

Example exam question

Week 8

Contents

| | |
|-----------------------------------------------|----------|
| Information | 1 |
| Instructions | 1 |
| Question 1 - HIV (40% of total mark) | 2 |
| Load data | 2 |
| Data properties | 2 |
| State vs sex | 2 |
| Mortality differences between States. | 3 |

Information

Week 8's mid-semester test will cover all content up to and including week 6. This includes topics such as:

- Data structure and bias.
- t-tests.
- chi-square tests.
- odds-ratios and relative risk.
- Missing data and clustering.
- One-way, two-way ANOVA and interaction effects.
- Regression, logistic regression, model selection, interpretation of coefficients.

It is important that you are able to comment and check on the assumptions of tests when performing analysis.

Instructions

- Use pre-filled rmarkdown file to help arrange your answers.
- Submit html file to turnitin.

Question 1 - HIV (40% of total mark)

Dr P. J. Solomon and the Australian National Centre in HIV Epidemiology and Clinical Research performed a study on HIV transmission and survival. They recorded the following characteristics of all patients who entered a HIV clinical of the course of the study.

| Variable | Description |
|----------|--------------------------------------------------------------------------------|
| state | Grouped state of origin: “NSW” includes ACT and “other” is WA, SA, NT and TAS. |
| sex | Sex of patient. |
| diag | (Julian) date of diagnosis. |
| death | (Julian) date of death or end of observation. |
| status | “A” (alive) or “D” (dead) at end of observation. |
| T.cat | Reported transmission category. |
| age | Age (years) at diagnosis. |

Load data

To load the data use the link <https://wimr-genomics.vip.sydney.edu.au/AMED3002/data/HIV.csv> as follows:

```
HIV = read.csv("https://wimr-genomics.vip.sydney.edu.au/AMED3002/data/HIV.csv")
```

Data properties

- How many variables and observations are in the dataset?
- Comment on the class of these variables and how they are stored in R.
- Is there any missing data in this dataset?

State vs sex

The researches are interested in the characteristics for the people being diagnosed in each state. Specifically, they would like to know whether there were differing numbers of men and women that were diagnosed in each state.

- What types of variables should sex and state be?
- What is an appropriate statistical test that could be used by the researchers to test this question?
- What is the corresponding null and alternate hypothesis?
- Construct a contingency table using the variables *sex* and *state* and comment on any striking features.
- Perform the appropriate test.
- Using a significance threshold of 0.05 what would you conclude from this test?
- What were the assumptions for this test? Comment on them in the context of the observed data.

Mortality differences between States.

The researchers would like to know if there is some difference between the states in the outcomes for HIV patients. They decide that they would like to see if the time between diagnosis and death are different between states.

- k. Create a new dataset containing only patients that died using the *status* variable.
- l. Create a new variable for the *time that patients survived* by subtracting *diag* from *death*.
- m. Visualise the time to death for the patients in each state using a boxplot. Comment on any striking features.
- n. Use an ANOVA to test whether the time between diagnosis and death are different between states.
- o. Using a significance threshold of 0.05 what would you conclude from this test?
- p. What were the assumptions for this test? Comment on them in the context of the observed data and fitted model.
- q. Are there any other tests that you could perform to help the researchers interpret these results. If yes, what would you inform the researchers?