

Caso10 - Registros médicos

Sebastian Matias Romero Davila

2024-07-13

Contents

<i>Base de datos inicial</i>	2
<i>Características Inherentes</i>	2
Exactitud	2
Compleitud	3
Consistencia	5
<i>Características Dependientes del sistema</i>	8
Disponibilidad	8
Portabilidad	8

Base de datos inicial

Cargamos librerías a emplear

```
library(tidyverse)
library(rio)
library(kableExtra)
```

Base: base_registros_medicos

```
registros_medicos <- import("base_registros_medicos.csv")
head(registros_medicos,10) %>%
kable(booktabs = TRUE,format = "latex")%>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_registro	id_paciente	fecha_admision	fecha_alta	costo	diagnostico	estado
432	148	2021-12-22	2023-03-06	12211.04	Neumonía	Hospitalizado
427	10	2022-05-15	2020-07-12	39787.55	COVID-19	NA
290	158	2022-10-15	2021-09-18	13363.47	COVID-19	Alta
448	71	2021-07-01	2022-04-07	11101.13	NA	Alta
483	89	2021-02-01	2023-03-26	45581.97	NA	Fallecido
310	44	2022-09-11	2022-10-09	28769.98	Fractura	Alta
466	187	2021-09-10	2020-02-17	43023.42	Gastritis	Fallecido
3	77	2022-05-22	2020-01-29	30320.48	NA	NA
77	259	2021-09-08	2023-04-09	16519.44	Gastritis	NA
-323	67	2021-10-09	2021-03-23	10229.33	Fractura	NA

Características Inherentes

Exactitud

Para la representación correcta de la realidad de los datos, empezamos identificando los tipos de variables

```
registros_medicos %>% glimpse()
```

```
## Rows: 800
## Columns: 7
## $ id_registro    <int> 432, 427, 290, 448, 483, 310, 466, 3, 77, -323, 51, 310~
## $ id_paciente    <int> 148, 10, 158, 71, 89, 44, 187, 77, 259, 67, 289, 112, 2~
## $ fecha_admision <chr> "2021-12-22", "2022-05-15", "2022-10-15", "2021-07-01",~
## $ fecha_alta     <IDate> 2023-03-06, 2020-07-12, 2021-09-18, 2022-04-07, 2023~~
## $ costo          <dbl> 12211.04, 39787.55, 13363.47, 11101.13, 45581.97, 28769~
## $ diagnostico    <chr> "Neumonía", "COVID-19", "COVID-19", NA, NA, "Fractura",~
## $ estado         <chr> "Hospitalizado", NA, "Alta", "Alta", "Fallecido", "Alta~
```

Luego, verificamos si existen inexactitudes

- id_registro: Se encuentran valores enteros negativos

- fecha_admision: Formato inadecuado
- fecha_admision > fecha_alta

```
medinexactos <- registros_medicos %>% filter(
  id_registro<0|substr(fecha_admision,3,3) == "-" |
  (fecha_admision>fecha_alta)
)
head(medinexactos,10) %>%
kable(booktabs = TRUE,format = "latex")%>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_registro	id_paciente	fecha_admision	fecha_alta	costo	diagnostico	estado
427	10	2022-05-15	2020-07-12	39787.55	COVID-19	NA
290	158	2022-10-15	2021-09-18	13363.47	COVID-19	Alta
466	187	2021-09-10	2020-02-17	43023.42	Gastritis	Fallecido
3	77	2022-05-22	2020-01-29	30320.48	NA	NA
-323	67	2021-10-09	2021-03-23	10229.33	Fractura	NA
440	284	2020-07-24	2020-03-15	27822.85	Fractura	Alta
483	25	2022-05-06	2021-02-08	10740.37	Neumonía	NA
371	92	2021-07-26	2020-03-01	6234.00	Neumonía	Fallecido
-316	137	31-12-2023	2020-12-17	9009.56	NA	Hospitalizado
500	169	2021-01-19	2020-04-01	34066.87	Neumonía	Hospitalizado

Complejidad

En la tabla se observan valores faltantes en varias de las filas.

```
registros_faltantes <- registros_medicos %>%
  filter(
    if_any(everything(), is.na)
  )
head(registros_faltantes,10) %>%
kable(booktabs = TRUE,format = "latex")%>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_registro	id_paciente	fecha_admision	fecha_alta	costo	diagnostico	estado
427	10	2022-05-15	2020-07-12	39787.55	COVID-19	NA
448	71	2021-07-01	2022-04-07	11101.13	NA	Alta
483	89	2021-02-01	2023-03-26	45581.97	NA	Fallecido
3	77	2022-05-22	2020-01-29	30320.48	NA	NA
77	259	2021-09-08	2023-04-09	16519.44	Gastritis	NA
-323	67	2021-10-09	2021-03-23	10229.33	Fractura	NA
310	112	2021-10-08	2023-11-23	32861.78	Fractura	NA
246	147	2021-04-12	2021-04-23	11466.99	COVID-19	NA
40	36	2021-04-19	2022-04-19	NA	Gastritis	Fallecido
483	25	2022-05-06	2021-02-08	10740.37	Neumonía	NA

Cantidad total de datos faltantes por cada columna en el conjunto de datos.

```
na_registros <- colSums(is.na(registros_medicos))
na_registros
```

```
##      id_registro      id_paciente fecha_admision      fecha_alta      costo
##           0           0           0           1           53
##      diagnostico      estado
##           167           208
```

Como hay una considerable cantidad de datos faltantes en las variables costo, diagnostico y estado, existe la ligera sospecha de encontrar alguna fila con las 3 variables vacías.

```
med_faltante <- registros_medicos %>% filter(
  is.na(costo) & is.na(diagnostico) & is.na(estado)
)
med_faltante %>% kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_registro	id_paciente	fecha_admision	fecha_alta	costo	diagnostico	estado
195	80	2020-12-24	2022-12-04	NA	NA	NA
45	29	2023-10-26	2020-03-23	NA	NA	NA
405	160	2023-06-27	2021-06-14	NA	NA	NA

No podemos rescatar estas filas, por lo que procedemos actualizar la base de datos sin considerarlas.

```
registros_medicos1 <- registros_medicos %>% filter(
  !(is.na(costo) & is.na(diagnostico) & is.na(estado)))
```

Entonces, ajustaremos los valores faltantes (NA) de la siguiente manera:

diagnostico: Se reemplazará con “Sin diagnóstico”.

estado: Se reemplazará con “Sin determinar”.

costo: Se asignará como -1 en caso de estar ausente.

```
registros_medicos2 <- registros_medicos1 %>% mutate(
  diagnostico = case_when(
    !is.na(diagnostico) ~ diagnostico,
    is.na(diagnostico) ~ "Sin diagnóstico"
  ),
  estado = case_when(
    !is.na(estado) ~ estado,
    is.na(estado) ~ "Sin determinar"
  ),
  costo = case_when(
    !is.na(costo) ~ costo,
    is.na(costo) ~ -1
  ),
)
head(registros_medicos2, 10) %>%
  kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_registro	id_paciente	fecha_admision	fecha_alta	costo	diagnostico	estado
432	148	2021-12-22	2023-03-06	12211.04	Neumonía	Hospitalizado
427	10	2022-05-15	2020-07-12	39787.55	COVID-19	Sin determinar
290	158	2022-10-15	2021-09-18	13363.47	COVID-19	Alta
448	71	2021-07-01	2022-04-07	11101.13	Sin diagnóstico	Alta
483	89	2021-02-01	2023-03-26	45581.97	Sin diagnóstico	Fallecido
310	44	2022-09-11	2022-10-09	28769.98	Fractura	Alta
466	187	2021-09-10	2020-02-17	43023.42	Gastritis	Fallecido
3	77	2022-05-22	2020-01-29	30320.48	Sin diagnóstico	Sin determinar
77	259	2021-09-08	2023-04-09	16519.44	Gastritis	Sin determinar
-323	67	2021-10-09	2021-03-23	10229.33	Fractura	Sin determinar

Consistencia

Previamente, se identificaron inconsistencias en las variables `id_registro`, `fecha_admision` y `fecha_alta`. Para abordar estas cuestiones, se realizaron las siguientes acciones:

```
registros_medicos3 <- registros_medicos2 %>% mutate(
  id_registro = abs(id_registro),
  fecha_admision = as.Date(case_when(
    substr(fecha_admision,3,3) == "-" ~ paste( substr(fecha_admision,7,10),
                                              substr(fecha_admision,4,5),
                                              substr(fecha_admision,1,2),
                                              sep = "-"),
    substr(fecha_admision,3,3) != "-" ~ fecha_admision)
  ),
  fecha_alta = as.Date(fecha_alta),
  fecha_temporal = fecha_admision,
  fecha_admision = case_when(
    fecha_admision > fecha_alta ~ fecha_alta,
    TRUE ~ fecha_admision
  ),
  fecha_alta = case_when(
    fecha_temporal > fecha_alta ~ fecha_temporal,
    TRUE ~ fecha_alta
  )
) %>% select(-fecha_temporal)
head(registros_medicos3,10) %>%
kable(booktabs = TRUE,format = "latex")%>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_registro	id_paciente	fecha_admision	fecha_alta	costo	diagnostico	estado
432	148	2021-12-22	2023-03-06	12211.04	Neumonía	Hospitalizado
427	10	2020-07-12	2022-05-15	39787.55	COVID-19	Sin determinar
290	158	2021-09-18	2022-10-15	13363.47	COVID-19	Alta
448	71	2021-07-01	2022-04-07	11101.13	Sin diagnóstico	Alta
483	89	2021-02-01	2023-03-26	45581.97	Sin diagnóstico	Fallecido
310	44	2022-09-11	2022-10-09	28769.98	Fractura	Alta
466	187	2020-02-17	2021-09-10	43023.42	Gastritis	Fallecido
3	77	2020-01-29	2022-05-22	30320.48	Sin diagnóstico	Sin determinar
77	259	2021-09-08	2023-04-09	16519.44	Gastritis	Sin determinar
323	67	2021-03-23	2021-10-09	10229.33	Fractura	Sin determinar

Nos piden crear variables para la duración de estadía y el mes de admisión

```
registros_medicos4 <- registros_medicos3 %>%
  mutate(
    estadia_dias = as.integer(trunc(difftime(fecha_alta, fecha_admision,
                                             units = "days"))),
    mes_transaccion = format(fecha_admision, "%B")
  )
head(registros_medicos4, 10) %>%
kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_registro	id_paciente	fecha_admision	fecha_alta	costo	diagnostico	estado	estadia_dias	mes_transaccion
432	148	2021-12-22	2023-03-06	12211.04	Neumonía	Hospitalizado	439	Diciembre
427	10	2020-07-12	2022-05-15	39787.55	COVID-19	Sin determinar	672	Julio
290	158	2021-09-18	2022-10-15	13363.47	COVID-19	Alta	392	Setiembre
448	71	2021-07-01	2022-04-07	11101.13	Sin diagnóstico	Alta	280	Julio
483	89	2021-02-01	2023-03-26	45581.97	Sin diagnóstico	Fallecido	783	Febrero
310	44	2022-09-11	2022-10-09	28769.98	Fractura	Alta	28	Setiembre
466	187	2020-02-17	2021-09-10	43023.42	Gastritis	Fallecido	571	Febrero
3	77	2020-01-29	2022-05-22	30320.48	Sin diagnóstico	Sin determinar	844	Enero
77	259	2021-09-08	2023-04-09	16519.44	Gastritis	Sin determinar	578	Setiembre
323	67	2021-03-23	2021-10-09	10229.33	Fractura	Sin determinar	200	Marzo

Por último, generamos un listado de pacientes con estadía anormalmente larga o corta en pacientes dados de alta, tomando en cuenta los cuantiles 1 y 3, fuera de ese rango se ubican los datos anormales. Este intervalo será calculado por el tipo de diagnóstico detectado.

```
quantiles_diagnosticos <- registros_medicos4 %>%
  filter(estado == "Alta" & costo != -1) %>%
  select(diagnostico, estadia_dias) %>%
  group_by(diagnostico) %>%
  mutate(
    cuantil1 = quantile(estadia_dias, 0.25),
    cuantil3 = quantile(estadia_dias, 0.75),
  ) %>% distinct(diagnostico, cuantil1, cuantil3)
head(quantiles_diagnosticos, 10) %>%
kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

diagnostico	cuantil1	cuantil3
COVID-19	302.00	833.00
Sin diagnóstico	189.00	700.00
Fractura	206.00	669.75
Gastritis	145.25	802.75
Neumonía	200.25	705.25

Obteniendo la siguiente tabla

```
pacientes_alta <- registros_medicos4 %>%
  filter(estado == "Alta" & costo != -1)

estadias_anormales <- pacientes_alta %>%
  select(diagnostico, estadia_dias, id_paciente) %>%
  group_by(diagnostico) %>%
  mutate(
    cuantil1 = quantile(estadia_dias, 0.25),
    cuantil3 = quantile(estadia_dias, 0.75),
    estadia_meses = trunc(estadia_dias/30),
  ) %>% ungroup() %>%
  filter(estadia_dias < cuantil1 | estadia_dias > cuantil3) %>%
  arrange(diagnostico, desc(estadia_dias), estadia_meses) %>%
  select(diagnostico, id_paciente, estadia_dias, estadia_meses)
head(estadias_anormales, 10) %>%
kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

diagnostico	id_paciente	estadia_dias	estadia_meses
COVID-19	214	1446	48
COVID-19	288	1171	39
COVID-19	130	1121	37
COVID-19	229	967	32
COVID-19	217	953	31
COVID-19	74	931	31
COVID-19	141	927	30
COVID-19	295	903	30
COVID-19	98	889	29
COVID-19	278	859	28

También, generamos una lista de los acientes con costos sospechosos

```
costos_pacientes_sospechosos <- registros_medicos4 %>%
  filter(estado == "Alta" & costo != -1) %>%
  select(id_paciente, costo) %>%
  group_by(id_paciente) %>%
  summarise(valor_total = sum(costo)) %>%
  arrange(desc(valor_total))
head(costos_pacientes_sospechosos, 15) %>%
kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_paciente	valor_total
77	177491.26
116	146179.80
18	114212.32
260	90819.55
290	82916.20
288	78581.20
229	77354.21
96	77269.64
201	75008.36
186	74729.68
194	67136.15
61	64481.35
113	62812.32
134	61575.94
214	60493.25

Características Dependientes del sistema

Disponibilidad

Medida en que los datos están disponibles y accesibles cuando se necesitan.

https://github.com/bastiannseef/E2422/tree/main/Caso10/data_procesada_caso10

Portabilidad

Para trasladar los datos evaluados bajo el criterio de calidad de la norma ISO/IEC 25012 entre sistemas, guardamos las bases de datos procesadas en archivos “.csv” y también, en formato “.rds” para garantizar eficiencia, portabilidad y seguridad.

Encontrará los siguientes archivos:

- data_caso10.csv: Data general evaluada por los criterios de calidad en formato “.csv”.
- data_caso10.rds: Data general evaluada por los criterios de calidad en formato “.rds”.
- estadias_anormales.csv: Data estadias anormales de pacientes en formato “.csv”.
- estadias_anormales.rds: Data estadias anormales de pacientes en formato “.rds”.
- costos_pacientes_sospechosos.csv: Data costos anormales de pacientes en formato “.rds”.
- costos_pacientes_sospechosos.rds: Data costos anormales de pacientes en formato “.rds”.

```
data.frame(registros_medicos3) %>%
  saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/datos_caso10.rds")

data.frame(estadias_anormales) %>%
  saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/estadias_anormales.rds")

data.frame(costos_pacientes_sospechosos) %>%
  saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/costos_pacientes_sospechosos.rds")
```



```
saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/pacientes_sospechosos.rds")

readRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/datos_caso10.rds")
export("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/data_caso10.csv")

readRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/estadias_anormales.rds")
export("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/estadias_anormales.csv")

readRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/pacientes_sospechosos.rds")
export("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso10/data_procesada_caso10/pacientes_sospechosos.csv")
```