# Final Exam – Take Home

# Bone Mass Density Screening

The data for your project comes from a study conducted in Youngstown, OH from 1997-2000. Health Professions professors and students screened anyone interested in obtaining an estimate of their bone mass density at health and county fairs. Subjects placed their ankle in a portable scanning machine to determine if they were at risk for osteoporosis. Over the course of the study, a total of 1792 subjects were screened. Your data set contains only a portion (250 observations) of the total data set. The file **osteo1a.xls** is available at http://csuohio.edu/holcombj/mth147/finalexam.htm Note that some variables may not be used for this assignment, but may be used for the take-home Final Examination.

Variable guide:

| | |
|---|---|
| **subjectno** | An identifying number to keep track of subjects |
| **sex** | 1=Female, 2=Male |
| **age** | Age in years |
| **weight** | Pounds |
| **height** | Inches |
| **bmd** | Bone Mass Density |
| **tscore** | Calculation of Risk for Osteoporosis |
| **fracture** | 0=No, 1=Yes |
| **osteo** | 0=No, 1=Yes at risk for Osteoporosis |
| **calcium** | 0=No, 1=Yes |
| **treat** | 0=No, 1=Yes |
| **count** | Column of 1's |

The variable **bmd** indicates whether a subject is at risk for osteoporosis. The lower the bmd, the greater the chance of having osteoporosis. The variable **osteo** indicates whether a subject was identified as being at risk for osteoporosis. A subject with osteoporosis is at high risk for fractures, especially fractures of the wrist, hip, and spine. A fall that leads to a fracture can be devastating for an elderly subject. The variable above treat indicates whether a subject is currently taking a medication such as estrogen, Fosxamax, Miacalcin, or Didronel which have all shown in clinical trials to increase bone mass density, or at least slow its deterioration. After screening, letters were sent to a subject's primary care physician if their T-score indicated they might have osteoporosis.

For this study, the population of interest is not the general public, but ambulatory subjects who are willing to participate in health fairs. Treat the dataset as a random sample of subjects from this population.

For the following tests of Hypothesis, be sure to state the hypotheses, the test statistic, the P-value or the P-value estimate, and your conclusion. Assume that the subjects are a random sample of first year students.

1. Is there sufficient evidence to conclude that over 20% of this population has had a fracture occur?
2. Is there sufficient evidence to conclude that the mean age of this population is over 50?
3. Is there sufficient evidence to conclude that being diagnosed with osteoporosis and reporting a fracture are independent?

Use the variable **age** to predict **bmd**:

1. Show the scatterplot.
2. Give the model for the regression line.
3. Give the $r^2$ value, interpret its strength, and interpret its meaning.
4. Use the model to predict the **bmd** when the **age** is 65. Are there any influential observations? Why or why not?
5. Include the residual plot. Are there any outliers? If so, how many?

Write a paragraph that describes your conclusions. Also, perform some kind of a test of hypothesis that I have not proposed. This could be a test involving a mean, a proportion, independence, or it could involve a regression analysis. Clearly state your null, alternative, test statistic, P-value, and conclusion.

Answers

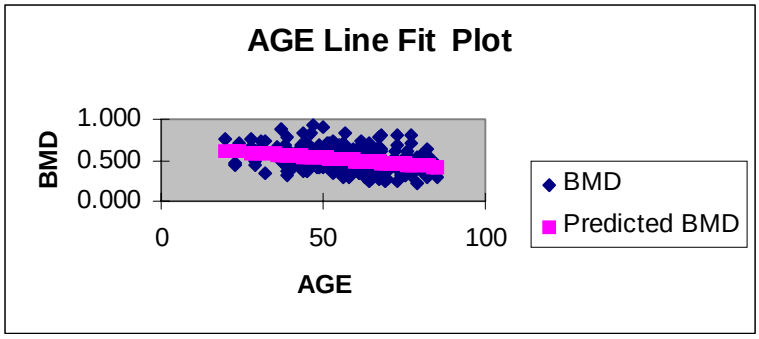
| *AGE* | |
|---|---|
| Mean | 57.336 |
| Standard Error | 0.869822 |
| Median | 58 |
| Mode | 64 |
| Standard Deviation | 13.7531 |
| Sample Variance | 189.1477 |
| Kurtosis | -0.37845 |
| Skewness | -0.24774 |
| Range | 65 |
| Minimum | 20 |
| Maximum | 85 |
| Sum | 14334 |
| Count | 250 |

| | |
|---|---|
| z | 8.433907 |

| Sum of COUNT | FRACTURE | | |
|---|---|---|---|
| OSTEO | 0 | 1 | Grand Total |
| 0 | 93 | 40 | 133 |
| 1 | 76 | 41 | 117 |
| Grand Total | 169 | 81 | 250 |

| | |
|---|---|
| phat | 0.324 |
| z | 4.901530373 |

| | |
|---|---|
| Chi-Square | 0.701 |
| df | 1 |
| p-value | 0.402 |




| *Regression Statistics* | |
|---|---|
| Multiple R | 0.307973 |
| R Square | 0.094847 |
| Adjusted R Square | 0.091198 |
| Standard Error | 0.129096 |
| Observations | 250 |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 0.68108 | 0.03507 | 19.42032 | 1.05E-51 |
| AGE | -0.00303 | 0.000595 | -5.09774 | 6.82E-07 |