**MÓDULO 5: ACTIVIDAD GRUPAL 1: MODELOS LINEA GENERALIZADOS. INFERENCIA**

**Alumnos:** Jorge Cano yYasmin Al Omar Mesnaoui

**Curso:** 2024-2025

Master en Big Data & Data Science

**DESCRIPCIÓN DE LA TAREA**

Contamos con los datos de número de casos de “affairs” (Encuentros amorosos fuera de la pareja / infidelidades).

La base de datos contiene:

* naffairs: number of affairs within last year.
* kids: 1=have children;0= no children.
* vryunhap: (1/0) very unhappily married.
* unhap: (1/0) unhappily married.
* avgmarr: (1/0) average married.
* hapavg: (1/0) happily married.
* vryhap: (1/0) very happily married.
* antirel: (1/0) anti religious.
* notrel: (1/0) not religious.
* slghtrel: (1/0) slightly religious.
* smerel: (1/0) somewhat religious.
* vryrel: (1/0) very religious.
* yrsmarr1: (1/0) >0.75 yrs.
* yrsmarr2: (1/0) >1.5 yrs.
* yrsmarr3: (1/0) >4.0 yrs.
* yrsmarr4: (1/0) >7.0 yrs.
* yrsmarr5: (1/0) >10.0 yrs.
* yrsmarr6: (1/0) >15.0 yrs.

**PREGUNTAS**

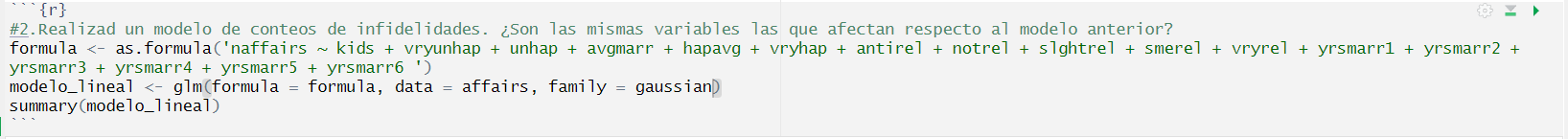
1. **¿Cuáles son las variables que definen la probabilidad de ser o no infiel?**
2. **Realizad un modelo de conteos de infidelidades. ¿Son las mismas variables las que afectan respecto al modelo anterior?**
3. **Argumentad la razón por la que no podemos incluir todas las variables yrsmarr dentro del modelo.**
4. **Calculad la frecuencia de infidelidades de una persona con más de 10 años de matrimonio, no-religioso, sin hijos. Sin saber el grado de satisfacción con el matrimonio.**
5. **¿Podríais calcular el 95% de confianza de la estimación anterior?**
6. **Calculad, bajo qué nivel de confianza, los residuos de la distribución pueden considerarse normales.**
7. **Calculad si la combinación de Años de Matrimonio e Hijos da nueva información a nuestro modelo.**
8. **Teniendo la combinación de Años de Matrimonio e Hijos metido en el modelo, ¿cuál sería el cambio en infidelidades de no tener hijos a tener hijos?**
9. **Calculad una variable que convierta las dummies de años de matrimonio en numérica. Calculad también esta misma variable al cuadrado. ¿En teoría hay alguna edad de matrimonio en la que cada año adicional suponga un descenso de las infidelidades?**

**1.  ¿Cuáles son las variables que definen la probabilidad de ser o no infiel?**

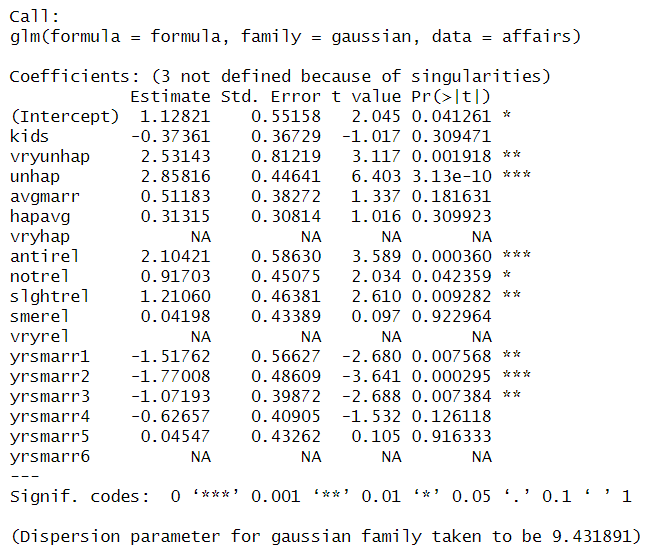
Las variables que definen la probabilidad de ser o no infiel son: kids, vryunhap, unhap, avgmarr, hapavg, vryhap, yrsmarr1 y yrsmarr6. La razón para escoger estas variables porque el hecho de tener hijos puede ser una razón de gran peso para ser infiel o no a tu pareja, también otra razón son si tu matrimonio es feliz o no; porque si tu estas dentro de un matrimonio infeliz hay más probabilidades de ser infiel a tu pareja. Otra razón puede ser los años en que una pareja está casada, porque hay personas que se aburren cuando la relación se vuelve monótona.

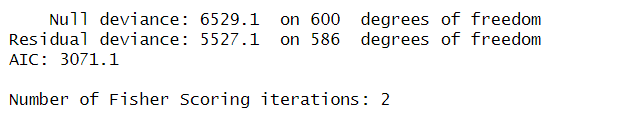
**2. Realizad un modelo de conteos de infidelidades. ¿Son las mismas variables las que afectan respecto al modelo anterior?**

**Código:**



**Resultados:**



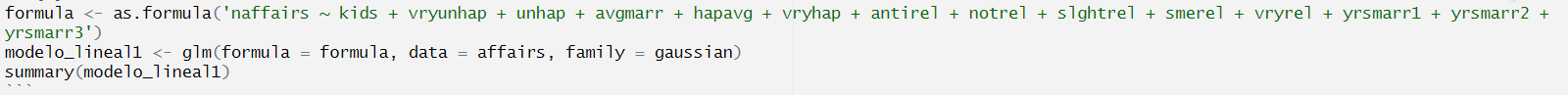


**Conclusiones:** Si creamos un modelo lineal en el que se incluye todas la variables, y se establece el nivel de significancia en 0.05; se puede ver que las variables vryunhap, unhap, antirel, notrel, slghtrel, yrsmarr1, yrsmarr2, y yrsmarr3. Esto significa que el hecho de estar en un matrimonio infeliz, no ser religioso o no ser muy religioso y llevar menos de 5 años casado pueden influenciar en el hecho de ser infiel a tu pareja. Si se comparan estos resultados con las variables con el anterior ejercicio, se observa que algunas variables coincidieron pero otras como la variable kids no, porque según el modelo el hecho de tener hijos no afecta directamente al hecho de ser infiel o no.

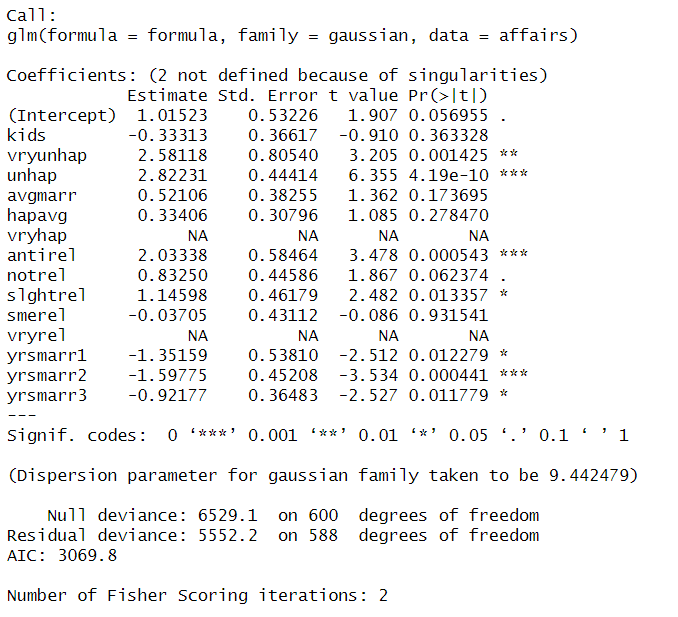
**3. Argumentad la razón por la que no podemos incluir todas las variables yrsmarr dentro del modelo.**

La razón por las que no se pueden incluir las todas las variables yrsmarr es por tu tema colinealidad, esto pasa cuando muchas variables son combinaciones lineales de otras variables; esto provoca problemas de redundancia en el modelo. Entonces para evitar esto hay que sacar las variables yrsmarr que no son significativa, y trabajar con las variables que si son de significancia como por ejemplo yrsmarr1, yrsmarr2 y yrsmarr3.

**Código:**

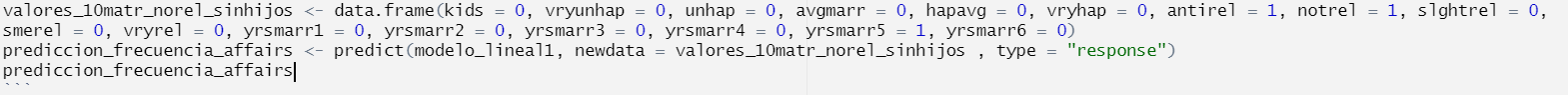


**Resultados:**



**4.** **Calculad la frecuencia de infidelidades de una persona con más de 10 años de matrimonio, no-religioso, sin hijos. Sin saber el grado de satisfacción con el matrimonio.**

**Código:**



**Resultados:**



**Conclusiones:**

La predicción es 3.88, representa la frecuencia de infidelidades con las condiciones especificadas. Para ver si esta predicción es alta o bajo, hay que calcular la media de infelidades; si la media es más baja que la predicción esto signifaca que la predicción es basntate alta.

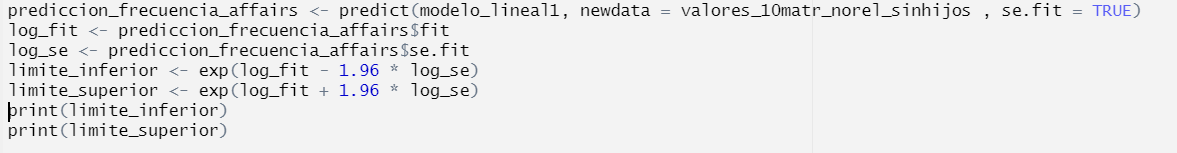




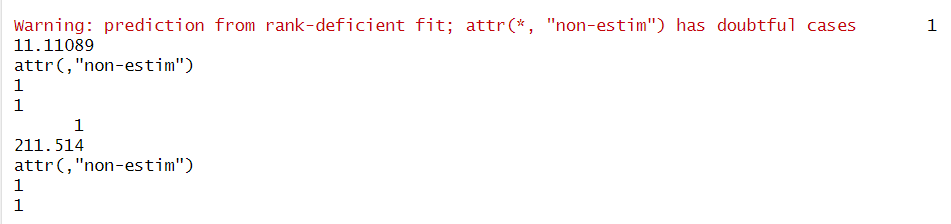
La media es de 1.45, por lo tanto la frecuencia de que una persona con más de 10 años de matrimonio, no religiosa y sin hijos sea infiel es bastante alta, porque el valor de la predicción es más grande que la media.

**5. ¿Podríais calcular el 95% de confianza de la estimación anterior?**

**Código:**



**Resultados:**



1. **Calculad, bajo qué nivel de confianza, los residuos de la distribución pueden considerarse normales**.

**Calculad si la combinación de Años de Matrimonio e Hijos da nueva información a nuestro modelo.**

**Teniendo la combinación de Años de Matrimonio e Hijos metido en el modelo, ¿cuál sería el cambio en infidelidades de no tener hijos a tener hijos?**

**Calculad una variable que convierta las dummies de años de matrimonio en numérica. Calculad también esta misma variable al cuadrado. ¿En teoría hay alguna edad de matrimonio en la que cada año adicional suponga un descenso de las infidelidades?**