Actividad Formativa 03

Análisis de datos de cuestionario (univariado)

Dany Lopez (dxlopez@ul.cl) - Ximena Catalán (xrcatala@uc.cl)

Contenido

1 Objetivo

Este material tiene como propósito proporcionar una guía para desarrollar análisis univariado de distintos tipos de ítems de cuestionarios utilizados en Educación Superior. Se trabajará con el cuestionario que utilizó para transcribir las respuestas de estudiantes ver aqui. En particular, usaremos los ítems de *caracterización* (CA01–CA04) y con los ocho ítems tipo Likert del módulo *Aprendizaje* (AP01–AP08).

Esta guía está pensada para que la pueda desarrollar en clases y también como material de estudio complementario fuera del horario oficial de clases.

> **!** Importante
>
> Esta guía está preparada para abordar los análisis utilizando EXCEL en conjunto con BlueSky (opción 1), y también utilizando el lenguaje R (opción 2). Sugerimos que aquellas personas que están comenzando en esta área que vean los ejemplos de la opción 1, y aquellas personas que ya cuentan con experiencia en análisis en R, que opten por la opción 2.

2 Introducción

2.1 Resumen tipo de ítems

Ya vimos que los ítems pueden medirse en escalas **nominales** (categorías sin orden), **ordinales** (categorías con orden) o **numéricas** (intervalo/razón). Es importante que la **codificación** de respuestas de los ítems respete el nivel de medición. Como ya vimos, en ítems ordinales (por ejemplo, Likert) es recomendable que sea explicito cuál serán las **etiquetas textuales** que se usarán y en caso de utilizar alguna codificación numérica, es crucial que se declare el **orden** de las categorías (por ejemplo, por medio de un diccionario).

En todos los casos, en la siguiente tabla se entrega un resumen de los tipos de ítems, su escala de medición, codificación recomendada, los típicos análisis estadísticos usados. La última columna muestra las **Actividades** asociadas a cada ítems que se verán más abajo (ver Actividades Prácticas). Note que no todos los ítems presentan actividades en virtud del tiempo destinado en clases. Si desea ahondar en estos aspectos, consulte al equipo docente.

Table 1: Tabla resumen de items

Tipo de ítem	Nivel de medición	Codificación recomendada	Análisis univariado típicos	Pruebas estadísticas frecuentes	Actividad sugerida
Dicotómico	Nominal (binario)	0/1 o Sí/No consistentes	tabla n–%, barras	Binomial exacta o z□prop; chi□cuadrado/Fisher	Act. 1
Selección única	Nominal (k factores)	Factor/etiquetas	tabla n–%, barras	Prueba χ^2 o Fisher	Act. 2
Múltiples respuestas	Nominal multidimensional	Una columna por alternativa (0/1)	% por persona y por respuesta	Proporciones por alternativa	-
Likert (5 puntos)	Ordinal	Factor ordenado (1–5) + etiquetas	distribución factores, barras apiladas al 100%	Wilcoxon 1□muestra; Mann–Whitney/Kruskal–Wallis; χ^2 multidimensional	Act. 3
Escala numérica	Intervalo/Razón	Numérica	media, mediana, DE, Q1–Q3, histograma, caja	t de 1 muestra; t Welch/ANOVA; Wilcoxon/M□Whitney/K□Wallis	-

2.2 Análisis cuantítativo de ítems

El análisis cuantitativo de ítems con distinta tipología requiere determinar procedimientos acordes al nivel de medición y al objetivo de la indagación. En un primer nivel, tenemos el análisis univariado, que describe cada ítem por separado de manera tal de caracterizar su distribución. En ítems dicotómicos y nominales se reportan frecuencias y porcentajes. En ítems ordinales tipo Likert se examinan además de la distribución por categorías, promedios, mediana y/o rangos (aunque para la estimación de promedios y mediana se requiere una justificación acorde). En items numéricos se resumen tendencia central y dispersión (promedio, desviación estándar, rango, curtosis, entre otros).

Por su parte, el análisis bivariado estudia relaciones entre dos variables, y los métodos estadísticos para realizarlo varía según la tipología de las variables. Entre variables categóri-

cas se emplean tablas de contingencia con pruebas de contrastes como pruebas χ^2 o Fisher y medidas de asociación como el coeficiente ϕ, V de Cramér u *odds ratio*. Entre ítems ordinales o numéricas se usan correlaciones (tipo Pearson y/ Spearman) y comparaciones de grupos mediante pruebas adecuadas a los supuestos (por ejemplo, prueba t, ANOVA, entre otros). En este tipo de análisis, las visualizaciones ayudan a comunicar estos resultados, por ejemplo, gráficos de barras y/o de barras apiladas para categóricas, histogramas, *boxplot* para ítems numéricas, y gráficos de dispersión o de líneas para relaciones continuas u ordinales. Este último aspecto lo veremos al final de la unidad I.

Cuando el interés es relacional entre múltiples variables, se consideran modelos que integran varios predictores y controlan covariables. Según el caso, pueden usarse modelos lineales o generalizados (lineal tradicional para respuestas continuas, logísticos para respuestas dicotómicas, multinomiales para Likert). También, se emplean análisis factoriales para comprender la estructura subyacente de un un set de ítems, y enfoques específicos para mediciones de atributos latentes (por ejemplo, modelos de respuesta al ítem y/o teória clásica de medición).

En esta guía, comenzaremos por los análisis más básicos, pero al mismo tiempo, muy fundamentales para comprender nuestros datos. El foco de la actividad estará en el análisis univariado de ítems, para luego en las siguientes semanas, inicialr con posteriores análisis bivariados y relacionales. Como ya dijimos anteriormente, el análisis univariado consiste en describir un caso en términos de una única variable, es decir, la distribución de los atributos que lo componen. Por ejemplo:

- Si se midiera el **sexo**, observaríamos **cuántos** de los sujetos son **hombres** y cuántos son **mujeres**.
- Al consultar a una persona si es la **primera en estudiar en su familia**, observaríamos cuántos de los sujetos son los primeros estudiantes en ingresar a la universidad y cuántos no son los primeros en su familia en estudiar en la universidad.

3 Actividades prácticas

A continuación se propone una actividad de análisis univariado en dos entornos complementarios, Excel y R, orientada a describir la distribución de variables categóricas, ordinales, de múltiples respuestas y numéricas. El objetivo es producir una tabla clara junto con un breve texto interpretativo. Para realizar estas actividades, usaremos el cuestionario con el que se transcribieron las respuestas de 30 estudiantes. Comenzaremos describiendo algunos de los ítems del módulo de caracterización para luego análizar los ítems del módulo de Aprendizaje. A continuación se describe la base de datos con la que se realizarán los análisis univariados.

3.1 Descripción de los datos

Los datos que usaremos provienen del módulo de Aprendizaje del cuestionario VOCES que se aplicó a 30 estudiantes de primer año de diversas carreras de una universidad privada.

A continuación se muestran los primeros 10 registros de la base de datos que utilizaremos para realizar los análisis univariados.

```
db <- readxl::read_excel('./01BasedeDatos/01_database.xlsx',
                         sheet='Base_datos',col_names =T, na=c("","  ")
                        )

codebook <- readxl::read_excel("./01BasedeDatos/01_database.xlsx",
                               sheet='Diccionario',col_names =T, na=c("","  ")
                              )
```

folio	CA01	CA02	CA03	CA04	AP01	AP02	AP03	AP04	AP05	AP06	AP07	AP08
XXX01	F	21	2022	No	5	5	1	5	2	2	2	1
XXX02	F	21	2022	No	NA	5	1	5	2	2	2	4
XXX03	F	21	2022	Si	5	1	5	4	5	3	1	2
XXX04	M	20	2021	Si	2	4	4	1	2	3	2	2
XXX05	No indica	20	2022	No	1	3	1	1	4	4	4	2
XXX06	No indica	20	2022	Si	5	2	2	3	2	4	4	1
XXX07	No indica	19	2022	No	3	2	1	2	1	1	1	3
XXX08	No indica	19	2023	No	2	1	3	2	1	2	5	2
XXX09	No indica	20	2022	No	5	3	2	4	5	4	2	2
XXX010	No indica	21	2021	Si	4	5	2	5	4	1	1	2

También se adjunta el diccionario para comprender el significado de las columnas y las escalas de medición de cada ítem.

Caracterizacion	Folio		FOLIO	identificador		
Caracterizacion	Sexo	Sexo	CA01	categorica		F = 'Femenino' M = 'Masculino' No indica = 'Prefiero no indicar'
Caracterizacion	Edad	Edad	CA02	continua		Edad medida en años
Caracterizacion	Período ingreso carrera	Año ingreso a la carrera	CA03	Entero		Variable medida en años
Caracterizacion	Primera persona estudios universitarios	¿Eres la primera persona en tu familia en estudiar en la Universidad?	CA04	binaria		Si No
Aprendizaje	Aprendizaje Superficial	He tenido problemas para encontrar el sentido a las cosas que tengo que estudiar.	AP01	Acuerdo-Desacuerdo	5	1 = Totalmente en desacuerdo 2 = En desacuerdo 3 = Ni de acuerdo ni en desacuerdo 4 = De acuerdo 5 = Totalmente de acuerdo
Aprendizaje	Aprendizaje Superficial	Muchas de las cosas que he aprendido permanecen en mi mente como ideas sin relación.	AP02	Acuerdo-Desacuerdo	5	1 = Totalmente en desacuerdo 2 = En desacuerdo 3 = Ni de acuerdo ni en desacuerdo 4 = De acuerdo 5 = Totalmente de acuerdo
Aprendizaje	Aprendizaje Superficial	Los temas que estudiamos son presentados de una manera tan complicada que a menudo no puedo entender qué significan.	AP03	Acuerdo-Desacuerdo	5	1 = Totalmente en desacuerdo 2 = En desacuerdo 3 = Ni de acuerdo ni en desacuerdo 4 = De acuerdo 5 = Totalmente de acuerdo
Aprendizaje	Aprendizaje Superficial	Tengo que estudiar una y otra vez cosas que realmente no me hacen mucho sentido.	AP04	Acuerdo-Desacuerdo	5	1 = Totalmente en desacuerdo 2 = En desacuerdo 3 = Ni de acuerdo ni en desacuerdo 4 = De acuerdo 5 = Totalmente de acuerdo
Aprendizaje	Aprendizaje profundo	Las ideas que he encontrado	AP05	Acuerdo-Desacuerdo	5	1 = Totalmente en

También, si lo deseas puedes mirar el video (Figure 5) donde se explica la base de datos y su diccionario.

https://youtu.be/BTeLsLwsM9g

Figure 1: Explicación base de datos y diccionario.

3.2 Análisis de la Sección Caracterización

3.2.1 Análisis Item CA01

Elabore una tabla de frecuencias para el ítem CA01 que incluya la frecuencia y porcentajes por categoría de respuesta. Indique cuántos de los sujetos eran del género masculino, y cuántos del género femenino, y cuántos prefirieron no indicaron. Una vez generada la tabla con esta variable, describa las tendencias que observa.

4 En Excel

Notará que puede ser complejo realizar explicaciones que permitan replicar los análisis en Excel. Vea el video y la explicación para replicar los resultados.

https://youtu.be/WjIuFNUY_D4

Figure 2: Análisis univariado en Excel.

5 En R

5.0.0.1 Forma 1

```
tabla_CA01 <- table(db$CA01)
tabla_CA01 <- prop.table(tabla_CA01)
tabla_CA01 <-round(tabla_CA01,2)*100
tabla_CA01 <-as.data.frame.array(tabla_CA01,2)

tabla_CA01$Genero <- rownames(tabla_CA01)
names(tabla_CA01) <- c("Porcentaje","Genero")
```

5.0.0.2 Forma 2

```
library(dplyr)

tabla_sexo <- db %>%
        dplyr::select(CA01) %>%
        dplyr::group_by(CA01) %>%
        dplyr::summarise(frecuencia = n()) %>%
        dplyr::mutate(total     = sum(frecuencia),
                porcentaje = round(frecuencia/total*100,2)) %>%
        dplyr::select(CA01, frecuencia, porcentaje)
```

La tabla se muestra a continuación

CA01	frecuencia	porcentaje
F	13	43.33
M	7	23.33
No indica	10	33.33

5.0.1 Análisis Item CA04

Elabore una tabla de frecuencias para el ítem CA04 que incluya la frecuencia y porcentajes por categoría de respuesta. Indique cuántos de los sujetos son las primeras personas en sus familias en estudiar en la universidad, y cuántos no. Una vez generada la tabla con esta variable, describa las tendencias que observa.

6 En Excel

Notará que puede ser complejo realizar explicaciones que permitan replicar los análisis en Excel. Vea el video y la explicación para replicar los resultados.

https://youtu.be/S41v5JEHay0

Figure 3: Análisis univariado en Excel.

7 En R

Ahora indicaremos cuántos de los sujetos son las primeras personas en sus familias en estudiar en la universidad, y cuántos no. El código es exactamente idéntico al usado en el

caso anterior para el ítem CA01. En este caso, solo deberemos cambiar CA01 por CA04 que es el código asociado al ítem de interés.

```
library(dplyr)

tabla_primeraPersona <- db %>%
                dplyr::select(CA04) %>%
                dplyr::group_by(CA04) %>%
                dplyr::summarise(frecuencia = n()) %>%
                dplyr::mutate(total     = sum(frecuencia),
                              porcentaje = round(frecuencia/total*100,2)) %>%
                dplyr::select(CA04, frecuencia, porcentaje)
```

La tabla se muestra a continuación

CA04	frecuencia	porcentaje
No	17	56.67
Si	13	43.33

7.1 Análisis de la Sección Aprendizaje

7.1.1 Interpretación cualitativa de los ítems

Responda las siguientes preguntas:

A) ¿Qué significa que un individuo seleccione: Completamente en desacuerdo en el ítem *AP01* y Completamente de acuerdo en el ítem AP05? Utilice la información del cuestionario del módulo de Aprendizaje junto con la tabla de especificaciones para responder a esta pregunta.

B) ¿Qué significa que un individuo seleccione: Completamente de acuerdo en los ítems *AP01, AP02, AP03 y AP04* y Completamente en desacuerdo en los ítems *AP05, AP06, AP07 y AP08*? Utilice la información del cuestionario del módulo de Aprendizaje junto con la tabla de especificaciones para reportar este caso.

7.1.2 Análisis Item AP01

Indique la proporción de sujetos que manifiestan distintos niveles de acuerdo con la siguiente afirmación: "He tenido problemas para encontrarle sentido a las cosas que tengo que estudiar. Elabore una tabla de frecuencias para el ítem AP01 que incluya la frecuencia y porcentajes por categoría de respuesta. Una vez generada la tabla con esta variable, describa las tendencias que observa.

8 En Excel

Notará que puede ser complejo realizar explicaciones que permitan replicar los análisis en Excel. Vea el video y la explicación para replicar los resultados.

https://youtu.be/grPsNCvui-s

Figure 4: Análisis univariado en Excel.

9 En R

Ahora indicaremos la proporción de sujetos que manifiestan distintos niveles de acuerdo con la siguiente afirmación: *"He tenido problemas para encontrarle sentido a las cosas que tengo que estudiar*. En este caso, al código usado solo deberemos cambiar CA04 por AP01

```
library(dplyr)

tabla_itemAP01 <- db %>%
                  dplyr::select(AP01) %>%
                  dplyr::group_by(AP01) %>%
                  dplyr::summarise(frecuencia = n()) %>%
                  dplyr::mutate(total     = sum(frecuencia),
                                porcentaje = round(frecuencia/total*100,2)) %>%
                  dplyr::select(AP01, frecuencia, porcentaje)
```

La tabla se muestra a continuación

AP01	frecuencia	porcentaje
1	5	16.67
2	7	23.33
3	6	20.00
4	3	10.00
5	8	26.67
NA	1	3.33

Podemos optimizar un poco más este código. Sería ideal que en la columna AP01 apareciera el enunciado del ítem, y que los niveles de la escala Likert, en vez de que aparezca de manera numérica, que aparezca la categoría que expresa cada nivel (que ya sabemos que estos números en realidad son solo etiquetas; note que esto aplica si solo si las digitación de la escala Likert se hace de manera numérica y no digitando explicitamente cada nivel en su forma original). Es decir nos gustaría hacer la siguiente conversión:

- 1 = Totalmente en Deacuerdo
- 2 = En desacuerdo
- 3 = Ni de acuerdo ni en Desacuerdo
- 4 = De acuerdo
- 5 = Completamente de acuerdo

Podemos crear un dataframe con esta información para luego utilizarla en la tabla que construimos anteriormente.

```
library(dplyr)

conversion_likert <- data.frame( likert_numero = c(1,
                                                   2,
                                                   3,
                                                   4,
                                                   5,
                                                   'NA'),
                                 likert_texto = c('Totalmente en Desacuerdo',
                                                  'En desacuerdo',
                                                  'Ni de acuerdo ni en Desacuerdo',
                                                  'De acuerdo',
                                                  'Completamente de acuerdo',
                                                  'Respuesta perdida')
                                 )
```

La construcción del dataframe se llama conversion_likert y se muestra a continuación

likert_numero	likert_texto
1	Totalmente en Desacuerdo
2	En desacuerdo
3	Ni de acuerdo ni en Desacuerdo
4	De acuerdo
5	Completamente de acuerdo
NA	Respuesta perdida

Entonces, ahora reemplazamos los valores numéricos de la escala Likert por su grado de acuerdo correspondiente. Para ello, utilizamos la función left_join()

```
library(dplyr)

tabla_itemAP01.editado <- tabla_itemAP01 %>%
                  dplyr::left_join(. ,conversion_likert,
                                   by=join_by(AP01 == likert_numero)) %>%
```

```
                dplyr::arrange(AP01) %>%
               select(likert_texto, frecuencia, porcentaje)
```

La tabla queda de la siguiente forma

likert_texto	frecuencia	porcentaje
Totalmente en Desacuerdo	5	16.67
En desacuerdo	7	23.33
Ni de acuerdo ni en Desacuerdo	6	20.00
De acuerdo	3	10.00
Completamente de acuerdo	8	26.67
Respuesta perdida	1	3.33

Finalmente, si quisieramos incluir el enunciado del ítem (*He tenido problemas para encontrarle sentido a las cosas que tengo que estudiar*) en vez de likert_texto, nos aprovecharemos de la información que se encuentra en nuestro diccionario. Veremos que en nuestro diccionario, la columna Pregunta tiene la información del enunciado para cada ítem. Luego, el siguiente código extrae el enunciado del ítem (Pregunta) y la reemplaza en la tabla construida anteriormente.

```
library(dplyr)

names(tabla_itemAP01.editado)[1] <- codebook %>%
                    dplyr::filter(item_codigo=='AP01') %>%
                    dplyr::select(Pregunta)
```

La tabla finalmente queda de la siguiente forma

He tenido problemas para encontrar el sentido a las cosas que tengo que estudiar.	frecuencia	porcentaje
Totalmente en Desacuerdo	5	16.67
En desacuerdo	7	23.33
Ni de acuerdo ni en Desacuerdo	6	20.00
De acuerdo	3	10.00
Completamente de acuerdo	8	26.67
Respuesta perdida	1	3.33

9.0.1 Análisis Items AP01-AP08

¿Qué proporción de respuestas se seleccionó en cada categoría para cada ítem del módulo Aprendizaje? Para responder a la pregunta, presente *una sola tabla* que exprese los recuentos y porcentajes por categoría de respuesta para cada uno de los ítems que compone el módulo. Notará que ahora lo que se le solicita es que realice múltiples análisis univariados para luego resumir toda la información en una sola tabla ¿Cómo organizaría la información?

10 En Excel

Este ejercicio puede resolverse aplicando el mismo procedimiento utilizado en el ítem AP01 y repitiéndolo para cada ítem.

11 En Excel y Bluesky

El enfoque únicamente con Excel, salvo que se disponga de Power Query, puede resultar laborioso. A continuación se presenta una forma más automatizada de aplicar el mismo procedimiento, generalizable a un número indefinido de ítems del mismo tipo. Para ello, usamos como herramienta secundaria el software BlueSky Statistics

https://youtu.be/Yj6KVqEj5G4

Figure 5: Análisis univariado múltiple ítems en Excel y en BlueSky.

12 En R

- Procesamiento de datos
- Convertimos la base de wide a long

```
library(dplyr)

# Extraemos los codigos de los items del modulo Aprendizaje
items  <- codebook %>%
            dplyr::filter(Dimension=='Aprendizaje') %>%
            dplyr::select(item_codigo)


# Convertimos todos los items AP01-AP08 double a caracter (dado que tenemos NA en AP01 codificado c

db  <- db %>% dplyr::mutate_if(is.double, as.character)

# Convertimos tabla de WIDE a LONG
db_long<-tidyr::pivot_longer(db,
                      cols = items$item_codigo,
                      names_to = 'items',
                      values_to = 'valores'
                     )
```

- Estadística descriptiva

```
library(dplyr)

tabla_resumen <- db_long %>%
            dplyr::select(items, valores) %>%
            dplyr::group_by(items, valores) %>%
              dplyr::summarise(frecuencia = n()) %>%
            dplyr::ungroup() %>%
            dplyr::group_by(items) %>%
              dplyr::mutate(total = sum(frecuencia),
                            porcentaje = round(frecuencia/total*100,2)
                            ) %>%
            dplyr::ungroup() %>%
            dplyr::select(items, valores, porcentaje) %>%
            tidyr::pivot_wider(., names_from  = 2,
                               values_from = 3)
```

El resultado es el siguiente

items	1	2	3	4	5	NA
AP01	16.67	23.33	20.00	10.00	26.67	3.33
AP02	23.33	13.33	16.67	23.33	23.33	
AP03	33.33	23.33	10.00	6.67	26.67	
AP04	20.00	16.67	10.00	23.33	30.00	
AP05	16.67	30.00	20.00	13.33	20.00	
AP06	16.67	30.00	10.00	23.33	20.00	
AP07	16.67	26.67	6.67	23.33	26.67	
AP08	16.67	26.67	20.00	26.67	10.00	

Sin embargo, hay un par de aspectos en la tabla que pueden mejorarse. Por ejemplo, sería ideal que en la columna items apareciera el enunciado del ítem, y que los niveles de la escala Likert en vez de que aparezca de manera numérica que aparezca la categoría que expresa cada nivel (que ya sabemos que estos números en realidad son solo etiquetas; note que esto aplica si solo si las digitación de la escala Likert se hace de manera numérica y no digitando explicitamente cada nivel en su forma original). Es decir nos gustaría hacer la siguiente conversión:

- 1 = Totalmente en Deacuerdo
- 2 = En desacuerdo
- 3 = Ni de acuerdo ni en Desacuerdo
- 4 = De acuerdo
- 5 = Completamente de acuerdo

Una manera de hacer lo anterior podría ser exportar en Excel (o en texto plano) la tabla de resumen para luego reemplazar manualmente el nombre de los ítems y sus categorías. Sin embargo, lo anterior podría generar un error humano o amenazar la replicabilidad de los resultados (recuerde que lo ideal es siempre dejar un registro de todo el flujo de análisis de datos, por lo que al usar R podemos dejar un registro de todo este flujo). Otra forma más robusta y que que alinea con la replicabilidad de los resultados se consigue utilizando nuestro diccionario. Veamos cómo integrar la información del diccionario en nuestro código.

Comenzaremos mostrando cómo cambiar el nombre de los ítems AP01, AP02,..., AP07, AP08 por los enunciados respectivos. A continuación se explican las líneas extras que se agregaron para incluir el enunciado en vez de los códigos AP01, AP02,..., AP07, AP08. Lo ideal es volver un poco atrás e incoporar la informacion del ítem en la tabla LONG que creamos. Luego

```
library(dplyr)

db_long_expand<-db_long %>%
            dplyr::left_join(                                          ①
                    codebook[, c("item_codigo", "Pregunta")],          ②
                     by = join_by(items == item_codigo),               ③
                    keep = FALSE) %>%                                  ④
             dplyr::relocate(Pregunta,                                 ⑤
                         .before  = items)
```

La tabla Long se ve así

folio	CA01	CA02	CA03	CA04	Pregunta
XXX01	F	21	2022	No	He tenido problemas para encontrar el sentido a las cosas que tengo que estudiar.
XXX01	F	21	2022	No	Muchas de las cosas que he aprendido permanecen en mi mente como ideas sin relación.
XXX01	F	21	2022	No	Los temas que estudiamos son presentados de una manera tan complicada que a menudo no pu
XXX01	F	21	2022	No	Tengo que estudiar una y otra vez cosas que realmente no me hacen mucho sentido.

Y ahora, considerar solo las dos columnas Pregunta y Valores para generar la tabla

```
library(dplyr)

tabla_resumen_v2<-db_long_expand %>%
            dplyr::select(Pregunta, valores) %>%
            dplyr::group_by(Pregunta, valores) %>%
             dplyr::summarise(frecuencia = n()) %>%
            dplyr::ungroup() %>%
            dplyr::group_by(Pregunta) %>%
             dplyr::mutate(total = sum(frecuencia),
```

```
            porcentaje = round(frecuencia/total*100,2)
            ) %>%
  dplyr::ungroup() %>%
  dplyr::select(Pregunta, valores, porcentaje) %>%
  tidyr::pivot_wider(., names_from  = 2,
                     values_from = 3)
```

Pregunta	1	2	3	4	5	NA
He tenido problemas para encontrar el sentido a las cosas que tengo que estudiar.	16.67	23.33	20.00	10.00	26.67	3.33
Intento relacionar lo que he aprendido en un curso con lo aprendido en otros.	16.67	26.67	20.00	26.67	10.00	
Las ideas que he encontrado en mis lecturas académicas me han hecho pensar y reflexionar profundamente sobre los temas que estoy aprendiendo.	16.67	30.00	20.00	13.33	20.00	
Los temas que estudiamos son presentados de una manera tan complicada que a menudo no puedo entender qué significan.	33.33	23.33	10.00	6.67	26.67	
Mientras voy leyendo material nuevo, trato de relacionarlo con lo que ya sé sobre el tema.	16.67	26.67	6.67	23.33	26.67	
Muchas de las cosas que he aprendido permanecen en mi mente como ideas sin relación.	23.33	13.33	16.67	23.33	23.33	
Observo cuidadosamente la evidencia para llegar a mi propia conclusión sobre lo que estoy estudiando.	16.67	30.00	10.00	23.33	20.00	
Tengo que estudiar una y otra vez cosas que realmente no me hacen mucho sentido.	20.00	16.67	10.00	23.33	30.00	

Finalmente, si ahora quisieramos incorporar la siguiente conversión:

- 1 = Totalmente en Deacuerdo
- 2 = En desacuerdo
- 3 = Ni de acuerdo ni en Desacuerdo
- 4 = De acuerdo
- 5 = Completamente de acuerdo

necesitamos invertir un poco más de esfuerzos para lograr una total replicabilidad y automatización de nuestros análisis. Necesitamos, por lo tanto, un tabla que contenga esta información. Esta tabla, por tanto, requiere que incluyamos más información a nuestro diccionario, lo que se logra incoporando una nueva hoja en nuestro EXCEL con la especificación de la escala de medición. Lo que se muestra a continuación:

```
escala_med <- readxl::read_excel('./01BasedeDatos/01_database.xlsx',
                  sheet='Escala_medicion_diccionario',col_names =T, na=c("","  ")
                 )
```

A continuación, se muestra la tabla con la información necesaria para hacer la conversión

escala_medicion	numero_niveles	nivel_item_numero	nivel_item_texto
Acuerdo-Desacuerdo	4	1	Totalmente en desacuerdo
Acuerdo-Desacuerdo	4	2	En desacuerdo
Acuerdo-Desacuerdo	4	3	De acuerdo
Acuerdo-Desacuerdo	4	4	Totalmente de acuerdo
Acuerdo-Desacuerdo	5	1	Totalmente en desacuerdo
Acuerdo-Desacuerdo	5	2	En desacuerdo
Acuerdo-Desacuerdo	5	3	Ni de acuerdo ni en desacuerdo
Acuerdo-Desacuerdo	5	4	De acuerdo
Acuerdo-Desacuerdo	5	5	Totalmente de acuerdo

Note que podríamos tener dos tipos diferentes de escalas likert de acuerdo/desacuerdo. Una que tenga 4 niveles y otra que tenga 5 niveles. Para este cuestionario, usaremos la que tiene 5 niveles.

```
library(dplyr)

escala_med_4nivel <- escala_med %>%
                dplyr::filter( escala_medicion=='Acuerdo-Desacuerdo' & numero_niveles== 5) %>%
                dplyr::select(nivel_item_numero, nivel_item_texto)
```

Luego, usamos esta información para cambiar el nombre de las columnas de la tabla resumen

```
names(tabla_resumen_v2)[2:(length(tabla_resumen_v2)-1)] <- escala_med_4nivel$nivel_item_texto
```

La tabla final, se muestra a continuación

Pregunta	Totalmente en desacuerdo	En desacuerdo	Ni de acuer
He tenido problemas para encontrar el sentido a las cosas que tengo que estudiar.	16.67	23.33	
Intento relacionar lo que he aprendido en un curso con lo aprendido en otros.	16.67	26.67	
Las ideas que he encontrado en mis lecturas académicas me han hecho pensar y reflexionar profundamente sobre los temas que estoy aprendiendo.	16.67	30.00	
Los temas que estudiamos son presentados de una manera tan complicada que a menudo no puedo entender qué significan.	33.33	23.33	
Mientras voy leyendo material nuevo, trato de relacionarlo con lo que ya sé sobre el tema.	16.67	26.67	
Muchas de las cosas que he aprendido permanecen en mi mente como ideas sin relación.	23.33	13.33	
Observo cuidadosamente la evidencia para llegar a mi propia conclusión sobre lo que estoy estudiando.	16.67	30.00	
Tengo que estudiar una y otra vez cosas que realmente no me hacen mucho sentido.	20.00	16.67	

De esta forma, solo tendríamos que editar nuestro diccionario y/o la especificación de la escala de medición que se encuentra en nuestro diccionario, y no el código. Lo que aquí

se construyó fue un sistema en el cual el diccionario se integra con la base de datos por medio del código. Luego, con este sistema es posible levantar una infraestructura de datos estandarizado en elque cada una de las partes (código + diccionario + base de datos),se integran armonicamente. Una gran ventaja de esta aproximación enfocado en el análisis de custionarario, es que cualquier modificación o ajuste que se quiera hacer al diccionario no implica alterar en absoluto el código para generar los resultados. En esa línea, este paradigma funciona por componentes o de manera modular lo que hace más fácil identificar en qué lugar (o módulo) podriamos llegar a tener potenciales errores que arreglar. Dos desvetajas de esta aproximación son: 1) puede resultar poco intuitivo si no se encuentra familiarizado con una programación orientada a componentes; 2) que necesita de manera obligatoria saber cómo se organizan y estructuran sus ítems en su cuestionario, que se soluciona una vez que se tiene la tabla de especicaciones.