TAKE-HOME CLASS QUIZ SOLUTIONS: DUE MONDAY NOVEMBER 11: USING LINEAR SYSTEMS FOR MEASUREMENT

MATH 196, SECTION 57 (VIPUL NAIK)

1. Performance review

27 people took this 4-question quiz. The score distribution was as follows:

- Score of 0: 2 people
- Score of 2: 9 people
- Score of 3: 9 people
- Score of 4: 7 people

The mean score was about 2.7.

The question-wise answers and performance review are as follows:

(1) Option (E): 21 people
(2) Option (D): 18 people
(3) Option (B): 17 people
(4) Option (D): 17 people

2. Solutions

PLEASE FEEL FREE TO DISCUSS *ALL* QUESTIONS.

The purpose of this quiz is two-fold. First, many of the ideas you saw early on in the course (in the second and third week) may be on the verge of fading out. Drawing from the best research on *spaced repetition* (see for instance `http://en.wikipedia.org/wiki/Spaced_repetition`) it's high time we tried recalling some of that stuff. But with a twist, because we can now use some concepts from later topics we've seen to refine our past understanding.

The second purpose is to prepare you for what we hope to eventually get to: a deep and rich understanding of linear algebra as it's *used*: in linear regressions, computing correlations, and more fancy applications like factor analysis and principal component analysis. The third question, in particular, relates to the central idea behind linear regression (specifically, ordinary least squares regression). The questions also relate, albeit not very directly, to the broad ideas behind factor analysis and principal component analysis.

For the questions here, assume two dimensions of a person's general cognitive ability: verbal and mathematical. Denote by g_v the person's general verbal ability and by g_m the person's general mathematical ability.

Various ability tests can be devised that aim to test for the person's abilities. However, it is not possible to construct a test that *solely* measures g_v or *solely* measure g_m. Different tests measure g_v and g_m to different extents. For instance, an ordinary numerical computation test might measure mostly g_m. On the other hand, a test similar to the quizzes in this course might measure both g_v and g_m a fair amount, given how much you have to read to answer the quiz questions.

Of course, the score on a given test could depend on a lot of factors other than general abilities. Some of them could be systematic: a student with poor mathematical abilities in general may have "trained for the test." As an example, using a calculus test to test for general mathematical ability might mean that people who have happened to take calculus do a lot better than people who haven't, but have similar general mathematical ability. Some are more ephemeral, such as students guessing answers, mood fluctuations, and other context-specific factors that affect scores.

For simplicity, we will assume that there are no systemic factors other than general verbal and general mathematical ability that the test is measuring. For even greater simplicity, assume that the (expected) test score is linear in g_v and g_m with zero intercept, i.e, the score is of the form $w_v g_v + w_m g_m$ where w_v and w_m are real numbers that serve as *weights*. Note that this assumes that there is no *interaction* between the verbal and mathematical skills in determining the scores.

The assumption may or may not be realistic. For instance, a question (such as those on this quiz!) that requires a lot of reading *and* strong math skills would probably have an expected score formula that is *multiplicative* in g_v and g_m: having zero or near-zero verbal ability means you will be unable to do the question, even if your mathematical ability is awesome. Similarly, having zero or near-zero mathematical ability means you will be unable to do the question, even if your verbal ablity is awesome. Multiplicatively separable functions are better suited to capture this sort of dependence. However, even if the test has questions of this sort, we can take logs on test scores and make them additively separable, so the additive model may still work well.

The assumption of *linearity* goes further, but this too might be realistic.

My goal is to use one or more tests in order to determine the true values of a student's g_v and g_m. Another formulation is that my goal is to determine the vector:

$$\vec{g} = \begin{bmatrix} g_v \\ g_m \end{bmatrix}$$

(1) I administer two tests to a student. The student's score s_1 on the first test is $2g_v + 3g_m$ while the score s_2 on the second test is $3g_v + 5g_m$. How do I recover g_v and g_m from s_1 and s_2?
 (A) $g_v = 2s_1 + 3s_2$, $g_m = 3s_1 + 5s_2$
 (B) $g_v = 2s_1 - 3s_2$, $g_m = 3s_1 - 5s_2$
 (C) $g_v = 5s_1 + 3s_2$, $g_m = 3s_1 + 2s_2$
 (D) $g_v = 5s_1 - 3s_2$, $g_m = 3s_1 - 2s_2$
 (E) $g_v = 5s_1 - 3s_2$, $g_m = -3s_1 + 2s_2$

Answer: Option (E)

Explanation: We have the following:

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} g_v \\ g_m \end{bmatrix}$$

We thus have:

$$\begin{bmatrix} g_v \\ g_m \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

Recall that, for a general 2×2 matrix, the inverse is given by:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

In this case, the determinant $ad - bc$ equals $(2)(5) - (3)(3) = 1$, so that the inverse is:

$$\begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$$

Plugging in, we get:

$$\begin{bmatrix} g_v \\ g_m \end{bmatrix} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

This simplifies to:

$$\begin{aligned} g_v &= 5s_1 - 3s_2 \\ g_m &= -3s_1 + 2s_2 \end{aligned}$$

Performance review: 21 out of 27 got this. 3 chose (A), 2 chose (D), 1 chose (C).

Historical note (last time): 23 out of 25 people got tihs. 1 each chose (B) and (C).

(2) In order to combat the problem of uncertainty about my model, I decide to administer more than two tests. I administer a total of n tests. The score on the i^{th} test is $s_i = w_{i,v}g_v + w_{i,m}g_m$. The score vector $\vec{s}$ has coordinates $s_i, 1 \le i \le n$.

If there is no measurement error and the student's actual score in each test equals the student's expected score, then we have a system of n simultaneous linear equations in 2 variables.

Let W be the matrix:

$$\begin{bmatrix} w_{1,v} & w_{1,m} \\ w_{2,v} & w_{2,m} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ w_{n,v} & w_{n,m} \end{bmatrix}$$

Assume that all entries of W are positive, i.e., each test tests to a nonzero extent for both verbal and mathematical ability.

Our goal is to "solve for" $\vec{g}$ the following vector equation:

$$W\vec{g} = \vec{s}$$

What is the necessary and sufficient condition on W so that the equation has at most one solution for $\vec{g}$ for each $\vec{s}$? If $\vec{s}$ arises from an actual $\vec{g}$, i.e., it is a true score vector, then note that there will be a solution.

(A) All the ratios $w_{i,v} : w_{i,m}$ are the same.
(B) All the ratios $w_{i,v} : w_{i,m}$ are different.
(C) At least two of the ratios $w_{i,v} : w_{i,m}$ are the same.
(D) At least two of the ratios $w_{i,v} : w_{i,m}$ are different.

Answer: Option (D)

Explanation: Recall that for a unique solution, the coefficient matrix needs to have full column rank, which in this case means rank 2. If all the ratios $w_{i,v} : w_{i,m}$ are the same, then the rank is 1. On the other hand, if two of the ratios differ, then those two rows alone give rank 2, and the remaining rows cannot affect the rank any more.

Performance review: 18 out of 27 got this. 6 chose (B), 2 chose (A), 1 chose (C).

Historical note (last time): 14 out of 25 people got this. 4 each chose (A) and (C). 3 chose (B).

(3) Use notation as in the previous question. Suppose that there is some measurement error, so that instead of getting the true score vector $\vec{s}$, I have a somewhat distorted score vector $\vec{t}$. How do I go about recovering my "best guess" for $\vec{s}$ from $\vec{t}$?

(A) Find the closest vector to $\vec{t}$ in the kernel of the linear transformation corresponding to W.
(B) Find the closest vector to $\vec{t}$ in the image of the linear transformation corresponding to W.
(C) Find the farthest vector from $\vec{t}$ in the kernel of the linear transformation corresponding to W.
(D) Find the farthest vector from $\vec{t}$ in the image of the linear transformation corresponding to W.

Answer: Option (B)

Explanation: Note that when talking of the kernel and image of W, we want to solve $W\vec{g} = \vec{s}$. Instead of $\vec{s}$, we have a distorted vector $\vec{t}$. Our best guess for $\vec{s}$ is the vector closest to $\vec{t}$ for which the equation can be solved, i.e., the vector closest to $\vec{t}$ in the image of the linear transformation corresponding to W.

Performance review: 17 out of 27 got this. 5 chose (A), 4 chose (C), 1 chose (D).

Historical note (last time): 14 out of 25 people got this. 8 chose (C), 3 chose (A).

(4) Suppose I want to introduce a *new* test that tests for both verbal and mathematical ability with expected score of the form $w_v g_v + w_m g_m$, but the values w_v and w_m are currently unknown. My strategy is as follows. I find two students. I administer a bunch of tests with *known* w_v and w_m values to those students. I use those tests to find the g_v and g_m values for both students. Then, I

administer the new test to both students and try to determine the values of w_v and w_m. Assume no measurement error.

Of course, I want the matrix W of the *known* tests to satisfy the condition of Question 2. What additional criteria would I wish of the two students I use for this in order to correctly determine g_v and g_m? Note that it will not be possible to be sure of this in advance, but one can still pick student pairs who are more likely to satisfy the criterion and thus avoid waste of effort.

(A) The students should have the same $g_v + g_m$ value.
(B) The students should have different $g_v + g_m$ values.
(C) The students should have the same $g_v : g_m$ ratio.
(D) The students should have different $g_v : g_m$ ratios.

Answer: Option (D)

Explanation: The reasoning is similar to that used for Question 2.

Let $g_{v,1}$ and $g_{m,1}$ be the g_V and g_m values for the first student. Let $g_{v,2}$ and $g_{m,2}$ be the g_v and g_m values for the second student. We want to find the weights w_v and w_m for the test. Explicitly, this means solving the following equation for unknowns w_v and w_m:

$$\begin{bmatrix} g_{v,1} & g_{m,1} \\ g_{v,2} & g_{m,2} \end{bmatrix} \begin{bmatrix} w_v \\ w_m \end{bmatrix} = \begin{bmatrix} \text{score of first student} \\ \text{score of second student} \end{bmatrix}$$

In order to have a unique solution, we need the coefficient matrix to have full column rank 2, i.e., we need the students' $g_v : g_m$ ratios to differ.

Performance review: 17 out of 27 got this. 9 chose (C), 1 chose (B).

Historical note (last time): 14 out of 25 got this. 9 chose (C), 1 each chose (A) and (B).