# IPM component 3

# Week 7: Evaluating arguments – Claims and supporting evidence

## Introduction

In these notes we look at how we can identify claims and supporting evidence made by STEM texts. Remember that

- a claim is a question for which I have an answer or solution,

and

- evidence is something which represents proof that my claim is true or valid.

In order to identify that any evidence supports a given claim one has to be familiar with the topic concerned. The only topics I am sufficiently familiar with is maths hence I can provide a more accurate assessment of evidence in support of claims for this topic only. For this reason, of the five examples below four of them are maths/stats related.

*A mathematics example*

In pure mathematics there is no such thing as a claim and evidence. Instead, we have theorems and proofs, where the proof might be considered as evidence in support of the claim. As an example of this consider proving that $\sqrt{2}$ is an irrational number.

> *Theorem*: The number $\sqrt{2}$ is irrational.
>
> *Proof:* We use a proof by contradiction. So, assume that $\sqrt{2}$ can be expressed as a rational in lowest terms. Hence, we can write $\sqrt{2} = p/q$ where $p$ and $q$ are integers and are co-prime. Then
>
> $$2 = \frac{p^2}{q^2} \quad \Longrightarrow \quad p^2 = 2q^2$$
>
> Therefore $p^2$ is even which implies $p$ is even. But this contradicts our assumption that $p/q$ is in its lowest terms. Another way of looking at this is that $p$ has to be an odd number for $p/q$ to be in lowest terms. But we have just found that $p^2$, and hence $p$, is even hence a contradiction. Therefore, $\sqrt{2}$ cannot be written as $p/q$.

*Analysis*

Here the claim is that $\sqrt{2}$ is an irrational number. The evidence justifying this claim is the proof, and the particular form of proof used (namely, proof by contradiction). In mathematics proofs can be seen as the argumentation which justifies our claim. How does the proof support the claim that $\sqrt{2}$ is irrational? By the fact that we assume the opposite of what we want to prove and then obtain a result which contradicts our assumption. In this case we assume that we can write $\sqrt{2}$ as a rational fraction $p/q$ which cannot be simplified any further. We then perform some algebra to find that our fraction can be simplified hence the contradiction, implying that our original assumption was false which now implies that $\sqrt{2}$ is irrational.

*A text from a maths, statistics and informatics journal*

Consider the text below taken from "Air passengers forecasting for Australian airline based on hybrid rough set approach", H. KUMAR SHARMA, K. KUMARI AND S. KAR, *Journal of Applied Mathematics, Statistics and Informatics*, Vol 14 (2018), No 1)

> "Accurate and reliable air passenger demand is very important for policy-making and planning by tourism management as well as by airline authorities. Therefore, this article proposed a novel hybrid method based on rough set theory (RST) to construct decision rules for long-term forecasting of air passengers. Empirical results show that the proposed method is highly accurate with the higher corrected classified accuracy. Also, forecasting accuracy of the proposed method is better than the other time series approaches.

*Analysis*

The problem with this paragraph of this text is that it is linguistically badly constructed:

- Firstly, it is not "accurate and reliable air passenger demand that needs to be considered for policy-making and planning". Rather it is accurate and reliable *information about* air passenger demand that needs to be considered for policy-making and planning.
- Secondly, the word "Therefore" at the end of line 2 is not appropriate for the continuity I believe the authors' are trying to make from the first sentence. A more appropriate form of words would be "to achieve this accuracy and reliability".
- Thirdly, in lines 3-4 we have "to construct decision rules for long-term forecasting of air passengers". This does not sound right. One does not "forecast air passengers". The use of the word "forecast" means you are trying to calculate an amount. So, one forecasts air passenger numbers. This makes much more sense.

An improved version of the first paragraph might read as follows:

> "Accurate and reliable information about air passenger demand is very important for policy-making and planning by tourism management as well as by airline authorities. To achieve this accuracy and reliability this article proposed a novel hybrid method based on rough set theory (RST) to constructs decision rules for long-term forecasting of the number of air passengers. Empirical results show that the proposed method is highly accurate with the higher corrected classified accuracy. Also, forecasting accuracy of the proposed method is better than the other time series approaches.

We might then say that the claim the authors are making is that the "the proposed method is highly accurate", and that the "forecasting accuracy of the [novel hybrid] method is better than the other time series approaches". The evidence for this claim comes from the authors' results. The authors use the statement "Empirical results show …" for this reason. A detailed reading of the paper (particularly any methodology, results and analysis, and discussion section) would be needed in order to understand what they mean by "better".

*<u>A Chemistry text</u>*

Consider the text below taken from "Geometry, Energy, and Some Electronic Properties of Carbon Polyprismanes: *Ab Initio* and Tight-Binding Study", Konstantin P. Katin, et. al., *Advances in Physical Chemistry*, Volume 2015

> We report geometry, energy, and some electronic properties of [n,4]- and [n,5]prismanes (polyprismanes): a special type of carbon nanotubes constructed from dehydrogenated cycloalkane $C_4$ - and $C_5$ -rings, respectively. Binding energies, interatomic bonds, and the energy gaps between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) have been calculated using density functional approach and nonorthogonal tight-binding model for the systems up to thirty layers. It is found that polyprismanes become more thermodynamically stable as their effective length increases. Moreover, they may possess semiconducting properties in the bulk limit.
>
> […] The data of numerical simulation obtained in this study indicate an increase of thermodynamic stability of PP as the number of layers $n$ increases. Binding energies estimated for both "infinite" tetra- and pentaprismanes are equal to 4.9 eV/atom. In

contrast to binding energies, HOMO-LUMO gaps decrease as the effective length of PP increases. In the bulk limit ($n\to\infty$) ΔHL is equal to 2.4 and 3.2 eV for [n,4]- and [n,5]prismanes, respectively. These values are near to the upper limit of characteristic semiconducting value; therefore, it becomes possible to tune electronic properties of the PP by doping or mechanical stresses that will be useful for nanoelectronic applications.

*Analysis*

Lines 1-3 simply describes the particular substance the authors will be investigating.

Lines 3-7 form one complete sentence. In my opinion this sentence is too long. A more readable way of writing this could be

> The following properties of the [n,4]- and [n,5]prismanes were calculated using density functional approach and nonorthogonal tight-binding model:
>
> - binding energies,
> - interatomic bonds,
> - the energy gaps between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).
>
> Systems up to thirty layers were considered.

Lines 7-9 represent claim statement. In fact, there are two claims here. The first one is seen by the language "it was found that … become more … as <*something happens*>". The second claim can be seen by the language "they may possess … properties".

Lines 11-19 represent the evidence in support of the claim (this part of the text comes in the conclusion of the paper). Even though I do not understand anything about the chemistry the authors are referring to I can still tell that this paragraph refers to evidence. This is because of the language and sentence structure used. This can be seen more easily if we grey out the scientific part of the text to leave only the language and structure of the sentences:

> The data of numerical simulation obtained in this study indicate an increase of thermodynamic stability of PP as the number of layers $n$ increases. Binding energies estimated for both "infinite" tetra- and pentaprismanes are equal to 4.9 eV/atom. In contrast to binding energies, HOMO-LUMO gaps decrease as the effective length of PP increases. In the bulk limit ($n\to\infty$) ΔHL is equal to 2.4 and 3.2 eV for [n,4]- and

[n,5]prismanes, respectively. These values are near to the upper limit of characteristic semiconducting value; therefore, it becomes possible to tune electronic properties of the PP by doping or mechanical stresses that will be useful for nanoelectronic applications.

*A statistics text: In-class exercise*

Identify the claim and evidence in the following text taken from "The Application of the Chi-Squared Test", R. L. Plackett, *The Mathematical Gazette*, Dec., 1971, Vol. 55, No. 394 (Dec., 1971), pp. 363-366.

"Suppose that we open a telephone directory and select any column which consists entirely of personal names. The last digit of each telephone number would seem to have no preferred value. For example, the middle column on p. 417 of the 1970 Newcastle upon Tyne directory contains 109 names, and the frequencies for 0, 1,..., 9 are as follows.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 11 | 10 | 8 | 11 | 10 | 11 | 11 | 16 | 13 | 8 |

The idea that there is no preferred digit can be formed into a statistical hypothesis that the probability for each digit is 1/10. We require a method of testing this hypothesis.

[…] If the probability for each digit is 1/10, then the frequency for each digit should average 10.9. We call this an expected frequency. Calculate

$$\mathrm{X}^2 = \frac{\sum(observed - expected)^2}{expected}$$

and refer the value to a table of $\chi^2$. In our example, the appropriate number of degrees of freedom is equal to one less than the number of frequencies, and large values of $\mathrm{X}^2$ are regarded as significant.

[…] The value of $\mathrm{X}^2$ for the data given is 4.49. Larger values of $\chi^2$ with 9 degrees of freedom occur with a probability of 0.87. This suggests that the hypothesis of equi-probability should be accepted. Our intuitive feeling about final digits is confirmed.

*A mathematical modelling text: In-class exercise*

Identify the claim and evidence in the text below which is adapted from “Risk-Taking in Critical Path Analysis”, William S. Jewell, *Management Science*, Jan., 1965, Vol. 11, No. 3, Series A, Sciences (Jan., 1965), pp. 438-443.

“In the simplest model of critical path analysis, one is given the known duration of each activity in a large project and a network which describes the precedence relationships among the activities. When the exact duration of each activity is not known in advance, the problem becomes more complicated. Early proponents of the PERT model proposed a "wait-and-see" type of problem, in which it is assumed that when the events of the network are scheduled, the actual durations of the activities will be available.

The problem is to find the distribution of the minimal project completion time, G(T). The search for the distribution G(T) seems, to this author, to be somewhat irrelevant to the actual project planning problem, since the uncertainties of the activity durations are rarely resolved when the project events are first scheduled. Instead, there is usually some element of risk still present. One approach to solving this problem is through chance constrained programming, in which various "risk guarantees" (probabilities of overrun) are assumed for each activity.

The approach we shall use in this paper is that of a "here-and-now" programming problem (see, for example, [4]). We shall assume that the planner must set up a fixed project schedule in the face of uncertain activity durations which become known only when the activity gets underway. Based upon the difference between the allotted time interval, and the actual "free" time needed by the job, corrective action may have to be taken to stay within the fixed schedule; the problem is to schedule the project so as to minimize the expected amount of extra effort expended to stay on schedule. This approach leads to a formulation which is similar to those in the cost-time models [5], [7] of critical path analysis.”

*A geology text: In-class exercise*

Consider the text below, which is adapted from “Studying the Deep Structure of Elbrus Volcano by Microseismic Sounding”, D. V. Likhodeev, et al., *Journal of Volcanology and Seismology*, 2017, Vol. 11, No. 6, pp. 413–418.

"This paper presents results from a study of the deep structure of the Elbrus edifice and adjacent areas using geophysical techniques. We confirmed the existence of a shallow magma chamber, derived a more accurate location of the chamber in the host rocks and its characteristic dimensions, and compared new results with known studies. More accurate estimates have been obtained for the temperature at the top of the magma chamber and new evidence is adduced concerning the deep structure of the fluid magmatic systems in the Elbrus volcanic area."

1) Is a claim being made in this text? If so, what is the claim?
2) Do the authors say that they have any evidence? If so, what evidence do they say they have?

*A geology text: In-class exercise*

Given the text below can you identify the claim and the evidence which justifies the claim? The text is adapted from "The Relationship between Glaciation and Volcanism: Numerical Simulation and the Holocene Volcanism in Kamchatka", A. G. Simakin, and Ya. D. Murav'ev, *Journal of Volcanology and Seismology*, 2017, Vol. 11, No. 3, pp. 187–205.

"This study is concerned with numerical simulation of the strain due to glaciation and glacial melting, when a magma zone (a layer containing inclusions of magma and magma cumulates) is present at the crust–mantle boundary. According to analytical solutions of this problem that involves viscous relaxation of an uncompensated depression at the place of the molten glacier, the depth to the zone of increased shear stresses beneath the depression is proportional to its width, while the relaxation duration is proportional to viscosity of the lithosphere and is a few thousand years. These fundamental estimates are corroborated by our numerical simulation. According to it, the magma zone at the Moho boundary shields the zone of increased shear stresses, limiting it from below. The maximum values (12–25 MPa) with glacial thickness 500–1000 m are reached at the top of this layer of low viscosity. The directions of maximum compression ($s_1$ ) as calculated for the time after the melting indicate that the magma that rises along dikes is displaced from the center of the magma lens toward its periphery. It is found that glacial unloading makes the dipping faults in the crust above the low-viscosity layer attractors for the rising magma. Glacial unloading accelerates, by factors of a few times, the magma generation in the mantle that occurs following the mechanism of adiabatic decompression, as well as

> facilitating the accumulation of mantle fluids in the zone of increased shear stresses at the boundary of the low viscosity layer. The magma traverses this deep fluid collector and increases the intensity and explosivity of eruptions at the beginning of an interglacial period. Our numerical simulation results are in general agreement with published data on Early Holocene volcanic eruptions that occurred after the second phase of the Late Pleistocene glaciation in Kamchatka.”

**The language and discourse of claims and evidence**

The examples above on claims and evidence involved certain types of vocabulary, phrasing and sentence building. The way in which this vocabulary and phrasing can be built is illustrated in the table on the next page. The aim of this table is to show you examples on an *underlying principle* of what constitutes *claims and evidence* language. This underlying principle is what you should aim to learn and understand. Then you will know how to write, in your own words, a claim and present evidence in support of that claim.

For example, consider the following, adapted from the mathematical modelling text above:

- Early proponents of the PERT adopted a "wait-and-see" model for the situation when the exact duration of each activity was not known in advance. The approach we shall use in this paper is that of a "here-and-now" model. We shall assume that the planner must set up a fixed project schedule in the face of uncertain activity durations, these becoming known only when the activity gets underway. This approach leads to a formulation which is similar to those in the cost-time models [5], [7] of critical path analysis.”

We can rewrite this as illustrated below, whilst still retaining the features and essence of *claims and evidence* type of writing:

- In this paper we propose an alternative model to the “wait-and-see” model adopted by early proponents of the PERT. This latter model was used in the situation when the exact duration of each activity of the critical path process was not known in advance. Our new “here-and-now” model, which leads to a formulation which is similar to those in the cost-time models [5], [7] of critical path analysis, assumes that planner sets up a fixed project schedule in the face of uncertain activity durations, these becoming known only when the activity gets underway.

Both paragraphs address the same concept to the same degree of detail, and state the same claim.

*The language and discourse of claims and evidence*

These results { support the idea of … / confirm the association between … / are consistent with data obtained in … / are in line with those of previous studies in … / … }

This study confirms { that X is associated with … / finding reported by Smith et al. (1989). / the increase in efficiency of … / the assumption that there is a correlation between … / … }

The test was successful as it was able to identify students who …

The present results are significant in at least two major respects.

There is significant improvement on the typical values reported in …

These values are a significant improvement
on the typical values reported in …

Such an approach was found to produce an optimal solution more
quickly than the general methods reported in …

The results show/demonstrate that …

**So what is a claim and what is evidence?**

By looking at as many examples as possible we can come to see that, in applied mathematics, engineering and science, claims are always supported by evidence. In general one

- has an idea that a particular phenomenon is true, whereupon we say that this is true. This truth statement is a claim statement, written in "claim statement language", about the truth of the phenomenon we are investigating;

- demonstrates evidence of the truth of the claim by having results, and analysis to explain these results. We then use “evidence phrasing” to demonstrate that what we are talking about is indeed evidence which supports our claim.

For example,

1. in mathematics the claim is the *theorem* (a statement that an equation or proposition is true), and the evidence is the *proof* (the steps of algebra or logical inference used to show the statement as true);
2. in statistics, the technique of hypothesis testing leads to a standard claim:
   - $H_0$ (the null hypothesis): There is no significant difference in/between …
   - $H_1$ (the alternative hypothesis): There is a significant difference in/between …

   In statistics we are looking for our experiment to produce a difference, so we would like $H_1$ to be confirmed. The evidence for or against $H_1$ is determined by the $p$ value (a probability value based on some calculation) for the given distribution we are using (say the normal distribution). If the $p$ value is greater than a known critical value one might think that $H_1$ is confirmed. However, statistics never deals with definite outcomes. Because of the probabilistic nature of statistics and the degrees of uncertainties involved in real data collection and analysis across different disciplines, we tend to speak in terms of there being (strong) evidence to reject accept $H_0$. This allows for the possibility that, under different conditions, $H_0$ might later be accepted. In general, and depending on the discipline (ecology versus engineering), statistical evidence may not always be clear cut.

   For example, if an engineering company manufactures 10000 nuts and bolts of length 20mm and diameter 5 mm respectively the company might want to make sure that these are produced to within a tolerance of 0.1mm (i.e. 2mm±0.1mm and 5mm±0.1mm). Since the machines which make the nuts and bolts are high quality machines which are highly calibrated there will be very little variability in the dimensions of the nuts and bolts over a given time period. The machine produces all nuts and bolts to within this tolerance because machines are inanimate objects which produces little variation in their outputs (however the more complex the manufacturing process the more prone the process is to error). Therefore, a statistically significant result can be more reliable in this context.

   However, if your discipline is zoology or ecology, and your experiment is to measure the wing span or echo frequency of a particular type of bat any statistically significant result could be less reliable in this context. There could be all sorts of problems relating to data collection, environmental factors or other which would affect conclusion made on the basis of your experiments because of the inherent (and natural) variability of living organisms.

3. in operational research the tool of critical path analysis (CPA) is used to optimise time or costs on a project (such as a construction project). The claim is that the method of critical path analysis minimises time or costs, and the evidence is the final CPA network (and the mathematics of optimisation which lies behind the CPA method) actually achieves this aim. Of course, the real evidence is in the fact that the same project would have taken longer or cost more had it been organised in any other manner.
4. In chemistry the general types of claims are __________ (fill in the gap) and the general types of evidence are __________ (fill in the gap).
5. In big data / data analytics the general types of claims are __________ (fill in the gap) and the general types of evidence are __________ (fill in the gap).
6. In your discipline of engineering the general types of claims are __________ (fill in the gap) and the general types of evidence are __________ (fill in the gap).

So, we see claims when we make categorical statements that "this or that" is true or valid, and we see evidence when we present the results of us testing our claim via some form of experiment, model, plan, etc. as confirmation of such claims. Therefore,

**Claim = An assertion or declaration of something as being true or valid**

(for which evidence is needed in order to confirm the claim).

What kind of aspect classify as evidence? Some are listed below.

1. *Appropriate use of methods/methodologies in your discipline*. Are the methods, experiments, techniques, methodologies, etc., sufficient to collect the data which would act as evidence for your claim?
    - For pure mathematics there are no data collection. In actual fact, the "data" here are numbers and functions and higher order mathematical objects. But all things track back to numbers, and these are mathematically and logically constructed rather than collected via experiment. As for "data analysis" this consists of algebra and other logical rules of inference so as to construct proofs or solve problems.

    - In statistics it is the use of appropriate statistical tests, significance testing (and calculated *p*-values compared to the theoretical critical *p*-value), confidence intervals, least square analysis, etc., with all the assumptions these entail.
    - In operational research it is the use of appropriate modelling methods (networks, critical path analysis, simplex method, etc…);
    - In the physical sciences it is the use of appropriate equipment, experimental methodologies and experiments.
2. *Positive effects*: Improved efficiency, reduced cost, greater accuracy, an increase or decrease in something, the ability to decide if a treatment has an effect (i.e. hypothesis testing in statistics), a strong correlation between two things, an optimal solution (simplex method or critical path analysis of shortest route problems, etc.);
3. *Reliability***:** Can the activity or experiment be replicated? Does it give the same (or, within an accepted margin of error, similar) results every time?
4. *Comprehensiveness***:** Have you and/or other people taken account, as much as possible, of the following
    - Have you collected as much data as possible, and as is reasonable, in order to validate the use of a particular method or methodology?
    - Is your data relevant? For example when the Hubble space telescope takes pictures of galaxies and stars it transmits this information back to Earth as an electronic signal. But this signal contains a lot of "stuff" that isn't the image of the galaxy or star. I.e. the signal contains "noise". This noise is not relevant and has to be separated from the information which relates to the image of the galaxy or star.
    - Have you used one or more methods, and done so appropriately, in order to come to a sufficiently compelling conclusion?
    - Have you adopted a critical mind to the results of your analysis, highlighting assumptions, limitations, negative results, confounding influences, etc. where relevant?

Ultimately, we can also say that evidence is data which has been analysed by some methods or methodology, the results of which are interpreted in a meaningful way. It is generally said that data and evidence are the same thing. However, a distinction can be made between them to the extent that we can say that data is not evidence. Data is analysed. The results/interpretation of this analysis then provide evidence in favour of, or against, a claim. Another way of saying this is that data is raw information with no judgement or opinion attached, whereas as evidence is about showing that the data proves or disproves a claim.

For example, we can collect data on 10000 people (such as height, weight, blood pressure, etc). This is simply data. This isn't evidence of anything (except that these people have these heights, weights, blood pressures, etc.). But if we now conduct a particular statistical test on this data (t-test, ANOVA, least squares analysis, significance test, etc.) then we are analysing the data in a particular way in order to determine whether or not such-and-such a hypothesis about this group is true. If the stats test produces a confirming result then we have *evidence* in favour of our hypothesis based on the initial *raw data*. Furthermore, evidence is usually always evidence in favour of a claim, or evidence which justifies a claim.

So we can also say that

**Evidence = The interpretation of data (based on our analysis)**
**which supports/confirms our claim.**
(this data being derived from the use of methods and methodologies
deemed sufficiently appropriate to be able to confirm the claim.)

As an aside the following is a definition of rationality as applied to claims and evidence

Claims and evidence are rational if we have
a systematic, logical, and coherent development of an idea (i.e. a theory)
which can be tested (i.e. via experimentation)
and which allows for categorical statements (i.e. claims)
so as to obtain relevant data which can be properly analysed (via methodologies and methods),
this producing the evidence required to justify the claims.

**Exercises**

1) Analyse (as best as possible) the statistics text 2 and the mathematical modelling text for claims and supporting evidence.

2) a) Find a paper from your own discipline and find its claim(s) and supporting evidence.

   b) Assuming you understand the technicalities of your chosen paper describe in more detail how the evidence justifies the author's claim. Is it related to improved/changed results, or to outcomes or reliability or consistency or effectiveness or efficiency or productivity or competence, etc.?

   Also, explain in plain English the methodology and analysis which justifies their claim (as I did on p1/2 for the mathematics example where I specified the type of proof used and how we arrived at a contradiction).