MATH1324 Introduction to Statistics

Assignment 2

Modelling the Distribution of Football Goals

## OverviewFree stock photo of grass, ...

The goal of assignment is simple. Your team must determine if football goals (soccer) fit a theoretical probability distribution. To do this, your team must decide what data to collect and how much. Your team must then go about collecting this data, tidying it up and importing it into R. Your team will then use R to summarise the empirical distribution of goals and compare it to a theoretical distribution. You will end by discussing the extent to which your theoretical distribution fits the empirical data and make recommendations regarding the modelling of football goals.

**Groups**

Students are permitted to work individually or in groups of up to 3 for Assignment 2. **Each individual or group must fill out the following form before 3/4/2017 to register their group details.** Submit the details of your group here.

[Individual and Group Registration Form](https://goo.gl/forms/lLtZP78QLRI1pcqR2)

**Submission**

Assignment 2 must be completed using the R Notebook template available here:

[R Notebook Template - Assignment 2](https://drive.google.com/a/rmit.edu.au/file/d/0Bwtqn_QygJ8_QXctMWJtSEVHRlk/view?usp=sharing)

**Reports are limited to 6 pages maximum (this includes code).** Information for using R Notebooks can be found here. The R Notebook template must be updated with your group details and your responses and code for the following sections. You must use the headings and chunks provided in the template. You can add more chunks to explain your approach if required.

This assignment is worth 5% and must be uploaded to the Assignment 2 Turnitin link by **23/04/2017**. **All group members must submit a copy of the report!** Group members that are not registered and do not submit a report will not be acknowledged. One group member’s submission will be marked and given feedback. It will be the responsibility of the marked group member to share the group’s feedback with the other group members. The other group members will receive a mark only.

Extensions will only be granted in accordance with the [RMIT University Extension and Special Consideration Policy](http://www1.rmit.edu.au/browse;ID=8r23xxpe4fne). No exceptions. Assignments submitted late will be penalised (see [Course Information](https://docs.google.com/a/rmit.edu.au/document/d/1gXDaYuLIFQXf1TQ6DFRoYV4IHb1XlY6wz-UiQiRqFt0/edit?usp=sharing) for further details).

The report must uploaded as a **PDF** with your code chunks showing. The easiest way to achieve this is to **Preview** your notebook → **Open in Browser** (Chrome) → Right click on the report in Chrome → Click **Print** and Select the **Destination** Option to **Save as PDF**.

## Collaboration

You are permitted to discuss and collaborate on the assignment with your classmates and other groups. However, the write-up of the report must be an individual/group effort. Assignments will be submitted through Turnitin, so if you’ve copied from a fellow classmate/group, it will be detected. It is your responsibility to ensure you do not copy or do not allow another classmate/groups to copy your work. If plagiarism is detected, both the copier and the student/group copied from will be responsible. It is good practice to never share assignment files with other students/groups. You should ensure you understand your responsibilities by reading the RMIT University website on [academic integrity](http://www1.rmit.edu.au/browse;ID=kkc202lwe1yv). Ignorance is no excuse.

## Report Section Descriptions

The report will be in a reproducible R Notebook format with written sections, R code and output. The report will be composed of the following sections (see Template above).

1. **Problem Statement [Plain text]**: Write a clear and concise problem statement that guides your investigation. A good problem statement explains the problem, the importance of solving it, and an outline of the approached taken to resolve it.
2. **Load Packages [R Chunk]:** This section is not marked.
3. **Data [Plain Text and R Chunk]:** Explain how and what data you collected. Import the data and prepare it for analysis. Show your code.
4. **Distribution Fitting [Plain Text and R Chunk] :** Compare the empirical distribution of football goals to a theoretical distribution. You need to do this visually and by using probability. Show your code.
5. **Interpretation [Plain text]:** Going back to your problem statement, what insight has been gained from the investigation? What are your suggestions for future investigations?

# 

# 

# Assignment 2 Marking Rubric

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria** | **Not acceptable**  **(0)** | **Needs Improvement**  **(1)** | **Good**  **(2)** |
| **Problem**  **(25%)** | No problem statement was provided OR the problem statement was inaccurate. | A problem statement was provided, but it was not clear or complete. | The group provides a clear and accurate problem statement of the report. |
| **Data**  **(25%)** | The data collection and management was insufficient or inappropriate for the purpose of the investigation. | There were some issues with the data collection and management that needed improvement. | Data collection and management was appropriate for the goals of the investigation. |
| **Fitting**  **(25%)** | The attempt to compare the empirical distribution to the theoretical distribution is inappropriate or missing. | The attempt to compare the empirical distribution to the theoretical distribution needed improvement. Explanation and/or the approach taken needed improvement. | The attempt to compare the empirical distribution to the theoretical distribution was appropriate and well explained. |
| **Interpretation**  **(25%)** | The interpretation of the fitted data was lacking or inappropriate. | Parts of the interpretation were appropriate, but improvement was needed relating the findings back to the problem and thinking critically about the investigation. | The interpretation demonstrated a clear understanding of the results within the context of the problem and limitations of the investigation performed. |