Randomised summative e-assessment using R-exams.

UKCOTS 2024, Manchester

Ilaria Bussoli, ib641@bath.ac.uk

Department of Mathematical Sciences, University of Bath

The setup

- Using the "[exams](#)" `R` package to create dynamic (randomised) cloze quiz questions (.R and .Rmd files).

- The randomisation arises from:
  - the ***internal*** data (numbers, text blocks, graphics, …)
  - the ***sub-questions*** generated in the question, and
  - the ***order*** in which the questions are presented.

- Two files:
  - RMarkdown for data and question-generation, and
  - R script to convert the files via `exams2moodle` call.
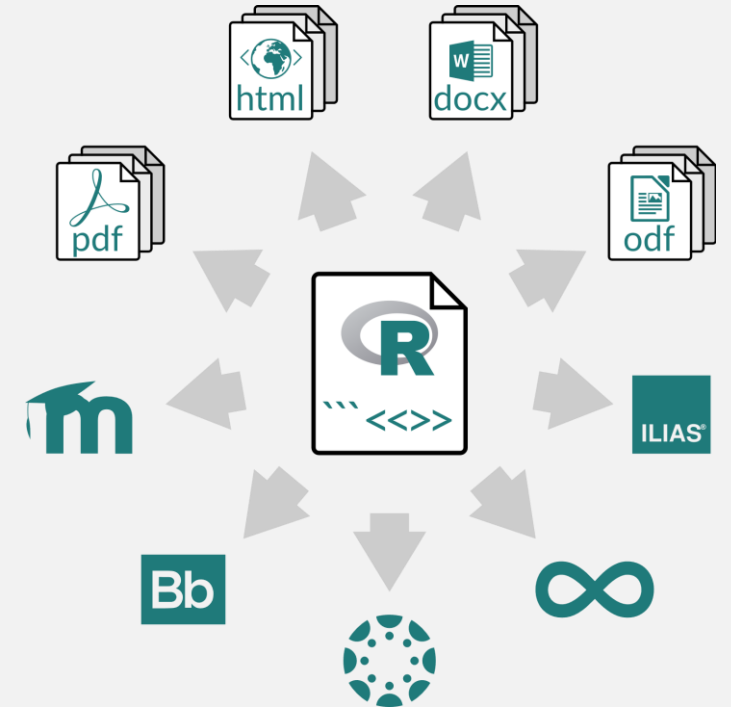
Core sections for a question

1. **Data generating** section in a R-chunk (e.g., `sample`, `rbinom`, `pnorm`, etc.).

2. **Question** and (optional) **solution** sections:
   - LaTeX / Markdown syntax for the text.
   - R-chunk to pass the list of questions generated in 1.
   - R-inline call to pass the list of correct answers generated in 1 to be associated to the solution section.

3. **Meta-information** section (e.g., the type of question, the list of solutions, and the name of the question).

End results

A medical team is convinced that the average basal temperature of (seemingly) healthy people has decreased over time, and is no longer 37 degrees (in Celsius). To demonstrate this, they measure the temperature of 24 randomly selected healthy subjects, finding an average temperature of 37.5 degrees and a sample standard deviation of 2.17 degrees. They want to run a hypothesis testing procedure such that the probability of finding a type 1 error is equal to 0.2.

Answer the following questions and, in case, help yourself with the appropriate quantile among those provided:

$z_{0.8} = 0.842;$ $z_{0.9} = 1.282;$ $t_{23,0.8} = 0.858$ $t_{23,0.9} = 1.319;$ $t_{22,0.8} = 0.858;$ $t_{22,0.9} = 1.321$

Note: Round up numerical answers to **3 decimal places** after the point. For example: $1.3335 \rightarrow 1.334$ and $1.3332 \rightarrow 1.333$.

Student A

a. Which option, among the three listed here, represents the hypotheses that the medical team wants to verify?

- ⦿ $H_0 : \mu = 37$ against $H_1 : \mu \neq 37$.
- ◯ $H_0 : \mu \geq 37$ against $H_1 : \mu < 37$.
- ◯ $H_0 : \mu \leq 37$ against $H_1 : \mu > 37$.

b. Compute the appropriate test statistic ts_{obs}: [   ]

A medical team is convinced that the average basal temperature of (seemingly) healthy people has changed over time, and is no longer 37 degrees (in Celsius). To demonstrate this, they measure the temperature of 24 randomly selected healthy subjects, finding an average temperature of 37.37 degrees and a sample standard deviation of 2.12 degrees. They want to run a hypothesis testing procedure such that the probability of finding a type 1 error is equal to 0.01.

Answer the following questions and, in case, help yourself with the appropriate quantile among those provided:

$z_{0.99} = 2.326;$ $z_{0.995} = 2.576;$ $t_{23,0.99} = 2.5$ $t_{23,0.995} = 2.807;$ $t_{22,0.99} = 2.508;$ $t_{22,0.995} = 2.819$

Note: Round up numerical answers to **3 decimal places** after the point. For example: $1.3335 \rightarrow 1.334$ and $1.3332 \rightarrow 1.333$.

Student B

a. Which option, among the three listed here, represents the hypotheses that the medical team wants to verify?

- ◯ $H_0 : \mu \geq 37$ against $H_1 : \mu < 37$.
- ◯ $H_0 : \mu = 37$ against $H_1 : \mu \neq 37$.
- ◯ $H_0 : \mu \leq 37$ against $H_1 : \mu > 37$.

b. Compute the appropriate test statistic ts_{obs}: [   ]

What to take away

Good "bits"

- One-for-all generators of questions.

- Automatic feedback.

- Automatic grading (summative e-assessment for ~100 Management students in Venice (2020)).

- Self-assessment and unique experience.

Bad "bits"

- Knowledge of R and Rmarkdown.

- Time consuming.

- Not compatible with bookdown/quarto books.