

refreg: un paquete de R para estimar rexións de referencia

Óscar Lado Baleato (oscarlado.baleato@usc.es)

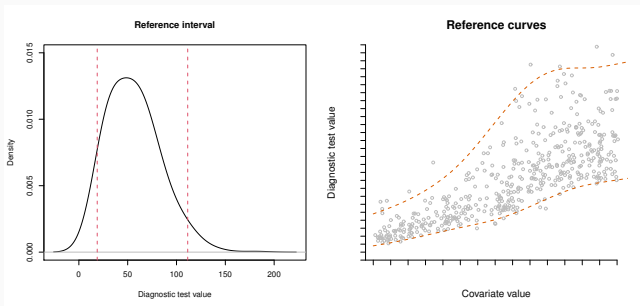
En colaboración con: Javier Roca Pardiñas, Carmen Cadarso Suárez e Francisco Gude Sampedro.

October 20, 2022

Introducción

Probas diagnósticas continuas

A maioría de decisións clínicas baseanse nos resultados de marcadores continuos. Estas probas requiren dun intervalo de referencia para a súa interpretación.



Intervalo de referencia

Un intervalo que contén o 95% dos resultados dos pacientes sans.

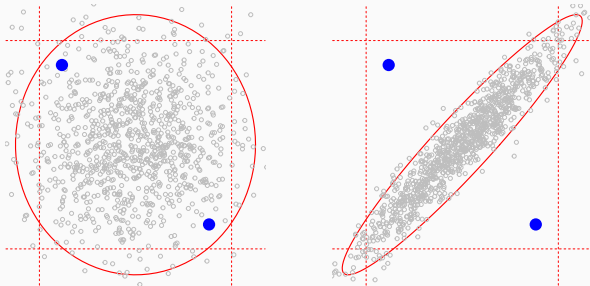
Curva de referencia

O mesmo pero condicionado a covariables.

Resultados de múltiples pruebas diagnósticas continuas

Para enfermedades diagnosticadas con máis dun marcador clínico → usanse varios intervalos de referencia univariantes.

Ignorar a correlación entre as probas **Menor especificidade e Maior sensibilidade.**



Rexión de referencia

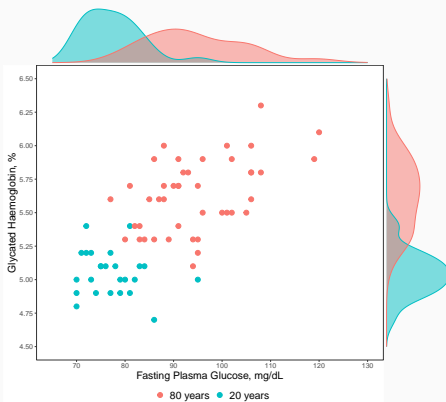
Unha rexión que contén a maioría dos valores conxuntos dos marcadores medidos na poboación san.

Rexións de referencia multivariadas condicionais

A distribución conxunta dos marcadores pode cambiar coas características dos pacientes.

Estudo AEGIS

Proxecto de investigación clínica para estudar a variabilidade dos marcadores glucémicos e a súa correlación na poboación xeral.



Modelo matemático

Modelo matemático

Para p covariables $\mathbf{X} = (X_1, \dots, X_p)$, e unha resposta bivalente continua $\mathbf{Y} = (Y_1, Y_2)$

Asumimos o seguinte modelo:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \mu_1(\mathbf{X}) \\ \mu_2(\mathbf{X}) \end{pmatrix} + \Sigma^{1/2}(\mathbf{X}) \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

$\Sigma^{1/2}(\mathbf{X}) \rightarrow$ sendo a descomposición de Cholesky de

$$\Sigma(\mathbf{X}) = \begin{pmatrix} \sigma_1^2(\mathbf{X}) & \sigma_{12}(\mathbf{X}) \\ \sigma_{12}(\mathbf{X}) & \sigma_2^2(\mathbf{X}) \end{pmatrix}$$

Residuos bivariantes $(\varepsilon_1, \varepsilon_2)$ que asumimos

- Independentes das covariables
- Media cero, varianza un e incorrelados.
- Cunha función de densidade coñecida $f(\varepsilon_1, \varepsilon_2)$

$$\mu_r(\mathbf{X}) = \alpha_r + \sum_{j=1}^p f_{jr}(X_j) \quad \sigma_r^2(\mathbf{X}) = \exp \left(\beta_r + \sum_{j=1}^p g_{jr}(X_j) \right) \quad \text{para } r = 1, 2$$

$$\rho(\mathbf{X}) = \tanh \left(\gamma + \sum_{j=1}^p m_j(X_j) \right)$$

- Estimación dos efectos suaves mediante splines penalizados (mgcv::gam).
- Para parámetros con restricciones (varianza, correlación) os modelos aditivos axustanse usando un algoritmo Newton-Rhapson modificado.
- Mediante remostraxe dos residuos podemos estimar intervalos de confianza dos efectos.

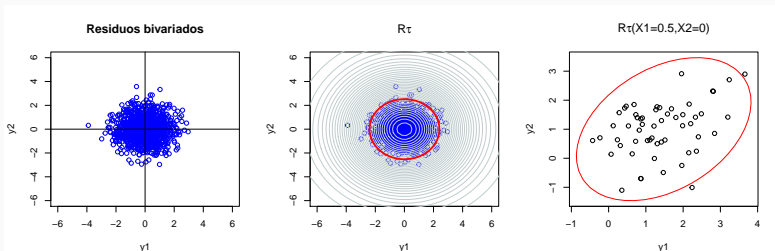
Rexión de referencia (I)

Para un valor das covariables \mathbf{X} a rexión de referencia condicional para (Y_1, Y_2) ven dada por:

$$R_\tau(\mathbf{X}) = \begin{pmatrix} \mu_1(\mathbf{X}) \\ \mu_2(\mathbf{X}) \end{pmatrix} + \Sigma^{1/2}(\mathbf{X})R_\tau \quad \text{for } \tau \in [0, 1]$$

onde $R_\tau = \{(u, v) \in \mathbb{R}^2 | f(u, v) \leq k\}$ sendo k o $1 - \tau$ cuantil de $f(\varepsilon_1, \varepsilon_2)$, tal que

$$P\left(\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \in R_\tau\right) = \tau$$

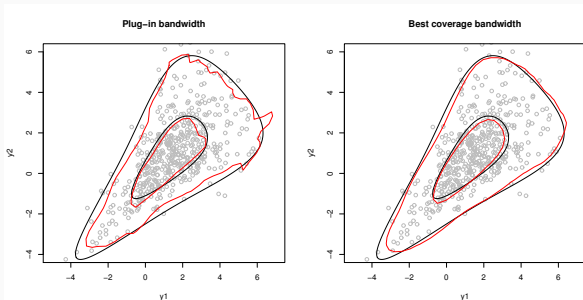


Rexión de referencia (II): estimación

Dada a mostra $\{(\hat{\varepsilon}_{i1}, \hat{\varepsilon}_{i2})\}_{i=1}^n$, o estimador da densidade f nun punto (u, v) ven dado por:

$$\hat{f}(u, v) = \frac{1}{n} \sum_{i=1}^n K_H \begin{pmatrix} u - \hat{\varepsilon}_{i1} \\ v - \hat{\varepsilon}_{i2} \end{pmatrix}$$

A ventá do estimador tipo núcleo **H** ten gran influencia na cobertura da rexión.



Uso do pacote refreg

Estrutura do paquete

O uso de dúas funcións permite estimar as rexións de referencia condicionais.

Function	Description
bivRegr	This function estimates the covariates effects on the means vector, and variance covariance matrix from a bivariate variable. Non linear effects might be estimated for continuous covariates using penalized spline smoothers.
bivRegion	This functions estimate a probabilistic/reference region for bivariate data. It is based on a kernel density estimation. It may be applied to a set of bivariate data points, or to a bivRegr object. In the former case, the function will estimate a bivariate reference region for the model standarized residuals.

As funcións *S3 summary*, *summary_boot*, *predict*, e *plot* poden aplicarse os obxectos *bivRegr* e *bivRegion*.

Estimación do modelo de localización-escala bivariado

R code

`dm_no = subset(aegis, aegis$dm == "no")` → pacientes sans

`mu1 = fpg ~ s(age)` → HbA1c e FPG dependentes da idade

`mu2 = hba1c ~ s(age)`

`var1 = ~ s(age)`

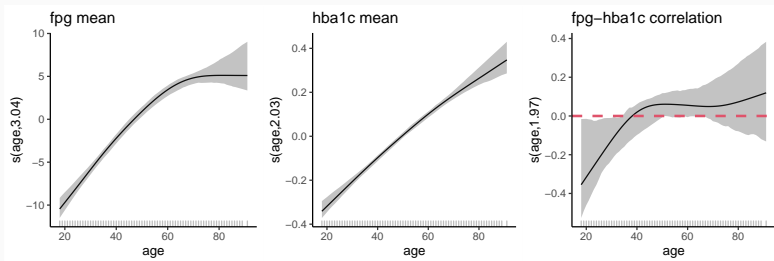
`var2 = ~ s(age)`

`rho = ~ s(age)`

`f = list(mu1,mu2,var1,var2,rho)`

`fit = bivRegr(formula,data=dm_no)` → Axuste

`s0 = summary_boot(fit,B=250,parallel=T)` → Efectos estimados



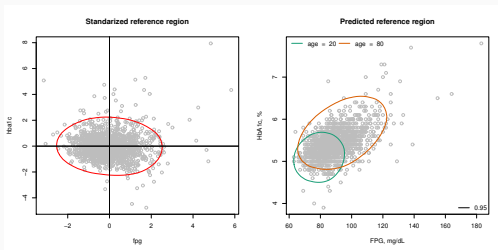
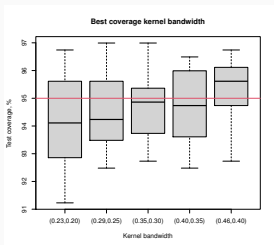
Estimación da rexión de referencia bivariada (I)

Código de R

```
region = bivRegion(fit, H_choice = "Hcov", tau=0.95)
```

```
plot(region, col = "grey", reg.lwd = 2, main = "Standarized reference region")
```

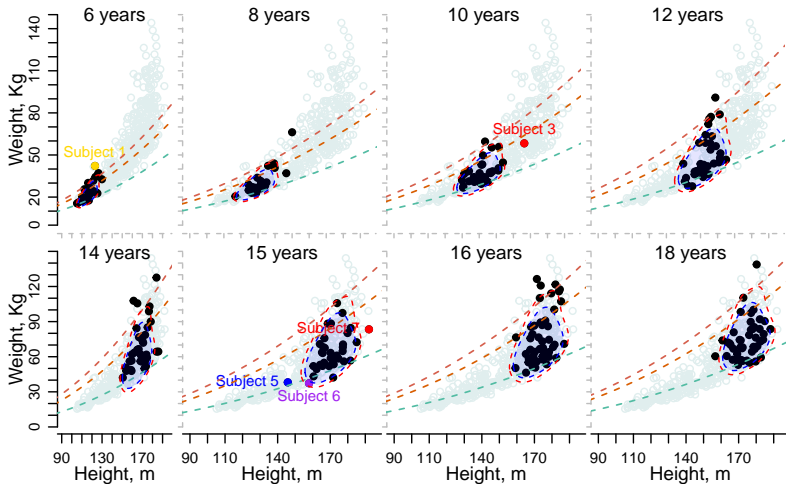
```
plot(region, cond=T, newdata = data.frame(age = c(20, 80)), reg.lwd = 2, col =  
"grey", legend = T, , xlab = "FPG, mg/dL", ylab = "HbA1c, %", main = "Predicted  
reference region")
```



Outras aplicacións

Pediatría: interpretación conjunta (talla,peso)

Males' bivariate growth chart



Research Square

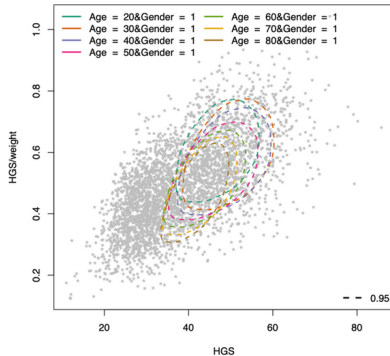
Search preprints

Short Report

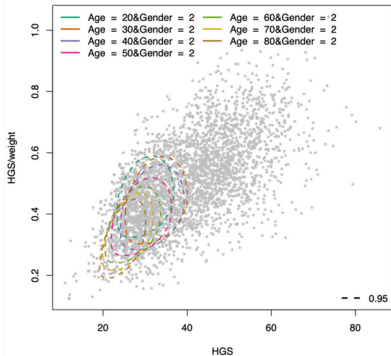
Handgrip strength conditional tolerance regions suggest the need for personalized sarcopenia definition: An analysis of the American NHANES database

Marcos Matabuena, Pedro Pugliesi Abdalla, Dalmo Roberto Lopes Machado, and 2 more

(A) - Men

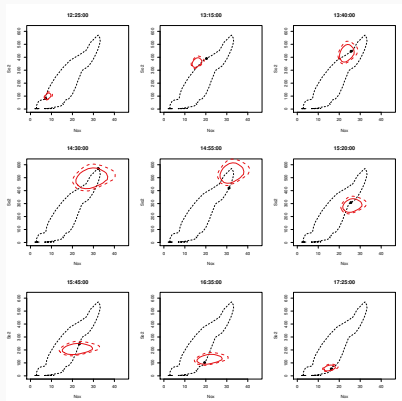


(B) - Women



Medio ambiente: predicción conjunta de contaminantes

Este paquete pode aplicarse en campos diversos como o medio ambiente.
Monitorización conxunta dos niveis de SO_2 e NO_x .



Roca-Pardiñas, J., Ordóñez, C., and Lado-Baleato, O. (2021). Nonparametric location–scale model for the joint forecasting of SO_2 and NO_x pollution episodes. *Stochastic Environmental Research and Risk Assessment*, 35(2), 231-244.

Futuras Melloras

- Inclusión de contrastes de especificación para efectos de covariables sobre a rexión.
- Medidas de axuste do modelo como BIC, función de perda bivariada.
- Métodos de selección automática de variables.
- Paralelización de procesos computacionalmente pesados. (posible computación GPU?)

Referencias

- Lado-Baleato, Ó., Roca-Pardiñas, J., Cadarso-Suárez, C., & Gude, F. (2021). Modeling conditional reference regions: Application to glycemic markers. *Statistics in Medicine*, 40(26), 5926-5946.
- Roca-Pardiñas, J., Ordonez, C., & Lado-Baleato, O. (2021). Non-parametric location–scale model for the joint forecasting of SO₂ and NO_x pollution episodes. *Stochastic Environmental Research and Risk Assessment*, 35(2), 231-244.

Gracias

Óscar Lado-Baleato está actualmente contratado por un contrato da plataforma de soporte a investigación clínica financiado por o instituto de Saúde Carlos III (ISCIII) PT20/00043. Este estudio foi financiado por ISCIII PI20/01069, ISCIII RD21/0016/0022, cofinanciado pola UE. Dentro do proxecto MTM2017-83513-R. E tamén con financiación da Xunta de Galicia: ED431C 2020/20 e IN607A/2021-2.