



**ESTADÍSTICA APLICADA II**

**PROYECTO**

“CORRELACIÓN ENTRE LAS CINTAS QUE GANARON UN OSCAR A MEJOR PELÍCULA, CON RESPECTO A LA PROPORCIÓN DE PALABRAS DICHAS POR MUJERES EN LA PELÍCULA Y LA CALIFICACIÓN QUE LE DIO LA AUDIENCIA EN LA PÁGINA DE ROTTEN TOMATOES”

**PROFESOR**

VÍCTOR MANUEL GUERRERO GUZMÁN

**EQUIPO 3**

ADRIANA ALAVEZ LUJANO 163480

SOFÍA OLIVA RUIZ 164595

DAVID ISAAC LÓPEZ ROMERO 173993

DIEGO CARLOS KRAFFT DE SILVA 173246

SOFÍA ALEJANDRA DÍAZ MIRANDA 172360

OTOÑO 2021

***Índice.***

[**Introducción**](#_xo11vnz4hq8e) **2**

[**Descripción de la información**](#_59curverub1a) **2**

[**Modelado e Implementación**](#_j2k8oq9uiowg) **3**

[**Interpretación de resultados**](#_7f2ci8abprn9) **10**

[**Conclusiones**](#_ypwc6xxtm1xg) **11**

[**Referencias bibliográficas**](#_t06vf7c9xdta) **12**

[**Apéndice o anexo**](#_ci3lz2hbca79) **12**

# Introducción

Los Premios Óscar son los más reconocidos en la industria cinematográfica. Fueron fundados en 1927 por la Academia de las Artes y las Ciencias Cinematográficas. Las categorías más reconocidas son: mejor actor, mejor actriz, mejor director y mejor película. Para el análisis que nos interesa en el desarrollo de este proyecto, nos enfocaremos en el premio a la mejor película. Esta categoría es considerada la más importante.

La página web Rotten Tomatoes tiene como función dar una crítica a películas de todo tipo. Lo hace dándole una calificación, la máxima puede ser 100. Asimismo, los aficionados pueden votar. Es por esto que hay dos calificaciones; la de los críticos y la de la audiencia.

Lo que buscamos en la construcción de este modelo es encontrar si existe una correlación entre las cintas que ganaron un Oscar a mejor película, con respecto a un ratio de palabras dichas por mujeres en la película y la calificación que le dio la audiencia en la página de Rotten Tomatoes que una de las páginas más reconocidas de críticas del cine.

# Descripción de la información

Planteamiento

Los Premios Óscar es un premio anual concedido ​en reconocimiento a la excelencia y activismo social de los profesionales en la industria cinematográfica que incluye actores, directores y escritores, ampliamente considerado el máximo honor en el cine. ​El Oscar se conoce, oficialmente, como el «Premio de la Academia al Mérito», y es el principal de los nueve premios que otorga dicha organización.son los más reconocidos en la industria cinematográfica. Fueron fundados en 1927 por la Academia de las Artes y las Ciencias Cinematográficas. Las categorías más reconocidas son: mejor actor, mejor actriz, mejor director y mejor película.

Las películas que ya han recibido este honor son 93% pertenecientes al género de drama y 2% al de acción y fantasía. Asimismo, 71% de los ganadores tenían un director o algún miembro del reparto que anteriormente había sido nominado para un Oscar y 52% de los ganadores provienen de los seis estudios de cine más importantes de Hollywood. También cabe mencionar que el 82% de los ganadores recibieron una calificación que supera al 80 sobre 100 por parte de Rotten Tomatoes y por la audiencia.

Variables

Para realizar el análisis integramos las siguientes variables:

* Proporción: Palabras dichas por una mujer (En las cintas nominadas a un Oscar a mejor película dentro de nuestra base de datos) dividido entre el número de palabras totales, es decir la suma de palabras dichas por hombres y mujeres
* Rotten Tomatoes: Calificación dada por los críticos de la plataforma rotten tomatoes
* Audiencia: Calificaciòn dada por la audiencia a través de rotten tomatoes

# Modelado e Implementación

**Modelo Inicial**

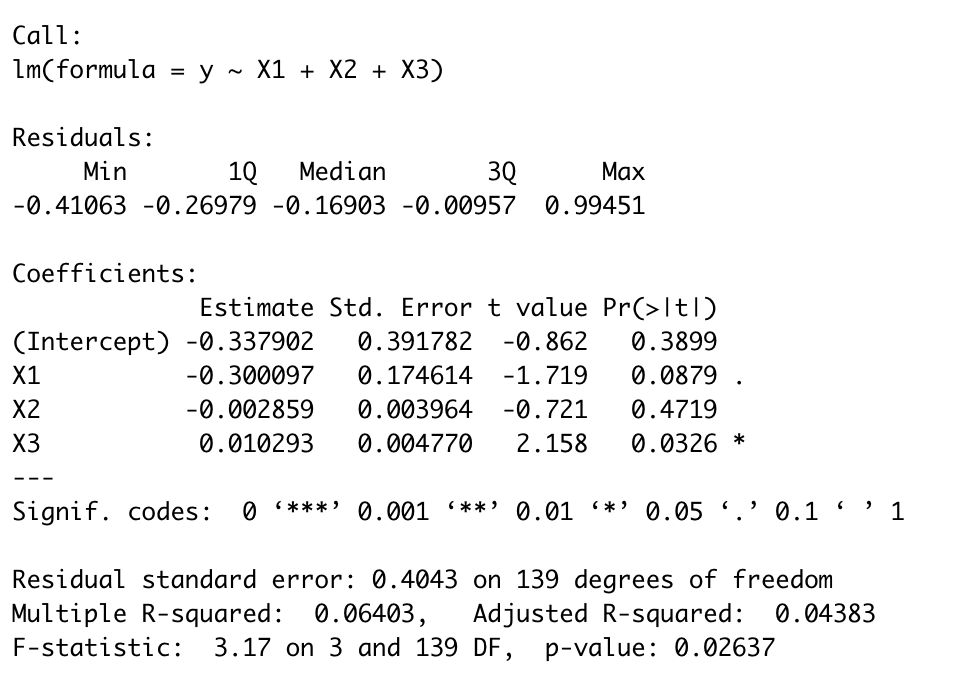
Y = β0 + β1 X1 + β2 X2 +β3 X3 + દ

Y = Probabilidad de ganar un Óscar a mejor película

X1 = Proporción

X2 = Calificación Rotten Tomato

X3 = Calificación Audiencia

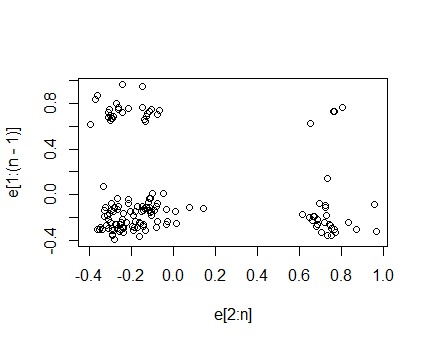


Los resultados de nuestro primer modelo fueron: R^2=0.043 y nuestra p-value=0.026. Esto quiere decir que nuestro modelo sólo está explicado por un 4.3%

**Validación**

**Autocorrelación**

No tenemos variables de ordenamiento. Por lo tanto, no hay autocorrelación



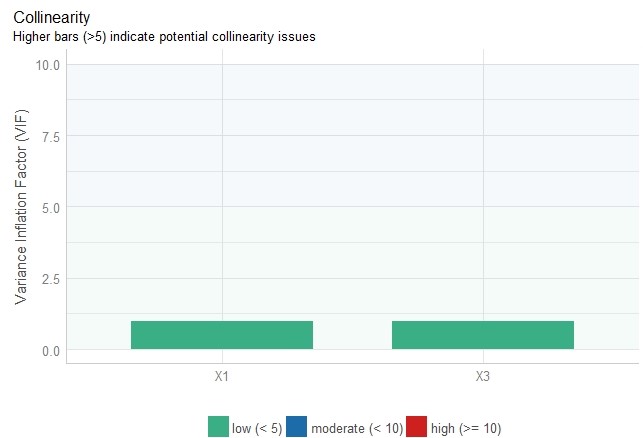
**Colinealidad**

La correlación entre las variables dependientes y la variable independiente es mayor que la correlación entre las variables dependientes entre sí

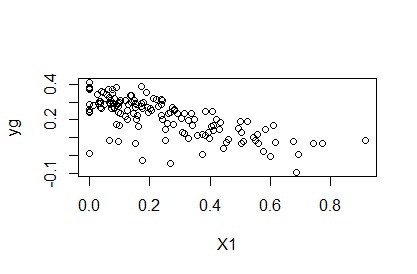
**cor(X1,X3) = 0.008291634**

**cor(X1,y) = 0.04403955**

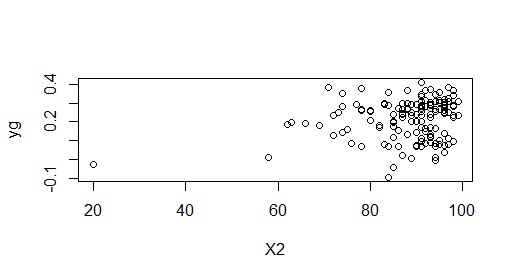
**cor(X3,y) = 0.2371696**

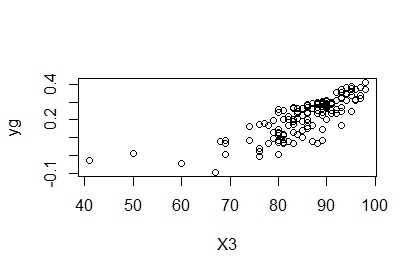
****

Podemos observar una tendencia lineal negativa para X1 y lineal positiva para X3 en las variables

****

La variable X2 no cumple los supuestos de linealidad en los variables por lo que la quitamos del modelo

****

****

**Modelo Corregido 1**

Y = β0 + β1 X1 + β3 X 3+ દ

Y = Probabilidad de ganar un Óscar a mejor película

X1 = Proporción

X3 = Audiencia

**Tabla de análisis de Devianza**

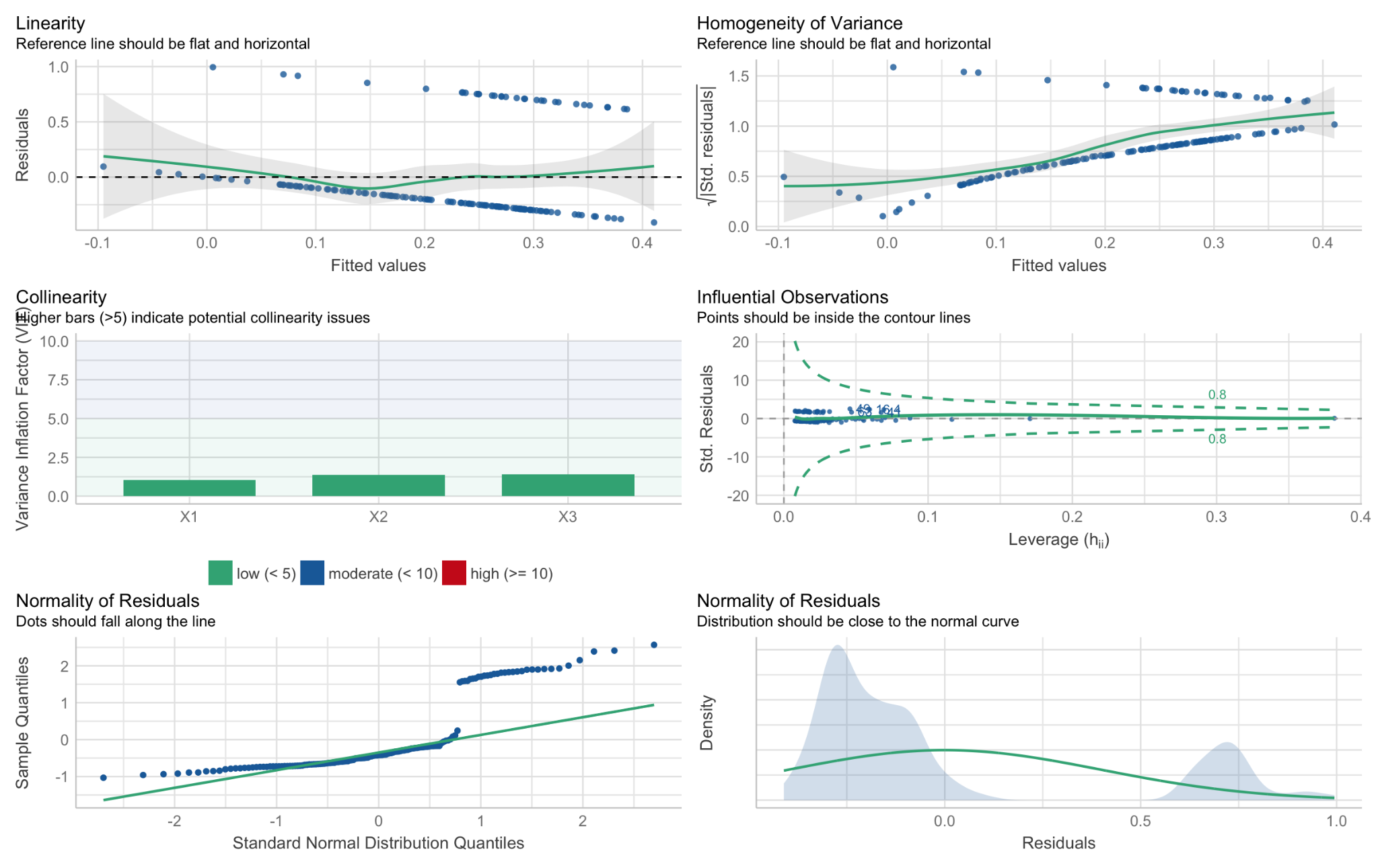
|  | **Grados de Libertad** | **Devianza** | **Residuo Grados de Libertad** | **Residuo Devianza** |
| --- | --- | --- | --- | --- |
| X1 | 1 | 0.7636 | 141 | 23.516 |
| X3 | 1 | 0.70598 | 140 | 22.81 |

**Nuestro análisis nos arrojó una tabla de devianza porque nuestra variable respuesta es dicotómica y la tabla nova usa variable de respuesta continua y predice el valor de la variable, mientras que una regresión logit usa variables dicotómicas y predice el valor sólo de la categoría.**

**Devianza: es una prueba de bondad de ajuste que comprara el modelo saturado, con el interés. El modelo saturado es aquel que ajusta todos los puntos y usa mayor número de parámetro.**

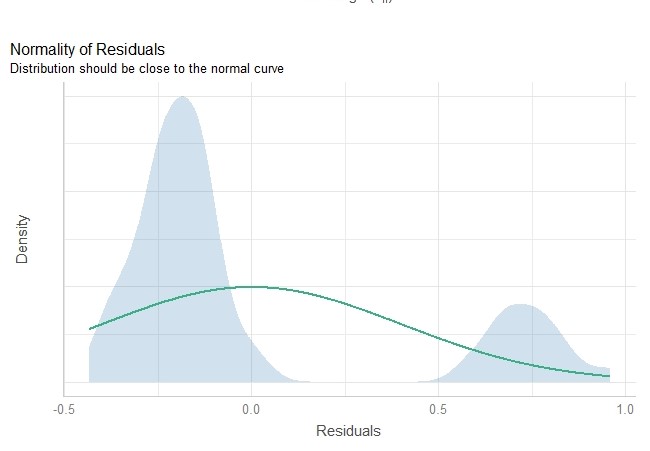
**Normalidad en los Residuos**

**Antes de la corrección del modelo**

****

Después de la correción del Modelo

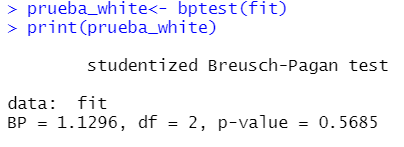
X110 , X35



Debido a la naturaleza de nuestras variables, nuestra mejor aproximación de la normal se presenta en la tabla anterior. Como nuestras variables son dicotómicas, obtenemos residuales que presentan características similares.

**Heterocedasticidad con y**

Debido a que la gráfica entre las variables y los residuos no era clara, decidimos no incluirla en el presente trabajo. De este modo, decidimos utilizar la prueba de de Breusch-Pagan, también conocida como prueba de White



Notamos que la prueba indica un valor-p = 0.5685 > 0.05. Para una prueba de hipótesis, donde la : varianza constante y : varianza no constante. Con el resultado obtenido, podemos concluir que existe mucha heteroscedasticidad en los datos, por lo que no se cumple la hipótesis de tener varianza constante. En consecuencia, tuvimos que modificar el modelo de nuevo para obtener una mejor aproximación de la realidad.

**Modelo Corregido 2**

Y = β0 + β1 X110/3 + β3 X 35/3+ દ

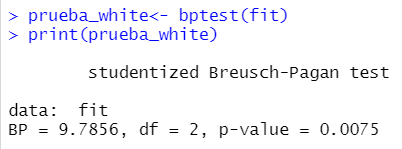
Y = Probabilidad de ganar un Óscar a mejor película

X1 = proporción

X3 = Audiencia

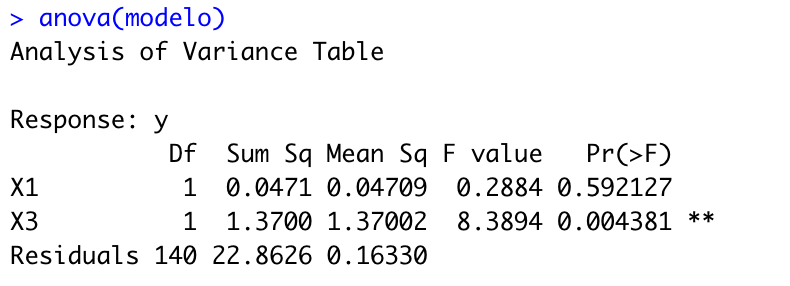
**Heterocedasticidad**

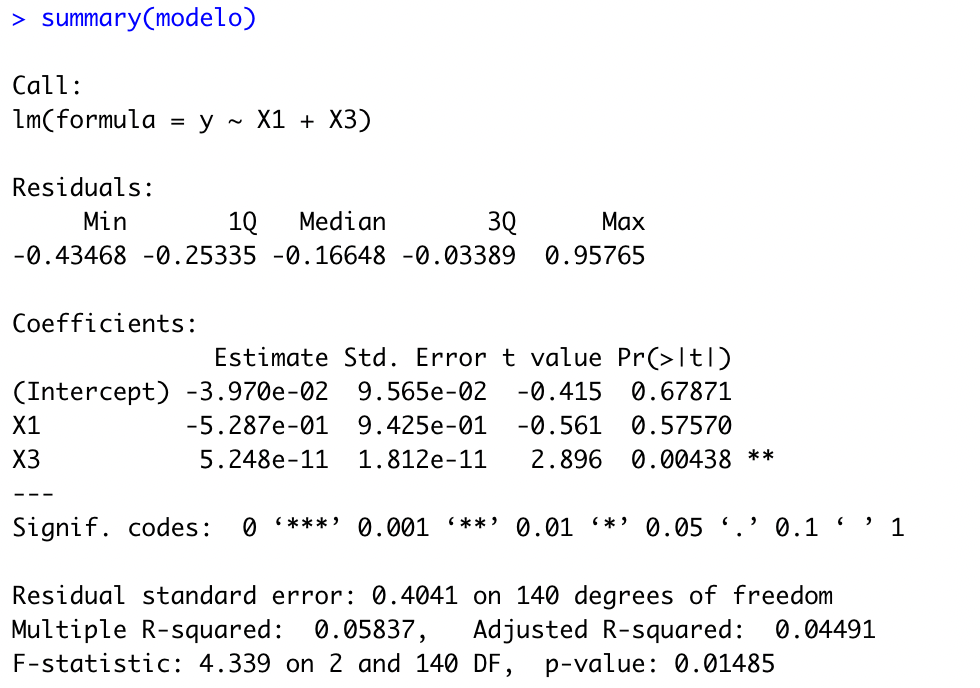
Para mantener consistencia con las pruebas aplicadas al modelo, nuevamente utilizamos la prueba de Breusch-Pagan.

****

Ahora obtuvimos un valor-p = 0.0075 < 0.05.

Para una prueba de hipótesis, donde la : varianza constante y : varianza no constante. Podemos decir que no rechazamos la hipótesis nula. De este modo, no hay heterocedasticidad en el nuevo modelo.

****

****

# Interpretación de resultados

**Resultado de los Coeficientes del modelo**

Y = -0.03970 -1.586\*X110/3 + 1.574e-10\*X 35/3 + દ

B0 es negativa porque tenemos más 0’s que 1’s en por lo que es normal que entre más películas nominadas haya para ganar un oscar menos probabilidad hay de ganarlo y se hace el ajuste al modelo

X1 tiene desde el inicio una tendencia negativa y X3 una tendencia positiva

# Conclusiones

Como vimos al principio no se lograron corregir todos los supuestos del modelo, por lo que no logra explicar lo que buscábamos

Se requiere un análisis teórico para ver este tipo de linealidades con variables dicotómicas

La calificación de la audiencia tiene una tendencia positiva con respecto a la probabilidad de ganar un oscar sin embargo es negativa respecto a la proporción de palabras dichas por una mujer.

# Referencias bibliográficas

*Premios Oscar: Historia y Curiosidades*. Muy interesante. En línea. 27/11/21.

<https://www.muyinteresante.es/cultura/especiales/premios-oscar-historia-y-curiosidades>

*Hay una fórmula para ganar los Oscar y está todo en las estadísticas*. Businessinsider. En línea. 27/11/21.

<https://www.businessinsider.es/hay-formula-ganar-oscar-todo-estadisticas-379059>

# *Apéndice* o *anexo*