HW 1. Variables and Data; Types of Variables

1. Name each of the variables and state the experimental unit. State the type of each variable.

a) Seventeen soil samples are taken from Quadrants A, B, C and D of a drilling site. For each sample the following data is obtained and recorded: depth at which sample was drawn, mass of sample to the nearest gram, quadrant sample was drawn from, and primary element by mass.

b) The following data is recorded for cars entering a campus parking lot: make, model, number of occupants, number of seatbelts in vehicle, distance displayed on odometer, diameter of front left tire.

2. State the type of each variable:
a) Time to complete a math test
b) Country of birth
c) Total cost of textbooks for this term
d) Weight of student's wallet
e) Student's height, rounded to the nearest inch
f) Employer from current or last job
g) Number of college or university course credits achieved
h) Last three symbols of postal code

HW 2. Bar Charts, Histograms, Stem and Leaf Plots

1. After the 2011 Canadian election, the number of federal seats were as follows. Construct a relative frequency bar chart.

Party	Frequency
Conservative	166
NDP	103
Liberal	34
Bloc Quebecois	4
Green	1

2. Hourly wage at current or last job: $11.00, 12.10, 15.50, 14.00, 13.00, 13.75, 11.75, 24.25, 15.65, 18.80, 12.25, 19.25, 17.80, 16.25, 22.50, 21.20, 23.00, 22.10$

Construct a frequency histogram with 5 classes. Recall that the left endpoint belongs to the class and the right endpoint does not. Describe the shape of the histogram.

3. Draw a stem and leaf plot using the first 2 digits as stems:
6435, 6464, 7009, 6485, 6555, 6672, 6881, 6422, 6813, 6772, 6745, 6479, 6517, 6871, 7004, 6763, 6645, 6391. Describe the shape of the plot.

HW 3. Mean, Median and Mode

1. Find the mean, median and mode for the set of temperature readings below.

Temperature ($^\circ C$)	Frequency
36.8	4
37.1	6
37.2	2
37.4	5
37.7	8

2. Find the mean, median and mode for the set of yields (in grams) from a chemical reaction.

Yield(g)	Relative Frequency
157	0.25
142	0.13
135	0.24
162	0.38

3. Estimate the mean of the test scores below. Note that you cannot find the exact mean without further information.

Score	Frequency
50 – 60	3
60 – 70	11
70 – 80	22
80 – 90	6
90 – 100	4

4. Given the data set: $301, 334, 321, 317, 318, 322, 325, 333, 308, 334, 307, 311, 325, 334, 326, 339$

Use a stem and leaf plot to find:
a) the mean
b) the median
c) the mode
d) the 4th smallest measurement

HW 4. Range, SD and Variance

1. Calculate the population SD, population variance, sample SD and sample variance for: $1.1, 1.5, 1.5, 1.9, 2.1, 2.7$.

2. Calculate the range, population SD, population variance, sample SD and sample variance for:

X	frequency
108	4
137	13
159	8
162	5

3. Which class's test scores are more tightly clustered? Explain.

Class A: $62, 65, 68, 71, 72, 74, 78, 81$

Class B: $66, 75, 75, 75, 75, 89, 89$

4. How do the mean and population variance change when each measurement in the following data set is multiplied by 0.2?
$5, 6, 8, 9, 16$

5. If a data set has $\sigma^2 = 0$, what can you conclude about the data set?

HW 5. Tchebysheff and Empirical Rules

1. A data set has $n = 300$ measurements; $\mu = 47$ and $\sigma = 4$. What can you say about the number of measurements:
a) between 40 and 54?
b) between 30 and 60?
c) bigger than 55?

2. For the data set below:
a) Does the empirical rule apply?
b) What does it predict for the proportion of measurements within σ of the mean?
c) What is the actual proportion of measurements within σ of the mean?

X	frequency
13	7
15	11
18	23
20	12
22	6

3. At a drilling site, the times to reach a specific depth are measured. The mean time is 10 hours, with a standard deviation of 12 hours. Use Tchebysheff's Theorem to find an interval which is guaranteed to contain at least 75% of measurements.

HW 6. Percentiles and Box Plot

1. For the following data set find:
a) the 20th percentile
b) the 50th percentile
c) the 74th percentile

X	frequency
4	7
6	16
10	15
13	12

2. Use a stem and leaf plot to find the five-number summary for the following data set:

$28, 56, 45, 41, 64, 63, 29, 27, 54, 33, 38, 33, 25, 44, 41, 55, 58, 32, 69, 23$

3. Write the five-number summary and draw a box plot for the following data set: $3, 9, 10, 2, 6, 7, 5, 8, 6, 6, 4, 9, 22$

HW 7. Correlation and the Regression Line

1. International Travel to the USA
(Data from: 2011 World Almanac)

X=Year	Y=Visitors (millions)
1991	42.7
1994	44.8
1997	47.8
2000	50.9

a) Find the regression line
b) Find the correlation coefficient
c) Plug in each of the x-values to the regression line to convince yourself that the fit is quite good
d) What value is predicted for $x = 1992$?
e) What value is predicted for $x = 2001$? Why might this prediction be less reliable than the one in d) ?

2. The Danger of Insufficient Data
a) Find the correlation coefficient and the regression line for the data below.
b) Confirm that all the data points lie on the parabola $y = x^2 + 2$
c) Add a few more data points on the parabola: $(-15, 227), (-5, 27), (20, 402)$ and calculate r again.

X	Y
1	3
2	6
3	11

3. For the blood pressure data below:
a) Find the regression line. Round a and b to 2 decimal places.
b) What age leads to a predicted systolic blood pressure of 152?

X=Age	Y=Systolic Blood Pressure
37	124
41	135
31	138
65	141
63	162
57	141
52	138
55	165

HW 8. Intro to Probability

1. A die is rolled. Let:
A be the event that the roll is at most two
B be the event that the roll is odd
C be the event that the roll is a multiple of 3
Find:
a) $P(A)$
b) $P(B)$
c) $P(C)$
d) $P(B \text{ or } C)$
e) $P(A \text{ and } B)$

2. Two dice are rolled. What is the probability that the rolls sum to less than 5 or exactly 8?

3. Below is the make-up of employees at an engineering firm. Find:
a) $P(\text{female})$
b) $P(\text{male or contract})$
c) $P(\text{female and permanent})$

	Male	Female
Contract	37	41
Permanent	98	55

4. Out of 62 job applicants, 35 have their P.Eng. qualification and 23 are fluent in French. Of those who are fluent in French, 17 have their P.Eng. qualification. What is the probability that an applicant has their P.Eng. but does not speak French?

Background on Cards: A standard deck of cards consists of 52 cards, divided into four suits (hearts, diamonds, clubs and spades). Each suit has 13 cards: Ace, $2, \ldots, 10$, Jack, Queen, King.

5. A card is randomly selected from a standard deck. What is the probability of getting a heart or an 8?

HW 9. Combinations and Permutations

1. Seven cards are selected simultaneously (all at once) from a standard deck. How many different outcomes are possible?

2. Seven cards are dealt in sequence from a standard deck. How many different outcomes are possible?

3. A committee consists of a chair, a treasurer and a secretary. How many different committees can be formed from a pool of 20 employees?

4. A class consists of 23 female students and 18 male students. A team of 4 students is formed. What is the probability that the team contains at least one male student and at least one female student?

5. A fair coin is tossed 8 times. What is the probability that between 2 and 4 heads appear?

HW 10. Unions, Intersections and Complements

1. A security system uses two devices: A and B. In the event of a break-in, the probability that it will be detected by Device A is 98%, by Device B is 97%, and by both devices is 96.5%.
a) If a break-in occurs, find the probability that it will be detected by at least one device.
b) If a break-in occurs, find the probability that it will not be detected.

2. A coin is tossed 3 times. Let A be the event that the first toss is heads. Let B be the event that exactly 2 heads appear in the three tosses. Find:
a) $P(A)$
b) $P(B)$
c) $P(A \cap B)$
d) $P(A \cup B)$
e) $P(\overline{A})$

3. Three cards are selected simultaneously from a standard deck. Find:
a) the probability that there is at least one heart selected
b) the probability that not all the cards are hearts

4. In this problem we are interested in the 365 possible birthdays, disregarding the year and ignoring February 29. Let's assume that each birthday is equally likely.
a) In a group of 2 people, what is the probability that they have different birthdays?
b) In a group of 3 people, what is the probability that they all have different birthdays?
c) In a group of n people, what is the probability that they all have different birthdays?
d) In a group of n people, what is the probability that at least two of them share a birthday?

HW 11. Conditional Probability and Independence

1. At your firm, client presentations are prepared using either Powerpoint or Keynote. A presentation is considered successful if the client signs a contract with your firm.

	Successful	Unsuccessful
Powerpoint	0.26	0.20
Keynote	0.36	0.18

Find the probability that:
a) a presentation is done in Powerpoint
b) a presentation is successful
c) a presentation is done in Keynote, given that it is successful
d) a Powerpoint presentation is unsuccessful

2. Your company accepts projects from the City of Victoria as well as private firms; 20% of projects come from the City and 80% come from private firms. Of the City projects, 85% are completed on time. Of the private projects, 75% are completed on time. Find the probability that:
a) a project comes from the City and is completed on time
b) a project is completed on time

3. Given $P(A) = 0.55$ and $P(B) = 0.2$, find $P(A \cup B)$ if A and B are independent events.

4. A coin is tossed 3 times. Let E be the event that at most one head appears. Let F be the event that at least one head and at least one tail appear. Are E and F independent events?

HW 12. Bayes' Rule

1. Manufacturing employees follow protocol 98% of the time. When protocol is followed, 1.2% of manufactured items are defective. When protocol is not followed, 3.8% of manufactured items are defective. What is the probability that manufacturing protocol was followed, given that an item is defective?

2. At your company 1% of employees use a certain illicit substance. During the mandatory semi-annual drug testing, 99.9% of users test positive and 99.8% of non-users test negative.

a) How probable is it that a non-user will test positive?
b) How probable is it that a positive result belongs to a non-user?

A flu outbreak occurs at the company. It is known that 18% of employees are infected. All employees are tested; 99.5% of infected people test positive and 99.3% of healthy people test negative.

c) How probable is it that a healthy employee will test positive?
d) How probable is it that a positive result belongs to a healthy employee?

HW 13. Random Variables

1. At your company, the maximum number of promotions an employee can get each year is 2. Let X be the number of promotions an employee gets this year. You are given that $P(X = 1) = 0.4P(X = 0)$ and $P(X = 2) = 0.2P(X = 0)$.

a) Find the probability distribution of X.
b) Find the expected value of X.
c) Find the standard deviation of X.

2. Two cards are drawn simultaneously from a standard deck. Let X represent the number of hearts chosen. Find the probability distribution of X.

3. Your company is considering competing for projects A and B.

Your company has a 35% probability of success if it competes for Project A. The cost of competing is $\$20,000$. If successful in the competition your company will earn $\$1.2$ million from the project.

Your company has a 42% probability of success if it competes for Project B. The cost of competing is $\$15,000$. If successful in the competition your company will earn $\$900,000$ from the project.

a) Let X be the net earnings (earnings from the project$-$cost of competition). Find the probability distribution of X for each of the two projects.
b) Which project has higher expected net earnings?
c) Which project has a lower value for the standard deviation of X? A lower SD indicates less risk in terms of net earnings.

HW 14. Binomial Distribution

1. A multiple choice test has 20 questions, each of which has 3 possible answers. A student guesses randomly on each question.
a) What is the probability that the student gets exactly 6 questions right?
b) What is the probability that the student gets between 5 and 7 questions right?
c) What is the probability that the student gets at most 18 questions right?
d) Let X be the number of questions the students gets right. If you were to draw the histogram for the probability distribution of X, what shape would it have?
e) What assumption(s) do you need to check in part d)?

2. You select eight cards in sequence from a standard deck. What is the probability that you selected exactly three clubs?

3. In the 2011 Canadian Federal election, 18.9% of ballots were votes for the Liberal Party. In total, 14.6 million votes were cast. In a random poll, 200 voters were contacted.
a) What is the approximate probability that 38 of these voters voted for the Liberal Party?
b) What assumption(s) do you need to check for the approximation to be valid?

HW 15. Poisson Distribution

1. X is a Poisson random variable with a mean of 5. What is the probability that X is greater than 3?

2. A web server receives an average of 12 requests per hour. What is the probability that the server receives at most 2 requests in the next 15 minutes?

3. X is a Poisson random variable and you are given that $P(X = 0) = 0.6$. What is the probability that X is equal to 3?

HW 16. Continuous Distributions: The Uniform Distribution

1. X is a continuous random variable that is uniformly distributed on the interval $3 < X \leq 12$. Find:
a) the probability that $4 \leq X \leq 6.3$
b) the probability that $X \geq 5$
c) the probability that $X = 7.7$
d) the probability that $5 < X < 15$
e) the 75th percentile

2. X is a continuous random variable that is uniformly distributed on the interval $a \leq X \leq b$.
a) Find a formula for the pth percentile. Your formula will be terms of a, b and p.
b) Use your formula to find the 50th percentile (the median).

HW 17. Normal Distribution

1. A brand of chocolate bar has a weight that is normally distributed with a mean of 85 grams and a variance of 0.64. Find the probability that a chocolate bar weighs:
a) between 84 and 86.2 grams
b) less than 83.5 grams

2. A type of ball bearing has a mean diameter of 12 mm with a standard deviation of 0.34 mm. The diameters are normally distributed. Find:
a) the 90th percentile of diameters
b) the 20th percentile of diameters

3. X is normally distributed. You are given that $P(X > 20) = 0.4$ and $P(X > 30) = 0.25$. Find the mean and standard deviation of X.

HW 18. Approximating the Binomial Distribution

1. A fair die is rolled 500 times. Find the (approximate) probability that an odd number appears:
a) between 240 and 250 times
b) more than 275 times

2. A company regularly orders too many boxes of parts because some boxes get delayed in transit. On average, 30% of boxes are delayed. Each box is shipped independently of the others. The company orders 250 boxes, but only needs 190. What is the (approximate) probability that they meet the demand?

HW 19. Central Limit Theorem

1. At a certain emergency room the average amount of time patients spend with a doctor is 7.1 minutes. The standard deviation is 5.2 minutes. Sixty patient records are selected at random.
a) Describe the sampling distribution of the mean visit time.
b) What assumption(s) do you need to check?
c) Find the probability that mean visit time is between 6 and 8 minutes.

2. At a certain firm, employees worked a mean of 45.5 hours last week, with a standard deviation of 6 hours. Work hours were normally distributed. 16 employees are selected at random.
a) What is the probability that their work hours totalled more than 730 hours?
b) What assumption(s) do you need to check?

3. 80% of all Canadian HR employees check references before hiring an individual. 100 Canadian HR employees were polled on whether they checked references for their last hire.
a) Describe the sampling distribution of $\hat{p}$, the proportion of people polled who checked references.
b) What assumption(s) do you need to check?
c) What is the probability that less than 77% of them checked references?

4. A machine is filling cans of pop. The volume per can is normally distributed with a standard deviation of 1.9 mL. What should the volume be set to on the machine (this is μ) in order to ensure that in a random sample of 12 cans, there is a 99% probability that the mean is at least 355 mL?

HW 20. Large Sample Confidence Intervals

1. At a paper factory, the paper length has a standard deviation of $\sigma = 0.02$ inches. In a random sample of 100 sheets, the mean is found to be 10.998 inches. Find a 98% confidence interval for μ and check any necessary assumptions.

2. A random sample of 500 vehicles is selected from BC's vehicle database. Of these, 68 are classified as SUV's. Find a 95% confidence interval for p and check any necessary assumptions.

3. A study looks at the effect of training on time (in minutes) to perform a simple task. Find a 99% confidence interval for $\mu_1 - \mu_2$ and check any necessary assumptions. What can you conclude?

Group 1 (After One Hour of Training):	$n = 30$	$\overline{x} = 15$	$s^2 = 16$
Group 2 (No Previous Training):	$n = 40$	$\overline{x} = 17$	$s^2 = 100$

4. A study looks at post-secondary completion in two different age brackets. Let x represent the number of people polled who earned a post-secondary qualification. Find a 90% confidence for $p_1 - p_2$ and check any necessary assumptions. What can you conclude?

Group 1 (Age 30-40):	$n = 180$	$x = 126$
Group 2 (Age 50-60):	$n = 110$	$x = 54$

5. What does the phrase "95% confidence" mean in the context of a 95% confidence interval for μ?

HW 21. One-Sided Confidence Bounds; Sample Size

1. Find a 95% upper confidence bound (UCB) and a 99% lower confidence bound (LCB) for μ given:

40 water samples in the inner harbour yielded a mean nitrate ion concentration of 25 ppm. The standard deviation was 5 ppm.

2. Paper length has $\sigma = 0.02$ inches. How large must a sample size be for a 98% confidence interval for μ of length less than 0.003?

3. Given $\hat{p} = 0.136$, how large must n be to give a 99% confidence interval for p of length less than 0.2?

4. Without any estimate for p, what is the minimum sample size you should take if you want a 95% margin of error for p of less than 0.04?

HW 22. Large Sample Hypothesis Tests

1. A certain material is required to have a mean breaking strength of exactly 70 pounds. A sample of 49 pieces reveals $\overline{x} = 69.1$ pounds and a SD of 3.5 pounds. Is there evidence that the requirements are not being met? Test with $\alpha = 0.01$

2. Is there evidence of a difference in mean fuel efficiencies (miles per gallon)? Test with $\alpha = 0.05$

Prius (Group 1): $n = 40$ $\quad \overline{x} = 59$ $\quad s = 2$
Smart (Group 2): $n = 40$ $\quad \overline{x} = 60$ $\quad s = 2$

3. At a ball bearing factory a sample of 1200 ball bearings reveals that 6 are defective. Is this evidence that the proportion of defective ball bearings is different than 0.006? Test with $\alpha = 0.05$

4. Are the proportions of hospital admissions for heart-related ailments the same for males and females? Test with $\alpha = 0.01$

Males (Group 1): $n = 1000$ $\quad \hat{p} = 0.052$
Females (Group 1): $n = 1000$ $\quad \hat{p} = 0.023$

5. A bottling company wants to test whether the mean fill volume is less than 355 mL. A sample of 32 cans reveals $\overline{x} = 354.4$ mL and $s = 1.9$ mL. Test with $\alpha = 0.02$

6. A shipping company wants to test whether the proportion of shipped items that are lost in transit is more than 0.004. A sample of 1800 records showed that 17 were lost in transit. Test with $\alpha = 0.05$

HW 23. Type I and Type II Error

1. a) Define a Type I error.
b) What is the probability of making a Type I error?

2. Define a Type II error.

3. You are testing the hypothesis H_0: $\mu = 102$ with $\alpha = 0.05$. You are given $\overline{x} = 103, s = 16$ and $n = 42$. What is the probability of not rejecting H_0 if the true value of μ is 103.8?

HW 24. Small Sample Inferences for the Mean

1. You are given the following measurements taken from a normally distributed population: 8.09 9.17 7.74 6.88 6.91 7.34 8.52 8.31

a) Find a 90% confidence interval for the population mean
b) Find a 90% upper confidence bound (UCB) for the population mean

2. A bottling company wants to test whether the mean fill volume is less than 355 mL. The filled volume is normally distributed. Here is a sample:
355.1 354.8 354.7 354.4 355.2 355.0 354.9

a) Test the hypothesis with $\alpha = 0.10$
b) What can you say about the p-value of the test?

3. A company claims that the average weight of their cereal boxes is 600 grams. The weights are normally distributed. Here is a sample:
598 597 589 601 588 602 581 585 593

a) Test the hypothesis with $\alpha = 0.05$
b) What can you say about the p-value of the test?

HW 25. Difference of Means in Small Samples

Earth's temperature (in degrees F) is measured from the air and from the ground on a single day.

Location	Air	Ground
1	47.3	46.9
2	48.1	45.4
3	37.9	36.3
4	32.7	31.0
5	26.2	24.7

1. Find a 95% confidence interval for $\mu_1 - \mu_2$ based on the data above. List all assumptions, and check them if possible.

2. Using the data above, test whether the mean air temperature is higher than the mean ground temperature with $\alpha = 0.05$. List all assumptions, and check them if possible.

3. 100 m freestyle swimming times (in seconds) were recorded for two competitors in 20 different races. At $\alpha = 0.05$ test whether their mean times are different. List all assumptions, and check them if possible.

Swimmer 1: $n = 10$ $\quad \overline{x} = 59.646$ $\quad s^2 = 0.027$
Swimmer 2: $n = 10$ $\quad \overline{x} = 59.627$ $\quad s^2 = 0.035$

HW 26. Inferences about σ^2

1. Test whether $\sigma^2 = 100$ at $\alpha = 0.02$ based on the following sample from a normal population: 1290 1302 1325 1281 1305

2. Find a 90% confidence interval for σ based on a sample of 9 measurements from a normal population: $\overline{x} = 188.112$ and $s = 7.201$.

3. Find a 95% upper confidence bound (UCB) for σ^2 based on a sample of 16 measurements from a normal population: $\overline{x} = 510.101$ and $s = 12.355$.

HW 27. Goodness-of-Fit Tests

1. The astrological signs of 264 Fortune 500 Executives appear below. At $\alpha = 0.05$ test whether astrological signs are equally represented among the executives.

Ar.	Taur.	Gem.	Can.	Leo	Vir.	Libra	Scor.	Sag.	Cap.	Aqua.	Pisces
23	24	18	23	24	19	18	21	19	22	24	29

2. Last year a certain university admitted 15% of applicants unconditionally, admitted 32% of applicants conditionally, and refused 53% of applicants. A sample of applications this year shows that 50 were admitted unconditionally, 167 were admitted conditionally and 83 were refused. Test at $\alpha = 0.05$ to see whether the proportions have changed from last year.

3. A coin is flipped 200 times, and exactly h heads appear. If we test at $\alpha = 0.01$, what values of h determine an unfair coin?