Practice problems for data, distributions, and correlation with partial answers

1. For the multivariate normal distribution given in the notes, , complete the following:
	1. Find f(**x**) at **x** = **μ** by actually multiplying out the necessary matrices (do not use dmvnorm()).

Notice that  is simply  when **x** = **μ**.

* 1. Find f(**x**) at **x** = [14, 19]′ by actually multiplying out the necessary matrices and by using dmvnorm().

> mu <- c(15, 20)

> sigma <- matrix(data = c(1, 0.5, 0.5, 1.25), nrow = 2, ncol = 2, byrow = TRUE)

> dmvnorm(x = x, mean = mu, sigma = sigma)

[1] 0.0851895

* 1. Plot the distribution and explore what happens to the distribution when changes are made to the mean vector and the covariance matrix. Make sure to examine the eigenvalues as well.
1. Simulate data from a multivariate normal distribution of your choice (it does not need to be bivariate). Compare the multivariate summary statistics to your **μ** and **Σ**. Examine what happens to the summary statistics relative to **μ** and **Σ** as N increases or decreases.
2. The covariance and correlation matrix are the same for standardized random variables. Why?

If z1 and z2 are standardized random variables, then they both have a mean of 0 and variance of 1. The correlation between the two variables is



1. With respect to the goblet data described in the notes (data is in goblet.csv of Section 2), complete the following:
	1. What is the experimental unit?

A goblet

* 1. Subject-matter researchers are interested in grouping goblets that have the same shape although they may have different sizes. One way suggested by Manly and Alberto (2016) is to adjust the data by dividing each measurement by X3 (height) and then complete additional analyses (shown later in our course). Below is how the data adjustment is done:

> goblet <- read.csv("C:\\chris\\goblet.csv")

> head(goblet)

 goblet x1 x2 x3 x4 x5 x6

1 1 13 21 23 14 7 8

2 2 14 14 24 19 5 9

3 3 19 23 24 20 6 12

4 4 17 18 16 16 11 8

5 5 19 20 16 16 10 7

6 6 12 20 24 17 6 9

> goblet2 <- data.frame(ID = goblet$goblet, w1 = goblet$x1/goblet$x3,

 w2 = goblet$x2/goblet$x3,

 w4 = goblet$x4/goblet$x3,

 w5 = goblet$x5/goblet$x3,

 w6 = goblet$x6/goblet$x3)

> head(goblet2)

 ID w1 w2 w4 w5 w6

1 1 0.5652174 0.9130435 0.6086957 0.3043478 0.3478261

2 2 0.5833333 0.5833333 0.7916667 0.2083333 0.3750000

3 3 0.7916667 0.9583333 0.8333333 0.2500000 0.5000000

4 4 1.0625000 1.1250000 1.0000000 0.6875000 0.5000000

5 5 1.1875000 1.2500000 1.0000000 0.6250000 0.4375000

6 6 0.5000000 0.8333333 0.7083333 0.2500000 0.3750000

* 1. Exhibit the adjusted data in a matrix.
	2. Find the multivariate summary statistics for the adjusted data.

> sigma.hat

 w1 w2 w4 w5 w6

w1 0.0588151251 0.032664837 0.017045208 0.02224044 0.0004540383

w2 0.0326648368 0.024199433 0.009612116 0.01494699 0.0003982420

w4 0.0170452077 0.009612116 0.017656026 0.01329360 0.0068352800

w5 0.0222404418 0.014946989 0.013293604 0.01940272 0.0018132298

w6 0.0004540383 0.000398242 0.006835280 0.00181323 0.0072358002

* 1. Compute the covariance of the adjusted X1 and X2 variables without the use of cov() or any function that immediately finds the covariance. In other words, program the actual formula into R that computes the covariance for two variables and apply it to data from these variables.