SOC3070 Análisis de Datos Categóricos

Trabajo 1

Información

Ponderación: 20% de la nota final del curso. Entrega: Desde el momento de entrega, los estudiantes tienen 3 semanas exactas de plazo para completar esta tarea. Responder la pregunta *bonus* NO es un requisito necesario para obtener puntaje completo. Responder incorrectamente la pregunta *bonus* no afectará negativamente la nota obtenida, pero responderla 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 el modelo lineal de probabilidad (LPM) y regresión logística 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 la probabilidad de percibir que los resultados de la competencia son "justos" (variable `fair`). 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_1.csv")
data_paper %>% glimpse()
```

```
## Rows: 660
## Columns: 7
## $ groupId  <dbl> 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...
## $ fair     <chr> "Fair", "Unfair", "Unfair", "Fair", "Fair", "Fair", "Unfai...
## $ T        <chr> "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...
## $ age      <dbl> 27, 53, 33, 33, 33, 28, 25, 46, 30, 23, 30, 44, 60, 65, 33...
## $ gender   <chr> "Female", "Female", "Male", "Female", "Female", "Male", "M...
```

Ejercicios

1. Usa un LPM para estimar el modelo implícito en el panel derecho de la figura 1 del artículo (interacción entre variables `T` y `W`). Como categorías de referencia usa `T==RA` y `W==0`. Presenta un `summary()` de los resultados.

```
lpm_1 <- lm(1*(fair=="Fair") ~ T*W, data=data_paper)
summary(lpm_1)
```

```
##
## Call:
## lm(formula = 1 * (fair == "Fair") ~ T * W, data = data_paper)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8313 -0.3916 -0.1358  0.3614  0.8642
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.39157    0.03448  11.357  < 2e-16 ***
## TRE1        -0.09036    0.05972  -1.513    0.131
## TRE2        -0.25576    0.06021  -4.248 2.47e-05 ***
## W            0.43976    0.04876   9.019  < 2e-16 ***
## TRE1:W      -0.10241    0.08445  -1.213    0.226
## TRE2:W      -0.11877    0.08515  -1.395    0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4442 on 654 degrees of freedom
## Multiple R-squared:  0.2178, Adjusted R-squared:  0.2119
## F-statistic: 36.43 on 5 and 654 DF,  p-value: < 2.2e-16
```

1.1 Interpreta el intercepto, el coeficiente asociado a `TRE1` y el coeficiente asociado a `TRE2:W`.

El intercepto corresponde a la probabilidad esperada de que un perdedor del tratamiento RA considere que los resultados del juego son justos. Dicha probabilidad es 0.39. El coeficiente asociado a `RE1` indica que la probabilidad esperada de que un perdedor del tratamiento RE1 perciba justicia es 9 puntos porcentuales menor que la un perdedor del tratamiento RA. En tanto, el coeficiente asociado a la interacción `TRE2:W` indica que la probabilidad de percibir justicia de los ganadores del tratamiento RE2 es $\beta_{RE2} + \beta_{RE2:W} = -0.37453$ puntos porcentuales (menor que) la de los ganadores del tratamiento RA y $\beta_W + \beta_{RE2:W} = 0.32$ puntos porcentuales mayor la de los perdedores del tratamiento RE2.

1.2 Manipula los resultados del modelo para obtener las probabilidades esperadas de que los ganadores – en cada uno de los tratamientos – sostengan que los resultados de la competencia son "justos" (los puntos azules en el panel derecho de la figura 1). Expresa formalmente las ecuaciones correspondiente a estas predicciones.

Formalmente:

- $\mathbb{P}(\text{Fair} \mid \text{T=RA, W=1}) = \beta_0 + \beta_W$
- $\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=1}) = \beta_0 + \beta_W + \beta_{RE1} + \beta_{RE1:W}$
- $\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=1}) = \beta_0 + \beta_W + \beta_{RE2} + \beta_{RE2:W}$

Implementación en R:

```
newx <- data_paper %>% data_grid(T,W=1,.model=lpm_1)
newy <- newx %>% mutate(pred_prob = predict(lpm_1, newdata = newx))
print(newy)
```

```
## # A tibble: 3 x 3
##   T         W pred_prob
##   <chr> <dbl>     <dbl>
## 1 RA        1     0.831
## 2 RE1       1     0.639
## 3 RE2       1     0.457
```

1.3. Agrega un control por `age` al modelo estimado en 1. Presenta un `summary()` de los resultados e interpreta el efecto de edad.

```
lpm_2 <- lm(1*(fair=="Fair") ~ T*W + age, data=data_paper)
summary(lpm_2)
```

```
##
## Call:
## lm(formula = 1 * (fair == "Fair") ~ T * W + age, data = data_paper)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97583 -0.36099 -0.06054  0.37496  0.91883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.194690   0.067828   2.870 0.004233 **
## TRE1        -0.080819   0.059581  -1.356 0.175427
## TRE2        -0.247604   0.059850  -4.137 3.98e-05 ***
## W            0.445920   0.048551   9.185  < 2e-16 ***
## age          0.005157   0.001534   3.363 0.000817 ***
## TRE1:W      -0.112677   0.084127  -1.339 0.180920
## TRE2:W      -0.133846   0.084749  -1.579 0.114747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4412 on 651 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.2296, Adjusted R-squared:  0.2225
## F-statistic: 32.33 on 6 and 651 DF,  p-value: < 2.2e-16
```

El coeficiente asociado a `age` indica que, independiente del tratamiento y de ganar o perder la competencia, cada año adicional de edad está asociado con un aumento de medio punto porcentual en la probabilidad de percibir que los resultados de la competencia son justos.

1.4 De acuerdo al modelo estimado en 1.3., ¿cuál sería la probabilidad esperada de que un perdedor de 10 años de edad en el tratamiento RE2 sostenga que los resultados de la competencia son "justos"?. Expresa formalmente la ecuación correspondiente a esta predicción.

$$
\begin{aligned}
\mathbb{P}(\text{Fair} \mid T = \text{RE2}, W = 0, age = 20) &= \beta_0 + \beta_{RE2} + \beta_{age} age && (1)\\
\mathbb{P}(\text{Fair} \mid T = \text{RE2}, W = 0, age = 20) &= 0.194690 - 0.247604 + 0.005157 \cdot 10 && (2)\\
&= -0.001344 && (3)
\end{aligned}
$$

2. Usa regresión logística para estimar el modelo implícito en el panel derecho de la figura 1 (interacción entre variables `T` y `W`). Presenta un `summary()` de los resultados.

```
logit_1 <- glm(1*(fair=="Fair") ~ T*W, family = binomial(link=logit), data=data_paper)
summary(logit_1)
```

```
##
## Call:
## glm(formula = 1 * (fair == "Fair") ~ T * W, family = binomial(link = logit),
##     data = data_paper)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8867  -0.9969  -0.5403   0.9472   1.9983
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.4407     0.1590  -2.772  0.00558 **
## TRE1         -0.4008     0.2873  -1.395  0.16293
## TRE2         -1.4099     0.3612  -3.903 9.50e-05 ***
## W             2.0358     0.2612   7.793 6.56e-15 ***
## TRE1:W       -0.6251     0.4215  -1.483  0.13808
## TRE2:W       -0.3585     0.4724  -0.759  0.44801
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 914.95  on 659  degrees of freedom
## Residual deviance: 759.13  on 654  degrees of freedom
## AIC: 771.13
##
## Number of Fisher Scoring iterations: 4
```

2.1 Interpreta el intercepto, el coeficiente asociado a `TRE1` y el coeficiente asociado a `TRE2:W`.

El intercepto corresponde al logaritmo natural de las odds (log-odds) de que un perdedor del tratamiento RA considere que los resultados del juego son justos. Dichas log-odds es -0.44. El coeficiente asociado a `RE1` indica que, en el caso de los perdedores del tratamiento RE1, las log-odds esperadas de percibir justicia es 0.4 puntos menor que la de los perdedores del tratamiento RA. En tanto, el coeficiente asociado a la interacción `TRE2:W` indica que las log-odds de percibir justicia de los ganadores del tratamiento RE2 es $\beta_{RE2} + \beta_{RE2:W} = -1.77$ puntos (menores que) la de los ganadores del tratamiento RA y $\beta_W + \beta_{RE2:W} = 1.68$ puntos mayores que la de los perdedores del tratamiento RE2.

2.2 Transforma e interpreta los coeficientes en 2.1. en términos de odds u odds-ratios. Presenta el desarrollo formal de una de estas odds-ratios.

Transformamos los coeficientes en odds ratios exponenciando los coeficientes originales:

```r
exp(summary(logit_1)$coefficients[,1])
```

```
## (Intercept)        TRE1        TRE2           W      TRE1:W      TRE2:W
##   0.6435644   0.6697613   0.2441758   7.6582418   0.5351968   0.6987556
```

Formalmente:

- $\ln \frac{\mathbb{P}(\text{Fair} \mid \text{T=RA, W=0})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RA, W=0})} = \beta_0$, por tanto:

$$\frac{\mathbb{P}(\text{Fair} \mid \text{T=RA, W=0})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RA, W=0})} = e^{\beta_0}$$

es decir: e^{β_0} corresponde a las odds de que un perdedor en el tratamiento RA perciba que el resultado del juego es "justo"

- $\ln \frac{\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=0})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=0})} = \beta_0 + \beta_{RE1}$, por tanto:

$$\frac{\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=0})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=0})} = e^{\beta_0+\beta_{RE1}} = e^{\beta_0}e^{\beta_{RE1}}$$

se sigue de esto que

$$\frac{\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=0})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=0})} \Bigg/ \frac{\mathbb{P}(\text{Fair} \mid \text{T=RA, W=0})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RA, W=0})} = e^{\beta_0}e^{\beta_{RE1}}/e^{\beta_0} = e^{\beta_{RE1}}$$

es decir, $e^{\beta_{RE1}}$ corresponde a la odds-ratio de percibir el resultado del juego como "justo" entre un perdedor en el tratamiento RE1 y un perdedor en el tratamiento RA. En particular, observamos que las odds de que un perdedor en el tratamiento RE1 perciba que el resultado del juego es "justo" son 0.67 veces las odds de que un perdedor en el tratamiento RA perciba el resultado del juego como "justo".

Por su parte, $e^{\beta_{RE2:W}}$ es un término de interacción que "comprime" las siguientes odds ratios tal como serían observadas en ausencia de interacción: (a) la odds-ratio de percibir el resultado del juego como "justo" entre un ganador en el tratamiento RE2 y un ganador en el tratamiento RA y (b) la odds-ratio de percibir el resultado del juego como "justo" entre un ganador en el tratamiento RE2 y perdedor en el mismo tratamiento:

$$(a) \quad \frac{\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=1})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=1})} \Bigg/ \frac{\mathbb{P}(\text{Fair} \mid \text{T=RA, W=1})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RA, W=1})} = e^{\beta_0}e^{\beta_{RE2}}e^{\beta_W}e^{\beta_{RE2:W}}/e^{\beta_0}e^{\beta_W} = e^{\beta_{RE2}}e^{\beta_{RE2:W}}$$

$$(b) \quad \frac{\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=1})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=1})} \Bigg/ \frac{\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=0})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=0})} = e^{\beta_0}e^{\beta_{RE2}}e^{\beta_W}e^{\beta_{RE2:W}}/e^{\beta_0}e^{\beta_{RE2}} = e^{\beta_W}e^{\beta_{RE2:W}}$$

2.3 Manipula los resultados del modelo para obtener las probabilidades esperadas de que los ganadores – en cada uno de los tratamientos – sostengan que los resultados de la competencia son "justos" (los puntos azules en el panel derecho de la figura 1). Expresa formalmente las ecuaciones correspondiente a estas predicciones.

- $\ln \frac{\mathbb{P}(\text{Fair} \mid \text{T=RA, W=1})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RA, W=0})} = \beta_0 + \beta_W$

por tanto,

$$\mathbb{P}(\text{Fair} \mid \text{T=RA, W=1}) = \frac{1}{1+e^{-(\beta_0+\beta_W)}}$$

- $\ln \frac{\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=1})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=0})} = \beta_0 + \beta_W + \beta_{RE1}$

por tanto,

$$\mathbb{P}(\text{Fair} \mid \text{T=RE1, W=1}) = \frac{1}{1+e^{-(\beta_0+\beta_W+\beta_{RE1}+\beta_{RE1:W})}}$$

- $\ln \frac{\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=1})}{1-\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=1})} = \beta_0 + \beta_W + \beta_{RE2}$

por tanto,

$$\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=1}) = \frac{1}{1+e^{-(\beta_0+\beta_W+\beta_{RE2}+\beta_{RE2:W})}}$$

Implementación en `R`:

```
newx <- data_paper %>% data_grid(T,W=1,.model=logit_1)
newy <- newx %>% mutate(pred_lor = predict(logit_1, newdata = newx), pred_prob = 1/(1 + exp(-pred_lor)))
print(newy)
```

```
## # A tibble: 3 x 4
##   T         W pred_lor pred_prob
##   <chr> <dbl>    <dbl>     <dbl>
## 1 RA        1    1.60      0.831
## 2 RE1       1    0.569     0.639
## 3 RE2       1   -0.173     0.457
```

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

```
logit_2 <- glm(1*(fair=="Fair") ~ T*W + age, family = binomial(link=logit), data=data_paper)
summary(logit_2)
```

```
##
## Call:
## glm(formula = 1 * (fair == "Fair") ~ T * W + age, family = binomial(link = logit),
##     data = data_paper)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2250  -0.9388  -0.4393   0.9542   2.1391
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.489302   0.353499  -4.213 2.52e-05 ***
## TRE1        -0.360049   0.290807  -1.238 0.215678
## TRE2        -1.399864   0.364577  -3.840 0.000123 ***
## W            2.105898   0.265516   7.931 2.17e-15 ***
## age          0.027242   0.008124   3.353 0.000798 ***
## TRE1:W      -0.686716   0.425902  -1.612 0.106879
## TRE2:W      -0.436466   0.477282  -0.914 0.360463
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 912.18  on 657  degrees of freedom
## Residual deviance: 746.51  on 651  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 760.51
##
## Number of Fisher Scoring iterations: 4
```

2.5 De acuerdo al modelo estimado en 2.4., ¿cuál sería la probabilidad esperada de que un perdedor de 10 años de edad en el tratamiento RE2 sostenga que los resultados de la competencia son "justos"?. Expresa formalmente la ecuación correspondiente a esta predicción. Compara este resultado con el obtenido en 1.4.

$$\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=0, age=10}) = \frac{1}{1+e^{-(\beta_0+\beta_{RE2}+\beta_{age}age)}}$$

$$\mathbb{P}(\text{Fair} \mid \text{T=RE2, W=0, age=10}) = \frac{1}{1+e^{-(-1.489302-1.399864+0.027242\cdot 10)}}$$

$$= 0.06806842$$

A diferencia del LPM, el sigmoide $1/(1+e^{-XB})$ garantiza que las probabilidades predichas siempre serán restringidas al rango 0-1.

2.6. De acuerdo al modelo estimado en 2.4., ¿cuál es *efecto* de edad sobre la probabilidad esperada de sostener que los resultados de la competencia son "justos" para un perdedor de 20 años de edad en el tratamiento RE2? ¿y para un perdedor de 50 años en el mismo tratamiento? ¿y para un perdedor de 50 años en el tratamiento RA? Expresa formalmente las ecuaciones correspondientes a estas cantidades. Compara esta respuesta con la respuesta dada en 1.3.

Recordar que:

$$\frac{\partial p_i}{\partial \text{age}} = \beta_{\text{age}} \times p_i(1-p_i)$$

donde $p_i = \mathbb{P}(\text{Fair}_i \mid \text{ T, W, age}) = 1/(1+e^{-(\beta_0+\beta_{RE1}RE1+\cdots+\beta_{RE2W}RE2*W)})$

Por tanto, para un perdedor de 20 años de edad en el tratamiento RE2,

$p_i = \mathbb{P}(\text{Fair} \mid \text{T=RE2, W=0, age=20}) = \frac{1}{1+e^{-(\beta_0+\beta_{RE2}+\beta_{age}20)}} = \frac{1}{1+e^{-(-1.489-1.399+0.027\cdot 20)}} = 0.088$

entonces:

$$\frac{\partial p_i}{\partial \text{age}} = \beta_{\text{age}} \times p_i(1-p_i) = 0.027 \cdot 0.088 \cdot (1-0.088) = 0.002$$

Por su parte, para un perdedor de 50 años de edad en el tratamiento RE2,

$p_i = \mathbb{P}(\text{Fair} \mid \text{T=RE2, W=0, age=50}) = \frac{1}{1+e^{-(\beta_0+\beta_{RE2}+\beta_{age}50)}} = \frac{1}{1+e^{-(-1.489-1.399+0.027\cdot 50)}} = 0.178$

entonces:

$$\frac{\partial p_i}{\partial \text{age}} = \beta_{\text{age}} \times p_i(1-p_i) = 0.027 \cdot 0.178 \cdot (1-0.178) = 0.004$$

Por último, para un perdedor de 50 años de edad en el tratamiento RA,

$p_i = \mathbb{P}(\text{Fair} \mid \text{T=RA, W=0, age=50}) = \frac{1}{1+e^{-(\beta_0+\beta_{age}50)}} = \frac{1}{1+e^{-(-1.489+0.027\cdot 50)}} = 0.465$

entonces:

$$\frac{\partial p_i}{\partial \text{age}} = \beta_{\text{age}} \times p_i(1-p_i) = 0.027 \cdot 0.465 \cdot (1-0.465) = 0.007$$

A diferencia del LPM, en el modelo de regresión logística el efecto de edad depende del valor de la misma variable edad y del valor de otras covariables.

3. **Bonus**:

3.1 Grafica la relación entre edad y la probabilidad esperada de que un perdedor en el tratamiento RE2 sostenga que los resultados de la competencia son "justos". Específicamente, grafica las probabilidades predichas derivadas del LMP en 1.3 y del modelo de regresión logística en 2.4.

Para implementar el gráfico en `R` sigue los siguientes pasos:

- En el código abajo remplaza `mi_lpm = nombre_tu_lpm` y `mi_logit = nombre_tu_logit` con el nombre del objeto que contiene el LPM y regresión logística, respetivamente. Aquí `lm(1*(fair=="Fair") ~ T + W + age, data=data_paper)` y `glm(1*(fair=="Fair") ~ T + W + age, family = binomial(link=logit), data=data_paper)` sirven sólo a modo de ejemplo.
- `mi_linear_pred_lpm` y `mi_linear_pred_logit` corresponden a las predicciones que producen dichos modelos.
- Remplaza `mi_pred_prob_lpm = transformacion_de_mi_linear_pred_lpm` y `mi_pred_prob_logit = transformacion_de_mi_linear_pred_logit` con el nombre del objeto que contiene las transformaciones de las predicciones de dichos modelos en términos de probabilidades. En el código abajo `mi_pred_prob_lpm = cos(mi_linear_pred_lpm) - 1` y `mi_pred_prob_logit = sin(mi_linear_pred_logit) + 1` sirven sólo a modo de ejemplo.

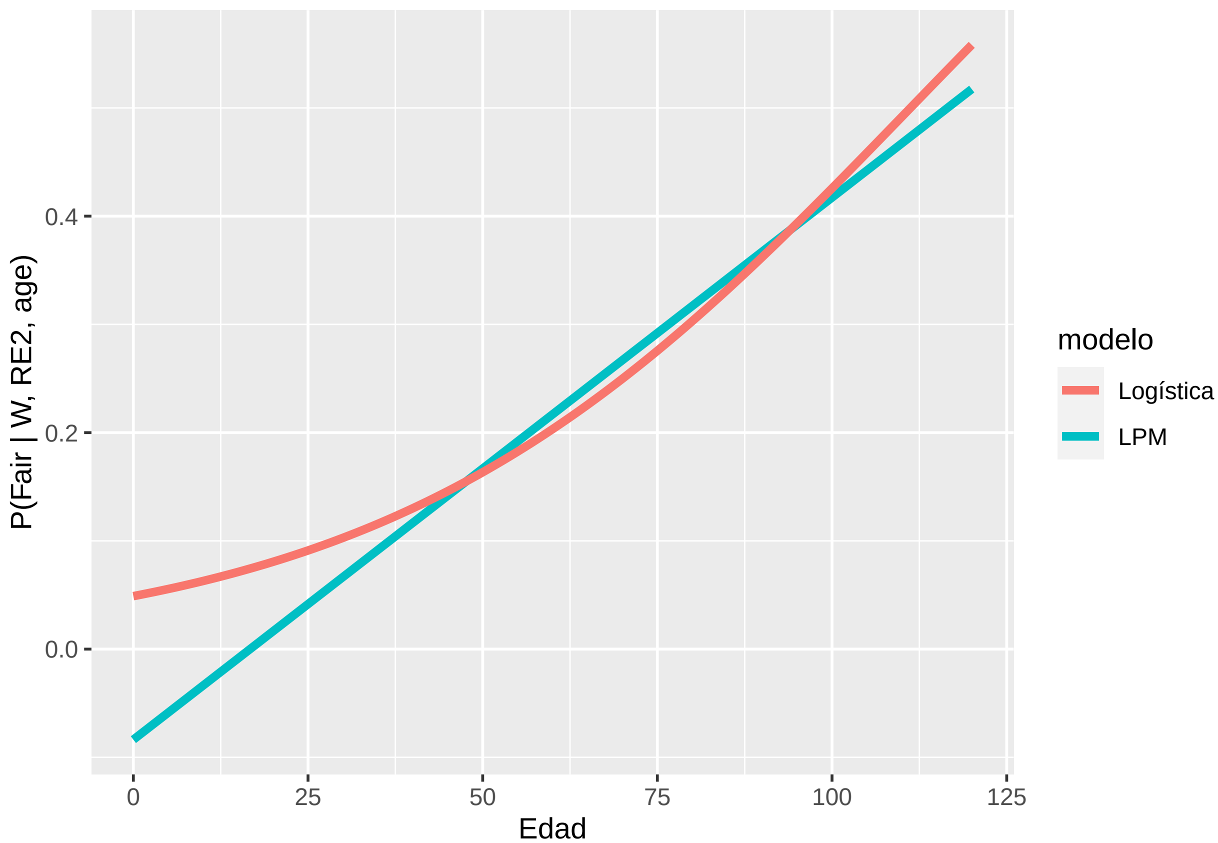
```r
mi_lpm <-  lm(1*(fair=="Fair") ~ T + W + age, data=data_paper)
mi_logit <- glm(1*(fair=="Fair") ~ T + W + age, family = binomial(link=logit), data=data_paper)

# crea un nuevo set de datos sobre los cuales crear predicciones
newx <- data_paper %>% data_grid(T="RE2",W=0, age=seq(0,120), .model=mi_lpm)

# crea valores predichos para el nuevo set de datos
mi_linear_pred_lpm = predict(mi_lpm , newdata = newx)
mi_linear_pred_logit = predict(mi_logit, newdata = newx)
mi_pred_prob_lpm = mi_linear_pred_lpm
mi_pred_prob_logit = 1/(1 + exp(-mi_linear_pred_logit))

newy <- newx %>% mutate(linear_pred_lpm = mi_linear_pred_lpm,
                        linear_pred_logit = mi_linear_pred_logit,
                        pred_prob_lpm = mi_pred_prob_lpm,
                        pred_prob_logit = mi_pred_prob_logit)

# crea gráfico
newy %>% ggplot(aes(x=age, y=pred_prob_lpm, colour="LPM")) +
  geom_line(size=1.5) +
  geom_line(aes(x=age, y=pred_prob_logit, colour="Logística"), size=1.5) +
  labs(y="P(Fair | W, RE2, age)", x="Edad", colour="modelo")
```



3.2 Inspecciona visualmente la figura producida en 3.1 y determina la edad aproximada en la cual encontramos el mayor *efecto* de edad sobre la probabilidad de que un perdedor en el tratamiento RE2 sostenga que los resultados de la competencia son "justos".

Para un perdedor en el tratamiento RE2 el mayor efecto marginal de edad sobre la probabilidad de sostener que los resultados de la competencia son "justos" ocurre al rededor de los 110 años de edad.

3.3 ¿Cuál es el mayor efecto posible de edad?

El mayor efecto de edad ocurre cuando $p_i = 0.5$, por tanto el mayor efecto marginal de edad es $\frac{\partial p_i}{\partial \text{age}} = \beta_{\text{age}}/4 = 0.027/4 = 0.007$