MATH-1230 (DUPRÉ) PRACTICE TEST PROBLEM ANSWERS

FIRST: PRINT YOUR LAST NAME IN LARGE CAPITAL LETTERS ON THE UPPER RIGHT CORNER OF THIS SHEET.

SECOND: PRINT YOUR FIRST NAME IN CAPITAL LETTERS DIRECTLY UNDERNEATH YOUR LAST NAME.

THIRD: WRITE YOUR FALL 2010 MATH-1110 LAB DAY DIRECTLY UNDERNEATH YOU FIRST NAME.

STANDARD INFORMATION: A **Standard Dice** has six faces and each face has a number of spots, that number being a positive whole number no more than six. The total number of spots on any pair of opposite faces is seven. Since there are exactly three pair of opposite faces, the total number of spots on a dice is twenty one. Therefore if we think of the spots as distributed evenly over the faces, there are on average 3.5 spots per face. A **Standard Deck of Cards** has 52 cards, four suits of 13 cards (denominations) each. The four suits are spades ($\spadesuit$), hearts($\heartsuit$), diamonds($\diamondsuit$), and clubs($\clubsuit$). Each suit has 3 face cards: Jack, Queen, King, and an Ace. In each suit, the cards which are not face cards each have a number of spots in the shape of the particular suit, the number of spots being a whole number no more than ten, the card with a single spot being the Ace of that suit. For example, the card with four spots each in the shape of a diamond is called "the four of diamonds($4\diamondsuit$)", whereas the card with the face labelled with "J" and a spade shaped spot is "Jack of spades($J\spadesuit$)". In many card games, an Ace can count as a denomination value of one or as a denomination higher King as a player desires. A Jack has denomination value eleven, a Queen has denomination value twelve, and a King has denomination value 13, in many card games. In the game of Black Jack or Twentyone, all face cards are given denomination value ten and all aces have denomination value eleven or one as desired by the player. A **Standard (American) Roulette Wheel** has a spinner with 38 slots which spins in a large bowl. A ball is sent rolling in opposite direction to the spin of the wheel near the upper rim of the bowl and as both the spinning wheel and ball slow down, the ball falls into one of the 38 slots. Two of the slots are colored green, one labelled zero (0) and the other labelled with two zeros (00), and referred to as "double zero". The remaining slots are each colored either red or black and numbered with the positive whole numbers no more than 36. At the roulette table a player can bet on zero or on double zero or on a specific positive whole number, or on even or on odd or on red or on black or on one through eighteen, or on one through twelve, or on 13 through 24, or on 25 through 36. There are therefore many possibilities for placing bets at the roulette wheel. The game of craps is played on a large oblong table with rounded ends and with vertical walls around the edges. The player who rolls the dice is called "the shooter" and must toss a pair of standard dice so as to hit the table's horizontal surface and bounce off the wall at the opposite end of the table back on the horizontal surface where it finally comes to rest. If any side is touching a vertical wall or is tilted off horizontal or goes off the table, then it does not count and the pair of dice must be tossed again-it is a "do over". Notice that it is virtually impossible to tell what faces will come up when a pair of dice is thrown, or which card will come up when a card is taken off the top of a shuffled deck, or what slot the ball will land on when the roulette wheel is spun. All the processes obey the laws of physics which are completely deterministic. However, the circumstances are guaranteed to make the use of physics to predict the outcomes virtually useless, as there is no practical way to keep track of the information required. Thus, the so-called "random" processes are really "randumb" processes, as they depend on reducing our information below what can be used to make any effective prediction. On the other hand, it is known that a very few people with practice have learned to throw the dice so as to have a fair amount of control over the outcomes and as of this time the major casinos do not admit that as a possibility so these few players have a tremendous advantage at the crap table.

1. Suppose that Sam guesses that his Master Charge card balance is 1450 dollars and his Visa card balance is 2500 dollars. What should he guess for the total he owes on both credit cards in order to be consistent?

ANSWER:
$1450 + 2500 = 3950$ dollars.

2. If a dice is in a box where you cannot see it and I look into the box and see it and tell you it is sitting with one of its faces on the floor of the box and one face on top-it is not tilted, then based only on this information (whether or not you believe it) what should you guess is the probability that the face on top has 4 spots?

ANSWER:
$\frac{1}{6}$

3. If a dice is in a box where you cannot see it and I look into the box and see it and tell you it is sitting with one of its faces on the floor of the box and one face on top-it is not tilted, then based only on this information (whether or not you believe it) what should you guess is the probability that the face on top has an even number of spots?

ANSWER:
$\frac{3}{6} = \frac{1}{2}$

4. If a dice is in a box where you cannot see it and I look into the box and see it and tell you it is sitting with one of its faces on the floor of the box and one face on top-it is not tilted, then based only on this information (whether or not you believe it) what should you guess is the probability that the face on top has an odd number of spots?

ANSWER:
$\frac{3}{6} = \frac{1}{2}$

5. If a dice is in a box where you cannot see it and I look into the box and see it and tell you it is sitting with one of its faces on the floor of the box and one face on top-it is not tilted and the top face has an even number of spots, then based only on this information (whether or not you believe it) what should you guess is the probability that the face on top has 4 spots?

ANSWER:
$\frac{1}{3}$

6. If a dice is in a box where you cannot see it and I look into the box and see it and tell you it is sitting with one of its faces on the floor of the box and one face on top-it is not tilted and the top face has an even number of spots, then based only on this information (whether or not you believe it) what should you guess is the probability that the face on top has 5 spots?

ANSWER:
0

7. If a dice is in a box where you cannot see it and I look into the box and see it and tell you it is sitting with one of its faces on the floor of the box and one face on top-it is not tilted and the top face has an even number of spots, then based only on this information (whether or not you believe it) what should you guess is the number of spots on the top face?

ANSWER:

The only possible values here are 2,4,6 and all three of these values are equally likely, so we should guess the average of these three numbers which is 4.

4

8. If a dice is in a box where you cannot see it and I look into the box and see it and tell you it is sitting with one of its faces on the floor of the box and one face on top-it is not tilted and the top face has an odd number of spots, then based only on this information (whether or not you believe it) what should you guess is the number of spots on the top face?

ANSWER:

The only possible values here are 1,3,5 and all three of these values are equally likely, so we should guess the average of these three numbers which is 3.

3

9. If X is a positive whole number that I have chosen and you think X is three times as likely to be even as odd, then what should you think is the probability that X is odd?

ANSWER:

It is the same as if you have a deck of cards and each has either ”even” or ”odd” written on it, what is the chance of drawing a card that has ”odd” written on it, given that there are three times as many cards with ”even” written on them as with ”odd” written on them. The simplest example is a deck of 4 cards where only one has ”odd” written on it and three each have ”even” written on it. From such a deck, the chance the top card has ”odd” written on it is clearly 1/4.

$\frac{1}{4}$

10. If X is a positive whole number with $X \leq 6$ that I have chosen and you think X is three times as likely to be even as odd, and if K is a symbol that stands for this information, then what is $E(X|K)$?

ANSWER:

Let A denote the statement that X is odd and let B be the statement that X is even. Then just as above, we know

$$P(A|K) = \frac{1}{4}$$

and therefore

$$P(B|K) = 1 - P(A|K) = \frac{3}{4}.$$

We also know that no odd number is any more likely than any other odd number, so

$$E(X|A\&K) = 3,$$

whereas since no even number is any more likely than any other even number,

$$E(X|B\&K) = 4.$$

Now we just apply our general formula that applies whenever there are several statements of which exactly on must be true. In this case we are dealing with, we know either A or B must be true and only one of these two statements can be true. Therefore

$$E(X|K) = E(X|A\&K)P(A|K) + E(X|B\&K)P(B|K) = (3)(\frac{1}{4}) + (4)(\frac{3}{4}) = \frac{15}{4}.$$

$\frac{15}{4}$

11. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the top card is a heart?

ANSWER:

Since all four suits have the same number of cards, that means all four suits are equally likely to be the suit of the card on top. This means the probability the top card is a heart is 1/4.

$\frac{1}{4}$

12. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the top card is a diamond?

ANSWER:

Since all four suits have the same number of cards, that means all four suits are equally likely to be the suit of the card on top. This means the probability the top card is a diamond is 1/4.

$\frac{1}{4}$

13. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the second card is a heart?

ANSWER:

Since all four suits have the same number of cards, that means all four suits are equally likely to be the suit of the card underneath the top card. This means the probability the second card is a heart is 1/4.

$\frac{1}{4}$

14. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the third card is a heart?

ANSWER:

Since all four suits have the same number of cards, that means all four suits are equally likely to be the suit of the third card from the top. This means the probability the third card is a heart is 1/4.

$\frac{1}{4}$

15. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the fourth card is a heart?

ANSWER:

Since all four suits have the same number of cards, that means all four suits are equally likely to be the suit of the fourth card from the top. This means the probability the fourth card is a heart is 1/4.

$\frac{1}{4}$

16. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the last card is a heart?

ANSWER:

Since all four suits have the same number of cards, that means all four suits are equally likely to be the suit of the fourteenth card. This means the probability the fourteenth card is a heart is 1/4.

$\frac{1}{4}$

17. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the second and fifth cards ares hearts?

ANSWER:

Use the multiplication rule:

$$E(XI_N|K) = E(X|A\&K)P(A|K)$$

as it applies to the special case of probability:

$$P(A\&B|C) = P(A|B\&C)P(B|C).$$

Let A be the event or statement that the second card is a heart and let B be the event or statement that the fifth card is a heart. Take C to be the statement that all cards are equally likely to be anywhere in the deck unless we are given specific information otherwise. Then

$$P(A|C) = \frac{1}{4}$$

whereas

$$P(A|B\&C) = \frac{12}{51} = P(B|A\&C).$$

We therefore conclude by the multiplication rule that

$$P(A\&B|C) = P(A|B\&C)P(B|C) = (\frac{12}{51})(\frac{1}{4}) = \frac{3}{51}.$$

$\frac{3}{51}$

18. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the fourth is a heart, given that the fifth is a spade?

ANSWER:

$\frac{13}{51}$

19. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the fourth is a heart given that the third is a spade?

ANSWER:

$\frac{13}{51}$

20. If fourteen cards are dealt from the top of a well shuffled deck of cards one after another without replacement, what is the chance that the fourth is a heart given that the third is a heart, the fifth is a heart, the sixth is a club, and the seventh is a club?

ANSWER:

$\frac{11}{48}$

21. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assisitant who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the first block drawn is red?

ANSWER:
$\frac{5}{20} = \frac{1}{4}$

22. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assisitant who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the last block drawn is red?

ANSWER:
$\frac{1}{4}$

23. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assisitant who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the fourth block drawn is red?

ANSWER:
$\frac{1}{4}$

24. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assisitant who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the second block drawn is red given that the first is red?

ANSWER:
$\frac{4}{19}$

25. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assisitant who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the first block drawn is red given that the second is red?

ANSWER:
$\frac{4}{19}$

26. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assisitant who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the second block drawn is red given that the first and third are both red?

ANSWER:
$\frac{3}{18}$

27. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assisitant who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the fifth block drawn is red given that the first and third are both red and the seventh is green?

ANSWER:
$\frac{3}{17}$

28. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assistaint who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the fifth block drawn is red given that the first and third are both red the seventh is green and the eighth is blue?

ANSWER:
$\frac{3}{16}$

29. A box contains 5 red blocks, 8 blue blocks, and 7 green blocks. A lab assisitant who is totally color blind is told to remove the blocks from the box one by one without replacement. What is the probability that the fifth and ninth blocks drawn are both red given that the first and third are both red the seventh is green and the eighth is blue?

ANSWER:
Let C be the statement of the problem setup and that the first and third are both red, the seventh is green and the eighth is blue. Let A be the statement that the ninth is red and let B be the statement that the fifth block drawn is red. From the multiplication rule of probability we know:

$$P(A\&B|C) = P(A|B\&C)P(B|C) = (\frac{2}{15})(\frac{3}{16}) = \frac{1}{40}.$$

$\frac{1}{40}$

30. Suppose that A, B and C are statements and that $P(A|B\&C) = .4$, that $P(A|C) = .7$, and that $P(B|C) = .6$. Calculate $P(B|A\&C)$.

ANSWER:
We know:

$$P(A|B\&C)P(B|C) = P(A\&B|C) = P(B\&A|C) = P(B|A\&C)P(A|C),$$

and therefore putting in the numbers,

$$(.4)(.6) = (.7)P(B|A\&C),$$

so

$$P(B|A\&C) = \frac{.24}{.7} = \frac{24}{70} = \frac{12}{35}$$

$\frac{12}{35}$ or .3428571429

31. Suppose that X and Y are unknowns with $\sigma_X = 3$ and $\sigma_Y = 10$ and with $\rho = .8$. Calculate $Cov(X, Y)$.

ANSWER:
We use the formula $Cov(X, Y) = \rho\sigma_X\sigma_Y$, so $Cov(X, Y) = (.8)(3)(10) = 24$.
24

32. Suppose that X and Y are unknowns with $\sigma_X = 3$ and $\sigma_Y = 10$ and with $\rho = .8$. Calculate the regression slope β_1 of Y on X.

ANSWER: The regression slope is given by the formula

$$\beta_1 = \frac{\rho\sigma_Y}{\sigma_X} = \frac{(.8)(10)}{3} = \frac{8}{3}.$$

$\beta_1 = \frac{8}{3}$

33. Suppose that X and Y are unknowns with $\sigma_X = 3$ and $\sigma_Y = 10$ and with $\rho = .8$. Suppose that $E(X) = 80$ and $E(Y) = 50$. Calculate β_0, the regression intercept.

ANSWER:
The regression slope is given by $\beta_0 = \mu_Y - \beta_1\mu_X = 50 - \frac{8}{3}80 = \frac{150-640}{3} = \frac{-490}{3}$.
$\beta_0 = \frac{-490}{3}$

34. Suppose that X and Y are unknowns with $\sigma_X = 3$ and $\sigma_Y = 10$ and with $\rho = .8$. Suppose that $E(X) = 80$ and $E(Y) = 50$. Calculate $E(Y|X = 60)$.

ANSWER:
We can use the regression equation

$$E(Y|X = 60) = \beta_0 + \beta_1(60) = \frac{-490}{3} + \frac{8}{3}(60) = \frac{480 - 490}{3} = \frac{-10}{3}.$$

Alternately, we can observe that as the mean of X is 80, the value of 60 is 20 units below the mean or 20/3 standard deviation units below the mean, so we would guess Y to be (.8)(20/3) standard deviation units below its mean of 50 and as Y has standard deviation 10 this means (8)(20/3) or 160/3 below 50 which is again -10/3.

$E(Y|X = 60) = \frac{-10}{3}$

Suppose that the Big Forest contains 3 bears(Papa Bear, Mama Bear, and Baby Bear) and Goldilocks. Papa Bear weighs 2000 pounds and eats 20 pounds of porridge a day, Mama Bear weighs 1750 pounds and eats 15 pounds of porridge a day, Baby bear weighs 200 pounds and eats 8 pounds of porridge a day, and Goldilocks weighs 50 pounds and eats 1 pound of porridge a day. Suppose we think of these three bears and Goldilocks as a population of size $N = 4$.

ANSWER NOTE: You can put the weights in a list in the calculator and the porridge consumption in another list so as to pair each weight with the corresponding porridge consumption. The the 2-Var Stats will give the means and standard deviations and the mean of the products. The lin reg(a+bx) then gives the regression equation and the correlation coefficient. We will work out each answer below without the calculator so you can see what happens in detail. It is best to choose symbols, so below we use X for weight and Y for porridge consumption.

35. What is the population mean weight?

ANSWER:
The mean weight is obviously 4000/4 which is 1000 pounds.

$\mu_X = 1000$

36. What is the population mean porridge consumption?

ANSWER:
The mean porridge consumption is obviously 44/4 which is 11 pounds

$\mu_Y = 11$

37. What is the standard deviation in weight?

ANSWER:
We can work in hundreds of pounds for convenience, so the population of weights in hundreds is 20,17.5,2, .5 with average 10 so the deviations are 10,7.5,-8, -9.5 and the squared deviations are therefore 100, 56.25, 64, 90.25 which average 77.625. The square root of this, 8.810505093, is the standard deviation, but remember we are calculating in hundreds, so the standard deviation in weight is 881.0505093, or to three significant digits, 881.

$\sigma_X = 881$

38. What is the standard deviation in porridge consumption?

ANSWER:
The mean for the porrigdge consumption is 11, so the deviations are 9, 4, -3, -10, and the squared deviations are therefore 81, 16, 9, 100 which average 206/4=51.5. The square root of this is the standard deviation, which is therefore 7.176350047.

$\sigma_Y = 7.18$

39. What is the correlation coefficient of the relationship between weight and porridge consumption?

ANSWER: The products of the deviations are (10)(9), (7.5)(4), (-8)(-3), (-9.5)(-10), where we again are calculating weights in units of hundreds. These products are 90, 30, 24, 95 which average 239/4 or 59.75. But since we used weight in units of hundreds here, this means the actuall covariance is one hundred times as much or 5975. Dividing this by the product of standard deviations gives the correlation coefficient $\rho = 5975/([7.176350047][881.0505093]) = .9450036557$ or $\rho = .945$.

$\rho = .945$

40. What is the covariance between weight and porridge consumption in this population?

ANSWER:
The products of the deviations are (10)(9), (7.5)(4), (-8)(-3), (-9.5)(-10), where we again are calculating weights in units of hundreds. These products are 90, 30, 24, 95 which average 239/4 or 59.75. Again this must now be multiplied by one hundred, so the correct answer is 5975.

Covariance: 5975

41. If Baby Bear gains 100 pounds, based on the regression of Y on X using the information given above, what would be a good guess as to Baby Bear's new porridge consumption?

ANSWER:
The regression slope is

$$Regression\ Slope = (.945)\frac{7.176350047}{881.0505093} = .0076972625,$$

which means that each additional one hundred pounds should cause us to guess an additional 0.76972625 pounds of porridge consumption, or 0.770, to three significant digits.
0.770

Suppose that X is an unknown which is a number in the set $S = \{2, 5, 7\}$. Suppose that $P(X = 2) = .3$, $P(X = 5) = .5$, and $P(X = 7) = .2$. Suppose that Y is an unknown which is a number in the set $\{3, 4\}$ with $P(Y = 3) = .6$ and $P(Y = 4) = .4$. Suppose that X and Y are independent of each other.

42. What is the moment generating function of X?

ANSWER: The moment generating function for X is

$$m_X(t) = (.3)e^{2t} + (.5)e^{5t} + (.2)e^{7t}.$$

43. What is the moment generating function of Y?

ANSWER:
The moment generating function of Y is

$$m_Y(t) = (.6)e^{3t} + (.4)e^{4t}.$$

44. What is the moment generating $X + Y$?

ANSWER:
Since X and Y are given to be independent, we know that the moment generating function of $X + Y$ is simply the product of that for X with that for Y, that is

$$m_{(X+Y)}(t) = m_X(t)m_Y(t).$$

We can easily multiply the two previous answers together to find now

$$m_{(X+Y)}(t) = (.18)e^{5t} + (.3)e^{8t} + (.12)e^{10t} + (.12)e^{6t} + (.2)e^{9t} + (.08)e^{11t}.$$

45. What are the possible values for $X+Y$ and their probabilities?

ANSWER:

Notice from this we now easily see that $X+Y$ is a number in the set $\{5,6,8,9,10,11\}$ and the probabilities of each value are easily read off of the moment generating function as these are the coefficients for each term. For instance, $P(X+Y=8)=.3$. It is very important to remember that X and Y **are here assumed to be independent.** You cannot do this when X and Y are correlated or dependent in any way. But as here X and Y are assumed independent, and as we have

$$m_{(X+Y)}(t)=(.18)e^{5t}+(.3)e^{8t}+(.12)e^{10t}+(.12)e^{6t}+(.2)e^{9t}+(.08)e^{11t},$$

we see the possible values are 5,6,8,9,10,11 with respective probabilities .18,.12,.3,.2,.12,.08. This could be calculated directly without too much difficulty, but here with moment generating functions the algebra does all the book keeping for us.

Suppose that X is a random variable and $X_1,X_2,X_3,...,X_n$ is a sequence of n observations of X thus forming a sample of size n. Let T be the sample total and let $\bar{X}$ be the sample mean. As usual, $E(X)=\mu_X$ is the expected value of X, and $SD(X)=\sigma_X$ is the standard deviation of X. Always, SD =Standard Deviation.

46. What is the expected value of T if $\mu_X=40$ and $n=5$?

ANSWER:

In general,

$$E(T)=n\mu_X,$$

so here $E(T)=5\cdot 40=200$.

47. What is the expected value of $\bar{X}$ if $\mu_X=40$?

ANSWER:

In general,

$$E(\bar{X})=E(X)=\mu_X,$$

so we have simply $E(\bar{X})=40$, in this case.

48. If $\sigma_X=10$ and $n=25$, and if the n observations are independent unknowns, then what is $SD(T)$?

ANSWER:

In general, for an independent random sample or, for short, an IRS,

$$SD(T)=\sqrt{n}\cdot\sigma_X,$$

so here we have

$$SD(T)=\sqrt{25}\cdot 10=5\cdot 10=50.$$

49. If $\sigma_X=10$ and $n=25$, and if the n observations are independent unknowns, then what is $SD(T)$?

ANSWER:

In general, for an IRS,

$$SD(\bar{X})=\frac{\sigma_X}{\sqrt{n}},$$

so here

$$SD(\bar{X})=\frac{10}{\sqrt{25}}=\frac{10}{5}=2.$$

50. Suppose $\sigma_X = 10$, and that the sample observations form a Simple Random Sample (for short, SRS) from a population of size $N = 101$ and that $n = 25$. Now what is $SD(T)$?

ANSWER:
In general, for a SRS, we have

$$SD(T) = \sqrt{\frac{N-n}{N-1}} \cdot SD(T)_{IRS} = \sqrt{\frac{N-n}{N-1}} \cdot \sqrt{n} \cdot \sigma_X,$$

so here we have

$$SD(T) = \sqrt{\frac{101-25}{101-1}} \cdot \sqrt{25} \cdot 10 = \sqrt{.76} \cdot 5 \cdot 10 = 50 \cdot \sqrt{.76} = 43.58898944.$$

51. Suppose $\sigma_X = 10$, and that the sample observations form a Simple Random Sample (for short, SRS) from a population of size $N = 101$ and that $n = 25$. Now what is $SD(\bar{X})$?

ANSWER:
In general for SRS, we have

$$SD(\bar{X}) = \sqrt{\frac{N-n}{N-1}} \cdot SD(\bar{X})_{IRS} = \sqrt{\frac{N-n}{N-1}} \cdot \frac{\sigma_X}{\sqrt{n}},$$

so here we have

$$SD(\bar{X}) = \sqrt{.76} \cdot \frac{10}{\sqrt{25}} = 2 \cdot \sqrt{.76} = 1.743559577.$$

52. Suppose we form U, the sum of the squared deviations from the sample mean, so

$$U = \sum_{k=1}^{n} (X_k - \bar{X})^2,$$

and assume IRS. If $\sigma_X = 10$, and $n = 25$, what is $E(U)$, the expected value of U?

ANSWER:
In general, with IRS,

$$E(U) = (n-1) \cdot \sigma_X^2,$$

with IRS, so here

$$E(U) = 24 \cdot (10)^2 = 2400.$$

53. Suppose we form U, the sum of the squared deviations from the sample mean, so

$$U = \sum_{k=1}^{n} (X_k - \bar{X})^2,$$

and assume IRS. If $\sigma_X = 10$, what is $E([1/(n-1)]U)$, the expected value of $[1/(n-1)]U$?

ANSWER:
Since $E(U) = (n-1) \cdot \sigma_X^2$, it follows that

$$E(\frac{U}{n-1}) = \sigma_X^2,$$

so here

$$E(\frac{U}{n-1}) = (10)^2 = 100.$$

Suppose that X and Y are both random variables with $E(X) = \mu_X,\ E(Y) = \mu_Y,\ SD(X) = \sigma_X,\ SD(Y)\ \sigma_Y$, and suppose that ρ is the correlation coefficient of X with Y. Suppose that a sample of size n is to be made with the pairs of observations being $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), ..., (X_n, Y_n)$.

54. Suppose that $\mu_X = 40,\ \mu_Y = 50,\ \sigma_X = 10,\ \sigma_Y = 8$, and suppose that $\rho = .6$. What is $Cov(X, Y)$, the covariance of X with Y?

ANSWER:

In general, recall that

$$Cov(X, Y) = \rho \cdot \sigma_X \cdot \sigma_Y.$$

So here,

$$Cov(X, Y) = (.6) \cdot (10) \cdot (8) = 48.$$

55. Suppose that $\mu_X = 40,\ \mu_Y = 50,\ \sigma_X = 10,\ \sigma_Y = 8$, and suppose that $\rho = .6$. Suppose that

$$W = \sum_{k=1}^{n} (X_k - \bar{X})(Y_k - \bar{Y}).$$

If $n = 25$, with IRS what is $E(W)$?

ANSWER:

In general, with IRS,

$$E(W) = (n - 1) \cdot Cov(X, Y),$$

so here we have

$$E(W) = 24 \cdot 48 = 8 \cdot 144 = 1152.$$

56. Suppose that $\mu_X = 40,\ \mu_Y = 50,\ \sigma_X = 10,\ \sigma_Y = 8$, and suppose that $\rho = .6$. Suppose that

$$W = \sum_{k=1}^{n} (X_k - \bar{X})(Y_k - \bar{Y}).$$

With IRS what is $E((1/(n - 1))W)$?

ANSWER:

Since in general with IRS we have $E(W) = (n - 1) \cdot Cov(X, Y)$, it follows that in general,

$$E(\frac{W}{n - 1}) = Cov(X, Y),$$

and therefore here we have

$$E(\frac{W}{n - 1}) = Cov(X, Y) = 48.$$

57. Let β_0 and β_1 be the regression coefficients for the linear regression of Y on X. Suppose that $\mu_X = 40,\ \mu_Y = 50,\ \sigma_X = 10,\ \sigma_Y = 8$, and suppose that $\rho = .6$. Give the values of β_0 and β_1.

ANSWER:

In general recall we have

$$\beta_1 = \rho \cdot \frac{\sigma_Y}{\sigma_X},$$

and

$$\beta_0 = \mu_Y - \beta_1 \cdot \mu_X,$$

so here,

$$\beta_1 = (.6) \cdot \frac{8}{10} = .48,$$

and

$$\beta_0 = 50 - (.48) \cdot 40 = 50 - 19.2 = 30.8.$$

Suppose that h is the non-negative function given by

$$h(x) = 1 - |x| = 1 - \sqrt{x^2},\ -1 \leq x \leq 1,$$

and $h(x) = 0$, **otherwise.**

58. Can h be a probability density function for some continuous random variable?

ANSWER:

To find out, we integrate h from negative infinity to infinity which is the same as from -1 to 1 and find the normalization constant, A. As $h(x) = 1 - (-x) = 1 + x$ for $-1 \leq x \leq 0$ and $h(x) = 1 - x$, for $0 \leq x \leq 1$, we have

$$A = \int_{-\infty}^{\infty} h(x)dx = \int_{-1}^{0} (1+x)dx + \int_{0}^{1} (1-x)dx = \frac{1}{2} + \frac{1}{2} = 1.$$

Notice the graph of h is a triangle of height 1 and base length 2, so the total area underneath the graph is obviously 1. This means that h is the probability density function for a distribution for a continuous random variable. In fact, if X is the horizontal coordinate of a point in the triangle which I know and you do not, then $f_X = h$. The answer here is therefore YES.

59. If X is a random variable with h as its pdf, then what is $E(X)$?

ANSWER: Obviously the region under the graph of h, the triangle, balances on $x = 0$, so from this we see $E(X) = 0$. Alternately, we can calculate. In general, if X has pdf equal to f_X, then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

so

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx,$$

and this means here that

$$E(X) = \int_{-\infty}^{\infty} x \cdot h(x)dx = \int_{-1}^{0} x(1+x)dx + \int_{0}^{1} x(1-x)dx$$
$$= \int_{-1}^{1} x dx + \int_{-1}^{0} x^2 dx - \int_{0}^{1} x^2 dx = 0.$$

60. If X is a random variable with h as its pdf, then what is $E(X^2)$?

ANSWER: We can here notice that $h(x)$ is an even function, since $h(x) = h(-x)$, and likewise, so is x^2, so therefore, $x^2 \cdot h(x)$ is also an even function. If e is any even function, then

$$\int_{-L}^{L} e(x)dx = 2\int_{0}^{L} e(x)dx,$$

so

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot h(x)dx = \int_{-1}^{1} x^2 \cdot h(x)dx = 2\int_{0}^{1} x^2(1-x)dx$$
$$= 2\int_{0}^{1} x^2 dx - 2\int_{0}^{1} x^3 dx = \frac{2}{3} - \frac{2}{4} = \frac{1}{6}.$$

61. If X is a random variable with h as its pdf, then what is $Var(X) = \sigma_X^2$, the varicance?

ANSWER: In general, $Var(X) = E(X^2) - (E(X))^2$. Since $E(X) = 0$, this means here

$$Var(X) = \frac{1}{6}.$$

Obviously, we now can find the standard deviation by taking the square root of the variance.

62. If X is the unknown with pdf given by $f_X = 2x \cdot I_{[0,1]}$, then what is the power series expansion for m_X, the moment generating function of X.

ANSWER:

We have $m_X = \mathcal{M}(f_X)$, where $\mathcal{M}$ is the integral transform

$$\mathcal{M}(h) = \int_{-\infty}^{\infty} e^{tx} h(x) dx,$$

for any function h which vanishes sufficiently fast at $\pm\infty$. Moreover, we have

$$\mathcal{M}(x \cdot h) = \frac{d}{dt}\mathcal{M}(h).$$

As we know

$$\mathcal{M}(I_{[0,1]}) = \frac{e^t - e^0}{t} = \frac{1}{t}\sum_{n=1}^{\infty}\frac{t^n}{n!} = \sum_{n=1}^{\infty}\frac{t^{n-1}}{n!} = 1 + \frac{t}{2} + \frac{t^2}{6} + \frac{t^3}{24} + ...,$$

therefore

$$\mathcal{M}(x \cdot I_{[0,1]}) = \frac{d}{dt}\mathcal{M}(I_{[0,1]}) = \sum_{n=2}^{\infty}\frac{(n-1)t^{n-2}}{n!} = \frac{1}{2} + \frac{1}{3}t + \frac{1}{8}t^2 + ...$$

and finally

$$m_X(t) = \mathcal{M}(2x \cdot I_{[0,1]}) = 2\mathcal{M}(x \cdot I_{[0,1]}) = 2\sum_{n=2}^{\infty}\frac{(n-1)t^{n-2}}{n!} = 1 + \frac{2}{3}t + \frac{1}{4}t^2 + ...$$

and from this we see instantly that

$$E(X) = \frac{2}{3}$$

and

$$E(X^2) = (2!)(1/4) = \frac{1}{2}.$$

Therefore

$$Var(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18},$$

and

$$\sigma_X = \frac{1}{3\sqrt{2}},$$

or σ_X is about .2357022604, slightly less than 1/4.

63. If X is the unknown with pdf given by $f_X = 3x^2 \cdot I_{[0,1]}$, then what is the power series expansion for m_X, the moment generating function of X.

ANSWER:

We have

$$\mathcal{M}(f_X) = 3\mathcal{M}(x^2 I_{[0,1]}) = 3\frac{d^2}{dt^2}\mathcal{M}(I_{[0,1]}).$$

Since (see answer in previous problem)

$$\frac{d}{dt}\mathcal{M}(I[0,1]) = \sum_{n=2}^{\infty} \frac{(n-1)t^{n-2}}{n!} = \frac{1}{2} + \frac{1}{3}t + \frac{1}{8}t^2 + \frac{1}{30}t^3 ...,$$

it follows that

$$\frac{d^2}{dt^2}\mathcal{M}(I[0,1]) = \sum_{n=3}^{\infty} \frac{(n-1)(n-2)t^{n-3}}{n!} = \frac{1}{3} + \frac{1}{4}t + \frac{1}{10}t^2 + \frac{1}{36}t^3 + \frac{1}{168}t^4 + ...,$$

and therefore

$$\begin{aligned} m_X(t) = 3\frac{d^2}{dt^2}\mathcal{M}(I[0,1]) &= 3\sum_{n=3}^{\infty} \frac{(n-1)(n-2)t^{n-3}}{n!} \\ &= 3\left(\frac{1}{3} + \frac{1}{4}t + \frac{1}{10}t^2 + \frac{1}{36}t^3 + \frac{1}{168}t^4 + ...\right), \\ &= \left(1 + \frac{3}{4}t + \frac{3}{10}t^2 + \frac{1}{12}t^3 + \frac{1}{56}t^4 + ...\right), \end{aligned}$$

from which we now see immediately that

$$E(X) = \frac{3}{4},$$

$$E(X^2) = (2!)\frac{3}{10} = .6,$$

$$E(X^3) = (3!)\frac{1}{12} = \frac{1}{2},$$

and

$$E(X^4) = (4!)\frac{1}{56} = \frac{24}{56} = \frac{3}{7}.$$

This means that the variance of X is

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{6}{10} - \frac{9}{16} = \frac{3}{80},$$

and this means in turn

$$\sigma_X = \sqrt{\frac{6}{160}} = \frac{\sqrt{.6}}{4},$$

or about .1936491673, which is slightly less than 1/5. Notice that in each of these examples (this problem and in the previous problem) we have $\mu_X + \sigma_X < 1$.

64. If X is the unknown with pdf given by

$$f_X = A[6 + 2(x-1)]I_{[1,2]},$$

then what is A and what is the moment generating function of X?

ANSWER:

The graph of f_X is a trapezoid with base 1 and vertical sides of heights 6 and 8, so its area is 7. Thus $A = 1/7$. We have

$$m_X(t) = \mathcal{M}(f_X) = A\mathcal{M}([4+2x]I_{[1,2]}) = 4A\mathcal{M}(I_{[1,2]}) + 2A\frac{d}{dt}\mathcal{M}(I_{[1,2]}).$$

Now,

$$\mathcal{M}(I_{[1,2]}) = \frac{e^{2t} - e^t}{t} = \frac{1}{t}\sum_{n=0}^{\infty}\left(\frac{(2t)^n}{n!} - \frac{t^n}{n!}\right) = \frac{1}{t}\sum_{n=0}^{\infty}\left(\frac{2^n - 1}{n!}\right)t^n$$
$$= \sum_{n=0}^{\infty}\left(\frac{2^n-1}{n!}\right)t^{n-1} = \sum_{n=1}^{\infty}\left(\frac{2^n-1}{n}\right)\frac{t^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty}\left(\frac{2^{n+1}-1}{n+1}\right)\frac{t^n}{n!}$$
$$= 1 + \frac{3}{2}t + \frac{7}{3}\cdot\frac{t^2}{2!} + \frac{15}{4}\cdot\frac{t^3}{3!} + ...$$

Therefore

$$\frac{d}{dt}\mathcal{M}(I_{[1,2]}) = \sum_{n=1}^{\infty}\left(\frac{2^{n+1}-1}{n+1}\right)\frac{t^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty}\left(\frac{2^{n+2}-1}{n+2}\right)\frac{t^n}{n!}$$
$$= \frac{3}{2} + \frac{7}{3}t + \frac{15}{4}\cdot\frac{t^2}{2!} + \frac{31}{5}\cdot\frac{t^3}{3!} + ...$$

Putting this together now gives

$$m_X(t) = A\left[4\sum_{n=0}^{\infty}\left(\frac{2^{n+1}-1}{n+1}\right)\frac{t^n}{n!} + 2\sum_{n=0}^{\infty}\left(\frac{2^{n+2}-1}{n+2}\right)\frac{t^n}{n!}\right]$$
$$A\sum_{n=0}^{\infty}\left[4\left(\frac{2^{n+1}-1}{n+1}\right) + 2\left(\frac{2^{n+2}-1}{n+2}\right)\right]\frac{t^n}{n!}.$$

For each n we must then have the n^{th} moment

$$E(X^n) = \frac{4}{7}\left(\frac{2^{n+1}-1}{n+1}\right) + \frac{2}{7}\left(\frac{2^{n+2}-1}{n+2}\right),$$

or using the first few terms of the series,

$$E(X) = \frac{4}{7}\cdot\frac{3}{2} + \frac{2}{7}\cdot\frac{7}{3} = \frac{6}{7} + \frac{2}{3} = \frac{32}{21} = 1.523809524,$$
$$E(X^2) = \frac{4}{7}\cdot\frac{7}{3} + \frac{2}{7}\cdot\frac{15}{4} = \frac{101}{42} = 2.404761905.$$
$$E(X^3) = \frac{4}{7}\cdot\frac{15}{4} + \frac{2}{7}\cdot\frac{31}{5} = \frac{137}{35} = 3.914285714.$$

Then

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{101}{42} - \frac{32^2}{21^2} = \frac{101}{42} - \frac{1024}{441} = \frac{73}{882} = .0827664399,$$

so

$$\sigma_X = .2876915708.$$

65. Suppose that the fish in my pond have mean length 20 inches with a standard deviation of 4 inches, and that we know nothing else about the distribution of lengths for this population of fish. What is the probability that a fish caught from this pond will be between 19 and 23 inches in length?

ANSWER: Since we only know the mean and standard deviation and nothing else about the distribution, the distribution we must use is the normal distribution. Use of any other distribution would be assuming more knowledge than we have. Let X denote the length of the fish we will catch. We know that $E(X) = \mu_X = 20$, so

$$P(19 \leq \bar{X} \leq 23) = \mathit{normalcdf}(19, 23, 20, 4) = .3720789949, \text{ or } .372$$

to three significant digits.

FINAL ANSWER: .372

66 AVERAGE. Suppose that the fish in my pond have mean length 20 inches with a standard deviation of 4 inches, and that we know nothing else about the distribution of lengths for this population of fish. What is the probability that an independent random sample of 5 fish caught from this pond will average between 19 and 23 inches in length?

ANSWER: Since we only know the mean and standard deviation and nothing else about the distribution, the distribution we must use is the normal distribution. Use of any other distribution would be assuming more knowledge than we have. We know that the sample mean $\bar{X}$ is expected to be $\mu_X = 20$, and for independent random sampling, the standard deviation of $\bar{X}_n$ is

$$SD(\bar{X}_n) = \frac{\sigma_X}{\sqrt{n}},$$

so here $\bar{X}$ is normal with mean 20 and standard deviation $4/\sqrt{5}$. Thus, here

$$P(19 \leq \bar{X} \leq 23) = \mathit{normalcdf}(19, 23, 20, 4/\sqrt{(5)}) = .6651587438,$$

or about .665, to three significant digits.

FINAL ANSWER: .665

66 TOTAL. Suppose that the fish in my pond have mean length 20 inches with a standard deviation of 4 inches, and that we know nothing else about the distribution of lengths for this population of fish. What is the probability that in an independent random sample, the total length of 5 fish caught from this pond will be between 95 and 110 inches in length?

ANSWER: Since we only know the mean and standard deviation and nothing else about the distribution, the distribution we must use is the normal distribution. Use of any other distribution would be assuming more knowledge than we have. For a sample of size n we have the expected value of T_n, the sample total, is

$$E(T_n) = n \cdot \mu_X,$$

so here we have

$$E(T_5) = 5 \cdot (20) = 100,$$

and for independent random sampling, we have the standard deviation $SD(T_n)$ of T_n is

$$SD(T_n) = \sqrt{n} \cdot \sigma_X,$$

which here is

$$SD(T_5) = \sqrt{5} \cdot (4).$$

Moreover, as X here is normal, the same is true for T_n and $\bar{X}$, so

$$P(95 \leq T_n \leq 110) = normalcdf(95, 110, 100, 4 \cdot \sqrt{(5)}\ ) = .5801486913,$$

or about .580, to three significant digits.

FINAL ANSWER: .580

67. Suppose that the fish in my pond have mean length 20 inches with a standard deviation of 4 inches, and that we know nothing else about the distribution of lengths for this population of fish. What is the shortest a fish can possibly be and still be in the top 20 percent of fish as regards length?

ANSWER: Since we only know the mean and standard deviation and nothing else about the distribution, the distribution we must use is the normal distribution. Use of any other distribution would be assuming more knowledge than we have. Now, we are not asking for a probability here, we are asking the reverse, namely you must find the value of x so that $P(X \leq x) = .8$ and $P(X \geq x) = .2$. It is customary to denote this x as $x_{.8}$, and to call this the 80^{th} percentile score. We are in the situation that we have the probability, and we want the limit. This in a sense is going backwards and that is the job of "invNorm" in your distribution menu. Thus here,

$$x_{.8} = invNorm(.8, 20, 4) = 23.36648493,$$

or about 23.4 inches long, to three significant digits. Notice that there may in fact be only one fish in the pond and these answers are the same, that is we are dealing with the state of our information. We can say that if a fish is caught there will be an 80 percent chance that it is less than 23.4 inches long.

FINAL ANSWER: 23.4

68. Suppose that I do not know the mean weight of the fish in my pond and I want to estimate it using the sample mean of an independent random sample. Suppose that I have a sample of 36 fish with sample mean $\bar{x} = 8.3$ pounds. Suppose that I know that the standard deviation in fish weight for the fish in my pond is 2 pounds. What is the margin of error in the 95 percent confidence interval for the true mean weight of the fish in my pond?

ANSWER: Since the sample size is over 30 and the sample is an IRS, it follows that $\bar{X}$ is normally distributed with standard deviation $2/\sqrt{36} = 2/6 = 1/3$, and this means the margin of error in the 95 percent confidence interval is M where

$$M = z_{(C=.95)} \cdot \frac{\sigma_X}{\sqrt{n}} = z_{(C=.95)} \cdot \frac{1}{3}.$$

To find z_C, here we can use the invNorm in the distribution menu,

$$z_{(C=.95)} = invNorm(.975, 0, 1) = -invNorm(.025, 0, 1) = 1.959963986.$$

This means

$$M = .6533213287, \text{ or } .653,$$

to three significant digits.

Alternately, you can use the z-interval in the calculator. Choose the stats option enter the information as directed, but use $\bar{x} = 0$, choose the C-level as .95, and in the readout, the positive number is M, which in this case is $M = .65332$ which gives again $M = .653$ to three significant digits.

FINAL ANSWER: .653

69. Suppose that I do not know the mean weight of the fish in my pond and I want to estimate it using the sample mean of an independent random sample. Suppose that I have a sample of 36 fish with sample mean $\bar{x} = 8.3$ pounds. Suppose that I do not know the standard deviation in the weight of fish in my pond, but the sample standard deviation of my sample is 2 pounds. What is the margin of error in the 95 percent confidence interval for the true mean weight of the fish in my pond?

ANSWER: This problem is the same as the preceding except that the standard deviation of the population is unknown and we use the sample standard deviation in its place. Whenever you do that, you must use the Student-t-distribution for $n - 1 = 35$ degrees of freedom in place of the standard normal distribution. Thus, the margin of error is now

$$M = t_{(C=.95)} \cdot \frac{s}{\sqrt{n}} = t_{(C=.95)} \cdot \frac{1}{3}.$$

To find t_C we can use the invT in the calculator,

$$t_{(C=.95)} = invT(.975, 35) = 2.030107868,$$

so

$$M = .6767026228, \text{ or } .677,$$

to three significant digits. Notice that the margin of error is increased due to the lack of knowledge of the population standard deviation, even though the sample standard deviation accidentally hits the population standard deviation on the nose. But we do not know that here. Alternately, you can use the t-interval in the calculator. Choose the stats option and enter $\bar{x} = 0$ and $s_x = 2$, and $n = 36$, and then the readout is $(-.6767,\ .6767)$ which means $M = .6767$ or again .677 to three significant digits.

FINAL ANSWER: .677

70. Suppose that I do not know the mean weight of the fish in my pond and I want to estimate it using the sample mean of an independent random sample. Suppose that I have a sample of 36 fish with sample mean $\bar{x} = 8.3$ pounds. Suppose that I know that the standard deviation in fish weight for the fish in my pond is 2 pounds. What is the significance of this data as evidence that the true mean weight of the fish in my pond actually exceeds 7.5 pounds?

ANSWER: As soon as we are asked for significance of data, we know we are dealing with a HYPOTHESIS TEST. Here we know the population standard deviations, so the significance is the value of p in the readout for the z-test in the test menu. We must choose the correct alternate hypothesis. Always remember, the alternate hypothesis is what you are trying to prove with the data. In this case, we are asking for the significance ($= P-$Value) as evidence that the true mean weight exceeds 7.5, so we enter $\mu_0 = 7.5$, and choose $\mu > \mu_0$ as our alternate hypothesis. The result is

$$P - Value = P(\bar{X}_n \geq \bar{x}_{data}) = P(\bar{X}_n \geq 8.3) = P(Z \geq z_{data}),$$

where z_{data} is the standard score for our sample mean under the null hypothesis H_0 which tells us to assume the true mean is $\mu_0 = 7.5$, so

$$H_0 : \mu_X \leq 7.5,$$

$$H_{alt} : \mu > 7.5,$$

$$z_{data} = \frac{8.3 - 7.5}{(2/\sqrt{36})} = \frac{6(.8)}{2} = 2.4,$$

$$P - Value = P(Z \geq 2.4) = .0081975289, \text{ or } .00820,$$

to three significant digits. Using the calculator test menu you begin by choosing the z-test, since you know the population standard deviation here. You then enter $\mu_0 = 7.5$, and $\sigma_x = 2$ as well as the data information asked for. Be sure to choose the correct alternate hypothesis in terms of μ_0. Thus, with $\mu_0 = 7.5$, the alternate hypothesis $\mu > 7.5$ becomes $\mu > \mu_0$, so we put the cursor on $> \mu_0$ and hit the enter button on the calculator. When we choose calculate and hit enter we see in the readout

$$Z - Test$$
$$\mu > 7.5$$
$$z = 2.4$$
$$p = .0081975289$$
$$\bar{x} = 8.3$$
$$n = 36$$

which again gives the same result as we calculated above for the significance of the data, again to three significant digits, the P-Value or significance of the data is .00820.

FINAL ANSWER: .00820

71. Suppose that I do not know the mean weight of the fish in my pond and I want to estimate it using the sample mean of an independent random sample. Suppose that I have a sample of 36 fish with sample mean $\bar{x} = 8.3$ pounds. Suppose that I do not know the standard deviation in the weight of fish in my pond, but the sample standard deviation of my sample is 2 pounds. What is the significance of this data as evidence that the true mean weight of the fish in my pond actually exceeds 7.5 pounds?

ANSWER: Notice that this is again a hypothesis test as we are asked for the significance ($= P-$Value) of the data. Just as for confidence intervals, when dealing with something you measure we must use z or t to test hypotheses concerning the mean, and the distinction between z and t is simply whether or not we know the true population standard deviation. Here we do not, and consequently we will use the $t-$ distribution for $n-1 = 35$ degrees of freedom. Thus, we have

$$t_{data} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{8.3 - 7.5}{2/6} = 3 \cdot (.8) = 2.4$$

and

$$P{-}Value = P(t_{35} \geq 2.4) = .5 - P(0 \leq t_{35} \leq 2.4) = .5 - tcdf(0, 2.4, 35) = .0109249267, \text{ or } .0109,$$

to three significant digits. Using the t-test in the test menu, enter $\mu_0 = 7.5$ which is the value of the true mean μ_X we assume under the null hypothesis, and after entering the information asked for from the data stats, just choose again the correct alternate hypothesis just as for the z-test, and the readout gives

$$\begin{gathered} T - Test \\ \mu > 7.5 \\ t = 2.4 \\ p = .01090249267 \\ \bar{x} = 8.3 \\ s_x = 2 \\ n = 36 \end{gathered}$$

of which we see the value of p is what we really want as that is the significance or $P-$Value of the data.

Notice here that the significance of the data is numerically a larger number than in the previous problem even though all the numbers were the same. This is because in this problem we have less knowledge as we are using the sample standard deviation in place of the true population standard deviation which is unknown to us. Therefore our whole argument is weakened and our data becomes less significant, which means numerically a larger value for the P-Value. Keep this in mind: very significant data has a very small P-Value. When we speak of "highly significant evidence", we mean evidence or data with a very small P-Value, whereas data with a large P-Value, say above .2 is insignificant. Generally, a P-Value should be down at or below .05 before we pay attention to the data.

FINAL ANSWER: .0109

72. Suppose that in my pond are found both red fish and blue fish and I do not know the true mean weight for either of these two populations. Suppose that I would like to make a 95 percent confidence interval for the amount the redfish on average exceed blue fish in weight. Suppose that I happen to know the population standard deviation in weight of red fish is 3 pounds whereas for the population of blue fish it is 2 pounds. Suppose also that I assume both populations are normal. If I have an independent random sample of 10 red fish having a mean weight of 8.5 pounds with standard deviation 2.8 pounds and an independent random sample of 12 blue fish having mean weight of 7.3 pounds with a standard deviation of 2.3 pounds, then what is the confidence interval for the difference in true means, $\mu_R - \mu_B$, where μ_R denotes the true mean weight of red fish and μ_B denotes the true mean weight of blue fish.

ANSWER: Since we are assuming the populations are normal, we do not need to worry about the small sample sizes. Also, as we know the population standard deviations (a fairly unrealistic situation), we can use the normal distribution to standardize the difference in sample means which we use to estimate $\mu_R - \mu_B$. Let's use subscripts R and B to denote quantities relating to the red fish and blue fish, respectively, and let $m = n_R = 10$ and $n = n_B = 12$. We have

$$SD(\bar{X}_R - \bar{X}_B) = \sqrt{\frac{\sigma_R^2}{m} + \frac{\sigma_B^2}{n}} = 1.110555417,$$

so the standardization of the mean difference $\bar{X}_R - \bar{X}_B$ random variable, denoted Z is

$$Z = \frac{\bar{X}_R - \bar{X}_B - (\mu_R - \mu_B)}{SD(\bar{X}_R - \bar{X}_B)} = \frac{\bar{X}_R - \bar{X}_B - (\mu_R - \mu_B)}{\sqrt{\frac{\sigma_R^2}{m} + \frac{\sigma_B^2}{n}}}.$$

With $C = .95$, we have

$$C = P(|Z| \leq z_C),\ z_C = invNorm([1 + C]/2, 0, 1) = invNorm(.975, 0, 1) = 1.959963986,$$

so the Margin of Error M is

$$M = z_C \cdot SD(\bar{X}_R - \bar{X}_B) = 2.176648621.$$

This means that the 95 percent confidence interval for the mean difference $\mu_R - \mu_B$ is

$$\mu_R - \mu_B = (8.5 - 7.3) \pm M = 1.2 \pm 2.18,$$

or

$$-.98 \leq \mu_R - \mu_B \leq 3.38.$$

Notice that since I know the population standard deviations, the sample standard deviations are ignored.

FINAL ANSWER: $(-.98,\ 3.38)$ **or** $-.98 \leq \mu_R - \mu_B \leq 3.38$

73. Suppose that in my pond are found both red fish and blue fish and I do not know the true mean weight for either of these two populations. Suppose that I would like to make a 95 percent confidence interval for the amount the redfish on average exceed blue fish in weight. Suppose that I do not know the population standard deviation in weight of red fish nor the population standard deviation of blue fish. Suppose also that I assume both populations are normal. If I have an independent random sample of 10 red fish having a mean weight of 8.5 pounds with standard deviation 2.8 pounds and an independent random sample of 12 blue fish having mean weight of 7.3 pounds with a standard deviation of 2.3 pounds, then what is the confidence interval for the difference in true means, $\mu_R - \mu_B$, where μ_R denotes the true mean weight of red fish and μ_B denotes the true mean weight of blue fish.

ANSWER: Since we do not know the population standard deviations, we must use the Student $t-$distribution in place of the standard normal when we use the data to estimate $SD(\bar{X}_R - \bar{X}_B)$. There are two possibilities now. If we assume that the redfish and bluefish have the same population standard deviations, that is if we assume $\sigma_R = \sigma_B$, then we would pool the variances from the sample data to form a single estimate of that common population standard deviation σ. We will take that option here. The pooled variance is denoted S^2_{pool} and is given by

$$S^2_{pool} = \frac{df_R \cdot s_R^2 + df_B \cdot s_B^2}{df_R + df_B},$$

where for each sample, the degrees of freedom df is simply the one less than the sample size. Therefore, $df_R = 9$ and $df_B = 11$, so the total degrees of freedom is $df = 9 + 11 = 20$, so

$$S^2_{pool} = \frac{(9) \cdot (2.8)^2 + (11) \cdot (2.3)^2}{20} = 6.4375$$

We can then compute

$$SD(\bar{X}_R - \bar{X}_B) \approx \sqrt{S^2_{pool}\left(\frac{1}{10} + \frac{1}{12}\right)} = 1.086373938.$$

The Margin of Error M is now

$$M = t_C \cdot SD(\bar{X}_R - \bar{X}_B),$$

where

$$t_C = invT(.975,\ 20) = 2.085963406,$$

so the margin of error is

$$M = (1.086373938)(2.085963406) = 2.26613628,\ \text{or } 2.27,$$

to three significant digits.

Now, this can be done rapidly in the calclator by choosing the 2-Samp T Interval in the calculator's TEST menu. It is on the second page of the menu, so you have to scroll down to get it. If you just want to find M, then enter 0 for each sample mean, enter the sample standard deviations and sample sizes and choose the Pool:Yes option. The read out is

$$2 - SampTInt$$
$$(-2.266,\ 2.2661)$$
$$df = 20$$
$$\bar{x}_1 = 0$$
$$\bar{x}_2 = 0$$
$$S_{x_1} = 2.8$$
$$S_{x_2} = 2.3$$

$$S_{XP} = 2.53722289$$
$$n_1 = 10$$
$$n_2 = 12$$

and form this we see

$$M = 2.2661, \text{ or } 2.27,$$

to three significant digits, the same result as before. Notice here I chose red fish as population 1 and blue fish as population 2. If I had reversed the two, it would have given the same result for the margin of error. Since the difference in means $\mu_R - \mu_B$ is estimated by the corresponding difference in sample means, $8.5 - 7.3 = 1.2$, it follows that the confidence interval is

$$1.2 - 2.2661 \leq \mu_R - \mu_B \leq 1.2 + 2.2661$$

or

$$-1.066 \leq \mu_R - \mu_B \leq 3.47.$$

If we want the confidence interval from the calculator directly, then we put the actual sample means in and the read out now gives for the confidence interval

$$(-1.066, \ 3.4661)$$

which is the same result, to three significant digits. Notice all these computation so far are predicated on the assumption that $\sigma_R = \sigma_B$. If we do not make that assumption, then we do not choose the pool option for variances, and then we must use the rather complicated formula (see the formula page) for the degrees of freedom and we should be aware that the distribution is not as accurately chi-square, so the accuracy is somewhat compromised. We will follow that option in the next problem.

FINAL ANSWER: $(-1.066, \ 3.47)$ **or** $-1.066 \leq \mu_R - \mu_B \leq 3.47$, **assuming** $\sigma_R = \sigma_B$

74. Suppose that in my pond are found both red fish and blue fish and I do not know the true mean weight for either of these two populations. Suppose that I would like to make a 95 percent confidence interval for the amount the redfish on average exceed blue fish in weight. Suppose that I do not know the population standard deviation in weight of red fish nor the population standard deviation of blue fish. Suppose also that I assume both populations are normal, but that their standard deviations may not be the same. If I have an independent random sample of 10 red fish having a mean weight of 8.5 pounds with standard deviation 2.8 pounds and an independent random sample of 12 blue fish having mean weight of 7.3 pounds with a standard deviation of 2.3 pounds, then what is the MARGIN OF ERROR in the confidence interval for the difference in true means, $\mu_R - \mu_B$, where μ_R denotes the true mean weight of red fish and μ_B denotes the true mean weight of blue fish.

ANSWER: Since we do not know the population standard deviations, we must use the $t-$distribution and use our sample data to estimate population variances. Now, we have

$$SD(\bar{X}_R - \bar{X}_B) = \sqrt{\frac{\sigma_R^2}{m} + \frac{\sigma_B^2}{n}}$$

but we do not know the standard deviations and they will be estimated with sample standard deviations. Thus, denoting the estimate of $SD(\bar{X}_R - \bar{X}_B)$ obtained by making these replacements with $\hat{\sigma}$, we have

$$SD(\bar{X}_R - \bar{X}_B) \approx \hat{\sigma} = \sqrt{\frac{(2.8)^2}{10} + \frac{(2.3)^2}{12}} = 1.106721886,$$

as the standard deviation estimate usually called the standard error in this situation. Then

$$M = t_C \cdot \hat{\sigma}$$

where the degrees of freedom used in the $t-$distribution is determined from the complicated formula

$$df = \frac{\left(\frac{\sigma_R^2}{n_R} + \frac{\sigma_B^2}{n_B}\right)^2}{\frac{1}{df_R}(\frac{\sigma_R^2}{n_R})^2 + \frac{1}{df_B}(\frac{\sigma_B^2}{n_B})^2}$$

by replacing all the population variances with sample variances from our data. That result is

$$df = 17.45212386.$$

Notice the result is not even a whole number. The conservative way to proceed here is to use less degrees of freedom when in doubt, so you can round down to the nearest whole. Also, you can calculate the result for both the lower round and the upper round and interpolate. If you use the 2-Samp TInterval in the calculator test menu, it does the calculation for you and gives the degrees of freedom in the readout. Just choose POOL: NO option. the readout is:

$$2 - SampTInt$$
$$(-2.33,\ 2.3304)$$
$$df = 17.45212386$$
$$\bar{x}_1 = 0$$
$$\bar{x}_2 = 0$$
$$S_{x_1} = 2.8$$
$$S_{x_2} = 2.3$$
$$n_1 = 10$$
$$n_2 = 12$$

from which we see the degrees of freedom is at most 17 in the conservative approach here, but it is also acceptable to calculate the margine of error using both 17 and 18 degrees of freedom and interpolate between the two according to the degrees of freedom given by the formula. Keep in mind that you do not really have the exact degrees of freedom since we used the sample variances in place of population variances in the formula giving the degrees of freedom. Thus,

$$t_C = invT(.975, 17) = 2.109815559,$$

so

$$M = t_C \cdot \hat{\sigma} = 2.334979054, \text{ or } 2.33,$$

to three significant digits. We can notice that this is slightly larger than the margin of error given in the readout and that is because the calculator actually uses the degrees of freedom stated in the readout instead of taking the more conservative approach. Also, it is the case that always

$$minimum\{df_R, df_B\} \leq df \leq df_R + df_B,$$

so you cannot get more degrees of freedom here than you can when you pool variances, and you can be assured that the degrees of freedom is at least as much as in the smaller sample. So, if you want to be lazy, or if you do not have the computing capability and are working with an old-fashioned table for the $t-$distribution, then you would just use the minimum degrees of freedom which in this case would be only 9.

If we calculate t_C using 18 degrees of freedom, the result is

$$t_C = invT(.975, 18) = 2.100922015,$$

so to interpolate between the results for both 17 and 18 degrees of freedom, since $df =$ 17.4512386, we have

$$t_C(df = 17.2512386) = 2.109815559 + (.4512386) \cdot [2.100922015 - 2.109815559] = 2.105802449,$$

which gives the margin of error

$$M = t_C \cdot \hat{\sigma} = 2.330537657$$

which is very close to the calculator output.

But, this is NOT what the calculator calculated. In fact, the calculator can calculate χ^2 for any positive number of degrees of freedom, and therefore, as the $t-$distribution is defined in terms of the standard normal and the χ^2-distributions,

$$t_d = \frac{Z_0}{\sqrt{\frac{\chi^2_d}{d}}},$$

where Z_0 is a standard normal independent of χ^2_d, it follows that the calculator can calculate the $t-$distribution for any number of degrees of freedom. Remember, the moment generating function for standard normal Z^2 is simply

$$m_{Z^2}(t) = \frac{1}{\sqrt{1-2t}},$$

and as χ^2_d for d degrees of freedom can be defined as the distribution for the sample total of d independent observations of Z^2, it follows that if W^2 has the χ^2_d-distribution, then

$$m_{W^2}(t) = \left(\frac{1}{\sqrt{1-2t}}\right)^d,$$

and this makes sense even if d is not an integer. Therefore, the moment generating function for χ^2_d for any positive number of degrees of freedom is given by the above formula and this gives the χ^2_d-distribution for any number of degrees of freedom. The formula is in your textbook on page 201. So, when you calculate t_C using invT in the menu, you just use the degrees of freedom calculated from the formula, $d = 17.45212386$. When you do, the result is

$$t_C = invT(.975, 17.45212386) = 2.105667502,$$

and this gives the margin of error

$$M = t_C \cdot \hat{\sigma} = (2.105667502)(1.106721886) = 2.330388309,$$

and to 5 significant digits, this is 2.3304, which is exactly the margin of error given by the calculator.

You can also notice here that to three significant digits, the answer is the same no matter whether you use 17 degrees of freedom or you use the exact number of degrees of freedom from the formula, which, keep in mind, is itself only an estimation as sample variances were used in place of population variances in the formula.

FINAL ANSWER: 2.33

75. Suppose that in my pond are found both red fish and blue fish and I do not know the true mean weight for either of these two populations. Suppose that I happen to know the population standard deviation in weight of red fish is 3 pounds whereas for the population of blue fish it is 2 pounds. Suppose also that I assume both populations are normal. If I have an independent random sample of 10 red fish having a mean weight of 8.5 pounds with standard deviation 2.8 pounds and an independent random sample of 12 blue fish having mean weight of 7.3 pounds with a standard deviation of 2.3 pounds, then what is the significance of this data as evidence that the true means satisfy , $\mu_R > \mu_B$, where μ_R denotes the true mean weight of red fish and μ_B denotes the true mean weight of blue fish.

ANSWER: Since we are assuming the populations are normal, we do not need to worry about the small sample sizes. Also, as we know the population standard deviations (a fairly unrealistic situation), we can use the normal distribution to standardize the difference in sample means which we use to estimate $\mu_R - \mu_B$. Let's use subscripts R and B to denote quantities relating to the red fish and blue fish, respectively, and let $m = n_R = 10$ and $n = n_B = 12$. We have

$$SD(\bar{X}_R - \bar{X}_B) = \sqrt{\frac{\sigma_R^2}{m} + \frac{\sigma_B^2}{n}} = 1.110555417,$$

so the standardization of the mean difference $\bar{X}_R - \bar{X}_B$ random variable, denoted Z is

$$Z = \frac{\bar{X}_R - \bar{X}_B - (\mu_R - \mu_B)}{SD(\bar{X}_R - \bar{X}_B)} = \frac{\bar{X}_R - \bar{X}_B - (\mu_R - \mu_B)}{\sqrt{\frac{\sigma_R^2}{m} + \frac{\sigma_B^2}{n}}}.$$

Notice that so far, our considerations are the same as if we were doing a confidence interval, but now we can use our null hypothesis and data to actually get a value of Z which we denote by z_{data}. The alternate hypothesis we are trying to prove is $\mu_R - \mu_B > 0$, and therefore the null hypothesis is $\mu_R - \mu_B \leq 0$, that is

$$H_0 : \mu_R - \mu_B \leq 0$$
$$H_{ALT} : \mu_R - \mu_B > 0,$$

and therefore our null hypothesis dictates that we assume the difference in means 0. Therefore

$$z_{data} = \frac{(8.5 - 7.3) - 0}{SD(\bar{X}_R - \bar{X}_B)} = \frac{1.2}{1.110555417} = 1.080540405.$$

The P-Value or significance of this data is then the probability that using the same sample method, the data would be as or more contradictory than what our actual data gives. Now, we are using $\bar{X}_R - \bar{X}_B$ to estimate $\mu_R - \mu_B$, and the inequality $\mu_R > \mu_B$ is equivalent to $\mu_R - \mu_B > 0$, and therefore the P-Value is

$$\begin{aligned} P - Value &= P(Z \geq z_{data}) = .5 - P(0 \leq z_{data}) \\ &= .5 - normalcdf(0, z_{data}, 0, 1) \\ &= .1399508371, \text{ or } .140, \end{aligned}$$

to three significant digits. Notice that since I know the population standard deviations, the sample standard deviations are ignored.

To do this in the calculator, we go to the 2-SampZTest in the TEST menu, and enter the information after choosing the stats option. Notice that you have to decide which population you want to call population1 and which to call population 2. This decision is arbitrary, but I find it least confusing to always pick as population 1 the population which you are trying to show has the larger population mean. Then the alternate hypothesis you pick is $> \mu_2$. The readout gives

$$\begin{aligned} 2 - SampZTest \\ \mu_1 > \mu_2 \\ z = 1.080540405 \\ p = 1.399508366 \\ \bar{x}_1 = 8.5 \\ \bar{x}_2 = 7.3 \\ n_1 = 10 \\ n_2 = 12 \end{aligned}$$

You can notice that the answer we calculated directly is slightly larger than that calculated by the calculator for the P-Value denoted p in the calculator readout. In fact, if you try approximating ∞ by 10000, you get

$$P(z_{data} \leq Z \leq 10000) = normalcdf(z_{data}, 10000, 0, 1) = 1.399508366,$$

so from this we conclude that the calculator is programmed to approximate ∞ as 10000 for the standard normal which introduces a slight error (underestimate) in the ninth decimal place. In any case, this is not a problem for getting three significant digits.

FINAL ANSWER: .140

76. Suppose that in my pond are found both red fish and blue fish and I do not know the true mean weight for either of these two populations. Suppose that I would like to establish that the redfish on average exceed blue fish in weight. Suppose that I do not know the population standard deviation in weight of red fish nor the population standard deviation of blue fish. Suppose also that I assume both populations are normal. If I have an independent random sample of 10 red fish having a mean weight of 8.5 pounds with standard deviation 2.8 pounds and an independent random sample of 12 blue fish having mean weight of 7.3 pounds with a standard deviation of 2.3 pounds, then what is the significance of this data as evidence that the true means satisfy , $\mu_R > \mu_B$, where μ_R denotes the true mean weight of red fish and μ_B denotes the true mean weight of blue fish.

ANSWER: Since we do not know the population standard deviations, we must use the Student $t-$distribution in place of the standard normal when we use the data to estimate $SD(\bar{X}_R - \bar{X}_B)$. There are two possibilities now. If we assume that the redfish and bluefish have the same population standard deviations, that is if we assume $\sigma_R = \sigma_B$, then we would pool the variances from the sample data to form a single estimate of that common population standard deviation σ. We will take that option here. The pooled variance is denoted S^2_{pool} and is given by

$$S^2_{pool} = \frac{df_R \cdot s^2_R + df_B \cdot s^2_B}{df_R + df_B},$$

where for each sample, the degrees of freedom df is simply the one less than the sample size. Therefore, $df_R = 9$ and $df_B = 11$, so the total degrees of freedom is $df = 9 + 11 = 20$, so

$$S^2_{pool} = \frac{(9) \cdot (2.8)^2 + (11) \cdot (2.3)^2}{20} = 6.4375$$

We can then compute

$$SD(\bar{X}_R - \bar{X}_B) \approx \sqrt{S^2_{pool}\left(\frac{1}{10} + \frac{1}{12}\right)} = 1.086373938.$$

Notice that so far, our considerations are the same as if we were doing a confidence interval, but now we can use our null hypothesis and data to actually get a value of t which we denote by t_{data}. The alternate hypothesis we are trying to prove is $\mu_R - \mu_B > 0$, and therefore the null hypothesis is $\mu_R - \mu_B \leq 0$, that is

$$H_0 : \mu_R - \mu_B \leq 0$$
$$H_{ALT} : \mu_R - \mu_B > 0,$$

and therefore our null hypothesis dictates that we assume the difference in means 0. Therefore

$$t_{data} = \frac{(8.5 - 7.3) - 0}{1.086373938} = 1.104592036.$$

The significance or P-Value of the data is therefore (look at the alternate hypothesis inequality):

$$P - Value = P(t \geq t_{data}) = .5 - P(0 \leq t \leq t_{data}) = .141224357, \text{ or } .141,$$

to three significant digits.

To do this in the calculator, simply choose the 2-SampTTest in the calculator TEST menu and enter the information. Be sure to keep straight which population you are regarding as population 1 and which you regard as population 2. I always think it is easiest to pick as population 1 the population you are trying to show has the largest true mean, so the alternate hypothesis is chosen $> \mu_2$. If you are merely trying to show the populations have different means, you would chose $\neq \mu_2$, and it does not matter so much, but its is still less confusing to take the population with the largest sample mean as population 1 so that t_{data} is not negative. In any case, the readout gives the P-Value as p. The readout is

$$2 - SampTTest$$
$$\mu_1 > \mu_2$$
$$t = 1.104592036$$
$$p = .141224357$$
$$df = 20$$
$$\bar{x_1} = 8.5$$
$$\bar{x_2} = 7.3$$
$$S_{x_1} = 2.8$$
$$S_{x_2} = 2.3$$
$$S_{X_p} = 2.53722289$$
$$n_1 = 10$$
$$n_2 = 12$$

giving the same results we calculated directly

FINAL ANSWER: .141

77. Suppose that in my pond are found both red fish and blue fish and I do not know the true mean weight for either of these two populations. Suppose that I would like to establish that the redfish on average exceed blue fish in weight. Suppose that I do not know the population standard deviation in weight of red fish nor the population standard deviation of blue fish. Suppose also that I assume both populations are normal but their population standard deviations may not be the same. If I have an independent random sample of 10 red fish having a mean weight of 8.5 pounds with standard deviation 2.8 pounds and an independent random sample of 12 blue fish having mean weight of 7.3 pounds with a standard deviation of 2.3 pounds, then what is the significance of this data as evidence that the true means satisfy , $\mu_R > \mu_B$, where μ_R denotes the true mean weight of red fish and μ_B denotes the true mean weight of blue fish.

ANSWER: Since we do not know the population standard deviations, we must use the $t-$distribution and use our sample data to estimate population variances. Now, we have

$$SD(\bar{X}_R - \bar{X}_B) = \sqrt{\frac{\sigma_R^2}{m} + \frac{\sigma_B^2}{n}}$$

but we do not know the standard deviations and they will be estimated with sample standard deviations. Thus, denoting the estimate of $SD(\bar{X}_R - \bar{X}_B)$ obtained by making these replacements with $\hat{\sigma}$, we have

$$SD(\bar{X}_R - \bar{X}_B) \approx \hat{\sigma} = \sqrt{\frac{(2.8)^2}{10} + \frac{(2.3)^2}{12}} = 1.106721886,$$

as the standard deviation estimate usually called the standard error in this situation. Then

$$t = \frac{\bar{X}_R - \bar{X}_B - (\mu_R - \mu_B)}{\hat{\sigma}},$$

where the degrees of freedom used in the $t-$distribution is determined from the complicated formula

$$df = \frac{\left(\frac{\sigma_R^2}{n_R} + \frac{\sigma_B^2}{n_B}\right)^2}{\frac{1}{df_R}(\frac{\sigma_R^2}{n_R})^2 + \frac{1}{df_B}(\frac{\sigma_B^2}{n_B})^2}$$

by replacing all the population variances with sample variances from our data. That result is

$$df = 17.45212386.$$

Notice the result is not even a whole number. The conservative way to proceed here is to use less degrees of freedom when in doubt, so you can round down to the nearest whole. Also, you can calculate the result for both the lower round and the upper round and interpolate. But, the $t-$distribution is actually defined for any positive real number of degrees. That is because the same is true for the χ^2-distribution. For as the squared standard normal Z^2 has moment generating function

$$m_{Z^2}(t) = \frac{1}{\sqrt{1-2t}},$$

it follows that the sum of d independent squared standard normals, say W_d^2, has moment generating

$$m_{W^2}(t) = \left(\frac{1}{\sqrt{1-2t}}\right)^d,$$

and this makes sense for any positive real number d. The corresponding distribution is given in your textbook in the table on page 201. Then, the $t-$distribution for d degrees of freedom is the distribution of

$$t_d = \frac{Z_0}{\sqrt{\frac{W_d^2}{d}}},$$

where W_d^2 is independent of the standard normal Z_0 with W_d^2 having the χ^2-distribution with d degrees of freedom. Therefore, we have the distribution we need, and we just use the null hypothesis to get the $t-$score of the data which we denote by t_{data}. The alternate hypothesis we are trying to prove is $\mu_R - \mu_B > 0$, and therefore the null hypothesis is $\mu_R - \mu_B \leq 0$, that is

$$H_0 : \mu_R - \mu_B \leq 0$$
$$H_{ALT} : \mu_R - \mu_B > 0,$$

and therefore our null hypothesis dictates that we assume the difference in means 0. This means that

$$t_{data} = \frac{\bar{x}_R - \bar{x}_B - 0}{\hat{\sigma}} = 1.084283247.$$

For the P-Value, looking at the alternate hypothesis, we see the inequality we need to compute the probability of is $t \geq t_{data}$, so

$$P - Value = P(t \geq t_{data}) = .5 - P(t \leq t_{data})$$
$$= .5 - tcdf(0, 1.084283247, 17.45212386)$$
$$= .1464958265, \text{ or } .146,$$

to three significant digits.

If you use the 2-Samp TTest in the calculator test menu, it does the calculation of the degrees of freedom for you and gives the degrees of freedom in the readout. Just choose POOL: NO option. Be careful to keep straight in your mind which population you are choosing to be population 1 and which is population 2 and be consistent throughout all your entries into the calclator. I usually find it best to choose population 1 to be the population with the largest sample mean so as to reduce the negative numbers. In the readout, the P-Value or significance is given the symbol p, and the degrees of freedom is given the symbol df. The readout is:

$$2 - SampTTest$$
$$\mu_1 > \mu_2$$
$$t = 1.084283247$$
$$p = .1464958265$$
$$df = 17.45212386$$
$$\bar{x}_1 = 8.5$$
$$\bar{x}_2 = 7.3$$
$$S_{x_1} = 2.8$$
$$S_{x_2} = 2.3$$
$$n_1 = 10$$
$$n_2 = 12$$

and you see the readout gives the exact same results as out direct calculations.

FINAL ANSWER: .146

78. Suppose that I do not know the mean weight of the fish in my pond and I want to estimate it using the sample mean of an independent random sample. Suppose that I have a sample of 36 fish with sample mean $\bar{x} = 8.3$ pounds. Suppose that I do not know the standard deviation in the weight of fish in my pond, but the sample standard deviation of my sample is 2 pounds. What is the margin of error in the 95 percent confidence interval for the true mean weight of the fish in my pond?

ANSWER: This problem is the same as the preceding except that the standard deviation of the population is unknown and we use the sample standard deviation in its place. Whenever you do that, you must use the Student-t-distribution for $n-1 = 35$ degrees of freedom in place of the standard normal distribution. Thus, the margin of error is now

$$M = t_{(C=.95)} \cdot \frac{s}{\sqrt{n}} = t_{(C=.95)} \cdot \frac{1}{3}.$$

To find t_C we can use the invT in the calculator,

$$t_{(C=.95)} = invT(.975, 35) = 2.030107868,$$

so

$$M = .6767026228, \text{ or } .677,$$

to three significant digits. Notice that the margin of error is increased due to the lack of knowledge of the population standard deviation, even though the sample standard deviation accidentally hits the population standard deviation on the nose. But we do not know that here. Alternately, you can use the t-interval in the calculator. Choose the stats option and enter $\bar{x} = 0$ and $s_x = 2$, and $n = 36$, and then the readout is $(-.6767,\ .6767)$ which means $M = .6767$ or again .677 to three significant digits.

FINAL ANSWER: .677

79. Suppose that I do not know the percentage of ducks that will vote for Donald for Mayor of Duckburg in an upcoming election. Suppose that in a simple random sample of 2000 citizens asked, 1078 say they will vote for Donald in the election. What is the 95 percent confidence interval for the true proportion of citzens of Duckburg who say they will vote for Donald in the upcoming election?

ANSWER: Here, the sample data is used to calculate first the sample total number of yes votes for Donald which we can call T_n, and then the true proportion is estimated by the sample proportion usually denoted $\hat{p}$ given by

$$\hat{p} = \frac{1}{n}T_n = \bar{X}_n,$$

that is to say, the same thing is being done here as in the previous two problems, except that now the basic unknown we are dealing with is simply the unkown

$$X = I_D,$$

where D denotes the statement that the person asked answers yes to voting for Donald.

Now, using $p = P(D)$, for the probability of D, which is the true proportion of voters who support Donald, and since $X^2 = X$ as X is an indicator, and

$$E(X) = E(I_D) = P(D) = p,$$

it follows that

$$Var(X) = Var(I_D) = E(X^2) - [E(X)]^2 = E(X) - p^2 = p - p^2 = p(1-p) = P(D) \cdot P(not\ D).$$

Also, since

$$0 \leq p \leq 1,$$

it follows that

$$p(1-p) \leq (1/2)^2 \leq 1/4,$$

and therfore

$$\sigma_X \leq \frac{1}{2}.$$

Even though T_n is discreet here, with such large $n = 2000$, we can assume that T_n and $\hat{p} = \bar{X}$ are normal. Therefore, M, the margin of error with 95 percent confidence is definitely no more than

$$M \leq z_{(C=.95)} \cdot \frac{(1/2)}{\sqrt{n}} = (1.959963986)\frac{.5}{\sqrt{2000}} = .0219130635, \text{ or } .0219,$$

to three significant digits.

Now, for such a large sample, we can be fairly sure to be reaonalby close to the true proportion (we see we are 95 percent sure we are within .0219), and the standard deviation as a function of the true proportion p does not change very much for slight change in p, so it is accepted statistical practice to use the estimate $\hat{p}$ from the sample data to calculate the value of σ_X thus getting an estimate which we can denote $\hat{\sigma}$ for σ_X here, that is

$$\hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})}.$$

This gives

$$M = z_{(C=.95)} \cdot \frac{\hat{\sigma}}{\sqrt{n}} = z_{(C=.95)} \cdot \sqrt{\frac{T_n(n-T_n)}{n^3}} = .0218463023, \text{ or } .0218.$$

Now, this last calculation is what is done in the calculator when you use the 1-Prop z-interval in the calculator's TEST menu. You enter the information and the readout gives

$$(.51715,\ .56085)$$
$$\hat{p} = .539$$
$$n = 2000.$$

This tells us that with 95 percent confidence

$$.5715 \leq p \leq .56085.$$

To find the margin of error here to compare this result with our previous result, notice that if the confidence interval for μ has the form (a, b), then that is the same as saying with confidence C, we know

$$a \leq \mu \leq b,$$

and if $\bar{x}$ is the value of the sample mean from the data, then

$$a = \bar{x} - M,$$

and

$$b = \bar{x} + M.$$

This means that if we know the confidence interval and we know the value of $\bar{x}$, then we easily find the margin of error

$$M = b - \bar{x}.$$

Here this means

$$M = .56085 - .539 = .02185,$$

which is again .219 to three significant digits.

However, this problem asked for the confidence interval itself, not the margin of error. Thus, we can get the answer from the readout directly, or calculate using

$$a = \hat{p} - M,\ \ b = \hat{p} + M,$$

and that result to three significant digits is agian

$$.517 \leq p \leq .561.$$

Returning to the margin of error here denoted M, we see that $M = .0219$ means that roughly we have an error of $2/100$, and we do not call this a two percent error, but rather an error of 2 percentage points. You should notice the distinction here. Whenever you speak of a percentage error you mean the fraction error divided by true amount, which would be rougly $.02/.539$ or about 4 percent here. That is here, an error of two percentage points is about a four percent error. However, when speaking of a percent error, we need to know the actual true value to know the percentage error, and we do not know that here, we can only guess its precise value to be around 4 percent. On the other hand, the margin of error for the confidence interval is very precisely known.

We should notice here that in using the normal approximation to the binomial, we should also use the continuity correction which we have here ignored and which your calculator ignores too. In fact, we should recall that if T_n is binomially distributed, then to approximate $P(a \leq T_n \leq b)$, where a and b are integers using the normal distribution, we use

$$P(a \leq T_n \leq b) = normalcdf(a - .5, b + .5, np, \sqrt{np(1-p)}\ ),$$

so we should really set this equal to the level of confidence C, so

$$C = normalcdf(a - .5, b + .5, np, \sqrt{np(1-p)}\ ).$$

Now, the margin of error in T_n is nM, and this means that the usual endpoints should be corrected to be

$$a - .5 = T_n - nM$$

and

$$b + .5 = T_n + nM,$$

and therefore the confidence interval for the proportion should really better be calculated as

$$a = \frac{T_n - nM + .5}{n} = \hat{p} - M + \frac{.5}{n}$$

and

$$b = \frac{T_n + nM - .5}{n} = \hat{p} + M - \frac{.5}{n}.$$

Now .5 divided by 2000 is .00025, so this slightly changes the confidence interval, but not in the significant digits. Thus, technically, our confidence interval is

$$(.539 - .0219130635 + .00025,\ .539 + .0219130635 - .00025) = (.5173369365,\ .5606630635)$$

which is the same result, to three significant digits. This means that for samples in the thousands, the confidence interval calculation can safely ignore the continuity correction in the normal approximation to the binomial.

FINAL ANSWER: $.517 \leq p \leq .561$

80. Suppose that I do not know the mean weight of the fish in my pond and I want to estimate it using the sample mean of an independent random sample. Suppose that I have a sample of 36 fish with sample mean $\bar{x} = 8.3$ pounds. Suppose that I know that the standard deviation in fish weight for the fish in my pond is 2 pounds. What is the significance of this data as evidence that the true mean weight of the fish in my pond actually exceeds 7.5 pounds?

ANSWER: As soon as we are asked for significance of data, we know we are dealing with a HYPOTHESIS TEST. Here we know the population standard deviations, so the significance is the value of p in the readout for the z-test in the test menu. We must choose the correct alternate hypothesis. Always remember, the alternate hypothesis is what you are trying to prove with the data. In this case, we are asking for the significance ($= P-$Value) as evidence that the true mean weight exceeds 7.5, so we enter $\mu_0 = 7.5$, and choose $\mu > \mu_0$ as our alternate hypothesis. The result is

$$P - Value = P(\bar{X}_n \geq \bar{x}_{data}) = P(\bar{X}_n \geq 8.3) = P(Z \geq z_{data}),$$

where z_{data} is the standard score for our sample mean under the null hypothesis H_0 which tells us to assume the true mean is $\mu_0 = 7.5$, so

$$H_0 : \mu_X \leq 7.5,$$

$$H_{alt} : \mu > 7.5,$$

$$z_{data} = \frac{8.3 - 7.5}{(2/\sqrt{36})} = \frac{6(.8)}{2} = 2.4,$$

$$P - Value = P(Z \geq 2.4) = .0081975289, \text{ or } .00820,$$

to three significant digits. Using the calculator test menu you begin by choosing the z-test, since you know the population standard deviation here. You then enter $\mu_0 = 7.5$, and $\sigma_x = 2$ as well as the data information asked for. Be sure to choose the correct alternate hypothesis in terms of μ_0. Thus, with $\mu_0 = 7.5$, the alternate hypothesis $\mu > 7.5$ becomes $\mu > \mu_0$, so we put the cursor on $> \mu_0$ and hit the enter button on the calculator. When we choose calculate and hit enter we see in the readout

$$Z - Test$$
$$\mu > 7.5$$
$$z = 2.4$$
$$p = .0081975289$$
$$\bar{x} = 8.3$$
$$n = 36$$

which again gives the same result as we calculated above for the significance of the data, again to three significant digits, the P-Value or significance of the data is .00820.

FINAL ANSWER: .00820

81. Suppose that I do not know the mean weight of the fish in my pond and I want to estimate it using the sample mean of an independent random sample. Suppose that I have a sample of 36 fish with sample mean $\bar{x} = 8.3$ pounds. Suppose that I do not know the standard deviation in the weight of fish in my pond, but the sample standard deviation of my sample is 2 pounds. What is the significance of this data as evidence that the true mean weight of the fish in my pond actually exceeds 7.5 pounds?

ANSWER: Notice that this is again a hypothesis test as we are asked for the significance ($= P-$Value) of the data. Just as for confidence intervals, when dealing with something you measure we must use z or t to test hypotheses concerning the mean, and the distinction between z and t is simply whether or not we know the true population standard deviation. Here we do not, and consequently we will use the $t-$ distribution for $n - 1 = 35$ degrees of freedom. Thus, we have

$$t_{data} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{8.3 - 7.5}{2/6} = 3 \cdot (.8) = 2.4$$

and

$$P{-}Value = P(t_{35} \geq 2.4) = .5{-}P(0 \leq t_{35} \leq 2.4) = .5{-}tcdf(0, 2.4, 35) = .0109249267, \text{ or } .0109,$$

to three significant digits. Using the t-test in the test menu, enter $\mu_0 = 7.5$ which is the value of the true mean μ_X we assume under the null hypothesis, and after entering the information asked for from the data stats, just choose again the correct alternate hypothesis just as for the z-test, and the readout gives

$$\begin{gathered} T - Test \\ \mu > 7.5 \\ t = 2.4 \\ p = .01090249267 \\ \bar{x} = 8.3 \\ s_x = 2 \\ n = 36 \end{gathered}$$

of which we see the value of p is what we really want as that is the significance or $P-$Value of the data.

Notice here that the significance of the data is numerically a larger number than in the previous problem even though all the numbers were the same. This is because in this problem we have less knowledge as we are using the sample standard deviation in place of the true population standard deviation which is unknown to us. Therefore our whole argument is weakened and our data becomes less significant, which means numerically a larger value for the P-Value. Keep this in mind: very significant data has a very small P-Value. When we speak of "highly significant evidence", we mean evidence or data with a very small P-Value, whereas data with a large P-Value, say above .2 is insignificant. Generally, a P-Value should be down at or below .05 before we pay attention to the data.

FINAL ANSWER: .0109

82. Suppose that I do not know the percentage of ducks that will vote for Donald for Mayor of Duckburg in an upcoming election. Suppose that in a simple random sample of 2000 citizens asked, 1078 say they will vote for Donald in the election. What is the significance of this data as evidence that the true proportion of citzens of Duckburg who say they will vote for Donald in the upcoming election actually exceeds .51?

ANSWER: Again, as we are asked for the significance of the data, we know we are dealing with a hypothesis test. In this case we are dealing with a question about the true proportion of citizens who say they will vote for Donald for Mayor of Duckburg. This means that our sample data is really consisting of the number of successes in n independent trials, where we regard it as a success when a citizen says he will vote for Donald for Mayor of Duckburg. The unknown here is the indicator of the event D, denoted I_D, just as in the previous problem where the estimate of the true proportion who say they will vote for Donald for Mayor of duckburg was made as a confidence interval with 95 percent confidence. Thus, the sample total T_n is binomially distributed. There, we used the fact that the binomial for large n is approximately normal. Likewise, we could do that here and in fact, that is what the 1-PropZ-test does, but it is not as accurate as it should be because the continuity correction for passing from the discrete count to the continuous normal has been left out, so the sample sizes have to be in the thousands for good accuracy.

To use the binomial distribution here is actually quite simple. Let p_D be the true proportion of citizens who say they will vote for Donald. Our hypothesis test is then

$$H_{ALT} : p_D > .51$$
$$H_0 : p_D \leq .51$$

and as usual, the inequality we calculate the probability of just mimics the alternate hypothesis. If T_n denotes the yes count, then it is binomially distributed and under the null hypothesis we assume the true proportion is $p_0 = .51$, so

$$P - Value = P(T_n \geq 1078 | T_n \ binomial, \ p_D = p_0, \ n = 2000)$$
$$= 1 - binomcdf(2000, .51, 1077) = .0050347322, \text{ or } .0503,$$

to three significant digits. If you use the 1-PropZTest in the test menu, the readout gives

$$1 - PropZTest$$
$$prop > .51$$
$$z = 2.594357777$$
$$p = .0047384234$$
$$\hat{p} = .539$$
$$n = 2000$$

and we see that the P-Value is off by quite a bit, since the answer to three significant digits is .00474, whereas with a direct calculation using the binomial distribution the result is .503 to three significant digits. To see what happened here, we have to ask what the calculator is actually calculating. It is trying to calculate the previous probability using the normal approximation to the binomial. The proper way to do that is to calculate

$$P - Value = P(T_n \geq 1078) = normalcdf(1077.5, \infty, (2000) \cdot (.51), \sqrt{(2000)(.51)(.49)})$$
$$= .5 - normalcdf((2000)(.51), 1077.5, (2000)(.51), \sqrt{(2000)(.51)(.49)})$$
$$= .0050557815, \text{ or } .506,$$

to three significant digits, which we see is only off in the third decimal place. If we replace the 1077.5 demanded by the continuity correction with 1078, in the previous calculation, the result is .0047384239, which agrees with the 1-PropZTest result to nine decimal places, so from that we can conclude that the calculator is programmed to ignore the continuity correction in using its normal approximation to the binomial for the z-test. This will cause error unless the sample size is around twenty thousand. On the other hand, the calculator's built in binomial distribution may also have error in it for large n, and to be able to choose between the .00503 and the .00506 would be difficult without knowing the details of the calculators program. These kinds of considerations must often be made in real life statistics, and the standard operating procedure should be to err on the conservative side. Here, that would mean choosing the value .00506 as that is larger and we do not want to make an unwarranted claim of a smaller P-Value. You might also notice here, that as far as the seriousness of an election is concerned, as the P-Value is very low in any one of these results, we would be inclined to think of the alternate hypothesis as fairly well established by this data. You always have to keep in mind that in many applications, the level of precision we are calculating here is often a waste of time in view of the overall implications of what you are doing. For instance, in dealing with elections, the significance of .005 is very strong evidence for the alternate hypothesis we are trying to prove, and there is virtually no real distinction in practical terms between the numbers .00474, .00503, and .00506. However, for your classwork in this class, we will stick to three significant digits in order to have a simple a uniform standard.

FINAL ANSWER: .00474, or .00503, or .00506

83. Suppose that I do not know the percentage of mice who support Goofy for Mayor of Duckburg but I have asked 10 mice and 7 say they support Goofy for Mayor of Duckburg whereas three support other candidates. What is the significance of this as evidence that the true proportion of mice who support Goofy for Mayor of Duckburg exceeds sixty percent?

ANSWER: This is a hypothesis test but the test statistic is the total count T of yes votes in an independent random sample of size only 10, so the count T is binomially distributed but the sample size is too small to apply the normal approximation to the binomial., that is, T may not be approximately normally distributed. This means that the binomial distribution in you calculator must be used to calculate the significance of this data, the 1-PropZTest in the TEST menu **CANNOT BE USED HERE.**

The hypothesis we are asking the data to prove is $H_A : p_G > .6$ so the null hypothesis is $H_0 : p_G \leq .6$, where p_G is the true percentage of mice that support Goofy for Mayor of Duckburg.

Our hypothesis test is:

$$H_0 : p_G \leq .6$$
$$\text{vs}$$
$$H_A : p_G > .6$$

and H_A of course is the alternate hypothesis whereas H_0 is the null hypothesis. Thus our null hypothesis tells us to assume the null hypothesis value of $p_0 = .6$ for the true proportion, so T is binomially distributed with $n = 10$ and success rate $p_0 = .6$. When we look at the alternate hypothesis, we see that the data significance is

$$\begin{aligned} P-Value = P(T \geq 7|n = 10, p_0 = .6) &= 1 - P(T < 7|n = 10, p_0 = .6) \\ &= 1 - P(T \leq 6|n = 10, p_0 = .6) \\ &= 1 - binomcdf(10, .6, 6) \\ &= .3822806016, \text{ or } .382, \end{aligned}$$

to three significant digits (and do not confuse significant digits with significant data). You can also notice here that this can be possibly more easily worked from the standpoint of mice who are against Goofy, as if p_A denotes the true proportion of mice who do not support Goofy for Mayor of Duckburg, then to say $p_G > .6$ is equivalent to saying $p_A < .4$, and our data tells us that 3 out of 10 are against Goofy. So the hypothesis test is also equivalent to the hypothesis test

$$H_0 : p_A \geq .4$$
$$\text{vs}$$
$$H_A : p_A < .4$$

and for this we assume the null hypothesis value of $p_0 = .4$ and looking at H_A now tells us to calculate the significance or P-Value as

$$P-Value = P(T \leq 3|n = 10, p_0 = .4) = binomcdf(10, .4, 3) = .3822806016, \text{ or } .382,$$

to three significant digits, the exact same result as before. Thus, you can always rearrange the hypothesis test logically so that the P-Value is calculated directly with the binomcdf in the calculator.

FINAL ANSWER: .382

84. Suppose that I want to make a 95 percent confidence interval for the true mean length of fish in my pond and I know that the length of fish in my pond is normally distributed with a standard deviation of at most $B = 8$ inches. What is the minimum size of an independent random sample from the population of fish in my pond required in order that the margin of error in my confidence interval be at most .25 inches?

ANSWER: Since the margin of error is M, where

$$M = \frac{z\sigma}{\sqrt{n}},\ z = invNorm(.975, 0, 1),$$

it follows that replacing σ by B will only increase M, so if we can choose n so as to make M be at most .25 using $B = 8$ in place of σ, then certainly the true σ will only serve to possibly reduce the margin of error but not increase it. Thus, in the equation

$$M = \frac{zB}{\sqrt{n}},$$

we simply solve for n and round up to the nearest whole number. The result is

$$n \geq \left(\frac{zB}{M}\right)^2,\ z = invNorm(.975, 0, 1),\ B = 8,\ M = .25,$$

so

$$n \geq \left(\frac{z \cdot 8}{.25}\right)^2 = 3933.653838,$$

which rounds up to 3934. Thus,

$$n \geq 3934.$$

FINAL ANSWER: 3934

85. Suppose that I want to make a 95 percent confidence interval for the true proportion of fish in my pond that are redfish. I want the margin of error to be at most .02. What is the miminum size of an independent random sample that will accomplish this.

ANSWER: The sampling for determination of a true proportion p is a special case of sampling a random variable which is an indicator of an event A, in this case, the event is that a randomly chosen fish in my pond is a redfish, and

$$\mu = E(I_A) = P(A) = p.$$

The sample mean in this case is the sample proportion, denoted $\hat{p}$, so if $X = I_A$ is the indicator, then T_n is the count of redfish in the sample and

$$\bar{X}_n = \frac{T_n}{n} = \hat{p}$$

is the sample proportion of redfish which I use to estimate the true proportion. Thus, here the margin of error M is given by

$$M = \frac{z\sigma_p}{\sqrt{n}},\ z = invNorm(.975, 0, 1),\ \sigma_p = \sqrt{p(1-p)} = \sigma_{I_A}.$$

Since $0 \leq p \leq 1$, and $\sigma_p = \sqrt{p(1-p)}$, it follows that

$$\sigma_p \leq \frac{1}{2}.$$

We therefore work this problem just like the previous problem using $B = 1/2$ and the .25 replaced by .02. The result is going to be roughly (since z is slightly less than 2),

$$n \geq \left(\frac{2 \cdot (1/2)}{.02}\right)^2 = \left(\frac{1}{.02}\right)^2 = (50)^2 = 2500,$$

that is in general, for a 95 percent confidence interval for a proportion, a conservative rough estimate that is easy, is simply take the inverse of the allowed margin of error and square it. To get the accurate answer, we calculate

$$n \geq \left(\frac{z \cdot (1/2)}{.02}\right)^2,\ z = invNorm(.975, 0, 1),$$

which gives, since $1/(.02) = 50$,

$$n \geq ((50)(.5) \cdot invNorm(.975, 0, 1))^2 = 2400.911766,$$

and therfore as n must be a whole number, we conclude

$$n \geq 2401.$$

FINAL ANSWER: 2401