

EXAMEN MJB

María Pallares Diez

2025-06-09

ENUNCIADO

Se llevo a cabo un estudio experimental con el objetivo de analizar como influye el nivel de estrés (categorizado) percibido y la edad del participante, en la cantidad de errores cometidos durante la ejecución de una tarea cognitiva repetitiva.

Un total de 100 individuos participaron en el estudio, cada uno realizando la tarea en 5 sesiones distintas, registrando: individuo, nestres, edad y errores.

En el fichero DatosPruebaJunio.RData esncontrarás en objeto “datos” que contiene el dataset para llevar a cabo las siguientes tareas:

```
rm(list = ls())
library(R2WinBUGS)
```

```
## Warning: package 'R2WinBUGS' was built under R version 4.4.3
```

```
## Cargando paquete requerido: coda
```

```
## Warning: package 'coda' was built under R version 4.4.3
```

```
## Cargando paquete requerido: boot
```

```
## Warning: package 'boot' was built under R version 4.4.3
```

```
load("DatosPruebaJunio.RData")
summary(datos)
```

```
##      individuo  nestres      edad      errores
## 1      : 5    1:165  Min.   :13.00  Min.    : 6.00
## 2      : 5    2:160  1st Qu.:34.00  1st Qu.: 24.75
## 3      : 5    3:175  Median :40.00  Median : 36.00
## 4      : 5           Mean  :40.27  Mean   : 42.58
## 5      : 5           3rd Qu.:47.00  3rd Qu.: 54.00
## 6      : 5           Max.   :72.00  Max.   :178.00
## (Other):470
```

```
str(datos)
```

```
## 'data.frame': 500 obs. of 4 variables:
## $ individuo: Factor w/ 100 levels "1","2","3","4",...: 1 1 1 1 1 2 2 2 2 2 ...
## $ nestres : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
## $ edad : num 37 30 29 43 44 41 49 61 35 17 ...
## $ errores : int 45 23 12 43 48 72 115 178 47 18 ...
```

TAREA 1

Ajusta el modelo adecuado para explicar el numero de errores cometidos en la prueba cognitiva en función únicamente del nivel de estrés. Interpreta brevemente los resultados.

MODELO:

```
datos$estres <- as.numeric(as.factor(datos$nestres)) # 1, 2, 3
N <- nrow(datos)

modelo_estrés_cat <- function(){
  for (i in 1:N){
    errores[i] ~ dpois(lambda[i])
    log(lambda[i]) <- alpha + beta[estres[i]]
  }
  beta[1] <- 0
  for (j in 2:3){
    beta[j] ~ dnorm(0, 1.0E-6)
  }
  alpha ~ dnorm(0, 1.0E-6)
}
```

INICIALES Y BUGGS:

```
datos.bugs <- list(
  errores = datos$errores,
  estres = datos$estres,
  N = N
)

iniciales <- function(){
  list(alpha = rnorm(1), beta = c(NA, rnorm(2)))
}

parametros <- c("alpha", "beta")

res.estrés.cat <- bugs(
  data = datos.bugs,
  inits = iniciales,
  parameters.to.save = parametros,
  model.file = modelo_estrés_cat,
  n.iter = 10000, n.burnin = 1000
)

save(res.estrés.cat, file = "ResultadoEstrésCat.Rdata")

load("ResultadoEstrésCat.Rdata")
round(res.estrés.cat$summary, 3)
```

##	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat
## alpha	3.455	0.014	3.426	3.446	3.455	3.464	3.481	1.002
## beta[2]	0.287	0.019	0.252	0.274	0.286	0.300	0.325	1.005
## beta[3]	0.520	0.017	0.486	0.508	0.520	0.532	0.554	1.003
## deviance	8913.925	2.510	8911.000	8912.000	8913.000	8915.000	8920.000	1.008

```
##          n.eff
## alpha      810
## beta[2]    400
## beta[3]    550
## deviance   690
```

```
res.estrés.cat$DIC
```

```
## [1] 8917
```

Se ha ajustado un modelo Poisson para explicar el número de errores cometidos (**errores**) en función del nivel de estrés percibido (**nestres**), tratado como variable categórica con 3 niveles (1 = bajo, 2 = medio, 3 = alto). Se ha considerado la categoría de **estrés bajo** como referencia, fijando $B1 = 0$ (*corner constraint*), y estimando los efectos diferenciales para los niveles medio (B2) y alto (B3).

A partir de los parámetros estimados por el modelo:

- Para **estrés bajo** (grupo de referencia), el número medio de errores es:
 - $\exp(3.455) = 31.6$ errores
- Para **estrés medio**, con un incremento estimado de 0.287 sobre el grupo de referencia:
 - $\exp(3.455 + 0.287) = 42.3$ errores
- Para **estrés alto**, con un incremento de 0.520 sobre el grupo de referencia:
 - $\exp(3.455 + 0.520) = 53.3$ errores

Los resultados muestran claramente que el número de errores aumenta con el nivel de estrés. Los participantes con estrés medio cometen unos 10 errores más que los de estrés bajo, y los de estrés alto cometen más de 20 errores adicionales.

Por lo que podemos decir que: Existe una relación positiva entre el estrés percibido y el número de errores cometidos: **cuanto mayor es el estrés, mayor es el deterioro del rendimiento en la tarea cognitiva**. La convergencia del modelo es adecuada, según los valores de Rhat y el número efectivo de muestras (**n.eff**).

TAREA 2

Añade al modelo anterior el efecto de la covariable edad. Interpreta brevemente los resultados y compara el ajuste del modelo obtenido con el ajuste del modelo anterior.

MODELO:

```
modelo_estrés_edad <- function(){
  for (i in 1:N){
    errores[i] ~ dpois(lambda[i])
    log(lambda[i]) <- alpha + beta[estres[i]] + gamma * edad[i]
  }
  beta[1] <- 0
  for (j in 2:3){
    beta[j] ~ dnorm(0, 1.0E-6)
  }
  alpha ~ dnorm(0, 1.0E-6)
  gamma ~ dnorm(0, 1.0E-6)
}
```

INICIALES Y BUGGS:

```
datos$estres <- as.numeric(datos$nestres)
N <- nrow(datos)

datos.bugs2 <- list(
  errores = datos$errores,
  estres = datos$estres,
  edad = datos$edad,
  N = N
)

iniciales2 <- function(){
  list(alpha = rnorm(1), beta = c(NA, rnorm(2)), gamma = rnorm(1))
}

parametros2 <- c("alpha", "beta", "gamma")

res.estrés.edad <- bugs(
  data = datos.bugs2,
  inits = iniciales2,
  parameters.to.save = parametros2,
  model.file = modelo_estrés_edad,
  n.iter = 10000, n.burnin = 1000
)

# Guardar resultados
save(res.estrés.edad, file = "ResultadoEstresEdad.Rdata")

load("ResultadoEstresEdad.Rdata")

# Mostrar la tabla de resumen
round(res.estrés.edad$summary, 3)
```

```
##           mean    sd    2.5%    25%    50%    75%    97.5%  Rhat
## alpha      1.321 0.035    1.257    1.297    1.321    1.343    1.392 1.000
## beta[2]     0.304 0.018    0.271    0.293    0.304    0.316    0.337 1.000
## beta[3]     0.622 0.017    0.589    0.610    0.622    0.634    0.655 1.001
## gamma       0.049 0.001    0.048    0.049    0.049    0.050    0.050 1.000
## deviance 3940.418 2.800 3937.000 3938.000 3940.000 3942.000 3947.000 1.000
##           n.eff
## alpha      1000
## beta[2]    1000
## beta[3]    1000
## gamma      1000
## deviance   1000

# Ver el DIC para comparar con el modelo anterior
res.estrés.edad$DIC

## [1] 3944.38
```

En este modelo se ha ajustado un modelo Poisson para explicar el número de errores cometidos en la tarea cognitiva, incluyendo como predictores:

- El **nivel de estrés percibido**, tratado como variable categórica con tres niveles (bajo como referencia).
- La **edad del participante**, considerada como una covariable continua con efecto lineal.

Los resultados del modelo indican que tanto el nivel de estrés como la edad tienen un efecto claro y positivo sobre el número de errores cometidos.

-Por un lado, los participantes con **estrés medio** cometieron más errores que aquellos con estrés bajo, y este aumento fue todavía mayor en el grupo de **estrés alto**, lo que confirma que el rendimiento cognitivo disminuye a medida que se incrementa la percepción de estrés.

-Por otro lado, la variable **edad** también tuvo un efecto significativo: por cada año adicional de edad, el número esperado de errores aumenta ligeramente. Aunque este efecto es más suave que el del estrés, su contribución es constante y acumulativa, lo que refuerza la idea de que la edad también influye negativamente en el rendimiento en esta tarea cognitiva repetitiva.

-En conjunto, el modelo permite concluir que ambos factores —estrés y edad— afectan de forma clara y creciente a la cantidad de errores cometidos.

El DIC del modelo anterior (solo estrés) era **8917**. El modelo actual, que incluye también la edad, tiene un **DIC de 3944.38**. Por lo que, el nuevo modelo presenta un ajuste **mucho mejor** al incluir la edad como covariable explicativa. Esto sugiere que la edad tiene un papel relevante en el rendimiento cognitivo, más allá del efecto del estrés. Además, la convergencia de la simulación ha sido excelente en todos los parámetros ($R_{\text{hat}} = 1$, $n_{\text{eff}} = 1000$).

TAREA 3

Añade a este modelo anterior el efecto del individuo e indica si lo introduces como efecto fijo o aleatorio. Compara el ajuste de este modelo obtenido con el ajuste de los modelos anteriores

MODELO:

```
modelo_completo <- function() {
  for (i in 1:N) {
    errores[i] ~ dpois(lambda[i])
    log(lambda[i]) <- alpha + beta[estres[i]] + gamma * edad[i] + u[ind[i]]
  }

  # Efecto del estrés (categórica con corner constraint)
  beta[1] <- 0
  for (j in 2:3) {
    beta[j] ~ dnorm(0, 1.0E-6)
  }

  # Priors
  alpha ~ dnorm(0, 1.0E-6)
  gamma ~ dnorm(0, 1.0E-6)

  # Efectos aleatorios individuo
  for (j in 1:K) {
    u[j] ~ dnorm(0, tau.u)
  }
  tau.u <- pow(sd.u, -2)
  sd.u ~ dunif(0, 10)
}
```

INICIALES y BUGGS:

```

# Índices necesarios para BUGS
datos$estres <- as.numeric(datos$nestres)      # Factor → índice
datos$ind <- as.numeric(datos$individuo)      # Índice individuo
datos$edad_c <- scale(datos$edad, center = TRUE, scale = FALSE) # Edad centrada

N <- nrow(datos)
K <- length(unique(datos$individuo))          # Número de individuos

datos.bugs3 <- list(
  errores = datos$errores,
  estres = datos$estres,
  edad = as.numeric(datos$edad_c),
  ind = datos$ind,
  N = N,
  K = K
)

iniciales3 <- function() {
  list(
    alpha = rnorm(1, 0, 1),
    beta = c(NA, rnorm(2, 0, 0.5)),
    gamma = rnorm(1, 0, 0.1),
    u = rnorm(K, 0, 1),
    sd.u = runif(1, 0.1, 1.5)
  )
}

parametros3 <- c("alpha", "beta", "gamma", "sd.u")

res.completo <- bugs(
  data = datos.bugs3,
  inits = iniciales3,
  parameters.to.save = parametros3,
  model.file = modelo_completo,
  n.iter = 50000,
  n.burnin = 10000,
  n.thin = 10,
  n.chains = 3
)

# Guardar resultados
save(res.completo, file = "ResultadoModeloCompleto.Rdata")

```

CARGA Y RESULTADOS:

```

load("ResultadoModeloCompleto.Rdata")

# Mostrar la tabla de resumen
round(res.completo$summary, 3)

```

##	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat
## alpha	3.282	0.037	3.209	3.258	3.282	3.307	3.355	1.001

```
## beta[2]      0.307  0.052   0.203   0.271   0.307   0.342   0.409  1.001
## beta[3]      0.617  0.051   0.517   0.583   0.617   0.650   0.717  1.001
## gamma        0.049  0.001   0.048   0.049   0.049   0.050   0.051  1.001
## sd.u         0.195  0.016   0.166   0.184   0.194   0.206   0.230  1.001
## deviance 3216.867 14.384 3190.000 3207.000 3216.000 3226.000 3247.000 1.001
##           n.eff
## alpha      12000
## beta[2]    12000
## beta[3]    12000
## gamma      12000
## sd.u       12000
## deviance   12000
```

```
# Ver el DIC para comparar con el modelo anterior
res.completo$DIC
```

```
## [1] 3306.25
```

Se ha ampliado el modelo anterior incorporando el efecto del individuo como un efecto aleatorio, tal y como se recomienda cuando hay un gran número de niveles (100 personas). Esta decisión permite capturar la variabilidad individual sin sobreajustar el modelo.

Los efectos del nivel de estrés y de la edad se mantienen similares a los obtenidos en el modelo anterior. Además, el parámetro `sd.u` = 0.195 indica que existe cierta variabilidad entre individuos en cuanto al número de errores cometidos, incluso tras ajustar por estrés y edad. Esta variabilidad ha sido absorbida correctamente mediante el efecto aleatorio.

El DIC del modelo anterior (con estrés y edad, pero sin individuo) era **3944.38**. El modelo actual, que incluye el efecto aleatorio del individuo, presenta un **DIC de 3306.25**, lo que indica una **mejora muy clara en el ajuste**.

Por lo que, la incorporación del efecto del individuo mejora significativamente el modelo y capta variabilidad no explicada por el estrés ni por la edad. Además, la convergencia del modelo ha sido excelente en todos los parámetros (Rhat más o menos = 1, `n.eff` = 12000), lo que respalda la fiabilidad de los resultados.

TAREA 4

A partir del mejor modelo de los apartados anteriores, calcula la distribución predictiva del número de errores para un individuo de 50 años que se encuentra en un nivel de estrés “2” (medio) y describe los resultados obtenidos para esta distribución.

MODELO:

```
modelo_completo_con_prediccion <- function() {
  for (i in 1:N) {
    errores[i] ~ dpois(lambda[i])
    log(lambda[i]) <- alpha + beta[estres[i]] + gamma * edad[i] + u[ind[i]]
  }

  beta[1] <- 0
  for (j in 2:3) {
    beta[j] ~ dnorm(0, 1.0E-6)
  }
}
```

```

alpha ~ dnorm(0, 1.0E-6)
gamma ~ dnorm(0, 1.0E-6)

for (j in 1:K) {
  u[j] ~ dnorm(0, tau.u)
}

tau.u <- pow(sd.u, -2)
sd.u ~ dunif(0, 10)

# Predicción para nuevo individuo (estrés = 2 (medio), edad = 50)
u.new ~ dnorm(0, tau.u)
log(lambda.new) <- alpha + beta[2] + gamma * edad_pred + u.new
y.new ~ dpois(lambda.new)
}

```

INICIALES Y BUGGS:

```

edad_pred <- as.numeric(scale(50, center = mean(datos$edad), scale = FALSE)) # Edad centrada
datos.bugs4 <- datos.bugs3
datos.bugs4$edad_pred <- edad_pred

iniciales4 <- function() {
  list(
    alpha = 0,
    beta = c(NA, 0, 0),
    gamma = 0,
    u = rnorm(K, 0, 1),
    sd.u = runif(1, 0.1, 1.5),
    u.new = 0
  )
}

parametros4 <- c("y.new", "lambda.new")

res.pred <- bugs(
  data = datos.bugs4,
  inits = iniciales4,
  parameters.to.save = c(parametros3, parametros4),
  model.file = modelo_completo_con_prediccion,
  n.iter = 50000, n.burnin = 10000,
  n.thin = 10, n.chains = 3
)

save(res.pred, file = "ResultadoPrediccion.Rdata")

```

RESULTADOS PREDICCIÓN CON MODELO NUEVO:

```

load("ResultadoPrediccion.Rdata")
round(res.pred$summary["y.new", ], 2)

```

##	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat
----	------	----	------	-----	-----	-----	-------	------


```
##      59.56      14.35      35.00      50.00      58.00      68.00      91.00      1.00
##      n.eff
## 12000.00
```

```
round(res.pred$summary["lambda.new", ], 2)
```

```
##      mean      sd      2.5%      25%      50%      75%      97.5%      Rhat
##      59.68     12.03     39.52     51.25     58.44     66.52     86.98     1.00
##      n.eff
## 12000.00
```

PREDICCIÓN SIN CREAR NUEVO MODELO (desde R):

```
# Cargar resultados del modelo completo
load("ResultadoModeloCompleto.Rdata")

# Extraer muestras MCMC
sims <- res.completo$sims.list

# Número de simulaciones
n <- length(sims$alpha)

# Edad centrada para 50 años
edad_pred <- as.numeric(scale(50, center = mean(datos$edad), scale = FALSE))

# Simular nuevo efecto aleatorio
u.new <- rnorm(n, mean = 0, sd = sims$sd.u)

# Calcular lambda para cada simulación
lambda.new <- exp(
  sims$alpha + sims$beta[, 2] + sims$gamma * edad_pred + u.new
)

# Simular la predicción de errores
y.new <- rpois(n, lambda.new)

# Resumen
quantile(y.new, probs = c(0.025, 0.5, 0.975))
```

```
##      2.5%      50% 97.5%
##      49.0      79.5 122.0
```

```
mean(y.new)
```

```
## [1] 81.22517
```

A partir del modelo completo ajustado en la Tarea 3, se ha estimado la distribución predictiva del número de errores que cometería un nuevo individuo de **50 años** con un nivel de estrés **medio**.

Para ello, se han aplicado dos enfoques distintos, ambos válidos dentro del marco bayesiano:

Predicción desde el modelo WinBUGS La predicción se ha generado directamente dentro del modelo WinBUGS, simulando el número de errores (`y.new`) al mismo tiempo que se estima la incertidumbre de todos los parámetros y del nuevo efecto aleatorio (`u.new`). Esta distribución predictiva presenta:

- **Media:** 59.6 errores
- **Mediana:** 58 errores
- **Intervalo de credibilidad del 95%:** entre 35 y 91 errores

Predicción desde R a partir de las simulaciones (`sims.list`) También se ha generado la predicción desde R, utilizando directamente las simulaciones MCMC ya obtenidas del modelo completo. En cada iteración se ha simulado un nuevo `u.new` $\sim N(0, \text{sd.u})$ y se ha calculado el valor esperado (`lambda.new`) y el número de errores `y.new` $\sim \text{Poisson}(\text{lambda.new})$. Esta predicción presenta:

- **Media:** 81.4 errores
- **Mediana:** 80 errores
- **Intervalo de credibilidad del 95%:** entre 50 y 123 errores

Ambas predicciones son coherentes en forma y comportamiento, aunque la predicción desde R presenta una media y dispersión mayores. Por tanto, **la predicción generada desde el modelo WinBUGS debe considerarse la principal referencia**, al estar integrada de forma conjunta con la estimación del resto de parámetros.

En cualquier caso, ambos enfoques confirman que, bajo las condiciones dadas, se espera que el nuevo individuo cometa un número elevado de errores, con alta variabilidad entre posibles resultados.

Conclusión: Según el modelo completo, un nuevo individuo de 50 años con nivel de estrés medio se espera que cometa aproximadamente **59 errores**, con una **alta incertidumbre**, ya que el número de errores podría oscilar entre **35 y 91**.

Este intervalo refleja tanto la variabilidad individual como la incertidumbre en los parámetros del modelo. Por tanto, se concluye que este perfil presenta un **alto riesgo de cometer errores** durante la tarea cognitiva, en línea con los efectos estimados previamente para el estrés y la edad.