Probability, Data Acquisition, Model Fitting

Andrew Hofstrand, Dan Matthews, Nathan Quarderer, Christine Roark, Nelson Weber

Project 1: Part A
Team E

The first half of Project 1 asked us to make several calculations using techniques in probability and statistics. Mathcad was used to help us solve many of the problems, and served as a convenient way to type-up our findings.

1. Someone [e.g. the State!] offers the following wager: Place a $1 bet and choose 5 numbers between 1 and 30. If all 5 numbers are guessed correctly (in any order), you win $1 million. The numbers do not repeat (they are unique).

   a. Is this a good bet? [calculate the odds]

Equation 1

$$\left[\left(\frac{1}{30}\right)\cdot\left(\frac{1}{29}\right)\cdot\left(\frac{1}{28}\right)\cdot\left(\frac{1}{27}\right)\cdot\left(\frac{1}{26}\right)\right]\cdot 5! = 7.017\times 10^{-6}$$

These numbers represent the odds of picking the next number or one divided by the total numbers left. The 5! allows these numbers to occur in any order.

   b. Suppose you only had to choose 4 numbers correctly. Is this a good bet? [odds?]

Equation 2

$$\left[\left(\frac{1}{30}\right)\cdot\left(\frac{1}{29}\right)\cdot\left(\frac{1}{28}\right)\cdot\left(\frac{1}{27}\right)\right]\cdot 4! = 3.649\times 10^{-5}$$

=(1/27,405)

These numbers represent the odds of picking the next number or one divided by the total numbers left. The 4! Allows these numbers to occurs in any order.

With odds being (1/142,000) and (1/27,400) respectively, these both appear to be very good bet's indeed (considering that you'd only have to lay down $142,000 to get back 1 million).

2. Find a coin in your pocket.

a. & b. Toss it 100 times. Record the number of heads. Collect the results from other students your team; enter these in your lab notebook.

Table 1

Name	# of Heads
Andrew	63
Chrissy	48
Dan	47
Nate	54
Nelson	44

c. Calculate the mean and standard deviation. Compare with the expected results (for a large number of trials).

Mean = Trials (1+2+3+4+5)/5 = 54+47+48+44+63= 51.2

Percent error = abs(actual – expected)/ actual * 100 = abs(51.2-50)/51.2 *100 = 2.4%

Standard Deviation = (np)^.5, (where n= number of chances and p = probability) = (100*.5)^.5 = 7.07

d. Calculate the probability that the number of heads ***exceeds*** 60.

$\sigma := 7.07 \qquad m := 50$

Equation 3

$$\frac{\displaystyle\int_{m-\sqrt{2}\cdot\sigma}^{m+\sqrt{2}\cdot\sigma} e^{\frac{-(x-m)^2}{2\cdot\sigma^2}}\,dx}{\displaystyle\int_{-\infty}^{\infty} e^{\frac{-(x-m)^2}{2\cdot\sigma^2}}\,dx} = 0.843$$

Using the equation above we get the answer that out of all the chances there is a 84.3% that the number of heads will not exceed 60 or be less than 40.

The probability that it will exceed 60 is found in the equation {1-(probability)}/2 = 1-.843= .157/2= .0785 -> 7.85 % that the number will exceed 60.

e. If you had flipped a coin 10^6 times, what is the probability that the number of heads would exceed 600,000?

The probability will be the same percentage as the equation above even though the decimal point has been moved on the numbers.

3. Suppose you have a drawer infinitely full of socks, red, green, blue in equal numbers.

a. What is probability of randomly choosing a pair of same color?

In this problem you have the probability times the each other:

probability of picking one sock is 1/1 or 1. probability of picking another sock of the same color is 1/3

multiplying together and you get 1*1/3=1/3 chance of pairing up the sock.

b. Now suppose there is exactly one pair of each color. What is the probability of a matching pair now?

This is a similar problem to part a

You have a 1/1 or one probability of picking the 1st sock. The second sock that you pick 1/5 because you only have one sock of the pair left and there are five socks left. This translates into a 1*1/5= 1/5 probability of matching up the socks.

c. Suppose there are 5 pairs Red socks, 10 pairs green, 20 pairs of blue socks. What is probability of picking 3 socks, all the same color?

First you figure out the properties of the red green and blue socks independently. Then you will add those probabilities together and you will get the probability of a total of three socks of the same color.

Red socks: 10/70 * 9/69 * 8/68 = 6/2737 or .0021921812

Green socks: 20/70 * 19/69 * 18/68 = 57/2737 or .0208257216

Blue socks: 40/70 * 39/69 * 38/68 = 494/2737 or .1804895871

Adding the probabilities together we get 2.2/1000 + 21/1000 + 180/1000 = 203/ 1000 = 1/4.97 or roughly a 1/5 probability of grabbing three socks the same color.

4. Galaxies have three morphological types: elliptical (40%), spiral (40%), and irregular (20%). Suppose they are randomly distributed in the Universe.

a. Pick a random elliptical galaxy. What is the probability that its nearest 3 neighbors are all spirals?

You can see the 40% or .4 of all the galaxies on the universe are spirals. This means you have a .4 chance that each of the galaxies are spirals (referring to the closest neighbors)

This works out to the equations .4 * .4 * .4 = .064 or a 8/125 chance of the neighbor being spirals.

b. Pick three random galaxies. What is the probability that none of the three are irregular?

Out of all of the galaxies in the universe 20% or .2 are irregular. All of the other galaxies make up the 80% or..8. Using the same idea as part a we get the equation

.8 * .8 * .8 = .512 or 64/125 or roughly ½ probability that none of the galaxies are irregular.

c. Consider sampling volumes containing ten galaxies each. How many boxes would you need to sample before having a 50% probability that none of the galaxies was a spiral?

For this problem we needed to figure out the number of galaxies that were not spirals. This is .6. Now to get the number of box that fit this probability we figured the equation

(.6 ^10)x = .5 Ten is the number of galaxies per box, x is the number of boxes and .5 ins the probability we want to figure. Now we solved for x

x = .5/(.6^10) = 82.69 or 83 boxes.

5. [Poisson statistics]. Suppose you observe a faint star, counting photons every second. The mean number of photons per second is five.

a. What is the probability of receiving 10 photons in a given second?

Using the Poisson distribution equation:

Equation 4

with a mean number of five photons per second (λ) $P(\lambda, k) := \frac{\lambda^k}{k!} \cdot e^{-\lambda}$

let: $\lambda := 5$ $k := 10$ then

$P(\lambda, k) = .018$

b. 20 photons? (k=20)

$P(\lambda, k)$ =2.641*10^-7

c. Zero photons? (k=0)

$P(\lambda, k)$ = 6.738*10^-3

d. What is the probability that 100 seconds will pass without any one second interval having a count of exactly zero photons?

{1-P(5,0)}^100=0.509 which corresponds to about a 51% probability that over a 100 second period, no one-second sub-intervals will have a count of exactly zero photons

6. Card tricks.

a. What is the probability of dealing an ace as the first card of a full deck?

There are 4 aces out of 52 cards. This translates to equation

$$\frac{4}{52} = 0.077$$

or a 1/13 chance of drawing an ace

b. Probability of dealing two straight aces?

This is the number aces divided by the number of cards time the number of aces left divided the number of remaining cards.

$$\left(\frac{4}{52}\right)\cdot\left(\frac{3}{51}\right) = 4.525\times 10^{-3}$$

$$= 1/220.9944751 \text{ or roughly } 1/221$$

c. Probability of dealing half the deck [26 cards] without a single ace?

Equation 5

$$\left[\left(\frac{48}{52}\right)\cdot\left(\frac{47}{51}\right)\cdot\left(\frac{46}{50}\right)\cdot\ldots\ldots\right]$$

This must be carried out for 1st 26 cards and can be expressed as follows:

$$\prod_{n=27}^{52}\frac{(n-4)}{n} = 0.055$$

d. Probability of a blackjack (ace plus a face card: K, Q, J or 10) in first two cards of a full deck?

This equation translate into the number of aces ~~dived~~ by the number of total cards time the number of ten value cards (i.e. Kings, Queens, Jacks, and Tens) divided by the total number of cards left. The 21 allows this to occur in any order.

Equation 6

$$\left[\left(\frac{4}{52}\right)\cdot\left(\frac{20}{51}\right)\right] + \left[\left(\frac{20}{52}\right)\cdot\left(\frac{4}{51}\right)\right] = 0.06$$ and can be expressed:

$$\left[\left(\frac{20}{51}\right)\cdot\left(\frac{4}{52}\right)\right]\cdot 2! = 0.06$$

This translates to about 1/17 or 1/18 chance of blackjack on the first two cards dealt from a full deck

e. Probability that a poker hand (5 cards) dealt from full deck will contain exactly one pair?

The total number of possible poker hands in a full deck of cards is (52*51*50*49*48)/5!. The first card that is dealt is any one of the 52 cards, the second is anyone of the remaining 51 cards, etc. Then you must divide by 5! to take into account that the order of the cards doesn't matter. The next thing we had to do is to find the number of hands that have exactly one pair: 13*6*48*44*40/3!. The number 13 represents the number of different face values in a standard deck. Six is the number of different pairs you can get from each face value without taking the order of the cards into account. Forty eight is the number of remaining cards that do not have the same face value as the pair. Forty four is the number of remaining cards that are different from the pair and the previously drawn card. The same concept is used to get 40. We divided by 3! to take into account that the order of the last three cards doesn't matter. We then divided the number of hands that contain pairs by the number of possible hands to get the probability:

Equation 7

$$\frac{\left[\frac{(13\cdot 6\cdot 48\cdot 44\cdot 40)}{(3!)}\right]}{\left[\frac{(52\cdot 51\cdot 50\cdot 49\cdot 48)}{5!}\right]} = 0.423$$

7. An astronomer observes an optical and x-ray flare simultaneously in an x-ray binary system. It is known that both optical and x-ray flares occur about once a day for an hour. The astronomer detected the simultaneous flare after observing the system continuously for 10 days What is the probability that this is a coincidence (i.e., that the simultaneous flares are physically unrelated)? *Note: 'simultaneous' means that there is some overlap in the flare times.*

When first analyzing this problem we see that in one particular day the chance that flare A (optical) can occur any time. It is during this time that flare B can occur in roughly a three hour time frame. This was because there was no limit on the length of time that they have to overlap. For example, if flare A occurred from the times 1:00 pm to 2:00 pm then flare B can start anywhere from 12:00:01 (noon and one second) to 1:59:59pm and go until 2:59:59pm. This leaves roughly a three hour window of time for occurrence. This is 3/24 or 1/8th of the day. Problems like this are often easier to work if you look at the chance of it not occurring using that idea we are left with 7/8ths of the day when it will not occur simultaneously. Since we are observing over a 10 day period we need the equation

7/8 * 7/8 * 7/8 * 7/8 * 7/8 * 7/8 * 7/8 * 7/8 * 7/8 * 7/8 or (7/8)^10 = .2630755762 or a 26% chance that it is not a coincidence. To find the probability that it is a coincidence we simply subtract this number from one and we end up with roughly a 75% chance that this is a coincidence.

8. What is the probability that in a room of 30 people, at least 2 people share the same birthday?

In this problem similarly to number 7 it works out better to figure out the chances of the incident not happening. I will first write down the equation and then tell you where I got the numbers.

Equation 8

$$\frac{\prod_{n=336}^{364} n}{365^{29}} = .294$$

The probability for the first person's birthday is 1, and the probability of the next person not having the same birthday is 364/365. For each successive person the numerator decreases by one all the way down to 336/365. The probability of it not happening is equal to .294 or roughly 2/7. This is a very high probability of this not happening. This probability of it happening is around 1 - .294 or about 70%. This means that approximately two thirds of the time you will have two people with the same birthday.

Project 1: Part B
Team E

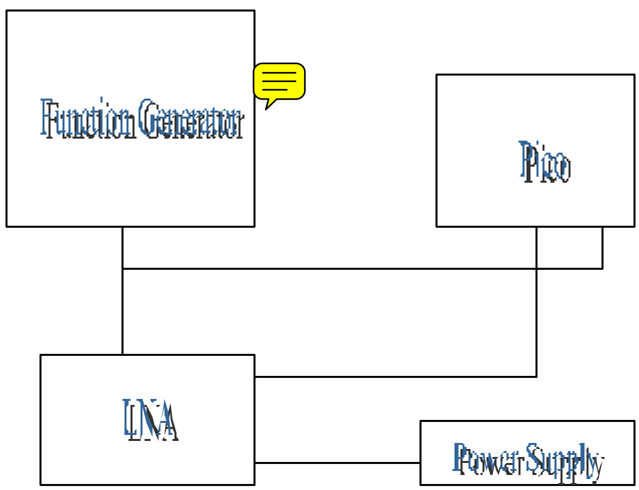
Data Analysis:

Introduction:

Part B of this project asked us to look at the properties of a Low Noise Amplifier, and then fit a simple model (function) to our data. Once our data was plotted, we then ~~went ahead and~~ calculated the uncertainties and the chi-square of our data.

Experiment:

We started by setting up our equipment as follows:



with the Pico device connected to the computer which houses the PicoScope software. With the function generator set to 1 MHz, the 40 dB attenuator depressed, and the sinusoidal waveform with no sweep selected, the power could be turned on. Once the setup was complete we were ready to begin taking data. With the Pico software running we displayed the oscilloscope and dB (power meter) for channels A and B separately, with the input and output levels differing by 20-30 dB.

Gain vs. Input Level

Beginning with our power level at -40 dBm, we recorded our input and output levels (in dB). With our output level displayed, we then calculated our peak-to-peak voltage. Using the definition of the dBm scale provided to us, along with Ohm's Law (***P = IV = V^2/R***) we were able to calculate the power level assuming a 600 ohm impedance. This calculated value was then compared to the dBm meter display.

Calculation of Peak-Peak Voltage:

V(naught) = Root Mean Square Voltage

$$V_0 := 95.7$$

$$\sqrt{2} \cdot V_0 = 135.34$$

$$2 \cdot 135.34 = 270.68$$

Following this calculation, we then went ahead and increased our power level by 3 dB increments, until the input level reached -10dB. By this point (-25 dBm) the output waveforms had noticeably distorted, with their parabaloidal peaks becoming increasingly flattened out.

Output as Function of Input

$$\mathbf{X1} := \begin{pmatrix} -47.66 \\ -44.80 \\ -41.88 \\ -38.60 \\ -35.84 \\ -32.84 \\ -29.70 \\ -26.62 \\ -23.72 \\ -20.72 \end{pmatrix} \quad \mathbf{Y1} := \begin{pmatrix} -14.12 \\ -11.16 \\ -7.94 \\ -4.94 \\ -2.32 \\ .016 \\ 2.88 \\ 5.18 \\ 7.04 \\ 8.48 \end{pmatrix} \quad \mathbf{s1} := \begin{pmatrix} .1 \\ .3 \\ .4 \\ .2 \\ .5 \\ .1 \\ .3 \\ .1 \\ .1 \\ .4 \end{pmatrix}$$

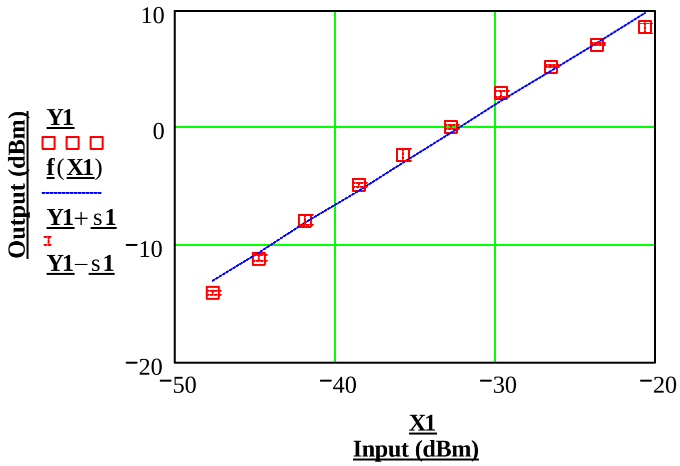
$$\mathbf{slope}(\mathbf{X1}, \mathbf{Y1}) = 0.851$$

$$\mathbf{intercept}(\mathbf{X1}, \mathbf{Y1}) = 27.449$$

$$\mathbf{m} := .851$$

$$\mathbf{b} := 27.449$$

$$\mathbf{f}(\mathbf{x}) := \mathbf{m} \cdot \mathbf{x} + \mathbf{b}$$



Uncertainty in Model Parameters:

$$\mathbf{N} := 10 \qquad \mathbf{i} := 0 .. \mathbf{N} - 1$$

$$D := \sum_{i} \frac{1}{\left(s1_i\right)^2} \cdot \left[\sum_{i} \frac{\left(X1_i\right)^2}{\left(s1_i\right)^2} \right] - \left[\sum_{i} \frac{X1_i}{\left(s1_i\right)^2} \right]^2$$

$$\mathbf{b} := \frac{1}{D} \cdot \left[\left[\sum_{i} \frac{\left(X1_i\right)^2}{\left(s1_i\right)^2} \right] \cdot \sum_{i} \frac{Y1_i}{\left(s1_i\right)^2} - \left[\sum_{i} \frac{X1_i}{\left(s1_i\right)^2} \right] \cdot \sum_{i} \frac{X1_i \cdot Y1_i}{\left(s1_i\right)^2} \right]$$

$$\mathbf{m} := \frac{1}{D} \cdot \left[\left[\sum_{i} \frac{1}{\left(s1_i\right)^2} \right] \cdot \sum_{i} \frac{X1_i \cdot Y1_i}{\left(s1_i\right)^2} - \left[\sum_{i} \frac{X1_i}{\left(s1_i\right)^2} \right] \cdot \sum_{i} \frac{Y1_i}{\left(s1_i\right)^2} \right]$$

$\mathbf{b} = 28.64$ $\qquad$ $\mathbf{m} = 0.888$

Calculation of Chi-Square:

$$c2 := \frac{1}{\mathbf{N} - 1} \cdot \sum_{i} \left(\frac{Y1_i - \mathbf{f}\left(X1_i\right)}{s1_i} \right)^2$$

comes out to about 22 (21.92)

Gain as Function of Input Level

$$\mathbf{x} := \begin{pmatrix} -50.80 \\ -47.74 \\ -44.80 \\ -41.88 \\ -38.60 \\ -35.84 \\ -32.84 \\ -29.72 \\ -26.88 \\ -23.74 \\ -20.72 \end{pmatrix} \quad \mathbf{y} := \begin{pmatrix} 33.68 \\ 33.62 \\ 33.64 \\ 33.94 \\ 33.66 \\ 33.52 \\ 33 \\ 32.6 \\ 31.92 \\ 30.78 \\ 29.2 \end{pmatrix} \quad s := \begin{pmatrix} .2 \\ .1 \\ .3 \\ .1 \\ .4 \\ .2 \\ .1 \\ .3 \\ .2 \\ .1 \\ .2 \end{pmatrix}$$

y= gain
x= input
σ = uncertainty

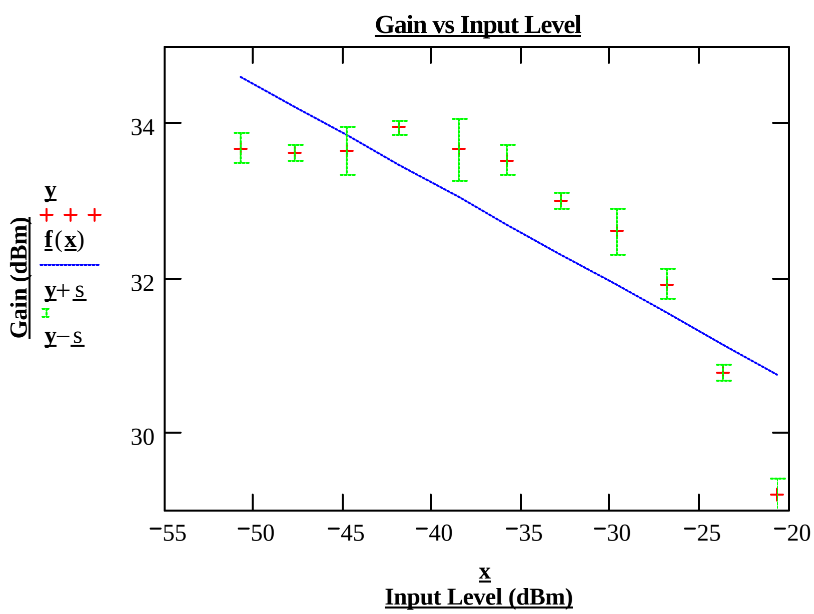
$\mathbf{slope}(\mathbf{x}, \mathbf{y}) = -0.128$

$\mathbf{intercept}(\mathbf{x}, \mathbf{y}) = 28.104$

$\mathbf{m} := -.128$

$\mathbf{b} := 28.104$

$\mathbf{f}(\mathbf{x}) := \mathbf{m} \cdot \mathbf{x} + \mathbf{b}$



Uncertainty in Model Parameters:

$\mathbf{N} := 11$

$\mathbf{i} := 0 .. \mathbf{N} - 1$

$$D := \sum_{\mathbf{i}} \frac{1}{\left(s_{\mathbf{i}}\right)^2} \cdot \left[\sum_{\mathbf{i}} \frac{\left(\mathbf{x}_{\mathbf{i}}\right)^2}{\left(s_{\mathbf{i}}\right)^2} \right] - \left[\sum_{\mathbf{i}} \frac{\mathbf{x}_{\mathbf{i}}}{\left(s_{\mathbf{i}}\right)^2} \right]^2$$

$$b := \frac{1}{D}\cdot\left[\left[\sum_i \frac{(x_i)^2}{(s_i)^2}\right]\cdot\sum_i \frac{y_i}{(s_i)^2} - \left[\sum_i \frac{x_i}{(s_i)^2}\right]\cdot\sum_i \frac{x_i\cdot y_i}{(s_i)^2}\right]$$

$$m := \frac{1}{D}\cdot\left[\left[\sum_i \frac{1}{(s_i)^2}\right]\cdot\sum_i \frac{x_i\cdot y_i}{(s_i)^2} - \left[\sum_i \frac{x_i}{(s_i)^2}\right]\cdot\sum_i \frac{y_i}{(s_i)^2}\right]$$

$m = -0.126$ $\quad$ $b = 28.167$

$$s_m := \sqrt{\frac{1}{D}\cdot\sum_i \left(\frac{1}{s_i}\right)^2} \qquad s_m = ■ \qquad s_b := \sqrt{\frac{1}{D}\cdot\sum_i \frac{(x_i)^2}{(s_i)^2}} \qquad s_b = ■$$

Calculation of Chi-square:

$$c2 := \frac{1}{N-1}\cdot\sum_i \left(\frac{y_i - f(x_i)}{s_i}\right)^2 \qquad c2 = 22.987$$

$$1 - pchisq[c2\cdot(N-1), N-1] = 0$$

During this part of the experiment we were asked to repeat the previous step, only this time we were to eliminate points that clearly deviated from our linear model. Our results were as follows:

$$x := \begin{pmatrix} -50.8 \\ -47.74 \\ -44.8 \\ -41.88 \\ -38.60 \\ -35.84 \end{pmatrix} \qquad y := \begin{pmatrix} 33.68 \\ 33.62 \\ 33.64 \\ 33.94 \\ 33.66 \\ 33.52 \end{pmatrix} \qquad s := \begin{pmatrix} .2 \\ .1 \\ .3 \\ .2 \\ .5 \\ .1 \end{pmatrix}$$

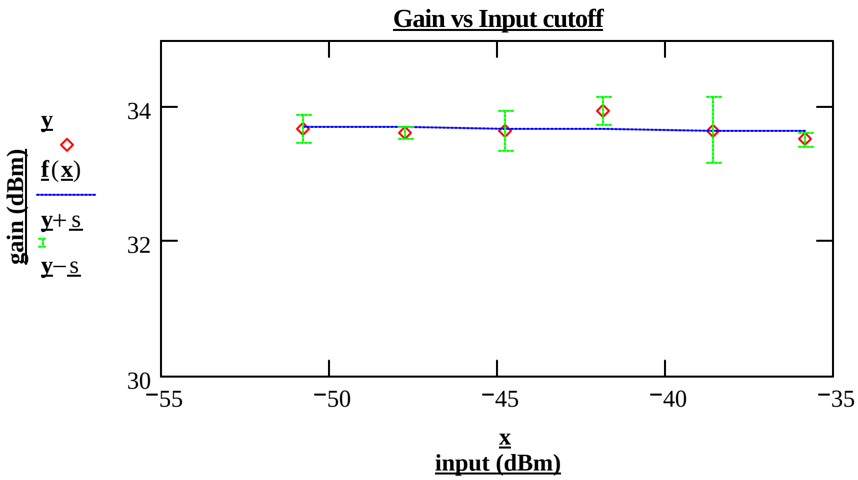
$intercept(x, y) = 28.104$

$m := -3.745\cdot(10^{-3})$

$slope(x, y) = -3.745 \times 10^{-3}$

$b := 33.515$

$f(x) := m\cdot x + b$



At first glance the non-linear graph of Gain vs. Input Level tells us that the gain begins to decrease around -38 dB. Using this, with the function, slope and intercept from the first part of our experiment, we were then able to calculate the 1 dB gain compression level for our amplifier. The calculation went as follows:

f(x)= m*x +b
m= .851
b= 27.449

1 dB compression level = f(x) = -4.889

At this point, we were asked to compare our results to the specifications listed on the manufacturer's website (www.minicircuits.com). According to the website, the suggested gain for our amplifier (ZFL-500LN) is +24 dB +/- .5 . While the 1 dB gain compression is advertised to be +5. Our data points to a gain near +33 dB, and a 1 dB gain compression of -5. These discrepancies may arise from the fact that our amplifier is old and abused, having been stretched beyond it's suggested parameters.

Gain as Function of Frequency

With our power level set to -40 dBm and a frequency of 2.0 MHz, we were ready to add a frequency meter to the output channel of our Pico display. Additionally, we added a spectrum analyzer assigned to the output channel. In the spectrum analyzer window appeared a large peak at 2 MHz, followed by several other (smaller) peaks representing the Fourier components of the signal. We attributed these other peaks to resonant frequencies and overall "noise" of the amplifier.

Following these observations, we went ahead and adjusted the frequency increments down to 200 KHz, and recorded the gain (output-input) at each frequency.

When plotting the gain as a function of frequency, we tried fitting a linear and quadratic model to our data. It was found that the gain was not constant with frequency from .2 to 2 MHz.

$$\mathbf{X} := \begin{pmatrix} 1994 \\ 1805 \\ 1595 \\ 1388 \\ 1191 \\ 982 \\ 805.6 \\ 602.7 \\ 398 \\ 199.1 \end{pmatrix} \qquad \mathbf{Y} := \begin{pmatrix} 33.82 \\ 33.72 \\ 33.56 \\ 33.58 \\ 33.52 \\ 33.64 \\ 33.76 \\ 33.76 \\ 33.92 \\ 34.2 \end{pmatrix}$$

$$\mathbf{slope}(\mathbf{X}, \mathbf{Y}) = -1.955 \times 10^{-4}$$

$$\mathbf{intercept}(\mathbf{X}, \mathbf{Y}) = 33.962$$

$$\mathbf{m} := -1.955 \cdot 10^{-4}$$

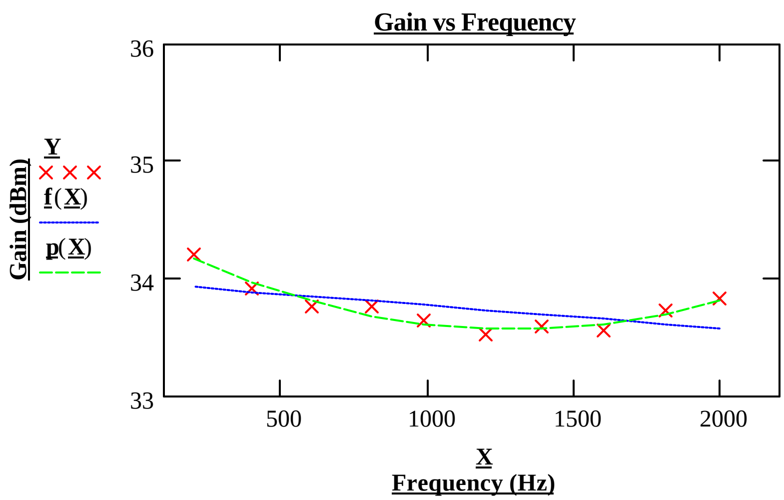
$$\mathbf{b} := 33.962$$

$$\mathbf{f}(\mathbf{X}) := \mathbf{m} \cdot \mathbf{X} + \mathbf{b}$$

$$\mathbf{s} := \mathbf{regress}(\mathbf{X}, \mathbf{Y}, 2)$$

$$\mathbf{s} = \begin{pmatrix} 3 \\ 3 \\ 2 \\ 34.416 \\ -1.329 \times 10^{-3} \\ 5.155 \times 10^{-7} \end{pmatrix}$$

$$\mathbf{p}(\mathbf{X}) := \left(5.155 \cdot 10^{-7} \cdot \mathbf{X}^2\right) + \left(-1.329 \cdot 10^{-3} \cdot \mathbf{X}\right) + 34.416$$



Though the linear model was not a good fit to our data, the quadratic model fits our data quite nicely.

Conclusion:

Part A of this project gave group members an opportunity to brush up on their probability and statistics problem solving skills. Part A also served as a tutorial for the Mathcad software, a skill that will no doubt be popping up again and again throughout this course. In Part B, we were asked to test the properties of a Low Noise Amplifier, and then fit appropriate models to our data, along with making calculations of uncertainty and confidence. Though our findings did not match up with those indicated by the manufacturers specifications, the project gave us a good opportunity to practice fitting models to experimental data (no matter how poor the data may be).

Certificate of Effort:

Though different tasks were preformed by different group members (some tasks harder than others), without the full effort of all the group members, this project would never have been completed on time. Despite the fact that some members had other priorities that kept from their duties as a group member, each member was able to complete his or her task on time. I therefore feel that all group members should receive the same grade on this project.