Question 1: I am hired to test the effectiveness of a flu vaccine on Berkeley students. I wish to vaccinate a simple random sample of students, but only half of them consent to be vaccinated. I vaccinate all those who consent (the treatment group), while retaining those who do not consent as a control group. I compare the percentage of the treatment group who get the flu to the percentage of the control group who get the flu using a two-sample z-test.

What is the major flaw in my experimental design? How could it be improved?

The histogram of incomes for employees of a large company is shown below:

Histogram of income

Question 2: What is the median income of the employees at the company?
Question 3: The correlation between height and weight for adults is 0.7. Body mass index (BMI) is equal to weight divided by weight squared. Multiple choice: the correlation between height and BMI is:
(a) equal to 0.7
(b) equal to -0.7
(c) equal to 0.3
(d) equal to 0.49
(e) cannot be determined from the information given

In a very large class, midterm scores are approximately normally distributed, with average 50 and SD 15. Final scores are also approximately normally distributed, with average 60 and SD 10. The scatter plot of final scores against midterm scores is football-shaped, with a correlation of 0.6.

Question 4: What percentage of students scored higher than 75 on the final?

Question 5: What is the regression prediction for the final score of a student who scored 45 on the midterm?

Question 6: What is the RMS error of the regression?

Question 7: I have a bag containing ten marbles: four blue and six green. I draw five marbles with replacement. What is the probability that exactly three of the marbles I draw are blue?

Question 8: Which of the following two events is more likely, and why?
(a) I toss a fair coin 100 times, and get exactly 50 heads.
(b) I toss a fair coin 1000 times, and get exactly 500 heads.
Note: You do not need to find the exact probabilities.

Question 9: I roll a biased die that has the following set of probabilities for each outcome:

<table>
<thead>
<tr>
<th>Number of spots</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20%</td>
</tr>
<tr>
<td>2</td>
<td>10%</td>
</tr>
<tr>
<td>3</td>
<td>10%</td>
</tr>
<tr>
<td>4</td>
<td>10%</td>
</tr>
<tr>
<td>5</td>
<td>20%</td>
</tr>
<tr>
<td>6</td>
<td>30%</td>
</tr>
</tbody>
</table>

What is the expected number of spots I see on one roll of the die?
I have a box containing 1000 tickets, each with a number written on it. I make 400 draws from the box with replacement. The average of these 400 draws is 10, while their SD is 30.

**Question 10:** Find a 95% confidence interval for the average of the tickets in the box.

I now examine all 1000 tickets. I find their average is 11.

**Question 11:** Can you now find a 95% confidence interval for the average of the tickets in the box? If so, calculate the confidence interval. If not, explain why you cannot.

**Question 12:** A mayoral election is to be held in a town where the population is 40% black, 30% white, 20% Latino and 10% Asian. Two survey companies take opinion polls on the mayoral election. Company A takes a quota sample of 400 black residents, 300 white residents, 200 Latino residents and 100 Asian residents, and asks them which mayoral candidate they prefer. The interviewers for Company A may select the members of the sample, as long as the quotas are met. Company B take a simple random sample of 1000 town residents. This include 390 black residents, 310 white residents, 195 Latino residents and 105 Asian residents.

Which has the better sampling design, Company A or Company B? Why?

Of a simple random sample of 240 Berkeley students, 150 support Proposition Z. Of a simple random sample of 200 Stanford students, 120 support Proposition Z.

**Question 13:** Find a 95% confidence interval for the percentage of all Berkeley students who support Proposition Z. (You do not need to use the correction factor.)

**Question 14:** To see if there is a statistically significant difference between the percentage of Berkeley students who support Proposition Z and the percentage of Stanford students who support Proposition Z, I wish to carry out a two-sample $z$-test. To do this, I need to estimate the standard error of the difference of the two sample percentages. What is this standard error?

**Question 15:** For my two-sample $z$-test, I end up getting a two-tailed $P$-value of 59%. What should I conclude from this? (If you accept or reject a hypothesis, then state that hypothesis.)

**Question 16 (hard):** I have an equal number of red, blue and green marbles in a large box. I draw four marbles from the box, with replacement. What is the probability I get at least one of each color?