TQSE2

Q1. For the information given in the table below, calculate the standardized value of a new observation: x = [12.0 0.78 190.2].

a. Does the standardized observation indicate that any of the measured variables may be an outlier (and explain how you know)?

b. If so, which one?

	Mean	Standard Deviation
Training	[20.1 -0.21 170.3]	[3.52 0.82 4.42]
Test	[15.6 0.60 200.3]	[0.73 0.76 4.59]
Validation	[10.2 1.01 195.7]	[1.34 0.05 8.67]

For the remainder of the questions, suppose you have 1000 observations of 6 process variables as predictors. The **eigenvalues** of the correlation matrix of the data are λ = 4.58, 1.03, 0.21, 0.15, 0.029, and 0.001.

c. What is the condition number of the predictor matrix for linear regression if we include all 6 process variables?

d. What is the minimum regularization parameter (α) needed to have a well-conditioned matrix in ridge regression?

e. If these data are transformed to the PC space, how many PCs will we need to explain at least 95% of the information?

Q2. If the table below gives the correlation coefficients of the six predictors with the output, indicate which variables you would include in a linear regression model and explain why?

	Correlation to y
x1	0.90
x2	0.15
x3	-0.63
x4	-0.78
x5	0.55
x6	-0.04