | (-.77, 4.66) |
| *age* | .58 (.01) \* | (.56, .60) |
| *rural (ref. urban)* | -.24 (.40) | (-1.03, .55) |
| *bmi* | 1.27 (.04) \* | (1.20, 1.35) |
| *diabetes (ref.no diabetes)* | 5.26 (.90) \* | (3.49, 7.03) |
| *R3* | 0.32 |  |
| No. of observations | 10349 |  |

a*ref*. = reference category

\**p < .05* (two-tailed)

Note: Standard errors are in parentheses

The results reveal that holding other factors in the model constant;

* Females report significantly lower systolic blood pressure than men. In particular, females report 4.11 mmHg lower systolic blood pressure than males.
* Blacks report significantly higher systolic blood pressure than White people. In particular, Black people report 2.46 mmHg higher systolic blood pressure than the Whites.
* A yearly increase in age results in a significant increase in systolic blood pressure by a factor of .58.
* A unit increase in BMI results in a significant increase in systolic blood pressure by a factor of 1.27.
* Individuals with diabetes report significantly higher systolic blood pressure than those without diabetes. In particular, individuals with diabetes report 5.26 mmHg higher systolic blood pressure than those without diabetes.

**Qualitative (binary) Data Analysis**

**a)**

The regression model is as follows:

ln [p/(1-p)] *=*β0 + β1*female2* + β2*race +*  β3*age +*  β4*rural +* β5*bmi +* β6*bpsystol* + ε

where p is the odds that an individual will have diabetes.

**b)**

Binary logistic regression. The outcome variable (*diabetes)* is categorical in nature and has only two levels.

**c)**

Table 2

*Results for Binary Logit Analysis of Diabetes*

|  |  |  |
| --- | --- | --- |
|  | *B* | Exp (*B)* |
| Intercept | -8.72 (.35) \* | -.00 |
| *Female2* (ref. a male) | .07 (.10) | 1.07 |
| *race (ref. White)* |  |  |
| *Black* | .60 (.13) \* | 1.83 |
| *Other* | .28 (.36) | 1.32 |
| *age* | .05 (.00) \* | 1.05 |
| *rural (ref. urban)* | .05(.10) | 1.05 |
| *bmi* | .06 (.01) \* | 1.06 |
| *bpsystol* | .01 (.00) \* | 1.01 |
| Pseudo *R3* | 0.11 |  |
| Likelihood ratio test | 458.03\* |  |
| No. of observations | 10349 |  |

a*ref*. = reference category

\**p < .05* (two-tailed)

Note: Standard errors are in parentheses

The results indicate that holding other factors in the model constant;

* The odds for presence of diabetes are 1.83 times higher for Blacks than White people.
* For every yearly increase in age, the odds for being diabetic increase by 1.05.
* For every unit increase in BMI, the odds for being diabetic increase by 1.06.
* For every unit increase in systolic blood pressure, the odds for being diabetic increase by 1.01.

**Appendix: STATA CODES**

use "/Users/kairu/Downloads/nhanes2 (4).dta"

##Question 1: Linear Regression

reg bpsystol i.female i.race age i.rural bmi i.diabetes

## Question 2: Qualitative Analysis

gen female2=female \*female

logit diabetes i.female2 i.race age i.rural bmi bpsystol