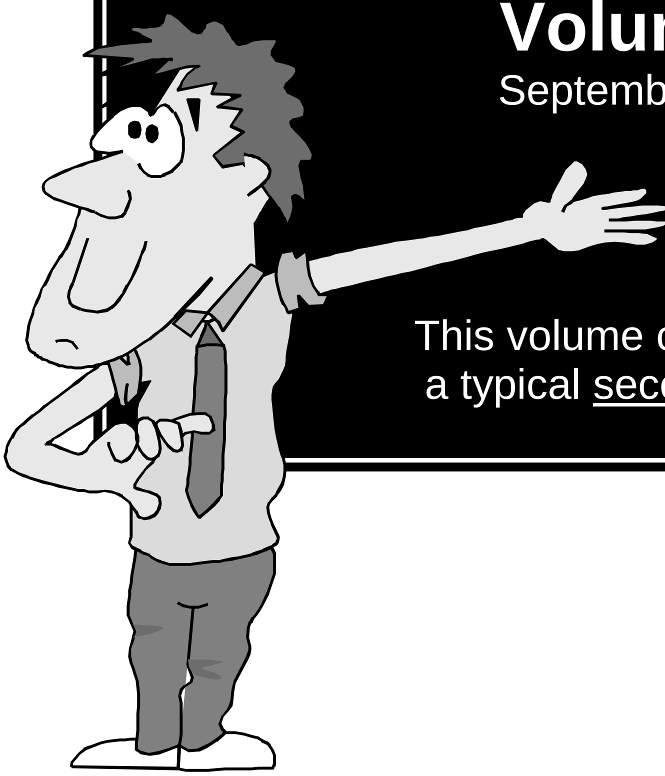
Grant's Tutoring

BASIC STATISTICS 1

Volume 2 of 3

September 2011 edition



This volume covers the topics on a typical <u>second midterm</u> exam.

Learn What You Need to Know

Know What You Need to Learn

While studying this book, why not hear Grant explain it to you? Contact Grant for info about purchasing **Grant's Audio Lectures**. Some concepts make better sense when you hear them explained.

Better still, see Grant explain the key concepts in person. Sign up for **Grant's Weekly Tutoring** or attend **Grant's Exam Prep Seminars.** Text or Grant (204) 489-2884 or go to **www.grantstutoring.com** to find out more about all of Grant's services. **Seminar Dates will be finalized no later than Sep. 25 for first term and Jan. 25 for second term.**

HOW TO USE THIS BOOK

I have broken the course up into lessons. Do note that the numbering of my lessons do not necessarily correspond to the numbering of the units in your course outline. Study each lesson until you can do all of my lecture problems from start to finish without any help. Then do the Practise Problems for that lesson. If you are able to solve all the Practise Problems I have given you, then you should have nothing to fear about your exams.

I also recommend you purchase the *Multiple-Choice Problems Set for Basic Statistical Analysis I (Stat 1000)* by Dr. Smiley Cheng available at The Book Store. The appendices of my book include complete step-by-step solutions for all the problems and exams in Cheng's book. Be sure to read the "Homework" section at the end of each lesson for important guidance on how to proceed in your studying.

You also need a good, but not expensive, scientific calculator. Any of the makes and models of calculators I discuss in Appendix A are adequate for this course. I give you more advice about calculators at the start of Lesson 1. **Appendix A in this book shows you how to use all major models of calculators.**

I have presented the course in what I consider to be the most logical order. Although my books are designed to follow the course syllabus, it is possible your prof will teach the course in a different order or omit a topic. It is also possible he/she will introduce a topic I do not cover. **Make sure you are attending your class regularly! Stay current with the material, and be aware of what topics are on your exam. Never forget, it is your prof that decides what will be on the exam, so pay attention.**

If you have any questions or difficulties while studying this book, or if you believe you have found a mistake, do not hesitate to contact me. My phone number and website are noted at the bottom of every page in this book. "Grant's Tutoring" is also in the phone book. **I welcome your input and questions.**

Wishing you much success,

Grant Skene

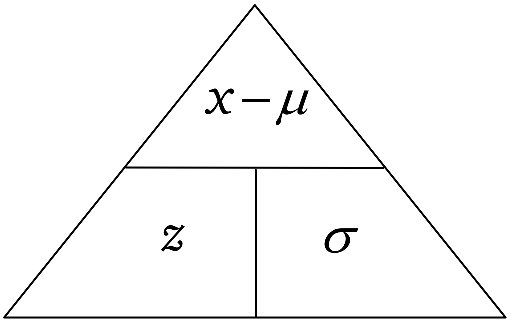
Owner of Grant's Tutoring and author of this book

SUMMARY OF KEY CONCEPTS IN LESSON 4

- Populations have **parameters**; samples have **statistics**.
- Know when to use which symbol.
  - A sample mean is $\bar{x}$ whereas a population mean is μ.
  - A sample standard deviation is s whereas a population standard deviation is σ.
  - When you are reading a sentence, if n, the sample size, is given, you are being given statistics. If there is no n in the sentence, you are being given parameters.
- Continuous quantitative populations are displayed by **density curves**.
  - Density curves cannot dip below the horizontal axis.
  - The total area under a density curve is 1 (100%).
- The **uniform distribution** has a rectangular density curve. The height of this rectangle is always the reciprocal of the rectangle's width (*height* = 1/*width*).
- An extremely important density curve is the bell curve displaying the **normal distribution**.
- If given or asked for any one of the **three Ps (proportion, percentage or probability)** for any density curve (a uniform distribution, a normal distribution, or whatever) you are being given or asked for the **area of a region** on that curve.
- Normal distributions follow **the 68-95-99.7 rule**.
  - 68% of the population is within 1 standard deviation of the mean.
  - 95% of the population is within 2 standard deviations of the mean.
  - 99.7% of the population is within 3 standard deviations of the mean.
- The **standard normal distribution, Z, has $\mu = 0$ and $\sigma = 1$: $Z \sim N(0, 1)$.**
  - A standardized z-score tells you how many standard deviations you are above or below the mean of a normal distribution.
- If a problem gives you a "standard normal distribution", you have a ***Z*-bell curve**; if a problem gives you simply a "normal distribution", you have an ***X*-bell curve**.

- Memorize the standardizing formula: $z = \dfrac{x-\mu}{\sigma}$.

- Perhaps, make a triangle like this to help you algebraically manipulate the standardizing formula:



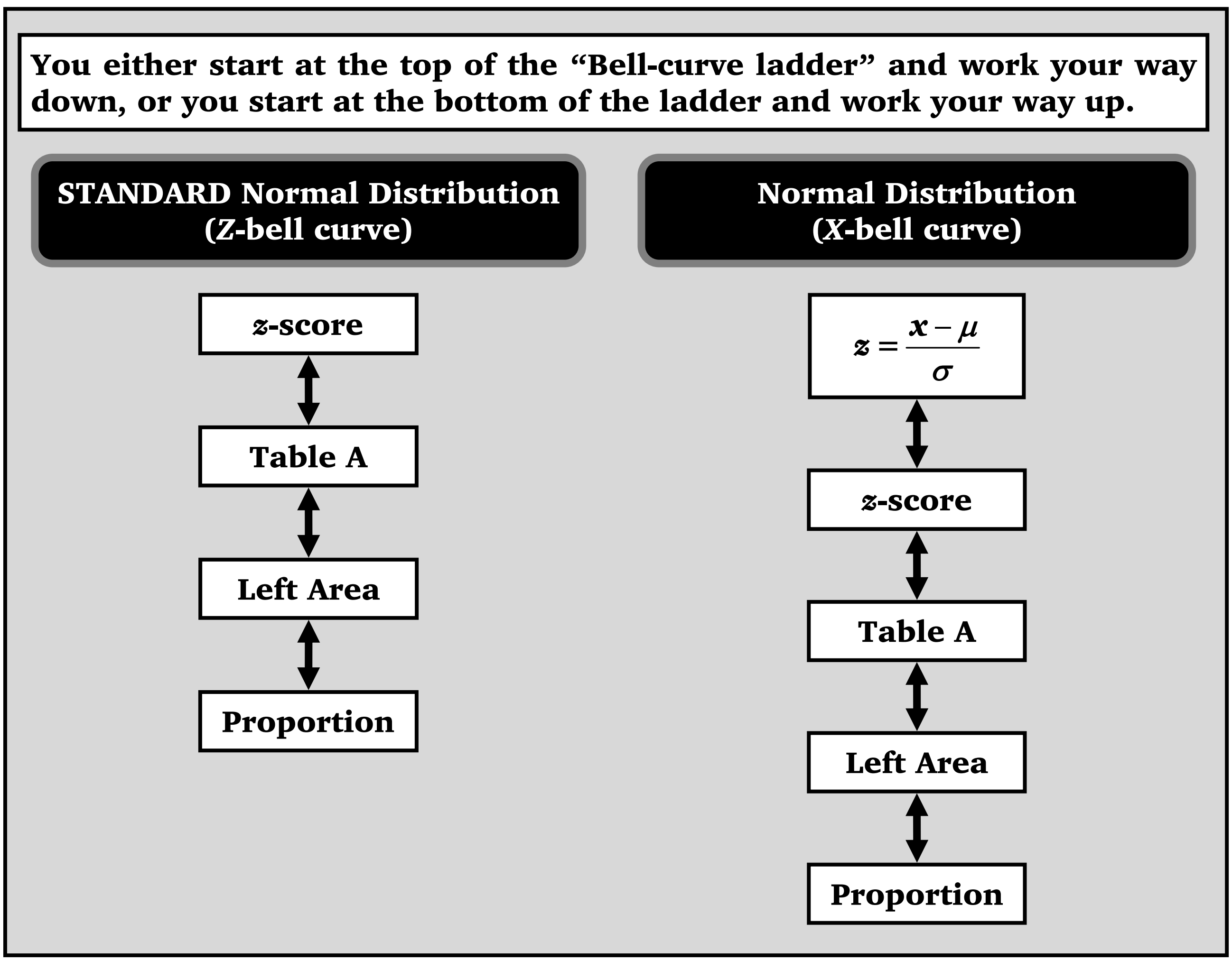
From this triangle we can see:

(1) $z = \dfrac{x-\mu}{\sigma}$ (of course)

(2) $\sigma = \dfrac{x-\mu}{z}$

(3) $x-\mu = z\sigma$

- All bell curve problems can be mapped out like this:



LECTURE PROBLEMS FOR LESSON 4

For your convenience, here are the 18 questions I used as examples in this lesson. Do not make any marks or notes on these questions below. Especially, do not circle the correct choice in the multiple choice questions. You want to keep these questions untouched, so that you can look back at them without any hints. Instead, make any necessary notes, highlights, etc. in the lecture part above.

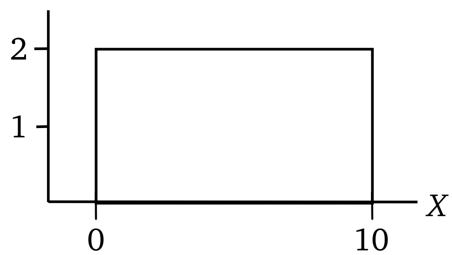
1. Each graph in (a) through (f) below depicts the distribution of a continuous quantitative random variable X. For each graph answer the following questions.

(i) Is the graph a properly defined density curve? Justify your answer.

(ii) If it is a properly defined density curve, what is the median of X?

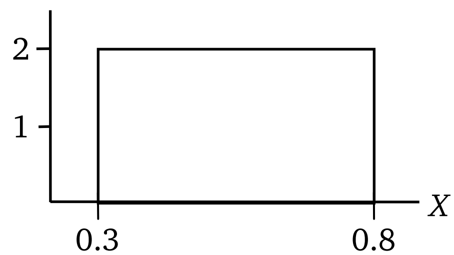
For your information, the area of a circle is πr^2 where r = the radius of the circle, and the area of a triangle is $\frac{1}{2}bh$ where b = the base length and h = the height of the triangle.

(a)



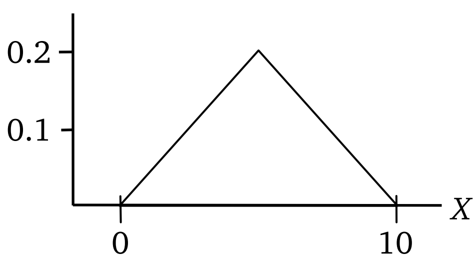
(Solution en page 200.)

(b)



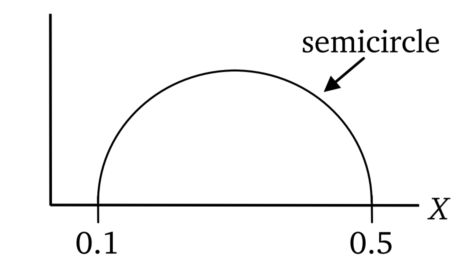
(Solution en page 201.)

(c)



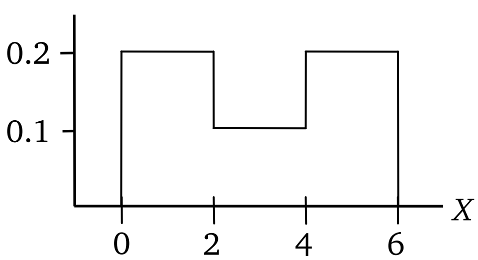
(Solution en page 201.)

(d)



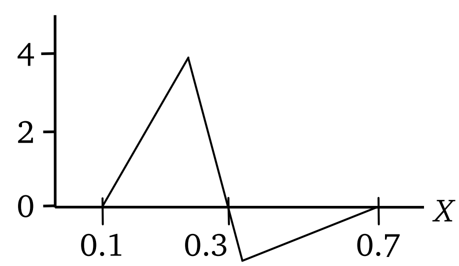
(Solution en page 202.)

(e)



(Solution en page 203.)

(f)



(Solution en page 203.)

2. The random variable X has a uniform distribution where $-2 \le X \le 6$. Determine the following proportions:

(a) $P(X \ge 4)$
(Solution en page 207.)

(b) $P(X < -1)$
(Solution en page 207.)

(c) $P(2 < X \le 5)$
(Solution en page 208.)

(d) $P(X > 1.2)$
(Solution en page 208.)

(e) $P(X = 3)$
(Solution en page 209.)

(f) $P(X > 7)$
(Solution en page 209.)

3. Assume the marks on a test are normally distributed with a mean of 64% and a standard deviation of 9%. Which of these statements are false?

(A) 68% of the class received a mark between 55% and 73%.
(B) A mark of at least 82% would be in the top 2.5% of the class.
(C) 84% of the class scored less than 73%.
(D) 20% of the class scored less than 46%.
(E) 81.5% of the class scored between 55% and 82%.

(See the solution on page 216.)

4. The amount of calories in a chocolate bar are normally distributed with an average of 250 calories. If 99.7% of all the bars have between 205 and 295 calories, then the standard deviation (in calories) is

(A) 45 **(B)** 20 **(C)** 10 **(D)** 90 **(E)** 15

(See the solution on page 217.)

5. For the standard normal distribution, find the proportion of observations that satisfy the following:

(a) $P(Z < -1.25)$ *(See the solution on page 221.)*
(b) $P(Z \le 2.37)$ *(See the solution on page 222.)*
(c) $P(Z \ge -3.16)$ *(See the solution on page 222.)*
(d) $P(Z > 0.48)$ *(See the solution on page 223.)*
(e) $P(-2.36 < Z \le -1.62)$ *(See the solution on page 224.)*
(f) $P(-2.93 < Z < 0.6)$ *(See the solution on page 225.)*
(g) $P(Z \le 4.02)$ *(See the solution on page 225.)*
(h) $P(-3.92 < Z < 1.83)$ *(See the solution on page 226.)*
(i) $P(Z = 1.25)$ *(See the solution on page 226.)*
(j) $P(|Z| < 1.65)$ *(See the solution on page 228.)*
(k) $P(|Z| > 2.12)$ *(See the solution on page 229.)*

6. For the standard normal distribution, find the z-score that satisfies the following proportions:

(a) $P(Z < z) = .1446$ *(See the solution on page 230.)*

(b) $P(Z \leq z) = .9871$ *(See the solution on page 231.)*

(c) $P(Z \geq z) = .0078$ *(See the solution on page 232.)*

(d) $P(Z > z) = .9767$ *(See the solution on page 233.)*

(e) $P(-z < Z < z) = .8324$ *(See the solution on page 234.)*

(f) $P(|Z| < z) = .7540$ *(See the solution on page 234.)*

(g) $P(|Z| > z) = .0060$ *(See the solution on page 235.)*

(h) $P(Z > z) = 15\%$ *(See the solution on page 236.)*

(i) $P(Z < z) = 10\%$ *(See the solution on page 236.)*

(j) $P(-0.12 < Z < z) = .4770$ *(See the solution on page 237.)*

7. List the first, second and third quartiles for the standard normal distribution.

(See the solution on page 238.)

8. The first and the eightieth percentiles for the standard normal distribution are, respectively:

(A) –2.33; 0.84 **(B)** –0.84; 2.33 **(C)** –2.33; –0.84 **(D)** 0.84; 2.33 **(E)** –2.3; 0.80

(See the solution on page 239.)

9. A vending machine is regulated to discharge an average of 7 oz. of coffee into an 8 oz. cup. If the amount of coffee dispensed is normally distributed with $\sigma = 0.3$ oz., what proportion of cups will overflow?

(A) 0.0004 **(B)** 3.33 **(C)** 0.4996 **(D)** 0.9996 **(E)** 0.0333

(See the solution on page 244.)

10. The heights of adult males are approximately normally distributed with mean 68 inches and standard deviation 2.5 inches. The percentage of this group who are taller than 5 feet is:

(A) 99.9% **(B)** 94.5% **(C)** 5.5% **(D)** 0.07% **(E)** 0

(See the solution on page 246.)

11. The average mark for a Basic Statistics class was 58.2% with a standard deviation of 6.4%. Assuming a normal distribution, the professor has decided that anyone in at least the fortieth percentile will be given a grade of C. What is the lowest mark that will receive a C?

(A) 40% **(B)** 50% **(C)** 56.6% **(D)** 57.6% **(E)** 59.8%

(See the solution on page 252.)

12. A graduate school program in English will admit only students with GRE (Graduate Record Examination) verbal ability scores in the top 30%. What is the lowest GRE score they will accept? Suppose the mean GRE score is 497 with standard deviation of 115 and that the distribution is normal.

(A) 437.2 **(B)** 616.6 **(C)** 556.8 **(D)** 612 **(E)** 382

(See the solution on page 254.)

13. An electrical firm manufactures light bulbs that have a lifetime that is normally distributed with $\mu = 800$ hours and $\sigma = 40$ hours. Of 300 bulbs, about how many will have lifetimes between 778 and 834 hours?

(A) 158 **(B)** 142 **(C)** 0.511 **(D)** 153 **(E)** 147

(See the solution on page 255.)

14. Bobby took the Scholastic Aptitude Test (SAT) and scored 1080. Kathy took the American College Test (ACT) and scored 28. It is known from past performance that the mean and standard deviation for SAT scores are 896 and 174, respectively. In addition, the mean for the ACT scores is 20.6 with a standard deviation of 5.2. The two distributions are bell-shaped. Which of the following statements is FALSE?

(A) Bobby's score is 1.057 standard deviations away from the mean SAT score.
(B) Kathy did relatively better than Bobby did.
(C) We can't compare scores from two different tests.
(D) Both students have scores that lie in the upper 16% of their respective distributions.
(E) Both students have scores that are at least in the eightieth percentile of their respective distributions.

(See the solution on page 257.)

15. A lathe produces washers whose internal diameters are normally distributed with mean equal to 0.373 inch and standard deviation of 0.002 inch. If specifications require the internal diameters be within 0.004 of 0.375 inch, what percentage of production will be unacceptable?

(A) 37.9% **(B)** 16.0% **(C)** 37.1% **(D)** 84.0% **(E)** 62.9%

(See the solution on page 258.)

16. A food processor packages instant coffee in small jars. The weights of the jars are normally distributed with a standard deviation of 0.3 ounces. If 5% of the jars weigh more than 12.492 ounces, what is the mean weight of the jars?

(A) 11.5 **(B)** 11.8 **(C)** 12 **(D)** 12.2 **(E)** 13

(See the solution on page 259.)

17. The contents of a soft drink are normally distributed with a mean of 450 ml. The first quartile is known to be 442 ml. What is the standard deviation?

(A) 10.5 ml **(B)** 11.9 ml **(C)** 13.2 ml **(D)** 32.1 ml **(E)** 4.3 ml

(See the solution on page 260.)

18. Human pregnancies last an average of 266 days with a standard deviation of 16 days. The distribution of pregnancies is approximately normal. What is the interquartile range?

(A) 21 days **(B)** 21.8 days **(C)** 22.8 days **(D)** 24.8 days **(E)** 25 days

(See the solution on page 262.)

SUMMARY OF KEY CONCEPTS IN LESSON 5

- The entire list of possible outcomes is called the **sample space**.
- All **probability models** must satisfy two conditions:
  - **The probability of any specific outcome or event must be somewhere between 0 and 1, inclusive (between 0 and 100%).** Put another way, for any event A, $\mathbf{0 \le P(A) \le 1}$.
  - **All possible outcomes together must have a probability of exactly 1 (or 100%).** Which is to say, since you have listed all the outcomes in the sample space, there is a 100% guarantee that one of these outcomes will occur (since it is impossible for anything else to occur). Put another way, for all sample spaces S, $\mathbf{P(S) = 1}$.
- Let X be a random variable from a discrete probability distribution, then:
  - The sum of all the probabilities is exactly 1: $\sum p(x) = 1$
  - The mean of X is $\mu = \sum x\, p(x)$.*
  - The variance of X is $\sigma^2 = \left(\sum x^2 p(x)\right) - \mu^2$.*
  - The standard deviation of X is $\sigma = \sqrt{\left(\sum x^2 p(x)\right) - \mu^2}$.*
- Some useful probability formulas to memorize:
  - **<u>The Complement Rule</u>:** $P(A^C) = 1 - P(A)$
  - **<u>The General Addition Rule</u>:** $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  - **<u>"Neither/nor" is 1 minus "or"</u>:** $P(\text{neither } A \text{ nor } B) = 1 - P(A \text{ or } B)$
  - **<u>The General Multiplication Rule</u>:** $P(A \text{ and } B) = P(A) \times P(B \mid A)$
  - **<u>If *A* and *B* are disjoint</u>, then** $P(A \text{ and } B) = 0$.
  - **<u>If (and only if) *A* and *B* are independent</u>, then** $P(A \text{ and } B) = P(A) \times P(B)$.
- If A and B are disjoint (mutually exclusive), they cannot be independent. If A and B are independent, they cannot be disjoint.

* The formulas for the mean, variance and standard deviation of a discrete random variable X are usually not taught in the Stat 1000 class. Check with your prof to see if you need to memorize them or not.

- Every time you toss a coin, every time you roll a die, every time you talk to a person, you are conducting a **trial**. **If you are using random sampling, you can generally assume each trial is independent.**
  - **Sampling <u>with</u> replacement** guarantees independent trials.
  - **Sampling <u>without</u> replacement** means each trial depends on the outcome of previous trials.
- The safest and most versatile way to solve probability problems is to first determine the sample space. That way you can actually see all the outcomes that fit the event in question. If more than one outcome fits, simply add the probabilities of the relevant outcomes together.
  - I think the best way to list a sample space is to make one or more **two-way tables**. The key to making a two-way table is to list the possible outcomes of your first trial along the top and the possible outcomes of your second trial down the side. Then, if necessary, the outcomes you have established for these first two trials act as the columns of a second two-way table where you list the possible outcomes of the third trial down the side. You continue in this vein for each additional trial until you have arrived at the entire sample space.
  - **If you know each outcome is equally likely**, then the probability of any event is simply a matter of counting all the outcomes that fit and dividing by the total number of possible outcomes in the sample space.
  - **If you know some outcomes are more likely than others**, use the multiplication rule to work out the probability of any particular outcome. You can do this because you are assuming each trial is independent.
    - For example, $P(AB) = P(A) \times P(B)$
    - For example, $P(ABC) = P(A) \times P(B) \times P(C)$
  - **Especially if you are given several questions to solve in a problem, you might consider listing the entire probability distribution first.** Which is to say, after having used two-way tables to determine the entire sample space, make a table where you list each outcome and the probability of each outcome, using the multiplication rule to determine each probability. That way you can actually see all the outcomes that fit the event in question. If more than one outcome fits, simply add the probabilities of the relevant outcomes together.
- **If you are given *P*(*A* and *B*)**, or if you are told ***A* and *B* are independent**, or if you are told ***A* and *B* are disjoint**, you can construct a **Venn diagram** to help you solve a probability problem.

LECTURE PROBLEMS FOR LESSON 5

For your convenience, here are the 18 questions I used as examples in this lesson. Do not make any marks or notes on these questions below. Especially, do not circle the correct choice in the multiple choice questions. You want to keep these questions untouched, so that you can look back at them without any hints. Instead, make any necessary notes, highlights, etc. in the lecture part above.

1. Below is a table showing the various prices a particular model of running shoe was sold for at a local sporting goods store. First the shoe was priced at $150, but then it was gradually reduced in price until the remaining stock was finally sold off at a sale price of $15. The table also includes the proportion of the stock that was sold at each price.

Price of the Running Shoe	$150	$125	$100	$75	$50	$15
Proportion Sold	0.05	0.16	k	$2k$	0.19	k

(a) What is the value of k if this is a properly defined probability distribution?

(See the solution on page 280.)

(b) What is the probability that a randomly selected purchaser did not pay full price ($150) for their running shoes?

(See the solution on page 280.)

(c) What is the probability that a randomly selected purchaser paid less than $100 for their running shoes?

(See the solution on page 281.)

(d) What was the average price paid for the running shoes?

(See the solution on page 282.)

(e) What was the standard deviation of the price paid for the running shoes?

(See the solution on page 282.)

2. For a specific population, $P(A) = 0.3$ and $P(B) = 0.5$.

(a) If A and B are disjoint, find the following probabilities:

(i) $P(A \text{ and } B)$ **(ii)** $P(A \text{ or } B)$.

(See the solution on page 292.)

(b) If A and B are independent, find the following probabilities:

(i) $P(A \text{ and } B)$ **(ii)** $P(A \text{ or } B)$.

(See the solution on page 292.)

3. Let us assume a woman is equally likely to give birth to a boy or girl and that each birth in a family is independent. Consider families that have exactly 3 children.

(a) List the sample space. *(See the solution on page 305.)*

(b) List the complete probability distribution. *(See the solution on page 306.)*

(c) What is the probability all three children in a family are the same sex?
(See the solution on page 306.)

(d) What is the probability at least two of the children are girls? *(Solution on page 307.)*

(e) What is the probability the youngest child is a boy? *(Solution on page 307.)*

(f) Let X be the number of boys in the family. Give the complete probability distribution of the discrete random variable X. *(Solution on page 308.)*

4. Let us assume, due to chemical contamination on an island in the North Atlantic, a woman on that island has a 70% chance of giving birth to a girl. Each birth in a family is independent. Consider families that have exactly 3 children.

(a) List the sample space. *(See the solution on page 309.)*

(b) List the complete probability distribution. *(See the solution on page 310.)*

(c) What is the probability all three children in a family are the same sex?
(See the solution on page 310.)

(d) What is the probability at least two of the children are girls? *(Solution on page 310.)*

(e) What is the probability the youngest child is a boy? *(Solution on page 311.)*

(f) Let X be the number of boys in the family. Give the complete probability distribution of the discrete random variable X. *(Solution on page 312.)*

5. Alice and Bob have been independently studying for their final exam. Based on their past performance, their professor estimates that Alice has an 80% chance of getting an A on the exam, while Bob has a 45% chance.

(a) What is the probability they both get an A on the exam?
(A) 0.35 **(B)** 0.53 **(C)** 0.47 **(D)** 0.64 **(E)** 0.36

(b) What is the probability only one of them gets an A on the exam?
(A) 0.35 **(B)** 0.53 **(C)** 0.47 **(D)** 0.64 **(E)** 0.36

(See the solutions on page 315.)

6. In Probomania, it has been found, among male voters, 45% vote Conservative, 30% vote Liberal and the rest vote Other. Among female voters, 35% vote Conservative, 50% vote Liberal and the rest vote Other. Assuming how a person decides to vote is independent of their spouse's decision:

(a) What is the probability a husband and wife both vote the same way?

(b) What is the probability at least one votes Liberal?

(See the solutions on page 317.)

7. A machine has 3 vital parts (*X*, *Y* and *Z*) which are independent of each other. If any part fails the machine will not operate. The probability *X* fails is 0.05, the probability *Y* fails is 0.10, and the probability *Z* fails is 0.08. What is the probability the machine operates?
(A) 0.0004 **(B)** 0.9996 **(C)** 0.7866 **(D)** 0.2134 **(E)** none of the above
(See the solution on page 318.)

8. A strand of Christmas tree lights has 20 bulbs. Each bulb has a 2% chance of burning out. If one bulb burns out, the strand will not light up at all. Assuming the bulbs burn out independently of each other, what is the probability the strand will not light?
(A) 0.0000 **(B)** 0.0200 **(C)** 0.4000 **(D)** 0.6676 **(E)** none of the above
(See the solution on page 320.)

9. In rolling a standard pair of fair dice, define the following events:

A: rolling a double (i.e. the same number on each die)
B: rolling a sum of nine (i.e. the two numbers add to 9)
C: rolling a 6 on the first die
D: rolling an odd sum
E: rolling an even sum
F: rolling a product of six (i.e. the two numbers multiply to 6)

(a) Find *P*(*A*), *P*(*B*), *P*(*C*), *P*(*D*), *P*(*E*), and *P*(*F*).
(See the solution on page 323.)
(b) Interpret *P*(*B*) so that a layman might understand.
(See the solution on page 324.)
(c) Find *P*(*B* and *D*). Are *B* and *D* independent? Explain.
(See the solution on page 324.)
(d) Find *P*(*B* or *D*).
(See the solution on page 326.)
(e) Are events *A* and *B* independent, disjoint, or neither? Explain.
(See the solution on page 326.)
(f) Are events *A* and *C* independent, disjoint, or neither? Explain.
(See the solution on page 327.)
(g) Which of events *A* through *F* are complements of each other? Explain.
(See the solution on page 328.)
(h) Which of events *B* through *F* are disjoint with *A*? Explain.
(See the solution on page 329.)

10. There are four World Cup Soccer quarterfinal games on the schedule. The games, as well as some selected probabilities of winning (in parentheses) are given below. Each game must have a winner; there can be no ties. Note that European teams are in bold.

Game 1: Argentina (0.6) vs. **Germany** (??)
Game 2: **France** (??) vs. Brazil (0.8)
Game 3: South Korea (??) vs. **Italy** (0.9)
Game 4: **England** (0.3) vs. **Portugal** (??)

(a) Let X = the winners of the four games. List the sample space of X.
(See the solution on page 331.)

(b) Attach probabilities to each outcome in the sample space.
(See the solution on page 332.)

(c) What is the probability both Argentina and Brazil win their games?
(See the solution on page 332.)

(d) What is the probability either France or South Korea win?
(See the solution on page 333.)

(e) What is the probability a European team wins every game?
(See the solution on page 334.)

(f) What is the probability a European team wins no more than one game?
(See the solution on page 334.)

11. In the game "Rock, Paper, Scissors", players simultaneously and independently display one of the three symbols with their hand. Three friends are playing one round of the game together. We will assume that each player selects each of the three symbols with equal probability. In this game, Rock beats Scissors, Paper beats Rock, and Scissors beats Paper.

(a) List the complete sample space of possible outcomes from one round of play.
(See the solution on page 335.)

(b) What is the probability only one player wins?
(See the solution on page 336.)

(c) Find the probability of each of the following events:

A = {first player selects Scissors}
B = {all three players select the same symbol}
C = {exactly two players select Rock}

(See the solution on page 336.)

(d) Find $P(A \text{ and } B)$, $P(A \text{ and } C)$ and $P(B \text{ and } C)$. Determine whether each pair of events is mutually exclusive, independent, or neither.
(See the solution on page 337.)

12. A slot machine has three reels that spin independently. Each reel has 10 equally likely symbols, as shown below:

Reel 1: 3 cherries, 4 oranges, 3 bells
Reel 2: 5 cherries, 3 oranges, 2 bells
Reel 3: 6 cherries, 3 oranges, 1 bell

(a) The outcome of interest is the set of three symbols showing on the three reels on any spin of the slot machine. What is the sample space for this experiment?

(See the solution on page 339.)

(b) What is the probability you win the jackpot (three bells)?

(See the solution on page 340.)

(c) Interpret the probability you calculated in (b) to someone with little or no background in statistics.

(See the solution on page 340.)

(d) What is the probability of all three reels showing a fruit?

(See the solution on page 340.)

13. Oscar has 12 socks in a drawer. They are all of the same style but of different colours, and they have not been paired up. Six of the socks are blue, four of the socks are white, and two of the socks are green. Assuming it is pitch black in the room, and he just reaches into the drawer and pulls two socks out randomly:

(a) What is the probability at least one of the socks is blue?

(See the solution on page 342.)

(b) What is the probability both socks are the same colour?

(See the solution on page 342.)

14. Given that events A and B are independent where $P(A) = .32$ and $P(B) = .17$, find the following probabilities:

(a) $P(A \text{ or } B)$

(See the solution on page 349.)

(b) $P(A \text{ and } B^C)$

(See the solution on page 349.)

(c) $P(A^C \text{ or } B)$

(See the solution on page 350.)

(d) $P(A^C \text{ and } B^C)$

(See the solution on page 350.)

15. Given that events A and B are mutually exclusive where $P(A) = .23$ and $P(B) = .51$, find the following probabilities:

(a) $P(A \text{ or } B)$

(See the solution on page 351.)

(b) $P(A \text{ or } B^C)$

(See the solution on page 352.)

(c) $P(A^C \text{ and } B)$

(See the solution on page 352.)

(d) $P(A^C \text{ and } B^C)$

(See the solution on page 353.)

(e) $P(A^C \text{ or } B^C)$

(See the solution on page 353.)

(f) $P(A \text{ and } B)$

(See the solution on page 353.)

16. At a local high school, 58% of the students have a part-time job; 17% of the students participate in school sports; 8% of the students have a part-time job <u>and</u> participate in school sports. If we were to select a student from the school at random, answer the following questions:

(a) What is the probability the student either has a part-time job, participates in school sports, or both?

(See the solution on page 355.)

(b) What is the probability the student neither has a part-time job nor participates in school sports?

(See the solution on page 355.)

(c) What is the probability the student has a part-time job but does not participate in school sports?

(See the solution on page 356.)

(d) Is whether or not a student has a part-time job independent of whether or not they participate in school sports?

(See the solution on page 356.)

17. Among other things, a market stall sells asparagus, beets, and cucumbers. Here are some interesting facts:

43% of customers buy asparagus.
81% of customers buy beets.
25% of customers buy asparagus and beets.
72% of customers buy asparagus or cucumbers.
19% of customers buy asparagus and cucumbers.
35% of customers buy beets and cucumbers.

(a) What is the probability a customer buys asparagus or beets?
(See the solution on page 358.)

(b) What is the probability a customer buys cucumbers?
(See the solution on page 359.)

(c) What is the probability a customer buys neither beets nor cucumbers?
(See the solution on page 360.)

18. A survey of Canadian sports fans determined the following facts:

77% of all fans follow hockey.
54% of all fans follow football.
82% of all fans follow **either** hockey **or** basketball.
36% of all fans follow **both** hockey **and** football.
13% of all fans follow **both** hockey **and** basketball.
8% of all fans follow **both** football **and** basketball.
5% of all fans follow all three.

(a) What is the probability a fan follows either hockey or football?
(See the solution on page 364.)

(b) What is the probability a fan follows only basketball?
(See the solution on page 364.)

(c) What is the probability a fan follows none of these three sports?
(See the solution on page 364.)

SUMMARY OF KEY CONCEPTS IN LESSON 6

- **The parameters of a binomial distribution are *n* and *p*.**
- The binomial distribution is a **discrete** distribution where each trial is **independent**. If we have a fixed number of trials *n* and if the probability of "yes" is the same for each trial, *p*, the random variable *X* has a binomial distribution where *X* = the number of "yeses". We can say $\boldsymbol{X \sim B\ (n, p)}$.
  - If we are given a **percentage**, a **proportion**, or a **fraction** we are given a value of ***p***. We will immediately suspect we have a binomial distribution at that point. All we need is a value for *n* to clinch it.
  - If we are rolling dice, tossing coins, or guessing on a test, we have a binomial distribution where we are expected to know the value of *p* ourselves. Again, we must have a specific number of trials *n* or else the problem is not binomial.
- The binomial probability formula is $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.
- If *X* has a **binomial** distribution, then the **mean of *X* = μ_X = *np*** and the **standard deviation of *X* = σ_X = $\sqrt{np(1-p)}$.**
- The **mean of $\hat{p}$ = $\mu_{\hat{p}}$ = *p*** and the **standard deviation of $\hat{p}$ = $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$**.
- Our **Rule of Thumb** says: If $N \geq 10n$, and if $np \geq 10$ and $n(1-p) \geq 10$, then both *X* and $\hat{p}$ in a binomial distribution are approximately normal.
- If we have boxed in an outrageous amount of *X* values in a binomial probability problem, we can bet our Rule of Thumb will tell us *X* is approximately normal, so we can use an ***X*-bell curve** to compute the approximate probability.
  - The standardizing formula for the variable *X* is $z = \frac{x - \mu_X}{\sigma_X} = \frac{x - np}{\sqrt{np(1-p)}}$.
- If we want to find the probability the sample proportion $\hat{p}$ is above, below or between some given amount(s), we can bet our Rule of Thumb will tell us $\hat{p}$ is approximately normal, so we can use a **$\hat{p}$-bell curve** to compute the approximate probability.
  - The standardizing formula for $\hat{p}$ is $z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$.

LECTURE PROBLEMS FOR LESSON 6

For your convenience, here are the 10 questions I used as examples in this lesson. Do not make any marks or notes on these questions below. Especially, do not circle the correct choice in the multiple choice questions. You want to keep these questions untouched, so that you can look back at them without any hints. Instead, make any necessary notes, highlights, etc. in the lecture part above.

1. Thirty-five percent of the voters in the last election voted Liberal. If you randomly selected ten voters from the last election, what is the probability exactly four of them voted Liberal?

(See the solution on page 382.)

2. A die is rolled seven times.

(a) What is the probability we roll a three four times?

(A) 0.0156 **(B)** 0.2857 **(C)** 0.4286 **(D)** 0.5714 **(E)** 0.8988

(See the solution on page 383.)

(b) What is the probability you get at least one 5?

(A) 0.2791 **(B)** 0.6093 **(C)** 0.7209 **(D)** 0.3907 **(E)** 1

(See the solution on page 385.)

3. A student is writing a multiple-choice Statistics exam. Each question has 5 choices and only one choice is correct. There are a total of 20 questions on the exam. If the student is simply guessing on every single question:

(a) What is the probability he just barely passes the exam (gets exactly 50%)?

(A) 0.5000 **(B)** 0.1762 **(C)** 0.0026 **(D)** 0.0020 **(E)** 0.0002

(See the solution on page 386.)

(b) What is the probability he passes the exam?

(A) 0.5000 **(B)** 0.1762 **(C)** 0.0026 **(D)** 0.0020 **(E)** 0.0002

(See the solution on page 387.)

4. A seed company has determined its seeds have a 90% chance of germinating. If 20 seeds are planted what is the probability more than 18 will germinate?

(A) 0.270 **(B)** 0.285 **(C)** 0.392 **(D)** 0.608 **(E)** 0.715

(See the solution on page 388.)

5. It is known 75% of the executives at a major multinational corporation are male. In a random sample of 8 executives from this corporation, what is the probability 3 or 4 of them are female?

(A) 0.2942 **(B)** 0.0865 **(C)** 0.2076 **(D)** 0.0231 **(E)** 0.1096

(See the solution on page 389.)

6. An airline determines 97% of the people who booked a flight actually show up in time to take their seat. Assuming this is true, what is the probability, in a randomly selected sample of 12 people who independently booked various flights, no more than 10 of them showed up?

(See the solution on page 390.)

7. A newspaper reports that one in three drivers routinely exceed the speed limit. Assuming this is true, we select a random sample of 30 drivers.

(a) What is the probability exactly half of them exceed the speed limit?

(See the solution on page 391.)

(b) What is the mean number of drivers in a sample of this size who routinely exceed the speed limit, and what is the standard deviation?

(See the solution on page 392.)

8. A mail-order company finds 7% of its orders tend to be damaged in shipment. If 500 orders are shipped:

(a) Compute the mean and standard deviation of the number of orders that would be damaged.

(See the solution on page 399.)

(b) Find the approximate probability between 30 and 50 orders (inclusive) will be damaged.

(See the solution on page 400.)

(c) Find the approximate probability at least 50 orders will be damaged.

(See the solution on page 401.)

9. A recent article claimed 40% of 12-year old American children are at least 5 pounds overweight. Assuming this is true, what is the probability a random sample of 600 children finds no more than 220 12-year old American children are overweight? Use the continuity correction.

(A) 0.0470 **(B)** 0.0475 **(C)** 0.9525 **(D)** 0.9530 **(E)** none of the above

(See the solution on page 403.)

10. In Big City only 35% of the voters in the last election were in favour of a one-time levy to cover the cost of sewer upgrades. During the current campaign a random sample of 575 voters will be selected. Assume the opinion has not changed since the last election.

(a) What is the mean and standard deviation of the number of voters sampled in favour of the levy?

(A) 0.35; 0.020 **(B)** 0.65; 0.020 **(C)** 201.25; 11.44
(D) 373.75; 11.44 **(E)** 350; 13.46

(See the solution on page 410.)

(b) What is the mean and standard deviation of the proportion of voters sampled in favour of the levy?

(A) 0.35; 0.020 **(B)** 0.65; 0.020 **(C)** 201.25; 11.44
(D) 373.75; 11.44 **(E)** 350; 13.46

(See the solution on page 410.)

(c) What is the probability the sample proportion will be between 30% and 40% in favour of the levy?

(See the solution on page 411.)

(d) What is the probability a random sample of 575 voters finds at least 225 in favour of the levy?

(See the solution on page 412.)

SUMMARY OF KEY CONCEPTS IN LESSON 7

- Populations have **parameters**; samples have **statistics**.
- The **Law of Large Numbers** states:
  - ***As n* gets larger, any statistic will come closer and closer to the value of the parameter it is estimating. The statistic will have less and less variability.**
  - For example, as n gets larger, $\bar{x}$ will come closer and closer to μ.
- The mean of $\bar{x}$ = $\mu_{\bar{x}} = \mu$ and the standard deviation of $\bar{x}$ = $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.
- If the population is normal (we have an *X*-bell curve), then the sample mean is also normally distributed (we have an $\bar{x}$-bell curve).
- The **Central Limit Theorem** states:
  - ***As n* gets larger, the distribution of the sample mean becomes closer and closer to the normal distribution.**
  - Even if the population is not normal, $\bar{x}$ will have an approximately normal distribution as long as n is large. Usually, $n \geq 15$ is large enough for the Central Limit Theorem to apply; $n \geq 40$ if a population is strongly skewed, or has outliers.
- The $\bar{x}$-bell curve standardizing formula is $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$.
- Be careful to distinguish between an *X*-bell curve problem and an $\bar{x}$-bell curve problem. If you are ever asked to find the probability that the sample *mean* or sample *average* is such and such, you are using an $\bar{x}$-bell curve.
- **$\hat{\theta}$ is an <u>unbiased estimator</u> of θ if the *mean* of $\hat{\theta}$ = $\mu_{\hat{\theta}} = \theta$.**
- The size of a population has very little to do with the variability of a statistic. It is the size of the sample that matters. If you are comparing two populations, you would prefer to take samples of the same size, so that their statistics will have about the same variability.
- The control limits for **$\bar{x}$ control charts** are $\mu \pm 3\frac{\sigma}{\sqrt{n}}$.
- Three signs a process may be **out of control** are the **one-point-out rule**, the **run-of-nine** rule, or any pattern or trend in the control chart.

LECTURE PROBLEMS FOR LESSON 7

For your convenience, here are the 8 questions I used as examples in this lesson. Do not make any marks or notes on these questions below. Especially, do not circle the correct choice in the multiple choice questions. You want to keep these questions untouched, so that you can look back at them without any hints. Instead, make any necessary notes, highlights, etc. in the lecture part above.

1. In the following, identify if the underlined number is a parameter or a statistic.

(a) A graduate student in family studies wants to know the proportion of married people at her university. She randomly selects 100 of the more than 20,000 students and finds 21% are married. In fact, for that particular university, 19% of the students are married.

(See the solution on page 420.)

(b) The distribution of heights of the adult male population has a mean of 69 inches. A random sample of 25 adult males gives a mean height of 67.5 inches.

(See the solution on page 420.)

2. A researcher intends to compare the proportion of Canadian adult males who are overweight to the proportion of American adult males who are overweight. Even though he knows the population of the United States of America is about ten times the size of Canada he decides to randomly select 1000 adult males from each country to conduct his research. Which of the following statements is true?

(A) No conclusion will be possible because the American sample size is too small.
(B) The research is pointless because we cannot compare Canadians to Americans.
(C) The Canadian sample proportion will have a considerably greater variability than the American sample proportion.
(D) The Canadian sample proportion will have a considerably smaller variability than the American sample proportion.
(E) The Canadian sample proportion will have about the same variability as the American sample proportion.

(See the solution on page 436.)

3. The Central Limit Theorem states:

(A) The distribution of the sample mean is normal, provided the population is normal.
(B) As the sample size increases, the sample mean becomes closer to the population mean.
(C) As the sample size increases, the distribution of the sample becomes closer to normal.
(D) The distribution of the sample is normal, provided the population is normal.
(E) As the sample size increases, the distribution of the sample mean becomes closer to normal.

(See the solution on page 437.)

4. A bottling plant produces soda bottles whose contents follow a normal distribution with mean 343 ml and standard deviation 5 ml. A consumer buys a six-pack. Assume this is a random sample.

(a) What is the probability a randomly selected bottle has more than 345 ml?
(A) .4000 **(B)** .3446 **(C)** .6554 **(D)** .1635 **(E)** .9798

(See the solution on page 439.)

(b) What is the probability the six-pack averages more than 345 ml per bottle?
(A) .4000 **(B)** .3446 **(C)** .6554 **(D)** .1635 **(E)** .9798

(See the solution on page 440.)

5. The mean income of households is 69 thousand dollars with a standard deviation of 24 thousand dollars. The population distribution is known to be strongly right-skewed.

(a) You select a random sample of 1,000 households and construct a histogram of that sample. You would expect the distribution of that histogram to be

(A) approximately normal with a mean of about 69 thousand dollars and a standard deviation of about 24 thousand dollars.
(B) approximately normal with a mean of about 69 thousand dollars and a standard deviation of about 759 dollars.
(C) approximately normal with a mean of about 69 thousand dollars and a standard deviation of about 759 thousand dollars.
(D) approximately right-skewed with a mean of about 69 thousand dollars and a standard deviation of about 24 thousand dollars.
(E) approximately right-skewed with a mean of about 69 thousand dollars and a standard deviation of about .759 thousand dollars.

(See the solution on page 442.)

(b) You select ten thousand random samples. Each sample consists of 1,000 households and you compute the sample mean each time. You now construct a histogram of the ten thousand sample means you computed. You would expect the distribution of that histogram of sample means to be

(A) approximately normal with a mean of about 69 thousand dollars and a standard deviation of about 24 thousand dollars.
(B) approximately normal with a mean of about 69 thousand dollars and a standard deviation of about 759 dollars.
(C) approximately normal with a mean of about 69 thousand dollars and a standard deviation of about .24 thousand dollars.
(D) approximately right-skewed with a mean of about 69 thousand dollars and a standard deviation of about 24 thousand dollars.
(E) approximately right-skewed with a mean of about 69 thousand dollars and a standard deviation of about .759 thousand dollars.

(See the solution on page 443.)

(c) What is the probability the mean income of one hundred randomly selected households will be between 65 and 75 thousand dollars?

(A) .1662 **(B)** .9463 **(C)** .0537 **(D)** .8338 **(E)** none of the above

(See the solution on page 444.)

(d) What is the probability a randomly selected household will have an income below $40,000?

(A) .0001 **(B)** .1131 **(C)** .1151 **(D)** .9999 **(E)** none of the above

(See the solution on page 445.)

6. The number of violent crimes reported per day in a large Canadian city follows a distribution that possesses a mean equal to 1.5 and a standard deviation equal to 1.7. The probability that, in the next 50 days, more than 95 violent crimes are reported is approximately equal to:

(A) 0.4515 **(B)** 0.0000 **(C)** 0.0485 **(D)** 0.1255 **(E)** 0.3745

(See the solution on page 446.)

7. The cost of individual long-distance phone calls for a company is a random variable with mean μ = \$3.20 and standard deviation σ = \$.80. The probability that 100 phone calls cost a total of no more than \$330 is

(A) 0.1056 **(B)** 0.3944 **(C)** 0.8944 **(D)** 1.0000 **(E)** none of the above

(See the solution on page 447.)

8. A factory produces widgets that are supposed to be normally distributed with a mean weight of 6.00 kg and a variance of 0.09 kg^2. To ensure the manufacturing process stays in control, every hour they randomly select 9 widgets and compute their mean weight. Here are the values for $\bar{x}$ for the last 16 hours:

6.12, 6.05, 5.97, 6.00, 5.82, 5.78, 6.03, 6.28,
6.30, 5.99, 5.99, 6.32, 6.11, 5.63, 5.85, 5.99

(a) Compute the $\bar{x}$ control limits for this process.

(See the solution on page 450.)

(b) Name two indicators a process has gone out of control.

(See the solution on page 450.)

(c) Construct an $\bar{x}$ control chart for this process and determine if the process is out of control. If so, when did the process go out of control?

(See the solution on page 451.)

PREPARING FOR THE SECOND MIDTERM EXAM

- If you have done all of the homework from all 4 lessons, you are now ready to start preparing for your second midterm exam. **Be sure to do <u>all</u> of the exams** from the Smiley Cheng *Multiple-Choice Problems Set for Basic Statistical Analysis I (Stat 1000)* available in the Statistics section of the UM Book Store (but not the final exams obviously). **Note that the course used to have only one midterm exam, so only certain questions in the old midterms and finals are appropriate. I will send you details of which questions are relevant to look at if you have signed up for Grant's Updates.** (I prefer to wait until the exam is approaching to make sure I know which old exam questions are relevant.) I suggest you start with the most recent exams and work your way backwards. The more recent exams are probably more indicative of what your exam will be like. The exams from the 90s are probably too easy, as the midterm has definitely gotten harder over the years.

- **The solutions to the final exams are here in Appendix D of my book starting on page D-1.** Solutions to the old term tests are included in Volume 1 of my book.

- **If your exam has a long answer section, be sure you do the long answer part first.** Time is sometimes an issue on the exam. If you are running out of time, you would rather be rushed as you are finishing off some multiple-choice questions (where you could always guess and hope) than feel rushed while trying to complete a more valuable long answer question. **A prepared student should have no fear of the long answer questions while there will undoubtedly be multiple-choice questions that will confuse any student.**

- **Never doubt yourself when answering a multiple-choice question.** If your answer is not one of the choices, simply select the closest choice and move on. Never waste your time redoing a question! If you have done it wrong, you are likely to still do it wrong the second time. You have other questions to do. Getting obsessed with one question, may mean not having time to answer two or three or more at the end. They are all worth the same marks, so leaving two or three blank at the end in order to vainly attempt to get one question right is just silly. If you have completed the exam, and still have time, by all means go back and try questions you had doubts about. Since you are now looking at the question fresh and with some distance, you have a much better chance of correcting your mistake (if you made one).

- **If the question is strictly theory, no math at all, you should never spend more than two minutes to make up your mind what choice to make.** Believe me. If you don't know the answer within one minute, they got you anyway, so just trust your gut, make a choice, and move on. That will buy you time to spend on the slower calculation questions.

APPENDIX A

HOW TO USE STAT MODES ON YOUR CALCULATOR

In the following pages, I show you how to enter data into your calculator in order to compute the mean and standard deviation. I also show you how to enter x, y data pairs in order to get the correlation, intercept and slope of the least squares regression line.

Please make sure that you are looking at the correct page when learning the steps. I give steps for several brands and models of calculator.

I consider it absolutely vital that a student know how to use the Stat modes on their calculator. It can considerably speed up certain questions and, even if a question insists you show all your work, gives you a quick way to check your answer.

If you cannot find steps for your calculator in this appendix, or cannot get the steps to work for you, do not hesitate to contact me. I am very happy to assist you in calculator usage (or anything else for that matter).

SHARP CALCULATORS
(Note that the EL-510 does not do Linear Regression.)

You will be using a "MODE" button. Look at your calculator. If you have "MODE" actually written on a button, press that when I tell you to press " [MODE] ". If you find mode written above a button (some models have mode written above the "DRG" button, like this: " [DRG] " with MODE above it) then you will have to use the " [2nd F] " button to access the mode button; i.e. when I say " [MODE] " below, you will actually press " [2nd F] [DRG] ".

BASIC DATA PROBLEM

Feed in data to get the mean, $\bar{x}$, and standard deviation, *s* (which Sharps tend to denote "*sx*").

Step 1: Put yourself into the "STAT, SD" mode.

Press [MODE] [1] [0] (Screen shows "Stat0")

Step 2: Enter the data: 3, 5, 9.
To enter each value, press the "M+" button. There are some newer models of Sharp that have you press the "CHANGE" button instead of the "M+" button. (The "CHANGE" button is found close by the "M+" button.)

3 [M+] (DATA) 5 [M+] (DATA) 9 [M+] (DATA)

You should see the screen counting the data as it is entered (Data Set=1, Data Set=2, Data Set=3).

Step 3: Ask for the mean and standard deviation.

[RCL] [4] ($\bar{x}$)

We see that $\bar{x}$ =5.6666...=5.6667.

[RCL] [5] (*sx*)

We see that *s*=3.05505...=3.0551

Step 4: Return to "NORMAL" mode. This clears out your data as well as returning your calculator to normal.

[MODE] [0]

LINEAR REGRESSION PROBLEM

Feed in *x* and *y* data to get the correlation coefficient, r, the intercept, a, and the slope, b.

Step 1: Put yourself into the "STAT, LINE" mode.

Press [MODE] [1] [1] (Screen shows "Stat1")

Step 2: Enter the data:

x	3	5	9
y	7	10	14

Note you are entering in pairs of data (the *x* and *y* must be entered as a pair). The pattern is first *x*, press "STO" to get the comma, first *y*, then press "M+" (or **"CHANGE")** to enter the pair; repeat for each data pair.

3 [STO] ((x,y)) 7 [M+] (DATA)

5 [STO] ((x,y)) 10 [M+] (DATA)

9 [STO] ((x,y)) 14 [M+] (DATA)

You should see the screen counting the data as it is entered (Data Set=1, Data Set=2, Data Set=3).

Step 3: Ask for the correlation coefficient, intercept, and slope. (The symbols may appear above different buttons than I indicate below.)

[RCL] [÷] (r)

We see that r=0.99419...=0.9942.

[RCL] [(] (a)

We see that a=3.85714...=3.8571.

[RCL] [)] (b)

We see that b=1.14285...=1.1429.

Step 4: Return to "NORMAL" mode. This clears out your data as well as returning your calculator to normal.

[MODE] [0]

CASIO CALCULATORS
(Note that some Casios do not do Linear Regression.)

BASIC DATA PROBLEM

Feed in data to get the mean, $\bar{x}$, and standard deviation, *s* (which Casios tend to denote "$x\sigma_{n-1}$" or simply "σ_{n-1}").

Step 1: Put yourself into the "SD" mode.

Press "[MODE]" once or twice until you see "SD" on the screen menu and then select the number indicated. A little "SD" should then appear on your screen.

Step 2: Clear out old data.

[SHIFT] [AC] (Scl) [=] (Some models will have "Scl" above another button. Be sure you are pressing "Scl", the "Stats Clear" button. (Some models call it "SAC" for "Stats All Clear" instead of Scl.)

Step 3: Enter the data: 3, 5, 9.
To enter each value, press the "M+" button.

3 [M+] (DT) 5 [M+] (DT) 9 [M+] (DT) (You use the "M+" button to enter each piece of data.)

Step 4: Ask for the mean and standard deviation.

[SHIFT] [1] ($\bar{x}$) [=]

We see that $\bar{x}$ = 5.6666...=5.6667.

[SHIFT] [3] ($x\sigma_{n-1}$) [=]

We see that *s* = 3.05505...=3.0551

(Some models may have $\bar{x}$ and $x\sigma_{n-1}$ above other buttons rather than "1" and "3" as I illustrate above.)

If you can't find these buttons on your calculator, look for a button called "S. VAR" (which stands for "Statistical Variables", it is probably above one of the number buttons).

Press: [SHIFT] [S. VAR] and you will be given a menu showing the mean and standard deviation. Select the appropriate number on the menu and press "=" (You may need to use your arrow buttons to locate the $\bar{x}$ or $x\sigma_{n-1}$ .options.)

Step 5: Return to "COMP" mode.

Press [MODE] and select the "COMP" option.

LINEAR REGRESSION PROBLEM

Feed in *x* and *y* data to get the correlation coefficient, r, the intercept, a, and the slope, b.

Step 1: Put yourself into the "REG, Lin" mode.

Press "[MODE]" once or twice until you see "Reg" on the screen menu and then select the number indicated. You will then be sent to another menu where you will select "Lin". (Some models call it the "LR" mode in which case you simply choose that instead.)

Step 2: Clear out old data.

Do the same as Step 2 for "Basic Data".

Step 3: Enter the data.

x	3	5	9
y	7	10	14

Note you are entering in pairs of data (the *x* and *y* must be entered as a pair). The pattern is first *x*, first *y*; second *x*, second *y*; and so on. Here is the data we want to enter:

3 [,] 7 [M+] (DT) 5 [,] 10 [M+] (DT) 9 [,] 14 [M+] (DT)

(If you can't find the comma button "[,]", you probably use the open bracket button instead to get the comma "[[(−−]". You might notice "[x_D, y_D]" in blue below this button, confirming that is your comma.)

Step 4: Ask for the correlation coefficient, intercept, and slope. (The symbols may appear above different buttons than I indicate below.)

[SHIFT] [(] (r) [=]

We see that r = 0.99419...=0.9942.

[SHIFT] [7] (A) [=]

We see that a = 3.85714...=3.8571.

[SHIFT] [8] (B) [=]

We see that b = 1.14285...=1.1429.

If you can't find these buttons on your calculator, look for a button called "S. VAR"

Press: [SHIFT] [S. VAR] and you will be given a menu showing the mean and standard deviation. **Use your left and right arrow buttons to see other options**, like "*r*". Select the appropriate number on the menu and press "=".

Step 5: Return to "COMP" mode.

Press [MODE] and select the "COMP" option.

HEWLETT PACKARD HP 10B II

BASIC DATA PROBLEM

Feed in data to get the mean, $\bar{x}$, and standard deviation, *s* (which it denotes "*Sx*").

Step 1: Enter the data: 3, 5, 9.
To enter each value, press the "Σ+" button.

[3] [Σ+] [5] [Σ+] [9] [Σ+] (As you use the "Σ+" button to enter each piece of data, you will see the calculator count it going in: 1, 2, 3.)

Step 2: Ask for the mean and standard deviation.
Note that by "orange" I mean press the button that has the orange bar coloured on it. The orange bar is used to get anything coloured orange on the buttons.

[orange] [7 $\bar{x},\bar{y}$]

We see that $\bar{x}$ = 5.6666...=5.6667.

[orange] [8 s_x, s_y]

We see that *s* = 3.05505...=3.0551

Step 3: "Clear All" data ready for next time.

[orange] [C C ALL]

LINEAR REGRESSION PROBLEM

Feed in *x* and *y* data to get the correlation coefficient, r, the intercept, a, and the slope, b.

Step 1: Enter the data:

x	3	5	9
y	7	10	14

Note you are entering in pairs of data (the *x* and *y* must be entered as a pair). The pattern is first *x*, first *y*; second *x*, second *y*; and so on.

[3] [INPUT] [7] [Σ+]

[5] [INPUT] [10] [Σ+]

[9] [INPUT] [14] [Σ+]

(As you use the "Σ+" button to enter each pair of data, you will see the calculator count it going in: 1, 2, 3.)

Step 2: Ask for the correlation coefficient, intercept, and slope.

[orange] [4 $\hat{x},r$] [orange] [K SWAP]

We see that *r* = 0.99419...=0.9942.
Note that the "SWAP" button is used to get anything that is listed second (after the comma) like "*r*" in this case.

The intercept has to be found by finding $\hat{y}$ when *x*=0:

0[orange] [5 $\hat{y},m$]

We see that *a* = 3.85714...=3.8571.
The slope is denoted "*m*" on this calculator:

[orange] [5 $\hat{y},m$] [orange] [K SWAP]

We see that *b* = 1.14285...=1.1429.

Step 3: "Clear All" data ready for next time.

[orange] [C C ALL]

TEXAS INSTRUMENTS TI-30X-II
(Note that the TI-30Xa does not do Linear Regression.)

BASIC DATA PROBLEM

Feed in data to get the mean, $\bar{x}$, and standard deviation, *s* (which it denotes "*Sx*").

Step 1: Clear old data.

[2nd] [DATA] (STAT) Use your arrow keys to ensure "*CLRDATA*" is underlined then press [ENTER =]

Step 2: Put yourself into the "STAT 1-Var" mode.

[2nd] [DATA] (STAT) Use your arrow keys to ensure "1-Var" is underlined then press [ENTER =]

Step 3: Enter the data: 3, 5, 9.

(You will enter the first piece of data as "X1", then use the down arrows to enter the second piece of data as "X2", and so on.)

[DATA] 3 [ENTER =] $(X1 = 3)$

[⇩] [⇩] 5 [ENTER =] $(X2 = 5)$

[⇩] [⇩] 9 [ENTER =] $(X3 = 9)$

Step 4: Ask for the mean and standard deviation.

Press [STATVAR] then you can see a list of outputs by merely pressing your left and right arrows to underline the various values.

We see that $\bar{x}$ = 5.6666...=5.6667.

We see that *s* = 3.05505...=3.0551

Step 5: Return to standard mode.

[CLEAR] This resets your calculator ready for new data next time.

LINEAR REGRESSION PROBLEM

Feed in *x* and *y* data to get the correlation coefficient, r, the intercept, a, and the slope, b.

Step 1: Clear old data (as in **BASIC DATA PROBLEM** at left).

Step 2: Put yourself into the "STAT 2-Var" mode.

[2nd] [DATA] (STAT) Use your arrow keys to ensure "2-Var" is underlined then press [ENTER =]

Step 3: Enter the data:

x	3	5	9
y	7	10	14

(You will enter the first *x*-value as "X1", then use the down arrow to enter the first *y*-value as "Y1", and so on.)

[DATA] 3 [ENTER =] [⇩] 7 [ENTER =] $(X1 = 3, Y1 = 7)$

[⇩] 5 [ENTER =] [⇩] 10 [ENTER =] $(X2 = 5, Y2 = 10)$

[⇩] 9 [ENTER =] [⇩] 14 [ENTER =] $(X3 = 9, Y3 = 14)$

Step 4: Ask for the correlation coefficient, intercept, and slope.

Press [STATVAR] then you can see a list of outputs by merely pressing your left and right arrows to underline the various values. **Note: Your calculator may have a and b reversed. To get a, you ask for b; to get b you ask for a.** Don't ask me why that is, but if that is the case then realize it will <u>always</u> be the case.

We see that r = 0.99419...=0.9942.

We see that a = 3.85714...=3.8571.

We see that b = 1.14285...=1.1429.

Step 5: Return to standard mode (as in **BASIC DATA PROBLEM** at left).

TEXAS INSTRUMENTS TI-36X
(Note that the TI-30Xa does not do Linear Regression.)

BASIC DATA PROBLEM

Feed in data to get the mean, $\bar{x}$, and standard deviation, s (which it denotes "σx_{n-1}").

Step 1: Put yourself into the "STAT 1" mode.

[3rd] [x ⇄ y] (STAT 1)

Step 2: Enter the data: 3, 5, 9.
To enter each value, press the "Σ+" button.

3 [Σ+] 5 [Σ+] 9 [Σ+] (As you use the "Σ+" button to enter each piece of data, you will see the calculator count it going in: 1, 2, 3.)

Step 3: Ask for the mean and standard deviation.

[2nd] [x^2] ($\bar{x}$)

We see that $\bar{x}$ = 5.6666...=5.6667.

[2nd] [$\sqrt{x}$] (σ_{xn-1})

We see that s = 3.05505...=3.0551

Step 4: Return to standard mode.

[ON/AC] (Be careful! If you ever press this button during your work you will end up resetting your calculator and losing all of your data. Use the [CE/C] button to clear mistakes without resetting your calculator. I usually press this button a couple of times to make sure it has cleared any mistake completely.)

LINEAR REGRESSION PROBLEM

Feed in *x* and *y* data to get the correlation coefficient, r, the intercept, a, and the slope, b.

Step 1: Put yourself into the "STAT 2" mode.

[3rd] [Σ+] (STAT 2)

Step 2: Enter the data:

x	3	5	9
y	7	10	14

Note you are entering in pairs of data (the *x* and *y* must be entered as a pair). The pattern is first *x*, first *y*; second *x*, second *y*; and so on.

3 [$x \rightleftarrows y$] 7 [Σ+]

5 [$x \rightleftarrows y$] 10 [Σ+]

9 [$x \rightleftarrows y$] 14 [Σ+]

(As you use the "Σ+" button to enter each pair of data, you will see the calculator count it going in: 1, 2, 3.)

Step 3: Ask for the correlation coefficient, intercept, and slope.

Note that this calculator uses the abbreviations "COR" for correlation, "ITC" for intercept and "SLP" for slope.

[3rd] [4] (COR)

We see that r = 0.99419...=0.9942.

[2nd] [4] (ITC)

We see that a = 3.85714...=3.8571.

[2nd] [5] (SLP)

We see that b = 1.14285...=1.1429.

Step 4: Return to standard mode.

[ON/AC]

TEXAS INSTRUMENTS TI-BA II Plus

Put yourself into the "LIN" mode.

[2nd] [8] (STAT) If "LIN" appears, great; if not, press [2nd] [ENTER] (SET) repeatedly until "LIN" does show up. Then press [2nd] [CPT] (QUIT) to "quit" this screen.

Note: Once you have set the calculator up in "LIN" mode, it will stay in that mode forever. You can now do either "Basic Data" or "Linear Regression" problems.

BASIC DATA PROBLEM

Feed in data to get the mean, $\bar{x}$, and standard deviation, *s* (which it denotes "*Sx*").

Step 1: Clear old data.

[2nd] [7] (DATA) [2nd] [CE/C] (CLR Work)

Step 2: Enter the data: 3, 5, 9.

(You will enter the first piece of data as "X1", then use the down arrows to enter the second piece of data as "X2", and so on. Ignore the "Y1", "Y2", etc.)

[DATA] 3 [ENTER =] (X1 = 3)

[⇓] [⇓] 5 [ENTER =] (X2 = 5)

[⇓] [⇓] 9 [ENTER =] (X3 = 9)

Step 3: Ask for the mean and standard deviation.

Press [2nd] [8] (STAT) then you can see a list of outputs by merely pressing your up and down arrows to reveal the various values.

We see that $\bar{x}$ = 5.6666...=5.6667.

We see that *s* = 3.05505...=3.0551

Step 4: Return to standard mode.

[ON/OFF] This resets your calculator ready for new data next time.

LINEAR REGRESSION PROBLEM

Feed in *x* and *y* data to get the correlation coefficient, r, the intercept, a, and the slope, b.

Step 1: Clear old data.

[2nd] [7] (DATA) [2nd] [CE/C] (CLR Work)

Step 2: Enter the data:

x	3	5	9
y	7	10	14

(You will enter the first *x*-value as "X1", then use the down arrow to enter the first *y*-value as "Y1", and so on.)

[DATA] 3 [ENTER =] [⇓] 7 [ENTER =] (X1 = 3, Y1 = 7)

[⇓] 5 [ENTER =] [⇓] 10 [ENTER =] (X2 = 5, Y2 = 10)

[⇓] 9 [ENTER =] [⇓] 14 [ENTER =] (X3 = 9, Y3 = 14)

Step 3: Ask for the correlation coefficient, intercept, and slope.

Press [2nd] [8] (STAT) then you can see a list of outputs by merely pressing your up and down arrows to reveal the various values.

We see that r = 0.99419...=0.9942.

We see that a = 3.85714...=3.8571.

We see that b = 1.14285...=1.1429.

Step 4: Return to standard mode.

[ON/OFF] This resets your calculator ready for new data next time.