HOMEWORK 1
MATH 463 - SPRING 2017

ALEX THIES

Comment: I collaborated with Sienna Allen, Joel Bazzle, Torrin Brown, Ashley Ordway, and Seth Temple on this assignment.

Here for a vector $\mathbf{y}$ and a subspace V, we denote by $\pi(\mathbf{y}|V)$ the projection of y on V. Also Π_V denotes the projection operator onto V.

Problem A The notation $N(\mu, \sigma^2)$ means the Normal distribution with mean μ and variance σ^2. Generate values $z_1, \ldots, z_{100}, e_1, \ldots, e_{100}, d_1, \ldots, d_{100}$ all from a $N(0,1)$ distribution. (Use the R function rnorm. Prior to generating these values, use set.seed with an argument you select, so that your data can be reproduced. Each student must use a different seed value.) Create values, for $i = 1, \ldots, 100$.

(1) $\mathbf{y}_i = 10z_i + e_i, \quad x_i = 8z_i + d_i$

(a) Create a scatterplot of y_i against x_i and superimpose the least-squares line.

(b) Suppose that x_{50} is changed to the value 5. Is the right-hand equation in (1) still true for $i = 50$?

The random vector (X, Y) has a multivariate Normal distribution if there is an $2 \times r$ matrix A and a vector $(W_1, W_2, \ldots, W_r)$ of independent standard Normal random variables such that that

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \mathbf{A} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_r \end{bmatrix}$$

(2) Suppose that Z, ε, δ are i.i.d. each with a standard Normal distribution, and $Y = 10Z + \varepsilon,\ X = 8Z + \delta$. Show that (X, Y) has a multivariate Normal distribution.

(a) If (X, Y) has a multivariate Normal distribution with $\mu_X = \mu_Y = 0$, the conditional distribution of Y given $X = x$ is $N(\rho \frac{\sigma_X}{\sigma_Y} x, (1 - \rho^2)\sigma_y^2)$, where $\rho = \operatorname{corr}(X, Y)$. Find the conditional distribution of Y given $X = x$ for the pair defined in (2).

(b) Find a fixed number b and Normal random variable γ so that, for any x the random variable

$$Y'_x = bx + \gamma$$

has the same distribution as Y given $X = x$.

(c) Generate $g_1, \ldots, g_{100}$, each with the same distribution as γ found above, and set

$$y'_i = bx_i + g_i.$$

Plot $\{y_i'\}$ against $\{x_i\}$. Can you distinguish this plot from the plot made in (a)? What is the least-squares line for this data, and how does it compare to the least-squares line for the data $\{(x_i, y_i)\}$?

(d) What does this exercise say about the ability to infer, based on observational data, the effect of an intervention to change the value of a single variable?

Solution.

(1)

(a) We provide the following plot as instructed, using seed $= 38703$.

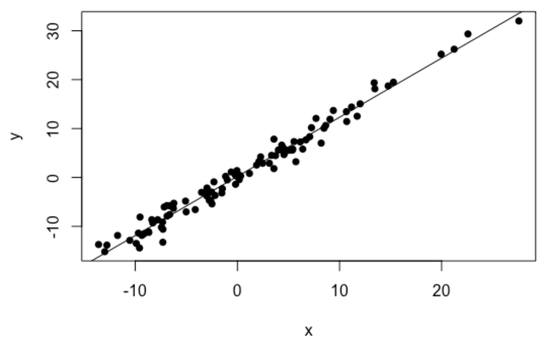

FIGURE 1. Scatter plot of random data, with least-squares line

(b) Given that x is a function of the random variables z and d the equation would not hold if we haphazardly changed a value of x without also changing its corresponding inputs. I think this follows in a fairly straightforward way from algebra, but I might be thinking about it too simplistically.

(2) We find that

$$\mathbf{A} = \begin{bmatrix} 10 & 1 & 0 \\ 8 & 0 & 1 \end{bmatrix}.$$

(a) Using the given formulae for the parameters of the conditional distribution for Y given $X = x$ is the following. Note that $\sigma_X = \sqrt{65}$, $\sigma_Y = \sqrt{101}$, $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, and $\mathrm{Cov}(X, Y) = 80$.

$$\begin{aligned} \frac{\sigma_Y}{\sigma_X}\rho x &= \frac{\sqrt{101}}{\sqrt{65}} \cdot \frac{80}{\sqrt{101}\sqrt{65}}, \\ &= \frac{80}{65}. \\ &\approx 1.231. \end{aligned}$$

$$\begin{aligned} \sigma_Y^2\left(1 - \rho^2\right) &= 101\left(1 - \left(\frac{80}{\sqrt{65}\sqrt{101}}\right)^2\right), \\ &\approx 2.5385. \end{aligned}$$

Hence, the conditional distribution for Y given that $X = x$ is $N(1.231x, 2.5385)$.

(b) Note that $\mathbb{E}(\gamma) = 0$, thus $\mathbb{E}(Y'_x) = \mathbb{E}(bx) + \mathbb{E}(\gamma) = bx$, thus from previous work we see that $b = 80/65$.

(c) We provide the following plot as instructed,

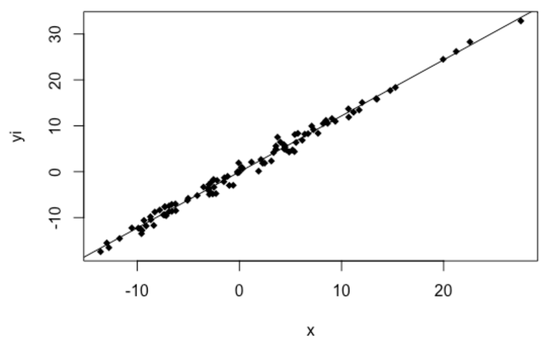

FIGURE 2. Scatter plot of slightly altered random data, with least-squares line

While some data points seem shifted slightly, the plots are difficult to tell apart.

(d) The best way that I can describe my thoughts on this, is that changing individual data points does not substantially effect our ability to make predictions with linear models.

$\square$

Problem B Load the R object gala by `load(url("http://pages.uoregon.edu/dlevin/DATA/gala.R"))` The variable $\mathbf{y}$ is given in the first column 'Species', and the variables x_i for $i = 1, 2, 3, 4, 5$ are given by the last four columns. This data records the number of species on islands in the Galapagos chain, along with other geographical and topological variables.

(a) Find the coefficients $b_0, b_1, b_2, b_3, b_4, b_5$ of the least-squares fit

$$y = b_0\mathbf{1} + b_1\mathbf{x_1} + b_2\mathbf{x_2} + b_3\mathbf{x_3} + b_4\mathbf{x_4} + b_5\mathbf{x_5} + \mathbf{e}$$

(where $e \perp \mathscr{L}\{\mathbf{1}, \mathbf{x_1}, \ldots, \mathbf{x_5}\}$) using the function lm in R.

(b) Plot $\mathbf{e}$ against $\hat{\mathbf{y}}$. What does this plot say about the fit of the least-squares linear function?

(c) Compute the least-squares coefficients in R using the matrix multiplication $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Note to get a matrix product in R, use `A%*%B`. The function `solve` can be used to invert a matrix.

Solution.

(a) We compute $\mathbf{b}$ and $\mathbf{e}$ in R and arrive at the following,

$$\mathbf{b} = \begin{bmatrix} 7.068221 \\ -0.023938 \\ 0.319465 \\ 0.009144 \\ -0.240524 \\ -0.074805 \end{bmatrix}$$

(b) We provide the following plot, as instructed.

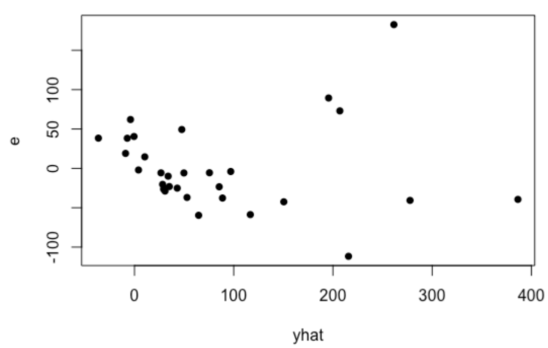

FIGURE 3. The slop as a function of the project $\hat{y}$

(c) We compute the least-squares coefficients in R, as instructed,

$$\mathbf{b} = \begin{bmatrix} 7.068220709 \\ -0.023938338 \\ 0.319464761 \\ 0.009143961 \\ -0.240524230 \\ -0.074804832 \end{bmatrix}$$

$\square$

Problem C Let $\mathbf{y} = (y_1, \dots, y_n)'$, $\mathbf{x} = (x_1, \dots, x_n)'$, $\mathbf{1} = (1, \dots, 1)'$ and $V = \mathscr{L}(\mathbf{1}, \mathbf{x})$.

(a) Use Gram-Schmidt orthogonalization on the vectors $\mathbf{1}$,$\mathbf{x}$ (in this order) to find orthogonal vectors $\mathbf{1}$,$\mathbf{x}^\star$ spanning V. Express $\mathbf{x}^\star$ in terms of $\mathbf{1}$ and $\mathbf{x}$, and find b_0,b_1 such that

$$\hat{\mathbf{y}} = b_0\mathbf{1} + b_1\mathbf{x}.$$

To simplify the notation let

$$\mathbf{y}^\star = \mathbf{y} - \pi(\mathbf{y}|\mathbf{1}) = \mathbf{y} - \bar{y}\mathbf{1},$$

$$S_{xy} = \langle \mathbf{x}^\star, \mathbf{y}^\star \rangle = \langle \mathbf{x}^\star, \mathbf{y} \rangle = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - \bar{x}\bar{y}n,$$

$$S_{xx} = \langle \mathbf{x}^\star, \mathbf{x}^\star \rangle = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - \bar{x}^2 n,$$

$$S_{yy} = \langle \mathbf{y}^\star, \mathbf{y}^\star \rangle = \sum_i (y_i - \bar{y})^2.$$

(b) Suppose

$$\hat{\mathbf{y}} = \pi(\mathbf{y}|V) = a_0\mathbf{1} + a_1\mathbf{x}^\star.$$

Find formulae for a_1 and a_0 in terms of $\bar{y}$, S_{xy}, S_{xx}.

(c) Express $\mathbf{x}^\star$ in terms of $\mathbf{1}$ and $\mathbf{x}$, and use this to determine formulas for b_1 and b_0 so that $y = b_0\mathbf{1} + b_1\mathbf{x}^\star$.

(d) Express $||\hat{\mathbf{y}}||^2$ and $||\mathbf{y} - \hat{\mathbf{y}}||^2$ in terms of S_{xy}, S_{xx}, and S_{yy}.

(e) Use the formula $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ for $\mathbf{b} = (b_0, b_1)'$ and verify that they are the same as those found in (c).

(f) For

$$\mathbf{y} = \begin{bmatrix} 2 \\ 6 \\ 7 \\ 8 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$$

find a_0, a_1, $\hat{\mathbf{y}}$, b_0, b_1, $||\mathbf{y}||^2$, $||\mathbf{y} - \hat{\mathbf{y}}||^2$. Verify that

$$||\hat{\mathbf{y}}|| = b_0\langle \mathbf{y}, \mathbf{1} \rangle + b_1\langle \mathbf{y}, \mathbf{x} \rangle,$$

and that $\mathbf{y} - \hat{\mathbf{y}} \perp V$.

Solution.

(a) We proceed with the Gram-Schmidt orthogonalization algorithm, note that $\mathbf{w}_1 = \mathbf{1}$.

$$\begin{aligned} \mathbf{w}_2 &= \mathbf{x} - \frac{\langle x, 1 \rangle}{\langle 1, 1 \rangle}\mathbf{1}, \\ &= \mathbf{x} - \frac{\sum x_i}{n}\mathbf{1}, \\ &= \mathbf{x} - \bar{x}\mathbf{1}, \\ &= \mathbf{x}^\star. \end{aligned}$$

Observe that $\mathbf{x}^\star$ is expressed in terms of $\mathbf{1}$ and $\mathbf{x}$ already. We will skip finding the b's until part (c).

(b) We compute the following,

$$\begin{aligned}\hat{y} &= a_0\mathbf{1} + a_1\mathbf{x}^\star,\\ &= a_0\mathbf{1} + a_1 x - a_1\bar{x}\mathbf{1},\\ &= (a_0 - a_1\bar{x})\,\mathbf{1} + a_1 x.\end{aligned}$$

Thus, we can see that $b_0 = a_0 - a_1\bar{x}$, $b_1 = a_1$.

(c) We compute the following,

$$\begin{aligned}\hat{y} &= \frac{\langle y, 1\rangle}{\langle 1, 1\rangle}\mathbf{1} + \frac{\langle y, x^\star\rangle}{\langle x^\star, x^\star\rangle}\mathbf{x}^\star,\\ &= \frac{\sum y_i}{n}\mathbf{1} + \frac{yx - y\bar{x}}{S_{xx}},\\ &= \bar{y}\mathbf{1} + \frac{S_{xy}}{S_{xx}}\mathbf{x}^\star.\end{aligned}$$

Thus, we can see that $a_0 = \bar{y}$ and $a_1 = \frac{S_{xy}}{S_{xx}}$

(d) We compute the following,

$$\begin{aligned}||\hat{y}||^2 &= \langle \hat{y}, \hat{y}\rangle,\\ &= \langle \bar{y}\mathbf{1}, \bar{y}\mathbf{1}\rangle + 2\left\langle \bar{y}\mathbf{1}, \frac{S_{xy}}{S_{xx}}x^\star\right\rangle + \left\langle \frac{S_{xy}}{S_{xx}}x^\star, \frac{S_{xy}}{S_{xx}}\right\rangle,\\ &= \bar{y}^2\langle 1, 1\rangle + \left(\frac{S_{xy}}{S_{xy}}\right)^2\langle x^\star, x^\star\rangle,\\ &= n\bar{y}^2 + \left(\frac{S_{xy}}{S_{xy}}\right)^2.\end{aligned}$$

$$\begin{aligned}||y - \bar{y}||^2 &= \langle y - \hat{y}, y - \hat{y}\rangle,\\ &= \langle y - \bar{y}\mathbf{1} - \frac{S_{xy}}{S_{xx}}x^\star, y - \bar{y}\mathbf{1} - \frac{S_{xy}}{S_{xx}}x^\star\rangle,\\ &= \langle y - \hat{y}, y - \hat{y}\rangle + 2\langle y - \bar{y}\mathbf{1}, -\frac{S_{xy}}{S_{xx}}x^\star\rangle + \left(\frac{S_{xy}}{S_{xx}}\right)^2\langle x^\star, x^\star\rangle,\\ &= S_{yy} - \left(\frac{S_{xy}}{S_{xx}}\right)^2.\end{aligned}$$

(e) We compute this in R.

(f) Upon running through the computations I was unable to verify any of what was asked. Given the lack of time remaining to complete the assignment, and my current zombie status due to lack of sleep, I'm calling it good.

$\square$

Problem D Let $\Omega = \mathbb{R}^4$, and

$$\mathbf{x}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix}1\\1\\1\\0\end{bmatrix},$$

and let $V_0 = \mathscr{L}(\mathbf{x}_4)$ for $\mathbf{x}_4 = 3\mathbf{x}_3 - 2\mathbf{x}_2$, $V = \mathscr{L}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$. Find Π_{V_0},Π_V, and Π_{V_1} for $V_1 = V_0^{\perp} \cap V$. For $\mathbf{y} = (0, 2, 14, 1)'$ find $\pi(\mathbf{y}|V_0)$, $\pi(\mathbf{y}|V_1)$, $\pi(\mathbf{y}|V)$.

Solution. We perform the majority of computations here in R, with the script attached. First, note that $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are mutually orthogonal, thus we can proceed without the need of Gram-Schmidt orthogonalization of $V = \mathscr{L}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ or $V_0 = \mathscr{L}(\mathbf{x}_1, \mathbf{x}_2)$. We compute the following,

$$\mathbf{x}_4 = \begin{bmatrix}1 & 1 & 3 & 0\end{bmatrix}^T.$$

Vector Spaces

$$V_0 = \operatorname{span}\begin{bmatrix}1 & 1\\1 & 1\\1 & 0\\1 & 0\end{bmatrix}, \qquad V = \operatorname{span}\begin{bmatrix}1 & 1 & 1\\1 & 1 & 1\\1 & 0 & 1\\1 & 0 & 0\end{bmatrix}$$

Projection Operators

$$\Pi_{V_0} = \begin{bmatrix}1/11 & 1/11 & 3/11 & 0\\1/11 & 1/11 & 3/11 & 0\\3/11 & 3/11 & 9/11 & 0\\0 & 0 & 0 & 0\end{bmatrix}, \quad \Pi_{V_1} = \begin{bmatrix}9/22 & 9/22 & -3/11 & 0\\9/22 & 9/22 & -3/11 & 0\\-3/11 & -3/11 & 2/11 & 0\\0 & 0 & 0 & 1\end{bmatrix},$$

$$\Pi_V = \begin{bmatrix}1/2 & 1/2 & -3/11 & 0\\1/2 & 1/2 & -3/11 & 0\\-3/11 & -3/11 & 2/11 & 0\\0 & 0 & 0 & 0\end{bmatrix}.$$

Projections

$$\pi\left(\mathbf{y}|V_0\right) = \begin{bmatrix}4\\4\\12\\0\end{bmatrix}, \quad \pi\left(\mathbf{y}|V_1\right) = \begin{bmatrix}-3\\-3\\2\\1\end{bmatrix}, \quad \pi\left(\mathbf{y}|V\right) = \begin{bmatrix}1\\1\\14\\1\end{bmatrix}.$$

□