Analysis

Your Name

2025-04-09

Table of contents

```
library(readxl)
library(gt)

data = read_excel("data.xlsx")
colnames(data)
```

```
 [1] "sexo"                    "edad"                    "estado_civil"            "univer
 [5] "grupo_carrera"           "ciclo"                   "promedio"                "nse"
 [9] "roles"                   "vive_con"                "ed_padre"                "ed_mad
[13] "sit_laboral_padres"      "region"                  "tipo_universidad"        "fp_hab
[17] "fp_soportesocial2"       "fp_habilidadessociales3" "fp_soportesocial4"       "fp_pla
[21] "fp_metas6"               "fp_soportesocial7"       "fp_planear8"             "fp_met
[25] "fp_metas10"              "fp_metas11"              "fp_habilidadessociales12" "fp_hab
[29] "fp_metas14"              "fp_planear15"            "fp_habilidadessociales16" "fp_hab
[33] "fp_planear18"            "fp_metas19"              "fp_soportesocial20"      "fp_sop
[37] "fp_planear22"            "fp_soportesocial23"      "fp_planear24"            "fo_con
[41] "fo_commitment2b"         "fo_expectance2c"         "fo_exploration3.1"       "fo_exp
[45] "fo_commitment5a"         "fo_commitment5b"         "fo_commitment5c"         "fo_con
[49] "fo_expectance7"          "fo_exploration8a"        "fo_exploration8b"        "fo_exp
[53] "fo_exploration8d"        "fo_exploration8e"        "fo_internalcontrol9a"    "fo_int
[57] "fo_internalcontrol9c"    "fo_internalcontrol9h"    "fo_expectance10a"        "fo_exp
[61] "fo_expectance10c"        "fo_expectance10d"        "fo_expectance10e"        "fo_exp
[65] "fo_value11a"             "fo_value11b"             "fo_value11c"             "fo_val
[69] "fo_value11e"             "fo_hopes3"               "fo_hopes4"               "fo_hop
[73] "fo_fears3"               "fo_fears4"               "fo_fears8"               "desere
[77] "tutoria"                 "habilidades_sociales"    "soporte_social"          "planea
[81] "metas"                   "internal_control"        "expectance"              "value
```

```
[85] "hopes"                    "fears"                    "exploration"              "commit
```

```r
corstars <- function(x, round_digits = 2, use = "pairwise.complete.obs") {
  require(Hmisc)

  # Compute correlation matrix with p-values
  rcorr_res <- Hmisc::rcorr(as.matrix(x), type = "pearson")
  r <- rcorr_res$r
  p <- rcorr_res$P

  # Create significance stars
  stars <- ifelse(p < 0.001, "***",
           ifelse(p < 0.01, "**",
           ifelse(p < 0.05, "*", "")))

  # Combine correlations and stars
  r_stars <- matrix(paste0(formatC(r, format = "f", digits = round_digits), stars),
                    nrow = nrow(r))
  rownames(r_stars) <- colnames(x)
  colnames(r_stars) <- colnames(x)

  # Set diagonal to 1.00 without NA or stars
  diag(r_stars) <- formatC(1, format = "f", digits = round_digits) # Explicitly set diagonal

  # Means and SDs
  means <- sapply(x, function(i) mean(i, na.rm = TRUE))
  sds <- sapply(x, function(i) sd(i, na.rm = TRUE))

  result <- as.data.frame(r_stars)
  result <- tibble::rownames_to_column(result, var = "Variable")
  result <- dplyr::left_join(
    tibble::tibble(Variable = names(means),
                   Mean = formatC(means, digits = round_digits, format = "f"),
                   SD = formatC(sds, digits = round_digits, format = "f")),
    result,
    by = "Variable"
  )

  return(result)
}
```

Limpieza de datos

```
data = data %>%
    mutate(
        sexo = ifelse(sexo == "Prefiero no decir", NA, sexo),
        estado_civil = ifelse(estado_civil == "NA", NA, estado_civil),
        estado_civil = ifelse(estado_civil == 'Soltero/a', "Soltero/a", "No Soltero"),
        grupo_carrera = ifelse(grupo_carrera == "NA", NA, grupo_carrera),
        promedio = parse_number(promedio),
        nse = ifelse(nse == 'Bajo', "Bajo", "Medio/Alto"),
        ed_padre = ifelse(ed_padre == "NA", NA, ed_padre),
        ed_madre = ifelse(ed_madre == "NA", NA, ed_madre),
        universidad = ifelse(universidad == "NA", NA, universidad),
        ed_padre = ifelse(ed_padre == "sin educacion", "secundaria", ed_padre),
        ed_madre = ifelse(ed_madre == "sin educacion", "secundaria", ed_madre),
        ed_padre = ifelse(ed_padre == "secundaria", "secundaria o menos", ed_padre),
        ed_madre = ifelse(ed_madre == "secundaria", "secundaria o menos", ed_madre)
        ) %>%
    mutate_at(vars(desercion, habilidades_sociales:metas), parse_number)
totales = data %>% select(sexo:tipo_universidad, desercion, habilidades_sociales:metas)
```

He agrupado algunos datos, porque los tamaños de la muestra en cada subgrupo son muy pequeños. Cuando no se usan para el analisis, los he mantenido (e.g., `vive_con`, `situacion_laboral_padres`).

Datos para metodo

```
totales %>% count(sexo)
```

```
# A tibble: 3 x 2
  sexo       n
  <chr>  <int>
1 Hombre   230
2 Mujer    532
3 <NA>       6
```

```r
mean(totales$edad)
```

```
[1] 21.45833
```

```r
sd(totales$edad)
```

```
[1] 2.538534
```

```r
min(totales$edad)
```

```
[1] 18
```

```r
max(totales$edad)
```

```
[1] 30
```

```r
totales %>% count(estado_civil)
```

```
# A tibble: 3 x 2
  estado_civil     n
  <chr>        <int>
1 No Soltero      28
2 Soltero/a      736
3 <NA>             4
```

```r
totales %>% count(universidad)
```

```
# A tibble: 9 x 2
  universidad                                                n
  <chr>                                                  <int>
1 Pontificia Universidad Católica del Perú                 179
2 Universidad Católica de Santa María                      181
3 Universidad Católica de Trujillo Benedicto XVI           139
4 Universidad Nacional Agraria de la Selva - Tingo María    55
5 Universidad Nacional Jorge Basadre Grohmann               14
6 Universidad Nacional de San Martín                        18
7 Universidad Nacional de Ucayali                           59
8 Universidad Nacional de la Amazonía Peruana              122
9 <NA>                                                       1
```

```r
totales %>% count(grupo_carrera)
```

```
# A tibble: 7 x 2
  grupo_carrera           n
  <chr>               <int>
1 arte y arquitectura    50
2 educación              27
3 gestion               155
4 humanidades            85
5 ingenieria            164
6 salud                 285
7 <NA>                    2
```

```r
totales %>% count(ciclo)
```

```
# A tibble: 9 x 2
  ciclo       n
  <chr>   <int>
1 Cuarto     64
2 Décimo     64
3 Noveno     78
4 Octavo     60
5 Quinto    119
6 Segundo    62
7 Sexto      75
8 Séptimo   111
9 Tercer    135
```

```r
mean(totales$promedio, na.rm = T)
```

```
[1] 14.74678
```

```r
sd(totales$promedio, na.rm = T)
```

```
[1] 1.810755
```

```r
min(totales$promedio, na.rm = T)
```

```
[1] 9.857
```

```r
max(totales$promedio, na.rm = T)
```

```
[1] 20
```

```r
totales %>% count(nse)
```

```
# A tibble: 2 x 2
  nse            n
  <chr>      <int>
1 Bajo         269
2 Medio/Alto   499
```

```r
totales %>% count(roles)
```

```
# A tibble: 2 x 2
  roles                n
  <chr>            <int>
1 estudia y trabaja   447
2 solo estudia        321
```

```r
totales %>% count(vive_con)
```

```
# A tibble: 4 x 2
  vive_con                                                          n
  <chr>                                                         <int>
1 Además de mi pareja y/o hijo(a)s, vivo con otros familiares      18
2 Además de mis padres y/o hermanos, vivo con otros familiares    150
3 Vivo con mi pareja y/o hijos                                     31
4 Vivo con mis padres y/o hermano(a)s                             569
```

```r
totales %>% count(ed_padre)
```

```
# A tibble: 5 x 2
  ed_padre              n
  <chr>             <int>
1 post grado           67
2 secundaria o menos   234
3 tecnico              234
4 universitario        206
5 <NA>                  27
```

```
totales %>% count(ed_madre)
```

```
# A tibble: 5 x 2
  ed_madre               n
  <chr>              <int>
1 post grado            61
2 secundaria o menos   239
3 tecnico              279
4 universitario        181
5 <NA>                   8
```

```
totales %>% count(sit_laboral_padres)
```

```
# A tibble: 4 x 2
  sit_laboral_padres        n
  <chr>                 <int>
1 No trabaja               35
2 Trabajo dependiente     353
3 Trabajo doméstico        43
4 Trabajo independiente   337
```

```
totales %>% count(region)
```

```
# A tibble: 3 x 2
  region     n
  <chr>  <int>
1 costa    332
2 selva    254
3 sierra   182
```

```
totales %>% count(tipo_universidad)
```

```
# A tibble: 2 x 2
  tipo_universidad     n
  <chr>            <int>
1 privada            499
2 publica            269
```

Missing data

Vamos a hacer listwise deletion. Voy a seleccionar todas las variables que entran al analisis, y nos quedamos unicamente con los datos completos.

Las variables demograficas que quedan son: `sexo edad estado_civil grupo_carrera nse ed_padre ed_madre region tipo_universidad`

```
totales = totales %>% select(-universidad, -ciclo, -roles, -vive_con, -sit_laboral_padres) %>
```

Correlations

```
totales %>% select(edad, promedio, desercion:metas) %>% corstars() %>% gt()
```

```
Loading required package: Hmisc
```

```
Attaching package: 'Hmisc'
```

```
The following object is masked from 'package:gt':

    html
```

```
The following objects are masked from 'package:dplyr':

    src, summarize
```

```
The following objects are masked from 'package:base':

    format.pval, units
```

Variable	Mean	SD	edad	promedio	desercion	habilidades_sociales	soporte_s
edad	21.33	2.52	1.00	-0.15***	0.03	0.12**	
promedio	14.71	1.79	-0.15***	1.00	-0.11*	0.02	0.1
desercion	2.43	1.76	0.03	-0.11*	1.00	-0.16***	-0.1
habilidades_sociales	5.00	1.42	0.12**	0.02	-0.16***	1.00	0.5
soporte_social	4.81	1.19	-0.00	0.16***	-0.17***	0.50***	
planear	5.19	1.25	0.08	0.11*	-0.23***	0.63***	0.5
metas	5.43	1.33	0.12**	0.06	-0.23***	0.71***	0.5

Proposed Model

```
library(lavaan)
```

```
This is lavaan 0.6-19
lavaan is FREE software! Please report any bugs.
```

```
model <- '
  # Direct effects
  metas ~ habilidades_sociales + soporte_social
  planear ~ metas + habilidades_sociales + soporte_social + promedio
  desercion ~ planear + promedio
'

fit <- sem(model, data = totales, estimator = "ML", missing = "listwise")

summary(fit, fit.measures = TRUE, standardized = TRUE, rsquare = TRUE)
```

```
lavaan 0.6-19 ended normally after 1 iteration

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        11

  Number of observations                           536

Model Test User Model:

  Test statistic                                 5.356
```

```
  Degrees of freedom                                 4
  P-value (Chi-square)                           0.253

Model Test Baseline Model:

  Test statistic                              1087.486
  Degrees of freedom                                12
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.999
  Tucker-Lewis Index (TLI)                       0.996

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2314.362
  Loglikelihood unrestricted model (H1)      -2311.684

  Akaike (AIC)                                4650.725
  Bayesian (BIC)                              4697.850
  Sample-size adjusted Bayesian (SABIC)       4662.933

Root Mean Square Error of Approximation:

  RMSEA                                          0.025
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.074
  P-value H_0: RMSEA <= 0.050                    0.748
  P-value H_0: RMSEA >= 0.080                    0.028

Standardized Root Mean Square Residual:

  SRMR                                           0.015

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
```