

BIOMETRÍA II

CLASE 13

INTEGRACIÓN DE MODELOS



Estructura de los datos

2

Tipo de estudio: experimental u observacional	Determina las conclusiones (causalidad o asociación)
Tipo de VR: continua, discreta, binaria	Define la distribución de probabilidades y por lo tanto el tipo de modelo y el método de estimación
Tipo de VE: continua, discreta, binaria, cuali	Solo cuali: anova Cuanti y cuali: regresión con dummies (ancova) Cuanti: regresión
Relación entre VR y VE	Modelos lineales o no lineales (en los parámetros)
Estructura de agregación (independencia vs bloques, anidamiento, medidas repetidas, parcela dividida)	Modelos con VE de efectos fijos vs modelos marginales /condicionales
Declaración de dependencia entre las observaciones	Modelos marginales o modelos condicionales

Protocolo

3

Parte aleatoria

1. Basándose en el diseño experimental o en el método de muestreo empleado, incluir efectos aleatorios (modelos condicionales) o estructura de la matriz de covarianza (modelos marginales) si existe nivel de agrupamiento entre las observaciones

Parte fija

3. Si se trata de un experimento: incluir en el modelo todos los términos que quedaron definidos por el diseño experimental. El modelo no se debería simplificar (independientemente de la significación de cada término)
4. Si se trata de un estudio observacional: utilizar algún criterio de selección de modelos a fin de identificar las VE significativas o importantes, las restantes se eliminan del modelo a menos que interese incluirlas por razones teóricas

Modelos en R

Fun- ción	Modelo	Estima- ción
lm	Modelo lineal con errores normales y varianza constante; generalmente es usado para regresión con VE cuantitativas	MCO
gls	Modelo lineal con errores normales. Permite modelar heterocedasticidad y distintas estructuras de matriz de covarianzas; no admite VE de efectos aleatorios	MV
lme	Modelo lineal con errores normales y VE de efectos fijos y aleatorios. Permite modelar heterocedasticidad y distintas estructuras de matriz de covarianza	MV
lmer	Idem anterior, pero no modela estructuras de matriz de covarianza	MV

Funciones genéricas

5

Función	Modelo
summary	Proporciona estimaciones de los parámetros del modelo en forma de regresión o anova (summary.lm y summary.aov respectivamente)
plot	Gráficos de diagnóstico de supuestos del modelo
anova	Permite comparar modelos anidados
update	Modifica el último modelo ajustado
coef	Proporciona los estimadores de los parámetros del modelo
fitted	Valores predichos por el modelo lineal
predict	Idem anterior
resid	Diferencia entre el valor observado y el predicho por el modelo
AIC	Compara modelos por AIC

Modelos lineales sin VE de efectos aleatorios

Modelo	lineal general	lineal general	lineal generalizado
Método de estimación	cuadrados mínimos	máxima verosimilitud restringida	máxima verosimilitud
distribución de la VR	normal	normal	Familia exponencial
heterocedasticidad	sensible	modelable	modelable
desbalanceo	sensible	robusto	robusto
supuestos	Independencia, normalidad, homocedasticidad, linealidad	Independencia, normalidad, linealidad	Independencia, linealidad, dispersión acorde a la distrib de la variable

Modelos lineales con VE de efectos aleatorios

Modelo	lineal general mixto	lineal general mixto	lineal general marginal
Método de estimación	cuadrados mínimos	máxima verosimilitud restringida	máxima verosimilitud
distribución de la VR	normal	normal	normal
Estructura de agregación	modelable pero arduo	Modelable mediante la inclusión de VE de efectos aleatorios	Modelable mediante la declaración de matriz de covarianza
Componentes de varianza	sí	sí	no
supuestos	normalidad, homocedasticidad, linealidad	Normalidad, linealidad; heterocedasticidad modelable	Normalidad, linealidad; heterocedasticidad modelable

Algunas pautas para una correcta experimentación

8

En el diseño:

- Aleatorización de los tratamientos
- Replicación en la escala adecuada. Atención a las seudorréplicas
- Determinar **a priori** el tamaño de muestra para una dada potencia y un efecto de tratamiento

Algunas pautas para un correcto análisis estadístico

9

- ❑ Recurrir a herramientas gráficas para tener idea del efecto de los tratamientos, variabilidad y posibles outliers
- ❑ Identificar tipo de VR, tipo de VE, si existen estructuras de agrupamiento de los datos
- ❑ Si se trata de un modelo con algún nivel de agregación de las observaciones, modelar primero la estructura aleatoria de los datos
- ❑ Centrar las VE si se desea darle sentido a la ordenada al origen
- ❑ Verificar los supuestos del modelo; patrones en los residuos pueden indicar la necesidad de incluir potencias o interacciones
- ❑ Evitar efectuar múltiples test estadísticos separados. Ajuste del nivel de significación
- ❑ Respetar el principio de marginalidad: siempre analizar primero la interacción de mayor orden; no excluir términos involucrados en interacciones
- ❑ No comparar p de distintas pruebas, ya que a menos que los GL sean iguales no son comparables

Algunas pautas para un correcto análisis estadístico

10

En las conclusiones:

- Alcance del estudio: aplicar la inferencia estadística sólo a la población de la cual se extrajo la muestra
- Diferenciar estudios experimentales de observacionales; atención con la causalidad
- Significación biológica vs estadística. Magnitud del efecto (diferencia de medias, pendiente), informar intervalos de confianza
- Presentar los resultados gráficamente utilizando las estimaciones del modelo

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3-14.

Presentación de resultados

Materiales & Métodos

11

- Diseño experimental
- Pruebas estadísticas
- Nivel de significación
- Software

Los resultados para (*variable respuesta*) fueron analizados mediante un modelo lineal general en un diseño (). Las variables explicatorias fueron (). Los supuestos () se estudiaron mediante (). El criterio utilizado para la selección de modelos fue ().

El nivel de significación empleado fue () / Se consideraron significativas aquellas pruebas con $p < ()$

Todos los análisis estadísticos fueron efectuados utilizando el programa estadístico (*R, R Core Team 2021*)

Plantas en cojín en la Puna

Análisis 1



12

- Las plantas en cojín son una de las formas de vida mejor adaptadas a las extremas condiciones de los ambientes de alta montaña
- Al proporcionar micrositios más adecuados para la adquisición de recursos, actuarían como nodrizas para el resto de las especies de la comunidad
- Se desea estudiar las características de los cojines de la especie *Laretia acaulis* que favorecen el establecimiento de plántulas de otras especies, a dos altitudes en la puna jujeña: baja (~2000 m.s.n.m) y alta(~ 3000 m.s.n.m)
- A cada altitud de la puna jujeña se seleccionaron al azar 65 cojines separados al menos 1 km. En cada uno se determinó la biomasa de otras especies vegetales (en gramos) y características de los cojines: diámetro (cm), altura máxima, temperatura del sustrato, potencial hídrico del suelo, distancia al cojín más cercano
- VR? VE? Tipo? Modelo? (asumiendo efectos aditivos) n? Base de datos

Modelando



13

1. Especule qué valores de VIF esperaría encontrar para las 5 VE cuantitativas
2. Si detecta colinealidad, explique qué consecuencias traería incluir a todas las variables en el modelo
3. Si detecta colinealidad, ¿se solucionaría centrando las variables?
4. ¿Cómo decidiría qué variables excluir?
5. Suponga que se retuvieron diámetro y altitud y que se sospecha que el efecto nodriza es mayor a mayor altitud. Plantee el modelo

Coefficients:

	Estimate	Std. Error	Pr(> t)	
(Intercept)	160.183	33.734	2.10e-04	***
diámetro	10.261	1.0835	2.23e-05	***
altitud.alta	-20.123	3.1471	2.18e-03	***
diámetro. altitud.alta	3.123	0.814	5.22e-03	***

6. ¿Interpretación de los coeficientes?
7. ¿Qué cambiaría si se centrara diámetro?
8. ¿Conclusiones biológicas?
9. ¿Y si se detecta que la variabilidad en la biomasa es mayor a menor altitud?

Análisis 2



15

- Los investigadores desean analizar las modificaciones microclimáticas que induce *L. acaulis* con respecto a los espacios abiertos.
 - Particularmente desean estudiar si existen cambios en el potencial hídrico matricial del suelo (PHMS).
 - Para ello, en cada uno de los cojines se midió el PHMS en el suelo en el cojín y en el suelo aledaño
-
1. Escriba el modelo condicional
 2. ¿Qué componentes de varianza pueden estimarse?
 3. ¿Qué diferencias tiene este modelo con el marginal?
 4. Si se midiese el PHMS a distintas distancias del cojín, ¿qué cambiaría?
 5. Si en cada sitio se tomasen observaciones a las 8, 12, 16 y 20 hs, ¿cómo se modificaría el modelo?