

**DEPARTMENT OF PHYSICS & ASTRONOMY**  
**PHAS3459 Mid-Term Exam**  
**16<sup>th</sup> November 2016**  
**09:45 - 12:45**

Please read the exam guidelines, rules, instructions and marking criteria at <https://moodle.ucl.ac.uk/mod/page/view.php?id=2016117> (linked from the PHAS3459 Examinations Moodle page).

This exam is worth 25% of your final mark for the course and is made up of two parts:

- 15 multiple-choice questions, worth 7.5% of your final mark;
- a programming exercise, worth 17.5% of your final mark.

The multiple-choice questions are given as a Moodle quiz.

The duration of the exam is 3 hours. For the programming exercise, you will write Java classes and methods to read data from a URL, analyse the data and present the results. The Java source code of your solution to the programming exercise should be uploaded using Moodle under the section headed "Exam I".

Each class should be uploaded as a separate file. Your classes should be created in a package called "exam1". This package must contain at least one "main" class called "MidTermExam": you should also include any other classes you create. You must upload all your classes used in your solution, including any you have copied or imported from earlier coursework modules. The code you upload **must be self-contained**: the marker must be able to compile and run it using only the classes uploaded and the Java API. If you use your own classes from earlier modules, make sure you copy them into the exam1 package and upload them along with any new classes you create during the exam.

**You are advised to read the entire exam paper before starting work.**

---

# DEPARTMENT OF PHYSICS & ASTRONOMY

## PHAS3459 Mid-Term Exam

### Programming Exercise

#### Background

In baseball, an entire industry exists in the collection and analysing of playing statistics. In Major League Baseball (MLB), statistics are collected in 3 main categories: hitting, pitching and fielding. In this exercise, you will analyse hitting statistics for all MLB players from the year 2001. *No prior knowledge of baseball is needed for this exercise: all information needed is given to you, and it is not necessary to understand the meaning of all the terms used.*

The statistics to be analysed can be found at the following URL:

<http://www.hep.ucl.ac.uk/undergrad/3459/exam-data/MLB2001Hitting.txt>

This file contains tab-delimited data, i.e. each field is separated by a tab. Care is needed when reading this file as some fields contain spaces. The first two lines are header lines: the first contains the name of each field in the statistics file, and the second is empty. Each of the remaining lines contains the following details for one MLB player:

- The player's name, in the form [LastName, FirstInitial] (Player);
- A three-letter code for the team that the player played for, unique to that team (Team);
- The player's Position (Pos);
- Games: the number of games the player played (G);
- At-bats: the total number of times over all games the player came up to bat (AB);
- Total Runs scored (R);
- Hits: the number of times the player managed to hit the ball and reach 1st base (H);
- Doubles: the number of times the player managed to reach 2nd base from a hit (2B);
- Triples: the number of times the player managed to reach 3rd base from a hit (3B);
- Home Runs: the number of home runs the player hit (HR);
- Runs Batted In: the number of runs that were scored due to a batters' action (RBI);
- Batting Average: the number of Hits divided by the number of At-Bats, H/AB (AVG);
- "On-Base Percentage" (not actually a percentage): the fraction of times the player managed to reach base, either by a successful hit, being walked or hit by a pitch (OBP);

When parsing the raw data, you may find it helpful to use a Scanner for each line, setting the delimiter to read tab-delimited rather than whitespace-delimited data accordingly:

```
Scanner s = new Scanner(line).useDelimiter("\t");
```

---

## Tasks

You should write a program using appropriate classes and methods to read the data from the URLs given above, store the data in suitable collection objects, and carry out the following tasks. When printing details of individual players, you should print their entire hitting statistics for the year in a clear and concise manner.

- Print the total number of players recorded in the file.
- Print details of the player who hit the most Home Runs for 2001.
- For each team, print:
  - the number of players with at least 10 At-Bats.
  - of the players with at least 10 At-Bats, the player with the highest “Slugging Percentage” (SLG); this is the total number of bases divided by the total number of At-Bats:

$$SLG = \frac{(1B) + (2 \times 2B) + (3 \times 3B) + (4 \times HR)}{AB}$$

Note that it is always quoted as this fractional value, not as an actual percentage, so should not be multiplied by 100.

- of the players with at least 10 At-Bats, the player with the highest On-Base Plus Slugging figure (OPS); this is the sum of the On-Base Percentage (OBP) and the Slugging Percentage (SLG):

$$OPS = OBP + SLG$$

Marks will be awarded for the following:

- Correctly loading and parsing the data from the given URL;
- Organising the data using appropriate `Collection(s)`;
- Correctly identifying the total number of players and the player who hit the most home runs;
- Grouping and filtering the data correctly by team and storing this data using the appropriate `Collection`;
- Appropriate use of `Exception` handling;
- Well commented code that is clearly structured and laid out and follows the Java indentation conventions.

More information on the marking criteria can be found at <https://moodle.ucl.ac.uk/mod/page/view.php?id=2016117>.

---

**END OF PAPER**