Webapp en Shiny para modelos de regresión
con polinomios fraccionarios de orden 1

Antonio Torres Ruiz
Isabel María Ortiz Rodríguez

Universidad de Almería

IX JORNADAS DE USUARIOS DE R
Granada, Noviembre de 2017

Índice

Índice

Polinomios fraccionarios

Un polinomio fraccionario de orden m se denota por PF_m y viene definido como:

$$\phi_m(x;\beta,p) = \beta_0 + \sum_{j=1}^{m} \beta_j H_j(x),$$

con $H_j(x)$:

$$\begin{cases} H_1(x) = x^{(p_1)} & \text{para } j = 1 \\ H_j(x) = \begin{cases} x^{(p_j)} & \text{si } p_j \neq p_{j-1} \\ H_{j-1}(x)\ln x & \text{si } p_j = p_{j-1} \end{cases} & \text{para } j = 2, \ldots, m \end{cases}$$

y

$$x^{(p)} = \begin{cases} x^p & p \neq 0 \\ \ln x & p = 0 \end{cases}$$

Royston y Altman (1994): tomar la potencia p en el conjunto

$$\mathcal{P} = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$$

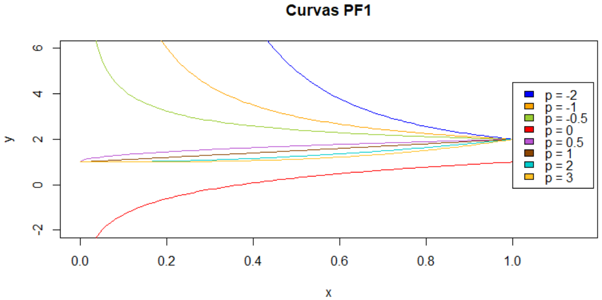
Transformación cuando $x < 0$:

$$\phi_m(x; \beta, p) = \beta_0 + \sum_{j=1}^{m} \beta_j H_j(x - c)$$

Características del polinomio fraccionario de orden 1

$$\mathrm{PF}_1(p) = \beta_0 + \beta_1 x^{(p)} = \begin{cases} \beta_0 + \beta_1 x^p & p \in \mathcal{P}, \quad p \neq 0 \\ \beta_0 + \beta_1 \ln x & p = 0 \end{cases}$$

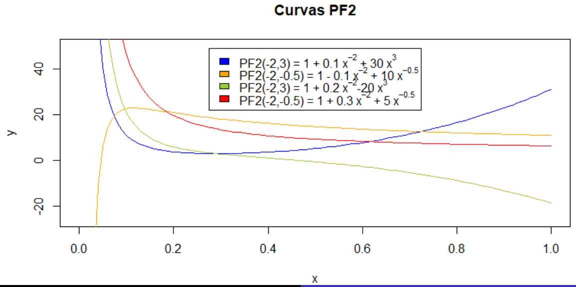
Representación para $\beta_0 = \beta_1 = 1$:



Características del polinomio fraccionario de orden 2

$$PF_2(p_1, p_2) = \begin{cases} \beta_0 + \beta_1 x^{(p_1)} + \beta_2 x^{(p_2)} & p_1, p_2 \in \mathcal{P}, \quad p_1 < p_2 \\ \beta_0 + \beta_1 x^{(p)} + \beta_2 x^{(p)} \ln x & p_1, p_2 \in \mathcal{P}, \quad p_1 = p_2 = p \end{cases}$$

Comportamiento gráfico para varios tipos de curvas:



Índice

Modelos de regresión con PF_1

Modelo de regresión con polinomios fraccionarios de orden 1:

$$Y = \beta_0 + \beta_1 X^{(p)} + \epsilon, \qquad X^{(p)} = \begin{cases} X^p & p \neq 0 \\ \ln X & p = 0 \end{cases},$$

con $p \in \mathcal{P} = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$.

Para una muestra de tamaño n:

$$Y_i = \beta_0 + \beta_1 X_i^{(p)} + \epsilon_i, \qquad i = 1, ..., n.$$

Hipótesis básicas sobre los errores aleatorios ϵ_i

$$Y_i = \beta_0 + \beta_1 X_i^{(p)} + \epsilon_i, \qquad i = 1, ..., n.$$

- $\mathrm{E}\left(\epsilon_i\right) = 0, \ i = 1, ..., n.$
- $\mathrm{Var}\left(\epsilon_i\right) = \mathrm{E}\left(\epsilon_i^2\right) = \sigma^2 = \text{cte}, \ i = 1, ..., n.$
- $\mathrm{Cov}\left(\epsilon_i, \epsilon_j\right) = \mathrm{E}\left(\epsilon_i \epsilon_j\right) = 0$, para todo $i \neq j$.
- Hipótesis adicional: $\epsilon_i \sim \mathcal{N}(0, \sigma^2), \ i = 1, ..., n.$

Residuos del modelo

- Muestra de tamaño n: $(X_i, Y_i), \quad i = 1, ..., n$.
- Valores observados: $Y_i, \quad i = 1, ..., n$.
- Modelo estimado: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X^{(p)}$.
- Valor estimado para X_i: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i^{(p)}$.
- Residuos: $e_i = Y_i - \hat{Y}_i, \quad i = 1, ..., n$.

A partir de ahora,

$$X^{(p)} = Z$$

Estimación de los parámetros del modelo

- Estimación de los parámetros β_0 y β_1:

$$\left\{ \begin{array}{l} \hat{\beta}_0 = \bar{Y} - \frac{S_{yz}}{S_{zz}^2}\bar{Z} \\ \hat{\beta}_1 = \frac{S_{yz}}{S_{zz}^2} \end{array} \right. \quad \text{con} \quad \left\{ \begin{array}{l} S_{yz} = \frac{\sum_{i=1}^n Y_i Z_i}{n} - \bar{Y}\bar{Z} \\ S_{zz}^2 = \frac{\sum_{i=1}^n Z_i^2}{n} - \bar{Z}^2 \end{array} \right.$$

- Estimador insesgado para σ^2:

$$\hat{\sigma}^2 = S_R^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

Intervalos de confianza $100(1-\alpha)\,\%$

- Para β_0:

$$\left[\hat{\beta}_0 - t_{n-2,1-\alpha/2}\sqrt{\frac{S_R^2}{n}\left(1+\frac{\bar{Z}^2}{S_{zz}^2}\right)}, \hat{\beta}_0 + t_{n-2,1-\alpha/2}\sqrt{\frac{S_R^2}{n}\left(1+\frac{\bar{Z}^2}{S_{zz}^2}\right)}\right]$$

- Para β_1:

$$\left[\hat{\beta}_1 - t_{n-2,1-\alpha/2}\sqrt{\frac{S_R^2}{nS_{zz}^2}}, \hat{\beta}_1 + t_{n-2,1-\alpha/2}\sqrt{\frac{S_R^2}{nS_{zz}^2}}\right]$$

- Para σ^2:

$$\left[\frac{(n-2)S_R^2}{\chi^2_{n-2,1-\alpha/2}}, \frac{(n-2)S_R^2}{\chi^2_{n-2,\alpha/2}}\right]$$

Contrastes de hipótesis

- Para β_0 y β_1:

Parámetro	β_0	β_1
Contraste	$\begin{cases} H_0 : \beta_0 = \beta_{00} \\ H_1 : \beta_0 \neq \beta_{00} \end{cases}$	$\begin{cases} H_0 : \beta_1 = \beta_{10} \\ H_1 : \beta_1 \neq \beta_{10} \end{cases}$
Estadístico	$T = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{\frac{S_R^2}{n}\left(1+\frac{\bar{z}^2}{S_{zz}^2}\right)}} \sim t_{n-2}$	$T = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\frac{S_R^2}{nS_{zz}^2}}} \sim t_{n-2}$
Región de rechazo	$\lvert T \rvert > t_{n-2,1-\alpha/2}$	$\lvert T \rvert > t_{n-2,1-\alpha/2}$

- Contraste $H_0 : \beta_1 = 0$. Tabla ANOVA:

Fuente de variación	SC	gl	MC	Estadístico
Regresión	SCR	1	MCR	F
Error	SCE	$n-2$	MCE	
Total	SCT	$n-1$		

Rechazar H_0 si $F > \mathcal{F}_{1,n-2,1-\alpha}$.

Intervalos de predicción

- Para la respuesta media en $X = X_0$ $(Z_0 = X_0^{(p)})$:

$$\left[\hat{Y}_0 - t_{n-2,1-\alpha/2}\sqrt{S_R^2\left(\frac{1}{n}+\frac{(Z_0-\bar{Z})^2}{nS_{zz}^2}\right)}, \hat{Y}_0 + t_{n-2,1-\alpha/2}\sqrt{S_R^2\left(\frac{1}{n}+\frac{(Z_0-\bar{Z})^2}{nS_{zz}^2}\right)}\right]$$

- Para la predicción en $X = X_0$ $(Z_0 = X_0^{(p)})$:

$$\left[\hat{Y}_0 - t_{n-2,1-\alpha/2}\sqrt{S_R^2\left(1+\frac{1}{n}+\frac{(Z_0-\bar{Z})^2}{nS_{zz}^2}\right)}, \hat{Y}_0 + t_{n-2,1-\alpha/2}\sqrt{S_R^2\left(1+\frac{1}{n}+\frac{(Z_0-\bar{Z})^2}{nS_{zz}^2}\right)}\right]$$

- Bandas de confianza y de predicción: unir los extremos de los intervalos anteriores calculados para cada valor de X.

Coeficiente de determinación R^2

$$R^2 = \frac{\text{SCR}}{\text{SCT}} = 1 - \frac{\text{SCE}}{\text{SCT}}$$

- $0 \leq R^2 \leq 1$.
- Cuanto mayor sea R^2 mejor será el modelo estimado.

Comprobación de las hipótesis del modelo

Homocedasticidad y linealidad

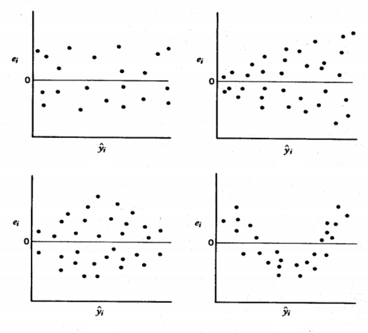
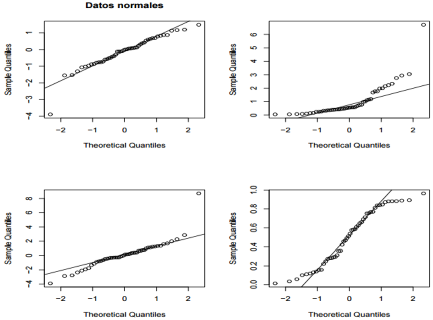


Gráfico qq-plot



Índice

Elección del mejor modelo PF_1

Medidas para comparar los modelos:

Medida	Expresión	Criterio
SCE	SCE	$\searrow$
MCE	$\frac{\text{SCE}}{n-2}$	$\searrow$
R^2	$1-\frac{\text{SCE}}{\text{SCT}}$	$\nearrow$
$\bar{R}^2$	$1-\left(1-R^2\right)\frac{n-1}{n-2}$	$\nearrow$
AIC	$n\ln\left(\frac{\text{SCE}}{n}\right)+n\ln(2\pi)+n+6$	$\searrow$

Índice

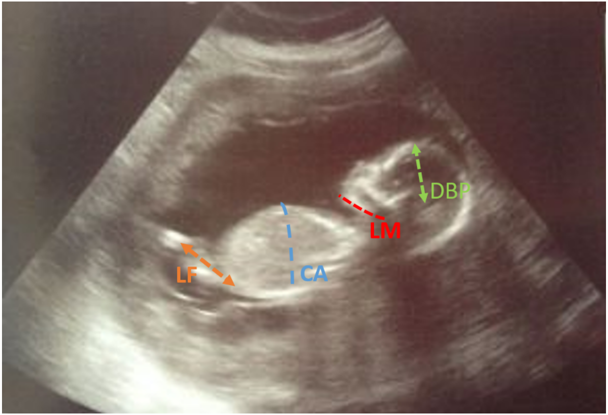
La aplicación desarrollada

- Enlace de la aplicación:

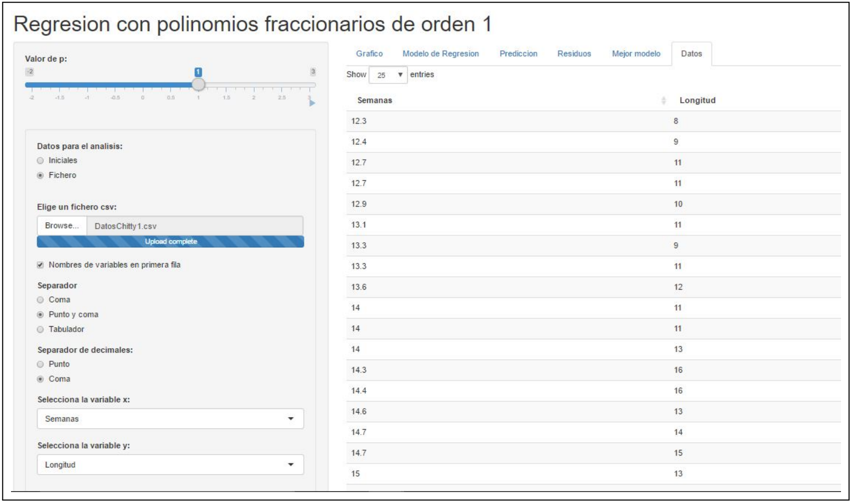
  `https://antoniotorres.shinyapps.io/shiny/`

Índice

Relación entre edad gestacional (X, semanas) y longitud de la mandíbula (Y, mm)



Introducción de los datos (Chitty et al. 1994):



Elección del mejor modelo:

Grafico | Modelo de Regresion | Prediccion | Residuos | Mejor modelo | Datos

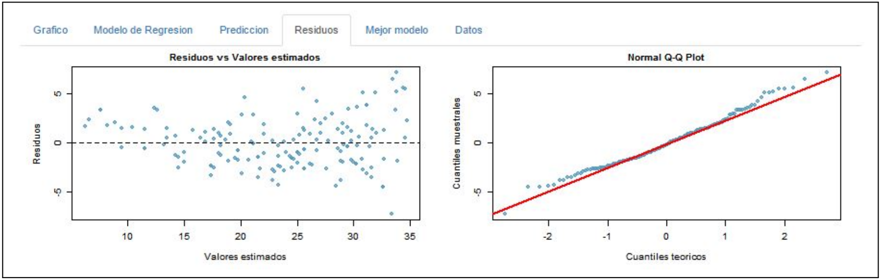
Medidas para comparar los modelos

p	S.C.E.	M.C.E.	R2	R2.Aj.	AIC	Comparar.AIC
-2.00	1489.58	9.55	0.83	0.83	808.88	-83.48
-1.00	1105.39	7.09	0.87	0.87	761.75	-36.35
-0.50	977.91	6.27	0.89	0.89	742.39	-16.98
0.00	897.65	5.75	0.90	0.90	728.86	-3.45
0.50	865.00	5.54	0.90	0.90	723.01	2.40
1.00	878.24	5.63	0.90	0.90	725.41	0.00
2.00	1027.21	6.58	0.88	0.88	750.16	-24.76
3.00	1304.30	8.36	0.85	0.85	787.89	-62.49

Mejor modelo

p	S.C.E.	M.C.E.	R2	R2.Aj.	AIC	Comparar.AIC
0.50	865.00	5.54	0.90	0.90	723.01	2.40

Comprobación de las hipótesis básicas:

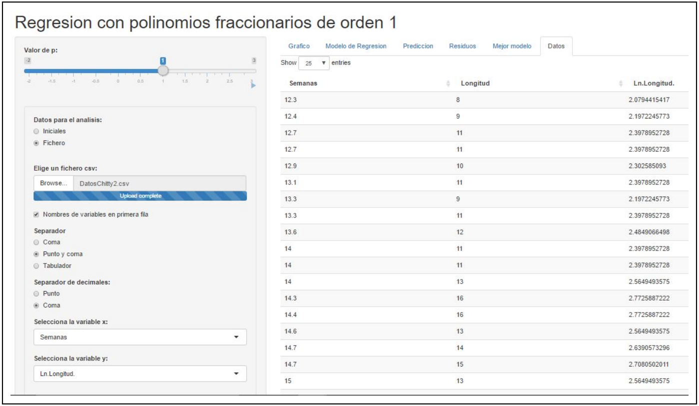


El modelo considerado no es el adecuado.

$$\boxed{Y = \ln(\text{Longitud})}$$

Relación entre edad gestacional (X) y $Y = \ln(\text{Longitud})$

Introducción de los datos:



Semanas	Longitud	Ln.Longitud.
12.3	8	2.0794415417
12.4	9	2.1972245773
12.7	11	2.3978952728
12.7	11	2.3978952728
12.9	10	2.302585093
13.1	11	2.3978952728
13.3	9	2.1972245773
13.3	11	2.3978952728
13.6	12	2.4849066498
14	11	2.3978952728
14	11	2.3978952728
14	13	2.5649493575
14.3	16	2.7725887222
14.4	16	2.7725887222
14.6	13	2.5649493575
14.7	14	2.6390573296
14.7	15	2.7080502011
15	13	2.5649493575

Elección del mejor modelo:

Grafico | Modelo de Regresion | Prediccion | Residuos | Mejor modelo | Datos

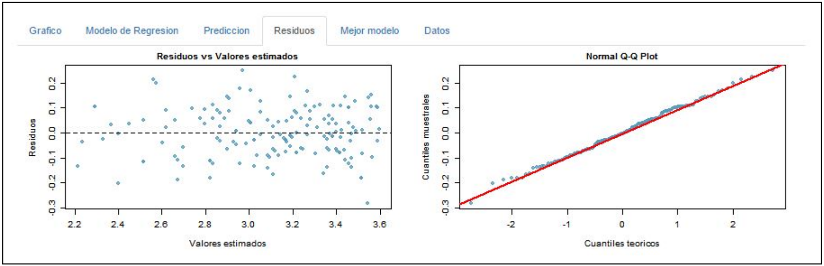
Medidas para comparar los modelos

p	S.C.E.	M.C.E.	R2	R2.Aj.	AIC	Comparar.AIC
-2.00	1.62	0.01	0.91	0.91	-269.19	60.19
-1.00	1.45	0.01	0.92	0.92	-286.33	77.33
-0.50	1.53	0.01	0.92	0.92	-277.94	68.94
0.00	1.72	0.01	0.91	0.91	-259.92	50.92
0.50	2.00	0.01	0.89	0.89	-235.80	26.80
1.00	2.37	0.02	0.87	0.87	-209.00	0.00
2.00	3.31	0.02	0.83	0.82	-156.17	-52.83
3.00	4.42	0.03	0.77	0.77	-110.55	-98.45

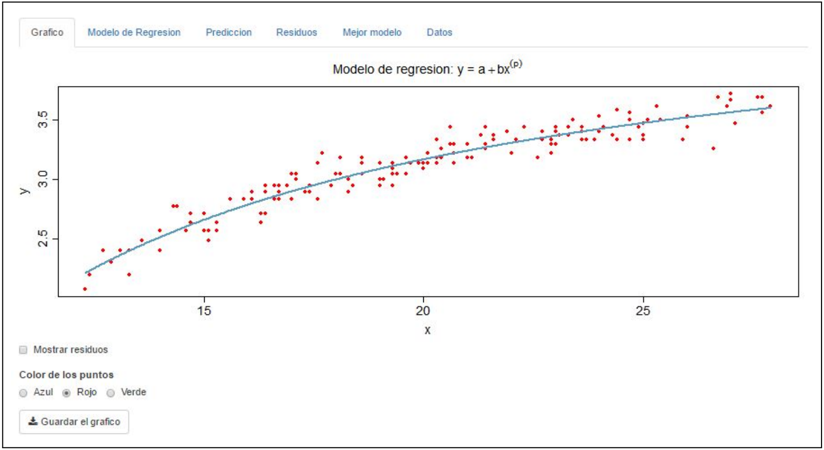
Mejor modelo

p	S.C.E.	M.C.E.	R2	R2.Aj.	AIC	Comparar.AIC
-1.00	1.45	0.01	0.92	0.92	-286.33	77.33

Comprobación de las hipótesis básicas:



Representación gráfica de las observaciones y el modelo estimado:



Estimación de parámetros, I.C. y contrastes:

Grafico | Modelo de Regresion | Prediccion | Residuos | Mejor modelo | Datos

Potencia del modelo

Potencia	Valor
p	-1

Modelo

Parametro	Estimacion	Error.est	I.C.inf	I.C.sup	t	P.valor
a	4.69	0.04	4.61	-31.79	125.87	0.00
b	-30.41	0.70	4.76	-29.02	-43.33	0.00

Guardar resultados

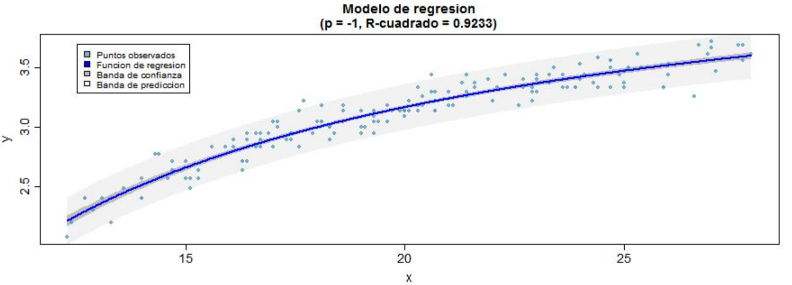
Tabla ANOVA del modelo

Fuente.variabiliad	S.C.	g.l.	M.C.	F.exp	P.valor
Modelo	17.50	1	17.50	1877.47	0
Residual	1.45	156	0.01		

Modelo estimado: $\widehat{\ln}(\text{Longitud}) = 4.69 - \frac{30.41}{\text{Semanas}}$

Bandas de confianza y predicción:

Grafico | Modelo de Regresion | Prediccion | Residuos | Mejor modelo | Datos



Prediccion

Nivel de confianza:

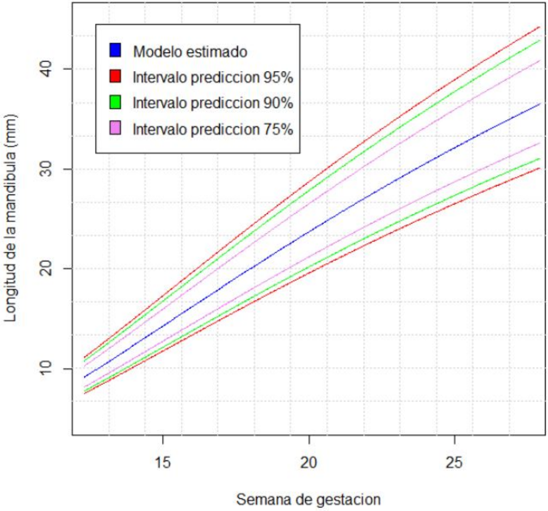
- 0.9
- 0.95
- 0.99

Introduce el valor de x:

10

x	prediccion	Limite.inferior	Limite.superior
10	1.65	1.44	1.85

Modelo con datos originales:



Índice

El paquete `mfp` (Sauerbrei et al. 2006)

Contraste:

$$\begin{cases} H_0 : \text{Modelo restringido} \\ H_1 : \text{Modelo completo} \end{cases}$$

mediante test de razón de verosimilitud:

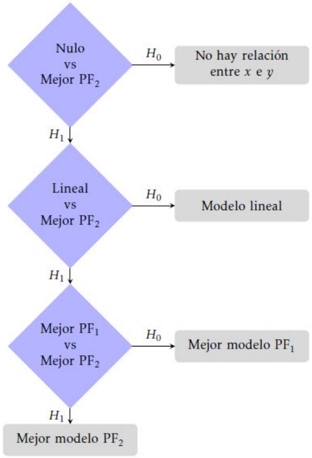
$$\mathrm{RV} = -2 \ln \frac{\hat{L}_0}{\hat{L}_1} \sim \chi^2_r$$

Ejemplo:

$$\begin{cases} H_0 : \text{Modelo lineal } (p = 1) \\ H_1 : \text{Modelo no lineal } (p \neq 1) \end{cases} \Leftrightarrow \begin{cases} H_0 : \mathrm{E}(Y) = \beta_0 + \beta_1 X \\ H_1 : \mathrm{E}(Y) = \beta_0 + \beta_1 X^{(p)}, \;\; p \neq 1 \end{cases}$$

$$\Rightarrow \mathrm{RV} \sim \chi^2_{3-2} = \chi^2_1$$

Esquema del proceso iterativo:



- Relación entre X e Y:

$$\begin{cases} H_0 : \text{Modelo nulo} \\ H_1 : \text{Mejor PF}_2 \end{cases}$$

- Linealidad:

$$\begin{cases} H_0 : \text{Modelo lineal} \\ H_1 : \text{Mejor PF}_2 \end{cases}$$

- Simplificación:

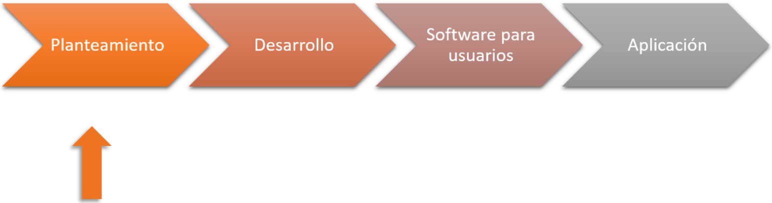
$$\begin{cases} H_0 : \text{Mejor PF}_1 \\ H_1 : \text{Mejor PF}_2 \end{cases}$$

Aplicación del algoritmo `mfp` al ejemplo

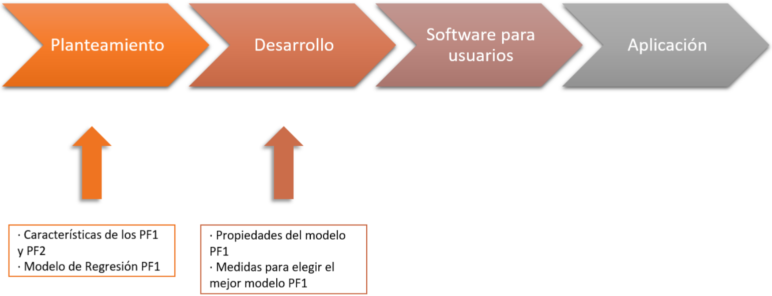
```
Deviance table:
                 Resid. Dev
Null model         18.95706
Linear model       2.372511
Final model        1.454316

Fractional polynomials:
  df.initial select alpha df.final power1 power2
x          4      1  0.05        2     -1      .


Transformations of covariates:
       formula
x I((x/10)^-1)

Rescaled coefficients:
Intercept        x.1
4.686        -30.407
```

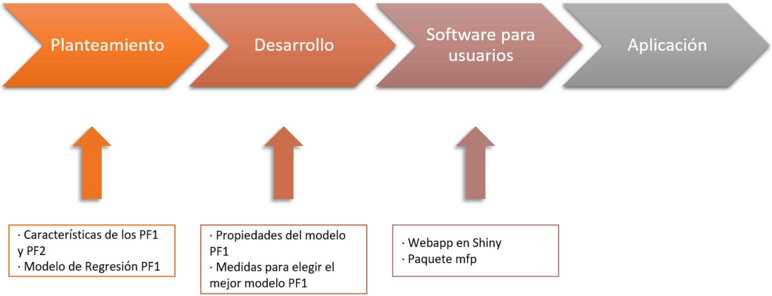
Índice

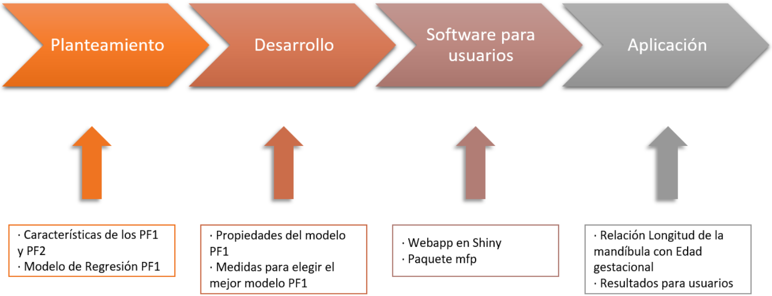

Planteamiento
Desarrollo
Software para usuarios
Aplicación



· Características de los PF1 y PF2
· Modelo de Regresión PF1


Planteamiento
Desarrollo
Software para usuarios
Aplicación
· Características de los PF1 y PF2
· Modelo de Regresión PF1
· Propiedades del modelo PF1
· Medidas para elegir el mejor modelo PF1


Planteamiento
Desarrollo
Software para usuarios
Aplicación
· Características de los PF1 y PF2
· Modelo de Regresión PF1
· Propiedades del modelo PF1
· Medidas para elegir el mejor modelo PF1
· Webapp en Shiny
· Paquete mfp


Planteamiento
Desarrollo
Software para usuarios
Aplicación
· Características de los PF1 y PF2
· Modelo de Regresión PF1
· Propiedades del modelo PF1
· Medidas para elegir el mejor modelo PF1
· Webapp en Shiny
· Paquete mfp
· Relación Longitud de la mandíbula con Edad gestacional
· Resultados para usuarios

Bibliografía I

L.C. Acosta
Ajuste de polinomios fraccionarios a curvas de crecimiento
Disponible en: *Enlace al texto completo*. [Consulta 13 nov. 2017]

J. Aparicio, M.A. Martínez, J. Morales
Modelos lineales aplicados en R
Disponible en: *Enlace al texto completo*. [Consulta 13 nov. 2017]

J.E. Callejas
Métodos de aproximación polinómica en el diseño experimental
Disponible en: *Enlace al texto completo*. [Consulta 13 nov. 2017]

L.S. Chitty, D.G. Altman, A. Henderson, S. Campbell
Charts of fetal size: 3, Abdominal measurements
Br. J. Obstet. Gynaecol., 101 (1994), 125–131

D. C. Montgomery, E. A. Peck, G. G. Vining
Introduction to linear regression analysis
Wiley, 1992

Bibliografía II

P. Royston, D.G. Altman
Regression using fractional polynomials of continuous covariates
J. Royal Stat. Soc., C 43 (1994), 429–467

W. Sauerbrei, C. Meier-Hirmer, A. Benner, P. Royston
Multivariable regression model building by using fractional polynomials.
Comput. Stat. Data An., 50 (2006), 3464–3485

Shiny: A web application framework for R
Disponible en: *Enlace web*. [Consulta 13 nov. 2017]

J. Zamora
Una metodología para manejar variables continuas en los modelos de pronóstico: Polinomios fraccionales
Disponible en: *Enlace al texto completo*. [Consulta 13 nov. 2017]

¡MUCHAS GRACIAS!