LaACES
Student
Ballooning
Course

Errors and Uncertainty
Part 2

LaACES
Student
Ballooning
Course

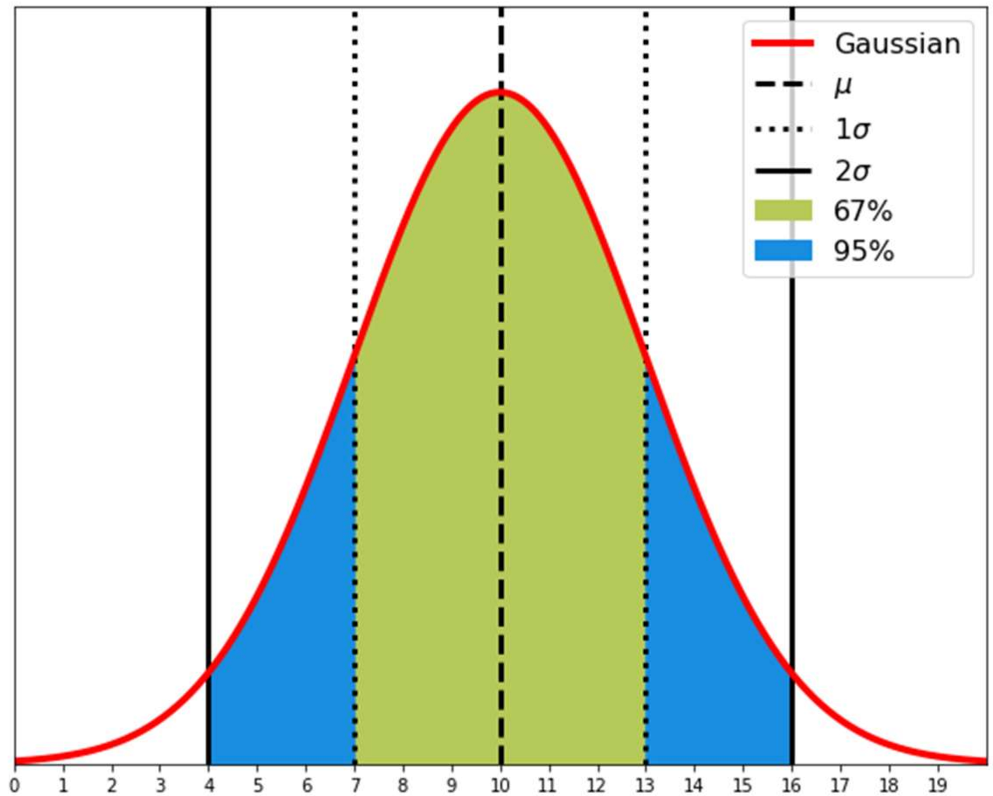
What is a Distribution

- Gives the relative chance(probability) getting a value when making 1 measurement of a particular quantity
  - As you make repeated measurements you are pulling more possible values out of the distribution
- You usually make a guess about the distribution for the measurement based on previous measurements, often assume the Normal Distribution
- When you make a single measurement you sampling the distribution, with multiple samples we can start to make more accurate statements about the distribution
- Shows how you should see the measurements to be distributed over all possible values if you were able to repeat the measurement and infinite amount of time
- Only addresses the random error, systematic error is assumed to be small or 0

LaACES Student Ballooning Course

Gaussian Distribution

- Most commonly used distribution
- Also called Normal Distribution
- $P(x) = k * e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$
- 3 constants (actually only 2) in equation
  - μ is the mean, the center and peak of the distribution and the most likely value
  - σ is the called variance which controls the width
  - The height is how to get that value when making a measurement
  - k is the normalization just scales the whole thing so that the sum (integral) is 1



A Gaussian distribution with μ=10 and σ=3, the vertical lines show 1σ and 2σ from the mean. The green region contains about 67% of the total area and the combined green and blue contain 95% of the total area.

LaACES Student Ballooning Course

The Mean

- There are in fact 2 means we want to think about
- μ the mean of the distribution (Could be thought of as the true mean)
  - May be called expectation value
  - Simply the weighted average by probability of all possible values in the distribution
- X the mean of the sample (i.e. the average value you measured)
  - $X = \sum x_i / N$: where x_i are all the individual measurements and N is the number of measurements
  - This is what is called an estimator, we are trying to estimate some property of the distribution based on finite number of measurements, as N becomes large the estimator approaches the true value

LaACES Student Ballooning Course

Standard Deviations

- Standard Deviation by itself is a somewhat imprecise that could have different meanings in different contexts/fields
- Because of this you want to be specific which one you are using (define the equation somewhere)
- Not the same as the σ used
- 3 Standard Deviations with 3 Different meanings

- Population Standard Deviation

$$\sigma_p = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$$

- Sample Standard Deviation

$$\sigma_s = \sqrt{\frac{\Sigma(x_i - X)^2}{N - 1}}$$

- Standard Deviation of the Mean

$$\sigma_m = \sqrt{\frac{\sigma_s^2}{N}}$$

LaACES Student Ballooning Course

Standard Deviation (Population)

- $\sigma_p = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$
- Note that it depends on μ (the true mean) which for many applications you do not know
- However if your measurements are the entire set of values you are interested in (the entire population) you could use this
- You are trying to estimate the σ of the distribution
- What if we just replace μ with X (sample mean)
- Doing this underestimates the error so we do not want to use it
- As N becomes very large σ_p will equal σ

LaACES Student Ballooning Course

Standard Deviation (Sample)

- $\sigma_s = \sqrt{\frac{\Sigma(x_i - X)^2}{N-1}}$
- We can correct σ_p for the bias
- Because N-1<N σ_s will always be larger than σ_p
- As N becomes the -1 doesn't really matter so for a large enough N, σ_p=σ_s
- Used when you only have a sample of the population or distribution (almost always the case)
- With 1 measurement you get $\frac{0}{0}$, which is undefined, but that makes sense because you can not make any meaningful statement about a distribution based on 1 sample other than saying that sample is in the distribution
- However correcting the bias does not mean σ=σ_p it just mean you are just as likely overestimate as underestimate σ
- As N becomes very large σ_s will equal σ

LaACES Student Ballooning Course

Standard Deviation of the Mean

- $\sigma_m = \sqrt{\frac{\sigma_s^2}{N}}$
- With $\sigma_{p,}\sigma_s$ you are making estimating the error in a single measurement (estimating σ)
- σ_m you estimating how close to X (your sample mean) is to μ (the true mean of the distribution), the error in X
- Unlike $\sigma_{p,}\sigma_s$ as N becomes very large σ_m will become zero

LaACES Student Ballooning Course

Example: Standard Deviation

- Suppose we make 5 measurements of the temperature a with a digital thermometer that reads out to 0.1 Degrees
- First calculate the sum

B7 =SUM(B2:B6)

	A	B	C	D	E
1		Value			
2		35.0			
3		35.0			
4		34.8			
5		34.5			
6		35.4			
7	Total	174.7			
8	N	5			
9	Mean	34.940000			

Example: Standard Deviation

- Suppose we make 5 measurements of the temperature a with a digital thermometer that reads out to 0.1 Degrees
- First calculate the sum
- Next calculate the mean

B9 =B7/B8

	A	B	C	D
1		Value		
2		35.0		
3		35.0		
4		34.8		
5		34.5		
6		35.4		
7	Total	174.7		
8	N	5		
9	Mean	34.940000		

LaACES Student Ballooning Course

Example: Standard Deviation

- Then you need to subtract the mean from each measurement

D6 =B6-B9

	A	B	C	D
1		Value		Value - Mean
2		35.0		0.0600
3		35.0		0.0600
4		34.8		-0.1400
5		34.5		-0.4400
6		35.4		0.4600
7	Total	174.7		
8	N	5		
9	Mean	34.9400		
10				

LaACES Student Ballooning Course

Example: Standard Deviation

- Then you need to subtract the mean from each measurement
- Then square each of them

F6 =D6^2

	A	B	C	D	E	F
1		Value		Value - Mean		Squared
2		35.0		0.0600		0.0036
3		35.0		0.0600		0.0036
4		34.8		-0.1400		0.0196
5		34.5		-0.4400		0.1936
6		35.4		0.4600		0.2116
7	Total	174.7				
8	N	5				
9	Mean	34.9400				

LaACES Student Ballooning Course

Example: Standard Deviation

- Then you need to subtract the mean from each measurement
- Then square each of them
- Sum the squares

F7 =SUM(F2:F6)

	A	B	C	D	E	F
1		Value		Value - Mean		Squared
2		35.0		0.0600		0.0036
3		35.0		0.0600		0.0036
4		34.8		-0.1400		0.0196
5		34.5		-0.4400		0.1936
6		35.4		0.4600		0.2116
7	Total	174.7		Sum of the Squared		0.4320
8	N	5				
9	Mean	34.9400				
10						
11						
12						

LaACES Student Ballooning Course

Example: Standard Deviation

- Then you need to subtract the mean from each measurement
- Then square each of them
- Sum the squares
- Divide by N-1

F8 =F7/(B8-1)

	A	B	C	D	E	F
1		Value		Value - Mean		Squared
2		35.0		0.0600		0.0036
3		35.0		0.0600		0.0036
4		34.8		-0.1400		0.0196
5		34.5		-0.4400		0.1936
6		35.4		0.4600		0.2116
7	Total	174.7		Sum of the Squared		0.4320
8	N	5		Divide by N-1		0.1080
9	Mean	34.9400				
10						

LaACES Student Ballooning Course

Example: Standard Deviation

- Then you need to subtract the mean from each measurement
- Then square each of them
- Sum the squares
- Divide by N-1
- And take the square root

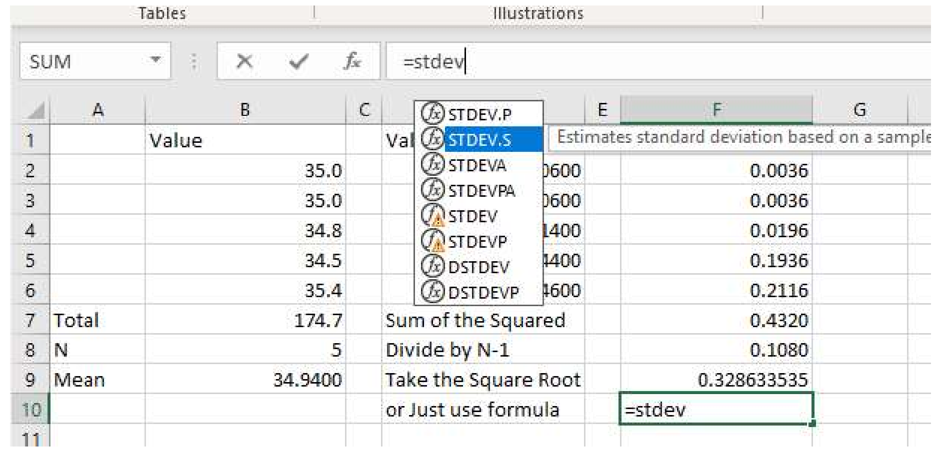
F9 =SQRT(F8)

	A	B	C	D	E	F
1		Value		Value - Mean		Squared
2		35.0		0.0600		0.0036
3		35.0		0.0600		0.0036
4		34.8		-0.1400		0.0196
5		34.5		-0.4400		0.1936
6		35.4		0.4600		0.2116
7	Total	174.7		Sum of the Squared		0.4320
8	N	5		Divide by N-1		0.1080
9	Mean	34.9400		Take the Square Root		0.328633535
10						
11						

LaACES Student Ballooning Course

Example: Standard Deviation

- However you could most programs have a built in standard deviation function
- But be careful to use the correct (sample not population)



LaACES Student Ballooning Course

Example: Standard Deviation

- However you could most programs have a built in standard deviation function
- But be careful to use the correct (sample not population)
- But can see gives the same result

F10 =STDEV.S(B2:B6)

	A	B	C	D	E	F
1		Value		Value - Mean		Squared
2		35.0		0.0600		0.0036
3		35.0		0.0600		0.0036
4		34.8		-0.1400		0.0196
5		34.5		-0.4400		0.1936
6		35.4		0.4600		0.2116
7	Total	174.7		Sum of the Squared		0.4320
8	N	5		Divide by N-1		0.1080
9	Mean	34.9400		Take the Square Root		0.328633535
10				or Just use formula		0.328633535

LaACES Student Ballooning Course

Example: Standard Deviation

- But what if we did the measurement with a bulb thermometer that could only has 0.5 deg resolution
- We get 0 for both the Standard Deviation and SD of the mean
- Does that mean no error?

F16

	A	B	C	D	E	F	G
1		Value		Value - Mean		Squared	
2		35.0		0.0000		0.0000	
3		35.0		0.0000		0.0000	
4		35.0		0.0000		0.0000	
5		35.0		0.0000		0.0000	
6		35.0		0.0000		0.0000	
7	Total	175.0		Sum of the Squared		0.0000	
8	N	5		Divide by N-1		0.0000	
9	Mean	35.0000		Take the Square Root		0	
10				or Just use formula		0	
11							
12				SD of Mean		0	
13							

LaACES Student Ballooning Course

What if my Standard Deviation is 0(or very small)?

- Let's say I measure the length of a metal bar with a ruler 10 times with a ruler marked in mm and I get 12 each time
- Calculating the σ_s you get 0 so I know the bar is exactly 12 mm, no uncertainty, down to the smallest fraction of a mm, right?
- NO! We have completely left out the other type of uncertainty, systematic
- Since the ruler is only marked in 1mm increments we would probably want to estimate the systematic error to be at least that large
  – Maybe you could argue 0.5mm but clearly if this was a digital measurement you couldn't go smaller than the last displayed digit
  – You would also want to include any accuracy given by the manufacturer specifications
- Need to estimate the systematic uncertainty and add it to the random uncertainty

LaACES Student Ballooning Course

Adding Errors

- Clearly the simplest solution would be to just add the errors together

$$\sigma = \sigma_1 + \sigma_2$$

- But we don't really expect them both to be at a max at the same time so can instead add them in quadrature

$$\sigma = \sqrt{(\sigma_1)^2 + (\sigma_2)^2}$$

- This assumes independent variables and normal distribution

LaACES Student Ballooning Course

Example: Adding Error

- Returning to our temperature example we can add the systematic and random errors
- $\sigma = \sqrt{(\sigma_{rand})^2 + \left(\sigma_{sys}\right)^2}$
- For the bulb thermometer its easy, the random error we calculated was 0 so:
- $\sigma = \sqrt{(0)^2 + 0.5^2}$=0.5°C
- For a less trivial example lets look at the digital thermometer
- $\sigma = \sqrt{(0.1)^2 + (0.3286)^2}$
  = 0.34347 °C

LaACES Student Ballooning Course

Propagation of Error

- If f is a function of variables (x_1,x_2,…)

$$\sigma_f = \sqrt{\left(\frac{\partial f}{\partial x_1}\sigma_{x1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\sigma_{x2}\right)^2 + \cdots}$$

- This is a generalization of the addition formula
- Again assumes independent variables and normal distribution
- Partial Derivative treat all other variables as constants and take the derivative of that one variable (feel free to look up the derivatives)

Example Propagation of Error

- We want to know the volume of a rectangular object with dimensions of 10 mm x 12 mm x 5 mm
- The error for each measurement is dominated by systematic for each is 0.5mm

- $V = l * w * h$
- $V = 10mm * 12mm * 5mm = 600\ mm^3$

LaACES Student Ballooning Course

Example Propagation of Error

- $V = l * w * h$
- V is a function of 3 variables l, w, and h
- $\sigma_V = \sqrt{\left(\frac{\partial V}{\partial l}\ \sigma_l\right)^2 + \left(\frac{\partial V}{\partial w}\sigma_w\right)^2 + \left(\frac{\partial V}{\partial h}\sigma_h\right)^2}$
- $\frac{\partial V}{\partial l} = w * h \quad \frac{\partial V}{\partial w} = l * h \quad \frac{\partial V}{\partial h} = l * h$
- $\sigma_V = \sqrt{(wh\sigma_l)^2 + (lh\sigma_w)^2 + (lw\sigma_h)^2}$

LaACES Student Ballooning Course

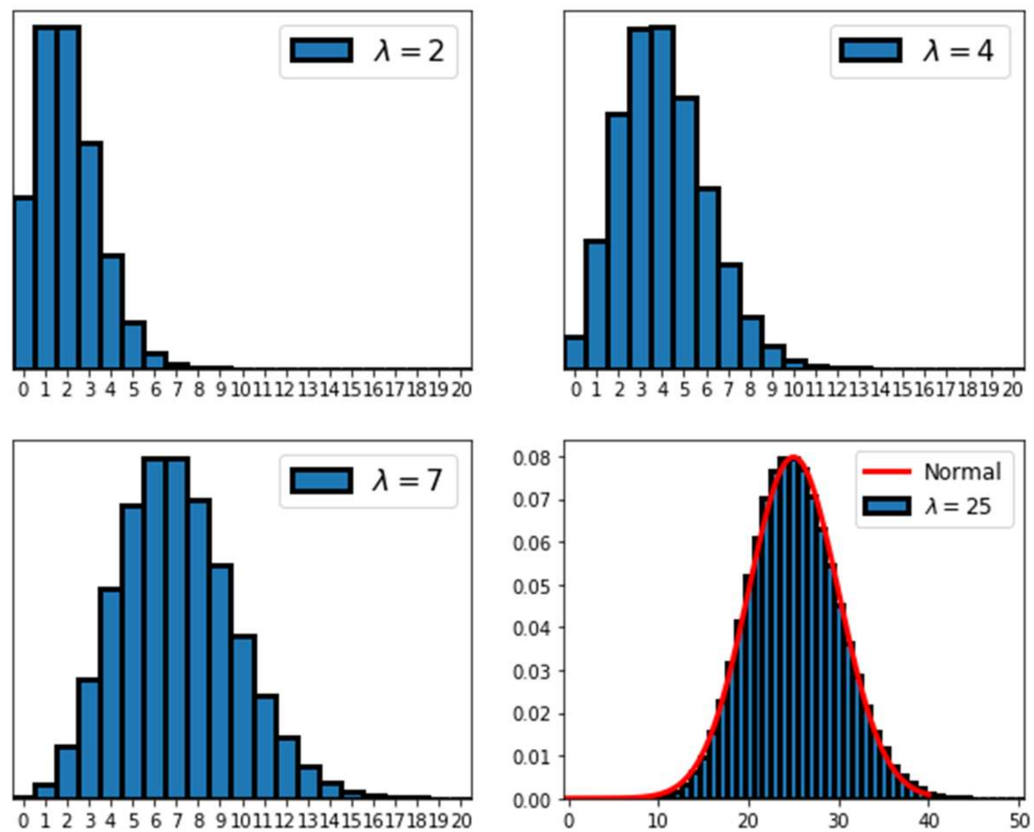
Example Propagation of Error

- $\sigma_V = \sqrt{(wh\sigma_l)^2 + (lh\sigma_w)^2 + (lw\sigma_h)^2}$
- Notice each term in parenthesis has units of volume
- $\sigma_l = \sigma_w = \sigma_h = 0.5\, mm$
- $\sigma_V = \sqrt{(12 * 5 * 0.5)^2 + (10 * 5 * 0.5)^2 + (10 * 12 * 0.5)^2}$
- $\sigma_V = 71.5\, mm^3$
- The volume is 600±70 mm^3
- More examples available in R05.02 Propagation of Error
- https://laspace.lsu.edu/laaces/wp-content/uploads/2020/08/R05.02_Propagation_of_Errors.pdf

LaACES Student Ballooning Course

Poisson Distribution

- Used for many types of counting measurements like radioactive decay, photon counting, river flooding..
- k is number of occurrences, λ is the average number of occurences
- $P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ (Probability of having 0 floods, 1 flood, k floods in the next 100 years, when on average have λ every 100 years)
- When λ is very large just becomes a normal distribution



Poisson distributions of various λ

But why Gaussian?

- If there are other distributions, why do we usually assume a Gaussian Distributions
- In the large number case (big n or λ) other distributions become close to a Gaussian
- There is good math for doing propagation and error handling
- It is a good model for many physical measurements
  - Can prove this is the case from a very many very small errors adding up from the Central Limit Theorem

LaACES Student Ballooning Course

Reporting Measurements

- If I think the error in a measurement is 0.5 mm does it make sense to report the average as 12.003mm
- The common practice is to round the error to 1 or 2 significant digit and then round the corresponding measurement to that digit
  - So we would report the values as 12.0±0.5mm
- However do not round intermediate values used for calculations because you do not want to have rounding errors compound
- Also errors should have the same units as the measurement
- You want to be clear about how you have calculated errors and what you mean with your ±, show your work

LaACES Student Ballooning Course

Putting it all together

- Let’s assume, we first did repeated temperature measurements at one temperature to show the random error is small compared to the 0.5 error from our bulb thermometer
- From the pixel size and signal width in Spectrogram software we estimate the systematic uncertainty of to be 11 Hz
- We decide we need to take 5 independent frequency measurements at each temperature

	Temp (C)	95.0	Error in T	0.5
	Frequencies (Hz)			
Beep #	1	2		
1	3767	5038	1271	
2	3724	4995	1271	
3	3746	4995	1249	
4	3746	5016	1270	
5	3767	4995	1228	
		Mean	1257.8	
		Std Dev	19.12328	
		SD Mean	8.552193	
		Syst Err	11	
		Total Freq		
		Error	13.93341	

LaACES Student Ballooning Course

Putting it all together

- We will calculate the mean of our 5 frequencies and use that as our fitting point
- We then need to calculate the standard deviation of that mean to determine the random frequency error for that mean value
- The we need to add the systematic frequency error to the random to find the total frequency error
- Doing this gives us our first datapoint (1257.8±14 Hz, 95.0 ±0.5°C)

	Temp (C)	95.0	Error in T	0.5
	Frequencies (Hz)			
Beep #	1	2		
1	3767	5038	1271	
2	3724	4995	1271	
3	3746	4995	1249	
4	3746	5016	1270	
5	3767	4995	1228	
		Mean	1257.8	
		Std Dev	19.12328	
		SD Mean	8.552193	
		Syst Err	11	
		Total Freq		
		Error	13.93341	

Calculating all our data points

LaACES Student Ballooning Course

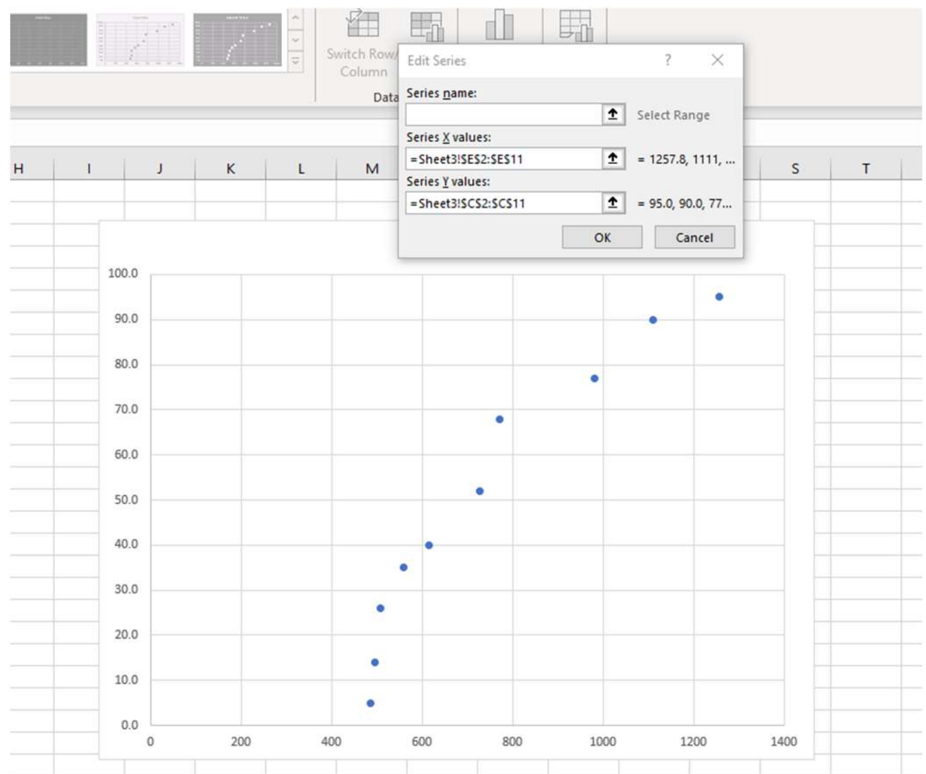
- Now repeat the process for all temperature and frequency measurements
- This gives us a set of x (frequency) and y (temperature) points with a horizontal and vertical error for each datapoint

	C	D	E	F	G
1	Temperature (C)	Error T	Frequency (Hz)	Error f	
2	95.0	0.5	1257.8	13.93341	
3	90.0	0.5	1111	16.90266	
4	77.0	0.5	981.8	15.44798	
5	68.0	0.5	771.6	11.84736	
6	52.0	0.5	728	11.7047	
7	40.0	0.5	616	12.21475	
8	35.0	0.5	560	13.01538	
9	26.0	0.5	508.4	12.19672	
10	14.0	0.5	495	11	
11	5.0	0.5	487.2	12.24908	
12					

LaACES Student Ballooning Course

Plot the points

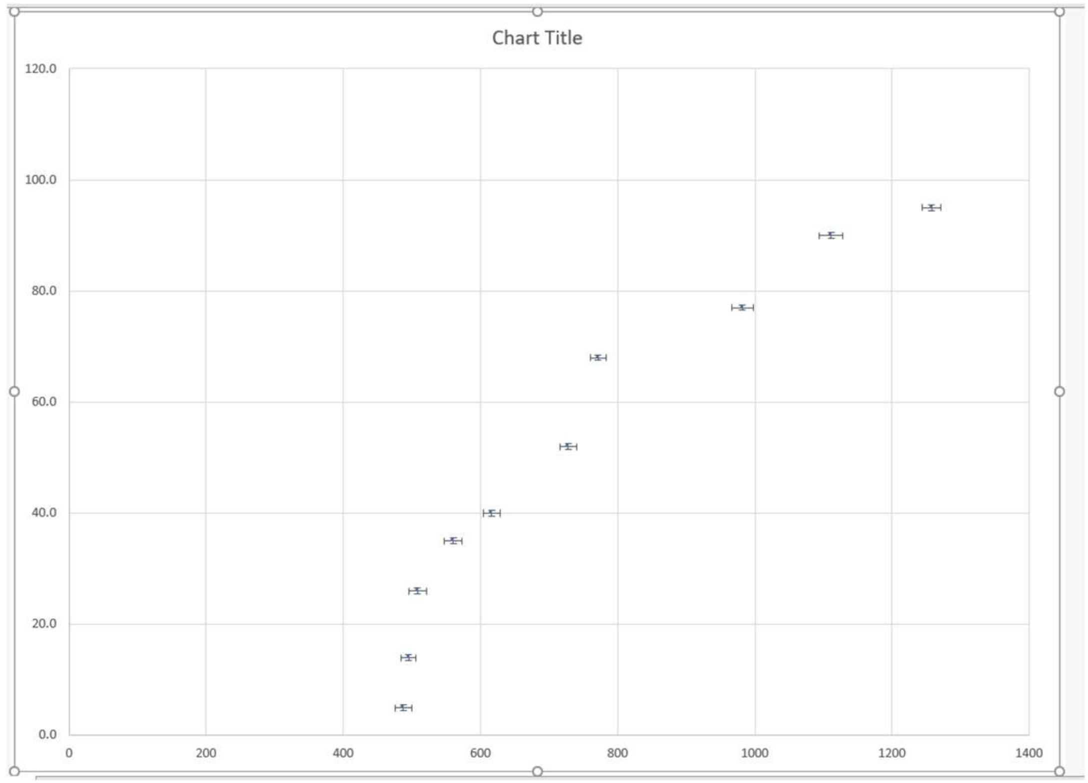
- Now we want to plot all of the points together
- Since the goal of the experiment is to be able to read a frequency and be able to tell what temperature the thermistor is, we want T as a function f
- So we pick f as the x values and T as the y values



LaACES Student Ballooning Course

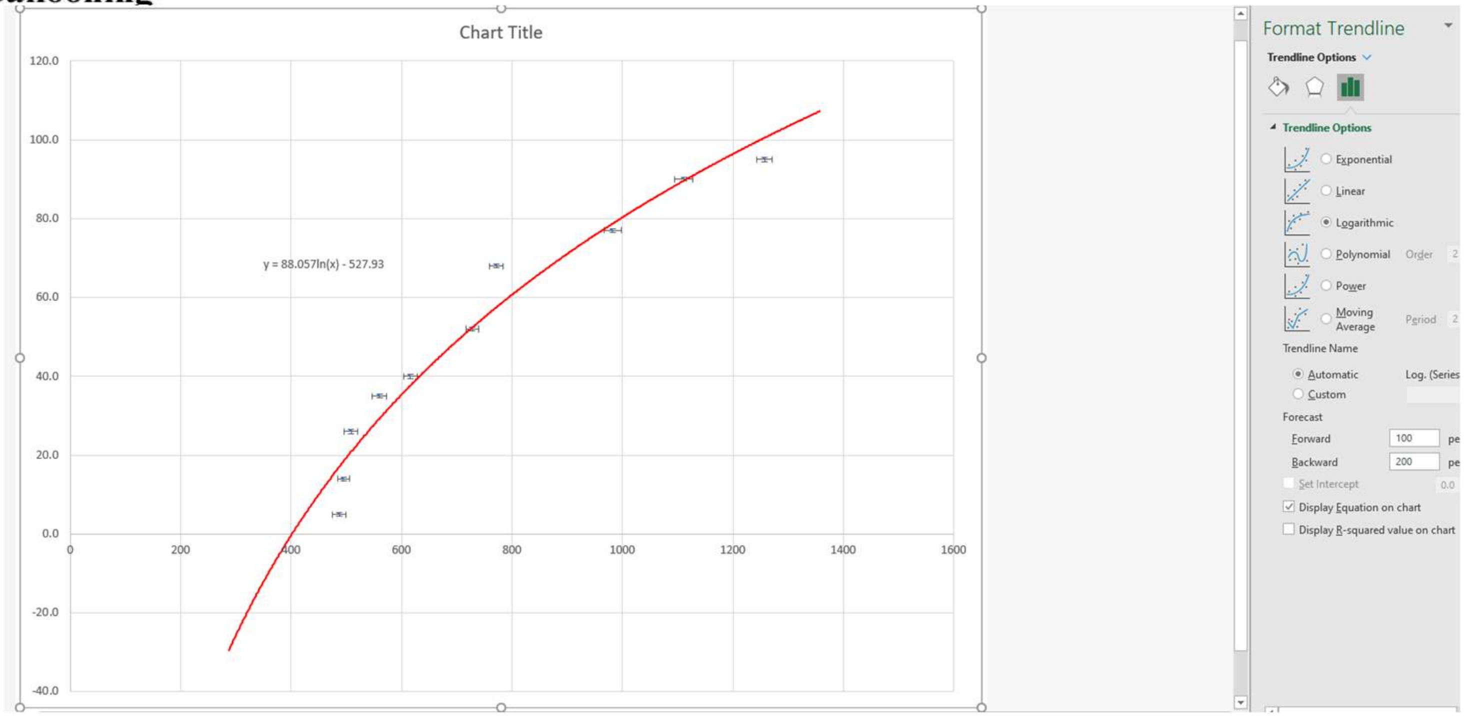
Add Error Bars

- We add the error bars to our plot
- Select in this case our error is symmetric, so we use the same value for both the positive and negative error
- Additionally we want to select the option that lets use specify values for error and not a percentage, standard deviation, or fixed value for example
- Add both horizontal and vertical error bars, using the errors we calculated
- It may be necessary to adjust to point marker size or add a caption if the errors are small



LaACES Student Ballooning

Add a trendline



- Finally we want to add a trendline
- Clearly the data is not linear
- Ideally, we would have some theoretical basis for picking a particular fit but we can also try seeing what matches the data
- Also probably want to show the equation of the fit on out plot

LaACES Student Ballooning Course

Finish the Plot and Draw Conclusions

- Don’t forget to add the axes titles, units, plot title, etc.
- If we have a good fit and correctly assessed our errors, we expect ~2/3 of our points error bars to overlap with our fit line (Remember 67% 1σ)
- Many less than 2/3
  - Maybe not a good fit function
  - Possibly underestimated errors, missed systematics
- Many more than 2/3
  - Too many constants in your function
  - Overestimated error, manufacturer specification often give “guaranteed to be this accurate” rather than a more scientific error