

91.30 Estadística Aplicada II

Trabajo Práctico Regresión Logística

Agustín Buttini

Padrón 104355

Docentes:

Prof. Ing. Mariano Bonoli Escobar,
JTP. Ing. María Fernanda Stewart-Harris

1C 2022

Índice

Desarrollo	2
Ejercicio 1.	2
Inciso a)	2
Inciso b)	5
Ejercicio 2.	7
Inciso a)	7
Inciso b)	10
Ejercicio 3.	10
Inciso a)	10
Inciso b)	10
Ejercicio 4.	11
Inciso a)	11
Inciso b)	12
Inciso c)	13

Introducción

Se analizan los datos de viaje de los clientes de una compañía aérea para estudiar los factores que determinan la satisfacción en los viajes de dichos clientes. Las variables independientes analizadas son el sexo, la edad, la lealtad del usuario percibida por la empresa, el tipo de viaje realizado (personal o de negocios), la distancia recorrida por viaje, la clase del viaje, la demora en la partida y la demora en el arribo.

Para el desarrollo del Trabajo Práctico se utilizó el dataset provisto por la cátedra en formato Excel y se realizó el código correspondiente en lenguaje R. Se adjuntan los scripts correspondientes según sean relevantes a la resolución.

Desarrollo

1. Estimación del modelo con todas las variables disponibles.

Importando el dataset de excel, se escribió un script para formar el modelo logístico, utilizando la satisfacción como variable de respuesta y todos los atributos restantes como variables explicativas. Se tuvieron en cuenta las variables clase y lealtad como variables ordinales y se codificaron de forma acorde:

```
factor_clase <- factor(datos_compania$clase, levels=c("Eco", "Eco Plus", "Business"), order = TRUE)
datos_compania$clase_ordinal <- factor_clase
factor_lealtad <- factor(datos_compania$lealtad, levels=c("Baja Lealtad", "Alta Lealtad"), order = TRUE)
datos_compania$lealtad_ordinal <- factor_lealtad
```

Se agregan al modelo las variables ordinales codificadas y se descartan sus variantes no ordinales, que aparecerían en forma de Dummy.

```
m1 <- glm(formula = as.factor(satisfecho) ~ . -clase -lealtad,
          data = datos_compania,
          family = binomial)
summary(m1)
```

Los coeficientes estimados obtenidos para cada variable son los siguientes:

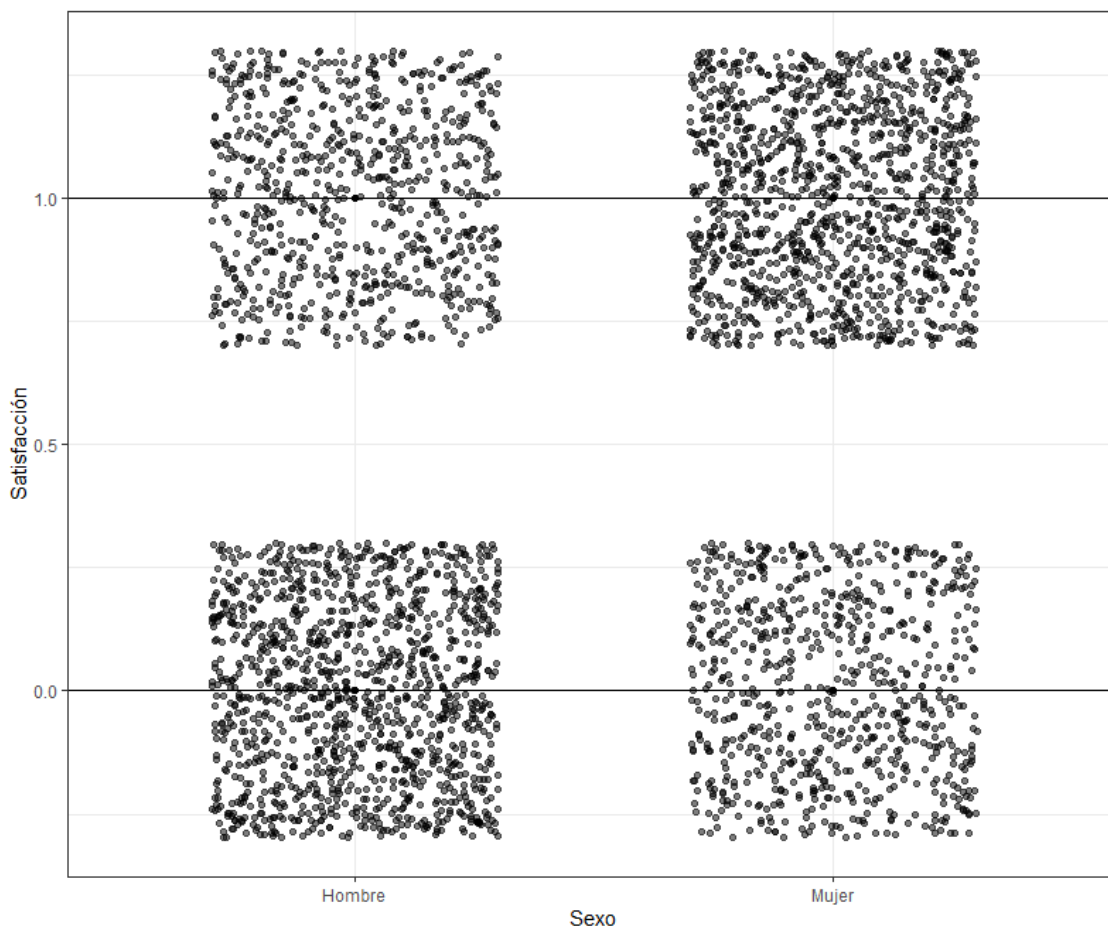
Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.096e+00	1.619e-01	-6.767	1.31e-11	***
sexoMujer	1.179e+00	7.475e-02	15.769	< 2e-16	***
edad	2.119e-03	2.578e-03	0.822	0.41100	
tipo_viajePersonal	-2.344e-01	9.832e-02	-2.384	0.01714	*
distancia	-5.383e-05	3.756e-05	-1.433	0.15179	
min_demora_partida	1.317e-03	3.549e-03	0.371	0.71065	
minutos_demora_arribo	-5.705e-03	3.498e-03	-1.631	0.10290	
clase_ordinal.L	9.041e-01	6.588e-02	13.724	< 2e-16	***
clase_ordinal.Q	3.331e-01	1.097e-01	3.037	0.00239	**
lealtad_ordinal.L	1.272e+00	7.914e-02	16.079	< 2e-16	***

Para los casos de las variables sexo y tipo_viaje el software construye variables tipo dummy (que pueden tomar valor 1 o 0 según corresponda), y utiliza los valores ausentes como referencia. Para el sexo la referencia es el valor "Hombre" y para tipo_viaje el valor de referencia es "Negocios".

- Analizar el signo de los coeficientes de las variables es útil para estimar de qué manera afectan a la variable de respuesta.

sexoMujer: $B = 1,179$

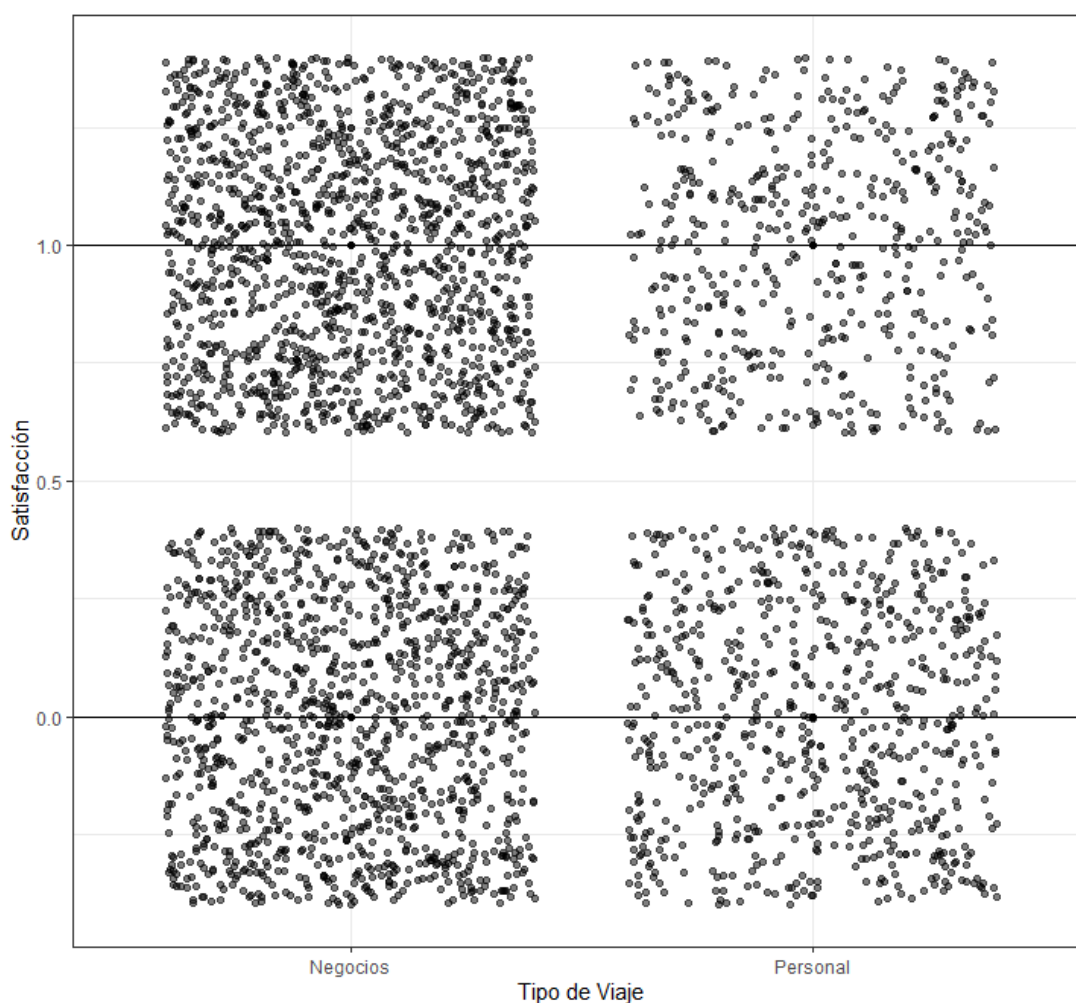
El valor positivo indica que aparentemente las mujeres experimentan un mayor grado de satisfacción que los hombres. Esto se ve en el siguiente gráfico, donde en las mujeres hay mayor concentración de puntos hacia la satisfacción y sucede lo opuesto para el caso de los hombres.



Una posible explicación de los distintos niveles de satisfacción entre hombres y mujeres es que los hombres encuentren incomodidad en el tamaño de los asientos debido a sus mayores proporciones físicas.

tipo_viajePersonal: $B = -0.2344$

Según la estimación del coeficiente de la dummy referente al tipo de viaje Personal, se observa que prevalece la insatisfacción en los viajes personales. Esto puede deberse a que las expectativas de los clientes son más altas cuando realizan viajes personales y el servicio no cumple con estas expectativas. Los clientes que realizan viajes de negocios tal vez lo hacen con más frecuencia y el servicio obtenido es lo suficientemente satisfactorio. Se muestra la densidad de satisfacción en el siguiente gráfico:



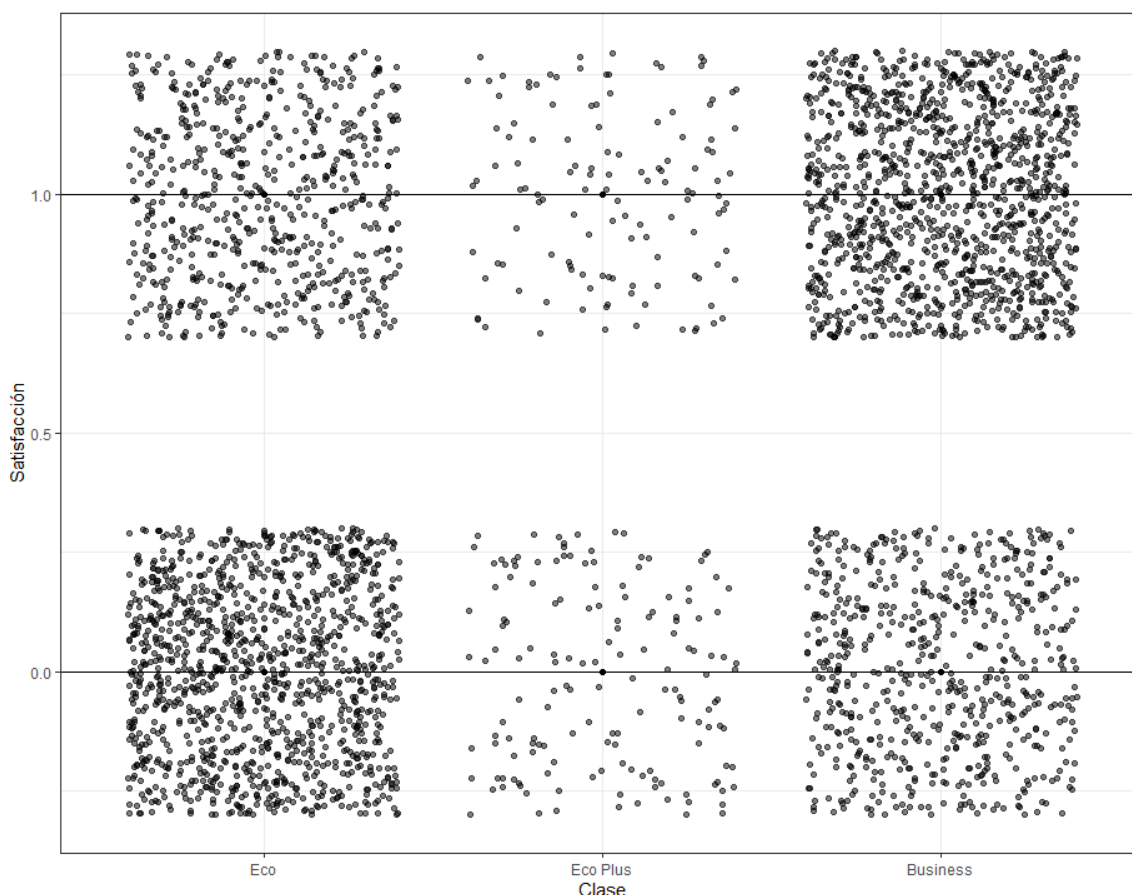
clase_ordinal.L $B = 0.9041$

clase_ordinal.Q $B = 0.3331$

Variable **clase_ordinal**:

Business	Eco Plus	Eco
0	0	1
0	1	1
1	1	1

Se sabe que **clase_ordinal.L** corresponde a Eco Plus, **clase_ordinal.Q** a Business y Eco es la referencia. Según la estimación en R, el caso donde la clase es Eco la satisfacción es menor. Lo mismo sucede en menor medida para la clase Eco Plus. Posiblemente esto se deba a que el servicio ofrece menos prestaciones cuando se trata de la clase Eco. Se ve en el gráfico que la clase Business tiene el mayor grado de satisfacción:



- b) Para analizar la significación de las variables se considera nivel de significación $\alpha = 0,05$, por lo que $Z_c = \pm 1,64485$ para hipótesis unilaterales.

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.096e+00	1.619e-01	-6.767	1.31e-11 ***
sexoMujer	1.179e+00	7.475e-02	15.769	< 2e-16 ***
edad	2.119e-03	2.578e-03	0.822	0.41100
tipo_viajePersonal	-2.344e-01	9.832e-02	-2.384	0.01714 *
distancia	-5.383e-05	3.756e-05	-1.433	0.15179
min_demora_partida	1.317e-03	3.549e-03	0.371	0.71065
minutos_demora_arribo	-5.705e-03	3.498e-03	-1.631	0.10290
clase_ordinal.L	9.041e-01	6.588e-02	13.724	< 2e-16 ***
clase_ordinal.Q	3.331e-01	1.097e-01	3.037	0.00239 **
lealtad_ordinal.L	1.272e+00	7.914e-02	16.079	< 2e-16 ***

sexoMujer: Basado en lo dicho anteriormente, se estima que las mujeres estén más cómodas en promedio por tener que acomodarse en los mismos asientos que los hombres, quienes tienen mayores proporciones físicas, aumentando así la satisfacción media en las mujeres. Esto se refleja en un coeficiente positivo.

$$H_0) \beta_{\text{sexoMujer}} \leq 0 \quad \text{vs} \quad H_1) \beta_{\text{sexoMujer}} > 0$$

$$CR) Z_{\text{obs}} \geq Z_c = 1,64485$$

Siendo $Z_{\text{obs}} = 15,769 > Z_c$ resulta sexoMujer significativa.

edad: podría suceder que a medida que aumenta la edad las necesidades de comodidad son mayores, por lo que el coeficiente estimado debería ser positivo.

$$H_0) \beta_{edad} \leq 0 \quad vs \quad H_1) \beta_{edad} > 0$$

$$CR) Z_{obs} \geq Z_c = 1,64485$$

Siendo $Z_{obs} = 0.822 < Z_c$ no rechaza H_0 y resulta edad no significativa.

tipo_viajePersonal: Cuando un cliente realiza un viaje personal es posible que preste más atención a la comodidad o la ausencia de la misma en comparación a un cliente que realiza un viaje rutinario de negocios. Se estima que el coeficiente será negativo.

$$H_0) \beta_{tipo_viajePersonal} \geq 0 \quad vs \quad H_1) \beta_{tipo_viajePersonal} < 0$$

$$CR) Z_{obs} \leq Z_c = -1,64485$$

Siendo $Z_{obs} = -2,384 < Z_c$ rechaza H_0 y resulta tipo_viajePersonal significativo.

distancia: Al decidir realizar el vuelo, el cliente ya tiene una predisposición a viajar una determinada distancia por lo que la misma no debería influir en la satisfacción. No hay razones estadísticas para suponer que la distancia viajada influye en la satisfacción del cliente, por lo que se toma hipótesis bilateral.

$$H_0) \beta_{distancia} = 0 \quad vs \quad H_1) \beta_{distancia} \neq 0$$

$$CR) Z_{obs} \leq Z_{c1} = Z_{0,025} = -1,96 \quad \cup \quad Z_{obs} \geq Z_{c2} = Z_{0,975} = 1,96$$

Siendo $Z_{c1} < Z_{obs} = -1,433 < Z_{c2}$, no rechaza H_0 y resulta distancia no significativa.

min_demora_partida/min_demora_arribo: al tratarse de demoras, se espera que el coeficiente sea negativo, una mayor demora muy probablemente esté asociada a un menor nivel de satisfacción del cliente.

$$H_0) \beta_{min_demora_partida} \geq 0 \quad vs \quad H_1) \beta_{min_demora_partida} < 0$$

$$CR) Z_{obs} \leq Z_c = -1,64485$$

Siendo $Z_{obs} = 0.371 > Z_c$ no rechaza H_0 y resulta min_demora_partida no significativo.

$$H_0) \beta_{min_demora_arribo} \geq 0 \quad vs \quad H_1) \beta_{min_demora_arribo} < 0$$

$$CR) Z_{obs} \leq Z_c = -1,64485$$

Siendo $Z_{obs} = -1,631 > Z_c$ no rechaza H_0 y resulta min_demora_arribo no significativo.

clase_ordinal.L (Eco Plus)/clase_ordinal.Q (Business): Las clases que no son Eco ofrecen servicios más completos, por lo que se estima que el nivel de satisfacción será mayor y el coeficiente será positivo.

$$H_0) \beta_{clase_ordinal.L} \leq 0 \quad vs \quad H_1) \beta_{clase_ordinal.L} > 0 \quad (\text{clase Eco Plus})$$

$$CR) Z_{obs} \geq Z_c = 1,64485$$

Siendo $Z_{obs} = 13,724 > Z_c$ rechaza H_0 y resulta clase_ordinal.L significativo

$$H_0) \beta_{clase_ordinal.Q} \leq 0 \quad vs \quad H_1) \beta_{clase_ordinal.Q} > 0 \quad (\text{clase Business})$$

$$CR) Z_{obs} \geq Z_c = 1,64485$$

Siendo $Z_{obs} = 3,037 > Z_c$ rechaza H_0 y resulta clase_ordinal.Q significativo.

lealtad_ordinal.L (Alta lealtad): Sabiendo que corresponde a Alta Lealtad, es posible afirmar que se espera que los clientes que tienen una lealtad más alta con la aerolínea tendrán un nivel más alto de satisfacción, por lo que se supone que el coeficiente será positivo.

$$H_0) \beta_{lealtad_ordinal.L} \leq 0 \quad vs \quad H_1) \beta_{lealtad_ordinal.L} > 0 \quad (\text{Alta lealtad})$$

$$CR) Z_{obs} \geq Z_c = 1,64485$$

Siendo $Z_{obs} = 16,079 > Z_c$ rechaza H_0 y resulta lealtad_ordinal.L Lealtad significativa.

2. Se realizan los métodos de step hacia adelante y hacia atrás en R mediante la observación del criterio de información de Akaike (AIC), que es un coeficiente comparativo que permite darle un valor al trade-off entre la bondad de ajuste y la complejidad entre modelos. Cuando menor sea el valor del AIC, más preciso será el modelo.

a) El código utilizado para realizar los steps fue:

```
mvacio <- glm(as.factor(satisfecho) ~ 1,
              data = datos_compania,
              family = binomial)
variables <- formula(as.factor(satisfecho) ~ sexo + lealtad_ordinal + edad + tipo_viaje +
                    clase_ordinal + distancia + min_demora_partida + minutos_demora_arribo)
mstepf <- stepAIC(mvacio, trace=TRUE, direction="forward", scope=variables)
mstepb <- stepAIC(m1, trace=TRUE, direction="backward", scope=variables)
```

donde mvacio es el modelo sin variables y el vector variables constituye el scope de donde el step reconocerá parámetros para agregarlos o quitarlos según se haga hacia adelante o hacia atrás.

Step Forward:

Para el Step Forward, se estimó un modelo vacío al cual se le fueron agregando variables iterativamente hasta llegar al valor mínimo de AIC.


```
Start: AIC=5519.44
as.factor(satisfecho) ~ 1
```

	Df	Deviance	AIC
+ clase_ordinal	2	5119.5	5125.5
+ lealtad_ordinal	1	5189.7	5193.7
+ sexo	1	5281.6	5285.6
+ edad	1	5441.2	5445.2
+ tipo_viaje	1	5466.6	5470.6
+ minutos_demora_arribo	1	5503.9	5507.9
+ min_demora_partida	1	5506.0	5510.0
+ distancia	1	5512.1	5516.1
<none>		5517.4	5519.4

A partir del modelo vacío, la función ordena y agrega las variables disponibles por la reducción producida en el AIC, y repite el proceso hasta que no agregar ningún parámetro constituya la mejor decisión.

```
Step: AIC=4527.96
as.factor(satisfecho) ~ clase_ordinal + lealtad_ordinal + sexo +
minutos_demora_arribo + tipo_viaje + distancia
```

	Df	Deviance	AIC
<none>		4512.0	4528.0
+ edad	1	4511.3	4529.3
+ min_demora_partida	1	4511.8	4529.8

El modelo obtenido es el que contiene las variables explicativas clase_ordinal, lealtad_ordinal, sexo, minutos_demora_arribo, tipo_viaje y distancia.

Obteniendo un summary del modelo de Step Forward:

```
> summary(mstepf)
```

Call:
 glm(formula = as.factor(satisfecho) ~ clase_ordinal + lealtad_ordinal +
 sexo + minutos_demora_arribo + tipo_viaje + distancia, family = binomial,
 data = datos_compania)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0037	-0.8161	0.5374	0.9738	2.8337

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-9.976e-01	1.093e-01	-9.130	< 2e-16 ***
clase_ordinal.L	9.069e-01	6.582e-02	13.778	< 2e-16 ***
clase_ordinal.Q	3.350e-01	1.096e-01	3.055	0.00225 **
lealtad_ordinal.L	1.289e+00	7.648e-02	16.857	< 2e-16 ***
sexoMujer	1.176e+00	7.467e-02	15.752	< 2e-16 ***
minutos_demora_arribo	-4.453e-03	9.700e-04	-4.590	4.43e-06 ***
tipo_viajePersonal	-2.454e-01	9.740e-02	-2.520	0.01174 *
distancia	-6.253e-05	3.593e-05	-1.740	0.08182 .

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 5517.4 on 3979 degrees of freedom
 Residual deviance: 4512.0 on 3972 degrees of freedom
 AIC: 4528

Number of Fisher Scoring iterations: 4

$AIC = 4528$

Cálculo de $R_{MF}^2 = 1 - \frac{4512}{5517,4} \rightarrow R_{MF}^2 = 0,182$

Se ve además que todas las variables del modelo resultan significativas.

Step Backward:

Para realizar el Step Backward, se inicia el proceso iterativo a partir del modelo m1 del punto anterior, y se retiran variables de forma análoga a las agregadas en el Step Forward:

```
Start: AIC=4531.14
as.factor(satisfecho) ~ (sexo + lealtad + edad + tipo_viaje +
  clase + distancia + min_demora_partida + minutos_demora_arribo +
  clase_ordinal + lealtad_ordinal) - clase - lealtad
```

	Df	Deviance	AIC
- min_demora_partida	1	4511.3	4529.3
- edad	1	4511.8	4529.8
<none>		4511.1	4531.1
- distancia	1	4513.2	4531.2
- minutos_demora_arribo	1	4513.8	4531.8
- tipo_viaje	1	4516.8	4534.8
- clase_ordinal	2	4710.4	4726.4
- sexo	1	4774.5	4792.5
- lealtad_ordinal	1	4798.7	4816.7

Al igual que en el step hacia adelante, se itera hasta alcanzar el menor valor posible de AIC:

```
Step: AIC=4527.96
as.factor(satisfecho) ~ sexo + tipo_viaje + distancia + minutos_demora_arribo +
  clase_ordinal + lealtad_ordinal
```

	Df	Deviance	AIC
<none>		4512.0	4528.0
- distancia	1	4515.0	4529.0
- tipo_viaje	1	4518.3	4532.3
- minutos_demora_arribo	1	4535.0	4549.0
- clase_ordinal	2	4712.9	4724.9
- sexo	1	4774.6	4788.6
- lealtad_ordinal	1	4831.6	4845.6

De esta forma se llega a un modelo con las variables explicativas: clase_ordinal, lealtad_ordinal, sexo, minutos_demora_arribo, tipo_viaje y distancia.

Summary del modelo:

```
> summary(mstepb)

Call:
glm(formula = as.factor(satisfecho) ~ sexo + tipo_viaje + distancia +
  minutos_demora_arribo + clase_ordinal + lealtad_ordinal,
  family = binomial, data = datos_compania)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0037  -0.8161   0.5374   0.9738   2.8337

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -9.976e-01  1.093e-01  -9.130 < 2e-16 ***
sexoMujer    1.176e+00  7.467e-02  15.752 < 2e-16 ***
tipo_viajePersonal -2.454e-01  9.740e-02  -2.520  0.01174 *
distancia    -6.253e-05  3.593e-05  -1.740  0.08182 .
minutos_demora_arribo -4.453e-03  9.700e-04  -4.590  4.43e-06 ***
clase_ordinal.L  9.069e-01  6.582e-02  13.778 < 2e-16 ***
clase_ordinal.Q  3.350e-01  1.096e-01   3.055  0.00225 **
lealtad_ordinal.L 1.289e+00  7.648e-02  16.857 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5517.4 on 3979 degrees of freedom
Residual deviance: 4512.0 on 3972 degrees of freedom
AIC: 4528

Number of Fisher Scoring iterations: 4
```

$$AIC = 4528$$

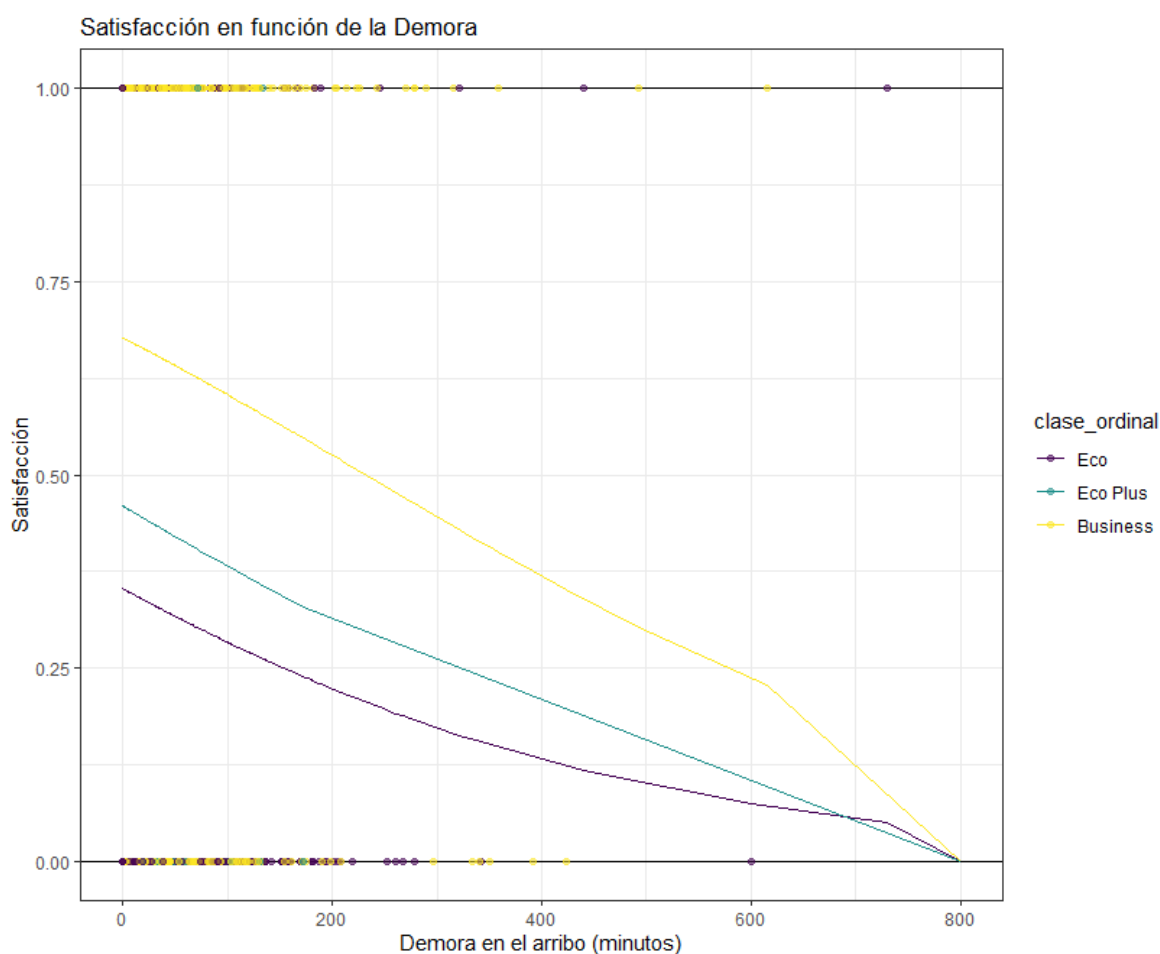
$$\text{Cálculo de } R_{MF}^2 = 1 - \frac{4512}{5517,4} \rightarrow R_{MF}^2 = 0,182$$

b) A partir de ambos métodos se obtuvo el mismo modelo, con el mismo valor de AIC y los mismos niveles de significación.

Aclaración importante: Las iteraciones fueron realizadas hasta alcanzar el menor AIC, pero no consideran los costos de poseer más variables. Puede ser conveniente el agregado de menos variables para el caso del Step Forward y la extracción de más variables para el Step Backwards, dado que los últimos ajustes del AIC para ambos casos son muy pequeños.

3. a) Para realizar la estimación del modelo se calculó la distancia media mediante la función `mean()` en R, se incluyeron las variables `clase_ordinal`, `distancia_promedio` y `minutos_demora_arribo`, se añadieron las estimaciones de los cuantiles `z` y las probabilidades `p` de esos cuantiles.

Con esta información, mediante el uso de la librería de R “ggplot2”, se representó gráficamente la evolución de la probabilidad de satisfacción en función del tiempo de demora en el arribo, separando las curvas según la clase en la que viajó cada cliente.



- b) En el gráfico anterior, se confirma empíricamente la suposición de que el nivel de satisfacción con el servicio aumenta según la clase.

Se aprecia cómo la clase Eco es la que tiene mayor cantidad porcentual de clientes insatisfechos y toma una forma algo asintótica a medida que aumentan los minutos de demora.

La media de Eco Plus se posiciona levemente por encima de la clase Eco y la forma de la curva es similar, pero tiene una baja cantidad de puntos como para extraer conclusiones estadísticas evidentes, y es debido a esto que en el gráfico la curva de Eco aparenta superar a la Eco Plus hacia el límite superior de minutos de demora.

Por otro lado, la clase Business tiene una media de satisfacción considerablemente mayor a Eco y Eco Plus. La probabilidad de satisfacción es mayor al 50% si la demora en el arribo es menor a 200 minutos y del 70% si la demora tiende a 0 minutos.

Para el negocio, esto puede mejorar de diversas maneras, por ejemplo:

- Priorizar la inexistencia de demoras para maximizar la probabilidad de satisfacción.
- Mejorar la calidad del servicio a los clientes, específicamente de las clases Eco y Eco Plus.
- Alternativamente a mejorar el servicio, una disminución del precio de las clases Eco y Eco Plus podría bajar las expectativas de los clientes, por lo que la satisfacción percibida aumentaría.

4. a) Se estimó un modelo con las variables del punto 3, y se dividió el dataset en uno de training y otro de testing utilizando las proporciones 75% / 25% respectivamente mediante la función de R `createDataPartition()`. La división se llevó a cabo mediante la distribución aleatoria de registros y se utilizó el dataset de entrenamiento para estimar el modelo m4.

Summary del modelo:

```
> summary(m4)

Call:
glm(formula = as.factor(satisfecho) ~ clase_ordinal + distancia +
    minutos_demora_arribo, family = binomial, data = training_dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6338  -0.9357  -0.6777   0.9598   2.2061

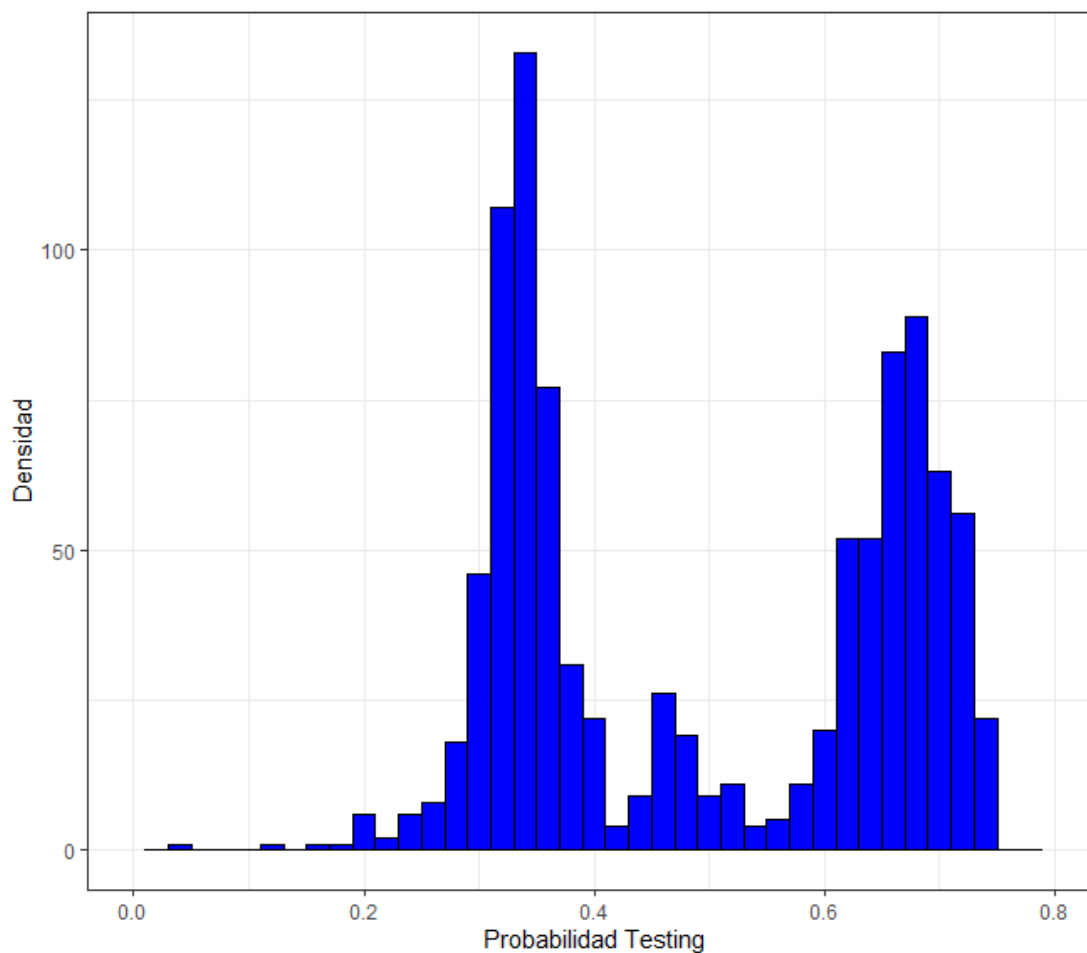
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.610e-01  9.161e-02   2.849  0.004389 **
clase_ordinal.L  1.003e+00  5.803e-02  17.292 < 2e-16 ***
clase_ordinal.Q  1.724e-01  1.189e-01   1.449  0.147215
distancia    -1.426e-04  3.837e-05  -3.717  0.000201 ***
minutos_demora_arribo -3.572e-03  1.023e-03  -3.491  0.000482 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

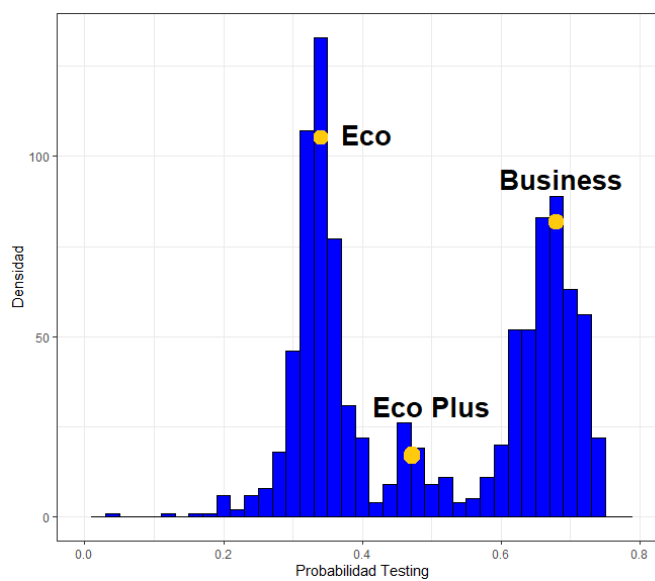
    Null deviance: 4137.8  on 2984  degrees of freedom
Residual deviance: 3800.1  on 2980  degrees of freedom
AIC: 3810.1

Number of Fisher Scoring iterations: 4
```

b) Para representar este modelo m_4 en un histograma se estimaron las probabilidades mediante los cuantiles obtenidos de la función `predict()` y se graficó mediante la librería `ggplot2`:



En este histograma se aprecia una clara mezcla de poblaciones por la aparición de múltiples picos en la distribución:



Se interpreta que estos picos representan una mezcla de poblaciones, ordenados de izquierda a derecha por la satisfacción promedio. Se ve como la clase Eco predomina en densidad de población y tiene el nivel de satisfacción mas bajo.

El segundo pico, por tener una densidad mucho menor se interpreta que es correspondiente a la clase Eco Plus, por la poca cantidad de registros de viajes en esta clase y por el nivel de satisfacción levemente superior a la clase anterior.

El tercer pico corresponde a la población de la clase Business, la cual presenta los mayores niveles de satisfacción.

c) Se construyó la matriz de confusión utilizando el corte de probabilidad = 0,60:

```
predicted <- as.factor(ifelse(predict_p > 0.6, 1, 0))
actual <- as.factor(ifelse(testing_dataset$satisfecho=="si",1,0))

confusionMatrix(data = predicted,
                  reference = actual)
```

La salida del código resulta:

```
Confusion Matrix and Statistics

              Reference
Prediction    0      1
      0  343  223
      1  137  292
```

Puesto en una tabla:

	$Y = 0$	$Y = 1$
$\hat{Y} = 0$	343	223
$\hat{Y} = 1$	137	292

```

Accuracy : 0.6382
95% CI : (0.6075, 0.6681)
No Information Rate : 0.5176
P-Value [Acc > NIR] : 1.086e-14

Kappa : 0.2799

Mcnemar's Test P-Value : 7.468e-06

Sensitivity : 0.7146
Specificity : 0.5670
Pos Pred Value : 0.6060
Neg Pred Value : 0.6807
Prevalence : 0.4824
Detection Rate : 0.3447
Detection Prevalence : 0.5688
Balanced Accuracy : 0.6408

'Positive' Class : 0

```

En la segunda parte de la salida de datos, se detallan los valores pedidos:

- La **exactitud** (Accuracy) estima la cantidad de clientes que fueron estimados correctamente en relación a la predicción del modelo.
Exactitud = 63,82%
- La **sensibilidad** (Sensitivity) permite calcular el porcentaje de clientes estimados como satisfechos en relación a los clientes que realmente están satisfechos.
Sensibilidad = 71,46%
- La **especificidad** (Specificity) estima el porcentaje de clientes estimados como insatisfechos en relación a los clientes que realmente están insatisfechos.
Especificidad = 56,70%
- La **precisión** (Neg Pred Value) determina la estimación del porcentaje de clientes satisfechos en relación a los estimados como satisfechos.
Precisión = 68,07%

Estos valores siempre mejoran cuanto más sube su valor.