1. Ajuste un modelo de regresión poisson, ¿por qué es adecuado para este tipo de problemas?

Table 1: Información en análisis

Publicaciones (Y_i)	Horas_dedicadas (X_1)	Experiencia (X_2)	Departamento (W)
12	17.19762	7.139031	Naturales
42	18.84911	6.989244	Sociales
32	27.79354	6.245845	Naturales
10	20.35254	5.789947	Naturales
22	20.64644	8.251362	Naturales

- Variables de conteo ; $A_y \geq 0$
- (2). Conteos son independientes
- (3). Varianza es constante

- Función de enlace : $\ln(\cdot)$: Para garantizar una respuesta positiva.
- Modelación: $Y_i \sim pois(\lambda_i)$; $f(y_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{\lambda_i!}$: Proceso de máxima verosimilitud.
- $E[Y_i] = \lambda$; $Var[Y_i] = \lambda$; Por MLE: $\hat{\lambda} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}$. No se modela directamente Y_i
- $\ln(E[Y_i|X\ldots]) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}$ (No se usa el término de error)
- $\ln(E[Y_i|X\ldots]) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 W$: $\ln(\lambda_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 W$
- $\ln(\lambda_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 I_{i2} + \beta_4 I_{i3}$: $\ln(\hat{\lambda}_i) = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 I_{i2} + \hat{\beta}_4 I_{i3}$

Naturales Sociales

```
# ──────────────
#     PRIMER PUNTO
# ──────────────
datos <- read.csv(file.choose())
modelo <- glm(Publicaciones ~ ., data = datos,
              family = poisson(link = "log"))
summary(modelo)
```

```
Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)            1.207557   0.059143   20.42   <2e-16
Horas_dedicadas        0.078322   0.002111   37.10   <2e-16
Experiencia            0.052331   0.002571   20.35   <2e-16
DepartamentoNaturales -0.339735   0.028296  -12.01   <2e-16
DepartamentoSociales   0.513795   0.023444   21.92   <2e-16
```

Significancia individual: $\begin{cases} H_0: \beta_j = 0 & j = 0, \ldots, 4 \\ H_1: \beta_j \neq 0 \end{cases}$

2. Realice la prueba de significancia global.

$\chi^2_c : D_0^2 - D^2$; $R_c : \{ \chi^2_{p,\alpha} > \chi^2_c \}$

$\chi^2_{k,\alpha}$; $k = 4$

$H_0: \beta_1 = \ldots = \beta_k = 0$

$H_1:$ Algún $\beta_j \neq 0$; $j = 1, \ldots, k$

- Modelo reducido: $\ln(\lambda_i) = \beta_0$

```
> anova(modelo_reducido, modelo, test = "Chisq")
Analysis of Deviance Table

Model 1: Publicaciones ~ 1
Model 2: Publicaciones ~ Horas_dedicadas + Experiencia + Departamento
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       299    2976.40
2       295     299.02  4   2677.4 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# ──────────────
#     SEGUNDO PUNTO
# ──────────────
# Definir un modelo completo y un modelo reducido
modelo_reducido <- glm(Publicaciones ~ 1, data = datos,
                       family = poisson(link = "log"))
# Se debe definir la función anova con el test chi- cuadrado
anova(modelo_reducido, modelo, test = "Chisq")
```

Interpretación β_j : la interpretación respecto a el promedio de $\ln(\lambda_i)$

- β_j, para $j > 0$, representa el cambio en el logaritmo de la media cuando x_j aumenta en una unidad y todas las demás covariables se mantienen constantes. Por lo tanto, e^{β_j} representa el cambio en la media.

  En el ejemplo anterior, β_0 no tiene interpretación práctica, pues la presión de la pistola de pintura es siempre mayor que cero y las unidades son producidas en alguna

- $\hat{\beta}_1$: el cambio en el logaritmo natural de la media (publicaciones) cuando se aumentan las horas dedicadas en una unidad, con todas las en una unidad y todas las demás covariables se mantienen constantes. Por lo tanto, e^{β_j} representa el cambio en la media.

En el ejemplo anterior, β_0 no tiene interpretación práctica, pues la presión de la pistola de pintura es siempre mayor que cero y las unidades son producidas en alguna de las dos plantas disponibles. En resumen, la variables predictoras no asumen el valor de cero bajo ninguna circunstancia.

Cuando se aumentan las horas cuadradas en una unidad, con todas las demás covariables constantes

4. ¿Cuál es un valor de λ estimado para este problema?

$$\boxed{\lambda_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 I_{i2} + \beta_4 I_{i4})}$$

(Todas las covariables son significativas)

$$\boxed{X_1 = 5;\ X_2 = 5;\ I_{i2}: \text{Naturales}}$$

```
> # ──────────────────────
> #      CUARTO PUNTO
> # ──────────────────────
> lambda <- exp(1.207 + 0.07 * 5 + 0.05 * 5 - 0.33 * 1)
> lambda
[1] 4.379787
```