Demo 1: ICU Stay after Cardiac Arrest

Introduction

Organising this data analysis

The demonstration is guided by three principles; **reproducibility**, **literate coding** and **standardised workflow**.

The aim is always to reduce the number of errors.

Reproducibility

No interactive analysis

- Data are read from the original Excel file and clean in an R script
- Cleaned data saved in RDS format
- Data exploration in an R script
- Report via an rmarkdown file

Literate coding

- File and section headers describe the analysis
- Consistent naming convention + descriptive names
- use of DF to denote a data frame or tibble
- use of the pipe
- packages loaded together at the start
- filenames defined together at the start
- spaces around `=`, `+` etc.
- alignment of calculations

Standardised workflow

Standard folder structure. Here is mine, ProjectFolder - data
- rawData
- rData
- docs
- R
- rmd
- temp

Diary file

New functions used in the demo

`glimpse()` - summarise the contents of a data frame or tibble
`is.na()` - return TRUE if missing (NA) and FALSE otherwise
`!` - negation or NOT, turns TRUE to FALSE & FALSE to TRUE
`!is.na()` - return TRUE if NOT missing
`median()` - `sum()`, `min()`, `max()` - return median, total, minimum, maximum
`quantile()` - return a quantile, i.e. value a proportion prob of the way through the sorted values, median(x) is the same as quantile(x, prob=0.5)
`wilcox.text()` - Wilcoxon rank sum test, also known as the Mann-Whitney test. Non-parametric test for difference in location between 2 groups.
`kruskal.test()` - Kruskal-Wallis nonparametric test for difference in location between 2 or more groups.
`scale_y_log10()` - in ggplot, use a log10 scale for the y-axis. Also scale_x_log10(), scale_x_log(), scale_x_reverse() and many others

Length of Stay in ICU

```
#=====================================================================
# File:         demo1.R
# Project:      icu (project folder C:/Projects/ISCB/icu)
# Author:       john thompson
# Date:         22 July 2022
# Description: Analysis of length of Stay in ICU
#=====================================================================
library(tidyverse)
library(readxl)
library(janitor)

excel_file <- "C:/Projects/RCourse/ISCB/data/rawData/iesu/Table 1 Enrica.xlsx"
clean_file <- "C:/Projects/Rcourse/ISCB/data/rData/cardiac.rds"

#---------------------------------------------------------------------
# Read & clean the data
#
read_excel( excel_file ) %>%
  remove_empty() %>%
  remove_constant() %>%
  clean_names( "lower_camel" ) %>%
  filter( !is.na(id) ) %>%
```

```
  saveRDS( clean_file )

#-----------------------------------------------------------------
# Read clean data
#
cardiacDF <- readRDS( clean_file )

#-----------------------------------------------------------------
# Data summary
#
cardiacDF %>%
  glimpse()

#-----------------------------------------------------------------
# Summarise ICU stay
#
cardiacDF %>%
  summarise( n       = n(),
             icuMean = mean(icuStay),
             icuStd  = sd(icuStay))

#-----------------------------------------------------------------
# Histogram of Days in ICU
#
cardiacDF %>%
  ggplot( aes(x = icuStay)) +
  geom_histogram( bins = 50, fill = "steelblue" ) +
  labs( x     = "Days",
        title = "Number of days spent in ICU following a Cardiac Arrest")

#-----------------------------------------------------------------
# Summarise ICU stay
#
cardiacDF %>%
  summarise( n         = n(),
             icuMin    = min(icuStay),
             icuQ1     = quantile(icuStay, prob=0.25),
             icuQ3     = quantile(icuStay, prob=0.75),
             icuMax    = max(icuStay) )
#-----------------------------------------------------------------
# Mortality & ICU stay
#
cardiacDF %>%
  group_by( hospDeath ) %>%
  summarise( n         = n(),
             icuMedian = median(icuStay),
             icuQ1     = quantile(icuStay, prob=0.25),
             icuQ3     = quantile(icuStay, prob=0.75)) %>%
  mutate( pct = 100 * n / sum(n) ) %>%
  select( hospDeath, n, pct, icuMedian, icuQ1, icuQ3)


#-----------------------------------------------------------------
```

```r
# Days in ICU by Death in Hospital
#
cardiacDF %>%
  ggplot( aes(x = icuStay)) +
  geom_histogram( bins = 50, fill = "steelblue" ) +
  labs( x = "Days in ICU") +
  facet_grid( hospDeath ~ . )

#-----------------------------------------------------------------
# Wilcoxon test of Days in ICU by Death in Hospital
#
cardiacDF %>%
  wilcox.test( icuStay ~ hospDeath, data = .)

#-----------------------------------------------------------------
# CVD score & ICU stay
#
cardiacDF %>%
  mutate( cvd = chronHeartFailure + hypertension + coronArteryDis ) %>%
  group_by( hospDeath, cvd) %>%
  summarise( n         = n(),
             icuMedian = median(icuStay),
             icuQ1     = quantile(icuStay, prob=0.25),
             icuQ3     = quantile(icuStay, prob=0.75))

#-----------------------------------------------------------------
# cvd and days in ICU for survivors
#
cardiacDF %>%
  filter( hospDeath == 0 ) %>%
  mutate( cvd = chronHeartFailure + hypertension + coronArteryDis ) %>%
  ggplot( aes(x = icuStay)) +
  geom_histogram( bins = 50, fill = "steelblue" ) +
  labs( x = "Days in ICU") +
  facet_grid( cvd ~ . )

#-----------------------------------------------------------------
# Kruskal-Wallis test of Days in ICU for survivors
#
cardiacDF %>%
  filter( hospDeath == 0 ) %>%
  mutate( cvd = chronHeartFailure + hypertension + coronArteryDis ) %>%
  kruskal.test( icuStay ~ cvd, data = .)

#-----------------------------------------------------------------
# ICU Stay & Time to Return of Spontaneous Circulation (RoSC)
#
cardiacDF %>%
  ggplot( aes(x = timeToRoscMin, y = icuStay) ) +
  geom_point( colour = "darkgreen") +
  scale_y_log10() +
  geom_smooth( method = "lm" , colour = "black") +
  labs( x = "Time to Spontaneous Circulation (mins)",
```

```
       y = "Days in ICU") +
  facet_grid( hospDeath ~ .)
```

Conclusions

The main lessons from this simple analysis are,

- Patients spend an average of 8 days in ICU
- The distribution is very skewed. Median 4 days, min 1 day, max 95 days
- 57% of ICU patients die in hospital
- Patients who die spend a shorter time in ICU, Median 3 days vs 7 days, p=7x10-12 (Wilcoxon Rank Sum test)
- No evidence that more the CVD score is associated with a longer ICU stay
- In patients who survive to discharge, a longer time to restoration of spontaneous circulation after the arrest is associated with a longer stay in ICU

Preparation of a Report

An html document describing the analysis is generated by `demo1_report.rmd`.

Steps in creating the report.

- Open a new rmarkdown script. RStudio: File - New File - R Markdown
- Remove the sample text and code, leave YAML header
- Adapt the YAML
- Add new section headers
- Copy selected code from your R script
- Add descriptive text
- knit everything together

Here is my complete R Markdown file.

````
---
title: "ICU Stay after a Cardiac Arrest"
author: "John Thompson"
date: "25/02/2022"
output: html_document
---

```{r setup, include=FALSE}
````
````

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
```

```{r echo = FALSE}
library(tidyverse)

clean_file <- "C:/Projects/RCourse/ISCB/data/rData/cardiac.rds"

cardiacDF <- readRDS(clean_file)
```

Background

The outcome of patients after sudden cardiac arrest (CA) remains poor, in particular because a large pr

In this analysis we look at the length of stay in ICU of cardiac arrest patients using data taken from,

Iesu, E., Franchi, F., Zama Cavicchi, F., Pozzebon, S., Fontana, V., Mendoza, M., ... & Taccone, F. S.
Acute liver dysfunction after cardiac arrest.
PLoS One, 13(11), e0206655..

Study Design

This retrospective study was performed in the Department of Intensive Care at Erasme Hospital, Brussels

ICU Stay

Figure 1 shows a histogram of the number of days in ICU for the 374 patients recruited into the study.

```{r hist_stay}
cardiacDF %>%
  ggplot( aes(x = icuStay)) +
  geom_histogram( bins = 50, fill="steelblue" ) +
  labs( x     = "Days in ICU",
        title = "Figure 1: Histogram of Length of Stay in ICU")
```

Summary statistics show that the median stay in ICU was 4 days, range 1 day to 95 days.
```{r summary_stay}
cardiacDF %>%
  summarise( n      = n(),
             icuMED = median(icuStay),
             icuMin = min(icuStay),
             icuQ1  = quantile(icuStay, prob=0.25),
             icuQ3  = quantile(icuStay, prob=0.75),
             icuMax = max(icuStay))
```

ICU Stay and Mortality

Patients who died in hospital include many who died in the ICU unit and so they have a shorter average s

```{r hist_mortality}
cardiacDF %>%


```

```
  ggplot( aes(x = icuStay)) +
  geom_histogram( bins = 50, fill = "steelblue" ) +
  labs( x = "Days in ICU",
        title = "Figure 2: ICU Stay for Patients who Survived (0) and those that died in hospital (1)")
  facet_grid( hospDeath ~ .)
```

Summary statistics show that the 57% of patients who dies in hospital had a median stay in ICU of 3 days

```{r summary_stay}
cardiacDF %>%
  group_by( hospDeath) %>%
  summarise( n      = n(),
             icuMED = median(icuStay),
             icuQ1  = quantile(icuStay, prob=0.25),
             icuQ3  = quantile(icuStay, prob=0.75)) %>%
  mutate( pct = 100 * n / sum(n)) %>%
  select( hospDeath, n, pct, icuMED, icuQ1, icuQ3)
```

In a wilcoxon rank sum test the distributions of length of stay are shown to be significantly different

```{r test_mortality}
cardiacDF %>%
  wilcox.test( icuStay ~ hospDeath, data = .)
```

ICU Stay and Pre-existing CVD

A score for pre-existing cardiovascular disease (CVD) was created and lenngth of stay in ICU was compare

```{r summary_cvd}
cardiacDF %>%
  mutate( cvd = chronHeartFailure + hypertension + coronArteryDis ) %>%
  group_by( hospDeath, cvd) %>%
  summarise( n      = n(),
             icuMED = median(icuStay),
             icuQ1  = quantile(icuStay, prob=0.25),
             icuQ3  = quantile(icuStay, prob=0.75))
```

ISU Stay and RoSC

The study recorded the time in minutes from the arrest to the return of spontaneous circulations (RoSC)

```{r rosc}
cardiacDF %>%
  ggplot( aes(x = timeToRoscMin, y = icuStay)) +
  geom_point( colour="steelblue" ) +
  scale_y_log10() +
  geom_smooth( method = "lm") +
  labs( y = "Days in ICU") +
  facet_grid( hospDeath ~ .)
```

Conclusion

In hospital mortality is high amongst patients who survive a cardiac arrest. The median stay in ICU for