Contents

Scenario 3

```
% Raymond/Lei Chi and Arav Sharma
```

Setup

```
clc
clear
close all

load('SATs.mat');
```

Calculations

```
meanMath = mean(SAT_Math, "omitnan");
meanVerbal = mean(SAT_Verbal, "omitnan");
varMath = var(SAT_Math, "omitnan");
varVerbal = var(SAT_Verbal, "omitnan");
covVM = cov(SAT_Math,SAT_Verbal, "omitrows");
% corrcoef = covVM(1,2) / (varMath * varVerbal);
```

For the entire range

page 151

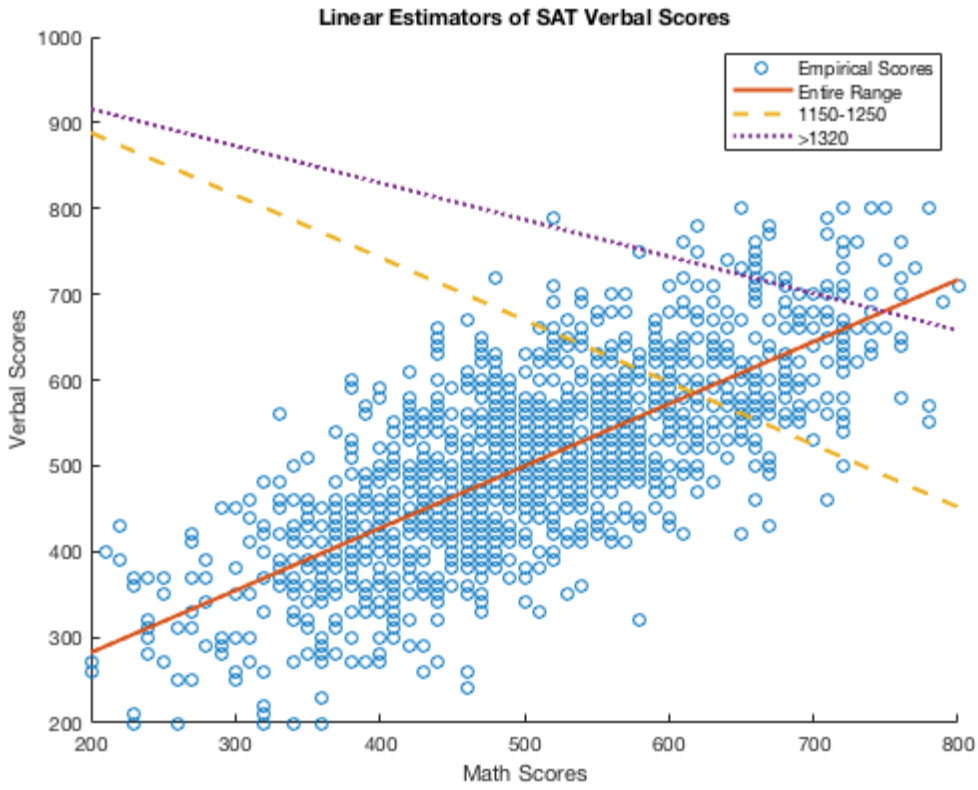
```
a_entire = covVM(1,2) / varMath;
b_entire = meanVerbal - a_entire * meanMath;
```

For scores between 1150 and 1250

```
subset1 = (SAT_Verbal + SAT_Math) >= 1150 & (SAT_Verbal + SAT_Math) <= 1250;

covMatrix_subset1 = cov(SAT_Math(subset1), SAT_Verbal(subset1));
varMath_subset1 = var(SAT_Math(subset1));

a_subset1 = covMatrix_subset1(1,2) / varMath_subset1;
b_subset1 = mean(SAT_Verbal(subset1)) - a_subset1 * mean(SAT_Math(subset1));
```

For scores above 1320

```
subset2 = (SAT_Verbal + SAT_Math) > 1320;

covMatrix_subset2 = cov(SAT_Math(subset2), SAT_Verbal(subset2));
varMath_subset2 = var(SAT_Math(subset2));

a_subset2 = covMatrix_subset2(1,2) / varMath_subset2;
b_subset2 = mean(SAT_Verbal(subset2)) - a_subset2 * mean(SAT_Math(subset2));


scatter(SAT_Math, SAT_Verbal);
hold on;

x = linspace(min(SAT_Math), max(SAT_Math), 100);
y_entire = b_entire + a_entire * x;
plot(x, y_entire, 'LineWidth', 2);

y_subset1 = b_subset1 + a_subset1 * x;
plot(x, y_subset1, '--', 'LineWidth', 2);

y_subset2 = b_subset2 + a_subset2 * x;
plot(x, y_subset2, ':', 'LineWidth', 2);

legend('Empirical Scores', 'Entire Range', '1150-1250', '>1320');
xlabel('Math Scores');
ylabel('Verbal Scores');
title('Linear Estimators of SAT Verbal Scores');
hold off;


% The estimators behave differently when computed over different ranges
% because the people who score in different ranges have different
% score profiles. Those who score highly (but not close to perfect)
% will tend to score higher in one category and around to average in
% another category, so it actually creates a negative correlation
% This is why only the absolute top scorers and average to low scorers
% will demonstrate a positive correlation between verbal and written scores
% This group also represents the majority (slightly above average is the
% the minority), so overall it is a positive correlation
```



Published with MATLAB® R2020a