

EXPERIENTIAL STATISTICS

(TI - 83 BASED)

Dr. John Loase

**Sponsored as an outgrowth of National Science
Foundation program 39155966**

Mathematical (Statistical Modeling)

INTRODUCTION

This manual is an outgrowth of a National Science Foundation grant in which I had the privilege of writing an applied statistics book for advanced mathematics students. The emphasis was on real world applications of statistical modeling and required several courses above the Calculus as prerequisites.

This short book is designed for all students to experience the importance of mathematics for their future personal and professional life. It is not written for the advanced mathematics student and focuses on the major topics of finance, correlation, linear regression, confidence intervals, and elementary hypothesis testing. These topics are the key elements that find their applications throughout the fields of education, business, and the health and social sciences.

This compact book includes step-by-step directions for performing the statistical and financial procedures with the TI-83 calculator. This reduces difficulties with arithmetic and algebra. It also teaches to the students' strengths – computers, calculators, and using existing programs to solve problems. We also will emphasize practice – not theory. We hope that you will learn through this course that major decisions in your life can be made with the help of Statistics. MATH IS POWER. This book has ten short chapters. Each chapter concludes with an experimental activity for the student to learn the material as an active participant. It is suggested that the student write a short paper describing his/her results of each of the ten recommended projects.

As my graduate professor Anthony Ventriglia once said,
"MATHEMATICS IS A PARTICIPANT SPORT - NOT A SPECTATOR SPORT."

Become involved. Experience the power of STATISTICS!

MATHEMATICS OF FINANCE

CHAPTER ONE - SIMPLE INTEREST

If you have \$1,000 in a savings account at 3% interest per year, you may be accused of being simple. To calculate your interest, just multiply $\$1,000 \times (.03) = \30 .

You can do better. These days some mutual funds are boasting of 20+ % returns. Of course, there is risk. Bonds return in the neighborhood of 6% interest, which would yield $\$1,000 \times (.06) = \60 per year. And you can shop and find bonds with very little risk.

An understanding of interest and developing a plan that maximizes return on your investment with a risk level appropriate for you is vital to your financial health. The real money is found when you compound interest. Your money grows exponentially. But before you deal with compound interest, let us get a little better at simple interest.

Suppose you have \$10,000 in a mutual fund. They advertise an average annual return of 27%. What can you expect for six months in interest?

Using the simple formula, $I = p \times r \times t$, p = principal - the amount of money you have in your account, r = your annual (yearly) rate of return or interest, t = time (in years).

Let us use \cdot for the \times sign. That way, we don't confuse \times - a letter for an unknown with \times - the multiplication sign.

For your mutual fund, your interest is calculated as follows:

$$27\% = .27$$

$$I = \$10,000 \cdot (.27) \cdot (6/12)$$

$I = \$1,350$, not bad for just sitting back and watching your money grow.

Now suppose you want to estimate your rich Uncle Louie's fortune after he brags about the \$18,000 he made in interest last year from his bonds that paid 6% interest. Please take a deep breath and don't panic. You have to use a little algebra - not much. First write down your formula, $I = p \cdot r \cdot t$. You know the interest is \$18,000. The time is 1 year, so $t = 1$. $I = 18,000 = p \cdot (.06)$, so $.06p = 18,000$. Think back to 9th grade algebra. Yes! Divide both sides by .06.

$$\frac{.06 p}{.06} = \frac{18,000}{.06} = \$300,000$$

Your Uncle Louie is indeed "in the money." Don't worry about toadying up to Uncle Louie for his future inheritance. He will sense you feigned interest like Sherlock Holmes recognizing Moriarity. It can't work. Start early and develop your own mathematical and statistical approach to your investments. By retirement you will be as well off as Uncle Louie. Uncle Louie took a lot of college math. So can you. But if you are pressed, master each topic from this abbreviated book. Each topic is of vital importance to your future fiscal and physical health.

CALCULATOR GUIDE

Turn on your calculator. (Press on.) Now turn it off. (Press 2nd. Then press on.) You know that things are not going to be trivially easy when you need to push 2nd on to turn something off.

Now let us go back and calculate your interest on your son's \$5,000 CD at 5.5% for this year.

1. Press on – if the screen is blank. The calculator automatically shuts off if you are not using it. I waited for four minutes, it hadn't gone to sleep yet.
2. Press 5,000.
3. Press X.
4. Press .055. This is 5.5%.
5. Press Enter.
6. The answer in the lower right screen is 275.

Now suppose you forgot your ninth grade algebra. You still want to calculate the riches of Uncle Louie, but you lack the algebraic facility (notice, I didn't say skill. It can come back quickly. Take a few college math courses.)

You have the equation. $.06p = 18,000$. You want to solve for p . Your TI-83 calculator has an equation solver as part of its menu. You can solve for any variable in an equation. Here are the steps:

1. Press MATH.
2. Use the arrow to move down the screen and stop at Solver.
3. Press Enter.

4. The screen says: eqn: $0 =$ [You must bring 18,000 to the left of the equation, since Solver requires that you set the expression $= 0$.] You see, even the TI-83 does not help you if you did not do your homework in 9th grade. Don't panic. Subtract 18,000 from both sides of the equation. You have the equation $.06 \cdot p - 18,000 = 0$,

To enter your expression, after eqn $= 0$.

1. Press .06.
2. Press X
3. Press ALPHA (this gives you the letters A - Z).
4. Press 8 (this gives you p).
5. Press - 18,000 (not(-), the (-) is only used for a negative number, not the operation minus).
6. Press Enter.

This gives you the equation: $.06 \cdot p - 18000 = 0$.

Calculators don't use proper commas, such as 18,000. IBM would not approve if you leave them out.

Next, take a guess. Let $p = 1000$ or so, any number will do.

The calculator gives you the upper and lower bounds of the answer. E99 means 10^{99} or 1 with 99 zeroes. One million has six zeroes. Even Uncle Louie's fortune won't require the full bounds of Solver.

Now you are ready to solve for Uncle Louie's fortune. Use the arrows and move the cursor to the letter (or variable) for which you want to solve. You want to solve for p . After your cursor is on p , you are ready to solve.

Then press ALPHA. Press Enter (Solve). Then press ALPHA again. Press Enter (Solve).

The answer is next to the letter for which you solved. $P = \$300,000$. Yes. It would have been a lot better if you had done your 9th grade homework and not had to rely on Math Solver after solving for zero in the equation. We pay for our transgressions. For penance, please do the exercises below using your TI-83 calculator.

Practice Exercises

1. Find the simple interest on a \$10,000 CD for 8 months at 4.5% interest (per year).
2. Find out how much you would have to own in bonds at 6% interest to earn \$20,000 per year in interest income. Use Math Solver. Also use elementary algebra.

Answers: 1) \$300; 2) \$333,333.33

EXPERIENCE 1

Write down all your debt, student loans, credit cards, automobile loans. Calculate your monthly interest on your current debt. If you have no debt, find someone who has debt and calculate their monthly interest charges. What have you learned?

CHAPTER TWO - COMPOUND INTEREST

We are finding dirt these days on even saintly Albert Einstein. First, we read letters from him to his first wife that detailed his cold, precise estimate of their divorce settlement with a frightening lack of emotion or feeling. Next, we found that he loved compound interest – one of the great inventions of humankind (according to Einstein).

Indeed, compound interest is vitally important to your future. Whether you are now training for your first job or planning for retirement, you need to understand the power and influence of compound interest.

Let us look at an example. Suppose I have currently 2000 shares of a stock called EMX planned for the college education of my children. My daughter who is 8 years old, starts college in $10\frac{1}{2}$ years. My son, who is 5, starts college in $13\frac{1}{2}$ years. The stock is currently trading at around \$29 per share. The reason that I bought the stock is that median growth rate of the corporation is estimated by analysts to be in excess of 25% over the next five years. [Dream on.]

Now suppose EMX averages 20% growth in the value of its stock over the next $10\frac{1}{2}$ and $13\frac{1}{2}$ years. Let us calculate the value of the stock earmarked for my daughter Heather and son John over the next $10\frac{1}{2}$ and $13\frac{1}{2}$ years.

The key element in the growth of the stock is the interest on interest or compounding of the stock's value over time. The formula for compound interest is:

The amount A accumulated on a principal (starting value of money) after n periods (usually years) at i interest rate (usually per year) is: $A = p \cdot (1 + i)^n$

Note \cdot means to multiply.

For our example, for my daughter $p = 29,000$, $i = 20\%$ (we hope), and $n = 10.5$.

$$A = 29,000 \cdot (1 + .20)^{10.5} = \$196,698.52, \text{ not bad for middle class families.}$$

Suppose EMX earns average stock growth of 10%. Let us see how much will be available for our daughter.

$$A = 29,000 \cdot (1 + .10)^{10.5} = 78,889.86.$$

This would probably not be enough for an elite private college, but would probably pay for a state university's room, board, and tuition for four years.

Let us glimpse my son's possible college fund from EMX. My son attends college in $13\frac{1}{2}$ years. So, let us first try $i = 20\%$ average earnings growth per year.

$$A = 29,000 \cdot (1 + .20)^{13.5} = \$339,895.03$$

So now you see the advantage of starting a regular savings plan early. My son would have \$143,196.51 more for his college education (assuming stock growth of 20%) simply by having the equal principal grow for 3 more years. Of course, the 20% earnings growth produces dramatic positive returns.

If EMX earned 10%, let us glimpse my son's college fund.

$A = 29,000 (1 + .10)^{13.5} = \$105,002.41$, not bad for a modest return of 10%. Of course, EMX could lose money. If we could survive the stress, let us

glimpse the unthinkable, that my beloved stock would lose 10% per year. My daughter would have after 10 ½ years the following:

$$A = 29,000 (1 + (-.10))^{10.5} = \$9,592.78$$

She would need a hefty scholarship to attend college at that rate. But you get the idea. Compound interest is exceptionally important, and its mastery can enhance your planning about many dimensions of your life.

One of my father's favorite lessons was how \$1,000 doubled ten times became 1 million dollars. Let us use the compound interest formula to verify my father's claim.

$$A = p (1 + i)^n \quad i = 100\% \text{ (a double)}, n = 10$$

$$A = 1000 (1 + 1.00)^{10} = \$1,024,000$$

It just made it with \$24,000 to spare.

Even if you are 18 years old, it would be sensible to begin saving and investing for retirement now. Suppose you have \$1,000 in a mutual fund that averages 10% growth per year. Suppose you decide to retire at 58, 40 years from now. Your \$1,000 would increase as follows:

$$A = \$1,000 \cdot (1.10)^{40} = \$45,259.26$$

Suppose you wait until 40 to start squirreling away your retirement savings. You only have 18 years of compound interest. Your total amount for retirement from the \$1,000 contribution could increase according to the formula:

$$A = \$1000 \cdot (1.10)^{18} = \$5,559.92$$

There is a big difference between \$45,259 and \$5,559. Start a regular program of savings and investment as early as possible.

Let us glimpse one final application of compound interest – the effects of inflation over 10, 20 or 30 years. This catastrophic weakening of the buying power of money if inflation were high, say 5-10%, would cause massive unrest and suffering, particularly for people living on fixed incomes. Suppose a cup of coffee costs \$1 in 2010. Suppose you retire today at 55. You have to plan for at least 30 years of retirement. Let us calculate what that cup of coffee is going to cost in 30 years at an average inflation rate of 10%.

$$A = p \cdot (1 + i)^n = \$1 (1.10)^{30} = \$17.45$$

If you retired with financial needs of \$40,000 per year in the year 2010, in 30 years you would need $\$40,000 \times (17.45) = \$697,976.09$ per year to stay even.

This is why Alan Greenspan, the Director of the Federal Reserve Bank, was so single-mindedly focused against inflation. Our entire nation's social fabric would unravel at 10% annual inflation. The massive and widespread suffering would be impossible to contain. Of course, we are in a different but very real economic crisis due to high risk investments, little regulation and real estate speculation and deceit.

Now let us glimpse a more realistic scenario, an average inflation rate of 3% over 30 years. To calculate the cost of that cup of coffee in 2040, use the following:

$A = \$1 \cdot (1.03)^{30} = \2.43 , not a bargain but achievable if you retire with enough of a financial cushion to guard against the ravages of inflation. To get an appreciation of what your \$40,000 annual retirement income feels like in 2040 dollars, divide \$40,000 by 2.43 = \$16,460.91. Sixteen thousand dollars does not go a long way in 2010. Forty thousand dollars a year won't go a long way in 2040. However, a facility with compound interest can guide you to an early commitment toward investment and savings. It also is essential for you to retire when your income will be able to afford you a secure retirement. It is no fun if your retirement dollars run out after ten years and you have to play scared for twenty plus years of your life, fretting about money. Math is power. And our next topic, annuities, is a virtual powerhouse.

Let us look in slow motion how you can compute the amount my daughter will have for college use if we have \$29,000 growing at 20% annually (we hope). She will need the funds in 10.5 years.

The formula is:

A = amount accumulated

P = principal (starting value) = \$29,000

i = interest rate (per year) = .20

n = number of years = 10.5

$A = p \cdot (1 + i)^n$

1. Press (on) - turn on your calculator.
2. Enter 29000 (principal today).

3. Enter X (multiplication).
4. Enter ((left parenthesis).
5. Enter $1 + .20 (1 + i)$.
6. Enter) (right parenthesis)
7. Enter ^ (this means you are raising the quantity in parenthesis to an exponent, which for this problem $= n = 10.5$).
8. Enter 10.5.
9. Press Enter.
10. Your answer is 196698.5151 or \$196,698.52 (rounded off to cents).

This shows you the power of compound interest.

Now try to verify the rest of the computations in this section on compound interest.

Practice Exercises

1. Suppose you buy \$5,000 worth of stock at age 21. If it averages 20% return (very lucky), how much would you have from this stock for retirement at age 65?
2. How much would you have for retirement if you invested the same \$5,000 in a bond that returned 6% annually?
3. How much would you have from the stocks for retirement if they lost 5% per year?

Answers: 1) \$15,238,591.62; 2) \$64,927.41; 3) \$523.37.

INVEST WISELY!

Now that you understand how to compute compound interest, it may be a good idea to learn the Finance menu of the TI-83. You can solve all the problems from this section by learning where to input the correct data. The TI-83 will then solve for the missing value.

For example, let us return to the first problem that we worked out – my daughter's funds for college use. Let us solve the same problem using Solver on the TI-83.

Remember $A = p(1 + i)^n$ $p = \$29,000$; $i = .20$; $n = 10.5$.

Steps for TI-83 Finance.

1. Press 2nd Finance. You will see a screen with 1:TVM Solver.
2. Use \rightarrow to go to 1:TVM Solver.
3. Press Enter. You will see a screen with: N, I%, PV, etc.
4. Enter 10.5 for n.
5. Enter 20 for I%
6. Enter (-) 29000 for PV = present value, the same as the principal invested today. The TI-83 requires that you use (-) for the amount already invested. This appears to be a quirk of the program. Don't fight it.
7. Press Enter for PMT = 0. There is no payment for this problem.

8. The P/Y means the number of payments per year. Leave it at 1. The C/Y means the number of compoundings per year. Leave this at 1 also.

9. Press ALPHA.

10. Press Enter.

If you have the TI-83 Plus, use the APPS key. Press 1: Finance, and the exact same steps appear.

The Finance program has computed your FV = Future Value = amount after 10.5 years = \$196,698.52. You have now learned how to check your work using Solver.

The TI-83 Solver can solve for any missing value, which could become tricky requiring logarithms if you didn't have Solver. For example, suppose you wanted to know how many years it would take you to become a millionaire if you had \$50,000 invested in a stock that was increasing by 20% per year.

Please read Chapter 14, Financial Functions, in the TI-83 Guidebook that comes with the graphing calculator. You can solve for any one of the five TVM variables if you put in the other four variables. The five TVM variables that we will use are:

1. N - total number of payment periods.
2. I% - annual interest rate.
3. PV - present value
4. PMT - payment amount.

5. FV – future value.

We use #6:P/Y number of payment periods per year and #7:C/Y number of compounding periods per year. These are set at the same value.

Now let us finish our problem – that is how long it will take you to become a millionaire. You start with \$50,000, increase this sum (if you are very lucky) by 20% per year, and want to solve for N – the number of years necessary to have a FV, future value, of one million dollars.

Please compute the answer by using the following steps:

1. Press 2nd Finance.
2. Press Enter. You see the display of TVM Solver.
3. Enter the values: N = ____ (you are solving for n); i = 20%;

PV = -50000; PMT = 0; FV = 1000000; P/Y = C/Y = 1. We assume that you are compounding your \$50,000 once per year at i = 20%.

4. Press ALPHA.
5. Press Enter.

Your answer is 16.43103715. Let us round $n = 16.43$. Now it is time to check our work. Remember $A = p(1 + i)^n$. $A = FV$; $p = PV$; $r = 1\%$, and $n = N$. Substitute $p = 5000$, $r = .20$, $n = 16.43$. You get an answer of \$999,810.92. You are about \$200 away from your million dollars, but don't lose sleep. You rounded off 16.43103715 to 16.43 and lost a little money. Since you had zero for a payment, you could use END or BEGIN (referring to the end or beginning of payment intervals) and obtain the same answer.

Now try a few more practice exercises with TVM Solver.

Practice Exercises

1. You plan to put your \$10,000 in either a Treasury note for 30 years with fixed rate of 6% or take your chances in the stock market. You use the historical stock market average return rate of 8% per year. How much will you expect to have in each after 30 years?

a) In your Treasury note investment you have \$57,434.91 at the end of 30 years.

b) In your stock market investment, you have \$100,626.57 at the end of 30 years.

Be careful, calculate how much you would have in your stock investment if you lost 5% per year. Answer: FV = \$2,146.39. Remember you can very easily lose money in the stock market. But if you are wise and lucky, you can earn a lot of money.

You can check the last answer by using the formula for compound interest.

$$A = p(1 + r)^n \quad r = -.05.$$

$$A = 10000(1 - .05)^{30} = 10000 \times (.95)^{30} = \$2,146.39.$$

2. Your Uncle Louie had one half million dollars that he just gave you in stocks for your inheritance. His family said that he started with \$25,000 forty years ago in the stock market and never touched the money. What interest rate

did Uncle Louie earn in the stock market? Use TVM Solver. $I = 7.78\%$. You may be able to improve on Uncle Louie's interest rate.

EXPERIENCE 2

From your previous analysis of monthly debt, calculate your total debt if you do not repay the principal over (a) 10 years; (b) 30 years. What have you learned?

CHAPTER THREE - ANNUITIES

You probably know that wealthy people don't make just one payment to their account and leave it alone. They are regularly saving money. They use annuities, which is a sequence of equal periodic deposits. The kind of annuity we will consider is the ordinary annuity where you make deposits at the same time interval.

Consider what happens if you make four deposits of \$100, one per year for the next three years. You start with \$100 today. How much will you have at the end of three years?

Let us first do the calculations the hard way, using $A = p(1 + r)^n$. Your first \$100 has accumulated to: $A_1 = 100(1 + .10)^3 = 100(1.1)^3 = \133.10 after three years.

Your second \$100 has accumulated to: $A_2 = 100(1.1)^2 = \$121$ after three years. Your third \$100 has accumulated to: $A_3 = 100(1.1)^1 = \$110$ after three years. Your fourth \$100 has just been put in after three years and is worth \$100. Or if you like to use the formula $A_4 = 100(1.1)^0 = 100(1) = \100 . Now you can add up $A_1 + A_2 + A_3 + A_4 = \$133.10 + \$121 + \$110 + \$100 = \464.10 .

Now let us use TVM Solver to solve these annuity problems.

1. Press 2nd FINANCE
2. Press 1: TVM Solver
3. Enter $N = 3$

$$I = 10\%$$

$$PV = -100$$

$$FV = \text{____}, \text{ this is our answer}$$

$$PMT = -100 \text{ each year}$$

$$P/Y = C/Y = 1$$

4. Put the cursor next to FV, your desired answer.
5. Press ALPHA
6. Press ENTER

Your answer is \$464.10. The TVM Solver makes your calculations a lot simpler and probably more accurate.

Suppose you want to save for your son's college education. Regular monthly savings is the key to accumulating an adequate college fund. How much would you have by your son's 18th birthday if he is three years old and you save \$100 every month in an account that pays 8% interest compounded monthly?

Enter the following data in TVM Solver.

- a) $N = (15 \text{ years}) \cdot (12 \text{ months per year}) = 180$
- b) $I = 8\%$ (interest rate each month)
- c) $PV = -100$
- d) $FV = \text{_____}$
- e) $PMT = -100$
- f) $P/Y = C/Y = 12$

FV = your accumulated college savings amounts to \$34,934.51. This is if your interest is compounded monthly.

Suppose you get lucky and find a stock fund that earns 20% on your \$100 monthly annuity. How much would you accumulate for your college fund at the end of 15 years?

You can use the previous data. Simply change $I = 20\%$ and solve. You have \$113,529.49 in your son's college fund. This is probably not enough for a fancy private college, but Merrill Lynch estimated that my son would need \$78,800 to afford four years at a state university in 2011-2014, or \$208,200 to afford four years at a private college.

Let us check their arithmetic. For a private college, Merrill Lynch estimates that our family save \$7,100 per year, assuming an annual after-tax investment return of 7.5%.

Enter the following data on TVM Solver:

1. $N = 13$ (years)
2. $I\% = 7.5$
3. $PV = -7100$
4. $PMT = -7100$
5. $FV = \underline{\hspace{2cm}}$
6. $P/Y = 1$
7. $C/Y = 1$

My result was \$165,898.04. Perhaps Merrill Lynch should have used $N = 14$, since my son was only four at the time. If I changed $N = 14$, we obtain a future value of \$185,440.39. This is still short of Merrill Lynch's result. You even have to check Merrill Lynch.

Now let us determine what I should be setting aside every year, starting today so that my son's college fund will have \$208,200 in thirteen years. Assume you start today with an initial payment of \$1,000.

1. $N = 13$
2. $I\% = 7.5$
3. $PV = -1000$
4. $PMT = \underline{\hspace{2cm}}$
5. $FV = \$208,200$
6. $P/Y = 1$
7. $C/Y = 1$

Your payment is quite high, \$9,883.90. If you were able to save for 20 years, notice the difference. Let $N = 20$. Your annual payment decreases to \$4,905.89. The key is to start saving early.

You can use TVM Solver for a great many financial activities. Suppose you want to check your mortgage payment on a house or your car payment.

Suppose you bought a car for \$15,000 and put down \$4,000. You agree to pay off your 12% (per year) loan in four years. Calculate your monthly payment.

Use TVM Solver.

1. $N = 48$ (months)
2. $I\% = 12$
3. $PV = -11000$
4. $PMT = \underline{\hspace{2cm}}$
5. $FV = 0$ (The loan will be paid off)
6. $P/Y = 12$
7. $C/Y = 12$

Your payment is \$289.67 per month.

Now check your mortgage payment. Suppose you have a \$197,000 mortgage with interest rate of 8.5% over 30 years. Use TVM Solver to check your current monthly mortgage payment.

Enter the following data:

1. $N = 360$
2. $I\% = 8.5$
3. $PV = -197000$
4. $PMT = \underline{\hspace{2cm}}$
5. $FV = 0$
6. $P/Y = 12$
7. $C/Y = 12$

Your mortgage payment is \$1,514.76 per month. Check all your loans for accuracy.

Practice Exercises

The detailed solutions for Chapter Three exercises are included in the Answer Key at the end of this book.

1. Calculate the value of your annuity that you start at age 21 and contribute \$1000 per year until retirement by age 65 if you:

- a) earn 8% interest per year
- b) earn 3% interest per year
- c) earn 20% interest per year.

Now you understand the secret to wealth - regular savings and good luck in the stock market.

2. Suppose you decide to buy a car that has a maximum car payment of \$300 per month for 48 months. How much car can you afford with no down payment if the interest rate is 6% per year?

3. If you purchase a house for \$250,000 and put down 20%, how much is your mortgage payment on a 30 year loan at 7.5% rate of interest?

Lower interest rates, such as 7.5%, enable you to buy more expensive homes or save money on your planned purchase. The interest rate in our economy is one of the most important economic statistics.

4. Suppose you decide to eventually build a home in the Adirondacks. You fall in love with a piece of land that costs \$69,000. If you put down 20%, how much is your monthly mortgage payment on a seven year loan?

EXPERIENCE 3

Estimate your salary in ten years. What portion of salary would you regularly put in an annuity? How much will you have after 30 years?

What have you learned?

CHAPTER FOUR - ELEMENTARY PROBABILITY

Probability is a concept that you already have familiarity with. The probability of a head in a coin flip is $\frac{1}{2}$. The probability of someone running a mile under one minute is near 0. The probability that having a good financial plan early will benefit you is nearly 100%.

Properties of Probability

1. The probability of an event lies between 0 and 1. This is written $0 \leq P(E) \leq 1$. The \leq sign (less than or equal) means that the probability could equal 0 or 1.

2. The sum of probabilities of all possible outcomes of an event equals 100% or 1.

3. The probability of an event not happening is $1 - P(E)$.

Sixty percent of baby boomers plan to retire early. If you asked a baby boomer if he/she plans to retire early, what do you think the chances are that they do not plan to retire early?

Well, this is a simple application of probability. $1 - 60\% = 100\% - 60\% = 40\%$.

Surveys show that two-thirds of baby boomers felt they could not invest for the long-term. What is the probability that your boomer acquaintance is investing for the long-term? Of course, the probability is $1 - \frac{2}{3} = \frac{1}{3}$. Things are pretty dismal unless you start a systematic, early retirement plan.

The median annual salary for Money Magazine readers is \$84,000 per year. This means 50% make more; 50% make less. The median is the middle number if salaries (or any set of data) are arranged in order (high to low or low to high). What is the probability that a reader picked at random makes less than \$84,000? (50%).

What is the probability that a reader picked at random makes more than \$84,000? $(1 - \frac{1}{2}) = 50\%$.

Definition

The way we frequently obtain probability is through frequencies. We count the number of times an event occurs and divide by the total number of events.

$$P(E) = \frac{\text{number of times E occurs}}{\text{number of trials}}$$

Usually, we are estimating more complicated events than coin flips. It is hard to get a handle on the true probabilities. As a result, we count.

For example, to obtain our median salary of Money readers of \$84,000 per year, we count the number of readers making more than \$84,000, say one million, and divide by total readership, say 2 million.

$$\frac{1 \text{ million}}{2 \text{ million}} = \frac{1}{2} = 50\%$$

Consider the risks of inflation on a baby boomer planning to retire at age 50. Look at the table below.

PLAN TO BEAT THE AVERAGES

How long should you expect your assets to last once you have retired?

Choosing the average life expectancy means you have a 50 percent chance of outliving your money. The table below is a safer benchmark. Your probability of outliving these figures was only 20 percent in the most recently published government life tables. To be really safe, add a couple more years.

At age.....	Plan to live another.....	
	Female	Male
45 yrs.	46 yrs.	43 yrs.
50	41	38
55	36	32
60	31	27
65	27	22

Source: National Center for Health Statistics Data for 1991

If you are a female baby boomer, your chances of living 41 years or more is roughly 20%. If you are a man, your chances of living more than 38 years is roughly 20%. It is a good idea to use tables that give you a high probability (80%) of having enough money in retirement.

Look at the table below for an analysis of what the effect is of 3%, 5% and 10% inflation on \$1 over 10, 20, 30 and 40 years.

YEARS	3%	5%	10%
10	1.34	1.63	2.59
20	1.81	2.65	6.73
30	2.43	4.32	17.45
40	3.26	7.04	45.26

If your 50 year old baby boomer retires, he/she has to expect to live (with approximately 80% probability) forty more years. The retirement income that

he/she starts with has to be divided by 3.26 to determine the real income he/she will have at age 90 if inflation increases by 3% per year. For example, if you retire with a comfortable \$70,000 retirement income at age 50, by age 90 you will have a real income of $\$70,000 \div 3.26 = \$21,472$.

As you can see, most 90 year olds will be struggling with \$21,472 in retirement income (in buying power after 40 years). If you plan young in life, you reap great benefits later in life, as we have already seen in mathematics of finance and specifically the power of compound interest and annuities.

Homework Probability (An answer key is found at the end of this book - Chapters 3-10)

1. What is the probability of living less than 46 years if you are a female retiring at age 45?
2. What is the probability of living an additional 27 years or more if you are a male retiring at age 60? What is the probability of living less than 27 years for this same person?
3. Using the data from the New York Times article "Vast Advance is Reported in Preventing Heart Illness," August 6, 1999:
 - a) Find the probability of dying of cancer in 1997.
 - b) Find the probability of dying of cancer in 1979.
 - c) Find the probability of dying from heart disease in 1950.
 - d) Find the probability of dying from heart disease in 1996.

4. In July, 1999 the unemployment rate was 4.3%. What percent of the American population of workers was employed.

EXPERIENCE 4

Interview your parents about their plans for retirement. Help them plan for their financial needs using the charts from this chapter. Also discuss with them the knowledge you have gained from reading The Millionaire Next Door. What have you learned? If it is not convenient to interview your parents, make up a fictitious account.

Vast Advance Is Reported In Preventing Heart Illnesses

By GINA KOLATA

The death rates from cardiovascular diseases have plummeted by 60 percent since 1950, a Federal agency is announcing today, making advances against this leading killer of Americans one of the major public health achievements of the 20th century.

It has been known for years that death rates from heart attacks and strokes have been falling. But this report, in summarizing the trends over a century, sharply illustrates what has been accomplished, health experts said.

In today's issue of Morbidity and Mortality Weekly Report, the Centers for Disease Control and Prevention in Atlanta said that the number of deaths from strokes had steadily declined since the beginning of the century while those from heart disease peaked in the 1960's and had been falling ever since.

For example, if the death rate from heart attacks in 1963 had been repeated in subsequent years, an additional 621,000 Americans would have died in 1996 alone, the report said. The decline in the death rate has continued throughout the 1990's.

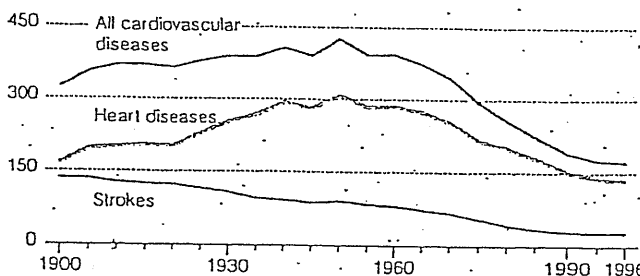
In 1950, the death rate from heart disease was 307.4 per 100,000 people.

BY THE NUMBERS

Limiting Cardiovascular Diseases

Changes in public health campaigns, including those against smoking, and medical advances have substantially decreased deaths.

Number of deaths per 100,000 people



Source: National Heart, Lung and Blood Institute

The New York Times

In 1996, it was 134.6. In 1950, the stroke death rate per 100,000 people was 88.8. In 1996, it was 26.5.

The decline "is a true success story of grand proportions," said Dr. Gilbert S. Omenn, a public health expert who is executive vice president for medical affairs at the University of Michigan.

Dr. David Jacobs, a professor of epidemiology at the University of Minnesota School of Public Health, said the decline in the death rates was "surely one of the great accomplishments of the century."

The decline was a major reason people were living longer, Dr. Jacobs said. But Dr. Omenn noted that it had made many Americans vulnerable to other maladies, like cancer.

According to the National Vital Statistics Report, the death rate from cancer has risen to 201.6 deaths per 100,000 in 1997 from 179.6 deaths per 100,000 in 1979.

Dr. Samuel H. Preston, a demographer who is dean of the School of Arts and Sciences at the University of Pennsylvania, said that when he was a graduate student in the 1960's no one would have expected the death rates from heart diseases to decline.

"They hadn't moved for several decades and there was a suggestion that they may be rising with affluence," Dr. Preston recalled. "There was a general aura of pessimism. Then all of a sudden in 1968, the rates began to go down very, very steadily and they have continued to go down."

The disease control centers said no single factor was responsible. Instead, it noted, the contributing factors constitute an almost humdrum list inspired by public health campaigns and clinical trials that showed the benefits of controlling the risks and treating the diseases.

One major factor was a decline in cigarette smoking, with 25 percent of adults smoking today compared with 42 percent 30 years ago. Other factors included better control of blood pressure, decreases in cholesterol levels and improved treatments of heart attacks and strokes.

It was a familiar story to those who study public health. The great killers of the 19th century were also controlled by a collection of societal changes, like maintaining clean water supplies and quarantining the sick, resulting in declines in diseases like tuberculosis, cholera and diphtheria before there were drugs to treat the illnesses or vaccines to prevent them.

Dr. Omenn predicted that diseases like cancer, arthritis, Alzheimer's disease and psychiatric diseases would be controlled not by a wonder drug like penicillin but by a mixture of preventive measures and treatments, some of which would involve changes in diet and behavior that might seem almost mundane.

Nonetheless, cardiovascular diseases remain a leading cause of death in the United States, with heart disease the No. 1 killer and strokes No. 3.

The disease control centers noted that more could be done to control the diseases even without new medical discoveries simply by getting the risk factors under control.

But Dr. Omenn noted that only half of those with cardiovascular disease had identifiable risk factors like smoking, high blood pressure, diabetes, high cholesterol levels or obesity. When he thinks about the potential to slash the cardiovascular death rates, further he thinks of that other half of the population whose disease is as yet unexplained by medical science.

CHAPTER FIVE - STANDARD DEVIATION/VARIANCE

Suppose you studied the want ads for accountants for a week and observed five jobs with starting salaries (in thousands) of \$40, \$35, \$40, \$50 and \$35. The range between the high salary of \$50 and \$35 is $50 - 35 = \$15K$. Also each salary is fairly close to the mean. To calculate the mean, add the five numbers and divide by 5.

$$\frac{40 + 35 + 40 + 50 + 35}{5} = \frac{200}{5} = \$40K$$

As you can see, the salaries are fairly close to \$40K.

Now suppose you observed these five starting salaries for accountants: \$100, \$20, \$15, \$30, \$35. The mean would still be the same \$40K. But there would be a larger average deviation from the mean. The concept of deviation from the mean is the essence of variance or standard deviation.

The variance (σ^2) of a population (the entire data set of a statistic of interest) is defined as:

$$\sigma^2 = \frac{\sum_{i=1}^n (X - \bar{X})^2}{n}$$

n = size of the sample
 X = individual score
 \bar{X} = population mean

We rarely know all the values of the population. In fact, much of Statistics is based on estimating the mean of the population based on a small sample ($n < 30$) or large sample ($n \geq 30$).

The sample variance, written as S^2 , is the sum of the squares of deviations from the mean, divided by $n - 1$.

$$S^2 = \frac{\sum (X - \bar{X})^2}{(n - 1)}$$

The sample standard deviation, S , is the square root of the variance.

Let us go back to our first example, where the five numbers for salary are: \$40, \$35, \$40, \$50, \$35. $\bar{X} = 40$.

You can use a table to simplify your calculations:

X	$X - \bar{X}$	$(X - \bar{X})^2$
40	40-40	$0^2 = 0$
35	35-40	$(-5)^2 = 25$
40	40-40	$0^2 = 0$
50	50-40	$10^2 = 100$
35	35-40	$(-5)^2 = 25$
$\Sigma X = 200$	$\Sigma(X - \bar{X}) = 0$	$\Sigma(X - \bar{X})^2 = 150$

This always equals 0

$$\text{The sample variance} = \frac{150}{5-1} = \frac{150}{4} = 37.5$$

To obtain the sample standard deviation, simply take $\sqrt{37.5} = 6.1$.

Usually, the larger standard deviation means greater variability within the sample. To compensate for samples with large numbers, like major league baseball players' salaries, you could use a statistic called the coefficient of variation:

$$\text{Coefficient of variation} = \frac{\text{standard deviation}}{\text{mean}} \times 100\%$$

To illustrate, use the standard deviation we have just calculated of 6.1, mean = 40. The coefficient of variation for this sample is:

$$\frac{6.1}{40} \times 100\% = 15\%$$

This 15% means fairly low variation.

There is a short-cut formula that eliminates the calculation of the $(X - \bar{X})^2$.

We will use this formula, which is:

$$\text{SHORT CUT FORMULA} \quad S^2 = \frac{n \sum X^2 - (\sum X)^2}{n(n-1)}$$

You will get slightly different answers for the variance depending upon whether you use the population or sample variance formula. If you take three semesters of Calculus and a Calculus based Probability and Statistics course, you will find out why. You will also understand the proofs for the formulas in the primer and become facile in an ever widening field of critical importance to our nation - Mathematical Modeling.

Rounding errors are frequently an issue in statistics. Drs. Jay Devore and Roxy Peck recommend using four or five digits of decimal accuracy beyond the decimal accuracy of the data values themselves (Introductory Statistics, 1994, p. 83). Then for the last step, round the answers to the same level of accuracy as the data.

Let us calculate our variance using the shortcut formula of the salary data:

$$n = 5 \text{ (salaries)}$$

$$\sum X^2 = 40^2 + 35^2 + 40^2 + 50^2 + 35^2 = 8150$$

$$\sum X = 40 + 35 + 40 + 50 + 35 = 200$$

$$(\sum X)^2 = 40,000$$

Note: $\sum X^2$ does not equal $(\sum X)^2$

$$\text{Variance (sample)} = S^2 = \frac{5(8150) - 40,000}{5(4)} = 37.5$$

$$\text{Standard deviation (sample)} = S = \sqrt{37.5} = 6.1$$

Let us try another example of variance. Consider the cost of a three bedroom house in the city you plan to work. Your largest expense is usually your house mortgage and taxes. This is vital statistics. You need to understand your finances and how to estimate your house expenses. Consider downsizing and moving to a highly rated city like Saranac Lake, New York. Houses are relatively inexpensive in Saranac Lake. On any given day you might have a wide variety of houses for sale in the \$60,000 - \$350,000 range. Many sell between \$60,000 and \$150,000 - quite a significant difference from high priced areas such as Westchester.

It is essential that you subscribe to the local newspaper of the town in which you plan to move. You can look at crime, job openings, and our immediate concern - house prices. The Adirondack Daily Enterprise on August 6, 1999 advertised homes for sale. Their prices (for the immediate Saranac Lake area) were \$345,000, \$215,000, \$145,000 and \$55,000. We can leave out the thousands and consider this housing sample of four as \$345, \$215, \$145, and \$55 K.

Let us use our TI-83 calculators to compute the variance and standard deviation of this sample of four house prices. Here are the steps to enter our data:

- 1) Press 2nd.
- 2) (
- 3) 345, 215, 145, 55
- 4) Press 2nd
- 5))
- 6) Press STO
- 7) Press 2nd
- 8) Press 1
- 9) Press Enter

Your data is stored in List 1. Now press

- 1) 2nd List
- 2) Use the ▲ key to go to highlight MATH
- 3) Use the ▼ key to go to Std Dev
- 4) Press Enter
- 5) Press 2nd 1
- 6) Press Enter

The answer is 122.3. The standard deviation is 122 if rounded to the nearest whole number (thousand).

Now use a similar process to obtain the mean.

The answer is 190 K. You have just seen how easy it is to analyze house prices. The mean house price was \$190K with standard deviation of \$122. The coefficient of variation was:

$$\frac{\text{S. Dev.}}{\text{Mean}} \times 100\% = \frac{122}{190} \times 100\% = 64\%$$

This is a very high coefficient of variation. The house prices are very widely variable. This was due to a small sample ($n = 4$) and unusually expensive house prices. Since I had sampled dozens of houses last year for an earlier book, I know that this one day sample of four does not represent the true house price situation in Saranac Lake. You should analyze your prospective area by taking a two month sample of house prices and preferably the actual selling price - not the price advertised in the newspaper or quoted by real estate agents.

Homework

1. Find the standard deviation, variance, and coefficient of variation for this sample of teacher salaries: \$30, \$40, \$36, \$28, \$42 K. Use both the definition formula and shortcut formula to compute variance.
2. Use your TI-83 calculator to confirm your answer to #1.
3. Your six stocks went up (in %) 0, 7, 8, 6, 10, and 28 last year.
 - a) Find the mean % gain for your six stocks
 - b) Find the variance, standard deviation, and coefficient of variation of this sample.
4. Physician salaries were sampled and found to be \$110, \$150, and \$135 K. Find the mean, variance, standard deviation and coefficient of variation three different ways.

EXPERIENCE 5

Obtain a sample of at least twenty house prices and at least 20 professional salaries in your city of possible relocation. Use a month or two months of samples from the local newspaper to obtain your data. Compute the mean, variance, standard deviation, and coefficient of variation for both samples. (Keep the data separate. Do not mix house prices and salaries.)

CHAPTER SIX - SAMPLING METHODS/CONFIDENCE INTERVALS

Most statistical analyses want to make an estimate of some population characteristic - for example the average salary you might make as a starting teacher in Mariana, Florida or Saranac Lake, New York. It is inefficient to ask every starting teacher from the previous year about their salary. It would be hard to find every one, very costly, and a tortuous process. The better way is to take a sample of say 30 or 60 or 100 starting salaries in the area. You can use that data to come up with an interval that with 95% confidence captures the true population mean (99% confidence if your are satisfied with a larger interval - say \$24,000 - \$46,000 as contrasted to a 95% confidence interval that might look like \$27,000 - \$43,000).

In order to ensure that your sample is well constructed, it must have certain characteristics, including:

1. Randomness

If we went to Greenwich (a wealthy town) to look at average house prices for the Northeast, we would have little evidence to draw valid conclusions. Greenwich is not representative of the Northeast. Its houses are among the highest priced in the nation.

You cannot make sound inferences as to the average house prices in the Northeast, unless you have a sample that is representative of the Northeast. You need a random sample.

A random sample means that each house in the Northeast has an equal chance of being picked. You could somehow put each index card with a house price into a very big hat, shuffle the cards well, and pick say 1000 cards. The key is to try to make your sample as random as humanly possible. You can never achieve perfect randomness, but you can easily eliminate glaring biases. If a sample is not random, it is biased.

If your population can be matched with numbers, you could use a random number generator from a standard calculator or computer to help you select your sample. Unfortunately, computer generated random numbers repeat every K times, so they are not random. But K can be a very large number if you use a well constructed random number generator.

2. Stratified Sampling

Sometimes you can break up your population into a set of n subpopulations, that we call strata. For example, you could break the Northeast into say 200 counties and take a random sample of each county house price.

Stratified sampling usually ensures that you have thought about the subgroups that make up your population and included each as part of your sample. For example, if there are three different subgroups of teaching that you are qualified for, say private school, public school and college, you would want to include each subpopulation of teacher in your sample. Otherwise you might limit your sample to public school teachers, enter private school teaching, and find that you had overestimated your starting salary. After all, public school

teachers generally make significantly more than private school teachers. Public defenders make much less than corporate lawyers. You shortchange yourself if you do not try to achieve a random sample with the underlying strata proportionately represented.

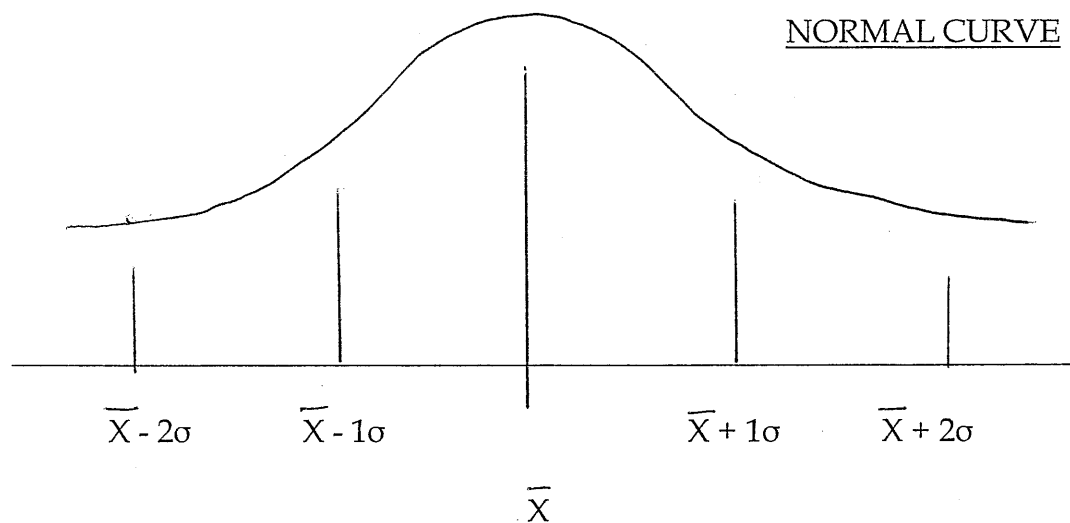
Confidence Intervals

Before we compute a confidence interval, there are two powerful concepts that should be introduced:

1. The Normal Curve - this curve is a basic and remarkable foundation for statistics. The curve below could be approximated by the graph $y = k \cdot 3^{-x^2}$. This is a poor approximation, but at least you get some idea of what the algebra is behind the normal curve below.

(Let $k = \frac{1}{\sqrt{2\pi}}$ and substitute values for $x = 0, \pm 1, \pm 2$ to glimpse the algebra

Behind the normal curve.)

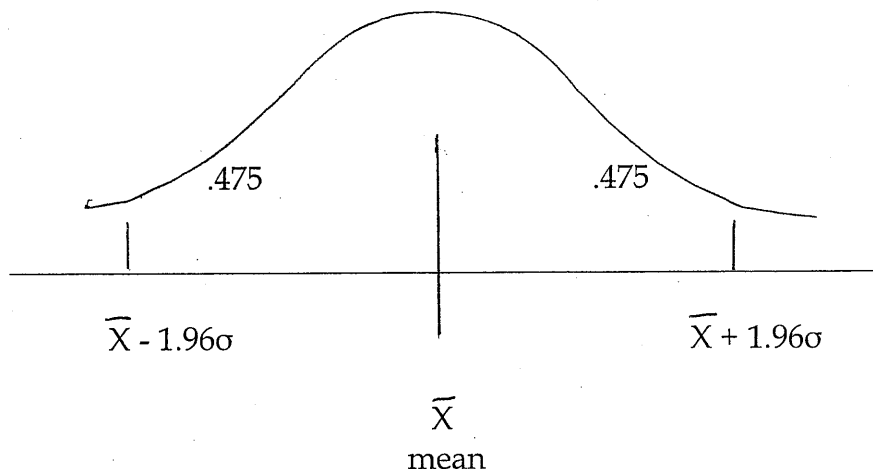


The probability of a value lying between the mean \bar{X} and positive infinity is 50%. The corresponding probability of a value lying between the mean \bar{X} and $-\infty$ is 50%.

The area under the normal curve is 100% = 1. This doesn't help us very much. What helps us is the fact that we can look up the probability of a value lying between any two values if they are expressed in standard form, Z measures where:

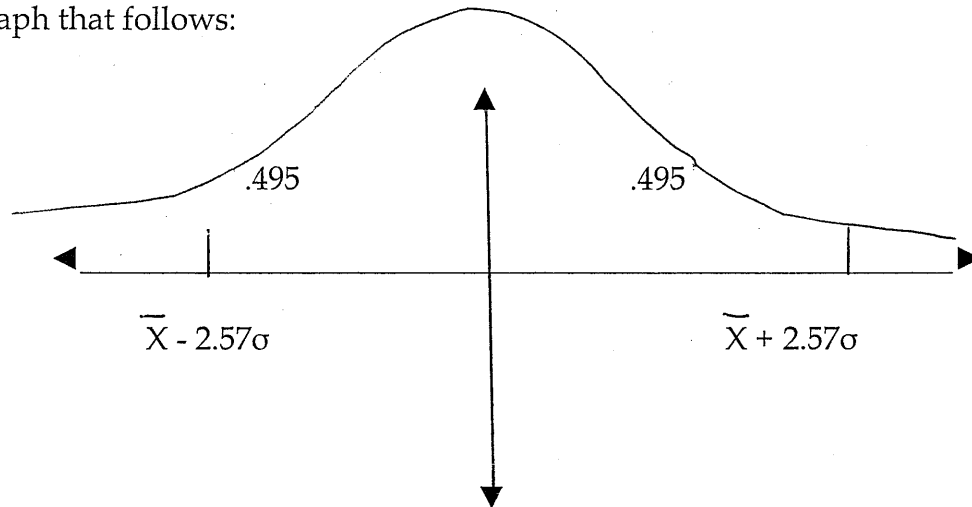
$$Z = \frac{X - \mu}{\sigma} \quad X = \text{value}; \mu = \text{population mean}; \sigma = \text{standard deviation}$$

The normal curve can be teamed with the central limit theorem to help us achieve a great many powerful statistical ideas. The central limit theorem tells us that when sample size is sufficiently large, say $n \geq 30$, the samples of \bar{X} (mean) can be approximated with probabilities from the normal curve. We will just use two normal curve probabilities below, the probability of being between the mean -1.96 (standard deviations) and the mean $+1.96$ (standard deviations). The probability of a value falling in this interval is 95%. See the graph below ($.475 + .475 = 95\%$)



The probability of a value lying between $\bar{X} + 2.57\sigma$ and $\bar{X} - 2.57\sigma$ is 99%.

See the graph that follows:



Even if the population from which you are sampling is not a normal distribution, you can always use this approximation. The normal curve is an amazingly generalizable and powerful invention, one of the most useful of the twentieth century.

We now can apply the central limit theorem and use the normal curve for our next result. If $n \geq 30$, the sampling distribution of \bar{X} (mean) can be described by a normal curve with sample mean \bar{X} and standard deviation σ/\sqrt{n} .

This gives us our two powerful formulas for large samples ($n \geq 30$).

95% Confidence Interval for Population Mean μ

$$\bar{X} \pm 1.96 (\sigma/\sqrt{n})$$

and 99% Confidence Interval for Population Mean μ

$$\bar{X} \pm 2.57 (\sigma/\sqrt{n})$$

Let's use these two results to calculate 95% and 99% confidence intervals for your home price if you move to Saranac Lake. Assume that you have inspected the Adirondack Daily Enterprise for several months and written down 50 house prices. You need to find the mean, say \$95,000. You next need to find the standard deviation, say \$25,000.

We can now calculate the 95% and 99% confidence intervals for the average home price in Saranac Lake. To calculate the 95% confidence interval,

$$95\% \text{ C.I.} = \bar{X} \pm 1.96 (\sigma / \sqrt{n}) = \$95,000 \pm 1.96 (25,000 / \sqrt{50})$$

$$95\% \text{ C.I.} = 95,000 \pm 1.96 (25,000 / 7.07107)$$

$$95\% \text{ C.I.} = 95,000 \pm 6929.644311$$

$$95\% \text{ C.I.} = (95,000 - 6929.64\dots, 95,000 + 6929.64\dots)$$

Now round to the nearest thousand at the end.

$$95\% \text{ C.I.} = (88,000, 102,000)$$

You can assume that your home in Saranac Lake will cost somewhere between \$88K and \$102K (with 95% confidence). The reason that you rounded to the nearest thousand at the end was that all your house prices were in thousands. If you ended up with a confidence interval that had house prices to the nearest penny, you would give people the illusion that you could estimate house prices much more accurately than you can. After all, most house prices rise or fall by \$10,000 as a result of price changes or negotiations. Always think of your audience and try to use statistics that make sense and reflect reality.

Now let us use the same data to compute a 99% confidence interval for home prices. This has to be a wider interval because you have more probability of being correct.

$$99\% \text{ C.I.} = \bar{X} \pm 2.57 (\sigma / \sqrt{n})$$

$$99\% \text{ C.I.} = 95,000 \pm 2.57 (25,000 / 7.07107)$$

$$99\% \text{ C.I.} = 95,000 \pm 9086.319326$$

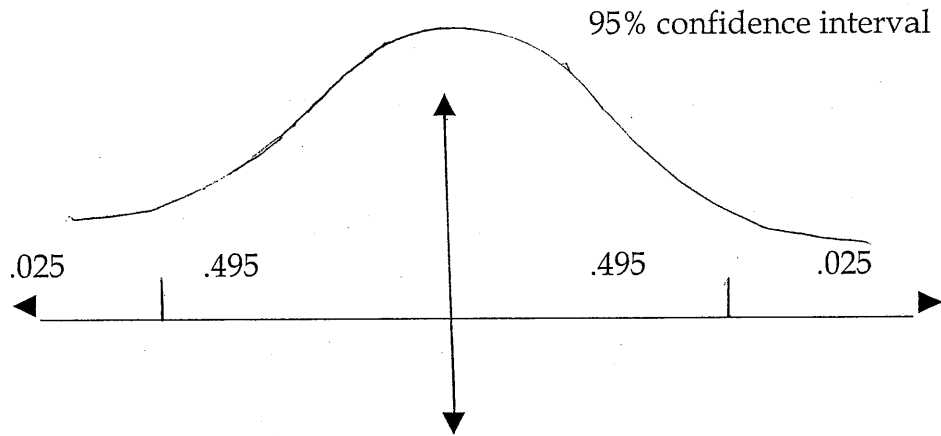
Round to thousands

$$99\% \text{ C.I.} = 95,000 \pm 9,000$$

$$99\% \text{ C.I.} = (86,000, 104,000)$$

You now have 99% confidence that your house will cost between \$86K and \$104K. If you wish more than 99% confidence, it usually is not practical. For example, you could be 100% confident that your house will cost somewhere between \$0 and \$1 billion. But it is difficult to plan with this type of confidence interval.

2. Small Sample Confidence Intervals - Frequently, we have to make analyses with smaller samples. If your sample is smaller than 30, you need to change the Z value of 1.96 or 2.57 to the corresponding value from a t-distribution. To determine the t value, use a t table and look up the value by determining df and p. The df (degrees of freedom) is (n-1), where n is sample size. Your p value is .025 (for 95% confidence) or .005 (for 99% confidence). Look at the graph that follows.



You now see the reason for relying on a p-value of .025 from the t-table to obtain a confidence interval (95%). There is 2.5% probability of landing in either the extreme left tail (house price below \$88K) or in the extreme right tail (house price above \$102K).

Let us calculate a 95% confidence interval for your starting salary if you have a sample of ten accountants' salaries from Saranac Lake and they are (in thousands):

22, 24, 18, 26, 28, 20, 26, 25, 29, 30

Calculate \bar{X} , σ using the data editor of the TI-83.

$$\bar{X} = 24.8$$

$$\sigma = 3.9$$

For 95% confidence, use $n = 10$, $df = (n - 1) = 10 - 1 = 9$

$$t_{.025, 9df} = 2.26$$

$$95\% \text{ C.I.} = 24.8 \pm 2.26 (3.9 / \sqrt{10})$$

$$95\% \text{ C.I.} = 24.8 \pm 8.814/3.162 = 24.8 \pm 2.787$$

$$95\% \text{ C.I.} = (22K, 28K)$$

3. Central Limit Theorem - Fun exercise: Let us use the random number program on your TI-83 to glimpse the genius of the central limit theorem. We are going to see the concept of the binomial density function. For a binomial random variable, we need something like a coin flip with two possible outcomes, heads or tails. The probability of a head does not change from coin flip to coin flip. We want to estimate the number of heads if we flip a coin say 100 times. The mean of the binomial random variable $\mu = np$, where $n = 100$, $p = \frac{1}{2}$

$$\mu = 100(1/2) = 50$$

The standard deviation of the random variable $\sigma = \sqrt{np(1-p)}$

For example, with $n = 100$, $p = \frac{1}{2}$ $\sigma = \sqrt{100(1/2)(1-1/2)} = \sqrt{25} = 5$

We can use our confidence interval knowledge to calculate the 95% confidence interval for the expected number of heads in 100 coin flips as:

$$\bar{X} \pm 1.96 [\sigma] = 50 \pm 1.96(5) = (40, 60)$$

With 95% confidence, the mean of heads will be between 40 and 60.

- a) Calculate the 99% confidence interval for # of heads.
- b) Use the Random Number key to simulate a coin flip as follows:
 - 1) Press MATH
 - 2) Go to PRB [Probability]
 - 3) Enter
 - 4) Now press enter 30 times

Write down a count of random numbers that are greater than .5. These are heads; random numbers less than .5 could be considered as tails. In the unlikely event of obtaining a perfect .5, discard this outcome. I obtained 13/30 numbers greater than .5 my first time. Repeat this exercise ten times. Suppose you now have a sample size of 300, $n = 300$, $p = 1/2$. Suppose you obtained 140 heads, 160 tails. Your confidence interval estimate could be determined as follows:

$$\mu = np = 300(1/2) = 150$$

$$\sigma = \sqrt{np(1-p)} = 6$$

$$95\% \text{ C.I.} = 150 \pm 1.96(6) = (138, 162)$$

You may conclude that there was a 95% probability of getting between 138 and 162 heads. If you obtained 140, you were within chance levels. If you obtained 200, your calculator was probably broken. This is way outside chance. This is one of the great many exciting applications of the central limit theorem. We obtained our value of 1.96 from the normal curve, the centerpiece of Statistics.

Please take several courses in Statistics to better understand the way Statistics can be applied to everyday decisions and improve your ability to plan.


Sample say 300 random numbers from 0-1 in a similar way to our example. Count the number of heads. Is your result within that predicted by a 95% confidence interval? Try a 99% confidence interval.

Homework

1. Calculate a 99% confidence interval for the data related to accountants' salaries. Would you expect the interval to be wider or narrower than the 95% confidence interval?

2. Use your TI-83 to enter your data for accountants' salaries. Obtain the mean, \bar{X} , and standard deviation, σ , as shown.

Next use the following to calculate the 99% confidence interval using your TI-83.

- a) STAT
- b) 
- c) Go to 8: T interval
- d) Enter
- e) Press Data
- f) List: L1
- g) Freq: 1
- h) C Level: 95
- i) Calculate
- j) Enter

The 95% confidence interval is now given: (22.023, 27.577). The TI-83 also gives $\bar{X} = 24.8$, $S_x = 3.881580434$, $n = 10$. You can check your computations using the TI-83.

3. Find a 95%, 99% confidence interval for your home price in Scarsdale if a recent New York Times advertised price list was as follows: (in thousands)

495, 650, 1300, 369, 599, 750, 375, 2400, 650, 475

4. Find a 95%, 99% confidence interval for your starting salary as a nurse if you sampled 36 starting salaries from Mariana and obtained an average of \$50K with standard deviation of 10K.

EXPERIENCE 6

Let us analyze your house price and salary estimate to determine whether you can afford your house. We will follow an example but ask you to subscribe to a local newspaper and obtain accurate housing and salary figures to ensure realistic estimates. Imagine living in Saranac Lake.

a) Your 95% confidence interval for house prices was calculated as (88K, 102K). To play it safe use an estimate of 102K. This way you can be 97.5% confident that your house will be cheaper.

b) Next use your salary estimate as accountant. Your 95% confidence interval was (22K, 27K). Take the 22K salary. You are 97.5% confident of making more.

c) Use your TI-83 Mathematics of Finance program to calculate your mortgage on your \$102K house. You would usually need 20K as a down payment. Your mortgage will be for approximately \$82K. Use TVM Solver as follows:

1. 2nd Finance [TVM Solver]
2. Enter
3. $n = 360$ (30 years, 12 months/year) You usually take a 30 year loan.
4. Let $I\% = \text{say } 8\%$. This is very conservative.
5. $PV = -82,000$. This means you start with debt (mortgage) of \$82K
6. $PMT =$ This is your monthly payment that SOLVER is used to determine. You can leave the 0 there.
7. $FV = 0$ When you pay off the loan in 30 years, your future debt is 0.
8. $P/Y = 12$ (payments per year)
9. $C/Y = 12$ (the bank compounds your interest 12 times per year)
10. BEGIN
11. Go to PMT
12. Press ALPHA Enter

Your monthly payment is calculated at \$597.70. You must also add in taxes. Taxes in Saranac Lake are around \$200/month on such a house. So your monthly payment = mortgage + taxes = \$800 approximately.

Your monthly salary is approximately $\$22,000/12 = \$1,833$.

Banks like to loan people mortgage money if the ratio of the monthly payment/salary is 25% or less. Consider your ratio: $800/1833 = 44\%$.

No bank would grant you a mortgage with a 44% ratio. You will need to rent, buy a much cheaper house, or have another salary to afford a house in Saranac Lake, one of the more affordable areas to live in New York State.

Imagine how impossible it is to start in an expensive area like Scarsdale.

Mortgages carry a lot of stress, so try to live within your means.

Your experiential project is to:

- 1) Obtain 95% confidence estimates of your future housing and salary using the local newspaper.

- 2) Use TVM SOLVER to determine whether you can afford your future house. Don't forget closing costs and insurance on your future home. If your mortgage and taxes together are 25% or lower of your monthly income, you will be very fortunate.

CHAPTER SEVEN - TESTS OF SAMPLE MEAN

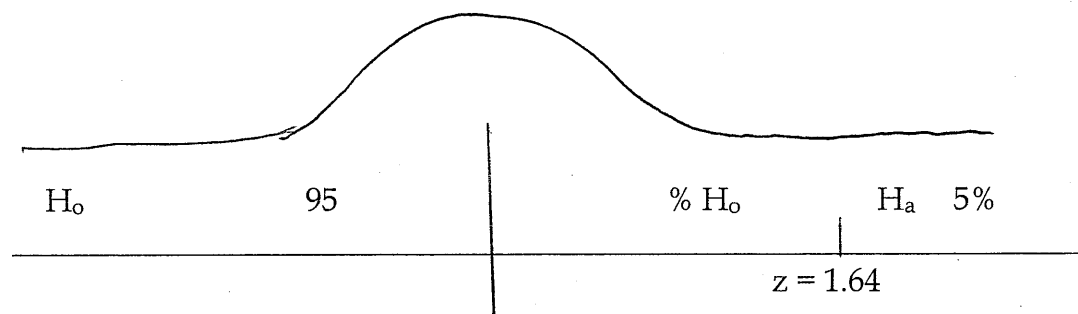
This is a course of Vital Statistics. It covers life and death matters, which impact your life. We are ready now to test hypotheses using our knowledge. A hypothesis is an educated guess. When I first began researching a book for baby boomers two years ago, I had a hypothesis that boomers would live longer if they retired in the Adirondack village of Saranac Lake than if they moved to Mariana, Florida. These were the two most highly rated small towns in New York State and Florida by Norman Crampton, The 100 Best Small Towns in America (Macmillan, 1995).

I sent away for the death records from the Adirondack Medical Center and Jackson County Health Department to determine whether Saranac Lake residents or Mariana residents lived above the national average of 75 years. The average for Mariana was 74.26 and the average for Saranac Lake was 75.27. We can use a simple formula to determine whether either was greater than (from a statistically significant perspective) the national average of 75. We need a brief introduction to the essential concept of significance. Please remember to take a complete course in Statistics, since we have reduced Statistics to bare essentials. There is no way for the reader to fully appreciate the enormous scope and power of Statistics from these eight highlights.

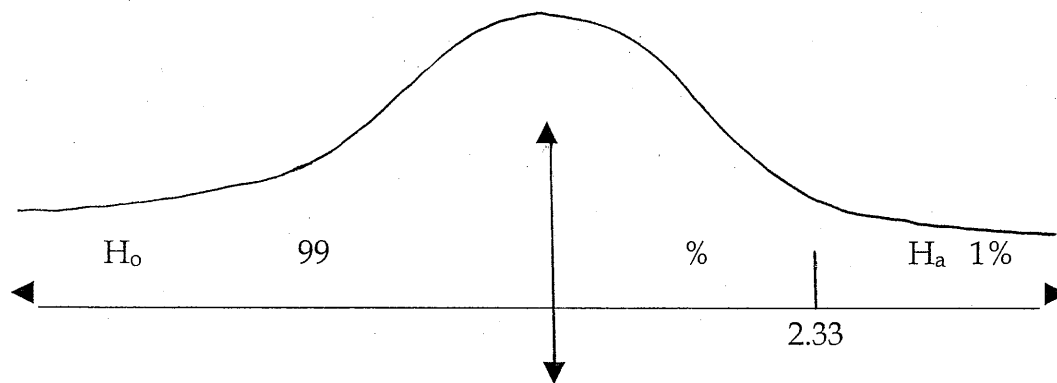
Statistical Significance

We are usually trying to establish that one value of a test statistic is significantly greater than another. For example, we know that Saranac Lake average life span of 75.27 based on a sample of say 30 is greater than 75. The question is whether the difference is statistically significant. We first need to set a level of Type I error. This level, usually 1% or 5%, means that even if we conclude that our test statistic is statistically significantly greater, there is a 1% or 5% chance that the result was due to chance. The sample simply may have been biased, not a reflection of the underlying superiority of one statistic. We are going to greatly simplify Statistics by confining our analyses to two possibilities, 1% type error or 5% Type I error and the hypothesis that one statistic is greater than another.

If sample size is large ($n \geq 30$), we only need two pictures below:



Picture I



Picture II

H_0 is always the hypothesis that the two means are equal (from a statistical perspective). The two numbers are hardly ever exactly equal, but H_0 means that they are too close to conclude that one is statistically significantly greater than the other. Two levels of statistical significance are common, 1% and 5%. These error levels, specifically Type I error levels, mean that 1% or 5% of the time sampling error will account for one sample to end up statistically significantly greater than the other even if they should be equal. Type II error, which you would learn a great deal about in an elementary Statistics class, means that you have concluded H_0 (means are equal) when they are not (H_a is true).

We can now use the formula for one sample mean hypothesis tests below:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

n = sample size
 \bar{X} = sample mean
 μ = population mean
 σ = standard deviation

Now let us test whether Saranac Lake's average life span 75.27 is statistically significantly greater than the national average of 75.

H_0 : (null hypothesis) Saranac Lake's life span is equal to that of the United States.

H_a : (alternative hypothesis) Saranac Lake's life span is greater than that of the United States.

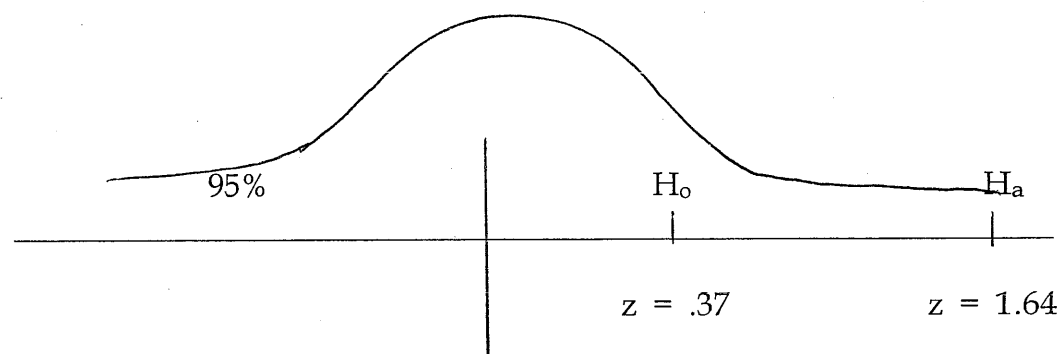
We arbitrarily select $\alpha = 5\%$, our confidence level is 95% and we use Picture I. If our z (computed value) is greater than 1.64, we conclude H_a ; any number less than 1.64, we conclude H_0 .

Calculate below: Let $\sigma = 4$, $n = 30$

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{75.27 - 75}{4/\sqrt{30}} = \frac{.27}{(4/5.477)}$$

$$z = \frac{.27}{.7303} = .37$$

Look at where $z = .37$ falls in Picture I, well to the left of 1.64.



We conclude that the average life span in Saranac Lake is the same as the average American life span. If we were to require $\alpha = 1\%$, we would reach the same conclusion using Picture II.

The average life span of the sample of Mariana residents was 74.26. Let us use $\alpha = 5\%$ to determine whether Mariana was higher than the national average.

(Of course, we will conclude H_0 since we are going to obtain a negative z value.)

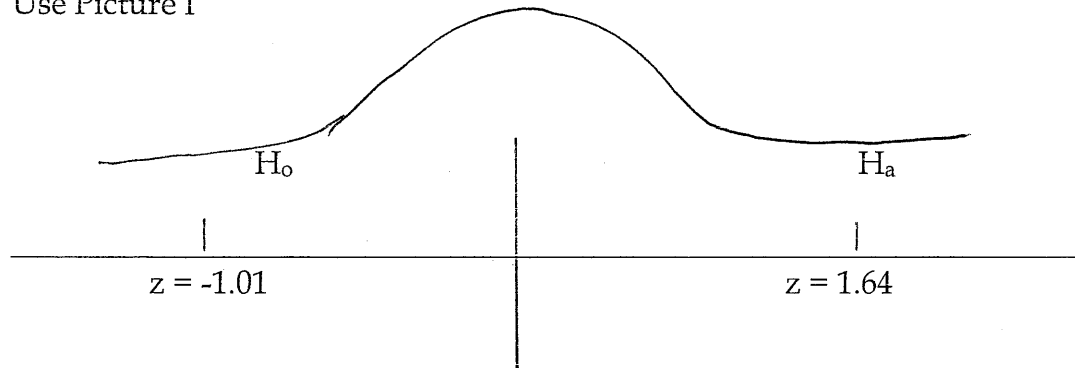
$$Z = \frac{74.26 - 75}{4/\sqrt{30}}$$

$\bar{X} = 74.26$ (Mariana mean); $\mu = 75$ population mean
 $\sigma = 4$ population standard deviation;
 $n = 30$ sample size

$$Z = \frac{.74}{.7303}$$

$$Z = -1.01$$

Use Picture I



Our conclusion is H_0 , the two means are equal (from a statistical perspective).

It would be natural to ask whether the national average is higher than the Mariana average. Of course, you now are testing a second hypothesis and must add another 5% Type I error to your accumulated error. Switch \bar{X} and μ in the formula. The result is summarized below:

$$Z = \frac{75 - 74.26}{4/\sqrt{30}} = 1.01$$

Since $1.01 < 1.64$, we conclude H_0 : The two means are equal. To be precise, the national average of 75 may have some rounding error, making our study somewhat inconclusive. However, you have just been introduced to one of the most powerful concepts of the twentieth century - Statistical Significance.

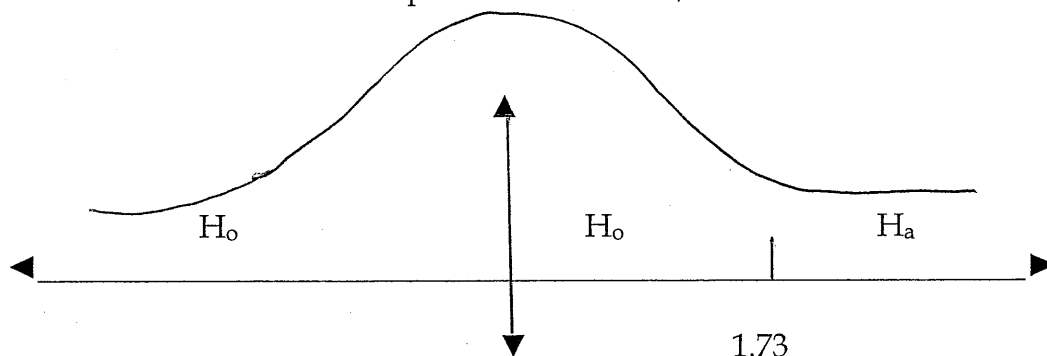
Small Sample Analysis

We often have samples that are small (fewer than 30). In order to perform the hypothesis test with one sample mean compared to a population mean, we simply have to adjust the critical value ($z = 1.64$ or 2.33). We can use the same formula to obtain a computed value, but we have to use the t-table with $(n-1)$ degrees of freedom to change our critical value. For example, let us change our original problem in only one element, sample size. Suppose we have an average life span in Saranac Lake of 75.27 but the sample size was 20. We compute the critical t value as follows:

$$t = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{75.27 - 75}{(4\sqrt{20})} = \frac{.27}{.8944}$$

$$t = .30$$

Now, if you want 95% confidence, $\alpha = \text{Type I error} = 5\%$, you look up $\alpha = .05$, $df = n-1 = 20-1 = 19$. Use the picture below. $t_{.05, 19df} = 1.73$



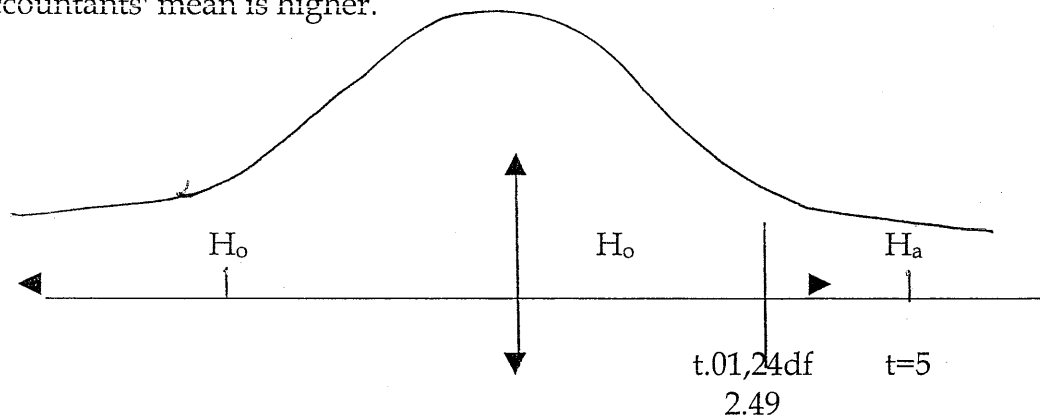
Since $.30 < 1.73$, we accept H_0 . There is no difference (statistically significantly) between Saranac Lake and the United States in longevity.

Let us analyze whether accountants make more than teachers, based on a sample of 25 accountants and the knowledge that all American teachers' average salary (mean) = \$35K. The sample result for accountants was a mean of \$45K, $\sigma = 10$. If we do not know the population standard deviation, we use the sample standard deviation. Usually we do not know population statistics since populations are typically very large. Consider the calculation for our problem below:

$$t = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{45 - 35}{10 / \sqrt{25}} = \frac{10}{2} = 5$$

If we select $\alpha = 1\%$, $t_{.01, 24df} = 2.49$, consider H_0 : the means are equal.

H_a : accountants' mean is higher.



Since $5 > 2.49$, we conclude H_a : accountants make statistically significantly higher salaries than teachers ($\alpha = .01$).

TI-83 Test for One Sample Mean

Our TI-83 has two separate menus, one for the z test (TI Guidebook 13-10) and one for the t-test (TI Guidebook 13-11). We have left out analyses of $\mu \neq \mu_0$ (means not equal). This is because the real world is usually testing whether one average is higher than another. We are typically testing whether one group lives longer, makes more money, is more satisfied.....than another. You can either use Data if you have sample data entered in L_1, L_2 , etc. Otherwise you can enter the data, \bar{X} , n , σ by the Stats option.

The calculated results will give you the same z or t that we calculated by hand. The major result is the p value. This leads to the probability that the results were statistically significant. For example, if $p = .3$, this is the Type I error associated with concluding that one mean was higher than the other. Since 5% is the maximum conventional α level, a p value of .3 means that the result was not statistically significant.

If you obtained a p value less than .05, the results are statistically significant with Type I error $\alpha = .05$. If you obtained a p value less than .01, your results are statistically significant at the .01 level. You may conclude that one mean is significantly greater than the other.

Let us use the TI-83 calculator to test whether a sample of 10 teachers' salaries in New York State were higher than the national average of \$35K. Our New York State sample was 28, 50, 60, 38, 85, 18, 62, 82, 90, 42.

1. Enter Sample in List 1

2. Go to STAT
3. Press ◀
4. Go to t-test (#2)
5. Enter
6. Input DATA
7. $\mu_o = 35$
8. List: L₁
9. Freq = 1 (each element in the sample is used once)
10. $\mu > \mu_o$ (Enter)

This tests whether the sample mean is greater than 35K

11. Calculate

The result is: $t = 2.62$

$$p = .013$$

$$\bar{X} = 55.5$$

$$S_x = 24.74$$

$$n = 10$$

This result means that New York State teachers' salaries are higher than the national average ($\alpha = .013886$).

So if you tested this hypothesis with $\alpha = .05$, you can conclude H_a , the sample mean was higher. If you set $\alpha = .01$, you have to conclude H_o , the two means are equal. This is because $.013886 > .01$. You are less than 99% confident

that the sample of teacher salaries was significantly greater than the national average.

Homework

Complete each problem with both the TI-83 and by calculating the computed t or z value from the given data.

1. Engineers are said to average \$50K per year. A sample of 36 engineering salaries revealed a mean of \$55K with standard deviation of 10K. Is the sample result greater than the population average? Let $\alpha = .05$, $\alpha = .01$.
2. The salaries of eight doctors in Saranac Lake were 50K, 36, 72, 64, 56, 60, 70, 80. Are they below the national average of 80K? Use the sample standard deviation for your population standard deviation. $\alpha = .05$, $\alpha = .01$
3. Saranac Lake house prices were found as 80K, 110, 65, 90, 64, 86, 120, 72 in July. Are they below the national average of \$100K, $\sigma = 10$? $\alpha = .05$, $\alpha = .01$
4. A sample of 100 health records of Alaska residents revealed a mean life span of 80, $\sigma = 10$. Is this result above the national average of 75? $\alpha = .05$, $\alpha = .01$

EXPERIENCE 7

Survey 30 or more students about any measure, their GPA at college, how many hours they study or work per week. Make a guess as to the mean before you gather data. Test whether the true mean is larger (or smaller) than your guess. Use $\alpha = .05$ and $.01$

CHAPTER EIGHT - TESTS OF TWO SAMPLE MEANS

This section will give you the tools to compare salaries between two different occupations. You could test whether house prices are higher in one town as compared to another. The more you plan and use Statistics, the less likely you are to encounter financial difficulties during your life.

I researched two cities that were considered to be very desirable places to live - Marianna, Florida and Saranac Lake, New York. These two choices were low in crime, low in median house price, and recommended retirement places for baby boomers.

I took a sample of house prices from each city. Saranac Lake had 30 houses to choose from. The median price was \$63,000. This means that half the houses cost more; half the houses cost less. My sample was from a decade ago. Prices have risen.

If you decided to move to Marianna, you had 45 houses from which to pick. The median price was \$92,000.

If you decided to move to Scarsdale, the median house price last year was \$645,000. You have to make a great deal of money to support a \$645,000 house price, which does not include taxes and house maintenance.

In order to compare samples of houses, the median is a good place to start, but it does not allow you to determine whether one sample mean is statistically significantly higher than another. You need the mean and standard deviation for

each sample. Then you substitute in the formula below if you are comparing two large samples (each sample has 30 or more elements).

Two Large Sample Difference of Mean Formula:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

\bar{X}_1 = sample mean for group 1

\bar{X}_2 = sample mean for group 2

σ_1^2 = sample variance for group 1

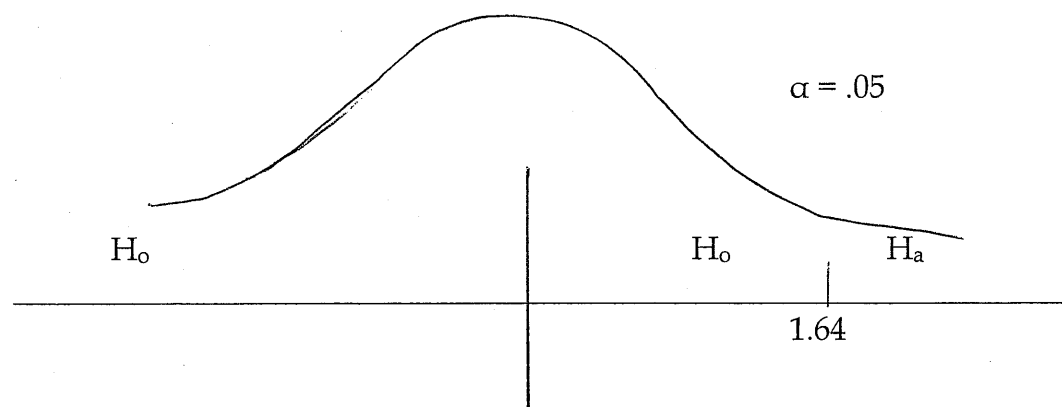
σ_2^2 = sample variance for group 2

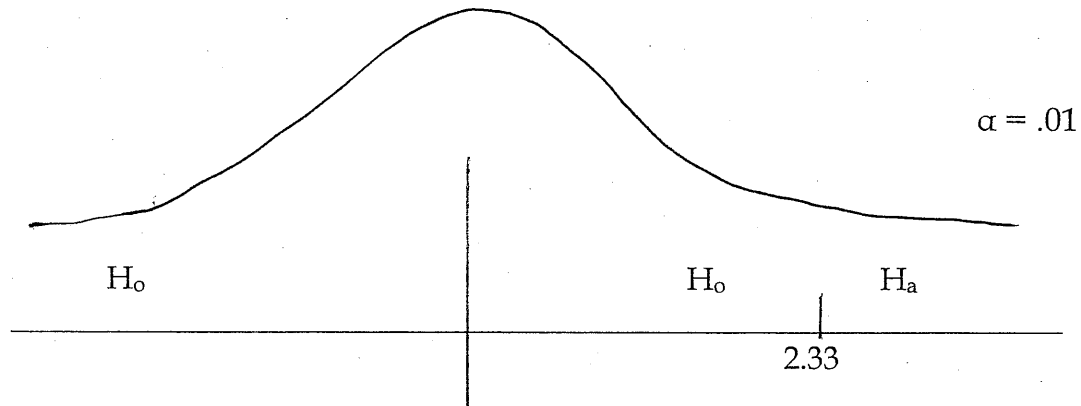
Then you test whether

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 > \mu_2$$

Using two pictures





To illustrate, suppose you are comparing whether attorneys make more than doctors. Suppose you find 40 attorneys' salaries which have a mean of 100K, $\sigma = 12$. We are being very flexible in using σ in place of S , the sample standard deviation. σ refers to the population standard deviation, which we rarely know. We usually substitute S for the population standard deviation σ . All data is in thousands.

A sample of 36 doctors earned an average income of 98K, $\sigma = 10$. Now substitute the values in the large sample formula below.

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{100 - 98}{\sqrt{12^2/40 + 10^2/36}} = \frac{2}{2.525}$$

$$Z = .79$$

If we let $\alpha = .05$, $.79 < 1.64$. See Picture I. Our conclusion is:

Accept H_0 : $\mu_1 = \mu_2$

OR there is no statistically significant Z value from which you could conclude that lawyers earn more than doctors. If $\alpha = .01$, $.78 < 2.33$. Your conclusion is the same.

Often you do not have one or both large samples. For example, biotechnology trials frequently use samples smaller than 30. Some hospitals, who are researching new drugs, have a limited number of patients. They must then take half of the sample for a control group, which receives a placebo (or no drug treatment).

We can easily adjust statistics to the small sample case by using the formula below:

Two Small Sample Difference of Means ($n_1 < 30$ or $n_2 < 30$)

Follow these steps:

1. Compute the pooled variance

First calculate σ_1^2 , σ_2^2 . Then find $S_p^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}$

2. Compute the two sample means, \bar{X}_1 and \bar{X}_2

3. Compute $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2(1/n_1 + 1/n_2)}}$

Now adjust your previous large sample critical value by looking up your t value ($\alpha = .05$ or $\alpha = .01$) with $(n_1 + n_2 - 2)$ degrees of freedom. We are limiting our discussion to the most common comparison: $H_0: \mu_1 = \mu_2$ against $H_a: \mu_1 > \mu_2$.

Please follow this example. Suppose cancer patients starting chemotherapy with drug A live 12 years longer on average, $\sigma_1 = 4$, sample size of 10. The placebo resulted in an average of 8 additional years, $\sigma_2 = 3$, sample size of 10. Can we conclude that drug A is effective in prolonging life?

Let $\alpha = .05$, which is what the FDA requires before a new drug can be marketed nationally. Substitute in the small sample t formula:

$$1. \quad \sigma_1^2 = 16, \sigma_2^2 = 9, n_1 = 10, n_2 = 10$$

$$S_p^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}$$

Substitute σ for S .

$$S_p^2 = \frac{(10-1)16 + (10-1)9}{10 + 10 - 2}$$

$$S_p^2 = \frac{9 \cdot 16 + 9 \cdot 9}{18} = 12.5$$

You should always obtain a value of S_p^2 somewhere between (or equal) to the values of S_1^2 and S_2^2 . The pooled variance S_p^2 is an average of the two variances and it enables us to use the t test, which adjusts for the small sample size.

$$2. \quad \bar{X}_1 = 12, \bar{X}_2 = 8. \text{ These are the values of the two sample means.}$$

$$3. \quad \text{Now calculate the computed t value.}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2 (1/n_1 + 1/n_2)}}$$

$$t = \frac{12 - 8}{\sqrt{12.5(1/10 + 1/10)}}$$

$$t = \frac{4}{1.5811} = 2.5298 = 2.53$$

Note, to calculate $\sqrt{12.5(1/10 + 1/10)}$, there are several ways to compute.

One way is as follows:

1. .1 + .1
2. Enter
3. \times 12.5 (multiply by 12.5). Enter
4. Press \wedge This is the exponent key.
5. .5 (the square root means the $\frac{1}{2}$ power or 5/10).
6. Enter

The answer is 1.58113883. Now look up your critical t in the t table.

$$\text{degrees of freedom} = n_1 + n_2 - 2$$

Set $\alpha = .05$, which is what the FDA requires for the approval of a new drug for cancer treatment (or virtually any new experimental trial).

$$df = 10 + 10 - 2$$

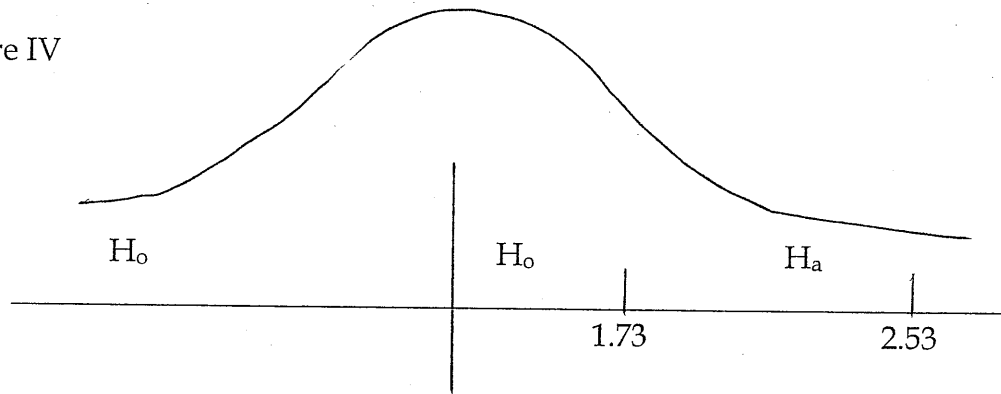
$$t_{.05, 18 df} = 1.73$$

Use Picture IV below with:

$H_0: \mu_1 = \mu_2$ (the means are equal from a statistical perspective)

$H_a: \mu_1 > \mu_2$ (people live significantly longer with the treatment utilizing drug A)

Picture IV



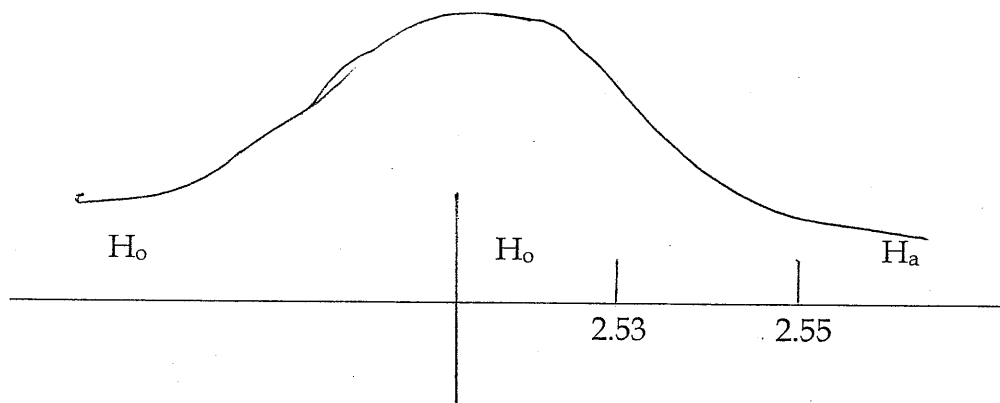
We conclude H_a : Since $2.53 > 1.73$, we accept H_a , the mean of Group I is significantly higher than Group II.

If we decided on an alpha value, $\alpha = .01$, this is a harder level of statistical significance to achieve. It is usually considered too stringent, since many effective treatments may fall short. After all, if you achieved a confidence level of 98% or $\alpha = .02$ for the efficacy of some medical advance, you would have to consider the method a failure at an $\alpha = .01$ level.

Now if we used $\alpha = .01$, look up the critical t value.

$$t_{.05, 18 \text{ df}} = 2.55$$

As a result, we would have to conclude that $\mu_1 = \mu_2$ or the drug did not work if we used $t_{.05, 18 \text{ df}} = 2.55$ for our test. Accept H_0 : $\mu_1 = \mu_2$ The two means are equal from a statistical perspective.



TI-83 Section

The TI-83 is programmed to enable the student to either enter the data in List 1 (Sample 1) or List 2 (Sample 2). Then you can go to the STAT menu and use either #2, the 2-sample Z test or #3, the 2-sample t test.

Let us re-do our previous example using the TI-83 according to the following steps. We summarize the data below:

Cancer patient data: $\bar{X}_1 = 12$

$$\bar{X}_2 = 8$$

$$n_1 = 10$$

$$n_2 = 10$$

$$\sigma_1 = 4$$

$$\sigma_2 = 3$$

$$\alpha = .05$$

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 > \mu_2$$

Now please follow the steps to program the TI-83 to complete the analysis:

1. Press STAT
2. Press \blacktriangleleft
3. Go to #4, 2 samp t test
4. Enter

5. Input stats. Enter

This is because you have the values $\bar{X}_1, \bar{X}_2, \sigma_1^2, \sigma_2^2$. If you have the data in List 1 and List 2, you would press Enter on Data.

Put in the following data:

- 6. $\bar{X}_1 = 12$
- 7. $S_{x1} = 4$
- 8. $n_1 = 10$
- 9. $\bar{X}_2 = 8$
- 10. $S_{x2} = 3$
- 11. $n_2 = 10$
- 12. We want to test $\mu_1 > \mu_2$, so go to $> \mu_2$. Enter
- 13. Pooled go to Yes. Enter
- 14. Calculate. Enter

Our t value confirms our work. $t = 2.5298$ $p = .0104$. As we know, we narrowly missed statistically significant results with $\alpha = .01$ ($p = .01$) but easily concluded that $\mu_1 > \mu_2$ at $\alpha = .05$.

Homework

- 1. Redo the comparison of attorney and physician salaries using the TI-83, $\alpha = .05$.
- 2. If Philadelphia's average house price is \$210K and Atlanta's is \$185K, where the sample size for Philadelphia was 100, $\sigma_1 = 50$, the sample size for Atlanta was 60, $\sigma_2 = 40$, can we conclude that Philadelphia's mean house price

is significantly greater than Atlanta's, $\alpha = .05$? Use the computed Z value with the formula and check your work with the TI-83.

3. The following is a list of salaries for teachers in public schools (List 1) and a list of salaries for private schools (List 2). Use both calculated t and the TI-83 to analyze whether public school teachers make significantly more than private school teachers. Use $\alpha = .05$ and $\alpha = .01$.

Public ($n_1 = 12$) 65, 55, 36, 80, 42, 33, 36, 38, 35, 40, 45, 50

Private ($n_2 = 11$) 26, 60, 35, 30, 53, 28, 30, 28, 28, 29, 29

4. Test whether people who live in the Adirondacks live longer than people who live in Mariana County based on an average life span in Saranac Lake of 75.27 with standard deviation 4, based on a sample of 20. In Mariana, the average life span was 74.26, with standard deviation of 4, based on a sample of 30. Use both the small sample t formula and the TI-83 to test your hypothesis with $\alpha = .05$ and $\alpha = .01$.

EXPERIENCE 8

Test any hypothesis with a sample from two populations. For example, do women have a higher GPA than men. Was your hypothesis confirmed or rejected by your study? What have you learned?

CHAPTER NINE - DIFFERENCE OF PROPORTIONS

Many times, there are only two possibilities in a category - success or failure in sports, defective or properly working in a mechanical process, divorced or continued married, or employed/unemployed. We can call P the sample proportion of success and define P as:

$$P = \frac{\text{number of successes in the sample}}{\text{total sample size}}$$

For example, if ten transmissions are defective from auto dealer A, the proportion of proper functioning transmissions is:

$$P = \frac{990}{1000} = .99$$

If you want to get several job offers, it may be necessary to send out hundreds, if not thousands, of resumes and cover letters. Sometimes the number of resumes it takes to have one job interview may be in the range of 100 to 1, sometimes 1000 to 1.

The ratio of successes P is defined as 1/100 (1%) or 1/1000 (.1%). You could use knowledge of job demand to your advantage in selecting your college major or your career change. For example, if 29 of 30 graduates of Columbia Law School obtained legal positions with starting salaries about \$80,000, your admission to law school would be a likely ticket to professional success. Of course, you should inspect what percentage of students who are admitted to Columbia Law School drop out. If 200 students enter and only 30 graduate, the 29/30 success ratio is not so impressive as a placement statistic.

It is natural to compare two ratios to decide whether one is significantly higher than another. For example, employers have to be sensitive to issues related to job discrimination. In Introductory Statistics (J. Devore and R. Peck, West, 1994, NY: p. 344-345.), the authors cite an article that discussed the court case Swain v. Alabama (1965) in which there was alleged discrimination against blacks in grand jury selection. The census data indicated that blacks constituted 25% of the grand jury pool and a random sample of 1050 people called for jury service yielded 177 blacks. We can use the following formula and our earlier two large sample Z pictures to test the hypothesis H_0 vs. H_a where:

$$H_0: P = \bar{P}$$

$$\bar{P} = \text{the sample proportion}$$

$$P = \text{the population proportion}$$

$$H_a: P > \bar{P}$$

The population proportion is significantly greater than the sample proportion.

One Sample Formula

Use this formula when you have one sample that you are comparing to a population statistic. The one sample z test for a population proportion is:

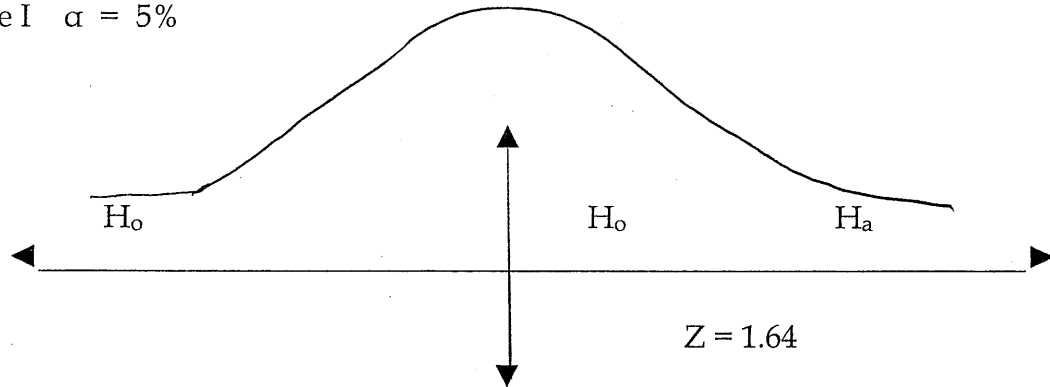
$$Z = \frac{P - \bar{P}}{\sqrt{\bar{P}(1-\bar{P})(1/n)}}$$

Use the population proportion for the denominator

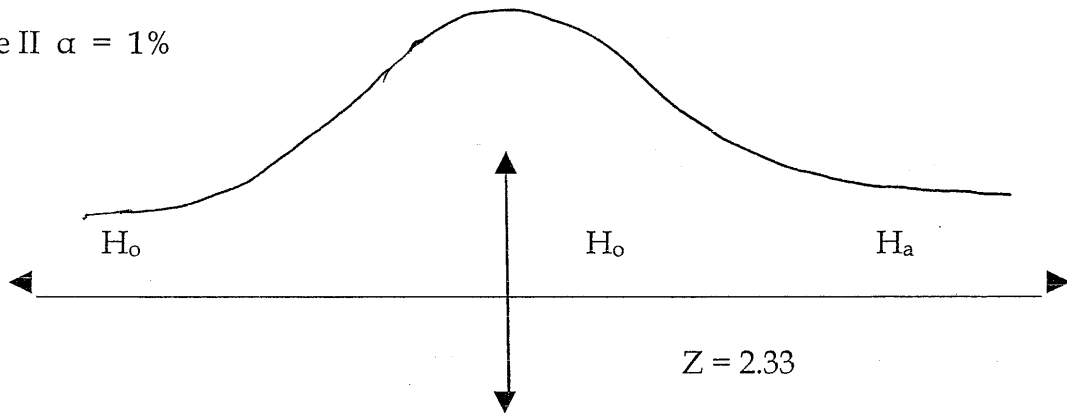
where n = sample size

You can use either Picture I or Picture II to test your hypothesis:

Picture I $\alpha = 5\%$



Picture II $\alpha = 1\%$



Now test whether our result of 177 blacks from a sample of 1050 is significantly less than the population proportion of 25%. To simplify the algebra, use the higher proportion as the first value. This avoids negative numbers and a shifting of Picture I and II to the negative side of the x-axis.

$$Z = \frac{P - \bar{P}}{\sqrt{P(1-P)(1/n)}}$$

$$P = 25\%$$

$$\bar{P} = 177/1050 = .1686$$

$$n = 1050$$

$$Z = \frac{.25 - .1686}{\sqrt{.25(1-.25)(1/1050)}}$$


$$Z = \frac{.0814}{.0134} = 6.07$$

This value is much higher than $Z = 2.33$ (Picture II). We can conclude that $H_a: P > \bar{P}$ with much higher confidence than 99%. Lawyers sometimes use 99% as the probability corresponding to "beyond a shadow of a doubt," a key element of criminal prosecution. We may conclude that there was discrimination against blacks at any conventional level of Type I error.

The court only looked at the numerator and observed a difference of only 8%. They concluded (incorrectly) that the difference was not large enough to establish a prima facie case (a case without further examination).

TI-83

Let us complete the statistical analysis (using the TI-83) of the discrimination case data. Use the following steps:

1. Press STAT
2. 
3. Go to #5: 1 Prop Z Test
4. Enter
5. P_o : .25
6. X: 177
7. n: 1050

8. Prop < P_o (population proportion)

We are testing whether our population is greater than our sample proportion. This is simply reversing the inequality sign. For example, $10 > 5$ is the same expression as $5 < 10$.

9. Enter

10. Calculate

Z = -6.0935563

P = .999999

The Z value is negative, because the calculator uses the value $(\bar{P} - P)$ in the numerator where we used $(P - \bar{P})$ to avoid negative numbers. The slight discrepancy between our Z of 7.02 and the calculator result is due to the calculator using 9 significant digits after the decimal. The P value is the main issue to keep in mind. The result (accept H_a) is statistically significant at any conventional level of α . You can have .999999 etc. confidence in your result. Discrimination is well demonstrated by our objective statistical analysis.

Two Sample Difference of Proportions

The stock market has made a great number of millionaires in the 1990s. Stocks like EMX and Dell have gone up way over 1000% in the past ten years.

One guide to assist you in picking the Dells and EMXs of the 2000s is to consult analysts' ratings. Analysts rate stock with 1 = strong buy, 2 = buy, 3 = hold, and 4 = sell. The best rating a stock could have is a mean of 1; the worst is a 4.

Analysts tend to be optimists so you might look at the fraction of analysts who rate a stock as a strong buy. On August 30, 1999, a visit to www.morningstar.com revealed the following analyst ratings.

$$\frac{\text{Strong Buys}}{\text{Total \# of Analysts}} = P$$

$$\text{Intel} \quad P = 15/36$$

$$\text{AOL} \quad P = 25/40$$

$$\text{Dell} \quad P = 13/32$$

It is natural to compare the strength of analyst ratings of two stocks, such as America On Line compared to Dell. The Z test for the difference of proportions is the test of choice. The test has several requirements, before you can use it. Consider the following:

$$n_1 = \text{sample size for sample 1}$$

$$n_2 = \text{sample size for sample 2}$$

$$P_1 = \frac{\text{\# of successes in sample 1}}{n_1} = \frac{x_1}{n_1}$$

$$P_2 = \frac{\text{\# of successes in sample 2}}{n_2} = \frac{x_2}{n_2}$$

$$\bar{P} = \text{Pooled Proportion} = \frac{x_1 + x_2}{n_1 + n_2}$$

Requirements:

1. $n_1 \geq 30$
2. $n_2 \geq 30$
3. $n_1 P_1 \geq 5$

$$4. \quad n_2 P_2 \geq 5$$

$$5. \quad n_2 (1 - P_2) \geq 5$$

Let us test our 5 requirements in the test of whether the rating of American On Line (AOL) is higher than the rating of Dell.

$$P_1 = \frac{x_1}{n_1} = \frac{25}{40}$$

$$P_2 = \frac{x_2}{n_2} = \frac{13}{32}$$

To test the five requirements:

$$1. \quad n_1 = 40 > 30$$

$$2. \quad n_2 = 32 > 30$$

$$3. \quad n_1 P_1 = 40 \cdot 25/40 = 25 > 5$$

$$4. \quad n_2 P_2 = 32 \cdot 13/32 = 13 > 5$$

$$5. \quad n_2 (1 - P_2) = 32(1 - 13/32) = 32(19/32) = 19 > 5$$

The five requirements are satisfied, so we can use the following test:

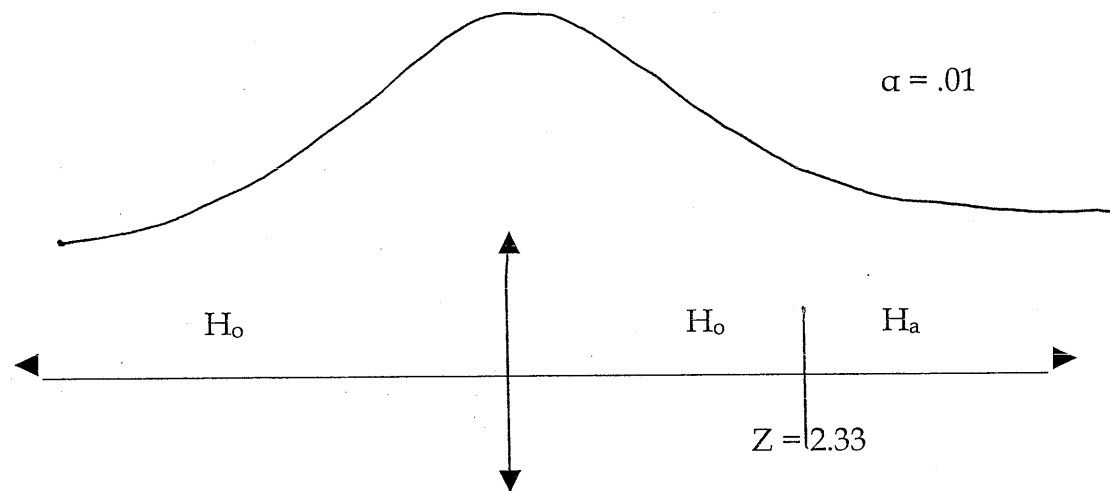
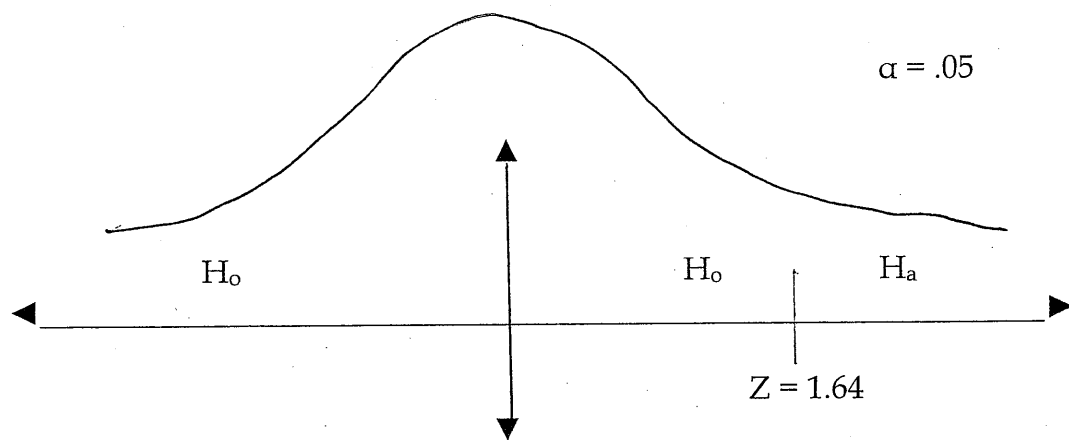
$$H_0: P_1 = P_2$$

The two proportions are equal.

$$H_a: P_1 > P_2$$

P_1 is significantly greater than P_2 .

Set $\alpha = .05$ or $.01$ and use Picture I or II below.



The formula for a 2 sample difference of proportion is:

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}}$$

Substitute: $P_1 = 25/40$, $P_2 = 13/32$, $n_1 = 40$, $n_2 = 32$

$$\bar{P} = \frac{25 + 13}{40 + 32} = \frac{38}{72}$$

$$Z = \frac{(25/40) - (13/32)}{\sqrt{\frac{38/72(1-38/72)}{40} + \frac{38/72(1-38/72)}{32}}}$$

$$Z = \frac{.625 - .406}{\sqrt{.0062 + .0078}} = \frac{.219}{.118}$$

$$Z = 1.86$$

Use Picture I and Picture II

If you select $\alpha = .05$, your conclusion is accept H_a : $1.86 > 1.64$

•• $P_1 > P_2$, AOL has a significantly higher rating than Dell.

If you select $\alpha = .01$, your conclusion is accept H_o : $1.86 < 2.33$

•• $P_1 = P_2$, AOL has an equal rating as Dell.

This remarkable result, two contradictory findings based on your selected α level, illustrates how essential an understanding of Statistics is to your financial and personal life. Most studies rely on an α level of 5%, so as to not reject significant results that usually fall short of the highly stringent $\alpha = .01$ level.

TI-83

Let us complete the same problem using the TI-83's statistical program.

Use the following steps:

- 1, STAT
2. ◀
3. Go to #6 - 2 Prop Z Test
4. Enter
5. X_1 : 25
6. n_1 : 40
7. X_2 : 13

8. n_2 : 32
9. $P_1 > P_2$ Enter
10. Calculate Enter

The result is $Z = 1.85$; the P value is .0323. This confirms our earlier result.

We concluded that $P_1 > P_2$ ($\alpha = .05$) and $P_1 = P_2$ ($\alpha = .01$). The precise level of alpha is .03. If you set alpha lower than .03, you fail to accept H_0 . If you set alpha higher than .03 ($\alpha = .05$), you can conclude $P_1 > P_2$.

Homework

1. Do more than 50% of analysts rate AOL as a strong buy? $\alpha = .05$, $\alpha = .01$. Use both the formula and the TI-83.
2. You are told that 10% of your resume letters will result in an interview. If you send out 1000 letters and receive 60 requests for an interview, is this significantly below the advertised claim of 10%? Let $\alpha = .05$. Use both the formula and the TI-83.
3. Intel has a strong buy rating from 15 analysts of 36 who have rated the stock (August 31, 1999). Is this significantly lower than the 25/40 rating AOL a strong buy? Use $\alpha = .05$, .01 and both formula and TI-83.
4. Presidential candidate A was given a favorable rating by 100 of 250 respondents. Candidate B was given a favorable rating by 90 of 300 respondents. Does Candidate A demonstrate significantly higher support at $\alpha = .05$? Use both the formula and TI-83.

EXPERIENCE 9

Ask two samples of thirty or more any question with a yes/no or true/false response. Compare the two resulting fractions. Was the % of the group response higher as you expected? Use $\alpha = .05$ and $.01$. What have you learned?

CHAPTER TEN - LINEAR CORRELATION/REGRESSION

The stock market has made a lot of people very wealthy in the 1990s. One of the key factors that has increased the value of stocks is the increase in the earnings of stocks in the 90s. For example, if IBM made an average of \$1 per share for 1990 and earned an average of \$5 per share in 2000, the stock may increase in price by a factor of 5.

Of course, there are many factors that relate to stock price besides earnings. The purpose of this section is to develop a method for determining the most accurate line, $y = mx + b$, that relates variables such as earnings (x) to y (stock price). Also, the Siamese twin of the line is the related topic, linear correlation, which measures the strength of the linear relationship between x and y .

Sometimes, there is a perfect linear relationship between two variables. For example, if you go to a store that charges 7% sales tax, the equation $\text{Tax} = y = .07 (\text{purchase}) = .07x$ is a deterministic relation between x and y . If you buy \$100 worth of goods, your tax is determined as follows:

$$y = .07x = .07(100) = \$7$$

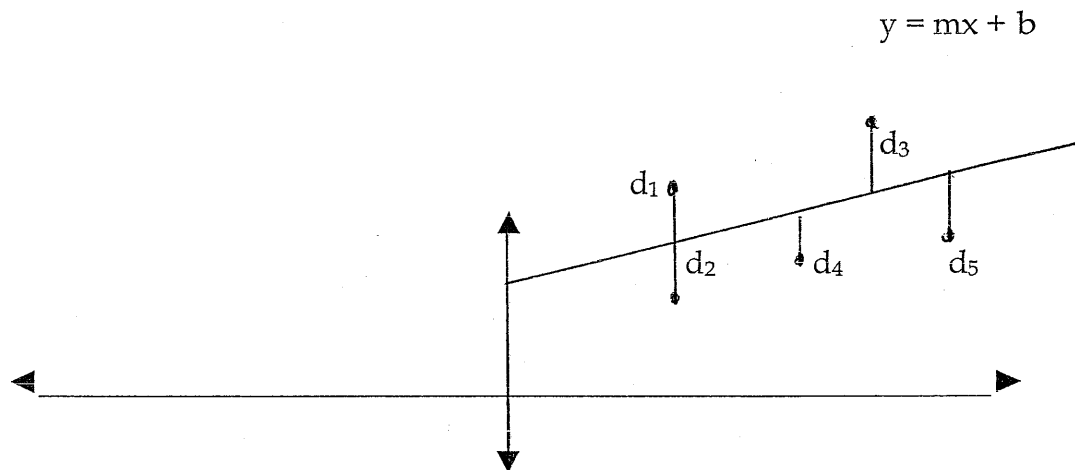
There is no error. You have a perfect prediction.

In real life, the prediction of one variable (y) based upon an independent variable (x) is rarely perfect. In fact, stock prediction could be improved by a complex regression method called multiple regression, where stock price $y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$. Some models use 5 predictors of stock price, while

others use dozens. This topic is too complicated for our brief introduction to key topics from Statistics.

Before you decide upon a method to link two variables, it is wise to plot the points (x,y) . The points are called the scatter plot and may show whether a linear relationship $(y = mx + b)$ exists between x and y or perhaps a quadratic relationship $(y = ax^2 + bx + c)$ or an exponential relationship $(y = e^x)$ or a variant.

We are going to use the principle of least squares. See the picture below.

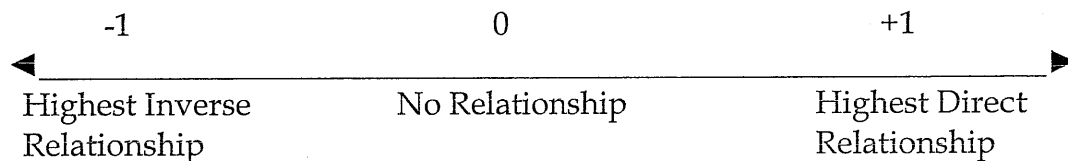


The d 's represent the distance between five points and the best fitting line $y = mx + b$, which minimizes the sum of the squares of the differences in y between the actual y value and the y value of the regression line.

Let us use data from Barrons (August 30, 1999), which gives the current price y of selected stocks and the analysts' expected earnings (per share) for the year 1999. The data for 5 stocks is presented below:

	<u>x (earnings)</u>	<u>y (price)</u>
America On Line	.60	99 1/8
EMX	1.06	60 9/16
General Electric	3.21	116 9/16
General Motors	4.55	65
IBM	3.91	124

Let us calculate the coefficients $y = mx + b$ and the correlation coefficient r for the data. First calculate the correlation coefficient, r .



The correlation coefficient, r , ranges between -1 and +1. The closer the r value is to -1, the higher the inverse relation between x and y . For example, if you are analyzing major league pitchers' salaries, there is a strong inverse relation between your earned run average (average number of runs you allow per game) and your salary. The lower the number of runs you allow, the higher your salary will be.

The relation between your level of education (x) and your future salary (y) is nearly a perfect 1. The more education you pursue (masters, medical degree, etc.), the more money you will likely make. No relation is a perfect +1 or -1, except for trivial deterministic relations.

The formula for calculating the correlation coefficient, r , is below:

$$r = \frac{n \sum (xy) - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

Complete the table for our given data, which we rounded to the nearest whole number to facilitate calculation.

	x	y	xy	x ²	y ²
AOL	1	99	99	1	9801
EMX	1	61	61	1	3721
GE	3	117	351	9	13,689
GM	5	65	325	25	4,225
IBM	4	124	496	16	15,376

Calculate the following:

$$n = 5 \text{ (# of pairs)}$$

$$\sum x = 14$$

$$\sum y = 466$$

$$\sum xy = 1,332$$

$$\sum x^2 = 52$$

$$\sum y^2 = 46,812$$

Substitute in the formula for r :

$$r = \frac{n \sum (xy) - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

$$r = \frac{5(1332) - (14)(466)}{\sqrt{5(52) - 14^2} \sqrt{5(46,812) - 466^2}}$$

$$r = \frac{136}{\sqrt{64} \sqrt{16,904}} = \frac{136}{1040} = .13$$

This is a low positive correlation. You have to look for other factors besides earnings to predict stock price with a good level of accuracy.

You can use the calculations from your correlation analysis to determine the least-squares line $y = a + bx$. This line is a predictor of y based upon values of x .

The slope of the least squares line is

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

Use the values from your correlation analysis to substitute:

$$n \sum xy - (\sum x)(\sum y) = 136$$

$$n \sum x^2 - (\sum x)^2 = 64$$

$$b = \frac{136}{64} = 2.1$$

To find the y intercept of $y = a + bx$, use the formula:

$a = \bar{y} - b\bar{x}$, where \bar{y} , \bar{x} are the means of y and x respectively.

$$\bar{y} = \frac{\sum y}{n} = \frac{466}{5} = 93.2$$

$$\bar{x} = \frac{\sum x}{n} = \frac{14}{5} = 2.8$$

$$a = \bar{y} - b\bar{x} = 93.2 - 2.1(2.8)$$

$$a = 87.3$$

The least squares line $y = a + bx = 87.3 + 2.1x$

You can use the line, $y = 87.3 + 2.1x$ to predict y based on various values of x . For example, if $x = 3$, we can predict $y = 87.3 + 2.1(3) = 93.6$

It is important to evaluate how well the least squares line predicts y . A standard method for evaluating the effectiveness of a least square line is the coefficient of variation. If the points in a scatter plot are close to the least squares line, the line is a good fit.

Again, please take a complete Statistics course. This topic will be covered in depth in a standard elementary statistics course.

First compute the residual sum of squares, SSRESID. This gives you the sum of the squares of the differences between your given y values and your predicted y values, given the line $y = a + bx$.

$$SSRESID = \sum y^2 - a \sum y - b \sum xy$$

Next compute $\sum (y - \hat{y})^2 = SSTO$

This gives you the squared differences between your y values and the mean.

You can use the short cut formula: $SSTO = \sum y^2 - \frac{(\sum y)^2}{n}$

Let us proceed with our calculations:

$$SSRESID = \sum y^2 - a \sum y - b \sum xy$$

$$SSRESID = 46,812 - 87.3(466) - 2.1(1332)$$

$$SSRESID = 3333$$

$$SSTO = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$SSTO = 46,812 - \frac{466^2}{5}$$

$$SSTO = 3381$$

We now compute the COEFFICIENT OF DETERMINATION

$$R^2 = 1 - \frac{SSRESID}{SSTO} = 1 - \frac{3333}{3381} = .01$$

R^2 ranges between 0 and 1.

$R^2 = .01$ for our least squares line

The closer the value of R^2 is to 1, the better prediction you can make of y based upon x. Since our R^2 value is so small, we conclude that we cannot very effectively predict stock price on the basis of earnings. Our current stock market is based on high growth stocks with currently low earnings. Our result confirms that earnings is not a key factor in determining a stock's price.

We can use the TI-83 to confirm our work. Let us work out the same problem with the TI-83.

1. Put the values of x in List 1: {1,1,3,5,4} \longrightarrow L1
2. Put the values of y in List 2: {99,61,117,65,124} \longrightarrow L2
3. STAT
4. CALC
5. Go to #4 Lin Reg (ax + b)
6. Enter
7. Enter

The result is: $y = ax + b$

$$a = 2.125$$

$$b = 87.25$$

This confirms our earlier result. To calculate the correlation, R , follow these steps:

1. 2nd catalog
2. Go to Diagnostician On
3. Enter
4. Enter
5. Go to STAT
6. CALC
7. Go to #4 Lin Reg ($ax + b$)
8. Enter
9. Enter

Your result is: $R^2 = .017$

$R = .13$ which confirms our conclusions that R^2 is a low value and our least squares line is not very accurate.

Homework

From Barrons (August 30, 1999)

	This Year's Earnings x	Next Year's Earnings y
AOL	.60	.60
EMX	1.06	1.37
GE	3.21	3.66
GM	4.55	8.15
IBM	3.91	4.49

1. Find the correlation between x and y - the two years' earnings.

Find R^2 - the coefficient of variation. Interpret the results.

	x(above)	y earnings growth (%)
AOL	1	.54
EMX	1	.29
GE	3	.14
GM	5	-.03
IBM	4	.15

2. First round x (from #1) to the nearest whole number.

Let y = earning's growth (in %) for the following year.

Compute the correlation R and the least squares regression line $y =$

$a + bx$, based upon the data. Use both the formula and the TI-83 for your

calculations. Interpret your results.

3.

x (education)		y (average annual salary - in thousands)
High School	0	20
Two Years College	1	25
Four Years College	2	47
Master's Degree	3	55
Law/Medical Degree	4	80

Let $x = 0$ (high school)

$x = 1$ (two years college)

$x = 2$ (four years college)

$x = 3$ (master's degree)

$x = 4$ (law/medical degree)

Calculate the correlation between your level of education and your eventual average annual salary. Also find the least squares line: $y = ax + b$. Use both the computing formula and the TI-83. Estimate y if $x = 2.5$.

4. OPEN QUESTION

Reverse x and y in problem #2. Do you get an equivalent equation with interchanged values of x and y or a slightly different equation? This problem is a BRAIN TEASER and was proposed by Dr. Ben Fusaro, the founder of the International Mathematical Modeling Competition (Florida State University).

EXPERIENCE 10

Obtain a sample of ten to thirty people and ask each to fill out the responses to two items. For example, their level of education (x) and salary (y).

1 - no high school; 2 = high school diploma; 3 = two years college; 4 = four years college; 5 = master's degree; 6 = doctorate or law degree.

Obtain both the linear correlation and regression equation $y = a + bx$. What have you learned?

ANSWER KEY

CHAPTER 3

1. (a) Start at 21 - 1000 per year until 65. 8% year.

$N = 44, I = 8, PV = -1000, PMT = -1000, FV = \text{answer}, P/Y = 1, C/Y = 1,$

END

ANS. 386,505.62

(b) 3% ANS. \$92,719.86

(c) 20% ANS. \$18,281,309.94

2. $N = 48, I = 6, PV = \text{ANS}, PMT = -300, FV = 0, P/Y = 12, C/Y = 12, \text{END}$

ANS. 12,774.10

3. $N = 360, I = 7.5, PV = -200,000, PMT = \text{ANS.}, FV = 0, P/Y = 12,$

$C/Y = 12, \text{END}$

ANS. 1398.43

4. $N = 84, I = \text{Answer since I left out I (my error)}, PV = -55,200, PMT =$

$846.67, FV = 0, P/Y = 12, C/Y = 12, \text{END}$

Int % = 7.5%

CHAPTER 4

1. 80%

2. 20%, 80%

3. (a) $\frac{201.6}{100,000} = .0020$

$$(b) \quad \frac{179.6}{100,000} = .0018$$

$$(c) \quad \frac{307.4}{100,000} = .0031$$

$$(d) \quad \frac{134.6}{100,000} = .0013$$

$$4. \quad 1 - .043 = .957$$

CHAPTER 5

$$1. \quad \begin{array}{lll} S_x & = & 6.099 \text{ Sample Standard Deviation} \\ \sigma_x & = & 5.455 \text{ Formal Def. St. Deviation} \end{array}$$

Variance

$$\text{Sample} \quad = \quad 6.099^2 = 37.0881$$

$$\text{Formal Def.} \quad = \quad 5.455^2 = 29.7570$$

$$\text{Mean} = \quad X = 35.2$$

Use Sample St. Deviation

$$\begin{array}{llll} \text{Coefficient} & = & \frac{6.099}{35.2} \cdot 100\% & = & 17.33\% \\ \text{of Variance} & & & & \end{array}$$

$$2. \quad \text{Check}$$

$$3. \quad (a) \quad X = 9.833$$

$$(b) \quad \begin{array}{lll} S_x & = & 9.517 \text{ st. dev.} \\ S_x^2 & = & 9.517^2 = 90.573 \end{array}$$

$$\text{coefficient of variation} = \frac{9.517}{9.833} \cdot 100\% = 96.816\%$$

$$4. \quad X = \text{mean} = 131.6667$$

$$S_x = \text{st. dev.} = 20.2072$$

$$S_x^2 = 20.2072^2 = 408.3309$$

$$\begin{array}{lcl} \text{coefficient} & = & \frac{20.2072}{131.6667} \cdot 100\% = 15.347\% \\ \text{of variation} & & \end{array}$$

Using the formal def. for variance will give slightly different results. The short cut formula is commonly used and gives a sound estimate of the population variance. The calculus based statistics course will clarify these two seemingly different answers.

CHAPTER 6

1. 99% C.I. = $24.8 \pm t_{.005, 9df}(3.8816/\sqrt{10})$
 = $24.8 \pm 3.250 (3.8816/\sqrt{10})$
 = $(20.811, 28.789) = (21, 29)$
2. Check
3. 95% C.I. = $(363, 1250)$
 99% C.I. = $(169, 1444)$
4. 95% C.I. = $50 \pm 1.96 (10/\sqrt{36}) = (47, 53)$
 99% C.I. = $50 \pm 2.57 (10/\sqrt{36}) = (46, 54)$

CHAPTER 7

$$1. \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{55 - 50}{10/\sqrt{36}} = 3$$

Accept H_a , $3 > 1.64$

Engineers are significantly greater than population mean, $\alpha = .05$.

Accept H_a , $3 > 2.33$

Engineers are significantly greater than population mean, $\alpha = .01$.

2. $S_x = 620.263$ (calculator)

$t = 3.878$

$P = .003$

They are below the national average if $\alpha = .05$ or $\alpha = .01$.

3. $t = 1.96$

They are below the national average if $\alpha = .05$ [barely - $P = .045$]

If $\alpha = .01$, they are statistically equal to national average.

4. $Z = \frac{80 - 75}{10/\sqrt{100}} = 5$

$5 < 1.64$, $5 > 2.33$. Accept H_a , Alaska residents live longer than US mean,

$\alpha = 5\%$ or 1%

CHAPTER 8

1. Redo with the TI-83

2. $Z = \frac{210 - 185}{\sqrt{50^2/100 + 40^2/60}} = \frac{25}{719} = 3.48$

Accept H_a - Philadelphia's mean house price is significantly greater than Atlanta - $\alpha = .05$.

3. $t = 2.25$

Accept H_a - Public school teachers make significantly more salary than private school teachers - $\alpha = .05$

Accept H_o - There is no statistically significant difference. $\alpha = .05$ or $\alpha = .01$.

4, $t = .87$

There is no statistically significant difference. $\alpha = .05$ or $\alpha = .01$

CHAPTER 9

$$1. \quad Z = \frac{.625 - .5}{\sqrt{.5(.5)/40}} = \frac{.125}{.079} = 1.58$$

Accept H_0 . $P = .50$. There is no statistically significant difference from $P = .5$,
 $\alpha = .05$ or $\alpha = .01$

$$2. \quad P = .10$$

$$Z = \frac{.10 - 60/100}{\sqrt{.10(.9)/1000}} = \frac{.04}{.0095} = 4.21$$

Accept H_0 . The claim is significantly too high. $\alpha = .05$

$$3. \quad Z = 1.82. \text{ Accept } H_a. \text{ Intel is significantly lower in rating - } \alpha = .05.$$

Accept H_0 . Intel and AOL have no significant difference in rating - $\alpha = .01$

$$4. \quad Z = 2.45$$

Accept H_a . Candidate A demonstrates significantly higher support -
 $\alpha = .05$ or $\alpha = .01$

CHAPTER 10

$$1. \quad R = .996$$

$$R^2 = .99$$

This is extremely high direct variation.

2. (a)

x	y
1	.54
1	.29
3	.14
5	-.30
4	.15

(b) $y = -.11x + .51$

$$R = -.88$$

There is a strong inverse relation (negative correlation) between stock price and earnings growth.

3. $R = .98$

$$y = 15x + 15.4$$

$$\text{if } x = 2.5, y = 52.9 = 53$$

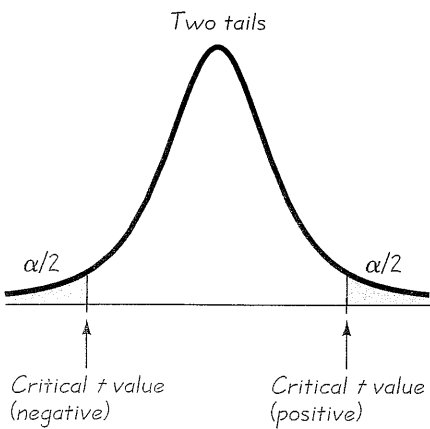
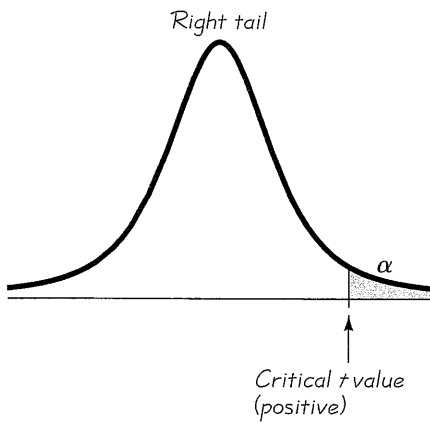
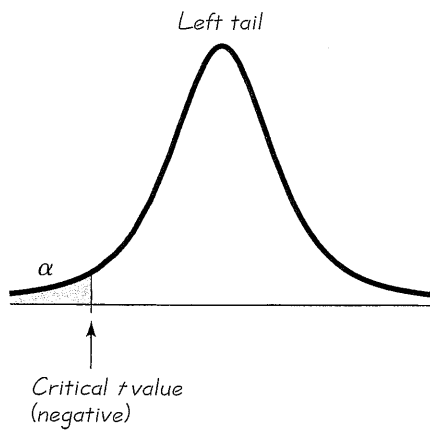
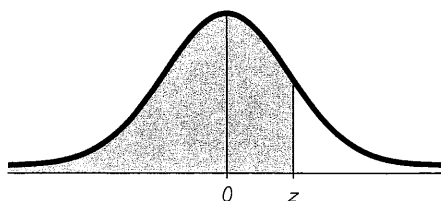


Table A-3 t Distribution: Critical t Values

Degrees of Freedom	0.005	0.01	Area in One Tail 0.025	0.05	0.10
	0.01	0.02	Area in Two Tails 0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.697	1.310
31	2.744	2.453	2.040	1.696	1.309
32	2.738	2.449	2.037	1.694	1.309
33	2.733	2.445	2.035	1.692	1.308
34	2.728	2.441	2.032	1.691	1.307
35	2.724	2.438	2.030	1.690	1.306
36	2.719	2.434	2.028	1.688	1.306
37	2.715	2.431	2.026	1.687	1.305
38	2.712	2.429	2.024	1.686	1.304
39	2.708	2.426	2.023	1.685	1.304
40	2.704	2.423	2.021	1.684	1.303
45	2.690	2.412	2.014	1.679	1.301
50	2.678	2.403	2.009	1.676	1.299
60	2.660	2.390	2.000	1.671	1.296
70	2.648	2.381	1.994	1.667	1.294
80	2.639	2.374	1.990	1.664	1.292
90	2.632	2.368	1.987	1.662	1.291
100	2.626	2.364	1.984	1.660	1.290
200	2.601	2.345	1.972	1.653	1.286
300	2.592	2.339	1.968	1.650	1.284
400	2.588	2.336	1.966	1.649	1.284
500	2.586	2.334	1.965	1.648	1.283
1000	2.581	2.330	1.962	1.646	1.282
2000	2.578	2.328	1.961	1.646	1.282
Large	2.576	2.326	1.960	1.645	1.282



POSITIVE z Scores

Table A-2 (continued) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	*	.9505	.9515	.9525	.9535
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	*	.9951
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.50 and up	.9999									

NOTE: For values of z above 3.49, use 0.9999 for the area.

*Use these common values that result from interpolation:

z Score	Area
1.645	0.9500
2.575	0.9950

Common Critical Values

Confidence Level	Critical Value
0.90	1.645
0.95	1.96
0.99	2.575

