1. Researchers want to assess the performance of two new driveway sealants. To do so, they first divide the country into six climatologically similar regions. Within each region, they obtain a simple random sample of twelve single-family homes belonging to owners willing to participate in this three year-long study. Six of the homes in each region are randomly chosen to test driveway sealant 1 and the other six receive driveway sealant 2. The sealants are compared with respect to their performance on a number of measures over the course of the test.

(a) (10 points) What sampling method was used to select the homes used in the study? Be specific, and support your answer.

ANS: Stratified random sampling, in which the six climatologically similar regions are the strata, and a simple random sample is chosen from each.

(b) (10 points) What kind of study is this? Be specific, and support your answer.

ANS: This is a controlled experiment since treatments (sealants) are imposed on experimental units (driveways) in order to observe a response (the measures used to compare the sealers).

(c) (10 points) Explain how blocking could be used to improve the study. Be specific: tell what the blocks are, how the blocking will be done and why it will be beneficial.

ANS: There is no unique answer. Here are two: (1) (The best solution) Divide each driveway into two sections. Apply one of the sealants (randomly chosen) to section 1 and the other to section 2. Here each driveway is a block. This will greatly reduce driveway to driveway variation. (2) Pair up the homes by locale and other factors (driveway in the same or nearly the same condition, etc.). Then each pair is a block. Randomly select one driveway in each block to receive sealant 1, and give the other sealant 2. This will reduce variation due to locale and the other factors defining the blocks, but is unlikely to reduce variation as effectively as (1), hence is not as good a solution.

2. Figure 1 displays a box and whiskers plot for a set of data.

(a) (10 points) Which of the following data sets could NOT have produced the plot in Figure 1 (For each one you conclude could not have produced the plot, give a reason):

i. 1 2 4 5 8 9 11 29 33 37 39 40 41 44 47
ii. 1 2 4 5 8 9 11 29 33 37 39 40 41 44 45
iii. 1 2 4 5 8 9 11 29 33 37 39 40 41 43 45
iv. 1 2 4 6 8 9 11 29 33 37 39 40 41 43 45
v. 1 3 4 5 8 9 11 29 33 37 39 40 41 43 45

ANS: i. could not have produced the plot because 47 would appear as an outlier. iii. could not have produced the plot because its median is 31, not 29 as shown on the plot. iv. could not have produced the plot because its Q1 is 6, not 5 as shown on the plot.

(b) (10 points) Is the box and whiskers plot in Figure 1 a good graphical summary for those data sets from part (a) that could have produced this plot? Justify your answer.

ANS: No, it is not a good graphical summary for either ii. or v. since the main feature of these data is that there are two distinct clusters: 1-11 and 29-45, and the box and whisker plot disguises this feature.
3. Diameter specifications for mylar sheaths produced by the Ebco Company are 15 ± 2 mm. Measurements taken from a large number of production units suggest the diameters $Y$ of the population of sheaths follow a normal distribution with mean 15.1 and standard deviation 0.9.

(a) **(10 points)** What is the population proportion of sheaths that fall within spec?

**ANS:** If $Y$ is the diameter of a randomly chosen sheath,

$$P(13 < Y < 17) = P\left( \frac{13 - 15.1}{0.9} < Z < \frac{13 - 15.1}{0.9} \right) = P(-2.33 < Z < 2.11) = 0.97727$$

(b) **(10 points)** The value of the $N(15.1, 0.9^2)$ density curve at 16.2 is one half its value at 15.1. Relate this fact to the relative likelihood of finding a sheath with a diameter close to 15.1 compared with finding one with diameter close to 16.2.

**ANS:** It is twice as likely to find one with a diameter close to 15.1 as it is to find one with diameter close to 16.2.

4. The eccentricity of a mouse tracking ball is a measure of how out-of-round it is. A perfectly round ball has value of 1, an out-of-round ball has a value greater than 1. Suppose the eccentricity of a population of mouse tracking balls is measured by the random variable $Y$ having density curve

$$p(y) = \begin{cases} 
0, & y < 1, \\
\frac{c}{y^5}, & y \geq 1 
\end{cases}$$

(a) **(10 points)** Find the value of the constant $c$ for the density curve.

**ANS:**

$$1 = \int_{-\infty}^{\infty} p(y) dy = \int_{1}^{\infty} \frac{c}{y^5} dy = -\frac{c}{4y^4} \bigg|_{1}^{\infty} = \frac{c}{4}$$

So $c = 4$.

(b) **(5 points)** A mouse tracking ball is considered unacceptable if its eccentricity is greater than 10. What proportion of the mouse tracking balls are unacceptable due to excessive eccentricity?

**ANS:**

$$P(Y > 10) = \int_{10}^{\infty} \frac{4}{y^5} dy = -\frac{1}{y^4} \bigg|_{10}^{\infty} = 0.0001$$

So the proportion is 0.1.

(c) **(10 points)** How much more likely is it to find a mouse tracking ball with eccentricity close to 5 than to find one with eccentricity close to 10?

**ANS:** It is $p(5)/p(10) = 2^5 = 32$, times as likely.
Figure 1: *Box and whiskers plot for question 2.*