Exercises

These exercises have been collated from many sources. Some are much more difficult than others, but all could appear on an exam. Some are answered explicitly in the textbook, but you should try to answer them without looking at the book.

1. The Monty Hall puzzle. There are three closed doors, behind one of which is a valuable prize, and behind two of which are approximately-valueless prizes. The contestant picks a door, which remains closed. The host, who knows where the prize is, opens one of the doors containing the low-value items. He then offers the contestant the chance to switch her choice to the remaining unopened door, or to stay with her initial choice.

Since the prize could be behind either remaining door, it often seems that the chances are equal of winning with either door. But this is not so. What are the probabilities of winning the valuable prize (i) by staying with the initial choice, or (ii) switching to the unopened door?

2. Consider a sample of n observations, $y_1, y_2, \ldots, y_n$, on some random variable Y. The empirical distribution function of this sample is a discrete distribution with n possible points. These points are just the n observed points, $y_1, y_2, \ldots, y_n$. Each point is assigned the same probability, which is just $1/n$, in order to ensure that all the probabilities sum to 1.

Compute the expectation of the discrete distribution characterized by the EDF, and show that it is equal to the **sample mean**, that is, the unweighted average of the n sample points, $y_1, y_2, \ldots, y_n$.

3. Suppose that X and Y are two binary random variables. Their joint distribution is given in the following table.

	$Y = 0$	$Y = 1$
$X = 0$	.16	.37
$X = 1$	.29	.18

What is the marginal distribution of Y? What is the distribution of Y conditional on $X = 0$? What is the distribution of Y conditional on $X = 1$?

Demonstrate the Law of Iterated Expectations explicitly by showing that $\mathrm{E}\big(\mathrm{E}(X \mid Y)\big) = \mathrm{E}(X)$. Let $h(Y) = Y^3$. Show explicitly that $\mathrm{E}\big(Xh(Y) \mid Y\big) = h(Y)\mathrm{E}(X \mid Y)$ in this case.

4. Using the following expression for the density $\phi(x)$ of the standard Normal distribution:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big),$$

show that the derivative of $\phi(x)$ is the function $-x\,\phi(x)$, and that the second derivative is $(x^2 - 1)\phi(x)$. Use these facts to show that the expectation of a standard normal random variable is 0, and that its variance is 1. These two properties account for the use of the term “standard.”

5. Let a random variable X_1 be distributed as $\mathrm{N}(0, 1)$. Now suppose that a second random variable, X_2, is constructed as the product of X_1 and an independent random variable Z, which equals 1 with probability $1/2$ and -1 with probability $1/2$. What is the (marginal) distribution of X_2? What is the covariance between X_1 and X_2? What is the distribution of X_1 conditional on X_2?

6. Consider two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of dimensions such that the product $\boldsymbol{AB}$ exists. Show that the i^{th} row of $\boldsymbol{AB}$ is the matrix product of the i^{th} row of $\boldsymbol{A}$ with the entire matrix $\boldsymbol{B}$. Show that this result implies that the i^{th} row of a product $\boldsymbol{ABC}\ldots$, with arbitrarily many factors, is the product of the i^{th} row of $\boldsymbol{A}$ with $\boldsymbol{BC}\ldots$.

7. Consider two invertible square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, of the same dimensions. Show that the inverse of the product $\boldsymbol{AB}$ exists and is given by the formula

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}.$$

8. Show that the transpose of the product of an arbitrary number of factors is the product of the transposes of the individual factors in completely reversed order:

$$(\boldsymbol{ABC}\cdots)^\top = \cdots \boldsymbol{C}^\top\boldsymbol{B}^\top\boldsymbol{A}^\top.$$

Show also that an analogous result holds for the inverse of the product of an arbitrary number of factors.

9. Define, explain, and discuss each of the following:

(a) mean squared error, bias and variance of an estimator

(b) the relationship among the quantities in (a)

(c) Type I error

(d) the Chebychev inequality

(e) symmetry and skewness

10. The number of alcoholic drinks consumed per week by undergrads at the University of Southern North Dakota has a mean of 22, and standard deviation of 4. Treat the population distribution as Normal. A random sample of students is taken.

a) find the probability of obtaining the misleading result that the mean number of drinks of the students in the sample is less than or equal to 14 per week if one samples (i) 1, (ii) 4, (iii) 25 students.

b) explain why the three answers differ, and illustrate with a graph.

c) if you did not know that the population was Normal, explain conditions under which you could give approximate answers to each part of question (a). Compute your approximate answers.

11. A pollster surveys n randomly sampled women, and another n randomly sampled men, asking them if they're going to vote yes or no, or have not decided, on a referendum to require a balanced government budget. Among women, 48% say yes; among men 54% say yes. Assume that all respondents give answers that genuinely reflect their intentions.

a) Find 95% confidence intervals for the proportions of men and women who will vote yes.

b) Assume that there are equal numbers of registered male and female voters and that an equal proportion will vote. Find a 95% confidence interval for the overall proportion of yes votes.

c) Test the hypothesis that the proportions of male and female voters who will vote yes is in fact the same. Give a P-value. Interpret the result.

12. A sociologist wants to know the opinions of employed adult women about government funding for day care. She obtains a list of 520 members of a local business and professional womens club and mails a questionnaire to 100 of these women selected at random. Sixty-eight questionnaires are returned. What is the population in this study?

a) all employed adult women;

b) all the members of a local business and professional womens club;

c) the 100 women who received the questionnaire;

d) all employed women with children ?

13. A sample of pounds lost, in a certain month, by individual members of a weight reducing clinic produced the following statistics:

Mean = 5 lbs.
Median = 4.5 lbs.
Mode = 4 lbs.
Standard deviation = 3.8 lbs.
First quartile = 2 lbs.
Third quartile = 8.5 lbs.

The correct statement is:

One fourth of the members lost exactly two pounds.
The middle fifty percent of the members lost from two to 8.5 lbs.
Most people lost 3.5 to 4.5 lbs.
All of the choices above are correct.

14. The Table below contains the total number of deaths worldwide as a result of earthquakes from 2000 to 2012.

Year	Total Number of Deaths
2000	231
2001	21,357
2002	11,685
2003	33,819
2004	228,802
2005	88,003
2006	6,605
2007	712
2008	88,011
2009	1,790
2010	320,120
2011	21,953
2012	768
Total	823,856

Use the Table to answer the following questions.

a) What is the proportion of deaths between 2007 and 2012?

b) What percent of deaths occurred before 2001?

c) What is the percent of deaths that occurred in 2003 or after 2010?

d) What is the fraction of deaths that happened before 2012?

e) What kind of data is the number of deaths?

15. The students in Ms. Ramirez's math class have birthdays in each of the four seasons. The Table shows the four seasons, the number of students who have birthdays in each season, and the percentage of students in each group. Construct a bar graph showing the number of students. Construct a bar graph showing the percentages.

Seasons	Number of students	Proportion of population
Spring	8	24%
Summer	9	26%
Autumn	11	32%
Winter	6	18%

16. There are 23 countries in North America, 12 countries in South America, 47 countries in Europe, 44 countries in Asia, 54 countries in Africa, and 14 in Oceania (Pacific Ocean region).

Let A = the event that a country is in Asia.

Let E = the event that a country is in Europe.

Let F = the event that a country is in Africa.

Let N = the event that a country is in North America.

Let O = the event that a country is in Oceania.

Let S = the event that a country is in South America.

If a country is selected at random, find the probabilities P(A), P(E), P(F), P(N), P(O), and P(S).

17. E and F are mutually exclusive events. P(E) = 0.4; P(F) = 0.5. Find P(E| F).

J and K are independent events. P(J| K) = 0.3. Find P(J).

U and V are mutually exclusive events. P(U) = 0.26; P(V) = 0.37. Find: P(U AND V), P(U| V), and P(U OR V).

Q and R are independent events. P(Q) = 0.4 and P(Q AND R) = 0.1. Find P(R).

17. A game involves selecting a card from a regular 52-card deck and tossing a coin. The coin is a fair coin and is equally likely to land on heads or tails.

If the card is a face card, and the coin lands on Heads, you win $6

If the card is a face card, and the coin lands on Tails, you win $2

If the card is not a face card, you lose $2, no matter what the coin shows.

Find the expected value for this game (expected net gain or loss).

Explain what your calculations indicate about your long-term average profits and losses on this game.

Should you play this game to win money?

18. Suppose that the value of a stock varies each day from $16 to $25 with a uniform distribution.

Find the probability that the value of the stock is more than $19.

Find the probability that the value of the stock is between $19 and $22.

Find the upper quartile – 25% of all days the stock is above what value? Draw the graph.

Given that the value of the stock is greater than $18, find the probability that its value is more than $21.

19. A personal computer is used for office work at home, research, communication, personal finances, education, entertainment, social networking, and a myriad of other things. Suppose that the average number of hours a household personal computer is used for entertainment is two hours per day. Assume the times for entertainment are Normally distributed and the standard deviation for the times is half an hour.

Find the probability that a household personal computer is used for entertainment between 1.8 and 2.75 hours per day.

Find the maximum number of hours per day that the bottom quartile of households uses a personal computer for entertainment.

20. The length of time, in hours, it takes an "over-40" group of people to play one football match is Normally distributed with a mean of two hours and a standard deviation of 0.5 hours. A sample of size $n = 50$ is drawn randomly from the population. Find the probability that the sample mean is between 1.8 hours and 2.3 hours.

21. The distribution of results from a cholesterol test has a mean of 180 and a standard deviation of 20. A sample size of 40 is drawn randomly.

Find the probability that the sum of the 40 values is greater than 7,500.

Find the probability that the sum of the 40 values is less than 7,000.

Find the sum that is one standard deviation above the mean of the sums.

Find the sum that is 1.5 standard deviations below the mean of the sums.

Find the percentage of sums between 1.5 standard deviations below the mean of the sums and one standard deviation above the mean of the sums.

22. The length of time a particular smartphone's battery lasts follows an exponential distribution with a mean of ten months. A sample of 64 of these smartphones is taken.

What is the standard deviation?

What is the distribution for the length of time one battery lasts?

What is the distribution for the mean length of time 64 batteries last?

What is the distribution for the total length of time 64 batteries last?

Find the probability that the sample mean is between seven and 11.

Find the 80th percentile for the total length of time 64 batteries last.

Find the inter-quartile range (IQR) for the mean amount of time 64 batteries last.

Find the middle 80% for the total amount of time 64 batteries last.

23. A random variable X has the Normal distribution $\mathrm{N}(54, 8)$.

Find the probability that $x > 56$.

Find the probability that $x < 30$.

Find the 80th percentile.

Find the 60th percentile.

24. A sample of 20 heads of lettuce was selected. Assume that the population distribution of head weight is normal. The weight of each head of lettuce was then recorded. The mean weight was 2.2 pounds with a standard deviation of 0.1 pounds. The population standard deviation is known to be 0.2 pounds.

Construct a 90% confidence interval for the population mean weight of the heads of lettuce.

Construct a 95% confidence interval for the population mean weight of the heads of lettuce.

What would happen if 40 heads of lettuce were sampled instead of 20, and the confidence level remained the same?

25. A sociologist claims the probability that a person picked at random in Times Square in New York City is visiting the area is 0.83. You want to test to see if the claim is correct. State the null and alternative hypotheses.

In a population of fish, approximately 42% are female. A test is conducted to see if, in fact, the proportion is less. State the null and alternative hypotheses.

26. A particular brand of tires claims that its deluxe tire averages at least 50,000 miles before it needs to be replaced. From past studies of this tire, the standard deviation is known to be 8,000. A survey of owners of that tire design is conducted. From the 28 tires surveyed, the mean lifespan was 46,500 miles with a standard deviation of 9,800 miles. Using a significance level $\alpha = 0.05$, is the data highly inconsistent with the claim?

27. From generation to generation, the mean age when smokers first start to smoke varies. However, the standard deviation of that age remains constant of around 2.1 years. A survey of 40 smokers of this generation was done to see if the mean starting age is at least 19. The sample mean was 18.1 with a sample standard deviation of 1.3. Do the data support the claim at the 5% level?

28. The mean number of sick days an employee takes per year is believed to be about ten. Members of a personnel department do not believe this figure. They randomly survey eight employees. The number of sick days they took for the past year are as follows: 12; 4; 15; 3; 11; 8; 6; 8. Let $x =$ the number of sick days they took for the past year. Should the personnel team believe that the mean number is ten?

29. Your statistics instructor claims that 60 percent of the students who take her Elementary Statistics class go through life feeling more enriched. For some reason that she can't quite figure out, most people don't believe her. You decide to check this out on your own. You randomly survey 64 of her past Elementary Statistics students and find that 34 feel more enriched as a result of her class. Now, what do you think?

30. The mean number of English courses taken in a two-year time period by male and female college students is believed to be about the same. An experiment is conducted and data are collected from 29 males and 16 females. The males took an average of three English courses with a standard deviation of 0.8. The females took an average of four English courses with a standard deviation of 1.0. Are the means statistically the same?