STT 315 201-202
Homework due Friday, 7-28-06

1. Binomial probability disrtibution having n = 5, p = 0.7.

a. Binomial coefficient = number of ways to choose 3 things from 5.

$$\binom{5}{3} = \frac{5!}{3!\ 2!} = 10$$

b. Determine p(3) = P(X = 3) from the formula for binomial probabilities.

$$\binom{n}{x} p^x q^{n-x} = \binom{5}{3} 0.7^3\, 0.3^2 = 0.3087$$ (compare with 0.309 in (e))

c. n = 5, p = 0.7, x = 3.

d. For the binomial

E X = n p = 5 (0.7) = 3.5

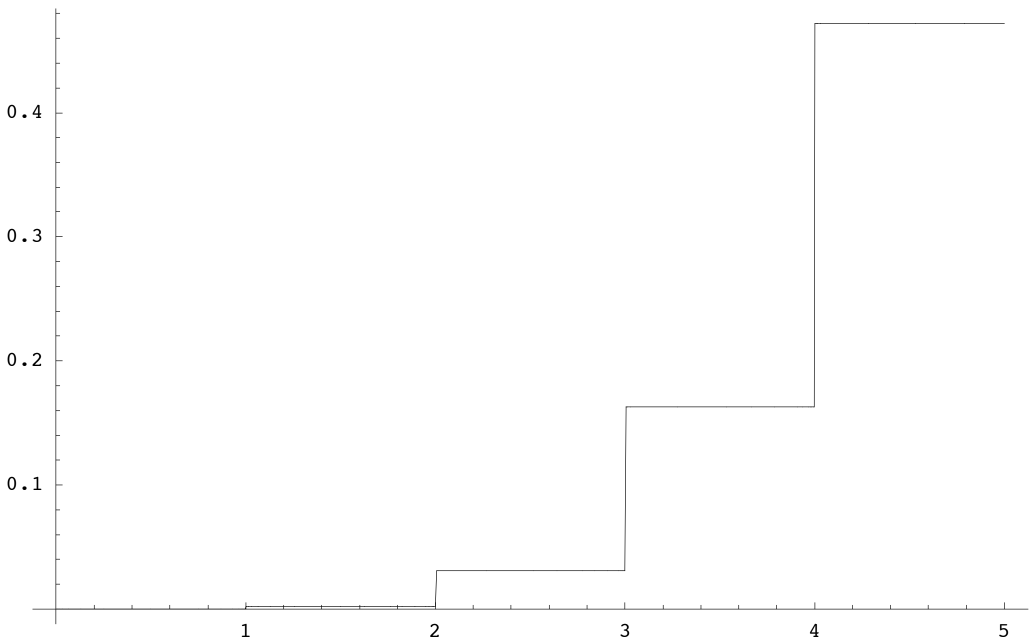
Var X = n p q = 5 0.7 0.3 = 0.105

SD X = $\sqrt{\text{Var}\, X} = \sqrt{0.105} = 0.324$

e. p(3) = P(X = 3) = F(3) - F(2) = 0.472 - 0.163 = 0.309

n		p = 0.7
5	x	
	2	F(2) = 0.163 = p(0) + p(1) + p(2)
	3	F(3) = 0.472 = p(0) + p(1) + p(2) + p(3)

2. a. The plot of F(x) for x from -1 to 5.



b. By eye p(3), the jump in F(x) at x = 3, is easily seen to be around 0.3, close to the value p(3) = 0.309 determined by formula and table.

c. Imagine we have taken 4 (instead of 5 as the problem was worded, for clarity) independent samples, each one a sample of 5 from a population having P(success) = p = 0.7. Further imagine that each of the 4 samples of n = 5 was scored x = the number of "successes" in that sample of 5. We then have 4 random variables X_1, ..., X_4 each of which follows the binomial n = 5, p = 0.7 distribution. I will now use the "Table lookup" method of accomplishing this same thing withou having to select the 4 samples of n = 5. This is called "simulation." To produce any one sample of the binomial n = 5, p = 0.7 what we do is select a random number from the table of random numbers. The first such number in the table is 1559 which we transform to 0.1559. We then look up the cumulative values F(x) for x = 0,1,2,3,4,5 until we find the x at which F(x) first exceeds 0.1559. That x = 2 since F(1) = 0.031 does not equal or exceed 0.1559 but F(2) = 0.163 does equal or exceed 0.1559. Continuing in this way we have the following random numbers from the first row of the table of random numbers together with the x values each generates by this method:

random number =	0.1559	0.9068	0.9290	0.8303
x =	2	5	5	4

It is as though we'd a sample of n = 5 containing 2 "successes" followed by a second sample of n = 5 with 5 "successes" and so forth. Notice that the binomial with n = 5 and p = 0.7 has mean n p = 3.5. Our little experiment of doing the sampling 4 times has produced 2, 5, 5, 4 which averages 16/4 = 4, close to the theoretical mean of 3.5.

3. 40% of sales are taxed. An independent sample of 10 sales contains a random number X taxed.

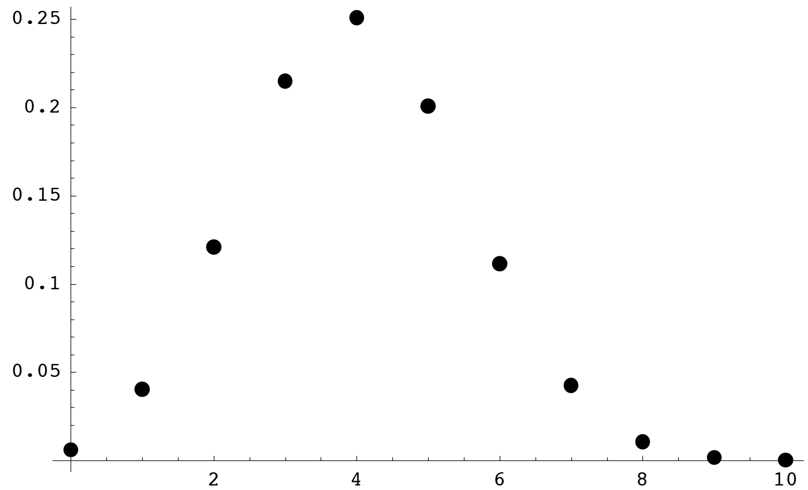
a. R.V. X takes values x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

b. The distribution of X (i.e. the list of possible values x and their probabilities p(x)) is the binomial distribution with n = 10 (independent trials) and p = 0.4 (probability of "success" at each trial).

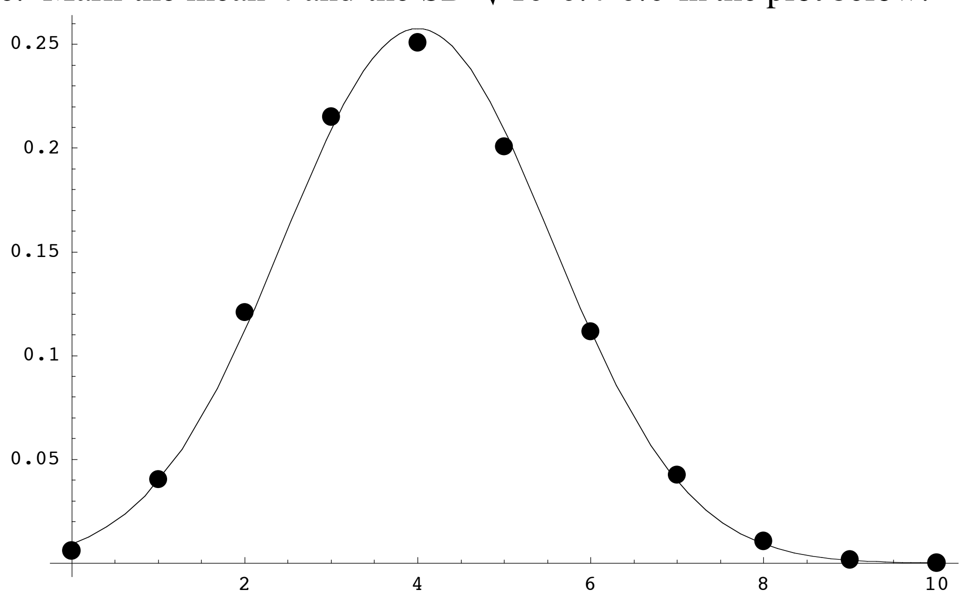
c. E X = n p = 10 (0.4) = 4, Var X = n p q = 10 0.4 0.6 = 2.4, SD X = $\sqrt{\text{Var}\,X} = \sqrt{2.4}$.

d. From the cumulative binomial table for n = 10 and p = 0.4 we obtain values p(x) = F(x) - F(x-1):

$$\begin{pmatrix} 0 & 0.006 \\ 1 & 0.0403 \\ 2 & 0.1209 \\ 3 & 0.215 \\ 4 & 0.2508 \\ 5 & 0.2007 \\ 6 & 0.1115 \\ 7 & 0.0425 \\ 8 & 0.0106 \\ 9 & 0.0016 \\ 10 & 0.0001 \end{pmatrix}$$



e. Mark the mean 4 and the SD $\sqrt{10\ 0.4\ 0.6}$ in the plot below.



f. Suppose your student number ends with "63." You then select row six and the third block of digits in that row. Those digits are 1197 and may be used as a random number 0.1197 in [0, 1]. We will use this random number to generate a sample X from the binomial distribution with n = 10 and p = 0.4. To do so, examine the cumulative distribution table and find the first x for which F(x) is at least 0.1197. This is x = 3 since F(3) = 0.382 is at least as large as 0.1197 but F(2) = 0.167 is not.

g. This was done above.

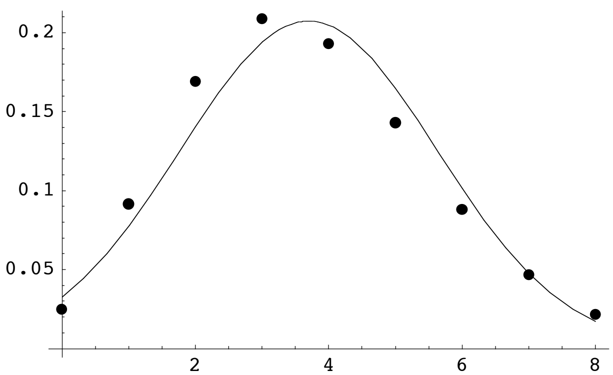
4. R.v. X has Poisson distribution with mean 3.7.

a. Using the Poiosson table, which is not cumulative, p(3) = P(X = 3) = 0.2087

x	μ
	3.7
3	0.2087

b. Using the formual p(3) = $e^{-3.7}\, 3.7^3 \;/\; 3! \;=\;$ 0.20872 (more accurate than (a)).

c. Our rule of thumb for approximating the Poisson by the normal is "$\mu \geq 3$." For the Poisson, SD = $\sqrt{\mu}$. I've plotted the Poisson values together with the normal density with mean 3.7 and SD $\sqrt{3.7}$. As you can see, it is a fairly good approximation.



d. Here is a CUMULATIVE table of Poisson probabilities for μ = 3.7. It was prepared by summing the Poisson probabilities from the table. From (3 f) my random number is 0.1197 (tours will likely differ). This leads to a simulated Poisson value of X = 3 since the third entry of the table below is the first to equal or exceed 0.1197.

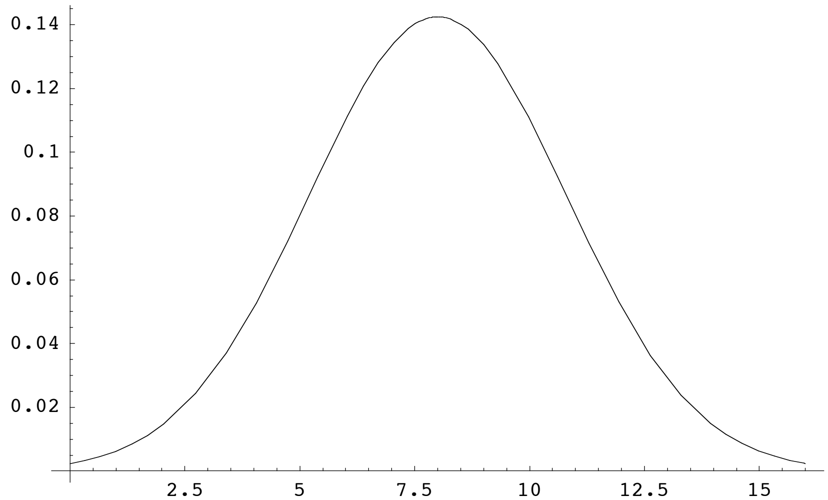
0	0
1	0.0247235
2	0.116201
3	0.285433
4	0.494153
5	0.687219
6	0.830088
7	0.918191
8	0.964759
9	0.986297

5. 2% of electic fans are defective and these events are independent. We've a shipment

of 400 fans. Let X denote the number of defective fans in the shipment.

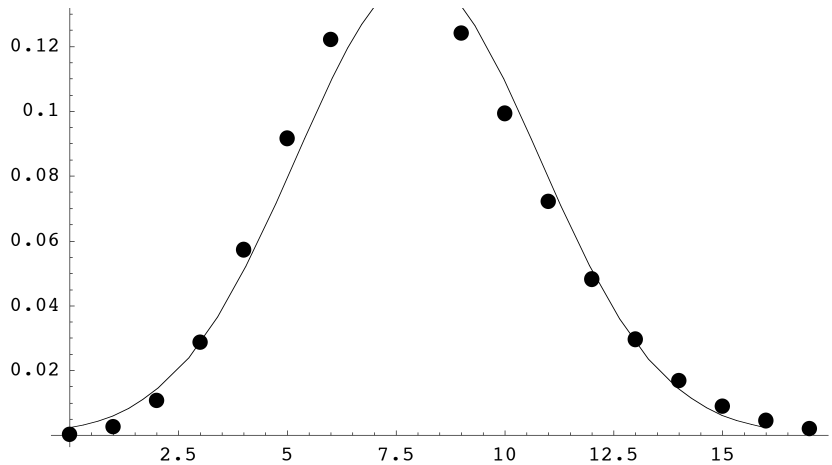
a. R.v. X has the binomial distribution with n = 400 and p = 0.02. Mean = n p = 8, Variance = n p q = 400 0.02 0.98 = 7.84, and SD = $\sqrt{7.84}$ = 2.8.

b. The 95% interval is 8 +/- 2 (2.8) (although 1.96 is a mor accurate choice than 2). Identify this interval, the mean, and the SD in the curve.



c. The Poisson approximation for mu = np = 8 with SD = $\sqrt{8}$ (for the Poisson the sd is the square root of the mean, remember). This is the same mean as the binomial but the sd is slightly different because we are using the sd for the Poisson.

d. I've computed lots of Poisson probabilities for $\mu = 8$ and plotted them with the normal having mean 8 and sd of $\sqrt{8}$.



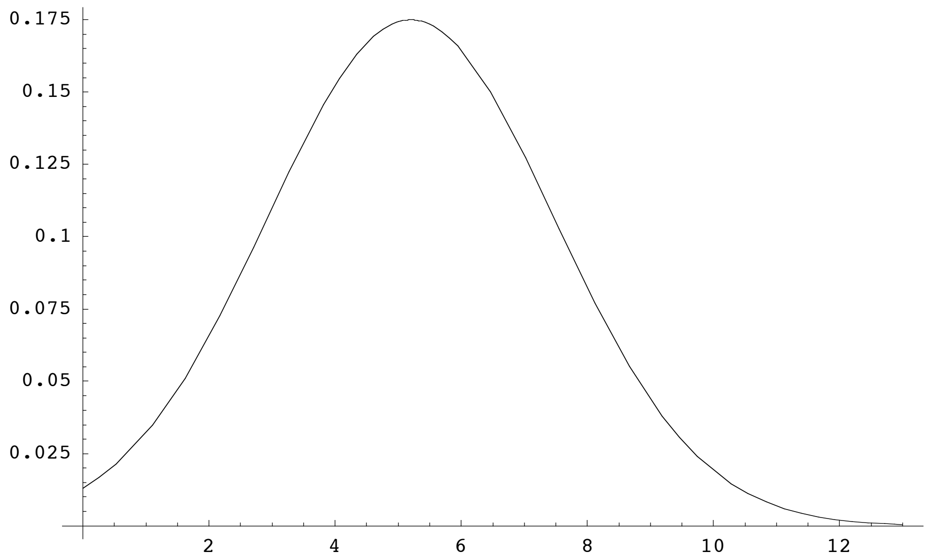
6. Cookies average 5.2 raisins each. This is from a batter making 1000 cookies into which have been randomly mixed 5,200 raisins. The number X of raisins in any particular coookie is approximately Poisson distributed with mean 5.2.

a. p(7) = P(X = 7) ~ $e^{-5.2}\, 5.2^7 \;/\; 7! \;=\; 0.112528$.

b. From the table

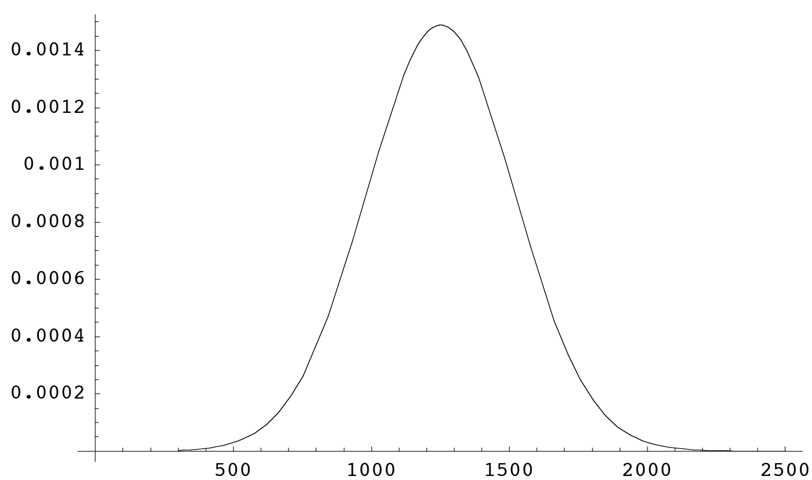
x	μ
	5.2
7	0.1125 (less accurate than (a))

c. Since the mean is at least 3 we invoke the normal approximation with mean 5.2 and sd $\sqrt{5.2}$. Place the mean, sd and 68 % interval in this picture.



7. R.v. X is the number of sales of a produce and is normally distriubted (at least approximately) with mean 1251 and sd 268.

a. Here is the density. Identify the mean and sd in the picture and a 95% interval.



b. Identify 34%, 16%, 47.5%, 2.5% interval rabnges in the above picture. For example, since mean +/ 2 SD is around 95% we know that 5% is in the combined tails. Since the picture is symmetric about the mean there must be around 2.5% to the right of mean + 2 SD (actually, it would be better to use 1.96 SD rather than 2 SD).

c. In the distribution of X, the standard score of 1380 is defined to be the number of SD units that 1380 differs from the mean. That is

standard score of 1380 = (1380 - 1251) / 268 = 0.4813
Standard scores will be negative for values below the mean.

d. P(1251 < X < 1380) = P(std score of 1251 < std score of X < std score of 1380)
= P(0 < Z < 0.48) = 0.1844.

z	0.08
0.1	0.1844

e. P(X < 1251) = 0.5 (the normal is symmetric about its mean). Note that P(X = 1251) = 0 since there is zero area above the single point x = 1251. Continuous models assign their probabilities by integration.

f. P(X > 1380) = 0.5 - P(1251 < X < 1380) = 0.5 - 0.1844

8. Inverse use of normal table.

a. Find z with P(0 < Z < z) = 0.34. This is easy since the interval +/1 contains 68% of the standard normal so the interval [0, 1] will contain 34%. From the table, look for 0.34 (or the slosest entry if it fails to exist) in the BODY of the z-table then read off z.

z	0.00	
1.0	0.3413	(closest value to 0.34 in the BODY of the table)

b. Find z with P(0 < Z < z) = 0.475. This is around z = 2 by the rule of thumb. What does the table give?

z	0.06
1.9	0.4750 (closest entry to 0.475 agrees to 4 decimals)

I does seem that z = 1.96 would be better than "2" used in the rule of thumb for 95% intervals.

c. Find z with P(0 < Z < z) = 0.31. We find the closest table entry to 0.31 in the BODY

z	0.08
0.8	0.3106

d. Sales of a product are approximately normally distributed with mean 1200 and sd 422. Find a 90% interval for sales. Hint: Begin with 0.90/2 = 0.45 and find the z value then convert to the scales of sales. Remember: z scores are in the scale of standard deviations from the mean. Here there is an apparent tie for "closest entry to 0.45 in the body of the table." Either z = 1.64 or z = 1.65 will be an acceptable answer for this exercise but the book often uses z = 1.645 for 90% of normal.

z	0.04	0.05
1.6	0.4495	0.4505

Suppose we choose z = 1.645. Then the 90% interval for **sales** would be around 1200 +/- 1.645 (422) which is around [506, 1894].

9. Combining independent normals. A basic property of normal r.v. is that scale and location changes and sums of them are again normal. This is a very useful feature of statistical process control/assurance. By eliminating unintended sourses of variation we typically produce output scores that have approximately normal distributions and are independent. Aggregates of these are again approximately normal. Another key property of normal distributions is that comparing them amounts to comparing two means and two sd. For these and other reasons the world of normal distributions is a natural result of scientific practices being extended these days to processes of all kinds.

a. Sales in July, August, September are independent and each is approximately normal. The respective means and sd are

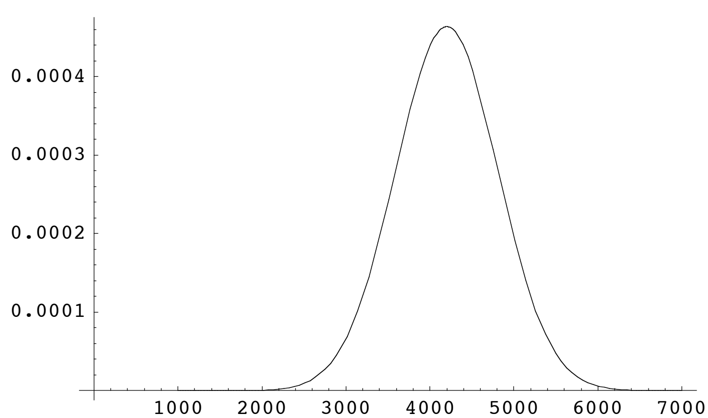
July	1200	300
Aug	1400	400
Sept	1600	700

Total sales is therefore normally distributed with

mean = 1200 + 1400 + 1600 = 4200

sd $= \sqrt{300^2 + 400^2 + 700^2} = 860.233$

Place the mean and sd in the sketch below.



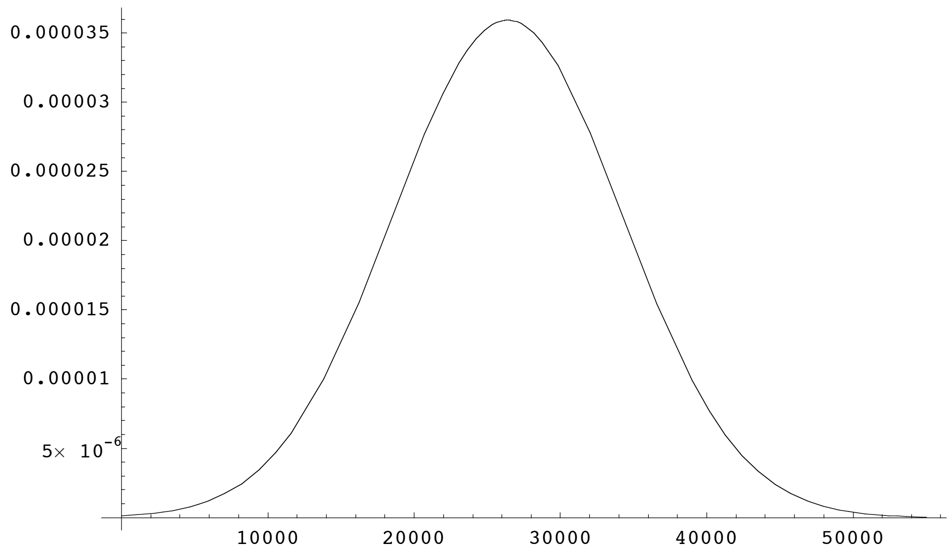
b. Re: July. Each sale brings in $37 gross. Production and other costs in July are normally distributed, independent of sales, with mean 18012 and sd 254. So the net profit for July is

net = 37 sales - costs

mean of net = 37 (1200) - 18012 = 26388

sd of net $= \sqrt{37^2\ 300^2 + 254^2} = 11102.9$

Notice that the variances ADD even though we are subtracting costs. Also, the means of sales and costs play no role in determining the sd of net profit from the sd of sales and costs. Independence is crucial to using "pythagoras' theorem."



c. A 95% interval for net profit is mean +/- 2 SD. This is

26388 +/- 11103 = [15285, 37491]

There is a lot of uncertainty in net profit for these figures!