

SOC3070 Análisis de Datos Categóricos

Trabajo 2

Información

Ponderación: 20% de la nota final del curso. Entrega: Desde el momento de entrega, los estudiantes tienen 2 semanas exactas de plazo para completar esta tarea. Responder las preguntas *bonus* NO es un requisito necesario para obtener puntaje completo. Responder incorrectamente las preguntas *bonus* no afectará negativamente la nota obtenida, pero responderlas correctamente mejorará la nota obtenida en un máximo de 0.7 puntos (o en la cantidad necesaria para obtener nota máxima si la nota original fuera superior a 6.3)

Introducción

En esta tarea usarán regresión logística multinomial (MNL) y regresión logística ordenada (OLR) para re-analizar un subconjunto de los datos utilizados en el artículo *“It’s not just how the game is played, it’s whether you win or lose”* (2019). Este estudio utiliza un experimento online para identificar el efecto causal de las desigualdades de oportunidades y de resultados sobre creencias acerca de las causas de la desigualdad y percepciones de justicia.

En particular, deberán estimar los efectos de las dos manipulaciones experimentales del estudio sobre las probabilidades de atribuir los resultados de la competencia al talento (“talent”), la suerte (“luck”) o circunstancias externas al individuo (“rules”). Estas respuestas se encuentran en la variable LTC. Una de las manipulaciones experimentales es el status de ganador o perdedor de los participantes (variable W). La otra es el nivel de desigualdad de oportunidades en la competencia, el cual es determinado por el tratamiento al que cada individuo fue asignado (variable T). Para esta tarea usarán el subconjunto de datos correspondientes a los tratamientos RA (“random exchange”, es decir, “level of redistribution” = 0), RE1 (“regressive exchange” con “level of redistribution” = 1) y RE2 (“regressive exchange” con “level of redistribution” = 2).

Para mayor contexto pueden revisar el artículo en el link indicado en el repositorio. Igualmente, los datos están disponibles en el repositorio del curso para ser descargados.

```
data_paper <- read_csv("data_t_2.csv") %>% mutate(W=if_else(W=="TRUE",1,0))
data_paper %>% glimpse()
```

```
## Rows: 660
## Columns: 8
## $ groupId <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2~
## $ playerId <dbl> 2, 1, 1, 2, 1, 2, 2, 1, 1, 2, 2, 1, 2, 1, 1, 2, 2, 1, 2, 1, 1~
## $ fair <chr> "Fair", "Unfair", "Unfair", "Fair", "Fair", "Fair", "Unfair", ~
## $ T <chr> "RA", "RA", "RA", "RA", "RA", "RA", "RA", "RA", "RA", "RA", "RA", "~
## $ W <dbl> 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1~
## $ LTC <chr> "luck", "luck", "luck", "luck", "luck", "luck", "luck", "luck", "luck~
## $ age <dbl> 27, 53, 33, 33, 33, 28, 25, 46, 30, 23, 30, 44, 60, 65, 33, 3~
## $ gender <chr> "Female", "Female", "Male", "Female", "Female", "Male", "Male~
```

Ejercicios

1. Usa regresión logística multinomial para estimar el modelo implícito en el panel derecho de la figura 2 (interacción entre variables T y W). Presenta un `summary()` de los resultados. Usa LTC = "rules" como categoría de referencia en la variable dependiente.

```
data_paper$LTC <- factor(data_paper$LTC, levels = c("rules","talent","luck"))
mnlr_1 <- multinom(LTC ~ T*W, data=data_paper)
```

```
## # weights:  21 (12 variable)
## initial  value 725.084111
## iter   10 value 613.113952
## final   value 608.467858
## converged
```

```
summary(mnlr_1)
```

```
## Call:
## multinom(formula = LTC ~ T * W, data = data_paper)
##
## Coefficients:
##      (Intercept)      TRE1      TRE2      W      TRE1:W      TRE2:W
## talent    0.1744688 -1.344576 -2.501769  1.1297749  0.11179555 -0.07875193
## luck      1.7430681 -1.303751 -1.873106 -0.3434807 -0.09584526 -0.02956029
##
## Std. Errors:
##      (Intercept)      TRE1      TRE2      W      TRE1:W      TRE2:W
## talent    0.2960105 0.4829239 0.6016807 0.3931272 0.6096407 0.7318028
## luck      0.2365513 0.3356499 0.3288228 0.3486892 0.5023573 0.4852516
##
## Residual Deviance: 1216.936
## AIC: 1240.936
```

1.1 Interpreta los coeficientes asociados a TRE1 y TRE2:W.

El coeficiente asociado a TRE1-talent indica que las log-odds de un un perdedor del tratamiento RE1 atribuya los resultados al talento en vez de a la reglas del juego son 1.35 puntos menores que la de los perdedores del tratamiento RA. Por su parte, el coeficiente asociado a TRE1-luck indica que las log-odds de que un perdedor del tratamiento RE1 atribuya los resultados a la suerte en vez de a la reglas del juego son 1.30 puntos menores que la de los perdedores del tratamiento RA.

En tanto, el coeficiente asociado a la interacción TRE2:W-talent indica que las log-odds de atribuir los resultados al talento en vez de a la reglas del juego por parte de los ganadores del tratamiento RE2 son 2.6 puntos menores que la de los ganadores del tratamiento RA ($\beta_{RE2-talent} + \beta_{RE2:W-talent} = -2.501769 - 0.07875193 = -2.580521$) y 1 punto mayores que la de los perdedores del tratamiento RE2 ($\beta_{W-talent} + \beta_{RE2:W-talent} = 1.1297749 + -0.07875193 = 1.051023$).

Asimismo, el coeficiente asociado a la interacción TRE2:W-luck indica que las log-odds de atribuir los resultados a la suerte en vez de a la reglas del juego por parte de los ganadores del tratamiento RE2 son 1.9 puntos menores que la de los ganadores del tratamiento RA ($\beta_{RE2-luck} + \beta_{RE2:W-luck} = -1.873106 - 0.02956029 = -1.902666$) y 0.4 puntos menores que la de los perdedores del tratamiento RE2 ($\beta_{W-luck} + \beta_{RE2:W-luck} = -0.3434807 - 0.02956029 = -0.373041$).

1.2 Transforma e interpreta los coeficientes correspondiente RE1-talent y RE1-luck en términos de odds-ratios (o probabilidades relativas). Presenta el desarrollo formal de una de estas odds-ratios.

Transformamos los coeficientes en odds ratios exponenciando los coeficientes originales:

```
exp(summary(mnlr_1)$coefficients)[,"TRE1"]
```

```
##      talent      luck
## 0.2606502 0.2715113
```

Formalmente:

$$\ln \frac{\mathbb{P}(\text{talent} \mid T=\text{RE1}, W=0)}{\mathbb{P}(\text{rules} \mid T=\text{RE1}, W=0)} = \beta_0 + \beta_{RE1}$$

Por tanto:

- $\frac{\mathbb{P}(\text{talent} \mid T=\text{RE1}, W=0)}{\mathbb{P}(\text{rules} \mid T=\text{RE1}, W=0)} = e^{\beta_0 - \text{talent} + \beta_{RE1} - \text{talent}} = e^{\beta_0 - \text{talent}} e^{\beta_{RE1} - \text{talent}}$

y

- $\frac{\mathbb{P}(\text{talent} \mid T=\text{RA}, W=0)}{\mathbb{P}(\text{rules} \mid T=\text{RA}, W=0)} = e^{\beta_0 - \text{talent}}$

se sigue de esto que

$$\frac{\mathbb{P}(\text{talent} \mid T=\text{RA}, W=0)}{\mathbb{P}(\text{rules} \mid T=\text{RA}, W=0)} \bigg/ \frac{\mathbb{P}(\text{talent} \mid T=\text{RE1}, W=0)}{\mathbb{P}(\text{rules} \mid T=\text{RE1}, W=0)} = e^{\beta_0 - \text{talent}} e^{\beta_{RE1} - \text{talent}} / e^{\beta_0 - \text{talent}} = e^{\beta_{RE1} - \text{talent}} = e^{-1.344576} = 0.26$$

es decir, el ratio entre la probabilidad de atribuir los resultados al talento y la probabilidad de atribuirlos a las reglas por parte de los perdedores del tratamiento RE1 es 1/4 el ratio observado para los perdedores del tratamiento RA.

1.3. Calcula la odds ratio de atribuir los resultados al talento en vez de a la suerte entre los perdedores del tratamiento RA vs los perdedores de RE2.

$$e^{\beta_{RE2} - \text{talent} - \beta_{RE2} - \text{luck}} = e^{(-2.501769 - -1.873106)} = 0.5333044$$

1.4 Manipula los resultados del modelo para obtener las probabilidades esperadas de que los participantes del tratamiento RA – tanto ganadores como perdedores – sostengan que los resultados de la competencia se deben principalmente al “talento”. Expresa formalmente las ecuaciones correspondiente a estas predicciones.

En términos generales sabemos que:

$$p_{ji} = \frac{e^{\beta_{j0} + \beta_{j1}x_{1i} + \dots + \beta_{jk}x_{ki}}}{1 + \sum_{j=1}^{J-1} e^{\beta_{j0} + \beta_{j1}x_{1i} + \dots + \beta_{jk}x_{ki}}}$$

Por tanto:

- $\mathbb{P}(\text{talent} \mid T=\text{RA}, W=0) = \frac{e^{\beta_0 - \text{talent}}}{1 + e^{\beta_0 - \text{talent}} + e^{\beta_0 - \text{luck}}} = e^{0.175} / (1 + e^{0.175} + e^{1.743}) = 0.15$
- $\mathbb{P}(\text{talent} \mid T=\text{RA}, W=1) = \frac{e^{\beta_0 - \text{talent} + \beta_W - \text{talent}}}{1 + e^{(\beta_0 - \text{talent} + \beta_W - \text{talent})} + e^{(\beta_0 - \text{luck} + \beta_W - \text{luck})}} = e^{0.175 + 1.13} / (1 + e^{0.175 + 1.13} + e^{1.743 - 0.343}) = 0.422$

Implementación en R:

```
newx <- data_paper %>% data_grid(T="RA", W, .model=mnlr_1)
newy <- cbind(newx, predict(mnlr_1, newdata = newx, type = "probs")) %>% dplyr::select(talent)
print(newy)
```

```
##      talent
## 1 0.1506064
## 2 0.4216893
```

1.5 Agrega un control por age al modelo estimado en 1. Presenta un summary() de los resultados.

```
mnlr_2 <- multinom(LTC ~ T*W + age, data=data_paper)
```

```
## # weights: 24 (14 variable)
## initial value 722.886886
```

```
## iter 10 value 609.186694
## iter 20 value 605.726489
## final value 605.726480
## converged
```

```
summary(mnlr_2)
```

```
## Call:
## multinom(formula = LTC ~ T * W + age, data = data_paper)
##
## Coefficients:
##      (Intercept)      TRE1      TRE2      W      age      TRE1:W
## talent  -0.3531519 -1.329350 -2.485030  1.1409125  0.0135904933  0.10567708
## luck     1.7593388 -1.326548 -1.873661 -0.3444602 -0.0004312425 -0.07253906
##      TRE2:W
## talent -0.10809386
## luck   -0.02791497
##
## Std. Errors:
##      (Intercept)      TRE1      TRE2      W      age      TRE1:W
## talent   0.5093950 0.4834789 0.6022103 0.3944464 0.010673148 0.6107055
## luck     0.4127734 0.3365347 0.3290754 0.3490795 0.008916198 0.5030607
##      TRE2:W
## talent 0.7334839
## luck   0.4862249
##
## Residual Deviance: 1211.453
## AIC: 1239.453
```

1.6. (Bonus 1) De acuerdo al modelo estimado en 1.5., ¿cuál es *efecto* (marginal) de edad sobre la probabilidad de que un perdedor de 20 años en el tratamiento RA atribuya los resultados al talento?

Sabemos que:

$$\bullet \frac{\partial p_{ij}}{\partial x_k} = p_{ij} \cdot \left(\beta_{jk} - \sum_{j=1}^{J-1} \beta_{jk} \cdot p_{ij} \right)$$

y

$$\bullet p_{ji} = \frac{e^{\beta_{j0} + \beta_{j1}x_{1i} + \dots + \beta_{jk}x_{ki}}}{1 + \sum_{j=1}^{J-1} e^{\beta_{j0} + \beta_{j1}x_{1i} + \dots + \beta_{jk}x_{ki}}}$$

Por tanto:

$$\begin{aligned} \mathbb{P}(\text{talent} \mid T=\text{RA}, W=0) &= \frac{e^{\beta_0 - \text{talent} + 20\beta_{\text{age} - \text{talent}}}}{1 + (e^{\beta_0 - \text{talent} + 20\beta_{\text{age} - \text{talent}}}) + (e^{\beta_0 - \text{luck} + 20\beta_{\text{age} - \text{luck}}})} \\ &= e^{-0.35 + 20 \cdot 0.014} / (1 + (e^{-0.35 + 20 \cdot 0.014}) + (e^{1.76 - 20 \cdot 0.00})) \\ &= 0.12 \end{aligned}$$

$$\begin{aligned}
\mathbb{P}(\text{luck} \mid T=\text{RA}, W=0) &= \frac{e^{\beta_0 - \text{talent} + 20\beta_{\text{age} - \text{talent}}}}{1 + (e^{\beta_0 - \text{talent} + 20\beta_{\text{age} - \text{talent}}}) + (e^{\beta_0 - \text{luck} + 20\beta_{\text{age} - \text{luck}}})} \\
&= e^{1.76 - 20 \cdot 0.00} / (1 + (e^{-0.35 + 20 \cdot 0.014}) + (e^{1.76 - 20 \cdot 0.00})) \\
&= 0.75
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathbb{P}(\text{talent} \mid T=\text{RA}, W=0, \text{age}=20)}{\partial \text{age}} &= p_{\text{talent } i} \cdot (\beta_{\text{age} - \text{talent}} - (\beta_{\text{age} - \text{talent}} * p_{\text{talent } i} + \beta_{\text{age} - \text{luck}} * p_{\text{luck } i})) \\
&= 0.12 * (0.0135904933 - (0.0135904933 * 0.12 + -0.0004312425 * 0.75)) \\
&= 0.0014784
\end{aligned}$$

Aproximación numérica:

```

delta = 0.1
grid <- data_paper %>% data_grid(T="RA",W=0,age=20, .model=mnlr_2)
grid_delta <- data_paper %>% data_grid(T="RA",W=0,age=20, .model=mnlr_2) %>% mutate(age = age + delta)

p_hat <- predict(mnlr_2, type="probs", newdata =grid) %>% as_tibble() %>%
  mutate(id = c("rules","talent","luck")) %>% rename(p_hat = value)

p_hat_delta <- predict(mnlr_2, type="probs", newdata =grid_delta) %>% as_tibble() %>%
  mutate(id = c("rules","talent","luck")) %>% rename(p_hat_delta = value)

p_hat_delta %>% left_join(p_hat, by=c("id")) %>% mutate(me_age = (p_hat_delta - p_hat)/delta)

## # A tibble: 3 x 4
##   p_hat_delta id      p_hat    me_age
##   <dbl> <chr>   <dbl>    <dbl>
## 1    0.130 rules   0.130 -0.000170
## 2    0.120 talent 0.120  0.00148
## 3    0.750 luck   0.750 -0.00130

```

2. Usa regresión logística ordinal para estimar el modelo implícito en el panel derecho de la figura 2 (interacción entre variables T y W). Presenta un `summary()` de los resultados. Usa el siguiente orden para la variable LTC: “rules”, “luck”, “talent” (de menos a más “meritocrático”).

```
data_paper$LTC <- ordered(data_paper$LTC, levels = c("rules","luck","talent"))
ologit_1 <- polr(LTC ~ T*W, data=data_paper)
summary(ologit_1)
```

```
##
## Re-fitting to get Hessian
## Call:
## polr(formula = LTC ~ T * W, data = data_paper)
##
## Coefficients:
##              Value Std. Error t value
## TRE1    -0.75329    0.2549 -2.9551
## TRE2    -1.39112    0.2620 -5.3087
## W         0.84860    0.2097  4.0461
## TRE1:W  -0.07632    0.3730 -0.2046
## TRE2:W  -0.74746    0.3739 -1.9990
##
## Intercepts:
##              Value Std. Error t value
## rules|luck  -1.3238  0.1568  -8.4446
## luck|talent  1.1972  0.1543   7.7586
##
## Residual Deviance: 1262.90
## AIC: 1276.90
```

2.1 Interpreta los coeficientes asociados a TRE1 y TRE2:W.

El coeficiente asociado a TRE1 indica que, en promedio, los perdedores del tratamiento RE1 tienen 0.8 punto menos en la variable latente de atribución meritocrática comparado con los perdedores del tratamiento RA.

En tanto, el coeficiente asociado a la interacción TRE2:W indica que los ganadores del tratamiento RE2 tienen, en promedio, 2.1 puntos menos en la variable latente de atribución meritocrática comparado con los ganadores del tratamiento RA ($\beta_{RE2} + \beta_{RE2:W} = -1.39112 + -0.74746 = -2.13858$) y 0.1 punto más que los perdedores del tratamiento RE2 ($\beta_W + \beta_{RE2:W} = 0.84860 + -0.74746 = 0.10114$).

2.2 Transforma e interpreta el coeficientes correspondiente a RE1 en términos de odds-ratios acumulativas . Presenta el desarrollo formal de dicha odds-ratio.

Para los perdedores del tratamiento RE1 el logit acumulado de la probabilidad de atribuir los resultados del juego a la suerte u otro motivo menos meritocrático (“rules” o “luck”) está dado por:

$$\ln \frac{\mathbb{P}(y \leq \text{luck} \mid T=\text{RE1}, W=0)}{\mathbb{P}(y > \text{luck} \mid T=\text{RE1}, W=0)} = \alpha_{\text{luck}|\text{talent}} - \beta_{RE1}$$

mientras que para los perdedores del tratamiento RA el mismo logit está dado por:

$$\ln \frac{\mathbb{P}(y \leq \text{luck} \mid T=\text{RA}, W=0)}{\mathbb{P}(y > \text{luck} \mid T=\text{RA}, W=0)} = \alpha_{\text{luck}|\text{talent}}$$

Por tanto, las correspondientes odds son:

$$\frac{\mathbb{P}(y \leq \text{luck} \mid T=\text{RE1}, W=0)}{\mathbb{P}(y > \text{luck} \mid T=\text{RE1}, W=0)} = e^{\alpha_{\text{luck}|\text{talent}}} \cdot e^{-\beta_{RE1}}$$

y

$$\frac{\mathbb{P}(y \leq \text{luck} \mid T=\text{RA}, W=0)}{\mathbb{P}(y > \text{luck} \mid T=\text{RA}, W=0)} = e^{\alpha_{\text{luck}|\text{talent}}}$$

Se sigue de esto que

$$\frac{\mathbb{P}(y \leq \text{luck} \mid T=\text{RE1}, W=0)/\mathbb{P}(y > \text{luck} \mid T=\text{RE1}, W=0)}{\mathbb{P}(y \leq \text{luck} \mid T=\text{RA}, W=0)/\mathbb{P}(y > \text{luck} \mid T=\text{RA}, W=0)} = e^{-\beta_{RE1}}$$

Este coeficiente los obtenemos de la siguiente forma:

```
exp(-(summary(ologit_1)$coefficients["TRE1",1]))
```

```
## [1] 2.12398
```

Es decir, las odds acumulada de atribuir los resultados del juego a la suerte u otro motivo menos meritocrático (“rules” o “luck”) de los perdedores del tratamiento RE1 son aproximadamente el doble de las odds de los perdedores del tratamiento RA.

2.3 Manipula los resultados del modelo para obtener las probabilidades esperadas de que los participantes del tratamiento RA – tanto ganadores como perdedores – sostengan que los resultados de la competencia se deben principalmente al “talento”. Expresa formalmente las ecuaciones correspondiente a estas predicciones.

$$\mathbb{P}(y = \text{talent} \mid T=\text{RA}, W=0) = 1 - \text{logit}^{-1}(\alpha_{\text{luck}|\text{talent}}) = 1 - 1/(1 + e^{-\alpha_{\text{luck}|\text{talent}}}) = 0.2319737$$

$$\mathbb{P}(y = \text{talent} \mid T=\text{RA}, W=0) = 1 - \text{logit}^{-1}(\alpha_{\text{luck}|\text{talent}} - \beta_W) = 1 - 1/(1 + e^{-(\alpha_{\text{luck}|\text{talent}} - \beta_W)}) = 0.4137196$$

Implementación en R:

```
newx <- data_paper %>% data_grid(T="RA",W,.model=ologit_1)
newy <- cbind(newx,predict(ologit_1, newdata = newx, type = "probs")) %>% dplyr::select(talent)
print(newy)
```

```
##      talent
## 1 0.2319727
## 2 0.4137196
```

2.4 Agrega un control por age al modelo estimado en 2. Presenta un summary() de los resultados.

```
ologit_2 <- polr(LTC ~ T*W + age, data=data_paper)
summary(ologit_2)
```

```
##
## Re-fitting to get Hessian
## Call:
## polr(formula = LTC ~ T * W + age, data = data_paper)
##
## Coefficients:
##              Value Std. Error t value
## TRE1      -0.762472  0.255964 -2.9788
## TRE2     -1.378425  0.262486 -5.2514
## W           0.846928  0.210196  4.0292
## age         0.008694  0.006685  1.3005
## TRE1:W    -0.058039  0.374110 -0.1551
## TRE2:W    -0.765743  0.375174 -2.0410
##
## Intercepts:
##              Value Std. Error t value
## rules|luck  -0.9962  0.2950   -3.3765
## luck|talent  1.5272  0.2985    5.1154
##
## Residual Deviance: 1258.10
```

```
## AIC: 1274.10
## (2 observations deleted due to missingness)
```

2.5. (Bonus 2) De acuerdo al modelo estimado en 2.5., ¿cuál es *efecto* (marginal) de edad sobre la probabilidad de que un perdedor de 20 años en el tratamiento RA atribuya los resultados al talento? Interpreta este resultado.

Talento es la categoría más alta, por tanto:

$$\frac{\partial \mathbb{P}(y_i = \text{talent} \mid T = \text{RA}, W = 0, \text{age} = 20)}{\partial \text{age}} = \beta_{\text{age}} \cdot \left(\mathbb{P}(y_i \leq \text{luck} \mid T = \text{RA}, W = 0, \text{age} = 20) \cdot (1 - \mathbb{P}(y_i \leq \text{luck} \mid T = \text{RA}, W = 0, \text{age} = 20)) \right)$$

$$\mathbb{P}(y_i \leq \text{luck} \mid T = \text{RA}, W = 0, \text{age} = 20) = \text{logit}^{-1}(\alpha_{\text{luck}|\text{talent}} - \beta_{\text{age}} 20) = 0.7946719$$

Por tanto,

$$\frac{\partial \mathbb{P}(y_i = \text{talent} \mid T = \text{RA}, W = 0, \text{age} = 20)}{\partial \text{age}} = 0.009 * 0.8 * (1 - 0.8) = 0.00144$$

Aproximación numérica

```
delta = 0.1
grid <- data_paper %>% data_grid(T="RA",W=0,age=20, .model=ologit_2)
grid_delta <- data_paper %>% data_grid(T="RA",W=0,age=20, .model=ologit_2) %>% mutate(age = age + delta)

p_hat <- predict(ologit_2, type="probs", newdata =grid) %>% as_tibble() %>%
  mutate(id = c("rules","luck","talent")) %>% rename(p_hat = value)

p_hat_delta <- predict(ologit_2, type="probs", newdata =grid_delta) %>% as_tibble() %>%
  mutate(id = c("rules","luck","talent")) %>% rename(p_hat_delta = value)

p_hat_delta %>% left_join(p_hat, by=c("id")) %>% mutate(me_age = (p_hat_delta - p_hat)/delta)

## # A tibble: 3 x 4
##   p_hat_delta id      p_hat      me_age
##   <dbl> <chr>   <dbl>      <dbl>
## 1     0.237 rules  0.237 -0.00157
## 2     0.558 luck   0.558  0.000152
## 3     0.205 talent 0.205  0.00142
```