# Regression Summary

**Numeracy**

Method

After the differential analysis, we have used univariate and multivariable generalized linear model to evaluate the association between self-reported health status and numeracy scores in ESP study. Specifically, multivariable generalized linear models are created to evaluate the independent effect of numeracy scores on the health status, after controlling demographical variables of interest (education, age categories, income categories, race, ethnicity). Weighs are added by education and income using manually iterative method. Similar analysis procedure has been performed to investigate the association between self-reported health status and numeracy scores in PIACC study. Specifically, multivariable generalized linear models are created to evaluate the independent effect of numeracy scores on the health status, after controlling same demographical variables of interest as ESP (education, age categories, income categories, race, ethnicity). Two additional predictors, problem solving and literacy, are also considered in further regression.

Results

The original health status has 5 levels after dropping all unknowns: Excellent/Very good/Good/Fair/Poor. Take clinical meaning into consideration, there are to dichotomizing method to have binary health status outcome: Good (611)/Not Good (152); Poor (737)/Not Poor(26). Since the situation is quite unbalanced and the regression result shows that numeracy is not a significant predictor, we choose the first dichotomizing method. The health status outcome has two levels: Good and Not Good.

**ESP**

Univariate regression results show that 6 variables are significant covariates of people’s health status: numeracy, age.cat, educ, income.cat, race, hisp. For multivariate regressions, stepwise selection method was conducted to choose the final model. Numeracy has significant effect on the self-reported health status (good / not good) with a power of 81%, after controlling education, age categories and income categories. This model has an AUC of 72%, and prediction error of 0.219 after randomly divided the dataset into training and test data by a proportion of 7:3.

Weights were applying on education and income since their distributions are significantly different from that of NYS data, and they are both significant covariates of people’s health status. The weighted regression results are quite similar to unweighted results: Numeracy has significant effect on the self-reported health status (good / not good) with a power of 57.4%, after controlling education, age categories, income categories and ethnicity. This model has an AUC of 72%, and prediction error of 0.219 after randomly divided the dataset into training and test data by a proportion of 7:3.

The regression summary table for ESP dataset is shown in the bellow table.

| ESP’s multivariate regression summary | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Response** | **Weights** | **Performance** | | | **Predictor** | **Covariates** | | | | | |
| **Outcome** | **Adjusted covariates** | **AUC** | **Power** | **Prediction Error Rate** | **numeracy** | **gender** | **age** | **educ** | **income** | **race** | **hisp** |
| health status: Good vs Not Good |  | 0.724 | 0.81 | 0.219 | \* |  | \* | \* | \* |  |  |
| health status: Good vs Not Good | education, income | 0.72 | 0.574 | 0.219 | \* |  | \* | \* | \* |  | \* |

**PIAAC**

Stepwise selection method was conducted to choose the multivariate regression model. Numeracy has significant effect on the self-reported health status (good / not good) with a power of 75%, after controlling education and income categories. The final model has an AUC of 67.6%, and prediction error rate of 14.8% after randomly divided the dataset into training and test data by a proportion of 7:3.

Additional stepwise regressions were conducted adding literacy and problem solving as predictors. Regression result shows that numeracy and problem solving has significant effect on the self-reported health status (good / not good) with a power of 77%, after controlling education and income. This model has an AUC of 69%, and prediction error of 14.8% after randomly divided the dataset into training and test data by a proportion of 7:3.

The regression summary table for PIAAC dataset is shown in the bellow table.

| PIAAC’s multivariate regression summary | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Response** | **Performance** | | | **Predictor** | **Covariates** | | | | **Additional Predictor** | |
| **Outcome** | **AUC** | **Power** | **Prediction Error Rate** | **numeracy** | **educ** | **income** | **race** | **hisp** | **problem solving** | **literacy** |
| health status: Good vs Not Good | 0.676 | 0.75 | 0.1475 | \* | \* | \* |  |  | / | / |
| health status: Good vs Not Good | 0.69 | 0.77 | 0.1475 | \* | \* | \* |  |  | \* |  |

**Visits**

Method

Univariate and multivariable generalized linear model to evaluate the association between self-reported health status and numeracy scores in ESP study. Specifically, multivariable generalized linear models are created to evaluate the independent effect of numeracy scores on doctor visiting frequency status, after controlling demographical variables of interest. (education, age categories, income categories, race, ethnicity). Weighs are added by education and income using manually iterative method.

Results

Univariate regression results show that 6 variables are significant covariates of people’s visits to doctors: gender, age.cat, educ, income.cat, race, hisp. Numeracy is not a significant predictor for visits to doctors. For multivariate regressions, stepwise selection method was conducted to choose the final model. Numeracy does not have significant effect for visits to doctors (often / not often), although gender, age, income and ethnicity are significant covariates. The final model has an AUC of 64.3%, and prediction error of 34.97% after randomly divided the dataset into training and test data by a proportion of 7:3.

Additional multinomial regression was conducted after adjusting the distribution of education and income with weights. Numeracy seems not significantly associate with visits to doctors (often/not often), while gender, age, education and income are significant covariates. The final model has an AUC of 64.6%, and prediction error of 37.25% after randomly divided the dataset into training and test data by a proportion of 7:3.

The regression summary table for ESP dataset is shown in the bellow table.

| **Response** | **Weights** | **Performance** | | | **Predictor** | **Covariates** | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Outcome** | **Adjusted covariates** | **AUC** | **Power** | **Prediction Error Rate** | **numeracy** | **gender** | **age** | **educ** | **income** | **race** | **hisp** |
| visits to doctors: Often vs Not Often |  | 0.643 | / | 0.3497 |  | \* | \* |  | \* |  | \* |
| visits to doctors: Often vs Not Often | education, income | 0.636 | / | 0.3725 |  | \* | \* | \* | \* |  |  |