|  |  |
| --- | --- |
| **1 º PRUEBA PIA** | **FECHA: 10/12/2024** |
| **NOMBRE: Diego J. López Hernández** | **NOTA:** |

**Nota 1 : Cada problema puntúa 2 puntos.**

**Nota2: El examen puntúa sobre 10. Elija 5 de los 6 problemas**

**Nota 3: La documentación de R y weka se depositara en el correo** [**rapariciog@campusdigitalfp.com**](mailto:rapariciog@campusdigitalfp.com)**. Los ejercicios en formato word o pdf**

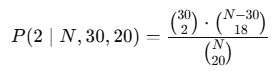
*1)*

*a) Se dispone de 40 piezas electrónicas y algunas de ellas no funcionan correctamente, aunque no se sabe cuántas hay defectuosas.*

*Para estimar el número de defectuosas se extrae al azar y sin reposición 5 de ellas y se observan 1 defectuosa. Estímese el número piezas defectuosas*

*b) Se desea estimar la población de trucha en el Pantano de la Tranquera . Para ello se capturan una muestra de 30 ejemplares y se marcan . Posteriormente se arrojan al pantano y tras 24 horas se extrae una muestra de 20 ejemplares y se observan 2 ejemplares marcados. Estímese la población de trucha*

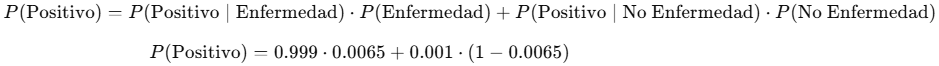
1. En el ejercicio al hacer los cálculos sale una estimación de 8 piezas defectuosas (40/5\*1=8).
2. En este ejercicio al estimar la población de truchas en la tranquera lo mas probable son 170 truchas.



*2) La prevalencia de algún tipo de enfermedad en España es de 65 personas cada 10000 habitantes . Si se posee la enfermedad una prueba diagnóstica la detecta con probabilidad del 99.9% y si no se posee la prueba la diagnostica por error en el 0.1% de los casos.*

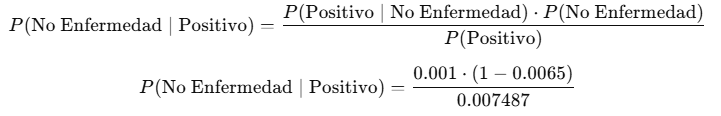
*a) Probabilidad que se diagnostique la enfermedad*

Al hacer el calculo de la probabilidad total sale que un 0,7487% de las veces se va a dar el caso de ser positivo en una enfermedad

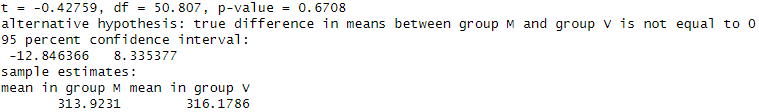


*b) Sabiendo que el diagnóstico es positivo. Calcula la probabilidad de que el diagnóstico sea erróneo.*

En este caso sale que un 13,26% de veces será erroneo el diagnostico si es



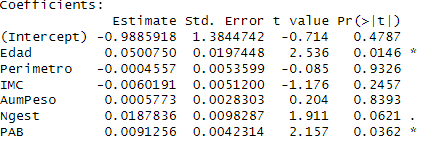
1. *Estúdiese las posibles diferencias entre PAB según el sexo del bebe . En caso de existencia de diferencias . Cuantifíquelas.Archivo diabetes.*

**

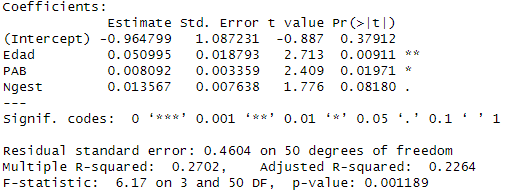
El p-value, al estar por encima de 0.05 quiere decir que no tiene relevancia el sexo del bebe con el perimetro abdominal.

1. *Efectúe un modelo de regresión lineal múltiple para estimar el número de PRN a través de las variables que crea oportunas. Validez del modelo. Interprete los parámetros de posición.Archivo diabetes.*

Aquí haremos una regresión en la cual no utilizaremos los valores que no sean numéricos como sexo o tipo de parto.

**

Aquí podemos comprobar que datos van a sernos útiles a la hora de estimar el PRN que en este caso sale Edad, Pab como útiles (podrían tener hasta 3 estrellas) y el ngest como algo que puede llegar a ser útil pero no mucho porque solo tiene un punto.



Aquí ya con estas variables utilizadas se puede comprobar que el p-value < 0.05 con lo cual estos valores son relevantes a la hora de estimar el PRN. Al hacerlas logarítmicas y doble logarítmicas se comprueba que el nguest pierde su valor como variable y cada vez se distancian mas del p-value.



Esta es la logarítmica en el que se puede ver un R-squared mejor y mas alejado del p-value que antes.

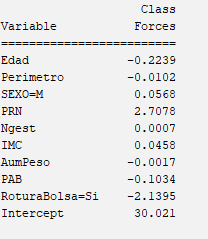


Y esta la doble logarítmica.

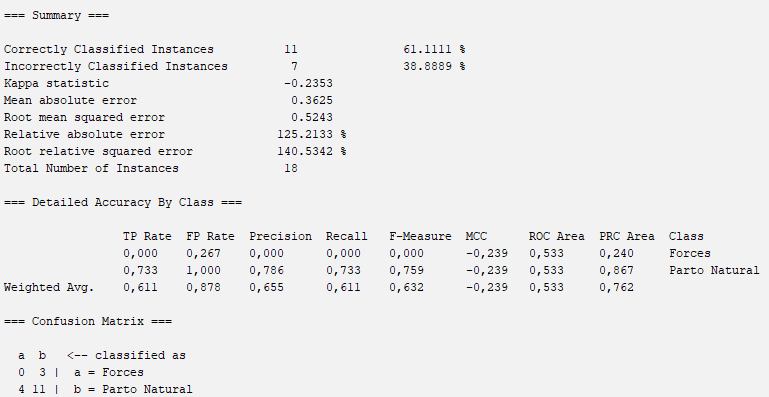
1. *Estímese el atributo Tipo .Archivo diabetes*

*Comentar la regresión utilizada y razones. Comentar resultados*

Este ejercicio lo e realizado en weka y lo primero a sido poner una regresión logística y ver que factores eran útiles en el proceso para estimar el tipo.

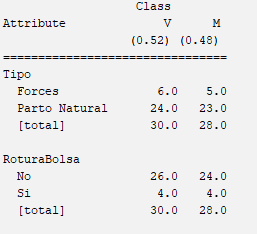


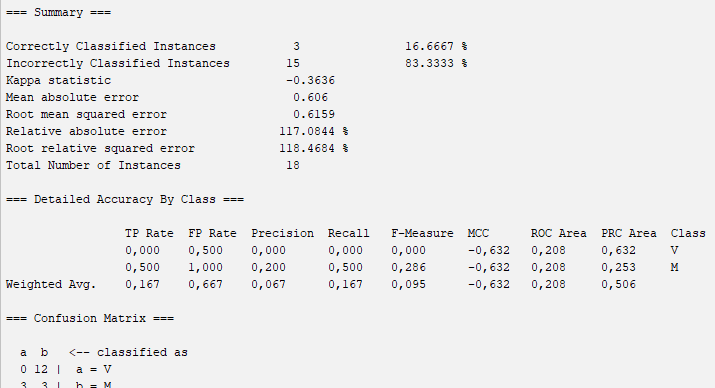
Aquí he podido comprobar los atributos mas útiles para el proceso y he ido haciendo pruebas hasta que he dado con que usando edad, PRN, PAB, roturabolsa se puede conseguir un Area ROC de 0,53. Obviamente no es útil esto, pero es lo máximo que he conseguido con esta función de regresión logística.



Aquí se puede ver como solo tiene una efectividad del 61% y que realiza mejor el parto natural que el de forces. El área de roc es de 0,533 y los verdaderos postivos de parto natural es un 73%.

1. *Aplicar Bayesiano Ingenuo en weka para predecir el sexo del bebe según las variables cualitativas del archivo diabetes. Interpretar Matriz de confusión, Indicador ROC así como las tasas TP y FP. Que clase predice mejor?*



Al usar solo las cualitativas que no tienen casi influencia en el SEXO se puede comprobar lo mal que sale este Bayesiano Ingenuo.

Awui se puede comprobar que esta muy mal falla mas que acierta. El área ROC es de 0,2 muy mala y los verdaderos positivos solo aciertan la mitad en mujeres mientas que los falsos positivos en mujeres son todos y en hombres la mitad.