Model testing

- Usually in your laboratory work, you know the function your data should fit.
- Sometimes you don't know the form the data should take.

Examples:

- Data may be poorly modelled by theory.
- You have data from numerical calculation that you wish to generate an approximate model for.
- You want to generate a prediction based on the data trend.

You can use statistical tests to determine which alternative model is best.

The most common tests are based on χ^2 calculations (chi squared).

'observed' value at k

'expected' value at point k

Recall:

$$\chi^2 = \sum_{k=1}^{n} \left(\frac{O_k - E_k}{\sigma_k} \right)^2$$

uncertainty in the observed value (error bar)

Consider a radioactive counting experiment:

- Here we make a histogram of the count values we see from all our trials.

Theoretically, you expect to see values E_k in bin k which can be predicted from the Poisson distribution.

- The actual value in each bin O_k are expected to vary about the theoretical value E_k.
  - The variation in O_k is expected to be within +/- $\sqrt{E_k}$.

'observed' 'expected'

So we expect:

$$\left|\frac{O_k - E_k}{\sqrt{E_k}}\right| \sim 1$$

uncertainty expected in the observed value

For some bins the ratio will be bigger than 1, some will be less,
On average it will be around 1, provided the data actually is given by the Poisson distribution.

Hence,

$$\chi^2 = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}$$

Only true for a Poisson distribution

expect that: $\chi^2 \sim n$

χ^2 ranges:

$\chi^2 = 0$:

- Requires $O_k = E_k$ for each value.
- Agreement between observed and expected is perfect.
- Essentially impossible when random variation is present.

$\chi^2 \sim n$:

- Agreement between observed and expected conforms to random statistics.
- The theoretical model for the "expected" variation is *probably* correct.

$\chi^2 \gg n$:

- There is more variation than would is expected due to random statistics.
- The model giving "expected" values is wrong.

- The Poisson distribution can be approximated by a Gaussian for large mean count values i.e. deviations small compared to the mean value.
- Consequently, we expect the χ^2 measure to be useful in most situations.

It turns out that the actual variation in values is given by yet another distribution function:

The χ^2 probability distribution function.

degrees of freedom ('almost' the number of data points)

$$X(\chi^2, \nu) = \frac{(\chi^2)^{\left(\frac{\nu}{2}-1\right)} \exp(-\chi^2/2)}{2^{\nu/2}\Gamma(\nu/2)}$$

Gamma function

- For large values of ν (i.e. a lot of data points) the χ^2 looks like a Gaussian.

Degrees of Freedom

Degrees of freedom: The number of observed data points minus the number of parameters computed from the data that is used in the calculation.

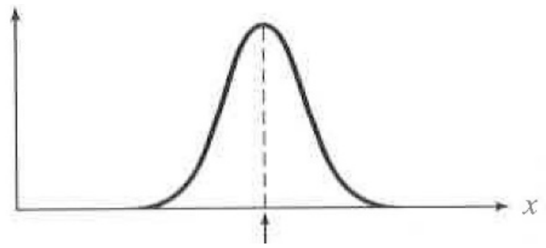
$$\nu = n - c$$

no. of data points

no. of constraints

E.g. Say we are measuring the range of projectile 40 times, and we are comparing the number of times it lands at a certain distance to a Gaussian distribution.

N



X (most probable distance)

$$N = A\,\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-X)^2}{2\sigma^2}}. \qquad \text{(A=40)}$$

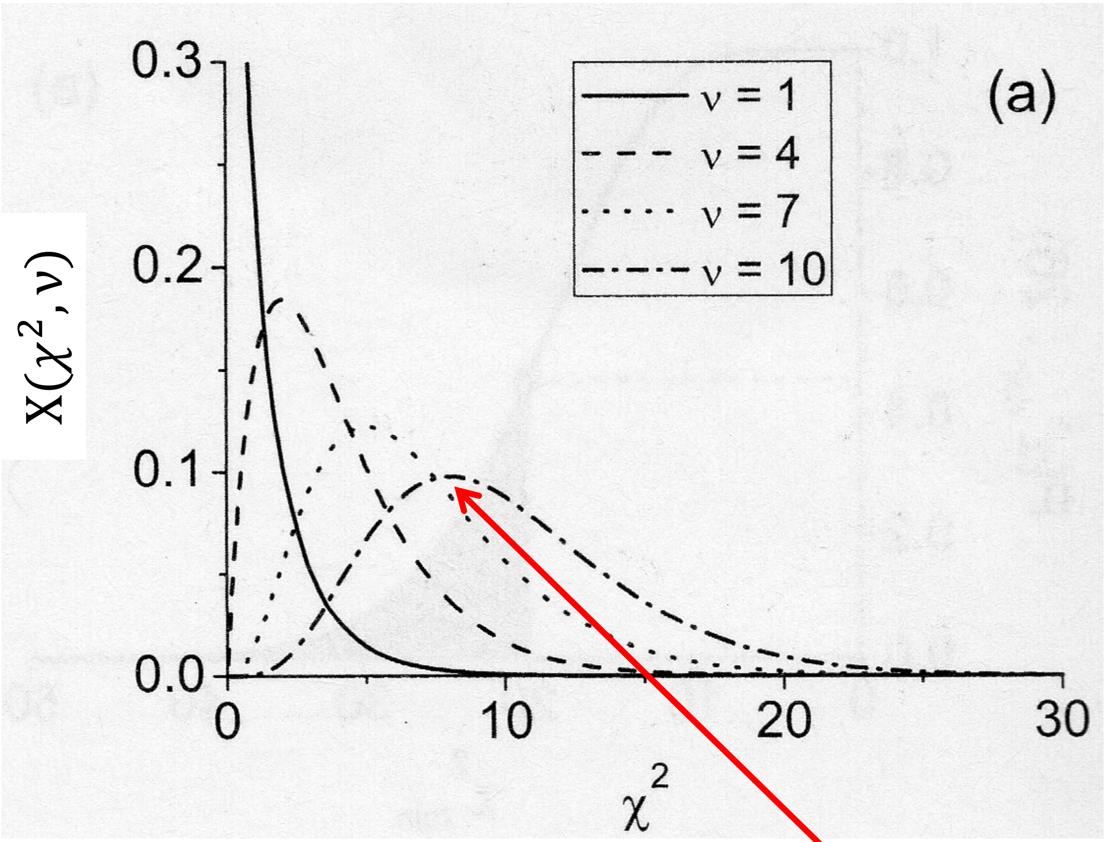
- Here there are 3 parameters we need to calculate the distribution (σ, X, A), which come from fitting the data.

- Hence our degrees of freedom (DoF) for the data set is 40 – 3 = 37.

- E.g. For fitting n data points to a straight line, $\nu = n - 2$ (fitting requires finding the intercept and a gradient).

The chi-squared probability distribution

Depends on the calculated χ^2 and the DoF (ν).



- The most probable value of χ^2 is the peak of $X(\chi^2, \nu)$, which depends on ν.
- For large ν the most probable value of χ^2 occurs at approximately $\nu - 2$.
- You can find tables of $X(\chi^2, \nu)$, but many software provide a function to calculate it.

χ^2

In Matlab: `X=chi2pdf(x,v)`

ν

The probability density of a given χ^2 value depends on ν.

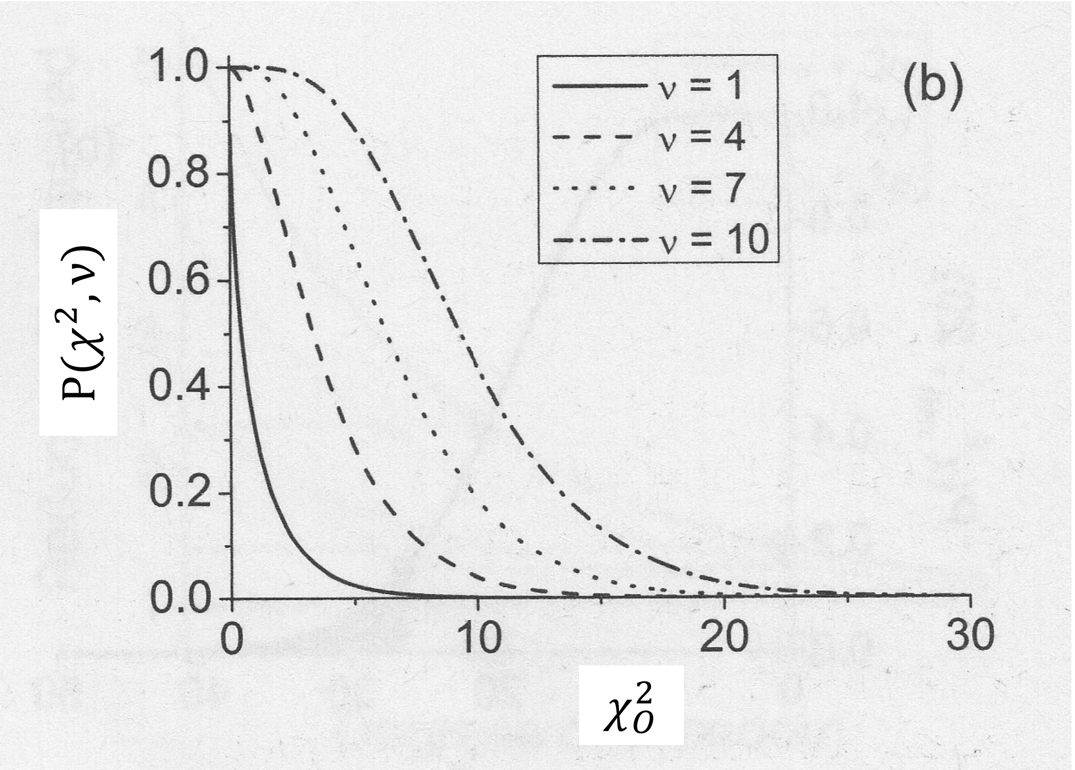
chi2pdf(8,10) = 0.0977

The chi-squared cumulative probability distribution

This can tell you how likely it is to exceed a given χ^2 value.

The P-value is the probability of that you should see a χ^2 larger than what you measured, χ_O^2

It is the integral of the χ^2 PDF from a given χ^2 to infinity.



For large ν: Expect $P \approx 0.5$

- This is because the PDF looks like a Gaussian at high ν and so is symmetric about the peak value.
- Often used in "hypothesis testing" where you rule out a model because the χ_O^2 is too large.
- Commonly use a P-value cut-off of 0.05 in hypothesis testing.

χ_o^2 calculates P greater than your χ_o^2

In Matlab: `P=chi2cdf(x,v,'upper')`

ν

An example:

If I have an experiment with 10 degrees of freedom (~ no. of data points) where I calculate a χ_o^2 of 10 . What is the probability that I'd expect a χ^2 value this big or greater?

`P=chi2cdf(10,10,'upper') = 0.4405` i.e. 44% (quite likely).

We expect a P value of ~ 0.5 when the deviations from the model are due to only random variation.

Say I observed a χ_o^2 of 20 for the same experiment. How likely is this due to chance?

`P=chi2cdf(20,10,'upper') = 0.0293` i.e. ~3 % (very unlikely).
It's likely the prediction model is wrong.

Chi-squared per degree of freedom

- Also known as the Reduced Chi-Squared metric.

As the most probable value of the χ^2distribution is approximately equal to the degrees of freedom (ν), a commonly used test is the reduced chi squared statistic:

$$\chi_\nu^2 = \frac{\chi_O^2}{\nu}.$$

- χ^2 for each data point is of order 1 and so you are likely to see $\chi^2 = n$ when the number of data points is large.
- A very small value for χ^2 is also unlikely and may suggest the error bars for each value have been underestimated.

For a reasonable fit expect to see $\chi_\nu^2 \approx 1$.

- If $\chi_\nu^2 \ll 1$ check your calculations for the uncertainties in the measurements.
- If $\chi_\nu^2 > 2$ for $\nu \approx 10$, or if $\chi_\nu^2 > 1.5$ for $\nu > 50$, then the fitting model is questionable.

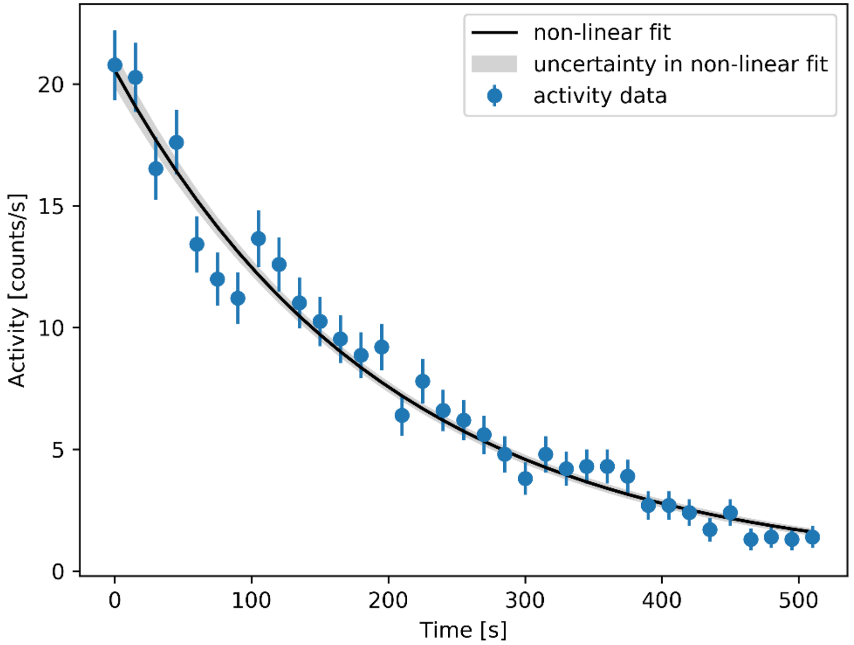
A reduced chi-squared test is often used (rather than calculating P-values) as it is easy to calculate.

Exercise 1

Example: Evaluating the quality of a fit

Consider a fit to exponential decay data.

Model: $y = a\,\exp(b\,x)$



- Calculate the reduced chi-square metric:

measured value at k

fit value at point k

$$\chi^2 = \sum_{k=1}^{n} \left(\frac{O_k - E_k}{\sigma_k} \right)^2$$

error bar at k

No. of data points: $n = 35$

Fitting equation has 2 parameters, therefore the degrees of freedom is:

$$\nu = n - 2 = 33$$

$\chi^2 = 35.6904$ $\qquad$ $\chi^2_\nu = \dfrac{\chi^2}{\nu} = 1.0815$

Conclusion:

- The model is a good fit.
- The χ^2 value suggest data deviates from the model in accordance with random variation.

Comparing different models for fit quality

We can use the reduced χ^2 metric to compare the fit quality of different models.

Example: Here we will compare several models for the period of a pendulum.

1. • The conventional model assumes only small angular deviations and has the result:

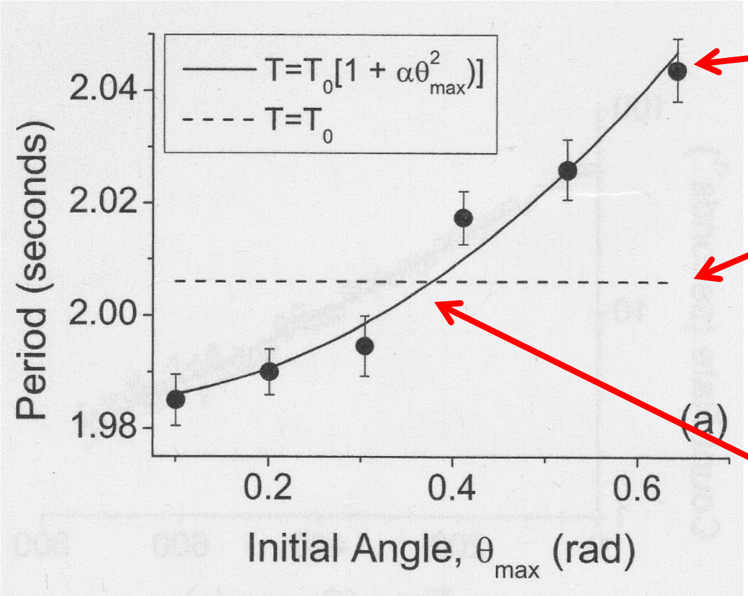
$$T_0 = 2\pi\sqrt{L/g}\,.$$

2. • The first-order corrected result for large initial angular displacement (θ_{max}) is:

$$T = 2\pi\sqrt{L/g}\left[1 + \frac{1}{16}\theta_{max}^2\right].$$

or $T = T_0\left[1 + \frac{1}{16}\theta_{max}^2\right]$.

3. • Let's also consider another model (not supported by theory): $T = T_0[1 + \alpha\theta_{\mathrm{max}}]$.



Data (n=6 data points)

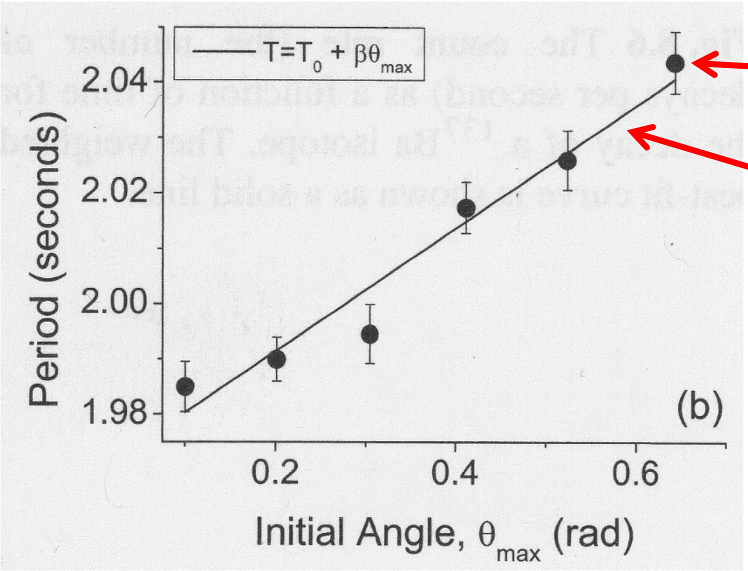
1. "fit" to $\boldsymbol{T = T_0}$

$$T = 2\pi\sqrt{L/g} = T_0$$

- Here the fit parameter is $\left(L/g\right)$.
- Hence, the DoF = (n – 1) = 5.

2. fit to $\boldsymbol{T = T_0\left[1 + \frac{1}{16}\theta_{max}^2\right]}$

- the DoF = (n – 2) = 4.



Data (n=6 data points)

3. fit is to $\boldsymbol{T = T_0[1 + \alpha\theta_{max}]}$

- the DoF = (n – 2) = 4.

Plots of the data and the fit to each of the three models.

Check the χ^2 metrics:

expected reduced chi-squared ~1

calculated for each model

expect P ~ 0.5

the most probable P value from the χ^2 distribution function

Model	ν	χ_O^2	χ_ν^2	P	$\chi^2_{most\ prob}$
$T = T_0$	5	107.2	21.2	1.6×10^{-21}	2.99
* $T = T_0\left[1 + \frac{1}{16}\theta_{max}^2\right]$	4	3.39	0.9	0.49	1.99
$T = T_0[1 + \alpha\theta_{\max}]$	4	4.39	1.1	0.36	1.99

can exclude this model

* This model gives the best metrics.

- Reduced chi-squared is close to 1
- P value is closest to 0.5
- The χ_0^2 is closest to the most probable value $\chi^2_{most\ prob}$

Occam's Razor

By adding more parameters to a fitting model a better fit to the data can be found.

Occam's Razor: (paraphrase)

When you have two competing theories that make exactly the same predictions, the simpler one is the better one to use.

Note: this is not rigorously defensible.

- The idea is that the simplest model that can explain observed behaviour is more likely to be predictive when new data is examined.
- "Overfitted" models are likely to diverge when data outside the original fitted range is provided.

Exercise 2

Testing distributions

Probably the most common use of chi-squared testing is showing that a distribution is **not** due to chance.

Common instances:

- pharmaceutical testing
- health sciences

Here, you test your observed distribution against the "null hypothesis" for rejection.

Null hypothesis: The distribution of values you see is just due to chance.

- Often the criterion chosen is "to reject the null hypothesis at the 5% level".
- Sometimes stated as "reject at $P < 0.05$".

- This means that if you find a P value less than 0.05, then you confirm there is a non-random effect occurring.
- Essentially, there is less than a 5% chance that your measurements are just due to randomness.

E.g. for testing a new drug, this means that it actually does something (probably)!

Practical testing of distributions

Here you compare your observations (the sample distribution) to a theoretical model (the parent distribution).

- This means you need to generate a theoretical histogram with the same binning as your results histogram.

Important considerations:

1.
   - Need to ensure that you have > 5 counts in every bin in your theoretical histogram.
   - Combine bins to achieve this (uneven binning).

2.
   - Only practical to test distributions with >20 total counts.

This is because a (typical) Gaussian distribution has $\nu = 3$ and so you need at least 4 bins (with a minimum of 5 counts) to avoid “overfitting”.

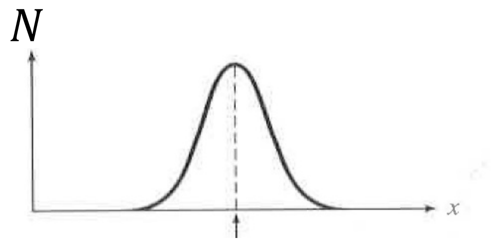
An example: Say I have an experiment firing a canon and I measure the range of the projectile (in m) 40 times. I expect that the variation is predicted by a Gaussian.
Is this a good model of my results?

x =

731	780	689	830	688	672	753	787
739	748	754	722	748	764	638	742
678	770	681	760	778	738	733	645
698	771	676	805	710	757	766	675
772	709	810	725	653	687	709	712

n=40

Step 1. Calculate the Gaussian distribution that fits the data:



X (most probable distance)

$$N = n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-X)^2}{2\sigma^2}}.$$

centre (X)
- given by the mean of the data.
- X = 730.08

width (σ)
- given by the standard deviation (n-1).
- σ = 46.77

Step 2. Calculate the theoretical distribution function

Here, we only have 40 counts, so lets choose the minimum of 4 bins.

Choose the bins so that at least 5 counts are expected in each bin.

As $\sigma \approx 48$ and the centre is at ≈730 , lets choose the bins as:

600-682 , 682-730, 730-778 , 778-900

note the bins aren't equal widths.

Calculate the expected values in each of these bins:

In Matlab, create the Gaussian function using the fit parameters:

make this a function of x only

use the . as x might be a vector

```
g=@(x) 40*1/(46.77*sqrt(2*pi))*exp(-(x-730.08).^2/(2*46.77^2))
```

Then numerically integrate between the bin limits to get the counts in each bin:

```
integral(g,600,682) = 5.9706
integral(g,682,730) = 13.8938
integral(g,730,778) = 13.9161
integral(g,778,900) = 6.1055
```

- Check that each bin has a count greater than 5.

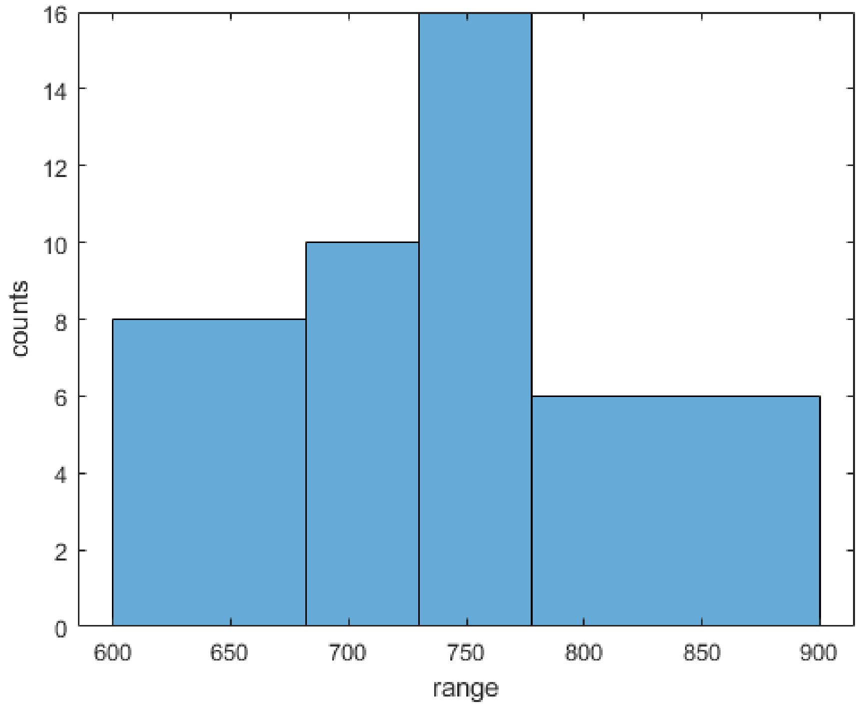
☑

Step 3. Make a histogram of the measured data

```
histogram(x,[600,682,730,778,900])
```

Matlab uses the bin edges to make the histogram

Bin	Counts
600-682	8
682-730	10
730-778	16
778-900	6



Step 4. Calculate the χ^2

As we are comparing counts in bins, we use the following χ^2 calculation:

$$\chi^2 = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k}$$

Bin	Observed counts (O_k)	Expected counts (E_k)	$\frac{(O_k - E_k)^2}{E_k}$
600-682	8	5.97	0.690
682-730	10	13.89	1.091
730-778	16	13.92	0.312
778-900	6	6.11	0.002

χ^2= 2.09

Step 5. Check the comparison metrics

Calculate the degrees of freedom:

- Here we used 3 degrees of freedom in calculating the expected distribution (n, σ, X).
- The comparison is made between the 4 bin values, so: $\nu = 1$

Reduced chi-squared value:

Consequently the reduced chi-squared value is: $\chi^2_\nu = 2.09$

As $\chi^2_\nu \approx 1$ it's possible that the range data conforms to a Gaussian distribution.

P-value:

$P = 0.148$

- Not very close to 0.5
- Suggests that there is ~15% chance that the deviations from a Gaussian are due to chance.

As $\nu = 1$, more data points and more comparison bins are recommended!

Exercise 3