Dpto. Matemática Aplicada y Estadística

Grado en IIAA y Grado en IHJ
Asignatura: **Estadística Aplicada. Curso 2011-2012**
Examen de prácticas de Febrero 2012

NOMBRE:..**APELLIDOS:**..
ESPECIALIDAD:..

La resistencia y lo fibroso del espárrago son de los principales determinantes de su calidad. Éste fue el enfoque de un estudio que aparecía en una revista científica. El artículo contenía los siguientes datos sobre x=fuerza cortante (en kg.) e Y=porcentaje del peso seco en fibra. En el fichero **esparrago.txt** se encuentran los datos correspondientes a las citadas variables. Una vez recuperado el fichero, responder a las siguientes cuestiones:

1. Calcular los siguientes descriptivos numéricos para ambas variables:

   a) Media muestral
   b) Rango intercuartílico
   c) Desviación típica muestral
   d) Coeficiente de variación
   e) Rango muestral

   ¿Cuál de las dos características presenta menor dispersión relativa? Razonar la respuesta.

2. Realizar un diagrama de caja y bigotes para cada una de las características e identificar cada una de las líneas que lo constituyen, así como los valores numéricos correspondientes. (*Copiar en el reverso de esta página los gráficos obtenidos*)

3. ¿Existe algún valor atípico? (si es así indicar cuál o cuáles). ¿Cuáles serían los valores admisibles entre los que se encontrarían los datos no atípicos para cada uno de los dos conjuntos de datos?

4. Realizar un histograma para la variable Y. Comentar las características más relevantes de este gráficos. (*Copiar el gráfico en el reverso de esta página*). ¿Qué medidas de centralización y dispersión consideras más adecuadas para resumir este conjunto de datos?

5. Admitiendo que la variable aleatoria Y=porcentaje del peso seco en fibra sigue una distribución normal de media 2.74 y de desviación típica 0.37, calcular las siguientes probabilidades:

   (a) $P(Y > 2.6)$
   (b) $P(2.5 < Y \leq 3.2)$

6. Proporcionar un intervalo de confianza al 97% para la media de la variable porcentaje del peso seco en fibra. Indicar la distribución de probabilidad que has utilizado para construir dicho intervalo.

7. ¿Podemos asumir que la media de la variable porcentaje del peso seco en fibra es iguaL a 2.7? Indicar el procedimiento utilizado y dar la respuesta a partir del $p - valor$ obtenido.

8. Se quiere determinar un modelo para explicar el porcentaje de peso seco de fibra en función de la fuerza cortante. ¿Qué tipo de modelo parece adecuado? Proponer un modelo concreto a partir de los datos y comentar la bondad del ajuste.

9. ¿Cuál sería el porcentaje de peso seco de fibra para un especimen que presenta una fuerza cortante de 99 kg? ¿Y para uno que presente una fuerza cortante de 250 kg.? ¿Son fiables estas estimaciones? Razonar la respuesta.

1/T

EXAMEN de PRÁCTICAS de FEBRERO 2012

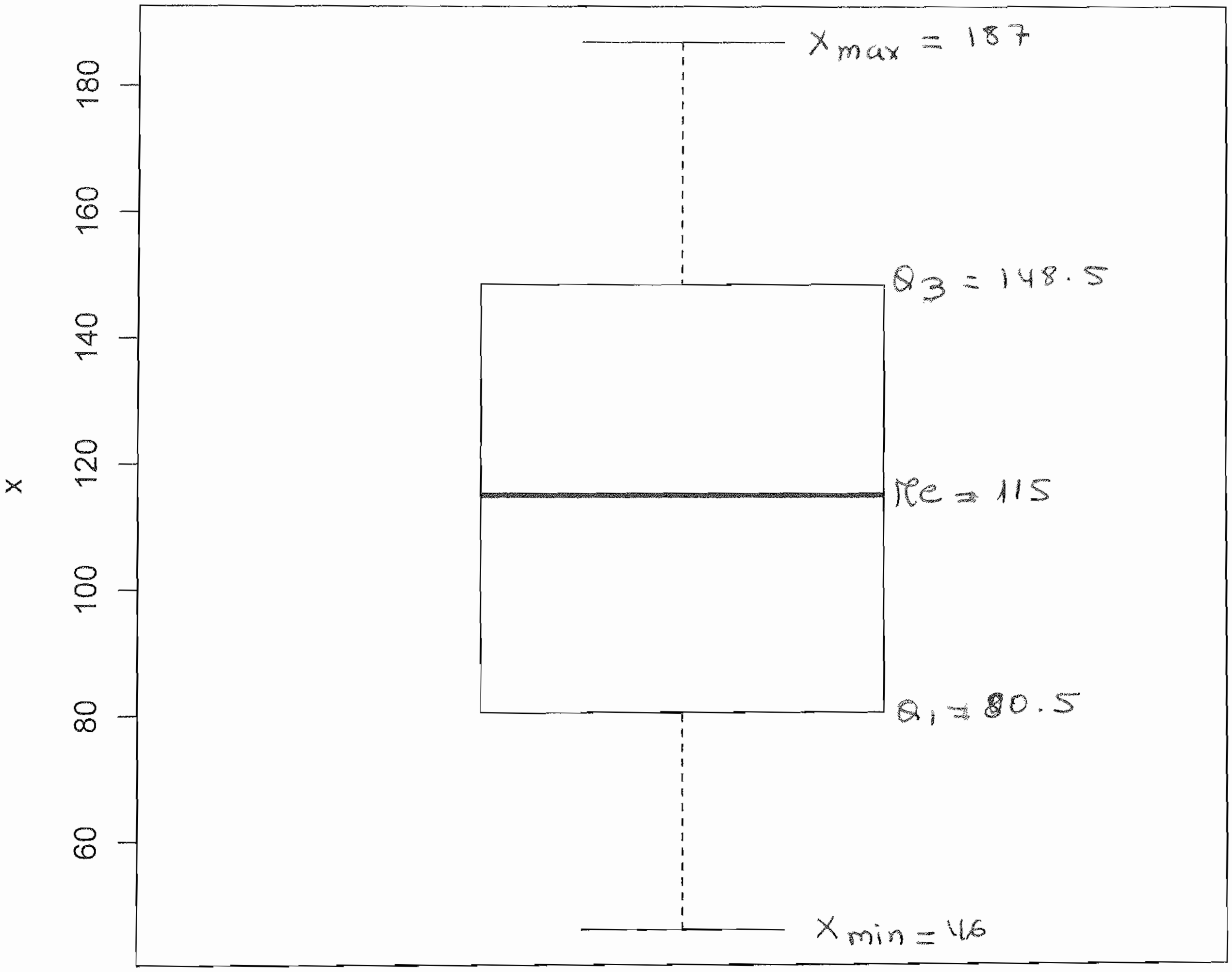
1. Descriptivos numéricos de x e y:

```
> numSummary(esparrago[,c("x", "y")], statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
        mean        sd    0%   25%    50%   75%   100%  n
x 116.454545 42.2697765 46.0 80.50 115.00 148.5 187.00 55
y   2.743273  0.3744134  2.1  2.49   2.66   3.0   3.85 55
```

- $\bar{x} = 116.45$; $S_x = 42.27$
- $\bar{y} = 2.74$; $S_y = 0.37$
- $\left\{ \begin{array}{l} RIQ_x = 148.5 - 80.5 = 68 \\ RIQ_y = 3 - 2.49 = 0.51 \end{array} \right.$
- $\left\{ \begin{array}{l} CV_x = \frac{S_x}{\bar{x}} = \frac{42.27}{116.45} = 0.36 \\ CV_y = \frac{S_y}{\bar{y}} = \frac{0.37}{2.74} = 0.14 \end{array} \right.$ Presenta menor dispersión relativa la característica Y.
- $\left\{ \begin{array}{l} R_x = X_{max} - X_{min} = 187 - 46 = 141 \\ R_y = Y_{max} - Y_{min} = 3.85 - 2.1 = 1.75 \end{array} \right.$

2º

DIAGRAMA de CAJA de la CARACTERÍSTICA X



3º Para la característica X:

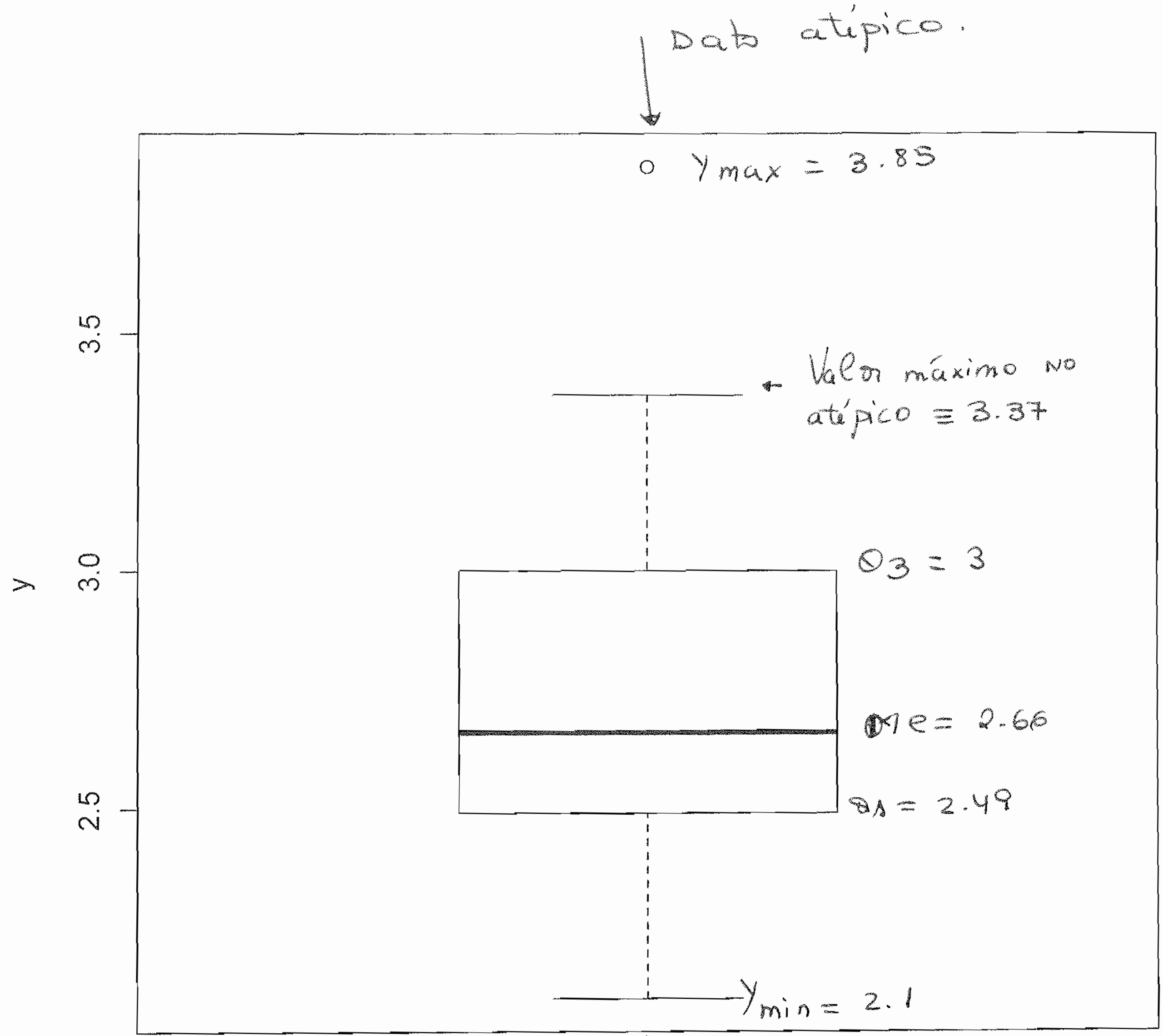
$X \rightsquigarrow \quad RIQ = 148.5 - 80.5 = 68$

$L_{sup} = 148.5 + \underbrace{1.5 * 68}_{102} = 250.5$

$L_{inf} = 80.5 - \overbrace{1.5 * 68}^{102} = -21.5$

$x_i \in [-21.5, 250.5]$ son datos NO ATÍPICOS.

DIAGRAMA de CAJA de la CARACTERÍSTICA Y



3º Para la CARACTERÍSTICA Y:

$Y \sim$ $RIQ_Y = 3 - 2.49 = 0.51$

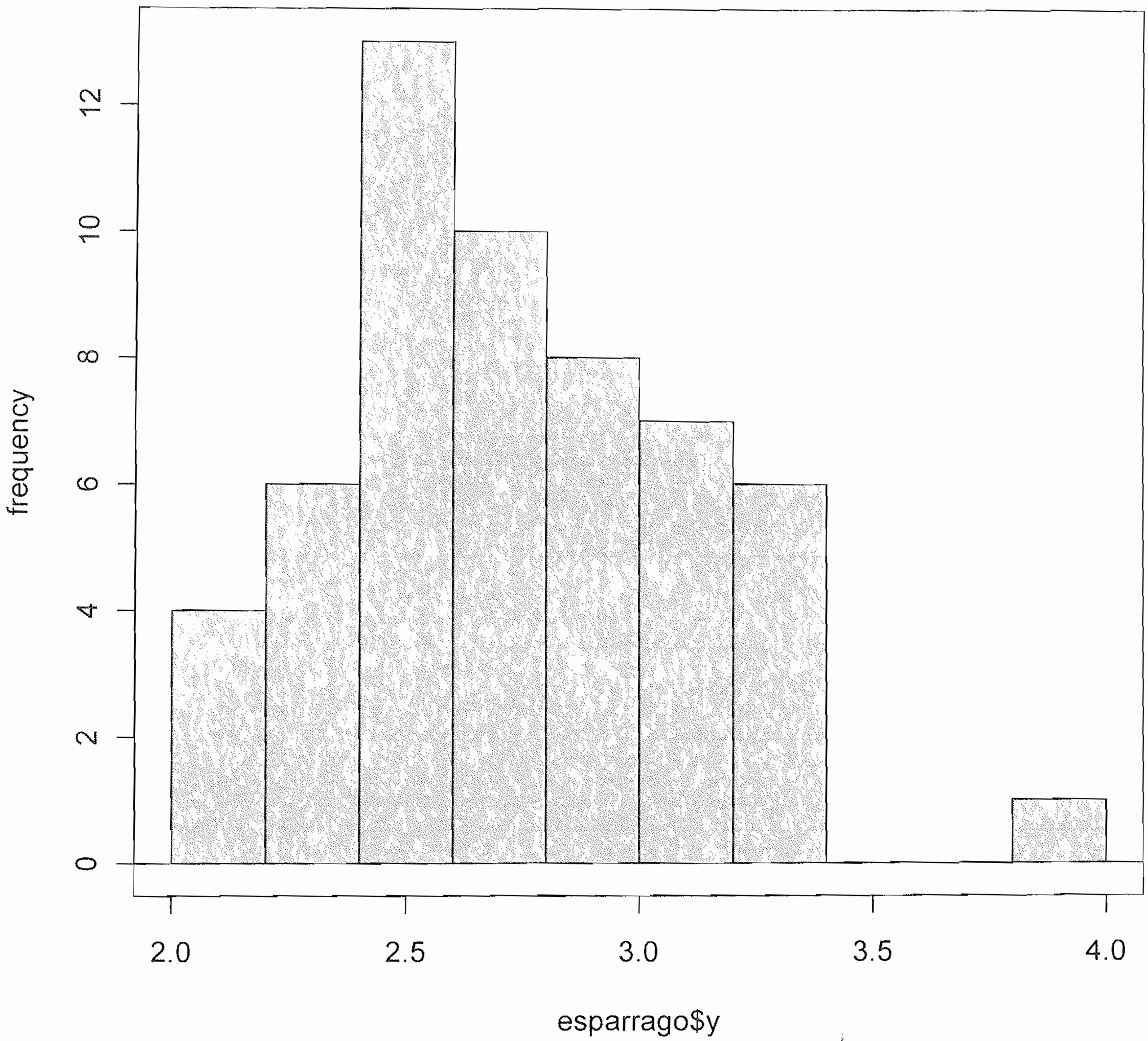
$L_{sup} = 3 + \underbrace{1.5 * 0.51}_{0.765} = 3.765$

$L_{INF} = 2.49 - \overbrace{1.5 * 0.51} = 1.725$

$y_i \in [1.725, 3.765]$ son datos NO ATÍPICOS.

4: HISTOGRAMA para Y:

Características: - Unimodal,
- Asimétrico a la derecha con existencia de dato atípico



medidas descriptivas adecuadas:

- m. centro: Mediana
- m. dispersión: RIQ

| Debido a la existencia del dato atípico.

5º $Y \sim N(\mu = 2.74, \sigma = 0.37)$

Ventana de instrucciones:

```
pnorm(c(2.6), mean=2.74, sd=0.37, lower.tail=FALSE)
pnorm(c(3.2), mean=2.74, sd=0.37, lower.tail=TRUE)
pnorm(c(2.5), mean=2.74, sd=0.37, lower.tail=TRUE)
```

Ventana de resultados:

```
> pnorm(c(2.6), mean=2.74, sd=0.37, lower.tail=FALSE)
[1] 0.6474252

> pnorm(c(3.2), mean=2.74, sd=0.37, lower.tail=TRUE)
[1] 0.8931109

> pnorm(c(2.5), mean=2.74, sd=0.37, lower.tail=TRUE)
[1] 0.2582828
```

- $P(Y > 2.6) = \boxed{0.6474}$

- $P(2.5 < Y \leq 3.2) = P(Y \leq 3.2) - P(Y \leq 2.5) =$

  $= 0.8931 - 0.2583 = \boxed{0.6348}$

6.º Intervalo de confianza para la media de Y:

```
> t.test(esparrago$y, alternative='two.sided', mu=0.0, conf.level=.97)

        One Sample t-test

data:  esparrago$y
t = 54.3374, df = 54, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
97 percent confidence interval:
 2.630742 2.855803
sample estimates:
mean of x
 2.743273
```

→ El IC al 97% para μ_Y es (2.63 , 2.86).

- Distribución de probabilidad utilizada es: $T = \dfrac{\bar{Y} - \mu_Y}{S_Y/\sqrt{n_Y}} \sim t_{54}$

7.º Contraste planteado para decidir si la media de y es igual a 2.7:

$$\left.\begin{aligned} H_0&: \mu_Y = 2.7 \\ H_1&: \mu_Y \neq 2.7 \end{aligned}\right\}$$

$t_0 = 0.8571$; p-valor $= 0.3952$ → No podemos rechazar H_0 → El valor medio de Y no es igual a 2.7.

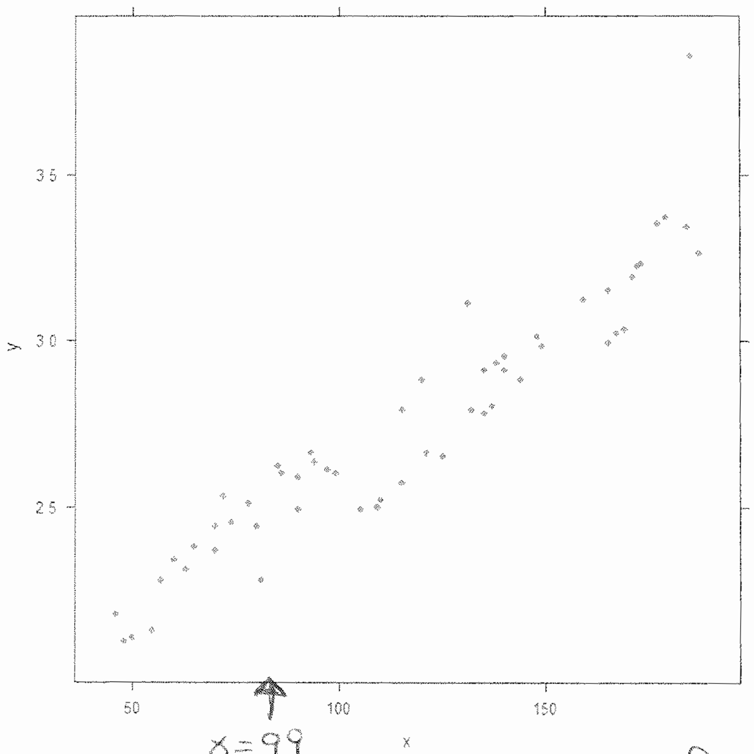
```
> t.test(esparrago$y, alternative='two.sided', mu=2.7, conf.level=.95)

        One Sample t-test

data:  esparrago$y
t = 0.8571, df = 54, p-value = 0.3952
alternative hypothesis: true mean is not equal to 2.7
95 percent confidence interval:
 2.642055 2.844491
sample estimates:
mean of x
 2.743273
```

T/ T

8º




```
> summary(RegModel.1)

Call:
lm(formula = y ~ x, data = esparrago)

Residuals:
     Min       1Q   Median       3Q      Max
-0.18091 -0.08851  0.00976  0.06760  0.53331

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.769079   0.049498   35.74   <2e-16 ***
x           0.008365   0.000400   20.92   <2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1242 on 53 degrees of freedom
Multiple R-squared: 0.8919  Adjusted R-squared: 0.8899
F-statistic: 437.5 on 1 and 53 DF,  p-value: < 2.2e-16
```

→ Dependencia lineal Positiva.

$Y = 1.769 + 0.008 * X$

$R^2 = 0.8919$ Ajuste bueno.

9º

→ • si $x = 99 \Rightarrow \hat{y} = 1.769 * 0.008 * 99 = \boxed{2.561}$

• si $x = 250 \notin$ Rango observado de los X's, entonces es no adecuado no hacer predicción pues no sería fiable