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 shows a box and whiskers plot for a set of 496 salaries.

   (a) **(10 points)** Is this box and whiskers plot an appropriate plot to summarize the pattern of variation of the data? Choose one: □YES □NO □CAN'T TELL. Explain your choice.

   ANS: *The correct answer is CAN'T TELL, because the pattern of variation could exhibit important features, such as multimodality, that the box and whiskers plot would hide.*

   (b) **(10 points)** Assuming the box and whiskers plot is an appropriate summary of the distribution, obtain a corresponding measure of location and spread. Tell what each measure means.

   ANS: *LOCATION: $Q_2 = 48378$ is the point separating the upper and lower half of the data. SPREAD: IQR= $76520 - 29243 = 47277$ is the range of the middle half of the data.*

3. A manufacturer of audio CDs claims the rate of defective disks is 0.015% (i.e., a proportion 0.00015 of all disks is defective.)

   (a) **(20 points)** Approximate the probability that in a shipment of 100,000 disks, 25 or more are defective.

   ANS: *If Y is the number of defective disks, and if the manufacturer's claim is true, then $Y \sim b(100000, 0.00015)$* **(10 points)** *We will use the normal approximation:*

$$P(Y \geq 25) = P(Y \geq 24.5) = P\left(\frac{Y - (100000)(0.000015)}{\sqrt{(100000)(0.000015)(1 - 0.000015)}} \geq \frac{24.5 - (100000)(0.000015)}{\sqrt{(100000)(0.000015)(1 - 0.000015)}}\right)$$

$$\approx P(Z \geq 2.453) = 0.0071$$

   (10 points)

   (b) **(5 points)** Justify the approximation you made in (a).

   ANS: *The approximation is justified, since both $np = (100000)(0.000015) = 15$, and $n(1 - p) = 99985$ exceed 10.*

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

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

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

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

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:

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

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?

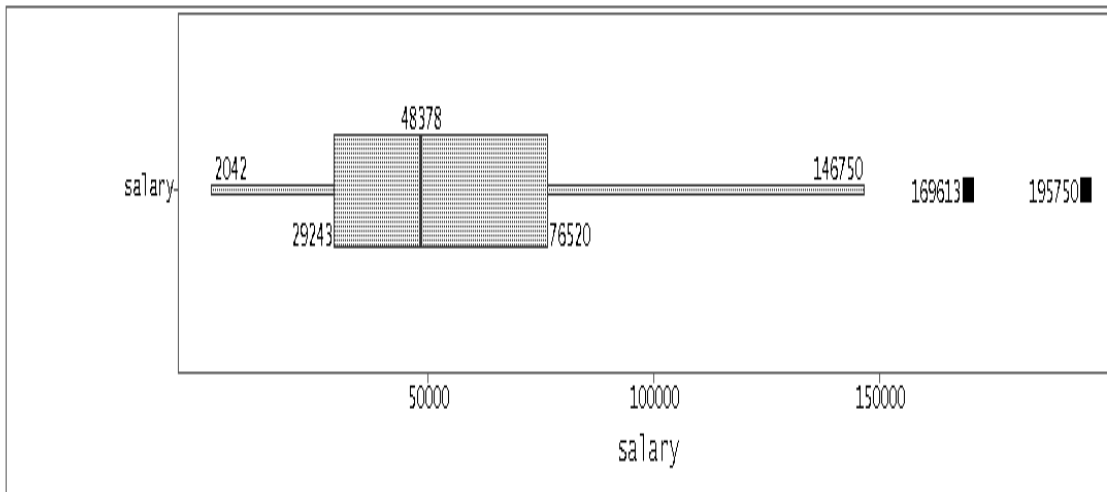**ANS:** *It is* $p(5)/p(10) = 2^5 = 32$, *times as likely.*

Figure 1: *Box and whiskers plot for a set of 496 salaries.*