

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) 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, donde W=1 corresponde a los ganadores). 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.

```
path <- url("https://raw.githubusercontent.com/mebucca/cda_soc3070/master/homework/t_2/data_t_2.csv")
data_paper <- read_csv(path) %>% 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, 2, 3, 3, 3, 3, 3, 3,
## $ playerId <dbl> 2, 1, 1, 2, 1, 2, 2, 1, 1, 2, 2, 1, 2, 1, 1, 2, 2, 1, 2, 1, 1, 2, 1, 2, 1, 2, 2, 1,
## $ fair <chr> "Fair", "Unfair", "Unfair", "Fair", "Fair", "Fair", "Unfair", "Unfair", "Fair", "Un
## $ T <chr> "RA", "RA", "RA", "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, 0, 0, 1, 0, 1, 1, 0,
## $ LTC <chr> "luck", "luck", "luck", "luck", "luck", "luck", "luck", "luck", "luck", "luck", "luck", "lu
## $ age <dbl> 27, 53, 33, 33, 33, 28, 25, 46, 30, 23, 30, 44, 60, 65, 33, 33, 45, 35, 34, 25, 61,
## $ gender <chr> "Female", "Female", "Male", "Female", "Female", "Male", "Male", "Male", "Male", "Female", "
```

Ejercicios

1. Usa regresión logística multinomial para estimar modelar las atribuciones de los resultados a las reglas/talento/suerte en función del estatus de ganador/perdedor y el tratamiento(variables T y W sin interacción). 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 ~ W + T, data=data_paper)
```

```
## # weights:  15 (8 variable)
## initial  value 725.084111
## iter   10 value 618.652365
## final   value 608.561074
## converged
```

```
summary(mnlr_1)
```

```
## Call:
## multinom(formula = LTC ~ W + T, data = data_paper)
##
## Coefficients:
##      (Intercept)           W      TRE1      TRE2
## talent    0.1622638  1.1390192 -1.252996 -2.555178
## luck      1.7585717 -0.3791515 -1.342930 -1.886472
##
## Std. Errors:
##      (Intercept)           W      TRE1      TRE2
## talent    0.2469750  0.2538146  0.2873209  0.3364911
## luck      0.1965451  0.1990359  0.2493574  0.2417923
##
## Residual Deviance: 1217.122
## AIC: 1233.122
```

1.1 Interpreta los coeficientes asociados a W-talent y TRE2-luck.

El coeficiente asociado a W-talent indica que las log-odds de que un ganador atribuya los resultados al talento en vez de a la reglas del juego son 1.14 puntos menores que la de los perdedores. Por su parte, el coeficiente asociado a TRE2-luck indica que las log-odds de que un jugador en el tratamiento RE2 atribuya los resultados a la suerte en vez de a la reglas del juego son 1.89 puntos menores que la de los jugadores del tratamiento RA.

1.2 Transforma e interpreta el coeficiente correspondiente a W-talent en términos de odds-ratios (o probabilidades relativas) y presenta el desarrollo formal de dicha odds-ratio.

Transformamos el coeficiente en odds ratios exponenciando el original:

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

```
## [1] 3.123703
```

Formalmente:

$$\ln \frac{\mathbb{P}(\text{talent} \mid W=1, T=t)}{\mathbb{P}(\text{rules} \mid W=1, T=t)} = \beta_{0-\text{talent}} + \beta_{W-\text{talent}} + \beta_{T-\text{talent}} * t$$

y

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

Por tanto:

$$\frac{\mathbb{P}(\text{talent} \mid W=1, T=t)}{\mathbb{P}(\text{rules} \mid W=1, T=t)} = e^{\beta_{0-\text{talent}}} \cdot e^{\beta_{W-\text{talent}}} \cdot e^{\beta_{T-\text{talent}}*t}$$

y

$$\frac{\mathbb{P}(\text{talent} \mid W=0, T=t)}{\mathbb{P}(\text{rules} \mid W=0, T=t)} = e^{\beta_{0-\text{talent}}} \cdot e^{\beta_{T-\text{talent}}*t}$$

se sigue de esto que

$$\frac{\mathbb{P}(\text{talent} \mid W=1, T=t)}{\mathbb{P}(\text{rules} \mid W=1, T=t)} \bigg/ \frac{\mathbb{P}(\text{talent} \mid W=0, T=t)}{\mathbb{P}(\text{rules} \mid W=0, T=t)} = e^{\beta_{0-\text{talent}}} \cdot e^{\beta_{W-\text{talent}}} \cdot e^{\beta_{T-\text{talent}}*t} / e^{\beta_{0-\text{talent}}} \cdot e^{\beta_{T-\text{talent}}*t} = e^{\beta_{W-\text{talent}}} = e^{1.1390192} = 3.123703$$

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 ganadores es más de 3 veces el ratio observado para los perdedores.

1.3. Calcula la odds ratio de atribuir los resultados al talento en vez de a la suerte entre ganadores y perdedores.

$$e^{\beta_{W-\text{talent}} - \beta_{W-\text{luck}}} = e^{(1.1390192 - -0.3791515)} = 1.518171$$

1.4 Manipula los resultados del modelo para obtener las probabilidades esperadas de que los ganadores del tratamiento RA sostengan que los resultados de la competencia se deben principalmente al “talento”. Expresa formalmente la ecuación correspondiente a esta predicción.

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:

$$\bullet \mathbb{P}(\text{talent} \mid T=\text{RA}, W=1) = \frac{e^{\beta_{0-\text{talent}} + \beta_{W-\text{talent}}}}{1 + e^{(\beta_{0-\text{talent}} + \beta_{W-\text{talent}})} + e^{(\beta_{0-\text{luck}} + \beta_{W-\text{luck}})}} = e^{0.1622638 + 1.1390192} / (1 + e^{0.1622638 + 1.1390192} + e^{1.7585717 - 0.3791515}) = 0.425$$

Implementación en R:

```
newx <- data_paper %>% data_grid(T="RA", W=1, .model=mnlr_1)
newy <- cbind(newx, predict(mnlr_1, newdata = newx, type = "probs"))
print(newy)
```

```
##           T W predict(mnlr_1, newdata = newx, type = "probs")
## rules   RA 1                      0.1156523
## talent  RA 1                      0.4249075
## luck    RA 1                      0.4594402
```

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

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

```
## # weights:  18 (10 variable)
## initial value 722.886886
## iter  10 value 608.696881
## final value 605.804429
## converged
```

```
summary(mnlr_2)
```

```
## Call:
## multinom(formula = LTC ~ W + T + age, data = data_paper)
##
## Coefficients:
##           (Intercept)           W           TRE1           TRE2           age
## talent  -0.3603167   1.1459554 -1.245324 -2.560805   0.0135646702
```

```
## luck      1.7741887 -0.3739084 -1.356209 -1.886723 -0.0004814554
##
## Std. Errors:
##      (Intercept)          W      TRE1      TRE2      age
## talent    0.4755874 0.2547221 0.2879200 0.3375763 0.010656556
## luck      0.3845134 0.1991846 0.2497436 0.2417767 0.008898293
##
## Residual Deviance: 1211.609
## AIC: 1231.609
```

1.6. (Bonus 1) De acuerdo al modelo estimado en 1.5., ¿cuál es *efecto* (marginal) de edad sobre la probabilidad de que un ganador de 30 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, calculamos la probabilidad en cuestión:

```
newx <- data_paper %>% data_grid(T="RA", W=1, age=30, .model=mnlr_2)
newy <- cbind(newx, predict(mnlr_2, newdata = newx, type = "probs"))
print(newy)
```

```
##      T W age predict(mnlr_2, newdata = newx, type = "probs")
## rules RA 1 30      0.1205729
## talent RA 1 30      0.3973563
## luck  RA 1 30      0.4820708
```

Luego, el efecto marginal de edad sobre la probabilidad de

$$\begin{aligned} \frac{\partial \mathbb{P}(\text{talent} \mid T=\text{RA}, W=1, \text{age}=30)}{\partial \text{age}} &= p_{\text{talent } i} \cdot (\beta_{\text{age}-\text{talent}} - (\beta_{\text{age}-\text{talent}} * p_{\text{talent } i} + \beta_{\text{age}-\text{luck}} * p_{\text{luck } i})) \\ &= 0.3973563 * (0.0135646702 - (0.3973563 * 0.0135646702 + -0.0004814554 * 0.4820708)) \\ &= 0.003340479 \end{aligned}$$

Implementación en R:

```
marginalEffects(mnlr_2, variables="age", newdata = newx)
```

```
##   rowid type group term      dydx   std.error statistic   p.value   conf.low   conf.high
## 1     1 probs rules  age -0.0006219113 0.0009431454 -0.6594013 0.5096381 -0.002470442 0.001226620
## 2     1 probs talent age  0.0033405062 0.0020599573  1.6216385 0.1048808 -0.000696936 0.007377948
## 3     1 probs luck  age -0.0027185949 0.0019625483 -1.3852372 0.1659799 -0.006565119 0.001127929
## predicted predicted_hi predicted_lo T W age LTC eps
## 1 0.1205729    0.1205693    0.1205729 RA 1 30 luck 0.0058
## 2 0.3973563    0.3973757    0.3973563 RA 1 30 luck 0.0058
## 3 0.4820708    0.4820551    0.4820708 RA 1 30 luck 0.0058
```

Es decir, para un ganador de 30 años en el tratamiento RA un aumento en un año de edad se traduce en un aumento en 0.003340479 puntos porcentuales en la probabilidad de atribuir los resultados al talento.

- Usando los datos de infidelidad, estima un modelo Poisson para la tasa de infidelidad a lo largo del matrimonio.

```
library("Ecdat")
data(Fair)
affairsdata <- Fair %>% as_tibble()
```

- La variable `nbaffairs` mide la cantidad de relaciones extra-matrimoniales que ha tenido una persona. La variable `ym` mide los años que una persona ha estado casada.
- Modela la cantidad de relaciones extra-matrimoniales que ha tenido una persona como función de su género (`age`), su felicidad en el matrimonio (`rate`) y la interacción entre ambas.

```
##
## Call:
## glm(formula = nbaffairs ~ rate * factor(sex), family = poisson(link = log),
##      data = affairsdata, offset = log(ym))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.8483  -1.7400  -0.9306  -0.3538   9.8012
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.63898    0.12575  -5.081 3.75e-07 ***
## rate           -0.31686    0.03650  -8.682 < 2e-16 ***
## factor(sex)male    0.32991    0.18779   1.757  0.079 .
## rate:factor(sex)male -0.08003    0.05408  -1.480   0.139
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 2828.8  on 600  degrees of freedom
## Residual deviance: 2662.4  on 597  degrees of freedom
## AIC: 3164.4
##
## Number of Fisher Scoring iterations: 7
```

10.1 Interpreta el efecto multiplicativo de la variable felicidad en el matrimonio sobre la tasa de infidelidad de los hombres.

Un aumento en una unidad de evaluación de felicidad en el matrimonio disminuye en un 33% ($e^{-0.31686-0.08003} - 1$) la tasa de infidelidad de los hombres.

10.2 Estima la cantidad esperada de relaciones extra-matrimoniales para una mujer en un matrimonio infeliz (`rate=2`) de 20 años de duración.

$$\mu_i = \theta_i \cdot n_i$$

$$\mu_i = e^{(\beta_0 + 2\beta_{rate})} \cdot ym_i$$

$$\mu_i = e^{(-0.63898 + 2 \cdot -0.31686)} \cdot 20 = 5.601488$$

Implementación en R:

```
predict(model_inf, newdata = data.frame(rate=2, sex="female", ym=20), type="response")
```

```
##      1
## 5.601511
```