Aprendizaje Automático - Trabajo Práctico 3

Gonzalo Castiglione - 49138

May 12, 2012

Objetivo: Aplicar diversos métodos estadísticos para aprender a hacer inferencia a partir de datos experiemtales.

1 Métodos de estadística paramétrica

1. Soluciones

   (a) medidas:

   (b)

	Largo Cépalo	Ancho Cépalo	Largo Pétalo	Ancho Pétalo
$\hat{\mu}$	6.5880	2.9740	5.5520	2.0260
$\hat{\sigma}$	0.6232	0.3160	0.5409	0.2692
E_{cm}	0.0081	0.0021	0.0061	0.0015

   (c) Intervalos de confianza para con un nivel de confianza de 0.95.

	Ancho	Largo	Ancho	Largo
I	2.8643	1.9657	3.0837	2.0863

2. Se tienen 80 componentes, de las cuales 12 son defectuosas. Por ser este experimiento una secuencia de ensayos Bernoulli, repetidos n veces, se lo puede considerar una distribución binomial.

   (a) La proporción de componentes no defetuosos de la muestra $= \bar{x}_{nd}\frac{80-12}{80} = 0.85$

      i. Un estimador $\hat{x}$ es un estimador insesgado para estimar a x si $E[\hat{x}] = p$.
      Sea $S_n = x_1 + x_2 + ... + x_n$. En donde cada x_i representa1 si el componente no esta defectuoso o 0 en caso contrario.
      $E[\bar{x}] = E((\sum x_i)/n) = \frac{1}{n}\sum E(x_i) = \frac{1}{n}np = p$.
      Por lo tanto este es un estimador *insesgado*.

      ii. Muestra: 68 mediciónes con $\{x_i, y_i\} = 1$ y 12 mediciónes con $\{x_i, y_i\} = 0$.
      $e_0 = (0 - 0.85)$ (para las 12 muestras defectuosas)
      $e_1 = (1 - 0.85)$ (para las 68 muestras no defectuosas)
      Por lo que el error cuádratico medio estaría dado por la fórmula:
      $E_{CM} = \sqrt{\frac{(1-0.85)^2*68+(0-0.85)^2*12}{80}} = \sqrt{\frac{1.53+8.67}{80}} \simeq 0.35$

   (b) Proporción de sistemas que funcionan correctamente $= \frac{\binom{80-12}{2}}{\binom{80}{2}} = \frac{2278}{3160} = 0.72$

3. código:

```
birds=[
...replace meditions here...
];
alpha=0.01
[h,p,ci,stats] = ttest(birds(:,2), birds(:,3), alpha)
```

Si alpha $>$ p-value $= 0.0117$

Entonces se rechaza H_0 (o sea que uD=0)

Si no

Entonces no se rechaza, es decir las plumas podrían ser iguales.

Estan en lo cierto, para alpha $= 0.05$ hay variación entre el color de las plumas

4. Solución

   (a) .

      i. Código de matlab utilizado

```
load brains.txt;
x = 1:28;
y = brains(:,1);
z = brains(:,2);
clf;
hold on;
plot(x, y, '*b;Peso Promedio en Kg;')
plot(x, z, '*r;Peso Cerebro Promedio en G;')
print('-dpng', './TotalWeightVsBrainWeight.png')
```

      (Ver figura 1 y 2 - Anexo)
      En una observación a simple vista, se puede ver que las mediciónes que se diferencian notablemente del resto son: $6, 7, 14, 15, 16$ y por supuesto, la 25.

      ii. No

   (b) Código de matlab utilizado

```
load brains.txt
regstats(brains(:,1), brains(:,2),'línear')
```

   (Ver figura 3 - Anexo)

   (c) La recta que mejor ajusta con los ejes x e y con la función log aplicados a ambos es: $y = a * x + b$. Con $a = 0.496$ y $b = 2.55$. Por lo tanto, para obtener la recta que ajusta a los puntos x e y sin aplicar las transformación, es aplicar la tranformación inversa al log, es decir $pow(10, n)$. Quedando así la curva que aproxima a los puntos como: $10^y = a * 10^x + b$. Despejando por y, $y = log(a * 10^x + b)$. En la figura 4 del Anexo se puede ver cuan bien se aproximan los valores.

   (d) En los gráficos del punto b, se puede observar que las mediciones que estan muy fuera del común de son no solo las$14, 15$ y 25 sinó que también la 6 y la 16. Por lo que fueron removidos de la tabla de valores. El gráfico obtenido luego de ajustados los valores se muestra en la fiegura 5 del Anexo.

2 Anexo

Gráficos de cada ejercicio:

1)

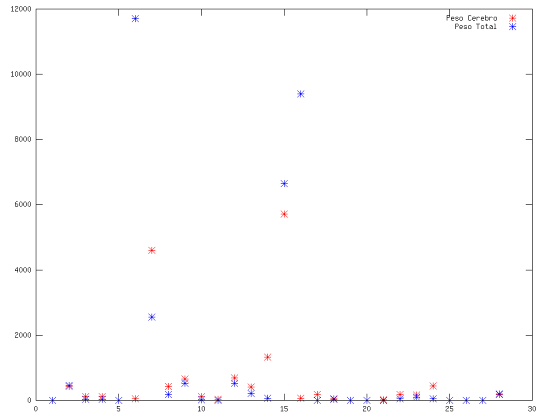

Figure 1: Peso del cerebro y peso total para cada medición en brains.txt*

* El valor del peso del cerebro de la medición 25 no se ve en la figura ya que se aleja demasiado del resto de los valores y el ajustar los ejes solo para mostrar ese valor produce que todas las demas mediciónes no puedan apreciarse correctamente.

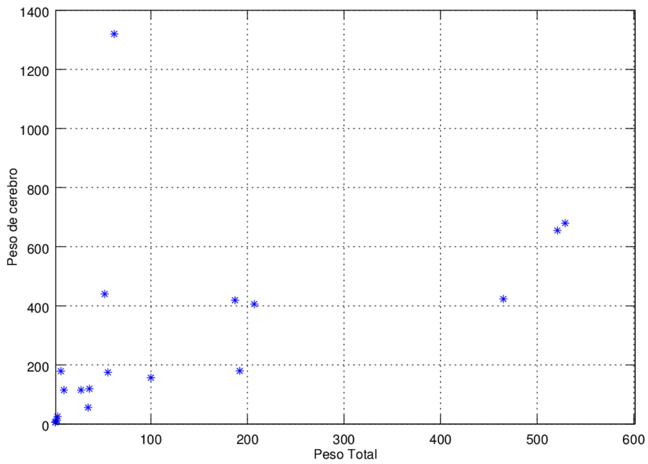

Figure 2: Peso total en Kg Vs peso del cerebro en G**

** Se removieron los valores para los 4 valores de x mayores a 2000 ya que ocultaban la visualización de todos los demás valores

4)

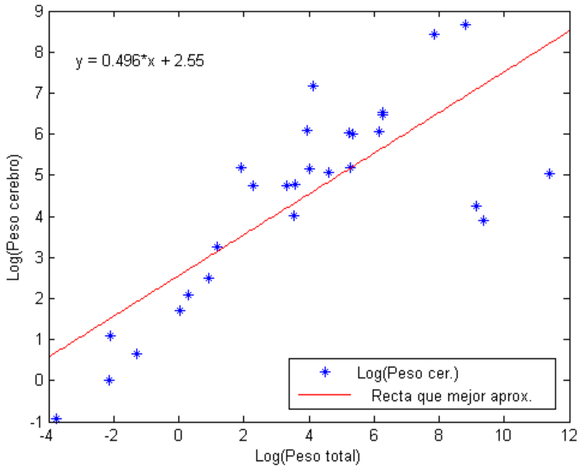

Figure 3: Logaritmo de ambas mediciónes y la línea que mejor los aproxima

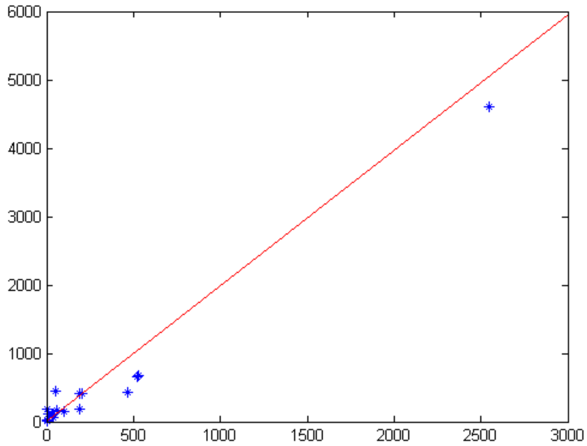

Figure 4: Logaritmo de ambas mediciónes y la línea que mejor los aproxima