Grant's Tutoring

BASIC STATISTICS 1

Volume 3 of 3

September 2011 edition

This volume covers the topics taught after your second midterm 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 FORMULAS IN THIS COURSE

Lesson 1. sample standard deviation $= s = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}}$

Lesson 2. correlation $= r = \dfrac{1}{n-1}\sum\left(\dfrac{x_i - \bar{x}}{s_x}\right)\left(\dfrac{y_i - \bar{y}}{s_y}\right) = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$

slope $= b = r\dfrac{s_y}{s_x}$ intercept $= a = \bar{y} - b\bar{x}$

Lesson 4. standardizing formula for X bell curves: $z = \dfrac{x-\mu}{\sigma}$

Lesson 5. $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

If A and B are independent: $P(A \text{ and } B) = P(A) \times P(B)$

Lesson 6. If X has a binomial distribution with parameters n and p, then the mean of X = $\mu_X = np$ and the standard deviation of X = $\sigma_X = \sqrt{np(1-p)}$.

The mean of $\hat{p} = \mu_{\hat{p}} = p$ and the standard deviation $= \sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.

Also, $P(X = k) = \dbinom{n}{k} p^k (1-p)^{n-k}$.

Lesson 7. The mean of $\bar{x} = \mu_{\bar{x}} = \mu$ and the standard deviation $= \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

Central Limit Theorem: If n is large, $\bar{x}$ is approximately normal.

Standardizing formula for $\bar{x}$ bell curves: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

Lesson 8. Confidence Intervals for μ: $\bar{x} \pm z^* \dfrac{\sigma}{\sqrt{n}}$ or $\bar{x} \pm t^* \dfrac{s}{\sqrt{n}}$

Sample size determination: $n = \left(\dfrac{z^* \sigma}{m}\right)^2$

Lesson 9. Test statistics for μ: $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ or $t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$

Lesson 11. Confidence interval for p: $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Sample size determination: $n = \left(\dfrac{z^*}{m}\right)^2 p^*(1-p^*)$

Test statistic for p: $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

STEPS FOR TESTING A HYPOTHESIS

Step 1. State the null and alternative hypotheses (H_0 and H_a), and so determine if the test is 2-tailed, upper-tailed, or lower-tailed.

Step 2. Use the given α (always use α = 5% if none is given) to get the **critical value (z^* or t^*)** from Table D and state the **rejection region**.

If the test is upper-tailed:	Reject H_0 if $z > z^*$ (or $t > t^*$)
If the test is lower-tailed:	Reject H_0 if $z < -z^*$ (or $t < -t^*$)
If the test is 2-tailed:	Reject H_0 if $z < -z^*$ or $z > z^*$ (or Reject H_0 if $t < -t^*$ or $t > t^*$)

Step 3. Compute the **test statistic (z or t)** using the appropriate formula, and see if it lies in the rejection region.

Test statistics: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ $\quad t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ $\quad z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

Step 4. (Only if specifically asked to do so.) Compute the ***P*-value**.

Draw a bell curve, mark the test statistic (found in Step 3), and shade the area as instructed by H_a. That area is the *P*-value.

Remember, a *P*-value is very handy to know if you are asked to make decisions for more than one value of α.

Reject H_0 if *P*-value < α.

Step 5. State your conclusion.

Either: Reject H_0. There is statistically significant evidence that the alternative hypothesis is correct. (Replace the underlined part with appropriate wording from the problem that says H_a is correct.)

Or: Do not reject H_0. There is no statistically significant evidence that the alternative hypothesis is correct. (Replace the underlined part with appropriate wording from the problem that says we are not convinced that H_a is correct.)

SUMMARY OF KEY CONCEPTS IN LESSON 8

- A **confidence interval for the population mean, μ,** can be computed using either:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \qquad \text{or} \qquad \bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

- **If σ is given, we can use z; if σ is not given we must use t.**
- The t distribution is a bell-shaped density curve centred at 0 but of a varying spread. The spread depends on the degrees of freedom. The larger the degrees of freedom get, the narrower the spread of the t curve. Ultimately, the t curve will look exactly like the standard normal curve z.
- When using t our **df = n – 1**. This is because t is using the sample standard deviation s, since σ is unknown, to compute the standard error of the sample mean.
  - The standard error of the sample mean is $SE_{\bar{x}} = \frac{s}{\sqrt{n}}$.
- When listing givens: if n is given in a sentence, you know you are being given statistics (sample info) such as the *sample* mean $\bar{x}$ or the *sample* standard deviation s. If n is not given in the sentence, you are being given parameters (population info) such as the *population* mean μ or the *population* standard deviation σ.
- We say that **inferences about the mean are robust** because we can generally trust them to give us reliable confidence intervals for μ and trustworthy conclusions from our hypotheses for μ (next lesson) even if our population is not normal. All we need is a random sample and a large enough sample size.
- Inferences about the mean are only reliable if the sample mean $\bar{x}$ is normally distributed. We know $\bar{x}$ is normal if the population is normal. If the population is not normal, $\bar{x}$ will still be approximately normal as long as n is large (by Central Limit Theorem). In general, as long as **$n \geq 15$**, we can assume $\bar{x}$ is approximately normal. If we believe the population has outliers or is strongly skewed, we want **$n \geq 40$** to be able to safely assume $\bar{x}$ is approximately normal.
- The *higher* our level of confidence, the *wider* our confidence interval becomes. Put another way, **as we increase our confidence level, our margin of error increases also**.
- The *larger* our sample size, the *narrower* our confidence interval becomes. Put another way, **as we increase our sample size, our margin of error decreases**.
- The larger the standard deviation we are using, the larger the margin of error of our confidence interval will be.

- Everything that comes after the "±" in a confidence interval formula is computing the margin of error m. There are three parts to the margin of error in a confidence interval for the mean:
  - The **critical value (z^* or t^*)** which is dictated by the **level of confidence** and is read off Table D (the t distribution table).
  - The **standard deviation (σ or s)**.
  - The **sample size (n)**.
- When we construct, for example, a 95% confidence interval for the mean, we are saying we are 95% confident the true value of the mean, μ, is somewhere in that interval. We never know for sure if our interval has indeed caught μ, but we do know that, **if we repeatedly took samples of the same size, and repeatedly used those samples to construct 95% confidence intervals for μ then, in the long run, 95% of the intervals we construct will contain the true mean, μ.** This is because we are assuming our sample mean, $\bar{x}$, is in the middle 95% of the curve. Since 95% of all the possible $\bar{x}$ values are in this region, 95% of the confidence intervals we construct will contain the population mean μ.
- **If we are using z^* and <u>if the level of confidence we are using is not given on Table D</u>, we will have to find z^* by reading Table A backwards. The level of confidence is the percentage shaded in the *middle* of the z-bell curve, leaving us with two equally sized tails. Establish the area in the left tail (knowing the total area is 100%), then look for the closest value you can find to that left tail area in the body of Table A. That will give us the negative z-score. Discard the negative sign and you have z^*.** This would never happen if you are using t^*. You can rest-assured the level of confidence you are given will always be on Table D when you are using t^*.
- To determine the sample size required to obtain a specific margin of error m for a confidence interval for μ, we use the sample size formula:

$$n = \left(\frac{z^*\sigma}{m}\right)^2$$

- Always use the **Paint-Can Principle** when computing sample size. Which is to say, you must always round UP↑ to the next whole number. Even if you have computed $n = 58.1$, you must round that up to $n = 59$ because 58 units are not enough, you need 58.1 units, so you will have to select 59 units.

LECTURE PROBLEMS FOR LESSON 8

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. We select a random sample of 25 adult males and find their mean weight to be 67.53 kg. Assume the population is normal with a standard deviation of 10 kg.

(a) A 95% confidence interval for the mean weight of adult males is

(A) (63.61, 71.45) **(B)** (62.73, 72.23) **(C)** (63.40, 71.66)
(D) (61.61, 73.45) **(E)** (65.00, 70.06)

(See the solution on page 472.)

(b) Interpret this confidence interval including an explanation of what 95% confidence really means so that a layman might understand.

(See the solution on page 473.)

2. An SRS of 25 adult Canadians finds they have an average annual income of 30 thousand dollars with a standard deviation of 11 thousand dollars. A 95% confidence interval for the true mean income is

(A) (29.120, 31.980) **(B)** (25.688, 34.312) **(C)** (25.459, 34.541)
(D) (25.468, 34.532) **(E)** (26.236, 33.764)

(See the solution on page 474.)

3. It is known a certain machine that fills "5-pound" sacks of sugar actually fills the sacks with an amount that varies from sack to sack with variance 0.0016 pounds squared. Suppose a sample of 30 sacks has an average weight of 5.08 pounds. A 98% confidence interval estimate for the mean of the population of all sacks filled by the process in its current state is:

(A) $5.08 - 1.960(.0073) \le \mu \le 5.08 + 1.960(.0073)$

(B) $5.08 - 2.462(.0040) \le \mu \le 5.08 + 2.462(.0040)$

(C) $5.08 - 2.326(.0040) \le \mu \le 5.08 + 2.326(.0040)$

(D) $5.08 - 2.326(.0073) \le \mu \le 5.08 + 2.326(.0073)$

(E) $5.08 - 2.462(.0073) \le \mu \le 5.08 + 2.462(.0073)$

(See the solution on page 475.)

4. Investigating toxins produced by molds that infect corn crops, a biochemist prepares extracts of the mold culture and measures the amount of the toxic substance. From six preparations, the following observations on toxic substances per gram of solution are obtained: 1.2, 0.8, 0.6, 1.1, 1.2, 1.1. The endpoints of a 95% confidence interval for the mean amount of toxic substances are (assuming a normal population):

(A) $1.0 \pm 2.571(0.10)$ **(B)** $1.0 \pm 1.960(0.10)$ **(C)** $.95 \pm 2.571(0.10)$
(D) $.95 \pm 1.960(0.10)$ **(E)** $1.0 \pm 2.447(0.10)$

(See the solution on page 476.)

5. A manufacturer of bran flakes claims a 30 gram serving of its cereal has a mean of 4.4 grams of dietary fibre with standard deviation of 0.25 grams. Five random 30 gram servings have the following amounts of dietary fibre: 4.5, 4.4, 4.3, 4.4, 3.9. Assuming the population is normally distributed with given standard deviation, a 90% confidence interval for the true mean amount of dietary fibre is

(A) $4.3 \pm 1.645(.105)$ **(B)** $4.3 \pm 2.132(.105)$ **(C)** $4.3 \pm 2.132(.112)$
(D) $4.3 \pm 1.645(.112)$ **(E)** $4.4 \pm 1.645(.112)$

(See the solution on page 477.)

6. We are investigating the lifetimes of a brand of batteries and we know the standard deviation is 8 hours. How large a sample size is necessary to estimate the average lifetime within 2.5 hours with 95% confidence?

(A) 39 **(B)** 40 **(C)** 16 **(D)** 256 **(E)** 1000

(See the solution on page 479.)

7. A company that manufactures O-rings for the space shuttle knows that when the manufacturing process is stable the critical dimension has a standard deviation equal to 0.065 millimetre and that the distribution of the measurements look very similar to the normal distribution. The sample size that would be required to estimate the process mean, μ, so that a 99% confidence interval would have width 0.036 mm is

(A) 4 **(B)** 14 **(C)** 50 **(D)** 51 **(E)** 87

(See the solution on page 480.)

8. A 95% confidence interval for a normal population with known standard deviation produces a margin of error of .004 based on a sample size of 100. If we wish to halve the margin of error (maintaining the same level of confidence), how large a sample must we take?

(A) 50 **(B)** 100 **(C)** 200 **(D)** 400

(E) cannot be determined without the variance

(See the solution on page 481.)

9. A confidence interval for the mean of a population with known variance is to be constructed. Which of the following statements is FALSE?

(A) A confidence interval using a sample size of 100 will be wider than a confidence interval using a sample size of 50.

(B) A 99% confidence interval will be wider than a 90% confidence interval.

(C) The larger the variance, the wider the confidence interval will be.

(D) A 90% confidence interval will be narrower than a 95% confidence interval.

(E) A confidence interval using a sample size of 100 will be narrower than a confidence interval using a sample size of 50.

(See the solution on page 482.)

10. A random sample of 15 car batteries maintained their charge for an average of 36.7 hours in laboratory tests. Assuming the population is normal with a standard deviation of 7 hours, a 92% confidence interval for the average length of time car batteries will maintain their charge is

(A) (33.16, 40.24) **(B)** (33.54, 39.86) **(C)** (33.73, 39.67)

(D) (32.04, 41.36) **(E)** impossible to determine with the tables we have

(See the solution on page 484.)

SUMMARY OF KEY CONCEPTS IN LESSON 9

- ❖ All hypothesis tests for the mean, μ, are reliable provided we are dealing with representative samples and that the sample mean, $\bar{x}$, is normally distributed. We want random samples and properly designed experiments to be assured of a representative sample. If the population is normal, we certainly know $\bar{x}$ is normally distributed. If the population is not normal, $\bar{x}$ will be approximately normal if n is large (by Central Limit Theorem).
- ❖ Memorize the **five steps to test a hypothesis** (listed on page 512).
- ❖ The **critical value** is either z^* or t^* and is read off of Table D.
- ❖ The **test statistic** is computed from the appropriate formula.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad \text{or} \qquad t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

- ❖ The ***P*-value** tells us the probability of getting our test statistic or something even more extreme, in the appropriate direction, assuming the null hypothesis is correct.
- ❖ The smaller the *P*-value is, the stronger the evidence to reject the null hypothesis.
- ❖ If you are ever given more than one level of significance α in a problem, compute the *P*-value so that you can make your decision more easily.
- ❖ **Reject H_0 if the *P*-value < α.**
- ❖ **If you are given a confidence interval, you can use it to test a two-tailed hypothesis.** Confidence intervals are of no use for one-tailed hypotheses.
  - ▪ The level of confidence dictates the level of significance you must use.
  - ▪ $\alpha = 1 - C$
  - ▪ If the hypothesized value μ_0 is inside the confidence interval for μ, we cannot reject H_0 since it is possible that H_0 is correct. If μ_0 is not inside the interval, we can reject H_0 since we are confident H_0 is wrong.

LECTURE PROBLEMS FOR LESSON 9

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. State the null and alternative hypotheses for the following problems. Also, state whether the test is lower-tailed, upper-tailed, or two-tailed.

(a) A real estate broker trying to sell a business claims the average daily receipts are at least \$3200, we want to test this claim.

(b) An engineer declares it will take an average of 150 man-hours to manufacture an airplane part, we want to see if he is right.

(c) A pizza restaurant claims it will deliver your pizza in an average of 30 minutes or less, you want to know if this is true.

(d) A tutor claims his students will score more than 70, on average, on their tests.

(See the solution on page 499.)

2. For the situations below, state the rejection region. Assume the populations are normally distributed.

(a) H_0: $\mu = 25$ vs. H_a: $\mu > 25$; $\sigma = 6$; $n = 30$; $\alpha = 5\%$.

(b) H_0: $\mu = 35$ vs. H_a: $\mu > 35$; σ is unknown; $n = 30$; $\alpha = 5\%$.

(c) H_0: $\mu = 25$ vs. H_a: $\mu < 25$; $\sigma = 16$; $n = 20$; $\alpha = 1\%$.

(d) H_0: $\mu = 25$ vs. H_a: $\mu \neq 25$; $\sigma = 6$; $n = 10$; $\alpha = 5\%$.

(e) H_0: $\mu = 25$ vs. H_a: $\mu \neq 25$; σ is unknown; $n = 18$; $\alpha = 1\%$.

(f) H_0: $\mu = 35$ vs. H_a: $\mu < 35$; σ is unknown; $n = 10$; $\alpha = 5\%$.

(See the solution on page 504.)

3. Return to question 1 above. Let us assume that, in each of parts (a) through (d), we decided to reject H_0. Give a carefully worded conclusion relevant to the problem. In addition, state the conclusion if we did not reject H_0 in each case.

(See the solution on page 509.)

4. The scores of students who wrote a provincial exam in mathematics are normally distributed with mean 75 and variance 36. A teacher believes his group of students is above average. The 25 students in his class scored an average of 79. Is the teacher's belief justified? Include the hypotheses, critical value, and test statistic in your answer.

(See the solution on page 514.)

5. A manufacturer of small electric motors asserts they will draw 0.8 amperes, on average, under normal load conditions. A sample of 16 of the motors was tested and it was found the mean current was 0.96 amperes with a standard deviation of 0.32 amperes. Assume the sample is normally distributed. At the 1% level of significance, are we justified in rejecting the manufacturer's assertion? Include the hypotheses, critical value, test statistic, and a carefully worded conclusion relevant to this problem in your answer.

(See the solution on page 519.)

6. Compute the *P*-value for the hypothesis tests you conducted in questions 4 and 5 above. Would the results be significant if you used a 10% level of significance for both? What if you used a 1% level of significance for both? Interpret the *P*-values for questions 3 and 4.

(See the solution on page 530.)

7. The *P*-value of a test of a null hypothesis is

(A) the probability, assuming the null hypothesis is true, that the test statistic will take a value at least as extreme (in the appropriate direction) as that actually observed.
(B) the probability, assuming the null hypothesis is false, that the test statistic will take a value at least as extreme (in the appropriate direction) as that actually observed.
(C) the probability that the null hypothesis is true.
(D) the probability that the null hypothesis is false.
(E) either 0.01, or 0.05, or 0.10.

(See the solution on page 531.)

8. In testing the hypothesis H_0: $\mu = 10$ versus H_a: $\mu \neq 10$ at $\alpha = .05$, we know the standard deviation is 6. A sample of size 36 has a mean of 7.87. The *P*-value for the test and our decision is:

(A) .0166; reject H_0. **(B)** .0166; fail to reject H_0. **(C)** .0332, reject H_0.
(D) .0332, fail to reject H_0. **(E)** none of the above.

(See the solution on page 532.)

9. A real estate broker trying to sell a luxury apartment block claims that the average size of the apartments is at least 3200 square feet. You are sceptical and measure 16 random apartments, finding the size to be an average of 2984 square feet with a standard deviation of 420 square feet. The *P*-value for the appropriate hypothesis test would be:

(A) between .025 and .01 **(B)** between .05 and .025
(C) between .10 and .05 **(D)** .0394 **(E)** .0197

(See the solution on page 533.)

10. A manufacturer of bran flakes claims a 30 gram serving of its cereal has a mean of 4.4 grams of dietary fibre with standard deviation of 0.25 grams. Five random 30 gram servings have the following amounts of dietary fibre: 4.2, 4.1, 4.3, 4.4, 3.9. Assuming the population is normally distributed with given standard deviation, does this sample give convincing evidence the mean amount of dietary fibre is less than claimed? Use a .05 level of significance.

(a) State the hypotheses.

(A) H_0: $\mu = 4.4$ vs. H_a: $\mu \neq 4.4$ **(B)** H_0: $\mu = 4.3$ vs. H_a: $\mu < 4.3$

(C) H_0: $\bar{x} = 4.18$ vs. H_a: $\bar{x} < 4.18$ **(D)** H_0: $\mu = 4.4$ vs. H_a: $\mu > 4.4$

(E) H_0: $\mu = 4.4$ vs. H_a: $\mu < 4.4$

(See the solution on page 534.)

(b) We would reject the null hypothesis if

(A) $t < -2.132$ **(B)** $t > 2.132$ **(C)** $z < -1.645$

(D) $|z| > 1.96$ **(E)** $z > 1.645$

(See the solution on page 535.)

(c) The value of the test statistic is

(A) –1.97 **(B)** –1.645 **(C)** –2.56 **(D)** –1.95 **(E)** 1.645

(See the solution on page 535.)

(d) The *P*-value of the test is

(A) 0.0512 **(B)** 0.0488 **(C)** 0.0244

(D) 0.0256 **(E)** between 0.05 and 0.25

(See the solution on page 535.)

(e) The conclusion is

(A) Don't reject H_0. There is convincing evidence the mean amount of fibre is 4.4 grams.

(B) Reject H_0. There is convincing evidence the mean amount of fibre is not 4.4 grams.

(C) Don't reject H_0. There is no convincing evidence the mean amount of fibre is less than 4.4 grams.

(D) Don't reject H_0. There is no convincing evidence the mean amount of fibre is not 4.4 grams.

(E) Reject H_0. There is convincing evidence the mean amount of fibre is less than 4.4 grams.

(See the solution on page 536.)

11. The average score for students who complete a university statistics course is 55. A random sample of 20 students who used a certain tutor's services to help in this course scored an average of 61 with a standard deviation of 10. Assuming the distribution is normal, perform a test of significance using the *P*-value approach to determine if the tutor has a positive effect. Be sure to include the hypotheses, test statistic, *P*-value, and a properly worded conclusion in your answer.

(See the solution on page 538.)

12. The average wait time for an MRI in the year 2000 was 240 days with a standard deviation of 60 days. A random sample of 40 patients who received an MRI last year found that they waited an average of 220 days. Can the government claim they have reduced waiting times? Use the *P*-value approach and assume the standard deviation is unchanged. Be sure to include the hypotheses, test statistic, *P*-value, and a properly worded conclusion in your answer.

(See the solution on page 540.)

13. If (2.5, 4.7) is an observed 99% confidence interval for the parameter μ.

(a) Which of these statements is true?

(A) The probability a sample lies between 2.5 and 4.7 is 0.99.
(B) The probability the sample mean lies between 2.5 and 4.7 is 0.99.
(C) 99% of the samples lie between 2.5 and 4.7.
(D) 99% of the true means μ lie between 2.5 and 4.7
(E) If many random samples (each of the same size) were taken, and a 99% C.I. calculated using each sample, then (in the long run) 99% of them would contain the unknown μ.

(See the solution on page 542.)

(b) A possible 95% confidence interval constructed from the same information is
(A) (2.3, 4.7) **(B)** (2.7, 4.5) **(C)** (2.7, 4.7) **(D)** (2.3, 4.9) **(E)** (2.5, 4.9)

(See the solution on page 544.)

(c) The sample mean used to construct this confidence interval is
(A) 1.96 **(B)** 2.5 **(C)** 4.7 **(D)** 3.6 **(E)** impossible to determine.

(See the solution on page 545.)

(d) We want to test H_0: $\mu = 2.6$ vs. H_a: $\mu \neq 2.6$. With this information we definitely

(A) reject H_0 at the 1% level of significance.
(B) fail to reject H_0 at the 5% level of significance.
(C) fail to reject H_0 at the 2.5% level of significance.
(D) fail to reject H_0 at the 1% level of significance.
(E) are unable to make a decision.

(See the solution on page 547.)

14. In a test of hypotheses, how should the level α be chosen if we want very strong evidence against the null hypothesis H_0 before rejecting it?

(A) very small **(B)** very large **(C)** smaller than the *P*-value
(D) larger than the *P*-value **(E)** none of the above

(See the solution on page 548.)

15. Suppose we have rejected the null hypothesis H_0: $\mu = \mu_0$, at a significance level $\alpha = .01$, where μ is the mean of a normal distribution with known variance. Now, using the same sample information, we want to test the same null hypothesis, but at significance level $\alpha = .05$.

(A) We will reject H_0.
(B) We will fail to reject H_0.
(C) We cannot make any decision until we obtain the value from the table.
(D) We cannot make any decision until we calculate the value of the test statistic.
(E) Both (C) and (D).

(See the solution on page 549.)

16. It is imperative an important part of a jet engine measure an average of 224 mm with a standard deviation of .00400 mm. Assuming the measures of these parts are normally distributed with the required standard deviation, the quality control inspector examines a random samples of 16 using *JMP*™. Here are her results:

Jet Engine Part

Moments		**Test Mean=value**		
Mean	224.0019	Hypothesized Value		224
Std Dev	0.0618	Actual Estimate		224.002
Std Err Mean	0.0155	df		15
upper 95% mean	224.0349	Std Dev		0.0618
lower 95% mean	223.9690	Sigma given		0.004
N	16.0000		z Test	
		Test Statistic	1.9375	
		Prob > |z|	0.0527	
		Prob > z	0.0263	
		Prob < z	0.9737	

(a) Should she be concerned? Include the hypotheses for this test, the test statistic, the *p*-value, and your conclusion.
(b) Give a 95% confidence interval for the true mean. Could this confidence interval be used to test the hypothesis in (a)? Justify your answer.

(See the solution on page 552.)

17. Forty-four 2-litre cartons of milk from a local dairy were randomly selected and the time in days it took until the milk went sour was recorded. Health code requirements are that a carton of milk must last an average of at least 40 days before going sour. At right are the results of a *JMP™* analysis on whether this dairy is meeting code.

(a) A histogram of the sample is strongly left-skewed. Is this a concern?

(b) What are the hypotheses and conclusion, including the p-value.

Test Mean=value		
Hypothesized Value		40
Actual Estimate		35.0909
df		43
Std Dev		11.1893
	t Test	
Test Statistic	-2.9102	
Prob > |t|	0.0057	
Prob > t	0.9971	
Prob < t	0.0029	

(See the solution on page 554.)

18. Coordination tests conducted on 22 subjects have been input to *JMP™* with the output at right obtained. Which of the following statements is correct?

Test Mean=value		
Hypothesized Value		80
Actual Estimate		75.238
df		21
Std Dev		7.675
	t Test	
Test Statistic	–2.2341	
Prob > |t|	0.0301	
Prob > t	0.9850	
Prob < t	0.0150	

(A) The P-value for testing H_0: $\mu = 75.238$ vs. H_a: $\mu \neq 75.238$ is 0.0301.

(B) The P-value for testing H_0: $\mu = 75.238$ vs. H_a: $\mu < 75.238$ is 0.0150.

(C) The P-value for testing H_0: $\mu = 75.238$ vs. H_a: $\mu > 75.238$ is 0.9850.

(D) One can reject the null hypothesis H_0: $\mu = 80$ at the 99% significance level when the alternative hypothesis is H_a: $\mu > 80$.

(E) One cannot reject the null hypothesis H_0: $\mu = 80$ at the 2.5% significance level when the alternative hypothesis is H_a: $\mu < 80$.

(See the solution on page 555.)

SUMMARY OF KEY CONCEPTS IN LESSON 10

- A **matched pairs experiment** applies both treatment A and treatment B to each unit, allowing you to compute the difference in treatment scores for each pair. A **two-sample experiment** selects one random sample of size n_1 to receive treatment A and a second independent random sample of size n_2 to receive treatment B. Make sure you can tell the difference between matched pairs data and two-sample data.
- **In the matched pairs *t* test**, there is only one set of data because we have computed the differences between the scores for each pair. We test $\boldsymbol{H_0\!: \mu = 0}$ where μ is the mean of the differences, using $\bar{x}$, the sample mean of our differences, and s, the standard deviation of the differences.
  - If you are doing a one-tailed test, use the sign of $\bar{x}$ to determine the appropriate way to point your alternative hypothesis.
- **In a two-sample *t* test** we have two of everything. Two sample sizes, n_1 and n_2, two sample means, $\bar{x}_1$ and $\bar{x}_2$, and two sample standard deviations, s_1 and s_2.
- In this course, if we have two-sample data, we will always use **the pooled two-sample *t* method.** The pooled method is **not robust**. It is only reliable if these conditions are met:
  - We must have two independent simple random samples.
  - Both populations must be normally distributed.
  - Both populations must have the same variance, $\sigma_1^2 = \sigma_2^2$.
- The pooled method has a t distribution with $\mathbf{df} = \boldsymbol{n_1 + n_2 - 2}$.
- The pooled method uses the **pooled sample variance,** $\boldsymbol{s_p^2 = \dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$.
- **A confidence interval for $\mu_1 - \mu_2$ is:** $\boldsymbol{(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{s_p^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$.
- **The test statistic for testing $H_0\!: \mu_1 = \mu_2$ is:** $\boldsymbol{t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}}$
- We will use ***The Price is Right Rule*** anytime Table D does not happen to have the degrees of freedom we require. Make do with the closest degrees of freedom that is on Table D without going over.

LECTURE PROBLEMS FOR LESSON 10

For your convenience, here are the 6 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. To determine whether an epilepsy drug is useful in treating children with severe learning problems, 5 children with a history of learning and behavioural problems are recruited. Each child was given a placebo for 3 weeks and the epilepsy drug for the other 3 weeks. After each 3 week period, all 5 children were given an IQ test. The results below were recorded. Set up the appropriate hypothesis test to determine if the drug improves IQ at the .10 level of significance. What are your assumptions and conclusions?

Child	1	2	3	4	5
IQ after Placebo	97	106	106	95	126
IQ after Drug	113	113	101	119	126

(See the solution on page 572.)

2. Two different methods to determine the fat content in meat are being compared. Twenty-five pieces of meat are randomly selected. Each piece of meat is then cut in half and one half is randomly selected to be analyzed using Method A, the other half is analyzed using method B, the difference in the two analyses is then recorded. The mean of these 25 differences is –1.02 with a standard deviation of 1.76. We are investigating whether the methods give different results at the 5% level of significance.

(a) What experimental design is being used and what would be the appropriate hypotheses?

(A) Completely randomized design; H_0: $\mu = -1.02$ vs. H_a: $\mu \neq -1.02$.

(B) Block design; H_0: $\mu = -1.02$ vs. H_a: $\mu \neq -1.02$.

(C) Matched pairs design; H_0: $\mu = -1.02$ vs. H_a: $\mu \neq -1.02$.

(D) Completely randomized design; H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$.

(E) Matched pairs design; H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$.

(See the solution on page 573.)

(b) The rejection region for the test is:

(A) $t > 1.711$ **(B)** $|t| > 2.064$ **(C)** $|t| < 2.064$

(D) $t < -1.711$ **(E)** $|z| > 1.96$

(See the solution on page 574.)

(c) The P-value is

(A) between .005 and .0025 **(B)** between .01 and .005

(C) between .025 and .01 **(D)** .0019 **(E)** .0038

(See the solution on page 575.)

3. Twelve samples of brand A produced a mean of 16 and a variance of 36. Fifteen samples of brand B have a mean of 18 and a variance of 33. The pooled sample variance is

(A) 34.32 **(B)** 294 **(C)** 1180 **(D)** 34.33 **(E)** 5.86

(See the solution on page 585.)

4. We are interested in comparing the annual salaries of registered nurses with at least 5 years of experience in two provinces. A random sample of 23 nurses in Manitoba fitting the criteria produced a mean salary of 36 thousand dollars per year with a standard deviation of 5 thousand dollars. A random sample of 21 nurses in Ontario fitting the criteria produced a mean salary of 40 thousand dollars with a standard deviation of 8 thousand dollars.

(a) Making appropriate assumptions, is there evidence the mean salary of registered nurses with at least 5 years experience differs in these two provinces at the 4% level of significance? Define your notation, and include your hypotheses, degrees of freedom, critical value, and test statistic in your answer.

(See the solution on page 590.)

(b) What is the *P*-value for the test you conducted in (a)?

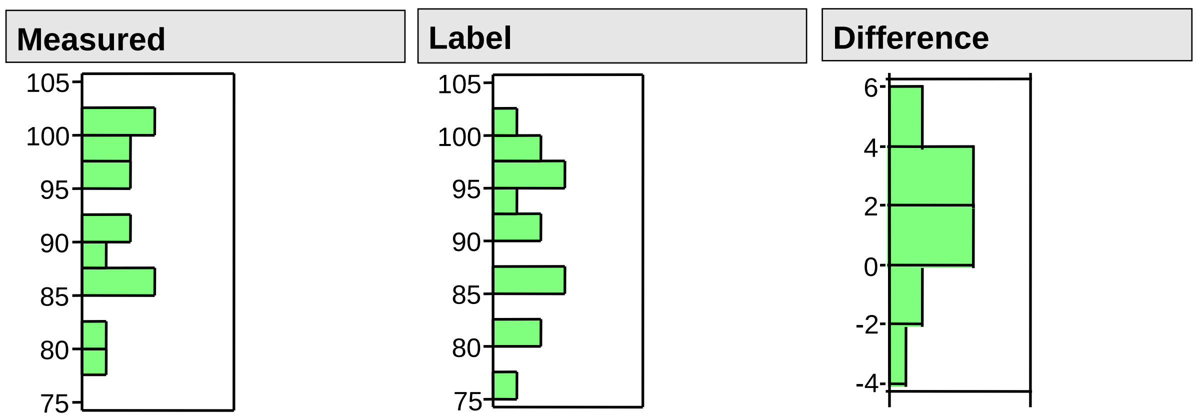
(See the solution on page 590.)

(c) Construct a 96% confidence interval for the difference in the mean salary of registered nurses with at least 5 years experience in these two provinces. Could you use this confidence interval to test the hypothesis in (a)? Explain.

(See the solution on page 592.)

5. A researcher wishes to determine if diet food labels tend to under-report the number of calories that the item contains. To investigate this, the researcher obtains a simple random sample of 15 diet food items. For each item, she measures the actual calorie content ("Measured") and compares it with the number of calories stated on the label ("Label"). From the information given, answer the questions below.

Food item	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Measured	82.0	88.0	99.5	96.5	91.4	100.0	86.5	97.6
Label	85.0	90.0	101.0	96.0	90.5	99.0	85.0	96.0
Difference	–3.0	–2.0	–1.5	0.5	0.9	1.0	1.5	1.6
Food item	**9**	**10**	**11**	**12**	**13**	**14**	**15**	
Measured	100.0	79.5	96.5	100.5	85.6	90.0	85.0	
Label	98.0	77.0	94.0	97.0	82.0	86.0	80.0	
Difference	2.0	2.5	2.5	3.5	3.6	4.0	5.0	



Moments (Measured)	
Mean	91.90667
Std Dev	7.19647
Std Err Mean	1.85812
upper 95% Mean	95.89193
lower 95% Mean	87.92140
N	15.00000

Moments (Label)	
Mean	90.43333
Std Dev	7.55661
Std Err Mean	1.95111
upper 95% Mean	94.61804
lower 95% Mean	86.24863
N	15.00000

Moments (Difference)	
Mean	1.473333
Std Dev	2.272024
Std Err Mean	0.586634
upper 95% Mean	2.731536
lower 95% Mean	0.215131
N	15.00000

(a) What would be the appropriate test to use in this problem and, with reference to the histograms provided, is the procedure valid?

(See the solution on page 595.)

(b) State the appropriate hypotheses to conduct this test defining your notation clearly.

(See the solution on page 595.)

(c) What conclusion can the researcher make at the 10% level of significance?

(See the solution on page 596.)

6. Two laboratory procedures for determining the amylase level in human body fluids are compared. The "New" method is less expensive than the "Standard" method, but may give different results. To test the procedures, samples from 20 subjects are randomly assigned 10 to each lab procedure with the following results (in units per millilitre). A variable called "New minus Standard" was also defined. The *JMP*™ analysis follows.

Standard	38	53	58	53	75	58	59	46	69	59
New	46	57	73	60	86	37	65	58	85	74
New minus Standard	8	4	15	7	11	-21	6	8	16	15

New	
Moments	
Mean	64.10000
Std Dev	15.84964
Std Err Mean	5.01210
upper 95% Mean	75.43825
lower 95% Mean	52.76175
N	10.00000

Standard	
Moments	
Mean	56.80000
Std Dev	10.49656
Std Err Mean	3.31930
upper 95% Mean	64.30885
lower 95% Mean	49.29115
N	10.00000

New minus Standard	
Moments	
Mean	7.300000
Std Dev	10.770845
Std Err Mean	3.406040
upper 95% Mean	15.005064
lower 95% Mean	-0.405064
N	10.000000

(a) What kind of test would be appropriate in order to determine if there is any difference in the methods. Use the given information to help in your explanation, and state the hypotheses and assumptions.

(b) What conclusion can we make about the methods at the 5% level of significance?

(c) Construct a 99% confidence interval for the difference between the means.

(d) Suggest an improved experimental design for this problem.

(See the solution on page 598.)

SUMMARY OF KEY CONCEPTS IN LESSON 11

- If we need to compute a sample proportion, $\hat{p}$, we are usually given a value of n followed by x, the number of yeses in our question of interest. Then $\hat{p} = \dfrac{x}{n}$.
- Recall, as we originally learned in Lesson 6:
  - The mean of $\hat{p}$ = $\mu_{\hat{p}} = p$.
  - The standard deviation of $\hat{p}$ = $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.
  - If $np \geq 10$ and if $n(1-p) \geq 10$, then $\hat{p}$ has an approximately normal distribution.
- A **statistical inference** is where we use sample statistics to *infer* (guess) what the population parameters might be like. Essentially, anytime you construct a confidence interval or test a hypothesis, you are making a statistical inference.
- In Lessons 8 and 9 we learned to make inferences about the mean, μ. Now we have learned to make inferences about the proportion, p.
- When you are asked to make a confidence interval, don't jump to conclusions. Are you asked to make a confidence interval for the *mean* or a confidence interval for the *proportion*? If it is a confidence interval for the mean, use either the z^* or t^* formula you learned in Lesson 8. If it is a confidence interval for the proportion, use this formula:

A confidence interval for *p* is $$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

- When you are asked to determine the sample size necessary to achieve a desired margin of error in a confidence interval for a *proportion*, use this formula:

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

  - If you have no idea what to use for p^*, use **the conservative estimate**. Let $p^* = 50\% = .5$.
- If you want to test a hypothesis about a *proportion*, rather than a mean, use this formula:

The test statistic for $H_0: p = p_0$ is $$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}.$$

LECTURE PROBLEMS FOR LESSON 11

For your convenience, here are the 6 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 a random sample of 450 homes it was found, in 90 cases, at least one resident watches *Survivor*. Give a 95% confidence interval for the proportion of homes that watch this program.

(A) (.199, .201) **(B)** (.163, .237) **(C)** (.169, .231)
(D) (.156, .244) **(E)** (.160, .240)

(See the solution on page 612.)

2. A government claims at least two-thirds of the people who claim refugee status are approved. An independent organization randomly selected 320 refugee claimants whose case had already been adjudicated. Of these, 180 had been given refugee status. Justifying any methods you use, is this significant evidence to reject the government's claim at the 5% level? Include the hypotheses, test statistic, critical value, *P*-value, and your conclusion in your answer.

(See the solution on page 615.)

3. It is believed a majority of farmers own their farms. In testing this hypothesis, 600 randomly selected farms revealed 315 were owned by the farmers. At the 1% level of significance, the test statistic and critical value are, respectively,

(A) $\dfrac{.025}{\sqrt{\dfrac{(.525)(.475)}{600}}}$; 1.960 **(B)** $\dfrac{.025}{\sqrt{\dfrac{(.525)(.475)}{600}}}$; 2.326

(C) $\dfrac{.025}{\sqrt{\dfrac{(.525)(.475)}{600}}}$; 2.576 **(D)** $\dfrac{.025}{\sqrt{\dfrac{(.5)(.5)}{600}}}$; 2.326

(E) $\dfrac{.025}{\sqrt{\dfrac{(.5)(.5)}{600}}}$; 2.576

(See the solution on page 616.)

4. From long experience, it is known a machine produces 30% defective tubes. After a modification is made to the machine, a random sample of 100 tubes is selected. It is found 22 of the sampled tubes are defective.

(a) In testing if the modification has improved the machine, what is the *P*-value for the test?

(A) .0268 **(B)** .0536 **(C)** .2200 **(D)** .0401 **(E)** .0802

(See the solution on page 617.)

(b) A 90% confidence interval for the proportion of defective tubes after the modification is:

(A) $.22 \pm 1.96\sqrt{\frac{(.30)(.70)}{100}}$ **(B)** $.22 \pm 1.96\sqrt{\frac{(.22)(.78)}{100}}$

(C) $.22 \pm 1.645\sqrt{\frac{(.30)(.70)}{100}}$ **(D)** $.30 \pm 1.645\sqrt{\frac{(.30)(.70)}{100}}$

(E) $.22 \pm 1.645\sqrt{\frac{(.22)(.78)}{100}}$

(See the solution on page 618.)

5. In the last election, the mayor received 60% of the vote. He wishes to determine what proportion of the electorate plans to vote for him in the upcoming election. Assuming the percentage is about the same, how many voters should be polled to estimate the proportion within 2% with 95% confidence?

(A) 61 **(B)** 106 **(C)** 172 **(D)** 122 **(E)** 2305

(See the solution on page 619.)

6. In order to estimate the unknown proportion of people who attend church on a regular basis, how large a sample do you need to draw if you want to construct a 90% confidence interval of width no more than 0.04?

(A) 1692 **(B)** 1691 **(C)** 3301 **(D)** 3328 **(E)** 3394

(See the solution on page 620.)

PREPARING FOR THE FINAL EXAM

- ❖ Once you have reviewed the lessons in this volume, you are ready to prepare for your final exam. Generally, very little that could have been on your first midterm exam will reappear on your final (you are looking at 5 or 6 multiple-choice questions from the first midterm on your final). I do think you should review the lessons in volume 2 of my book quite thoroughly though. Then use the old final exams to review. In addition, you should study the midterm exams you actually wrote. If you understand all the questions that were on your midterms, you should consider yourself well enough prepared in those areas. For those of you in distance education, your exam will be more comprehensive (although still putting more weight on the material covered in my Volumes 2 and 3) so be sure to review volume 1 of my book as well.

- ❖ **Be sure to do <u>all</u> of the Final Exams** in *Multiple-Choice Problems Set for Basic Statistical Analysis 1 (Stat 1000)* by Smiley Cheng. 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 course has definitely gotten harder over the years. **The solutions to all the old final exams are here in Appendix D of my book starting on page D-1.**

- ❖ **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.

- ❖ **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.

PREPARING FOR THE FINAL EXAM (continued)

❖ **If a multiple-choice question on your exam is strictly theory, no math at all, you should never spend more than two minutes to make up your mind what choice to make.** That will buy you time to spend on the slower calculation questions. If you don't know the answer within two minutes, face it, you don't know the answer, so it is time to trust your gut and move on. The fact is, if you do know your stuff, the right answer will present itself within 30 seconds. Never spend time thinking about a particular choice until you have gone through all the choices. Try to quickly eliminate choices that are obviously wrong and then use your remaining time considering the choices that remain.

❖ When you are going over old exams and come across theory questions, or questions where they ask you which of the statements are True or False, once you know what the correct answer is, **read the question and the TRUE choices over and over again. Never, ever, re-read a false choice.** You don't want your mind to be cluttered with things that aren't true. I say read the question and all the true choices over three times in a row at least. Wipe out of your mind whatever you thought or guessed when you first looked at the question. Otherwise, you risk remembering what you guessed, if you ever see the question again, rather than remembering what is really true. If all you have read is true statements, then, whatever you recognize on an exam is sure to be a true statement, too. Also, when you are studying, if the correct choice appears to be very similar to another choice, look closely and try to see what is actually different in those two choices and what makes the one choice correct and the other choice wrong.

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] " (MODE).

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 ⇄ y] 7 [Σ+]

5 [x ⇄ y] 10 [Σ+]

9 [x ⇄ 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.