



Proyecto Econometría

Geoconda Molina y Debbie Echanique

30 de diciembre de 2023

Índice general

0.1. Datos	2
0.2. Descripción del método	3
0.3. Presentación de resultados	8
0.4. Conclusiones	17

0.1. Datos

Para la ejecución del presente proyecto, se optó por fundamentar el análisis en los datos correspondientes a las Ventas Globales de Videojuegos, los cuales fueron adquiridos de la reconocida plataforma Kaggle. Dicha plataforma proporciona acceso gratuito a bases de datos, convirtiéndola en una fuente confiable para la obtención de información relevante.

El objetivo primordial de este proyecto consiste en llevar a cabo un estudio sobre las ventas globales de videojuegos mediante la aplicación del método de regresión múltiple en el ámbito de la econometría. Para lograr este propósito, se pretende establecer relaciones entre variables independientes y una variable dependiente específica, con la finalidad de obtener conclusiones sustanciales acerca de la capacidad explicativa de dichas variables. Este análisis se llevará a cabo mediante la aplicación de pruebas de hipótesis, lo que permitirá evaluar la robustez y significancia de las relaciones identificadas.

La variable seleccionada como dependiente para este estudio es *Global*, que representa las ventas totales obtenidas durante el periodo de análisis. Con el fin de abordar la complejidad inherente a la diversidad de datos, se han considerado dos modelos diferentes. En uno de ellos, se explorarán exclusivamente variables numéricas, mientras que en el otro se incluirán variables categóricas.

La división en dos modelos distintos proporcionará la oportunidad de realizar análisis separados y detallados, permitiendo una comprensión más profunda de la relación entre las variables seleccionadas y las ventas globales de videojuegos. Este enfoque estratégico facilitará la identificación de patrones, tendencias y posibles correlaciones que contribuirán significativamente a la interpretación de los resultados obtenidos.

0.2. Descripción del método

El método que utilizamos en el trabajo es el método de regresión lineal múltiple para tratar de explicar el número de ventas globales de videojuegos. Se realizaron dos modelos, en el primero se tomó como variables independientes a North.America, Japan, Rank y Review y para el segundo modelo se consideraron las variables Genre, Platform, Rank y Review. La selección de estas variables para ambos modelos se las hizo de acuerdo a los papers "Factors that Impact Video Game Sales" y "Predicting Global Video-Game Sales". Para el caso del segundo modelo como Genre y Platform son dos variables categóricas, se asignó para Genre como base a Action y se crearon las siguientes variables binarias: Shooter, Sports y Misc y para Platform se tomó como base a PS2 y se crearon las variables: PS3, X360 y Wii.

Además se estimó los parámetros de cada modelo $(\beta_1, \beta_2, \dots)$ por MCO y se realizaron las siguientes pruebas de hipótesis:

- Para ver si β_k es significativo, se empleo la siguiente prueba de hipótesis:
 - $H_0: \beta_k = 0$
 - $H_1: \beta_k \geq 0$

Para determinar si se acepta o no la hipótesis nula calculamos primero:

$$t = \frac{\hat{\beta}_k}{\sqrt{V(\hat{\beta}_k)}}$$

y se obtuvo el estadístico t (t_c) con un nivel de confianza de $1 - \alpha$ y $n-K$ grados de libertad y si se tiene que $t > t_c$ se rechaza la hipótesis nula.

- Para ver si los parámetros son conjuntamente significativos. En este caso se tomaron para ver si son conjuntamente significativas entre 2 y 3 parámetros. De manera general se tiene la siguiente prueba de hipótesis:
 - $H_0: \beta_0 = 0, \beta_1 = 0, \beta_2 = 0, \dots$
 - $H_1: H_0$ es falsa

En este caso debemos calcular el modelo restringido y se emplearon los siguientes estadísticos para determinar si se acepta o no la hipótesis nula:

- Estadístico F: Para lo cual se calculó:

$$F = \frac{(SSE_r - SSE_{nr})/9}{SSE_{nr}/(n - K)}$$

y también de la siguiente forma:

$$F = \frac{(R_{nr}^2 - R_r^2)}{(1 - R_{nr}^2)} * \frac{(n - K)}{q}$$

y se determinó F_c con q y n-K grados de libertad. Se rechaza la hipótesis nula si $F > F_c$. El segundo estadístico que se empleo es el ML, para lo cual primero debemos calcular el modelo restringido y luego de esto obtenemos los residuos de este modelo para luego realizar un modelo de regresión lineal múltiple con los residuos como variable dependiente sobre todas las variables independientes del modelo original, para de esta forma obtener:

$$ML = n * R_u^2$$

y este valor se lo compara con el valor del estadístico χ_q^2 .

- Sobre combinaciones lineales de parámetros, en donde se calculó:

$$t = \frac{\hat{\beta}_j - \hat{\beta}_k}{\sqrt{V(\hat{\beta}_j) - \hat{\beta}_k}}$$

donde $j \neq k$ y el estadístico t (t_c) con un nivel de significancia de $1 - \alpha$ y n-K grados de libertad, y se rechaza la hipótesis nula si se tiene que $t > t_c$.

Por otra parte, calculamos los siguientes intervalos de confianza:

- Intervalos de confianza para los parámetros (β):

$$\hat{\beta}_k - t_c * \sqrt{V(\hat{\beta}_k)} \leq \beta_k \leq \hat{\beta}_k + t_c * \sqrt{V(\hat{\beta}_k)}$$

- Intervalo de confianza para las predicciones:

$$\hat{\theta}_o - tc * \sqrt{V(\hat{\theta}_o)} \leq \theta_o \leq \hat{\theta}_o + tc * \sqrt{V(\hat{\theta}_o)}$$

- Intervalos para la predicción de una observación en particular:

$$\hat{y}_o - tc * \sqrt{V(\hat{e}_o)} \leq y_o \leq \hat{y}_o + tc * \sqrt{V(\hat{e}_o)}$$

Para el primer modelo:

Tenemos que el modelo poblacional esta dado de la siguiente forma:

$$Global = \beta_1 + \beta_2 * North.America + \beta_3 * Japan + \beta_4 * Review + \beta_5 * Rank + \mu$$

Donde:

- North.America: Es la variable que representa el número de unidades vendidas en Norte América y está expresada en millones.
- Japan: Es la variable de unidades vendidas en Japón, en millones.
- Review: Es la variable del puntaje de crítica del juego en una escala del 1 al 10.
- Rank: Es la variable que representa la posición del videojuego de acuerdo a las ventas globales.

Se calculo SST, SSE, SSR, R^2 y R^2 ajustado. Se realizaron las siguientes pruebas de hipótesis:

- β_k significativo:
Se realizó esta prueba sobre β_k con $k \in 2, \dots, 5$
- Parámetros conjuntamente significativos, se realizaron las siguientes pruebas:
 - Ho: $\beta_2 = 0, \beta_3 = 0$
H1: Ho es falsa
 - Ho: $\beta_2 = 0, \beta_4 = 0$
H1: Ho es falsa
 - Ho: $\beta_2 = 0, \beta_3 = 0, \beta_5 = 0$
H1: Ho es falsa
- Combinaciones lineales de parámetros:
 - Ho: $\beta_2 - \beta_3 = 0$
H1: Ho es falsa
 - Ho: $\beta_4 - \beta_5 = 0$
H1: Ho es falsa
 - Ho: $\beta_2 - \beta_4 = 0$
H1: Ho es falsa
 - Ho: $\beta_3 - \beta_5 = 0$
H1: Ho es falsa

Y se determinaron los siguientes intervalos de confianza:

- Intervalos de confianza para los parámetros:
Se realizó sobre todos los β_k con $k \in 1, 2, \dots, 5$
- Intervalo de confianza para las predicciones: Para este intervalos se tomó los siguientes valores específicos:
 - North.America= 15
 - Japan=22
 - Review=91
 - Rank=54

- Intervalo para una observación en particular, se tomó los siguientes valores:
 - North.America= 31
 - Japan=10
 - Review=27
 - Rank=25

Para el segundo modelo

Como se mencionó anteriormente, para este caso se consideraron dos variables categóricas *Genre* que se refiere al género del Videojuego y la variable *Platform* que es la plataforma en la que se jugó el juego, así, el modelo obtenido después de crear las variables binarias para cada categoría es el siguiente:

$$Global = \beta_0 + \beta_1 Misc + \beta_2 Shooter + \beta_3 Sports + \beta_4 Review + \beta_5 Rank + \beta_6 PS3 + \beta_7 X360 + \beta_8 Wii$$

De donde podemos tomar en cuenta que las variables *Misc*, *Shooter* y *Sports* son las variables creadas apartir de la variable categórica *Genre* y las variables *PS3*, *X360* y *Wii* son las variables creadas apartir de la variable categórica *Platform*

Al comenzar a analizar el modelo de regresión antes mencionado, primero se realizó la prueba para ver cuáles son las variables significativas de forma individual para el modelo, es decir, se queria realizar la siguiente prueba de hipótesis:

- $H_0: \beta_k = 0$
- $H_1: H_0$ es falsa

Para todo $k = 0, 1, \dots, 8$ y tomando como criterio para rechazar o aceptar la hipótesis, el valor p , el cual se obtuvo gracias a la función *summary* inplemetada en R.

El paso siguiente tiene que ver con cuáles serían los intervalos de confianza al 95 porciento de los coeficientes β que acompañan a las variables, para ello, una forma más rápido de hacerlo es utilizando la funcion *confint* con en nivel de significancia requerida, asi:

```
# intervalos de confianza para los betas
intervalos_confianza <- confint(modelo, level = 0.95)
print(intervalos_confianza)
```

Una vez encontrados los intervalos de confianza, se procedió a realizar pruebas de hipótesis conjuntas, es decir, se realizaron pruebas de hipótesis para ver si varias variables son significativas o no para el modelo pero de manera conjunta.

Para ello se realizó la prueba F y la prueba ML entre dos y tres variables distintas, es decir, se realizaron las pruebas de hipótesis:

- Ho: $\beta_i = \beta_j = 0$
H1: Ho es falsa
- Ho: $\beta_i = \beta_j = \beta_s = 0$
H1: Ho es falsa

En la prueba F para aceptar o rechazar la hipótesis nula, se contruye:

$$F = \frac{R_{nr}^2 - R_r^2}{1 - R_{nr}^2} * \frac{n - k}{q}$$

de donde:

R_{nr}^2 es el coeficiente de determinación del modelo no restringido (Modelo original)

R_r^2 es el coeficiente de determinación del modelo restringido (H0 es cierta)

n número de datos analizados

k número de coeficientes β en el modelo

q número de variables consideradas en la hipotesis nula $H0$

Así, se compara F con el estadístico F_c que es el estadístico F de Fisher $F_{q,n-k}$

Ahora para la prueba ML (Multiplicador de Lagrange), se crea el modelo restringido que será el modelo donde la hipótesis nula $H0$ es cierta, es decir, se elimina del modelo original las variables de las cuales se desea hacer la prueba de hipótesis.

```
# Problemas: si es conjuntamente significativo b_1 y b_2
# modelo restringido
modelo_res <- lm(Global ~ Rank+Review+Sports+PS3+X360+Wii, data = data_categ_1)
```

Paso seguido, se obtienen los residuos del modelo

```
#obtenemos los residuos
residuos_res <- modelo_res$residuals
```

Para luego crear el modelo auxiliar $residuos = \sum_{i=0}^8 \beta_i * X_i$, es decir, los residuos del modelo restringido sobre todas las variables del modelo original

```
# hagamos el modelo auxiliar
modelo_aux <- lm(residuos_res ~ Rank+Review+Sports+Misc+Shooter+PS3+X360+Wii, data = data_categ_1)
```

Así, finalmente se contruye el valor $ML = n * R_{\mu}^2$, de donde R_{μ}^2 es el coeficiente de determinación del modelo auxiliar y se realiza la comparacion con el valor chi cuadrado χ_q^2 para aceptar o rechazar la hipótesis nula planteada.

0.3. Presentación de resultados

Para el primer modelo:

Se obtuvieron las siguientes estimaciones de los β :

Coefficients:	
	Estimate
(Intercept)	3.679e-01
North.America	1.649e+00
Japan	2.317e+00
Review	-6.329e-04
Rank	-2.176e-04

De donde se tienen las siguientes interpretaciones:

- $\hat{\beta}_2$: En promedio, manteniendo todas las otras variables constantes, un incremento de una unidad (unidad en millones) de venta de un videojuego en Norte América genera un incremento de 1.649e+00 millones de ventas globales.
- $\hat{\beta}_3$: En promedio, manteniendo todas las otras variables constantes, un incremento de una unidad (unidad en millones) de venta de un videojuego en Japon genera un incremento de 2.317e+00 millones de ventas globales.
- $\hat{\beta}_4$: En promedio, manteniendo todas las otras variables constantes, un incremento de un punto en Review genera un incremento de -6.329e-04 de ventas globales.
- $\hat{\beta}_5$: En promedio, manteniendo todas las otras variables constantes, un incremento de un punto de venta de un videojuego en Rank genera un incremento de -2.176e-04 de ventas globales.

Los signos obtenidos eran los esperados, ya que para las dos primeras variables (North.America y Japan) al aumentar las unidades de ventas en estos lugares va a incrementar las ventas globales, por otra parte, para la variables Review tenemos que si el juego se encuentra con una crítica alta, lo que implica que el juego no es muy bueno, se tiene que las ventas globales van a disminuir y por último tenemos para la variable Rank que si el videojuego se encuentra en una posición alta, donde 1 se considera la mejor posición y así sucesivamente va bajando, tenemos que mientras más alta se encuentre el videojuego en la lista peor es y por tanto las personas no van a querer comprarlo lo que disminuye las ventas globales.

El valor de R^2 es:

0.964137

El valor de R^2 ajustado es:

0.9638963

El valor de SSE es:

422.3509

El valor de SST es:

11776.77

El valor de SSR es:

11354.42

Por otra parte, los resultados obtenidos para las pruebas de hipótesis son los siguientes:

■ Significativos:

Se obtuvieron los siguientes valores t:

β	t
β_2	84.68155
β_3	21.02937
β_4	-0.1833992
β_5	-3.063862

y el valor de t_c es 1.647414, por lo tanto tenemos que:

- β_2 : Es significativo
- β_3 : Es significativo
- β_4 : No es significativo
- β_5 : No es significativo

■ Párametros conjuntamente significativos:

- $H_0: \beta_2 = 0, \beta_3 = 0$
 $H_1: H_0$ es falsa

Se obtuvo el siguiente modelo restringido:

$$\text{Global} = \text{Review} + \text{Rank} + \mu$$

y los siguientes valores de los estadísticos:

- Estadístico F
 - ◊ Usando SSE:

$$F=6170.144$$

◊ Usando R^2 :

$$F=6170.144$$

◊ $F_c=3.010841$

○ ML

◊ ML= 573.3108

◊ MLc= 5.99146

Tenemos que $F > F_c$ y también que $ML > ML_c$ por lo tanto se rechaza la hipótesis nula, es decir que β_2 y β_3 son conjuntamente significativas.

- Ho: $\beta_2 = 0, \beta_4 = 0$

H1: Ho es falsa Se obtuvo el siguiente modelo restringido:

$$\text{Global} = \text{Japan} + \text{Rank} + \mu$$

y los siguientes valores de los estadísticos:

○ Estadístico F

◊ Usando SSE:

$$F=3587.752$$

◊ Usando R^2 :

$$F=3587.752$$

◊ $F_c=3.010841$

○ ML

◊ ML= 554.909

◊ MLc= 5.99146

Tenemos que $F > F_c$ y también que $ML > ML_c$ por lo tanto se rechaza la hipótesis nula, es decir que β_2 y β_4 son conjuntamente significativas.

- Ho: $\beta_2 = 0, \beta_3 = 0, \beta_5 = 0$

H1: Ho es falsa

Se obtuvo el siguiente modelo restringido:

$$\text{Global} = \text{Review} + \mu$$

y los siguientes valores de los estadísticos:

○ Estadístico F

◊ Usando SSE:

$$F=5270.012$$

◇ Usando R^2 :

$$F=5270.012$$

◇ $F_c = 2.619854$

○ ML

◇ $ML = 554.909$

◇ $ML_c = 5.99146$

Tenemos que $F > F_c$ y también que $ML > ML_c$ por lo tanto se rechaza la hipótesis nula, es decir que β_2, β_3 y β_5 son conjuntamente significativas.

■ Combinaciones lineales de parámetros:

- $H_0: \beta_2 - \beta_3 = 0$

H_1 : H_0 es falsa

Se obtuvo el valor de t igual a -5.527253 y el valor t_c es igual a 1.647414, por lo tanto se acepta la hipótesis nula.

- $H_0: \beta_4 - \beta_5 = 0$

H_1 : H_0 es falsa

Se obtuvo el valor de t igual a -0.1209283 y el valor t_c es igual a 1.647414, por lo tanto se acepta la hipótesis nula.

- $H_0: \beta_2 - \beta_4 = 0$

H_1 : H_0 es falsa

Se obtuvo el valor de t igual a 82.94039 y el valor t_c es igual a 1.647414, por lo tanto se rechaza la hipótesis nula.

- $H_0: \beta_3 - \beta_5 = 0$

H_1 : H_0 es falsa

Se obtuvo el valor de t igual a 21.02848 y el valor t_c es igual a 1.647414, por lo tanto se rechaza la hipótesis nula.

Por otra parte se obtuvieron los siguientes intervalos de confianza:

■ Intervalos de confianza para los parámetros:

$$-0,1288061 \leq \beta_1 \leq 0,8646354$$

$$1,61667 \leq \beta_2 \leq 1,680829$$

$$2,135864 \leq \beta_3 \leq 2,499009$$

$$-0,006318442 \leq \beta_4 \leq 0,00505262$$

$$1,61667 \leq \beta_5 \leq 1,680829$$

- Intervalo de confianza para las predicciones, con los siguientes valores es:

- North.America= 15
- Japan=22
- Review=91
- Rank=54

es:

$$72,23653 \leq \theta_o \leq 79,79031$$

- Intervalo para una observación en particular, para los siguientes valores:

- North.America= 31
- Japan=10
- Review=27
- Rank=25

es

$$72,54777 \leq y_o \leq 76,71422$$

Para el segundo modelo:

Recordemos el modelo que se está estimando:

$$Global = \beta_0 + \beta_1 Misc + \beta_2 Shooter + \beta_3 Sports + \beta_4 Review + \beta_5 Rank + \beta_6 PS3 + \beta_7 X360 + \beta_8 Wii$$

Así, los β de la estimación son:

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4.3832038	1.5492104	2.829	0.004823	**
Rank	-0.0036533	0.0002957	-12.356	< 2e-16	***
PlatformPS3	0.2584559	0.4493037	0.575	0.565350	
PlatformWii	1.7525002	0.4896929	3.579	0.000373	***
PlatformX360	0.4066203	0.4273609	0.951	0.341755	
GenreMisc	-0.0585138	0.5253726	-0.111	0.911356	
GenreShooter	0.2763647	0.4463488	0.619	0.536044	
GenreSports	0.3520346	0.4187446	0.841	0.400861	
Review	0.0130858	0.0176388	0.742	0.458456	

De donde se tienen las siguientes interpretaciones:

- $\hat{\beta}_4$ y $\hat{\beta}_5$: que son los coeficientes de las variables *Review* y *Rank* respectivamente, tienen la misma interpretación que en el modelo anterior.
- $\hat{\beta}_1$: En promedio, manteniendo todas las otras variables constantes, que el videojuego sea de la categoría "Misterio" genera un cambio negativo en las ventas globales de 0,059
- $\hat{\beta}_2$: En promedio, manteniendo todas las otras variables constantes, que el videojuego sea de la categoría "Shooter" genera un cambio positivo en las ventas globales de 0,2763
- $\hat{\beta}_3$: En promedio, manteniendo todas las otras variables constantes, que el videojuego sea de la categoría "Sports" genera un cambio positivo en las ventas globales de 0,3520

Por otro lado:

- $\hat{\beta}_6$: En promedio, manteniendo todas las otras variables constantes, que el videojuego sea jugado en la plataforma "PS3" genera un cambio positivo en las ventas globales de 0,2584
- $\hat{\beta}_7$: En promedio, manteniendo todas las otras variables constantes, que el videojuego sea jugado en la plataforma "X360" genera un cambio positivo en las ventas globales de 0,4066
- $\hat{\beta}_8$: En promedio, manteniendo todas las otras variables constantes, que el videojuego sea jugado en la plataforma "Wii" genera un cambio positivo en las ventas globales de 1,7525

Una vez interpretados los coeficientes del modelo de regresión, podemos observar los valores p de la tabla de coeficientes para cada uno de los β , así podemos concluir que las variables: *Rank* y *Wii* además del *intercepto* son variables significativas para el modelo pues su $vp < 0,05$.

Luego, los intervalos de confianza para cada uno de los β , son:

	2.5 %	97.5 %
(Intercept)	1.340586652	7.425820975
Rank	-0.004233981	-0.003072567
PlatformPS3	-0.623967152	1.140879026
PlatformWii	0.790753450	2.714246937
PlatformX360	-0.432707685	1.245948291
GenreMisc	-1.090334713	0.973307077
GenreShooter	-0.600254995	1.152984419
GenreSports	-0.470371076	1.174440280
Review	-0.021556466	0.047728144

Se realizaron diferentes pruebas de hipótesis para descubrir significancia conjunta entre las variables, para ello se utilizó la prueba F para unas y la prueba ML para otras, así:

1. Usando la prueba F

a) Probemos si es conjuntamente significativo β_2 y β_3

Al encontrar el estadístico $F = 0,3863169$ y $F_c = 3,010943$, notamos que se cumple que:

$$F < F_c$$

Por lo que se acepta H_0 , es decir, la categoría Shooter y Sports no son conjuntamente significativos

b) Probemos si es conjuntamente significativo β_1 y β_4

Al encontrar el estadístico $F = 0,2911129$ y $F_c = 3,010943$, notamos que se cumple que:

$$F < F_c$$

Por lo que se acepta H_0 , es decir, la categoría Misterio y la variable Review no son conjuntamente significativos

c) Probemos si es conjuntamente significativo β_5 y β_8

Al encontrar el estadístico $F = 86,97604$ y $F_c = 3,010943$, notamos que se cumple que:

$$F > F_c$$

Por lo que se rechaza H_0 , es decir, la categoría Wii y la variable Rank son conjuntamente significativos

d) Probemos si es conjuntamente significativo β_1 , β_3 y β_4

Al encontrar el estadístico $F = 0,5617209$ y $F_c = 2,619955$, notamos que se cumple que:

$$F < F_c$$

Por lo que se acepta H_0 , es decir, la categoría Misterio, Sports y la variable Review no son conjuntamente significativos

e) Probemos si es conjuntamente significativo β_1 , β_5 y β_8

Al encontrar el estadístico $F = 58,18198$ y $F_c = 2,619955$, notamos que se cumple que:

$$F > F_c$$

Por lo que se rechaza H_0 , es decir, la categoría Wii, Misterio y la variable Rank son conjuntamente significativos

2. Usando la prueba ML

a) Probemos si es conjuntamente significativo β_1 y β_2

Al encontrar el estadístico $ML = 0,5280828$ y usar la función *pchisq* implementada en R para encontrar el valor p, con el cual se rechazará o aceptará la hipótesis nula, así:

```
v_p <- pchisq(ML, 2, lower.tail = FALSE)
v_p
```

Se obtienen que $vp = 0,7679417$, es decir:

$$vp > 0,05$$

Por lo que se acepta H_0 , es decir, la categoría MISC Y SHOOTER no son significativos conjuntamente

b) Probemos si es conjuntamente significativo β_3 y β_4

Se tiene que: $ML = 1,287297$ y $vp = 0,525372$, es decir:

$$vp > 0,05$$

Por lo que se acepta H_0 , es decir, la categoría Sports y la variable Review no son significativos conjuntamente

c) Probemos si es conjuntamente significativo β_3 , β_2 y β_1

Se tiene que: $ML = 1,179811$ y $vp = 0,7578503$, es decir:

$$vp > 0,05$$

Por lo que se acepta H_0 , es decir, la categoría Sports, MICS, SHooter no son conjuntamente significativos.

d) Probemos si es conjuntamente significativo β_1 , β_2 y β_4

Se tiene que: $ML = 1,188453$ y $vp = 0,755775$, es decir:

$$vp > 0,05$$

Por lo que se acepta H_0 , es decir, la categoría MICS, SHooter y la variable review no son conjuntamente significativos

Interacciones entre variables binarias:

Se consideraron las siguientes variables:

■ Action y Misc

Tenemos lo siguiente:

$$\begin{aligned} E(Global/Action = 1) &= \beta_0 + \beta_4 Review + \beta_5 Rank + \beta_6 PS3 + \beta_7 X360 + \beta_8 Wii \\ E(Global/Misc = 1) &= \beta_0 + \beta_1 + \beta_4 Review + \beta_5 Rank + \beta_6 PS3 + \beta_7 X360 + \beta_8 Wii \end{aligned}$$

De donde se sigue que la diferencia en el valor esperado de la v.dependiente Global entre las categorías Action y Misc es:

$$\beta_1 = -0,0585138$$

■ Misc y Shooter

Tenemos lo siguiente:

$$E(Global/Misc = 1) = \beta_0 + \beta_1 + \beta_4 Review + \beta_5 Rank + \beta_6 PS3 + \beta_7 X360 + \beta_8 Wii$$

$$E(Global/Shooter = 1) = \beta_0 + \beta_2 + \beta_4 Review + \beta_5 Rank + \beta_6 PS3 + \beta_7 X360 + \beta_8 Wii$$

De donde se sigue que la diferencia en el valor esperado de la v.dependiente Global entre las categorías Misc y Shooter es:

$$\beta_1 - \beta_2 = -0,0585138 - 0,2763647 = -0,3348785$$

- Misc y PS3:

Tenemos el siguiente valor esperado:

$$E(Global/Misc = 1, PS3 = 1) = \beta_0 + \beta_1 + \beta_4 Review + \beta_5 Rank + \beta_6$$

- Sports y Wii:

Tenemos el siguiente valor esperado:

$$E(Global/Sports = 1, Wii = 1) = \beta_0 + \beta_3 + \beta_4 Review + \beta_5 Rank + \beta_8$$

- Action y X360:

Tenemos el siguiente valor esperado:

$$E(Global/Action = 1, X360 = 1) = \beta_0 + \beta_4 Review + \beta_5 Rank + \beta_7$$

0.4. Conclusiones

Al tener un valor de $R^2 > 0.80$ para el primer modelo, se tiene que es un buen modelo.

Los parámetros β_2 y β_3 que corresponden a las variables North.America y Japan son significativos para el primer modelo, por otra parte las pruebas de hipótesis realizadas ver si son conjuntamente significativas se obtuvo que todas lo son.

En las pruebas de hipótesis tenemos que β_5 no es significativo, pero al realizar la prueba de hipótesis conjunta entre β_2, β_3 y β_5 tenemos que es significativo, por lo que se debe también analizar si son conjuntamente significativos y no solo ver si son individualmente significativos antes de decidir sacar o no una variables del modelo.

Para el segundo modelo se encontró que el valor $R^2 = 0,2298$, es decir puesto que es un valor menor a 0,8 podemos concluir que no es un modelo que explique bien a las ventas globales de videojuegos.

Al realizar las pruebas de hipótesis para demostrar significancia de las variables en el modelo se encontró que la categoría *Wii* de plataforma y la variable *Rank* e *Intercepto* son las únicas variables que son individualmente significativas para el segundo modelo. Por otro lado, al realizar las pruebas de hipótesis para verificar significancia conjunta entre las variables, se puede concluir que las variables *Wii* y *Rank* son conjuntamente significativas, mientras que las demás variables no son significativas conjuntamente.

Además, se encontró que cualquier variable con la que se intente buscar significancia conjunta usando las variables *Wii* y *Rank*, será también significativa.

Al realizar la interacción entre las variables binarias que pertenecen a una misma categoría, se tiene que la diferencia del valor esperado de la v.dependiente Global va a ser la resta de sus respectivos parámetros (β) de cada una de las variables, a excepción de la variable que se le toma como base, en este caso la diferencia del valor esperado de la variable base y otra variable va a ser el parámetro que le corresponde a la segunda variable.

Bibliografía

- [1] Kaggle. "*Global Video Game Sales and Reviews*",
Recuperado de: <https://www.kaggle.com/datasets/thedevastator/global-video-game-sales-and-reviews>
- [2] Autor desconocido. (n.d.). *Factors that Impact Video Game Sales*. CAUSEweb.,
Recuperado de: <https://www.causeweb.org/usproc/sites/default/files/usclap/2017-2/FactorsthatImpactVideoGameSales.pdf>
- [3] Yufa, A., Yu, J. L., Chan, H., and Berger, P. D. (2019). *Predicting Global Video-Game Sales*. *Journal of Research in Business and Management*, 7(3), 60-64.,
Recuperado de <https://www.questjournals.org/jrbm/papers/vol7-issue3/I07036064.pdf>