

N741: Homework 5

Melinda K. Higgins, PhD.

February 27, 2017

Homework 5 - DUE March 15, 2017

For this homework, we'll work with the “Wong” dataset built in to the `car` package. The “Wong” data frame has 331 row and 7 columns. The observations are longitudinal data on recovery of IQ after comas of varying duration for 200 subjects. The data are from Wong, Monette, and Weiner (2001) and are for 200 patients who sustained traumatic brain injuries resulting in comas of varying duration. After awakening from their comas, patients were periodically administered a standard IQ test, but the average number of measurements per patient is small ($331/200 = 1.7$). *To get more info type `??Wong`.*

The 7 variables in the dataset are:

- `id`
 - patient ID number.
- `days`
 - number of days post coma at which IQs were measured.
- `duration`
 - duration of the coma in days.
- `sex`
 - a factor with levels Female and Male.
- `age`
 - in years at the time of injury.
- `piq`
 - performance (i.e., mathematical) IQ.
- `viq`
 - verbal IQ.

Load dataset in from car package

```
library(car)

## Warning: package 'car' was built under R version 3.3.3
## Warning: replacing previous import 'lme4::sigma' by 'stats::sigma' when
## loading 'pbkrtest'
data(Wong)

library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:car':
##
##      recode
##
## The following objects are masked from 'package:stats':
##
```

```
##      filter, lag
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

# add an age group variable
Wong$agegrp <- case_when(
  (Wong$age > 0 & Wong$age <= 10) ~ 1,
  (Wong$age > 10 & Wong$age <= 20) ~ 2,
  (Wong$age > 20 & Wong$age <= 30) ~ 3,
  (Wong$age > 30 & Wong$age <= 40) ~ 4,
  (Wong$age > 40 & Wong$age <= 50) ~ 5,
  (Wong$age > 50 & Wong$age <= 60) ~ 6,
  (Wong$age > 60 & Wong$age <= 70) ~ 7,
  (Wong$age > 70 & Wong$age <= 100) ~ 8)

# convert to factor, add code levels and labels
Wong$agegrp <- factor(Wong$agegrp,
  levels = c(1,2,3,4,5,6,7,8),
  labels = c("Ages 1-10",
             "Ages 11-20",
             "Ages 21-30",
             "Ages 31-40",
             "Ages 41-50",
             "Ages 51-60",
             "Ages 61-70",
             "Ages 71-100"))
```

Using this dataset, and today's demos complete the following tasks:

1. Make a table of non-parametric statistics (median and IQR) for the number of days and duration grouped by sex. You'll be using `summarise()` from the `dplyr` package. For a given variable `x` you'll use `median(x, na.rm=TRUE)`, `quantile(x, 0.25, na.rm=TRUE)`, and `quantile(x, 0.75, na.rm=TRUE)`. Give the table a title using the `caption=` option and update the column names with something nice using the `col.names=` option in the `knitr::kable()` command.

ANSWER KEY

```
t1 <- Wong %>%
  group_by(sex) %>%
  summarise(durmed = median(duration, na.rm=TRUE),
            dur25 = quantile(duration, 0.25, na.rm=TRUE),
            dur75 = quantile(duration, 0.75, na.rm=TRUE),
            daymed = median(days, na.rm=TRUE),
            day25 = quantile(days, 0.25, na.rm=TRUE),
            day75 = quantile(days, 0.75, na.rm=TRUE))

knitr::kable(t1,
  col.names=c("Sex",
              "Median Duration",
              "Q1 Duration",
              "Q3 Duration",
              "Median Days",
              "Q1 Days",
              "Q3 Days"),
  caption="Non-parametric Summary By Sex - Days and Duration")
```

Table 1: Non-parametric Summary By Sex - Days and Duration

Sex	Median Duration	Q1 Duration	Q3 Duration	Median Days	Q1 Days	Q3 Days
Female	4	1	11	135	58.50	361.00
Male	7	1	18	163	59.75	431.25

2. Make a table of parametric statistics (mean and SD) for the performance outcomes `piq` and `viq` grouped by `sex`. Like the table above, you'll be using `summarise()` from the `dplyr` package. Now you'll use `mean(x, na.rm=TRUE)` and `sd(x, na.rm=TRUE)`. Give the table a title using the `caption=` option and update the column names with something nice using the `col.names=` option in the `knitr::kable()` command.

ANSWER KEY

```
t1 <- Wong %>%
  group_by(sex) %>%
  summarise(piqmean = mean(piq, na.rm=TRUE),
            piqsd = sd(piq, na.rm=TRUE),
            viqmean = mean(viq, na.rm=TRUE),
            viqsd = sd(viq, na.rm=TRUE))

knitr::kable(t1,
             col.names=c("Sex",
                         "Mean PIQ",
                         "SD PIQ",
                         "Mean VIQ",
                         "SD VIQ"),
             caption="Parametric Summary By Sex - PQI and VIQ")
```

Table 2: Parametric Summary By Sex - PQI and VIQ

Sex	Mean PIQ	SD PIQ	Mean VIQ	SD VIQ
Female	89.18310	17.99866	94.35211	14.24690
Male	87.11154	14.25658	95.13077	14.02281

3. Make a table containing the frequencies and relative percentages for `agegrp`. Use the example we did in class to help guide you.

ANSWER KEY

```
# find the sample size or number of rows in dataset
ss <- length(Wong$agegrp)

# group by each level of agegrp
# find the raw counts using n()
# then compute the relative % by dividing by ss
t2 <- Wong %>%
  group_by(agegrp) %>%
  summarise(freq = n(),
            pct = n()*100/ss)

# make a summary table
```

```
knitr::kable(t2,
  col.names = c("Age Group",
    "Frequency",
    "Percent"),
  caption="Frequency Table for Age Group")
```

Table 3: Frequency Table for Age Group

Age Group	Frequency	Percent
Ages 1-10	1	0.3021148
Ages 11-10	53	16.0120846
Ages 21-10	144	43.5045317
Ages 31-10	48	14.5015106
Ages 41-10	42	12.6888218
Ages 51-10	27	8.1570997
Ages 61-70	14	4.2296073
Ages 71-100	2	0.6042296

4. Make a regression model (Model 1) for the performance IQ (piq) using **age** and **sex**. Put the regression model results into a table.

ANSWER KEY

```
m1 <- lm(piq~sex+age, data=Wong)
sm1 <- summary(m1)
knitr::kable(sm1$coefficients,
  caption="Model 1")
```

Table 4: Model 1

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	86.8868114	2.5796740	33.681314	0.0000000
sexMale	-2.1651531	2.0258169	-1.068780	0.2859546
age	0.0743977	0.0600527	1.238874	0.2162781

5. Make a second regression model (Model 2) for performance IQ (piq) using **age** and **sex** plus **days** and **duration**. Put the regression model results into a table.

ANSWER KEY

```
m2 <- lm(piq~sex+age+days+duration, data=Wong)
sm2 <- summary(m2)
knitr::kable(sm2$coefficients,
  caption="Model 2")
```

Table 5: Model 2

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	88.0961373	2.6462149	33.2913764	0.0000000
sexMale	-1.7252891	2.0152576	-0.8561135	0.3925638
age	0.0542142	0.0604989	0.8961181	0.3708509
days	0.0011534	0.0007457	1.5468461	0.1228705
duration	-0.1026657	0.0328189	-3.1282468	0.0019172

6. Finally, make a table showing the results from the `anova()` command comparing Model 1 and Model 2 you made above using the example we did in class as a guide.

ANSWER KEY

```
m1m2 <- anova(m1,m2)
row.names(m1m2) <- c("Model 1","Model 2")
knitr::kable(m1m2,
              caption = "Compare M1 and M2")
```

Table 6: Compare M1 and M2

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Model 1	328	74967.59	NA	NA	NA	NA
Model 2	326	72585.66	2	2381.933	5.348922	0.0051796

7. STUDENT CHOICE - pick either a `htmlwidget` from <http://gallery.htmlwidgets.org/> or do a “flex-dashboard” using the templates at <http://rmarkdown.rstudio.com/flexdashboard/> as a guide.

ANSWER KEY - answers will vary by student

References

Wong, P. P., Monette, G., and Weiner, N. I. (2001) Mathematical models of cognitive recovery. *Brain Injury*, 15, 519–530.

Fox, J. (2016) *Applied Regression Analysis and Generalized Linear Models*, Third Edition. Sage.