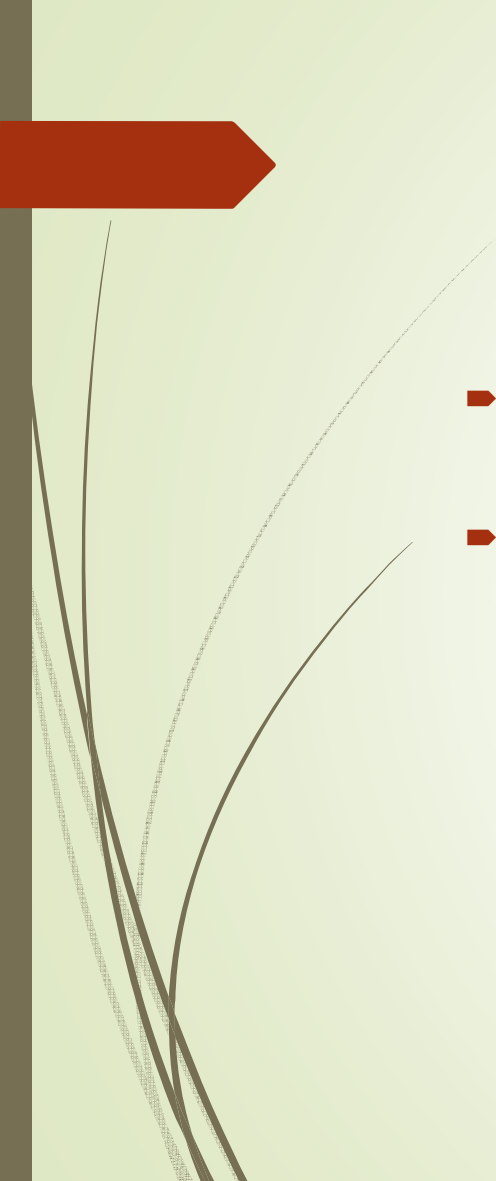


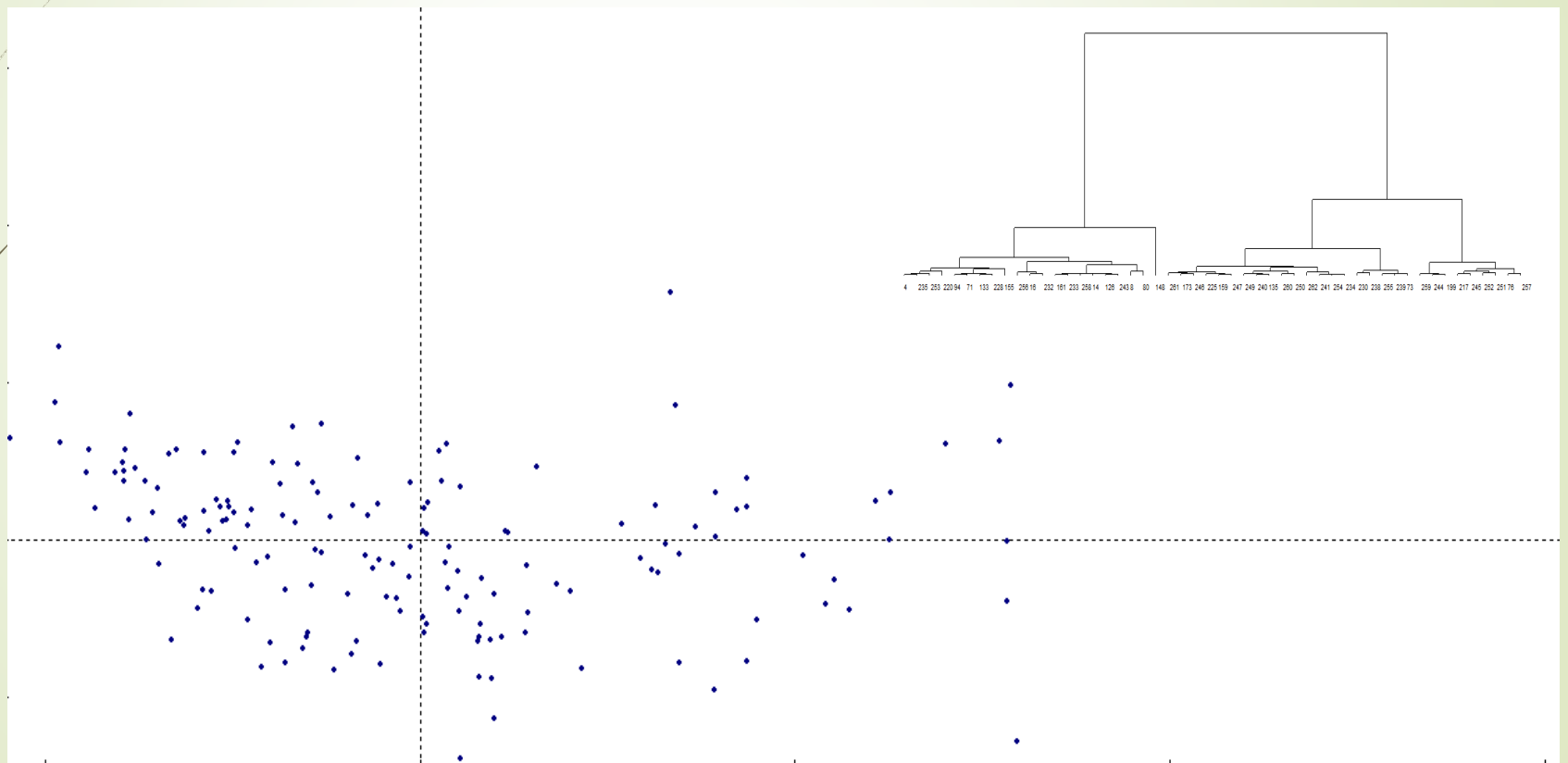
# **TEST DE HOMOCEDASTICIDAD PARA OBSERVACIONES PREVIAMENTE CLASIFICADAS**

Alfredo Baronio – Ana Vianco

Universidad Nacional de Río Cuarto - Universidad Nacional de Villa María

- 
- OBJETIVO: proponer un test para el contraste de homocedasticidad.
  - HIPOTESIS: es posible identificar la estructura de heterocedasticidad a partir de la clasificación de los individuos en clases mutuamente excluyentes, incorporar esa información en el modelo considerando la aplicación de variables ficticias y obtener estimaciones homocedásticas.

# Econometría con observaciones previamente clasificadas



## idea inicial...

- 1. especificar y estimar  $X_{1i} = \beta_1 + \sum_{j=2}^p \beta_j X_{ji} + \varepsilon_j \quad \forall i = 1, \dots, n \quad j = 2, \dots, p$
- 2. clasificación jerárquica de Ward:  $n \times p$  en  $K$  grupos
- 3. en los  $K$  grupos: estimar el modelo y calcular la varianza
- 4. contrastar  $H_0: \sigma_r^2 = \sigma_s^2 \quad \forall r \neq s; r < s \quad r, s = 1, 2, \dots, K$

*La cantidad de hipótesis a docimar:  $K(K - 1)/2$*

- 5. utilizar el estadístico  $F^* = \frac{e_r' e_r / (n_r - p)}{e_s' e_s / (n_s - p)} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_s^2} \quad \forall r \neq s \quad r, s = 1, 2, \dots, K$
- 6. regla de decisión:

si  $F_{\alpha/2; (n_r - p, n_s - p)} \leq F^* \leq F_{(1-\alpha/2); (n_r - p, n_s - p)}$  se acepta  $H_0$

si  $F^* < F_{\alpha/2; (n_r - p, n_s - p)} \circ F^* > F_{(1-\alpha/2); (n_r - p, n_s - p)}$  se rechaza  $H_0$

Si solo se aceptan  $G < K(K - 1)/2$  hipótesis planteadas, se pueden reunir los grupos homocedásticos

# test de homocedasticidad

- 1. especificar y estimar  $X_{1i} = \beta_1 + \sum_{j=2}^p \beta_j X_{ji} + \varepsilon_j \quad \forall i = 1, \dots, n \quad j = 2, \dots, p$
- 2. clasificación jerárquica de Ward:  $n \times p$  en  $K$  grupos
- 3. en los  $K$  grupos: estimar el modelo y calcular la varianza
- 4. obtener estimación centrada de  $\sigma^2$

$$\overline{S^2} = \frac{(n_1-p)S_1^2 + (n_2-p)S_2^2 + \dots + (n_K-p)S_K^2}{n_1 + n_2 + \dots + n_K - pK} \quad S_k^2 = \frac{e_k' e_k}{n_k - p} \quad \forall k = 1, \dots, K,$$

- 5. obtener estimación por intervalo con el estadístico  $\frac{\chi^2_{(n_1+n_2+\dots+n_K-pK)}}{(n_1+n_2+\dots+n_K-pK)} = \frac{\overline{S^2}}{\sigma^2}$

$$P\left(\frac{(n_1 + n_2 + \dots + n_K - pK) \overline{S^2}}{\chi^2_{(1-\alpha/2), (n_1+n_2+\dots+n_K-pK)}} < \sigma^2 < \frac{(n_1 + n_2 + \dots + n_K - pK) \overline{S^2}}{\chi^2_{(\alpha/2), (n_1+n_2+\dots+n_K-pK)}}\right) = 1 - \alpha$$

➤ 6.  $H_0$ : Homocedasticidad  $\sigma^2 = \sigma_k^2 \quad \forall k = 1, \dots, K$

➤ 7. La regla de decisión es:

Si para todo  $k = 1, \dots, K$ ,  $\frac{(n_1 + \dots + n_K - pK) \bar{S}^2}{\chi^2_{(1-\alpha/2), (n_1 + \dots + n_K - pK)}} \leq S_k^2 \leq \frac{(n_1 + \dots + n_K - pK) \bar{S}^2}{\chi^2_{(\alpha/2), (n_1 + \dots + n_K - pK)}}$  ACEPTA  $H_0$

Si para algún  $k = 1, \dots, K$ ,  $S_k^2 < \frac{(n_1 + \dots + n_K - pK) \bar{S}^2}{\chi^2_{(1-\alpha/2), (n_1 + \dots + n_K - pK)}}$  o  $S_k^2 > \frac{(n_1 + \dots + n_K - pK) \bar{S}^2}{\chi^2_{(\alpha/2), (n_1 + \dots + n_K - pK)}}$  RECHAZA  $H_0$

- si solo se aceptan  $G < K$  grupos homocedásticos, se puede proceder a reagrupar las observaciones que dan lugar a errores homocedásticos en un solo grupo.
- La identificación de los grupos permitiría incorporar variables ficticias al modelo, salvando el error de especificación y posibilitando estimaciones homocedásticas.

# aplicaciones

- $X_{1i} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i \quad \forall i = 1, \dots, 100$
- Se clasifican las 100 muestras en  $K = 3$  grupos
- Se especifica y estima el modelo en cada grupo y se calcula

$$S_1^2 = \frac{e_1' e_1}{n_1 - p} = 2951,01 \quad S_2^2 = \frac{e_2' e_2}{n_2 - p} = 2371,80 \quad S_3^2 = \frac{e_3' e_3}{n_3 - p} = 2566,06$$

- La estimación centrada puntual de la varianza:

$$\overline{S^2} = \frac{(n_1 - p)S_1^2 + (n_2 - p)S_2^2 + \dots + (n_K - p)S_K^2}{n_1 + n_2 + \dots + n_K - pK} = 2620,23$$

- La estimación del intervalo de confianza para la varianza:

$$P(1998.97 < \sigma^2 < 3585.54) = 0.95$$

- Se acepta  $H_0: \sigma^2 = \sigma_1^2 = \sigma_2^2 = \sigma_3^2$

- $X_{1i} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i; \quad \forall i = 1, \dots, 156$
- Se clasifican las 156 muestras en  $K = 3$  grupos
- Se especifica y estima el modelo en cada grupo y se calcula

$$S_1^2 = \frac{e_1' e_1}{n_1 - p} = 18.76 \quad S_2^2 = \frac{e_2' e_2}{n_2 - p} = 6.39 \quad S_3^2 = \frac{e_3' e_3}{n_3 - p} = 2.24$$

- La estimación centrada puntual de la varianza es

$$\overline{S^2} = \frac{(n_1 - p)S_1^2 + (n_2 - p)S_2^2 + (n_3 - p)S_3^2}{n_1 + n_2 + n_3 - pK} = 7.36$$

- La estimación del intervalo de confianza para la varianza

$$P(5.91 < \sigma^2 < 9.40) = 0.95$$

- Se rechaza  $H_0: \sigma^2 = \sigma_1^2 = \sigma_2^2 = \sigma_3^2$





# conclusión

- El estudio se inspira en el contraste de Goldfeld y Quandt y se complementa con las técnicas de clasificación de la información y partición.
- La prueba propuesta conserva toda la información, no transforma las variables y no utiliza resultados asintóticos.
- Limitante: insuficientes grados de libertad en las clases.