Remember that the purpose of the exam is to show me how you think about these problems. Accordingly, please show your work, use mathematical notation correctly, and justify your answers.

1. We’ve talked a lot about randomization in this class, and indeed a table of random numbers is included at the back of this exam.

(a) (10 points) Explain in a single sentence why randomization is often incorporated into the design of experiments and samples.

Randomization is used to help eliminate bias.

(b) (10 points) To talk intelligibly about randomness, we generally use the language of probability. Describe in simple words what it means to say that “the probability of getting Type II diabetes at some point in your life is 10%.”

If you were to randomly select a very large group of people who are “like you,” and not how many of them get diabetes, then about 10% will come down with the disease.

2. (10 points) Suppose you find yourself one day debating with your statistically minded spouse the best way to choose a name for your first born child. Remembering your Math 160 class, you propose to choose three letters at random from the alphabet, and then choose a name whose initials correspond to these letters. Your spouse points out that in order to give each letter a fair chance of being chosen, you’ll need to select a simple random sample of size three. Use the random number table at the end of this exam to select a simple random sample of size three from the set of 26 English language letters,


Explain the exact procedure you followed to select your sample. (Bonus: what will you call your child?)

Assign letters: A=0, B=1, etc. I use line 16 of random digit table. The 1st set of digit is 59, 14, 19, which correspond to I, O, and S. The child’s name will be

Igor Octavious Sergiush.
3. **(20 points)** Explain what is wrong with each of the following experimental or sampling procedures, and what you might do to correct things.

(a) A study compares two marketing campaigns to encourage people to eat less meat. The first campaign is launched in Montana and the second is launched in California.

> Atitudes to meat eating are probably very different in CA and MT. I would launch the campaigns in the same market to ensure comparable results.

(b) In order to get a sense of student exam-taking ethics, a teacher asks students to raise their hand if they have ever cheated on an exam.

> Students might be reluctant to admit to cheating, and this will lead to response bias. Anonymous feedback, via a survey, would be better.

(c) Forty smokers are to be used to measure the effectiveness of a smoking-cessation drug. Twenty women are assigned to receive the treatment, and twenty men are assigned to be the control.

> Women and men may respond differently. Better to have 10 women and 10 men, like the treatment, 10 women and 10 men as the control.

(d) A medical researcher is interested in investigating the effects of a new blood pressure drug. She gives herself 3mg/day for a month and records her blood pressure.

> Lacks rejection comparison, a control. Better to give the treatment to a high subject.
4. Suppose the age at which a ghost acquires its first set of chains is a normally distributed random variable, with mean 120 and standard deviation 20.

(a) (7 points) What is the probability that a randomly selected ghost receives his first set of chains before he turns 100?

\[ X = \text{age of first chain} \]
\[ P(X < 100) \approx 0.16 \]

(b) (8 points) What percentage of ghosts are between 100 and 150 years old when they get their first set of chains?

\[ P(100 < X < 150) = \frac{150 - 100}{20} = \frac{50}{20} = \frac{2.5}{2.5} = 0.9738 \]

5. Suppose that 30% of all students at UPS like to listen to bluegrass music.

(a) (3 points) What is the probability that a randomly chosen UPS student does not like bluegrass?

\[ 0.70 \]

(b) (12 points) Suppose you choose two students at random, ask each of them whether or not they like bluegrass, and keep track of the sequence of responses. What is the sample space? Also write down the probability distribution for this sample space. i.e. for each element in the sample space, specify the probability that it occurs. (Hint: assume that the responses of the two students are independent of one another.)

\[ \begin{array}{cccc}
X_1 & Y_1 & Y_2 & N_2 \\
\ P & .3^2 & .3(.7) & .7^2 \\
.09 & .21 & .21 & .49
\end{array} \]
(c) \textbf{(5 points)} What is the probability that if you choose two students at random, one likes bluegrass and one doesn't?

\[
P(NY \text{ or } YN) = 0.2(1.42) = 0.284
\]

6. Suppose \( X \) is a random variable whose distribution is as follows:

\[
\begin{array}{c|c|c|c}
\text{x} & -1 & 0 & 4 \\
\hline
\text{p} & 0.5 & 0.25 & 0.25
\end{array}
\]

(a) \textbf{(5 points)} Calculate the mean of \( X \)

\[
\mu_x = -1(0.5) + 0(0.25) + 4(0.25) = -0.5 + 1 = 0.5
\]

(b) \textbf{(5 points)} Calculate the variance of \( X \)

\[
\sigma_x^2 = (-1-0.5)^2(0.5) + (0-0.5)^2(0.25) + (4-0.5)^2(0.25) = \frac{9}{4} \cdot \frac{2}{4} + \frac{1}{4} \cdot \frac{1}{4} + \frac{49}{4} \cdot \frac{1}{4} = \frac{1}{18} \cdot \left[ 18 + 1 + 49 \right] = \frac{66}{18} = \frac{33}{9} = \frac{11}{3} = 4.444
\]

(c) \textbf{(5 points)} Suppose \( Y \) is a different random variable with mean \( \mu_y = 0.5 \) and \( \sigma_y^2 = 0.75 \). What are the mean and variance of the random variable \( X + Y \)?

\[
\mu_{X+Y} = \mu_x + \mu_y = 0.5 + 0.5 = 1
\]

\[
\sigma_{X+Y}^2 = \sigma_x^2 + \sigma_y^2 = 4.444 + 0.75 = 5.194
\]