

Guía de trabajos prácticos Bioestadística II: Análisis de poder y primera tarea

Derek Corcoran

March 14, 2018

Contents

Obejtivos del práctico	1
Matriz de confusión	1
Calculo de poder en R	2
Tarea	2

Obejtivos del práctico

- Entender cálculos de poder en base a matriz de confusión
- Primera tarea de práctico

Matriz de confusión

La matriz de confusión es una herramienta de toma de desiciones, en el caso especial de la toma de desiciones tenemos la siguiente matriz de confusión

Table 1: Tabla de confusión de errores

	Hipótesis nula cierta	Hipótesis alternativa cierta
Acepto hipótesis nula	No hay error	Error tipo 2
Acepto hipótesis alternativa	Error tipo 1	No hay error

Esto puede ser facilmente ejemplificado con el problema de una alarma de humo, en este caso cuando la alarma suena y no hay fuego y suena la alarma tenemos un error de tipo 1, en cambio si hay fuego y la alarma no suena tenemos un error de tipo 2

	No hay fuego	Hay fuego
No suena alarma	No hay error	Error tipo 2
Suena alarma	Error tipo 1	No hay error

Poder y matriz de confusión

- Probabilidad de que suene la alarma cuando no hay fuego
 - α usualmente 5%
 - una de cada 20 alarmas es falsa
 - ¿Cuál es el α de una alarma de auto?

- Probabilidad de que no suene la alarma cuando hay fuego
 - β si es 10% uno de cada 10 fuegos no es detectado
 - poder es $1 - \beta$ confianza de que fuegos son detectados

Calculo de poder en R

Para hacer calculos de poder en ANOVAS de una y dos vías en R, utilizamos el paquete *pwr2*. En este paquete podemos utilizar la función *pwr.1way* para determinar el poder de un ANOVA de una vía, los argumentos de esta funcion son:

- *K*: El número de grupos a testear
- *n*: Número de individuos por grupo
- *Alpha*: Nivel de significancia
- *Delta*: Valor mínimo a detectar
- *Sigma*: Desviación estandar de la muestra

Para calculos precisos de n necesarios para muestras usar la siguiente app <https://derek-corcoran.shinyapps.io/MinimosCuadrados/>

Tarea

El problema

Una compañía que genera pesticidas descarga parte de sus desechos a un río. La ONG **RioSano**, dice que ha notado una alza en la mortalidad de los patos cortacorriente (*Merganneta armata*) del río.

Ante esto la empresa contrata un científico, el cual hace una estimación de la mortalidad de patos en 10 zonas del río en que descargan sus desechos, y lo compara con otros dos ríos no contaminados. Este científico dice que no hay diferencias significativas en la mortalidad de los patos de los ríos con desechos y sin desechos con una confianza del 95%. Para esto muestra como evidencia la figura 1 y tabla 3 e incluso hace públicos sus datos en el archivo *MuestraPatos.csv*.

Table 3: Tabla de ANOVA de una vía de la mortalidad de patos de los tres ríos

term	df	sumsq	meansq	statistic	p.value
rio	2	301.2531	150.62656	2.359899	0.1136188
Residuals	27	1723.3436	63.82754	NA	NA

La ONG *RioSano* lo contrata para determinar la validez del estudio y si es necesario generar un estudio extra. Ante esto:

1. Genere una matriz de confusión del problema y explique en este contexto que significaría el alfa y beta para este problema, y cual consideraría más relevante.
2. Diseñe el estudio que le gustaría hacer, determinando cuantas áreas debe muestrear por río, estime un delta mínimo que le gustaría determinar y el beta con el que se siente seguro y determine el n mínimo necesario para ese estudio. Justifique su respuesta
3. Dado este n mínimo realice lo siguiente
 - Realice un muestreo de n muestras por tipo de río del archivo *Patos.csv*

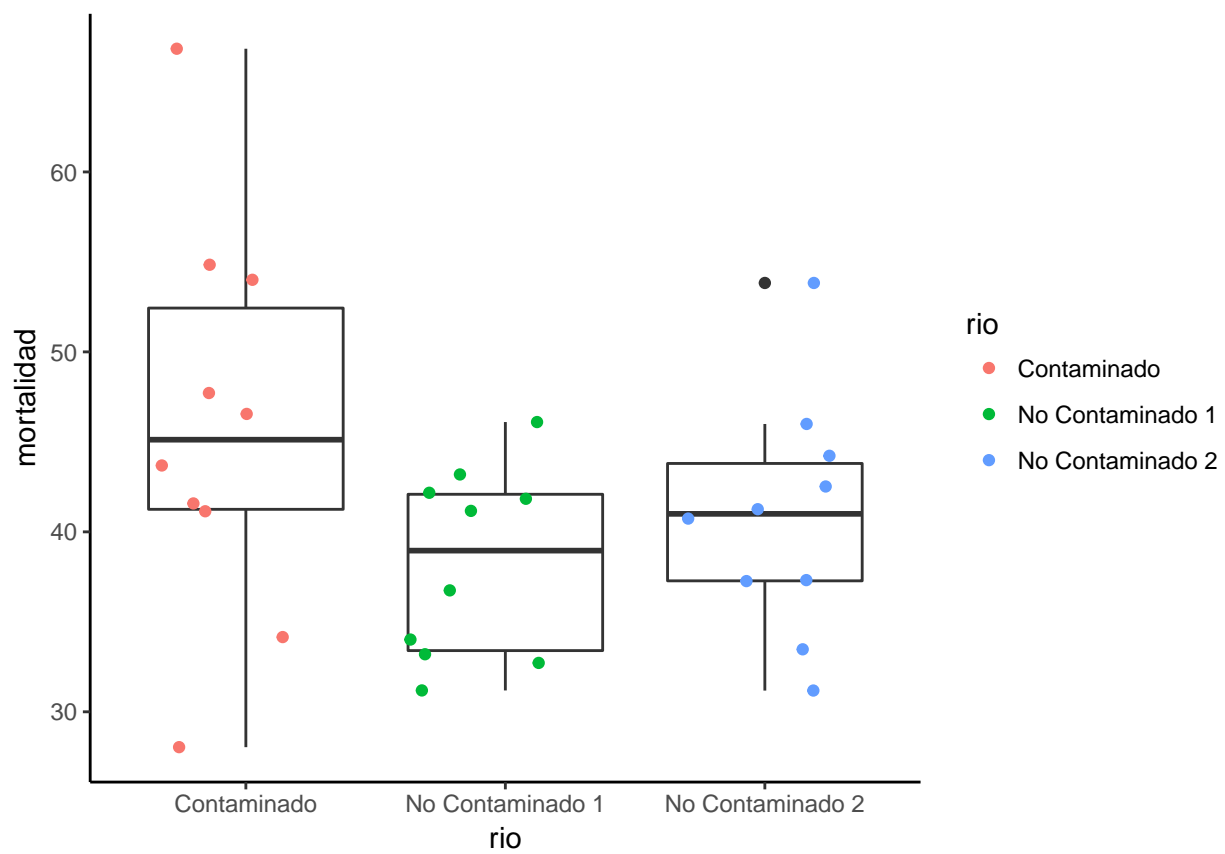


Figure 1: Mortalidades calculadas en 10 zonas de tres ríos

- Genere gráficos y tablas exploratorias de los datos de su muestreo y describalas
 - Revise los supuestos del ANOVA para su base de datos tanto gráficamente como con tests y determine si se puede realizar el anova
 - Diga si según su diseño hay diferencias significativas en la mortalidad de patos entre los ríos
4. Cada zona a muestrear requiere de un monitoreo exhaustivo, que tiene un costo de 500.000 pesos (esto es 1.500.000 de pesos si consideramos los 3 ríos). La ONG *RioSano* consiguió 20.000.000 de pesos para este estudio. Dadas esas limitaciones, genere un balance de α , β y n dada esa limitación para hacer el mejor estudio posible dadas las consecuencias, justifique su respuesta.

Genere un informe para la ONG *RioSano* incorporando estos 5 puntos e incluya una introducción, metodología, resultados, discusión-conclusión y bibliografía, envíe el script de como generó los resultados