

# Regresión y ANOVA: Pulgones \*

García Prado, Sergio  
sergio@garciparedes.me

21 de octubre de 2017

## 1. Descripción del conjunto de datos

El conjunto de datos sobre el cual se va a realizar el análisis de la varianza se refiere a una serie de mediciones sobre el número de pulgones por planta de trigo. El experimento fue realizado recogiendo 40 plantas (muestras aleatorias que supondremos independientes) de trigo, durante un periodo de 6 semanas.

Para la realización de este análisis se ha utilizado la plataforma SAS [2], en concreto la *University Edition*. En este caso, el conjunto de datos ha sido suministrado en forma de fragmento de código, el cual se incluye en la figura 2. El conjunto de datos sigue una estructura tabular de 240 filas (referidas a cada observación) y 3 columnas (referidas a la **semana**, identificador de **muestra** en esa semana y **recuento** de pulgones en dicha observación) tal y como se muestra en la figura 1. El código *SAS* utilizado en este caso se muestra en la figura 3.

Obs	semana	repet	recuento
1	1	1	12
2	1	2	1
3	1	3	6
4	1	4	1
5	1	5	5

Figura 1: Visión preliminar del conjunto de datos pulgones

## 2. Cuestiones

El objetivo general del estudio es el siguiente: “**Se trata de analizar si existen diferencias en el número de pulgones por planta entre las diferentes semanas**”, para lo cual se proponen una serie de sub-objetivos que se tratarán de responder en las siguientes secciones.

---

\*URL: <https://github.com/garciparedes/anova-pulgones>

- 2.1. ¿Es adecuado utilizar un modelo de un factor para ello? Haz un análisis descriptivo de los datos por semanas y valora las hipótesis que se asumen en el modelo.

[TODO ]

- 2.2. Realiza el contraste de igualdad de medias y analiza los residuos. ¿Qué conclusiones sacas?

[TODO ]

- 2.3. Realiza el test de Levene. ¿Te sorprende el resultado?

[TODO ]

- 2.4. Transforma la respuesta mediante  $\log(\text{recuento} + 1)$  y repite el apartado 2.2. ¿Qué cambios observas?

[TODO ]

- 2.5. Realiza el test de kruskal-Wallis sobre los datos originales para contrastar la igualdad de medias

[TODO ]

### 3. Código fuente

```

data pulgones;
  do semana=1 to 6;
    do repet=1 to 40;
      input recuento @@;
      output;
    end;
  end;
datalines;
12 1 6 1 5 7 1 1 2 1 20 0 9 7 0 12 2 0 0 2 8 0 11 2 21 0 3 18 2 2 6 6
5 1 12 0 3 1 1 18 40 16 32 15 44 41 43 53 67 21 6 31 15 11 21 40 15 50
17 32 24 7 25 11 64 22 50 27 3 46 45 10 8 27 34 19 86 83 17 36 86 63
20 68 55 42 24 29 20 27 26 63 40 46 7 15 10 30 46 26 15 42 6 28 7 9 5
35 6 9 108 38 35 64 21 20 62 25 0 0 29 2 3 0 4 2 6 7 5 4 6 0 0 5 1 3 2
2 2 5 0 1 1 0 3 1 2 0 3 3 18 7 21 0 0 0 2 3 0 40 5 7 0 0 0 1 1 2 1 0
25 1 0 0 0 0 0 0 0 5 0 2 0 0 0 2 0 0 0 4 0 0 0 0 2 0 0 0 0 2 1 0 0 1 7
0 0 0 4 1 5 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
;
run;

```

Figura 2: *Código SAS*: Lectura del conjunto de datos

```

proc print data=pulgones (obs=5) n;
run;

```

Figura 3: *Código SAS*: Vista preliminar del conjunto de datos

## Referencias

- [1] BARBA ESCRIBÁ, L. Regresión y ANOVA, 2017/18. Facultad de Ciencias: Departamento de Estadística.
- [2] SAS® SOFTWARE INSTITUTE. Sas. <https://www.sas.com/>.