



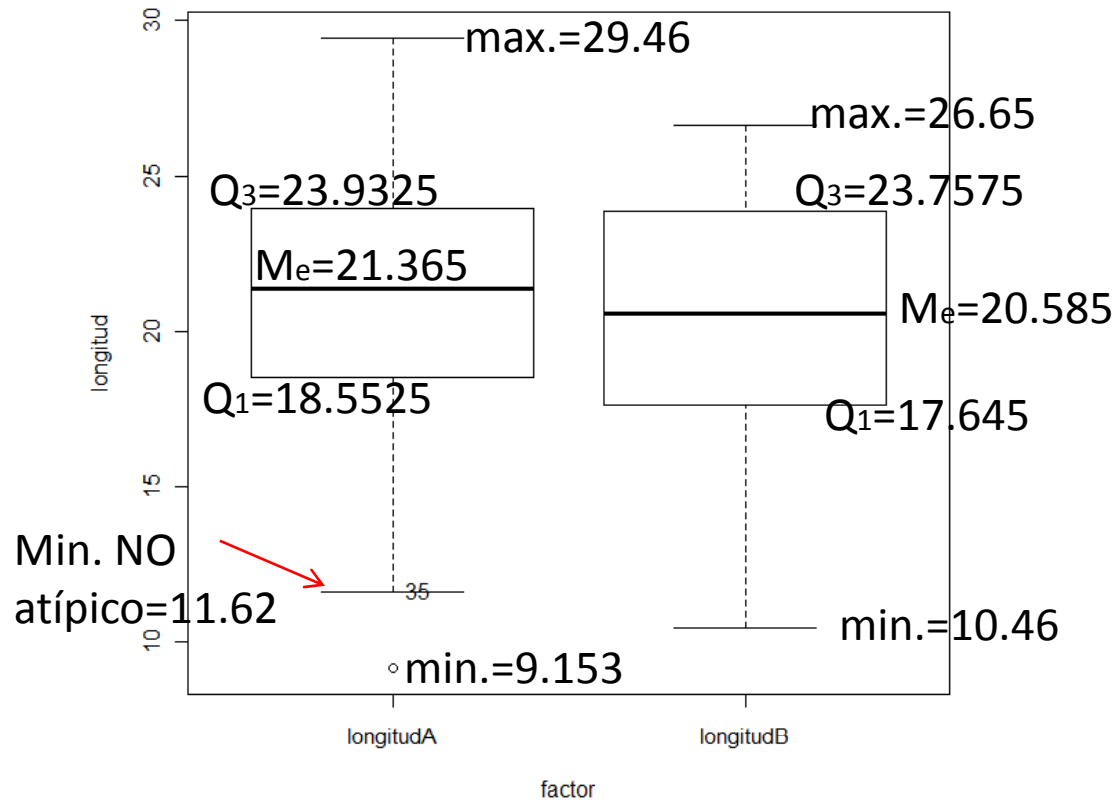
NOMBRE:..... APELLIDOS:.....
ESPECIALIDAD:.....

Para estudiar el efecto de la temperatura, en °C, sobre el vigor durante la germinación, se dispusieron semillas de dos variedades de alfalfa (**A** y **B**) en germinadores a distintas temperaturas. A los 6 días se midió la longitud de las plántulas de cada uno de los dos tipos, en mm., y los datos que se obtuvieron se encuentran en el fichero **alfalfa.txt**. Después de importar los datos, se pide:

1. Realizar un diagrama de caja y bigotes comparativo (los dos gráficos en uno) para los datos observados de la longitud de las plántulas de alfalfa de la **variedad A** y de la **variedad B** e identificar cada una de las líneas que lo constituyen, así como los valores numéricos correspondientes a las mismas. (**Copiar el gráfico obtenido de manera aproximada**). Comenta las características más relevantes: compara tendencias centrales, dispersión, simetría, presencia de atípicos, etc. A partir de los resultados anteriores, ¿qué medidas de centralización y dispersión consideras más adecuadas para resumir cada conjunto de datos? Justificar la respuesta y dar el valor numérico de estos descriptivos.
2. Según el diagrama anterior, ¿existen datos atípicos en alguna de las dos variables? Si la respuesta es positiva, indicar los valores admisibles entre los que se encontrarían los datos no atípicos para dicha(s) variable(s).
3. Supongamos que la variable "longitud de las plántulas de la **variedad A**" sigue una distribución normal de media $\mu = 21.15$ y de desviación típica $\sigma = 4.43$, determinar la probabilidad de que la longitud de las plántulas del **tipo A** sea mayor a 19 mm e inferior a 24 mm.
4. Proporcionar un intervalo de confianza al 97% para la longitud promedia de las plántulas de la **variedad A**. Indicar qué distribución de probabilidad has utilizado para obtener el citado intervalo.
5. A partir de los valores observados, ¿podemos afirmar al 95% de confianza que la longitud de las plántulas de la **variedad A** es significativamente mayor a 19.5 mm.? Indicar la prueba realizada y responder a partir del p-valor correspondiente.
6. A partir de los datos observados, ¿existen diferencias significativas entre la longitud media de las plántulas de alfalfa de la **variedad A** y de la **variedad B**? Responder a esta pregunta realizando un contraste de hipótesis. Indicar el valor del estadístico del contraste y del p-valor de la prueba. ¿Qué decisión tomaría al 95% de confianza?
7. Se desea proponer un modelo que permita predecir el comportamiento de la variable "longitud de las plántulas de la **variedad A**" a partir de la temperatura. ¿Qué modelo parece adecuado? (**Copiar de manera aproximada el gráfico en el que has basado tu respuesta**).
8. Realizar un ajuste lineal con el fin de predecir el comportamiento de la variable "longitud de las plántulas de la **variedad A**" a partir de la temperatura. Determinar los parámetros de la recta ajustada así como el valor del coeficiente de determinación, dando una interpretación del mismo.
9. Para una temperatura de 32°C, proporcionar una estimación de la "longitud de las plántulas de la **variedad A**" a los 6 días. ¿Es fiable esta estimación? Razonar la respuesta.

PRACTICAS DE SEPTIEMBRE 2014: Fichero de datos **alfalfa.txt**

1.-



Observamos que **longitudA** presenta una tendencia central (**mediana**) de mayor magnitud que **longitudB**. Por otra parte, la **longitudB** presenta una dispersión (**rango intercuartílico o amplitud de la caja**) de mayor magnitud que **longitudA**. La **longitudA** (si excluimos el dato atípico) presenta simetría mientras que **longitudB** presenta asimetría a la izquierda pues el bigote inferior es el más largo que el superior. Sólo **longitudA** presenta un dato atípico.

Con lo cual, podemos tomar la media aritmética como medida de centralización y la desviación típica como medida de dispersión en **longitudB** y la mediana y el rango intercuartílico, respectivamente, en la **longitudA**. Sin embargo como el dato atípico de la **longitudA** no está muy alejado del resto de los datos también podríamos tomar la media aritmética y la desviación típica.

```
> numSummary(alfalfa[,c("longitudA", "longitudB")], statistics=c("mean")
      mean      sd      0%      25%      50%      75%      100%  n
longitudA 21.14532 4.427471  9.153 18.5525 21.365 23.9325 29.46 40
longitudB 20.58125 3.989617 10.460 17.6450 20.585 23.7575 26.65 40
```

2.-

La variable `longitudA` presenta un dato atípico, el dato 35 cuyo valor es 9.135.

Valores **No atípicos** para la variable `longitudA`:

$$RIQ=23.9325-18.5525=5.38$$

$$[Q1-1.5*RIQ, Q3+1.5*RIQ]=[18.5525-1.5*5.38, 23.9325+1.5*5.38]=$$
$$=[10.4825, 32.0025]$$

Entonces la variable `longitudA` presenta un dato atípico: 9135

3.- Supongamos que $longitudA \rightarrow N(\mu = 21.15, \sigma = 4.43)$

```
> pnorm(c(19,24), mean=21.15, sd=4.43, lower.tail=TRUE)
[1] 0.3137221 0.7399985
```

Entonces,

$$p(19 < longitudA < 24) = \\ = 0.7399985 - 0.3137221 = 0.4262764$$

4.- Intervalo de confianza al 97% para la media de la variable `longitudA` es:

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

```
One Sample t-test
```

```
data: alfalfa$longitudA
```

```
t = 30.2057, df = 39, p-value < 2.2e-16
```

```
alternative hypothesis: true mean is not equal to 0
```

```
97 percent confidence interval:
```

```
19.56854 22.72211
```

```
sample estimates:
```

```
mean of x
```

```
21.14532
```

Para construir el intervalo de confianza de la media de la variable `longitudA` utilizamos la distribución:

t_{39}

5.- Nos planteamos el contraste:

$$\left. \begin{array}{l} H_0 : \mu_{longitudA} = 19.5 \\ H_1 : \mu_{longitudA} > 19.5 \end{array} \right\}$$

```
> t.test(alfalfa$longitudA, alternative='greater', mu=19.5, conf.level=.95)
```

```
One Sample t-test
```

```
data: alfalfa$longitudA
t = 2.3503, df = 39, p-value = 0.01196
alternative hypothesis: true mean is greater than 19.5
95 percent confidence interval:
 19.96584      Inf
sample estimates:
mean of x
 21.14532
```

Obtenemos un p-valor=0.01196<0.05 ➡ Rechazamos H0 con una gran confianza

➡ El nivel medio de la longitud de las plantas tipo A es significativamente MAYOR a 19.5 mm.

6.- Nos piden que comparemos las medias de las variables longitudA y longitudB, esto es, que planteamos el contraste:

$$\left. \begin{array}{l} H_0 : \mu_{longitudA} = \mu_{longitudB} \\ H_1 : \mu_{longitudA} \neq \mu_{longitudB} \end{array} \right\}$$

```
> t.test(longitud~factor, alternative='two.sided', conf.level=.95, var.equal=FALSE, data=Datosapilados_alfalfa)

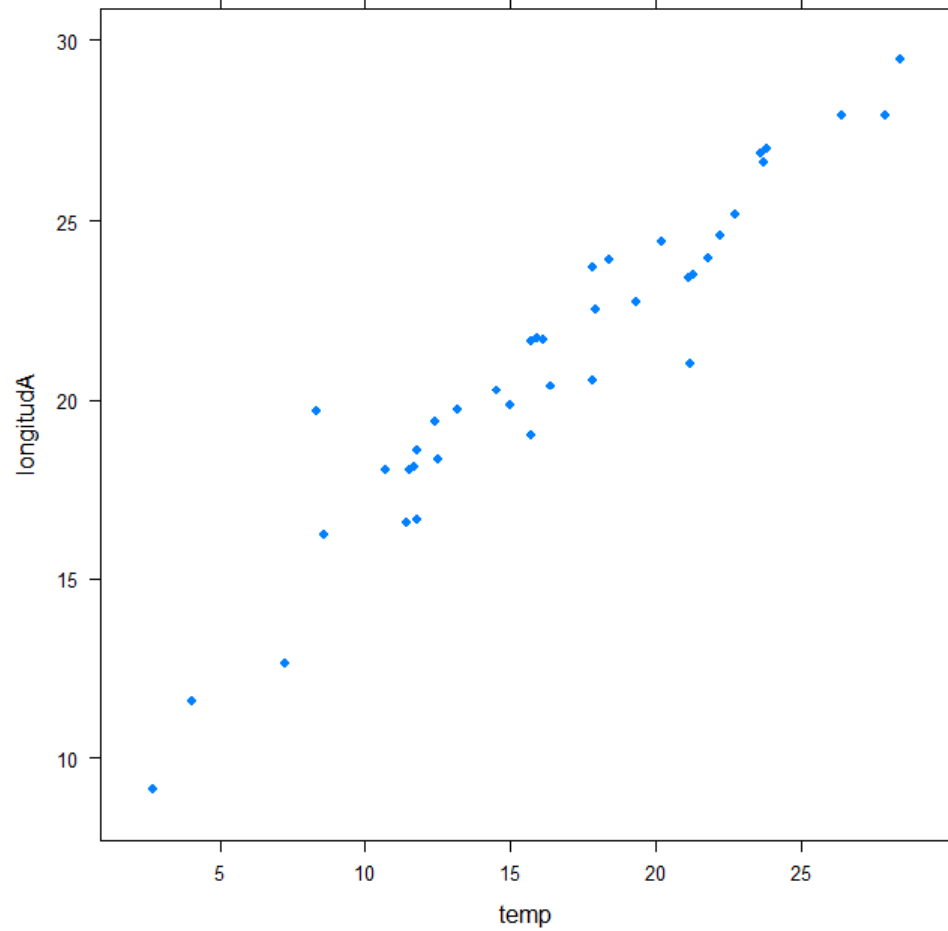
Welch Two Sample t-test

data: longitud by factor
t = 0.5986, df = 77.169, p-value = 0.5512
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.312281  2.440431
sample estimates:
mean in group longitudA mean in group longitudB
      21.14532           20.58125
```

Obtenemos un p-valor=0.5512 \geq 0.05 \rightarrow NO podemos rechazar H0.

\rightarrow No hay evidencias de que la longitud media de ambas semillas sean distintas.

7.-



Vemos que la relación entre **longitudA** y la temperatura es aparentemente lineal con dependencia positiva.

8.-

```
> RegModel.1 <- lm(longitudA~temp, data=alfalfa)

> summary(RegModel.1)

Call:
lm(formula = longitudA ~ temp, data = alfalfa)

Residuals:
    Min       1Q   Median       3Q      Max
-3.3056 -0.8989  0.2423  0.8454  4.1592

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.83023    0.60250   16.32  <2e-16 ***
temp         0.68516    0.03422   20.02  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.32 on 38 degrees of freedom
Multiple R-squared: 0.9134, Adjusted R-squared: 0.9111
F-statistic: 400.8 on 1 and 38 DF, p-value: < 2.2e-16
```

$$\rightarrow \textit{longitudA} = 9.83023 + 0.68516 * \textit{Temp}$$

$$R^2 = 0.9134$$

El ajuste es muy bueno, ya que el 91.34% de la *longitudA* está explicado por la temperatura.

- 9.- El valor estimado para longitud de las plántulas del tipo A a los 6 días cuando la temperatura es de 32°C es:

$$\textit{longitudA} = 9.83023 + 0.68516 * 32 = 31.75535\textit{mm}.$$

Esta estimación no podemos tomarla como fiable pues, aunque el ajuste es bueno, la temperatura de 32°C NO pertenece al rango observado de las temperaturas=[2.7°C , 28.4°C]