Informatics 1: Data & Analysis

Lecture 20: Course Review

Ian Stark

School of Informatics
The University of Edinburgh

Friday 25 March 2016
Semester 2 Week 10

THE UNIVERSITY *of* EDINBURGH

Plan !

This is Teaching Week 10 of Semester 2, next week is Week 11, and the teaching block ends on Friday 1 April.

Your final Inf1-DA tutorial is next week, in which your tutor will return your coursework assignment with marks and feedback.

But only if you go to the correct tutorial group!

Find the course web page, Tutorial tab, "group membership" link.

The rest of this lecture includes: a review of exam arrangements; advice on preparation and practice; and a summary of topics covered in the course.

The final lecture, on Tuesday, will review specific past exam questions.

Exam: Time and Place !

Informatics 1: Data & Analysis will be assessed by a single two-hour written examination.

Date: Monday 2 May 2016
Time: 0930–1130
Place: The Pleasance Sports Hall *and*
Dance Studio (Gym 4) St Leonard's Land
(Allocation of students to venues still to be published)

This information for course code INFR08015 is current at 2016-03-24; please check the link on the Inf1-DA web page nearer to the date to check this and to confirm all of your exams.

Exam: Format

As in previous years, the exam will have three compulsory questions.

- Read all questions before beginning the paper
- You don't need to do the questions in order
- Don't assume a question is only using one part of the course
- If you get stuck on one question: don't waste too much time on it; do go on to the next question; come back later. Don't give up!

All questions include information about marks available for each section.

Adding up the marks you achieve for each question gives your total mark for the exam, and for Inf1-DA overall.

Exam: Grades

The Inf1-DA exam is graded on the standard University of Edinburgh scale:

Mark	Grade	Description
70–100	A	Excellent
60–69	B	Very Good
50–59	C	Good
40–49	D	Satisfactory
30–39	E	Marginal fail
20–29	F	Clear fail
10–19	G	Bad fail
0–9	H	Bad fail

Students who do not pass at this sitting may resit in August.

Passing this course is essential for progression to second-year study.

All passing marks and grades are recorded on your transcript. However, final degree grade is based only on performance in Year 3 onwards.

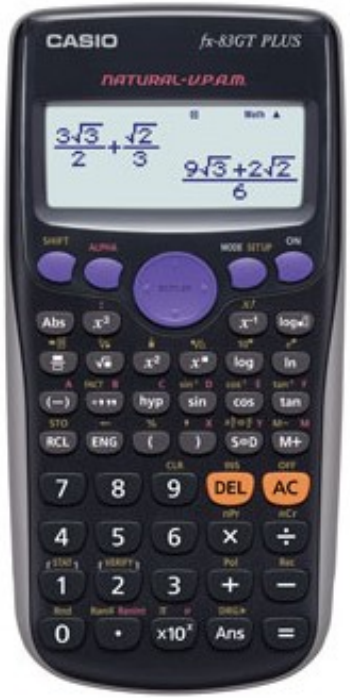
Exam: Calculators

You may use a scientific calculator in the Inf1-DA exam, and I recommend that you do take one with you.

There is a standard list of permitted calculators, linked from the Inf1-DA web pages.

Most of these calculators perform some algebra, but they are not programmable and do not draw graphs.

If you need a different type of calculator (or other assistive technology) then please do get in touch with either me, your personal tutor, the student disability service, or the ITO.



"The FX-83GT Plus is the UK's #1 Scientific Calculator allowed in every UK exam where a calculator can be used."
— Casio UK

No other electronic devices

You may take in a calculator from the authorised list, but no other devices.

Exam Hall Regulations

...

10. The use of mobile devices/personal electronic equipment is not permitted. Mobile devices must be switched off during an examination. These should be placed in your bag and should not be on your person. Mobile devices are those which store/display data or connect to the internet, such as a mobile telephone, smart watches, smart glasses or any other communications equipment.

...

Take care with this: make sure you do not accidentally carry in any connected device.

Exam: Closed Book

The Inf1-DA exam is *closed book* — you may not take in any books, notes, or other reference material into the examination hall.

If you wish to take an English-language dictionary into the exam then please contact the ITO to arrange for permission.

There is no formula sheet or book provided. However, where necessary questions will include tables of critical values or other essential material.

Past Exam Papers

Past Questions

All tutorial exercises have Examples and Solutions from past exam papers.

These are chosen to match each course topic. Solutions include comments about different possible answers, and notes on important things to include.

Past Papers

The University Library keeps a full set of past papers online.

http://exampapers.ed.ac.uk

For Inf1-DA, the ITO web pages also have sets of sample solutions.

In each year there are two Inf1-DA exams, from the main and resit diets.

All questions since the 2009 exam are appropriate for the current syllabus.

Using Past Exam Papers

Past papers are a good source of practice material, and I strongly recommend you attempt as many of these questions as you can.

However, please note the following.

- The key exam preparation task is to *master the course material:* understand it and be able to apply it in practice.

  Conveniently, that's also the best possible outcome for you from this course

- Memorizing the answers to previous questions is not a helpful way to tackle future ones.

- Attempting past exam questions yourself, to test your knowledge, and identify any gaps, is a much more effective way to learn.

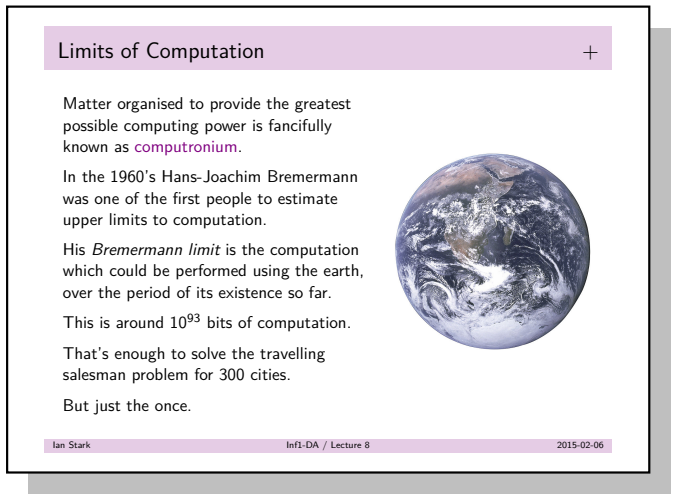
If you are puzzled by a past question or the solution provided, ask on *Piazza* or email me directly.

Examinable Material

Unless otherwise specified, all of the following material is examinable:

- Topics covered in lectures
- Directed reading distributed in lectures
- Topics covered in the weekly exercise sheets

All non-examinable additional material in lectures is marked by being on purple-shaded slides with a "+" in the top right-hand corner.


Limits of Computation +

Matter organised to provide the greatest possible computing power is fancifully known as computronium.

In the 1960's Hans-Joachim Bremermann was one of the first people to estimate upper limits to computation.

His *Bremermann limit* is the computation which could be performed using the earth, over the period of its existence so far.

This is around 10^{93} bits of computation.

That's enough to solve the travelling salesman problem for 300 cities.

But just the once.

Ian Stark Inf1-DA / Lecture 8 2015-02-06

Burning Questions

- Will this be on the exam?
- Is this examinable?
- Do I need to know this?
- Is it important?

Topic Summary

The entity-relationship model, ER diagrams. The relational model, SQL DDL. Translating an ER model into a relational one. Relational algebra, tuple-relational calculus, SQL queries; translating between all three.

Semistructured data models and the XPath data tree. XML documents. Schema languages and DTDs. Relational data converted into XML. XPath as a query language.

Corpora: what they are and how they are made; examples. Annotations and tagging. Concordances, frequencies, n-grams, collocations. Methods for machine translation.

Information retrieval: what it is, evaluating and comparing performance of IR systems; the vector space model and cosine similarity measure.

Data scales, summary statistics, population vs. sample; hypothesis testing and significance; correlation coefficient, χ^2 test.

Some Specific Items

Corpora

In general it is the principles of corpora that are examinable, rather then the precise details of individual corpora. Similarly, you should be familiar with the principles underlying POS-tagging and syntactic annotation, but you do not need to know detailed linguistics or specific tag sets.

You should however, be able to give examples of a corpus or a POS tag.

The CQP tool was used in a tutorial, so is examinable — although again for general principles and use, not every detail of syntax.

Statistics

You are not expected to memorize critical value tables; however, you should be able to use one if provided.

You are expected to know the formulas for the various statistics used, and to be able to calculate with them.

XML and XPath

XML trees Nodes: root, internal, leaves. Parents, children, ancestors, descendants, siblings. XPath node types: root, element, text, attribute.

XML documents Nested start and end tags; text content; tag attributes. Ordering of nodes matters; of attributes doesn't. Well-formed documents.

Validation Specifying a hierarchy and attribute set. Validating document D against schema S. Schema languages, such as DTD.

XPath navigation Path expressions identify a set of nodes. Navigation: axes, node tests, predicates.

Relational Database Modelling

Relational models

- Relations: Tables matching schemas
- Schema: A set of field names and their domains
- Table: A set of tuples of values for these fields

Student

matric	**name**	**age**	**email**
s0456780	John	18	john@inf
s0378435	Helen	20	helen@phys
s0412375	Mary	18	mary@inf
s0189034	Peter	22	peter@math

Takes

matric	**code**	**mark**
s0456780	inf1	71
s0412375	math1	82
s0412375	geo1	64
s0189034	math1	56

Course

code	**title**	**year**
inf1	Informatics 1	1
math1	Mathematics 1	1
geo1	Geology 1	1
dbs	Database Systems	3
adbs	Advanced Databases	4

Tuple-Relational Calculus

The tuple-relational calculus (TRC) is a mathematical language for expressing queries over a relational database.

Standard TRC Idioms

To pick out some tuples from an existing table:

$$\{\, \mathrm{R} \in \mathrm{Table} \mid \exists S \in \mathrm{OtherTable}, \mathrm{T} \in \mathrm{FurtherTable} \,.\, \langle \mathit{Test} \rangle \,\}$$

To obtain tuples not directly appearing in any other table:

$$\{\, \mathrm{R} \mid \exists S \in \mathrm{OtherTable}, \mathrm{T} \in \mathrm{FurtherTable} \,.\, \langle \mathit{Test} \rangle \,\}$$

In both cases $\langle \mathit{Test} \rangle$ is a Boolean expression using R, S, T,... and including assertions like $(S.\mathrm{field} = \mathrm{R}.\mathrm{field}) \wedge \ldots$ to match R with other records.

Example: Records from Existing Table

All records for students more than 19 years old

$$\{ S \in \text{Student} \mid S.\text{age} > 19 \}$$

The set of all tuples S in the table "Student" with field "age" greater than 19.

All records for students taking math1

$$\{ S \in \text{Student} \mid \exists T \in \text{Takes} \,.\, S.\text{mn} = T.\text{mn} \wedge T.\text{code} = \text{math1} \}$$

The set of all tuples S in the table "Student" for which there is a tuple T in "Takes" linking the matriculation number of S to course "math1".

Example: Building New Records

Results for a single course

$$\{ \mathrm{R} \mid \exists \mathrm{T} \in \mathrm{Takes} \,.\, \mathrm{T.code} = \text{"math1"} \wedge \mathrm{R.mn} = \mathrm{T.mn} \wedge \mathrm{R.mark} = \mathrm{T.mark} \}$$

The set of all tuples R where there is a "Takes" tuple T with code "math1" such that R and T have matching matriculation number and mark.

Students on courses

$$\{ \mathrm{R} \mid \exists \mathrm{S} \in \mathrm{Student}, \mathrm{T} \in \mathrm{Takes}, \mathrm{C} \in \mathrm{Course} \,.\, \mathrm{S.mn} = \mathrm{T.mn} \wedge \mathrm{T.code} = \mathrm{C.code} \wedge \mathrm{R.name} = \mathrm{S.name} \wedge \mathrm{R.title} = \mathrm{C.title} \}$$

The set of all tuples R where there is a "Student" tuple S, a "Takes" tuple T, and a "Course" tuple C with matching matriculation numbers and course codes, and where R takes student name from tuple S and course title from tuple C.

Instructions

Quite separately from the technical content of your degree studies, the following is an important high-level skill which it's important that you can demonstrate by the time you graduate. Given a task to be completed:

1. Find the instructions.
2. Read the instructions.
3. Choose from:

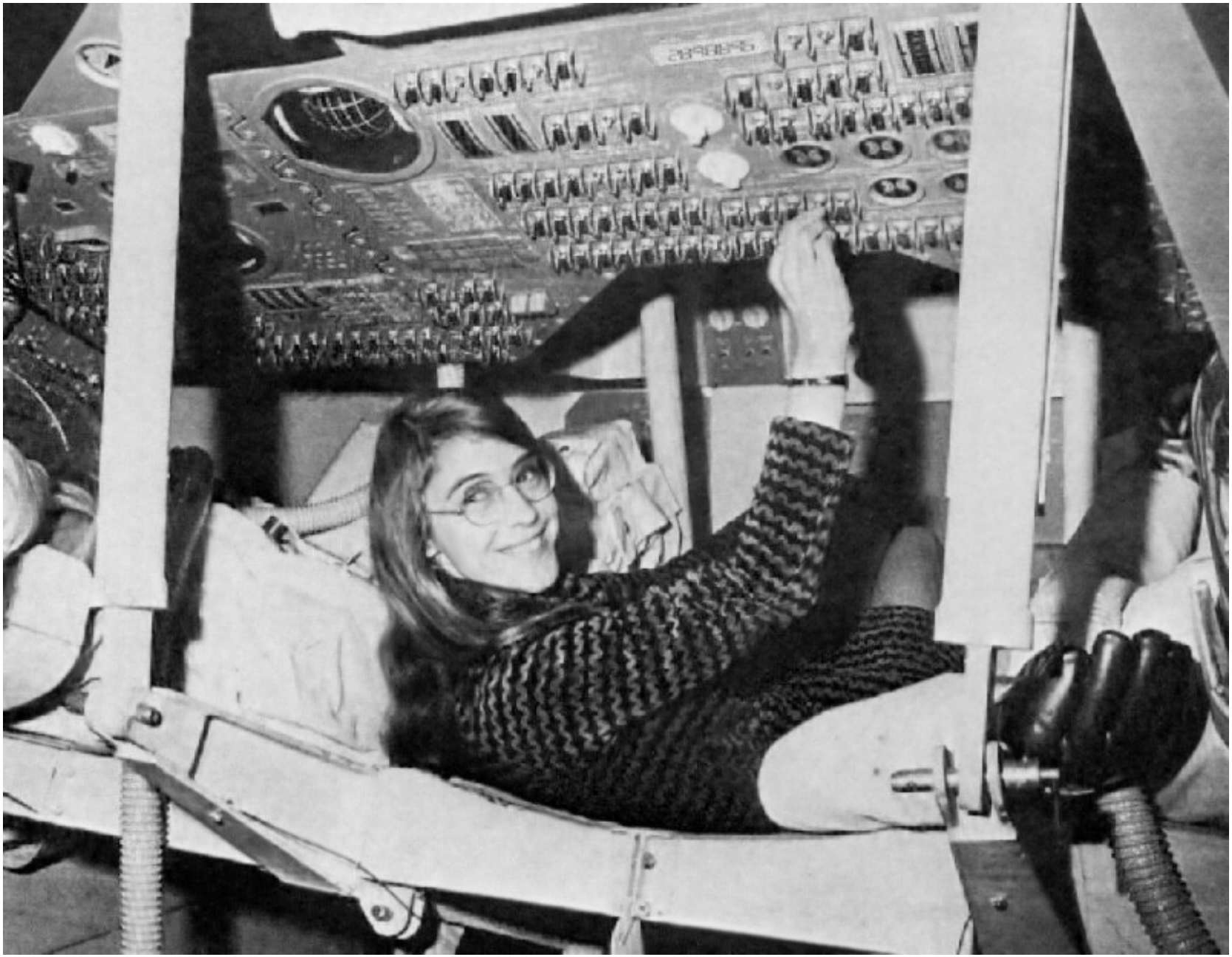
   Either Follow the instructions as directed

   Or Recognise that you are not following the instructions, and be prepared to address any consequences.

   Both of these are viable choices.

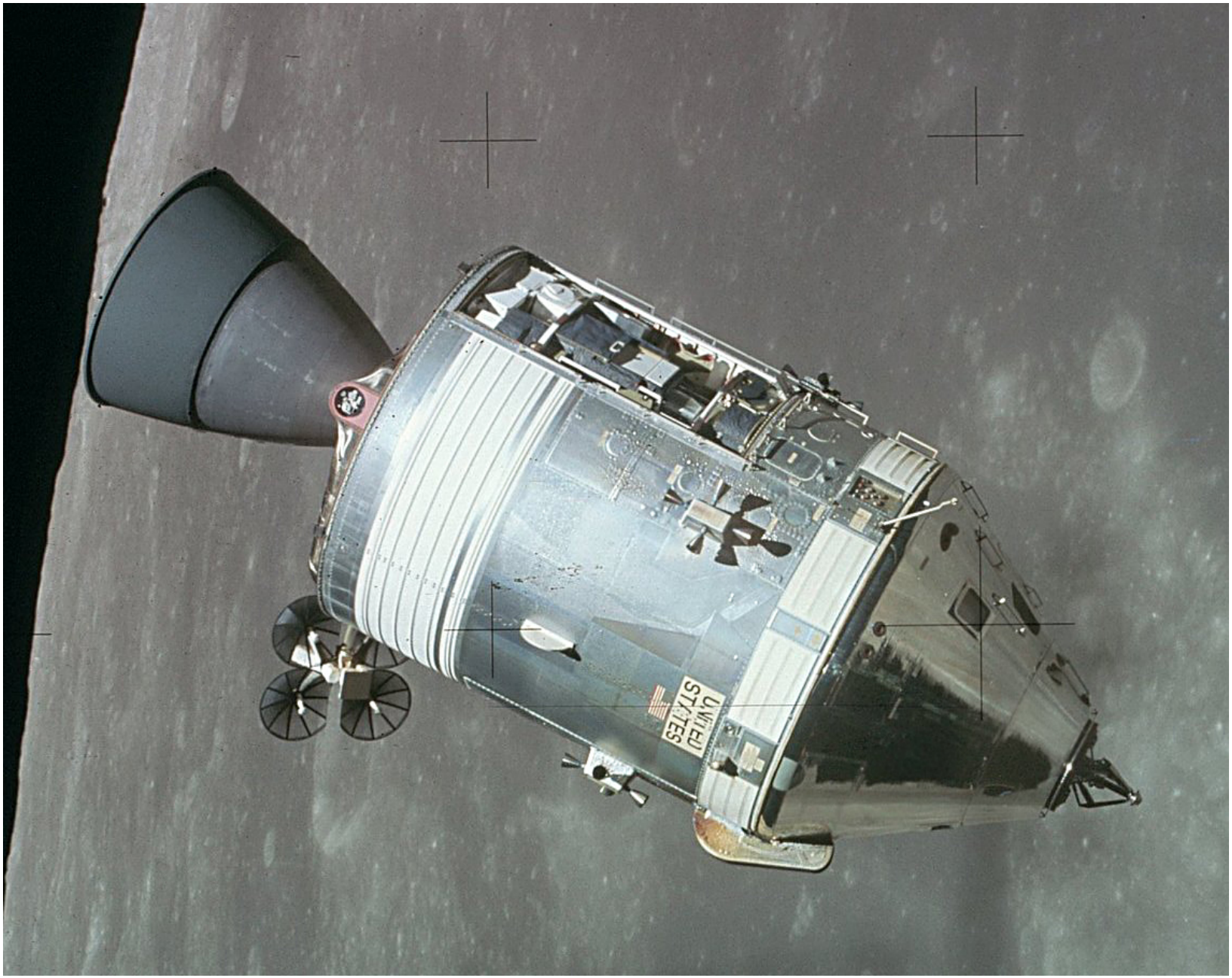
Do not underestimate this skill: it is much harder than many expect.

Actual Rocket Scientist



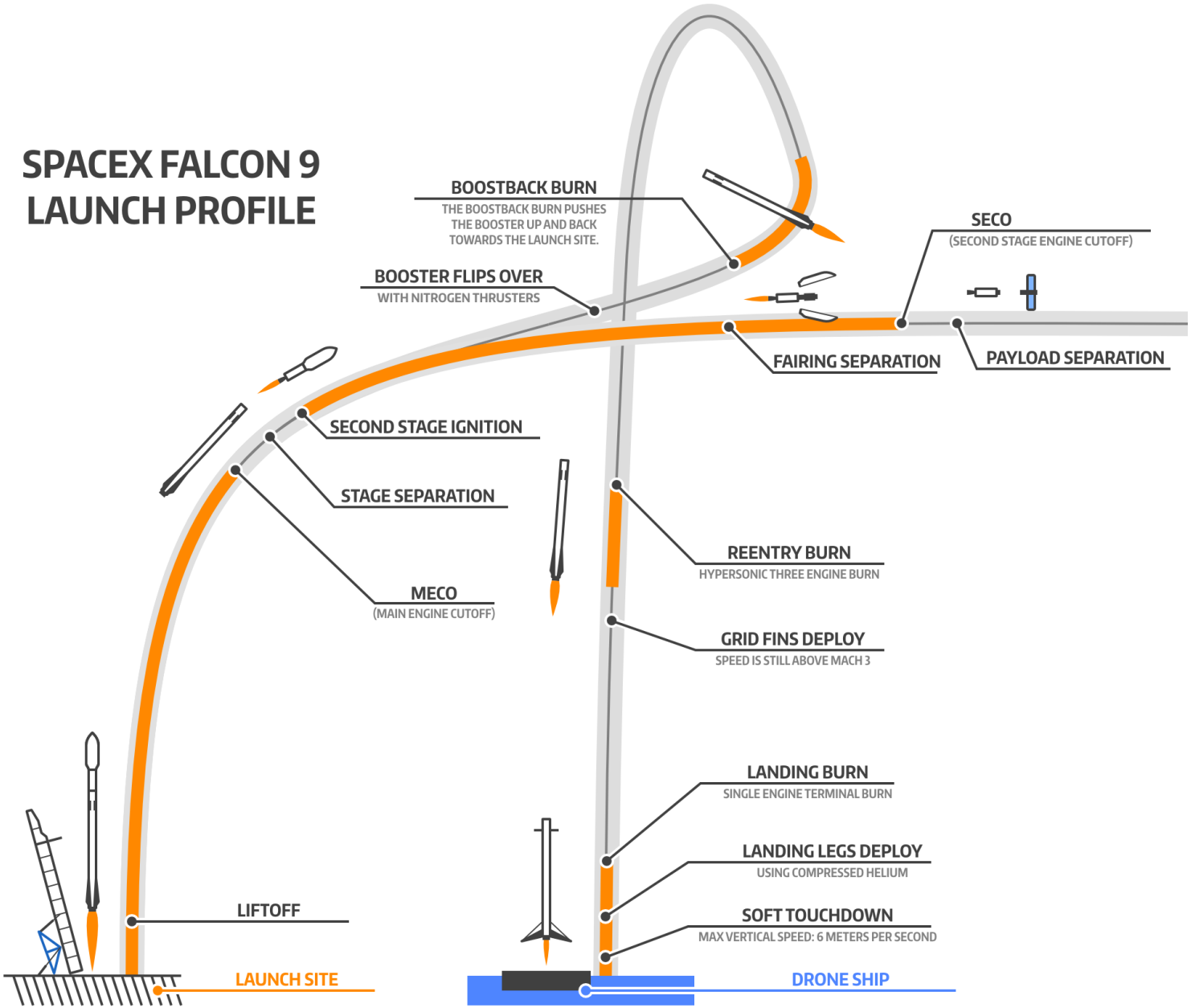
Actual Rocket Scientist

Apollo Command Module in Lunar Orbit



SpaceX Falcon 9 Going Up

SpaceX Falcon 9 Launch Profile



SpaceX Falcon 9 Coming Down

Rocket Science

RTQ

The most important rule of every examination ever:

Read The Question

Read the question. Answer the question. Go back and read the question again to make sure you have answered the right question. Repeat.

Don't be fooled: this is easy to forget. Sometimes you will get it wrong.

Follow These Instructions

- Download the Inf1-DA resit exam paper from August 2014 and the main exam paper from May 2015.
- Read the following two questions

August 2014	Question 1
May 2015	Question 2

  One of these runs on to a second page.
- Write out answers to both questions.
- Bring your solutions along on Tuesday.

In the lecture on Tuesday I shall work through and explain sample solutions for these past questions.