Grant's Tutoring

BASIC STATISTICS 1

Volume 1 of 3

September 2011 edition

This volume covers the topics on
a typical first 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 CONCEPTS IN LESSON 1

- Know the difference between a **quantitative variable** and a **categorical variable**.
  - If the variable is quantitative, is it **continuous** or **discrete**?
  - If the variable is categorical, is it **nominal** or **ordinal**?
- Use a **histogram**, **stemplot**, or **boxplot** to display the distribution of a quantitative variable.
- Use a **bar chart** or **pie chart** to display the distribution of a categorical variable.
- Use a **time series** to display data that has been collected as time goes by in order to identify trends, if any, in that data.
- When making a stemplot, we can choose to make a **split stemplot** to count by fives rather than tens if that will display the data better. We can also **trim the data** (cut away the last digit) if that will make the number of stems more manageable.
- When making a boxplot, we can choose to use a **modified boxplot** (also called an **outlier boxplot**) to display any outliers by dots in order to not exaggerate the length of the whiskers.
- If we want to compare two separate distributions graphically, we can use a **back-to-back stemplot** or **side-by-side boxplots**. Side-by-side boxplots can even be used to compare three or more distributions on one graph.
- Discuss the **shape**, **centre** and **spread** of a quantitative variable when summarizing its distribution.
  - **Shape:** Consulting any graphs you have, are there any **outliers**? How many **peaks** are there, if any? Is the distribution **symmetric**, **left-skewed** or **right-skewed**?
  - **Centre:** Three measures of centre are **mean**, **median** and **mode.** For symmetric data, the mean and median are the same. **When data is skewed or has outliers, the mean is pulled away from the median in the direction of the skew or the outliers. A median is a more trustworthy measure of centre in these situations.**
  - **Spread:** Three measures of spread are **range**, **interquartile range** and **standard deviation**. When using a median to measure the centre, use the range and interquartile range to measure the spread. When using a mean to measure the centre, use the standard deviation to measure the spread.
- The **five-number summary** is ***Min, First Quartile, Median, Third Quartile, Max***.

- Know how to use both **the $\frac{n+1}{2}$ Rule** and **the Finger Method** to locate the median and/or quartiles and use whichever you prefer for a given data set. Remember, use a cleaver to slice the ordered data in half at the median's location. If that cleaver hits a piece of data, it has been smashed to smithereens, excluding it from both the lower and upper halves when it comes to counting your way to the quartiles.
- The **1.5 × *IQR* Rule** can be used to determine outliers. Any value that is lower than $Q_1 - (1.5 \times IQR)$ or higher than $Q_3 + (1.5 \times IQR)$ is an outlier.
- The **sample mean** is denoted $\bar{x}$ and the **sample standard deviation** is denoted s. The **sample variance** is s^2.
- Memorize these formulas: $\bar{x} = \frac{\sum x}{n}$ and $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$.
- If you are ever asked to transform data by adding, subtracting, multiplying, and/or dividing each value by some constant, remember how that transforms any measure of centre or spread. **Any transformation you are performing on the data will also transform the measures of centre in the same way. However, only multiplication or division transforms a measure of spread. Addition or subtraction by a constant has no effect on spread.**
  - If we are transforming X into Y via the formula $Y = AX + B$, then
    - ***Centre of* $Y = A$ (*Centre of* X) $+ B$**

      (The mean, median, or mode would be transformed this way.)
    - ***Spread of* $Y = A$ (*Spread of* X)**

      (We do not add the constant "B"; standard deviation, range, or interquartile range would be transformed this way.)

LECTURE PROBLEMS FOR LESSON 1

For your convenience, here are the 20 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. A survey asked the following questions:

- What is your eye colour?
- Rate your boss on a scale from 1 to 10 where 1 is awful and 10 is wonderful.
- How much do you weigh (in kilograms)?
- What is the 3-digit area code for your home phone number?
- How far do you live from work or school? (less than 5 km, between 5 km and 10 km, more than 10 km)
- What method of transportation to work/school do you normally use? (car, bus, bike, walking, other)
- What is your annual income?
- How many times in a typical month do you eat at a restaurant (including take-out or delivery)?
- Did you vote in the last federal election?

For each of the above survey questions, identify if the variable is quantitative (if so, is it discrete or continuous?) or categorical (if so, is it ordinal or nominal?). In addition, what graph could you use to display the data gathered in each case?

(See the solution on page 14.)

2. You record the age, marital status, earned income, and sex of a sample of 1463 people. The number of variables you have recorded is:

(A) 1463.
(B) five: age, marital status, income, sex, and number of people.
(C) four: age, marital status, income, and sex.
(D) two: age and income; marital status and sex are not variables because they are not a numerical quantity.
(E) none: because no one has any business asking such personal questions.

(See the solution on page 15.)

3. The table below shows the number of people (in millions) living on farms in a certain country over the years.

Year	1910	1920	1930	1940	1950	1960	1970	1980	1990
Population	1.6	2.9	2.8	2.5	1.8	1.3	1.3	0.9	0.6

Construct a time series for this data and comment on what you see.

(See the solution on page 16.)

4. The test scores (out of 100) for a random sample of 50 students who wrote a statistics midterm exam are as follows:

75	88	47	66	78	45	73	66	77	100
64	61	77	87	66	92	86	57	80	70
52	84	80	79	66	92	72	83	50	84
65	75	77	79	79	57	63	51	44	59
84	77	44	81	61	77	57	75	3	52

(a) Construct a frequency table. *(Solution on page 23.)*

(b) Construct a relative frequency table. *(Solution on page 24.)*

(c) Construct a histogram. *(Solution on page 26.)*

(d) Construct a stemplot. *(Solution on page 28.)*

(e) Construct a split stemplot. *(Solution on page 30.)*

(f) Discuss the shape of the distribution. Are there any outliers? *(Solution on page 36.)*

(g) Find the median test score. *(Solution on page 37.)*

(h) Find the first and third quartiles. *(Solution on page 40.)*

(i) Find the Range and Interquartile Range. *(Solution on page 41.)*

(j) State the five-number summary. *(Solution on page 42.)*

(k) Justify the outliers (if any) mathematically. *(Solution on page 43.)*

(l) Draw a boxplot. *(Solution on page 44.)*

(m) Draw a modified (outlier) boxplot. *(Solution on page 46.)*

(n) Find the mode of the distribution. *(Solution on page 47.)*

(o) Find the mean of the distribution. *(Solution on page 49.)*

(p) Find the standard deviation of the distribution. *(Solution on page 51.)*

5. Find the first, second and third quartiles for the data sets below.

(a) 3, 8, 84, 51, 23, 13, 18, 15, 18, 4, 16, 4, 9.

(See the solution on page 55.)

(b) 24, 27, 26, 22, 23, 27, 22, 18, 21, 10.

(See the solution on page 60.)

6. Find, by hand, the mean, variance and standard deviation of this data (show all your work): 6, 12, 9, 8, 5, 14, 2

(See the solution on page 61.)

7. In order to analyze the overall pattern of a distribution, the three things we should discuss are:

(A) the mean, median and mode.

(B) the interquartile range, range and variance.

(C) the number of peaks, the outliers and the shape of the distribution.

(D) the shape of the distribution, the centre and the spread.

(E) the outliers, the influential observations and the lurking variables.

(See the solution on page 63.)

8. The annual salary (in thousands of dollars) of a random sample of male and female workers in the construction industry is shown below. Construct a back-to-back stemplot for this data and discuss your observations.

Males: 29 32 32 27 46 24 45 50 47 36 35 30 28 88 37 38 52 43
Females: 28 39 29 23 32 29 18 22 38 40 26 17 33

(See the solution on page 67.)

9. The five-number summary for a sample of 60 observations is 27, 45, 50, 62, 101. We can say:

(A) The sample is clearly symmetrical.

(B) The mean is 50.

(C) There are no outliers.

(D) Any data values below 19.5 or above 87.5 are outliers.

(E) Any data values below 24.5 or above 75.5 are outliers.

(See the solution on page 68.)

10. The first and third quartiles for a random sample of 200 observations are 48 and 77, respectively. Three of the observations are 3, 120, and 121. Consider these statements:

(I) 3 is an outlier.
(II) 120 is not an outlier.
(III) 121 is an outlier.

(A) Only (I) is true. **(B)** Only (II) is true.
(C) Only (III) is true. **(D)** Only (I) and (III) are true.
(E) (I), (II) and (III) are all true.

(See the solution on page 69.)

11. In measuring the effectiveness of a new drug treatment for cancer patients, 11 patients were tracked after the treatment to see if the cancer returned. The number of years they were cancer free was recorded, and they were given "*N*" for no cancer if they went at least 10 years without a recurrence. The data was: 3.7 9.8 5.5 *N* 1.2 *N* 7.8 8.8 7.5 *N* 2.9. The median of the data is

(A) 6.5 **(B)** 7.8 **(C)** 7.5 **(D)** *N* **(E)** impossible to determine

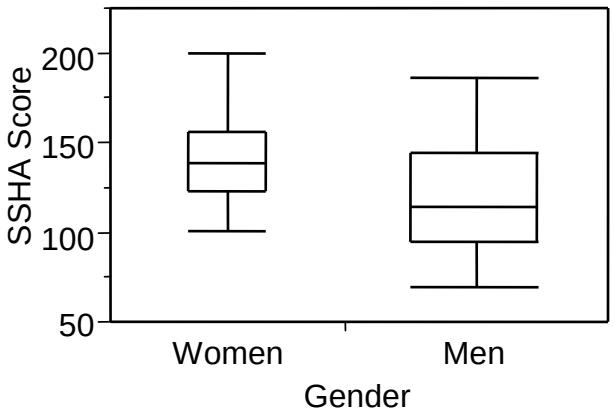
(See the solution on page 70.)

12. The times (in seconds) for 9 subjects to complete a task were as follows: 16 7 3 21 12 7 26 22 45. The interquartile range and variance, respectively, are equal to:

(A) 19; 12.9 **(B)** 17; 12.9 **(C)** 15; 165.5 **(D)** 17; 165.5 **(E)** 19; 165.5

(See the solution on page 72.)

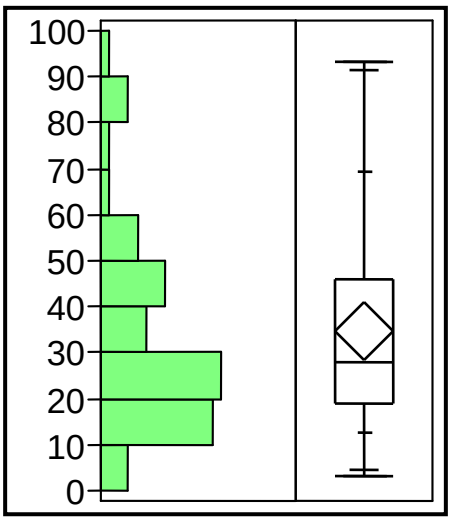
13. The survey of study habits and attitudes (SSHA) is a psychological test that measures the motivation, attitude toward school, and study habits of students. Scores range from 0 to 200. *JMP*™ produced side-by-side boxplots at right to compare the scores on the test for a random sample of men and women at the University of Manitoba. Which statement below is <u>false</u>?



(A) The women's scores have a narrower spread than the men's.
(B) The interquartile range for the men's scores is approximately 50.
(C) More men wrote this test than women.
(D) The women tend to score higher than the men on this test.
(E) The mean test score for the men is less than their median score.

(See the solution on page 73.)

14. At right is the distribution of annual income (in thousands of dollars) of a sample of Canadian adult males as displayed by *JMP*™.



(a) From this information we would conclude:
- **(A)** the sample was clearly not random.
- **(B)** the distribution is symmetrical.
- **(C)** the distribution is left-skewed.
- **(D)** the distribution is right-skewed.
- **(E)** men make more money than women.

(See the solution on page 75.)

(b) The best way to summarize this data is:
- **(A)** the mean and standard deviation.
- **(B)** the mean and variance.
- **(C)** the mean, the median and the mode.
- **(D)** the five-number summary.
- **(E)** none of the above.

(See the solution on page 75.)

(c) The mean and median, respectively, of this distribution are approximately
(A) 28; 35 **(B)** 35; 28 **(C)** 45; 32 **(D)** 32; 45 **(E)** 40; 32

(See the solution on page 76.)

15. The table below shows the scores for the 60 people who participated in a recent 18-hole golf tournament (the winner scored −1 or 1 below par).

Score	−1	0	+1	+2	+3	+4	+5	+6	+7
Frequency	1	0	1	4	3	7	10	25	9

Looking at the distribution of scores, we can conclude:
- **(A)** The median is +3 and the mean is higher.
- **(B)** The median is +3 and the mean is lower.
- **(C)** The median and mean are both +3.
- **(D)** The median is +6 and the mean is lower.
- **(E)** The median and mean are both +6.

(See the solution on page 78.)

16. The stemplot at right shows the number of weeks a sample of 40 patients in Manitoba had to wait before receiving hip replacement surgery. The five-number summary for this data is:

(A) 7, 26, 35, 42, 52
(B) 0, 26.5, 35, 42.5, 69
(C) 7, 26, 35, 42, 69
(D) 7, 26.5, 35, 42.5, 69
(E) 7, 26.5, 35, 40.5, 69

(See the solution on page 79.)

Stem	Leaf
7	
6	9
6	
5	
5	12
4	67789
4	0011234
3	555589
3	13344
2	55667789
2	2
1	589
1	4
0	7

17. The mean daily high temperature for June in a particular city is 20 °C with a standard deviation of 2.7 °C. What would the mean and standard deviation be in Fahrenheit? (Hint: The formula $F = 1.8C + 32$ converts Celsius into Fahrenheit.)

(A) 68 & 3.86 **(B)** 68 & 36.86 **(C)** 68 & 4.86 **(D)** 68 & 34.7
(E) It is impossible to determine without the original data.

(See the solution on page 83.)

18. After a particularly difficult exam, the average mark was 47%, the median was 40%, and the interquartile range was 10%. The prof decides to add 20% to everyone's mark. After this addition, which of the following statements is FALSE?

(A) The first quartile could now be as low as 50%.
(B) The third quartile could now be as high as 70%.
(C) The median mark is now 60%.
(D) The interquartile range is now 30%.
(E) The average mark is now 67%.

(See the solution on page 84.)

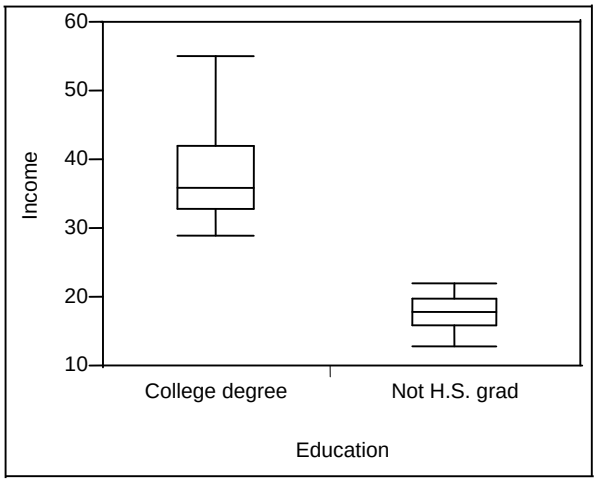
19. A student's average after writing six tests is 68%. She just got her seventh test back with a mark of 82%. What is her average mark now?

(A) 74% **(B)** 71% **(C)** 75% **(D)** 72% **(E)** 70%

(See the solution on page 85.)

20. Below is a *JMP*™ printout comparing the annual income (in thousands of dollars) of two random samples of 35 year-old full-time workers. One sample is of 10 people who have a college degree, the other is 11 people who did not graduate from high school.

Oneway Analysis of Income By Education



Quantiles

Level	Minimum	10%	25%	Median	75%	90%	Maximum
College degree	29	29.1	33	36	42	54.3	55
Not H.S. grad	13	13.2	16	18	20	21.8	22

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
College degree	10	38.1000	7.97844	2.5230	32.393	43.807
Not H.S. grad	11	18.0000	2.82843	0.8528	16.100	19.900

What is the mean and median income of all 21 workers?

(A) \$27,571 and \$27,000, respectively
(B) \$27,571 and \$22,000, respectively
(C) \$27.571 and \$27, respectively
(D) \$28,050 and \$27,000, respectively
(E) \$28,050 and \$22,000, respectively

(See the solution on page 88.)

SUMMARY OF KEY CONCEPTS IN LESSON 2

- A regression analysis can be performed on (x, y) data pairs provided **both x and y are quantitative variables**.
- The **explanatory variable** is x and the **response variable** is y. We believe x can explain y's response. We hope to use x to predict y.
- We use a **scatterplot** to see if x and y have either **a positive or a negative association** (or neither), and if the trend (if any) is **linear** or **nonlinear**.
  - A rising trend is a positive association (y gets larger as x gets larger).
  - A falling trend is a negative association (y gets smaller as x gets larger).
- If we do believe we have a linear trend, we can confirm it by computing the **correlation coefficient, r.**
  - $$r = \frac{1}{n-1}\sum\left[\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)\right]$$
  - **r measures the strength and direction of the linear relationship between x and y.**
  - **$-1 \le r \le 1$; the closer r is to -1 or 1, the stronger the linearity.**
  - **r has no units.**
  - r will be identical regardless of which of the two variables is considered x and which is considered y.
  - If you change your mind as to which units to use when measuring x and y, the value of r will not change; r is independent of the units used for x and y.
- The **coefficient of determination** is **r^2**
  - **r^2 is the percent of the y's variation explained by the regression with x.**

❖ The **least-squares regression equation** is $\hat{y} = a + bx$. This equation draws the best fit line through our scatterplot.

- $b = r\dfrac{s_y}{s_x}$ and $a = \bar{y} - b\bar{x}$
- The **intercept, *a*,** is where the regression line intercepts the "$x = 0$" line
- ***a* is the predicted value of *y* when $x = 0$.**
- The **slope, *b*,** is ***Rise/Run***; as x runs 1 unit, y rises (or falls) b units.
- ***b is* the amount *y* changes for each additional unit of *x*.**

❖ Extending the regression line beyond the scatterplot in an attempt to predict y for an x value beyond the data range is an **extrapolation**. Extrapolations are unreliable because there is no guarantee the linear trend continues.

❖ **Residual** $= e = y - \hat{y}$

- A residual is the difference between the observed value of y and the predicted value of y.
- If x and y truly have a linear relationship, a residual plot will have no pattern.
- If a residual plot has a pattern, x and y could not have a linear relationship; linear regression should not have been attempted.
- The sum of the residuals is 0.
- The sum of the squares of the residuals has been minimized. That is what we mean by "least-squares" regression or the method of "least-squares".

❖ When we find an unexpected correlation between x and y, it is undoubtedly due to a **lurking variable**.

- A strong correlation between children's *test scores* and *shoe size* is because *age* is a lurking variable, for example.

❖ Any observation on a scatterplot that is far to the left or right of the cluster could be an **influential observation**. If removing that observation considerably changes the least-squares regression equation (changes the value of a and/or b), it is definitely an influential observation.

LECTURE PROBLEMS FOR LESSON 2

For your convenience, here are the 9 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. A researcher believes that the fuel economy of a car can be predicted by its speed. A car was driven on an oval track for two hours at five different speeds and then had its fuel economy measured. The results, together with their means, were tabulated as shown below.

						Mean
Fuel Economy (mpg)	19	13	12	8	7	11.8
Speed (mph)	10	20	30	40	50	30

(a) Identify the explanatory and response variables. *(See the solution on page 113.)*

(b) Make a scatterplot and comment on its pattern (if any). *(Solution on page 115.)*

(c) Compute the correlation coefficient and interpret it. *(Solution on page 116.)*

(d) Compute the coefficient of determination and interpret it. *(Solution on page 119.)*

(e) Determine the least-squares regression equation and draw its line on your scatterplot found in (b). *(Solution on page 122.)*

(f) Interpret the slope of the least-squares regression line. *(Solution on page 123.)*

(g) Predict the fuel economy at a speed of 35 mph, or explain why this is not a reasonable calculation. *(Solution on page 124.)*

(h) Predict the fuel economy at a speed of 80 mph, or explain why this is not a reasonable calculation. *(Solution on page 124.)*

(i) Without calculation, predict the fuel economy at a speed of 30 mph.
(See the solution on page 125.)

(j) Compute the residual for the above data when the speed was 20 mph.
(See the solution on page 126.)

(k) At what speed would we predict a fuel economy of 9 mpg? *(Solution on page 129.)*

2. For which set of data below would the correlation be -1?

(A)

x	5	10	15
y	12	18	24

(B)

x	5	10	15
y	5	10	15

(C)

x	5	10	15
y	12	10	8

(D)

x	5	10	15
y	12	8	24

(E)

x	5	10	15
y	28	18	14

(See the solution on page 130.)

3. Which of the following statements does not contain a blunder?

(A) A study found a strong positive correlation (0.89) between the colour of a person's shirt and the number of mosquito bites received in a 3 hour period.

(B) There is a strong negative correlation (-1.23) between the number of alcoholic beverages consumed the night before an exam and the mark achieved on the exam.

(C) The correlation between a man's leg length and his time to climb two flights of stairs is 0.32 minutes.

(D) The correlation between the height of a student and their GPA is -0.53 suggesting that taller students tend to have a lower GPA than shorter students.

(E) The correlation between elementary school children's shoe size and their score on a standard spelling test is 0.85 suggesting that having big feet causes you to spell better.

(See the solution on page 132.)

4. In each case below a regression analysis was performed on data. An alteration was made to the data, and a regression was done again. Which of the following alterations would give two different values of r, the correlation coefficient?

(A) A regression on people's height (in inches) and weight (in pounds) was performed. Now the same data is converted into metric units (centimetres and kilograms, respectively), and the regression is performed again.

(B) A regression on height versus weight used height as the explanatory variable. Now the same data is used, but we use weight as the explanatory variable.

(C) A regression on automobile deaths for the years 1991, 1992, to 2004 was performed. Now the same data is used, but we decide to code 1991 as year 1, 1992 as 2, etc.

(D) A study on crop yield (bushels per acre) versus amount of fertilizer (litres per acre, l/a) was designed. Ten farms used no fertilizer; ten used 1 l/a; ten used 2 l/a; ten used 3 l/a; and ten used 4 l/a. The crop yields for the 50 farms were recorded and a regression analysis was performed. Now the same data is used, but instead of using the individual farms, the mean crop yield for the ten farms receiving the same amount of fertilizer is computed, giving us an average crop yield for no fertilizer, 1 l/a, 2 l/a, 3 l/a, and 4 l/a. A regression on mean crop yield versus amount of fertilizer is performed on these 5 values.

(E) None of the above.

(See the solutionon page 134 .)

5. Below is the *JMP*™ output for 12 random samples of a particular model of used car. The selling price of the particular car (in dollars) and the distance travelled according to its odometer reading (in thousands of kilometres) were recorded.

Variable	Mean	Std Dev	Correlation	Number
Odometer Reading	48	6.96	-0.87	12
Selling Price	6500	1997.52		

(a) The appropriate least-squares regression equation would be:

(A) $\hat{y} = 67.70 + 0.0030x$ **(B)** $\hat{y} = 18{,}485.12 - 249.69x$

(C) $\hat{y} = -5485.12 - 249.69x$ **(D)** $\hat{y} = 67.70 - 0.0030x$

(E) $\hat{y} = 28.5 - 0.0030x$

(See the solution on page 136.)

(b) Which one of the following statements is TRUE?

(A) A car with a lot of kilometres on its odometer tends to be more expensive.

(B) 87% of the selling price is predicted by the regression equation.

(C) -87% of the selling price is predicted by the regression equation.

(D) The selling price drops by 25 cents per kilometre on the odometer.

(E) The selling price drops by 250 dollars per kilometre on the odometer.

(See the solution on page 137.)

6. A least-squares regression equation for a random sample is $\hat{y} = 120 - 12.5x$. The explanatory variable is the amount of alcohol consumed (in ounces) and the response variable is the score on a physical coordination test. Participant A in the study consumed 2 ounces of alcohol and scored 100 on the test. Which of the following statements is false?

(A) We would predict a person who drinks no alcohol would score 120 on the test.

(B) For each additional ounce of alcohol consumed, we would predict the score would drop 12.5 points.

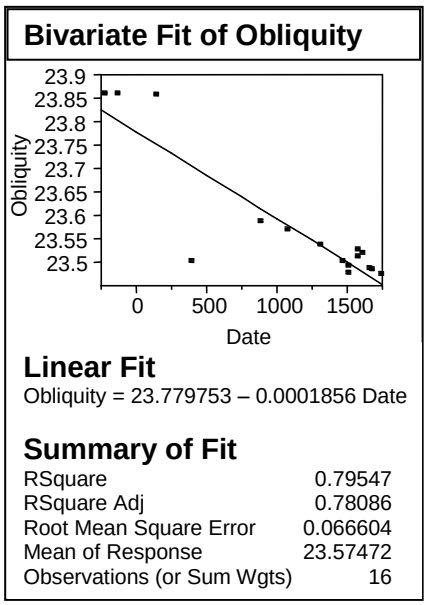
(C) The residual for Participant A is 5.

(D) The sum of the squares of the residuals has been minimized.

(E) There is clearly a strong negative linear correlation between the score on a physical coordination test and the amount of alcohol consumed.

(See the solution on page 138.)

7. The angle of the earth's rotation is called its "obliquity" and is measured in degrees. Measurements have been made at various dates in earth's history. Let us assume that the measurements were accurate. A regression of the obliquity by date was performed by *JMP*™ and the results are at right. Which of the following statements are false?



(A) The measurement made about the year 490 is an outlier.

(B) Each year the obliquity increases by 0.0001856 degrees.

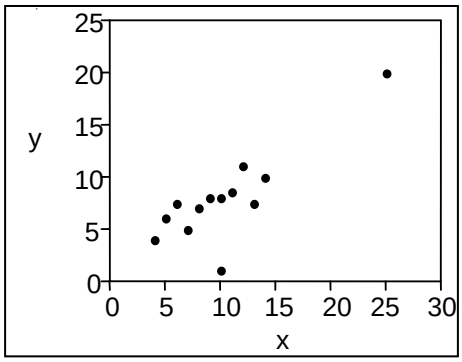
(C) The correlation between date and obliquity is -0.8919.

(D) The measurement at about the year 900 has a negative residual.

(E) About 79.5% of the variation in the earth's obliquity can be explained by this regression line.

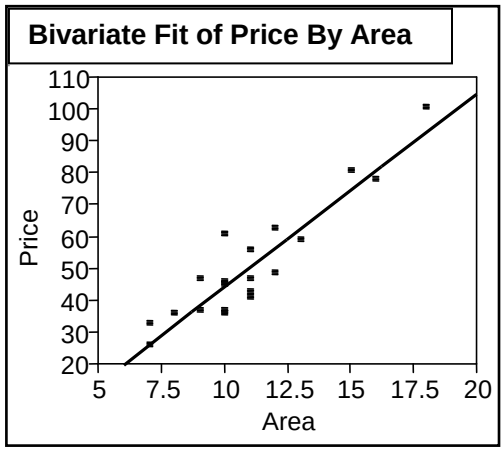
(See the solution on page 140.)

8. For the scatterplot shown at right, circle and label any point you believe is an outlier, influential observation, or both.



(See the solution on page 142.)

9. At right is a scatterplot and least-squares regression line examining the effect the area of a condominium (in hundreds of square feet) has on the selling price (in thousands of dollars). The regression line is



(A) $\hat{y} = 20 + 6x$

(B) $\hat{y} = -17 + 6x$

(C) $\hat{y} = 20 + 40x$

(D) $\hat{y} = 5 - 6x$

(E) $\hat{y} = -50 + 12x$

(See the solution on page 143.)

SUMMARY OF KEY CONCEPTS IN LESSON 3

- In order to get insights about a population we must select a **representative sample**.
- Bad sampling methods which inevitably do not give us representative samples are **voluntary response samples** and **convenience samples**.
- Good sampling methods are **simple random samples**, **stratified random samples**, and **multistage samples**.
- Possible sources of bias in a sample are **question bias**, **response bias**, **nonresponse**, and **undercoverage**.
- Depending on what a researcher wants to examine, they might use a **survey or opinion poll**, an **observational study**, or an **experiment**.
  - An **observational study** does not impose conditions on the **experimental units or subjects** but merely observes the responses to the conditions that were already present or chosen by the subjects themselves.
  - An **experiment** imposes specific treatments on the **experimental units or subjects**.
- The three principles of experimental design are **randomization**, **replication** and **control of outside factors**.
- To study the **factor(s) of interest** the researcher breaks each factor up into two or more **factor levels** in order to compare the effects on the **response variable**.
  - If there are two or more factors, multiply the factor levels together to determine the number of **treatments** the experiment will require.
- The most straightforward experimental design is the **completely randomized design**. More specialized experiments are **block design** and **matched pairs design**.
  - In a **block design**, we separate our experimental units into two or more distinct blocks. These blocks are based on some inherent characteristic within the units and cannot just be created randomly. We then perform the entire experiment on Block 1, repeat the experiment on Block 2, and so on for all the blocks.
  - In a **matched pairs design** each unit is given both treatments either at the same time if possible, or one first, the other later and the difference in response is recorded.

LECTURE PROBLEMS FOR LESSON 3

For your convenience, here are the 11 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. Some television stations take quick polls of public opinion by announcing a question during a program and asking the viewers to call one of two telephone numbers to register their opinions as "yes" or "no". Telephone companies make available "1-900" numbers for this purpose and dialling such a number results in a small charge to the caller's phone bill. One such call-in poll found that 73% of those who called in favour the "Free-Guard Zone" in Canadian Curling. The results of this poll suggest approximately

(A) 73% of all curlers favour using the "Free-Guard Zone".
(B) 73% of all curlers watching that program favour using the "Free-Guard Zone".
(C) 73% of all people watching the program favour the "Free-Guard Zone".
(D) 73% of all people watching the program are curlers.
(E) none of the above.

(See the solution on page 161.)

2. In a telephone survey, after being asked, "Did you vote for the Liberals in the last provincial election?", the person refuses to answer and hangs up the telephone. This is an example of

(A) response bias **(B)** question bias **(C)** voluntary response sampling
(D) nonresponse **(E)** confounding

(See the solution on page 161.)

3. In a telephone survey, a person is asked to say which category their income falls into and they say it is between \$50,000 and \$75,000 when, in fact, they earn less than \$30,000. This is an example of

(A) response bias **(B)** question bias **(C)** voluntary response sampling
(D) nonresponse **(E)** confounding

(See the solution on page 162.)

4. To determine the average age of students attending a particular university, a researcher selects simple random samples of 50 students from each of the university's faculties. This is an example of

(A) an observational study **(B)** multistage sampling **(C)** block design
(D) stratified random sampling **(E)** matched pairs design

(See the solution on page 162.)

5. To determine the average age of faculty members at a university, a researcher randomly selects 10 faculties and randomly selects three departments from each of those 10 faculties. The researcher then takes simple random samples of 5 faculty members from each of those selected departments. This is an example of

(A) an observational study **(B)** multistage sampling **(C)** block design
(D) stratified random sampling **(E)** matched pairs design

(See the solution on page 162.)

6. Does regular exercise reduce the risk of a heart attack? Here are two ways to study this question. Explain clearly which design is more trustworthy.

(I) A researcher finds 4000 men over 40 who have not had heart attacks. She interviews the men and splits them into 2 groups, those who exercise regularly and those who do not, and she follows both groups for 5 years.

(II) Another researcher finds 4000 men over 40 who have not had heart attacks and are willing to participate in a study. She randomly assigns 2000 of the men to a regular program of supervised exercise. The other 2000 continue their usual habits. The researcher follows both groups for 5 years.

(See the solution on page 165.)

7. Identify the experimental units, factor(s), factor levels, the number of treatments, the blocking variable (if applicable) and the response variable in each of the following experiments. Also, identify what kind of experimental design is being used. Finally, draw a diagram outlining the experiment.

(a) A company that produces 35 mm film for cameras is interested in determining the effect of a new film coating on the contrast of the pictures that are produced. A total of 24 Pentax PZ-10 cameras are available for the experiment. Some of the cameras will use film with the standard coating, while other cameras will use the film with the new coating. A measure of contrast will then be determined for one picture from each roll of film used. *(See the solution on page 173.)*

(b) A professor from the Animal Science Department at the State University wants to determine the effects of varying the type of dog food and the time during which the food is served to the dog on the body fat of the dog. A total of 48 dogs, all from the same breed and roughly the same age will be used in the experiment. Three types of dog food (A, B and C) and two serving times (early and late) will be considered. The reduction in fat after 2 weeks will be measured on each dog.

(See the solution on page 175.)

(c) A heating and air conditioning company is examining the effect of using a new air conditioner compressor on the time required to cool 2000-square-foot houses by 10 degrees. A total of 24 houses (14 bungalows and 10 two-storey houses) will be used in the experiment. The air conditioners will be fitted with one of the two compressors (either the compressor that is currently in use or the new compressor).

(See the solution on page 178.)

(d) An education specialist is interested in determining the effectiveness of different teaching methods on the time required for bus drivers to learn how to operate the *TourNado*, a new highway bus. A total of 63 bus drivers, each with 10 years of experience driving highway buses (21 from Greyhound Lines, 24 from Grey Goose Lines, 18 from STC Lines) will participate in the experiment. Three types of teaching methods will be considered (75% in- class instruction and 25% training on the bus, a 50-50 split between in-class and bus instruction, 25% in-class instruction and 75% training on the bus). Drivers will then be tested on the operating procedures, and test scores will be compared.

(See the solution on page 179.)

(e) A consumer magazine wants to determine if there is any difference in the wear between Goodyear and Michelin tires under normal use. One hundred people are willing to participate in an experiment where their cars will be fitted with both of the tires to be tested. They will then be free to drive their cars as they always would for 2 years. At the end of this time, the wear in the tires will be measured. Assume no tires get punctured or in any way need to be replaced during the experiment.

(See the solution .on page 181)

8. An experiment is performed to assess weight gain of pigs on two types of feed. 40 pigs are used in the experiment, 10 from each of 4 farms. The two feeds are randomly assigned to 5 pigs on each farm. The response, block, treatment, and experimental unit are respectively:

(A) pig, feed, weight gain, farm
(B) weight gain, farm, feed, pig
(C) pig, farm, feed, weight gain
(D) pig, feed, farm, weight gain
(E) weight gain, feed, farm, pig

(See the solution on page 183.)

9. A medical researcher wishes to determine the effect of the stretching program and the stretching time on the time required for a person to run 100 metres. Undergraduate physical education students are available for the experiment. Three stretching programs (A, B and C) and two stretching times (5 minutes and 15 minutes) will be considered. The time (in seconds) required for each student to run 100 metres will be measured.

(a) The experimental units, factor(s), and response are, respectively

(A) stretches, length of the race, the time to run 100 metres
(B) stretching program and stretching time, the students, the time to run 100 metres.
(C) the students, the 5 stretches, the time to run 100 metres
(D) the students, the stretching program and stretching time, the time to run 100 metres
(E) none of the above.

(See the solution on page 184.)

(b) To get adequate replication, the researcher wants 15 students randomly assigned to each treatment. The number of treatments in the experiment and the total number of students needed for the experiment are, respectively

(A) 6; 90 **(B)** 6; 60 **(C)** 5; 75 **(D)** 6; 15 **(E)** 3; 45

(See the solution on page 184.)

10. An experimenter wishes to test whether or not two types of fish food (A & B) promote equivalent weight gain during a 2-month feeding program. The experimenter has 2 identical fish tanks (1 & 2) to put fish in and is considering methods for assigning the 40 tagged fish to the tanks. To properly assign the fish, one procedure would be to:

(A) Put all the odd tagged numbered fish in one tank, the even in the other, and give food type A to the odd numbered ones.
(B) Obtain pairs of fish whose weights are virtually equal at the start of the experiment and randomly assign one to the group tank 1, the other to tank 2 with the feed assigned at random to the tanks.
(C) Obtain pairs of fish whose weights are virtually equal at the start of the experiment and put the heavier of the pair into tank 2.
(D) Assign the fish at random to the tanks and give feed type A to tank 1.
(E) Not to proceed as in (B), because using the initial weight is a non-random process. Use the initial length of the fish instead.

(See the solution on page 185.)

11. The Helsinki Heart Study was a randomized comparative experiment that asked whether taking the drug gemfibrozil (which lowers cholesterol levels) would prevent heart attacks in middle-aged men with high blood cholesterol. In this study, 2051 men took gemfibrozil; 56 of them had heart attacks during the five years of the study. Another 2030 men took a placebo, and 84 of these men had heart attacks. The **Type** of experimental design that was used and the **Response** variable are, respectively:

(A) A Matched Pairs design; amount of gemfibrozil taken.
(B) A randomized Block design; whether or not the subject had a heart attack.
(C) A simple random sample; amount of cholesterol of the subject.
(D) A completely randomized design; level of cholesterol of the subject.
(E) A completely randomized design; whether or not the subject had a heart attack.

(See the solution on page 186.)

PREPARING FOR THE FIRST MIDTERM EXAM

- If you have done all of the homework from all 3 lessons, you are now ready to start preparing for your first midterm exam. **Be sure to do <u>all</u> of the Term Tests** 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 several questions in the old midterms must be omitted (you will cover those topics later in the course). Again, 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 these term tests are here in Appendix C of my book starting on page C-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. **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: "[MODE] [DRG]") 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] [MODE] [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] [STAT DATA] Use your arrow keys to ensure "*CLRDATA*" is underlined then press [ENTER =]

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

[2nd] [STAT DATA] 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] [STAT DATA] 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] [STAT 1 x ⇄ y]

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] [$\bar{x}$ / x^2]

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

[2nd] [σ_{xn-1} / $\sqrt{x}$]

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] [STAT 2 Σ+]

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] [COR 4]

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

[2nd] [ITC 4]

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

[2nd] [SLP 5]

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 "S*x*").

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.