COMPARACIÓN Y ORDENAMIENTO DE *K* ECSIS: UNA PRUEBA ÓMNIBUS Y UN PROCEDIMIENTO BOOTSTRAP

Sergio Juárez
Universidad Veracruzana

ITAM, Noviembre, 2018

Agradezco al Dr. Luis Enrique Nieto la invitación

CONTENIDO

- Motivación: Estudio de medición de satisfacción de los estudiantes de la Universidad Veracruzana
- Instrumento: Modelo ECSI (European Customer Satisfaction Index)
- Problema: Comparación de ECSIs y su ordenamiento
- Resultados del estudio en la Universidad Veracruzana
- Comentarios finales

UNIVERSIDAD VERACRUZANA (UV)

- Institución de educación superior con mas de 78,000 estudiantes inscritos en 174 programas de pregrado y 132 programas de posgrado (https://www.uv.mx/numeralia/).
- El ranking de las mejores universidades mexicanas, coloca a la UV en los lugares 12, 15, 12 y 14 de un total de 50, para los años 2014, 2015, 2016 y 2017, respectivamente.

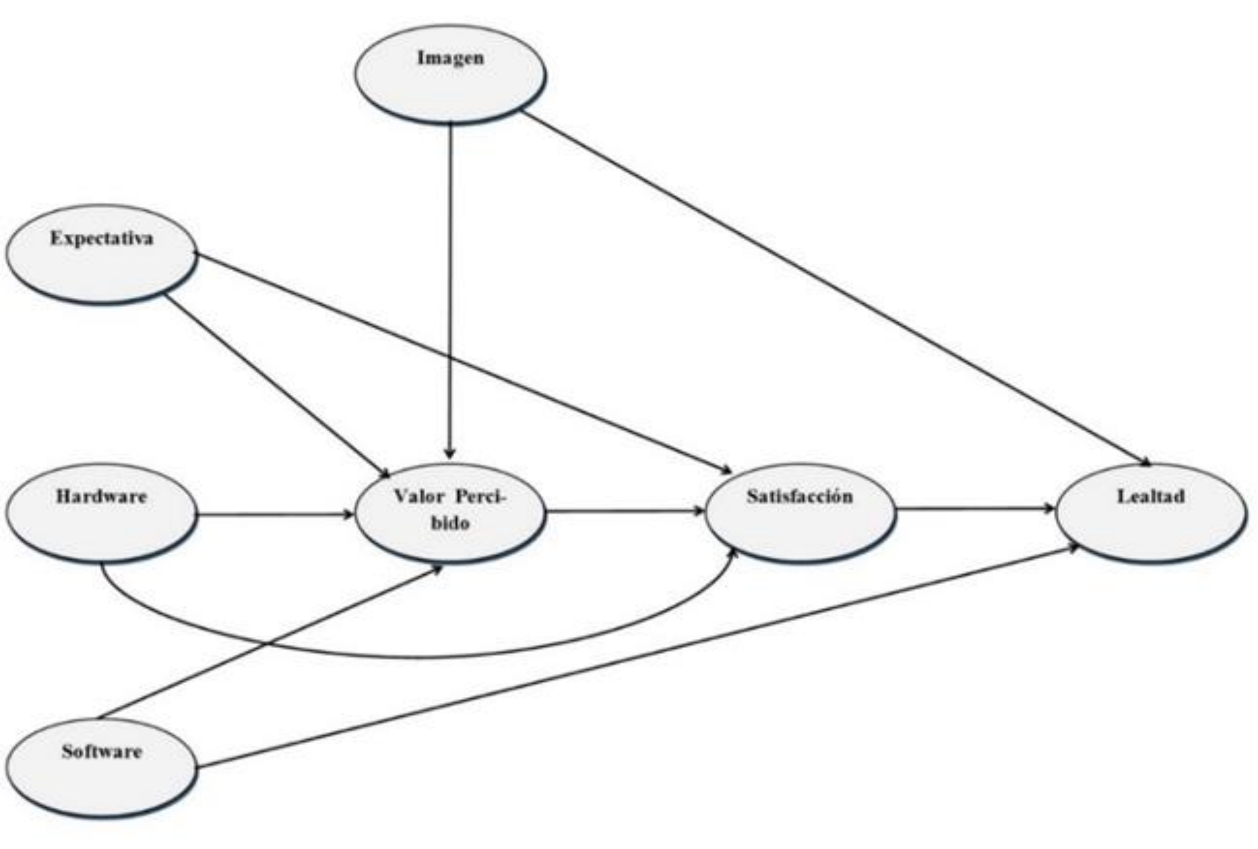
INSTALACIONES DE LA UV

INSTALACIONES DE LA UV

ESTUDIO DE SATISFACCIÓN DE ESTUDIANTES

- En el 2016 se conduce un estudio para la medición de la satisfacción de los estudiantes con los aspectos académicos que les ofrece la institución.
- Población de interés: Carreras que se ofrecen dentro del área Económico-Administrativa en sus cinco regiones.
- Muestra aleatoria estratificada de estudiantes de licenciatura. Estratos determinados por Carrera-Región.
- Metodología: Índice de Satisfacción de Clientes Europeo (ECSI)

MODELO ECSI
ECSI TECHNICAL COMMITTEE (1998)



ESTUDIOS SIMILARES

Los estudios de satisfacción son comunes en las instituciones de educación superior. Algunos de los que se han hecho con la metodología ECSI:

Autores	Institución de educación superior
Martensen et al. (2000)	Escuela de Negocios Aarhus, Dinamarca
Duque Zuluaga (2003)	Universidad de Barcelona, España
Alves y Raposo (2004)	Universidad de Beira, Portugal
Chitty y Soutar (2004)	Universidad de Australia
Østergaard y Kristensen (2005)	Escuela de Negocios Aarhus, Dinamarca
Chiandotto et al. (2007)	Universidad de Florencia, Italia
Zhan y Gao (2008)	Universidad de China
Brown y Mazzarol (2009)	Cuatro universidades de Australia
Balzano y Trinchera (2011)	Universidad de Italia
Duarte, Raposo, y Alves (2012)	Universidad de Portugal
Álvarez y Vernazza (2013)	Universidad de la República, Uruguay
Sosa Galindo, Juárez Cerrillo y Cruz Kuri (2014)	Universidad Veracruzana, México
Viñán Andino (2015)	Escuela Politécnica Superior del Chimborazo, Ecuador.

ECUACIONES DEL MODELO

Modelo estructural

- $\mathrm{V} = \beta_{1,0} + \beta_{1,1}\mathrm{I} + \beta_{1,2}\mathrm{E} + \beta_{1,3}\mathrm{HW} + \beta_{1,4}\mathrm{SW} + \varepsilon_1,$
- $\mathrm{S} = \beta_{2,0} + \beta_{2,1}\mathrm{E} + \beta_{2,2}\mathrm{V} + \beta_{2,3}\mathrm{HW} + \varepsilon_2,$
- $\mathrm{L} = \beta_{3,0} + \beta_{3,1}\mathrm{I} + \beta_{3,2}\mathrm{S} + \beta_{3,3}\mathrm{SW} + \varepsilon_3.$

Sean $\eta = (\mathrm{V}, \mathrm{S}, \mathrm{L})'$, $\xi = (\mathrm{I}, \mathrm{E}, \mathrm{HW}, \mathrm{SW})'$ y $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)'$; se hacen los supuestos $\mathrm{E}(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \Sigma_{\varepsilon,}$, $\mathrm{Cov}(\eta, \varepsilon) = 0$, $\mathrm{Cov}(\xi\,, \varepsilon) = 0.$

ECUACIONES DEL MODELO

Modelo de medición para las variables latentes endógenas η

$$Y = \Lambda_Y \eta + \delta_Y,$$

Y es el vector de variables observables que operacionalizan a las variables latentes endógenas η. La matriz Λ_Y contiene los coeficientes de regresión. El término δ_Y es un vector de errores que satisface $\mathrm{E}(\delta_Y) = 0$, $\mathrm{Var}(\delta_Y) = \Sigma_{\delta_Y}$ y $\mathrm{Cov}(\eta, \delta_Y) = 0$.

Modelo de medición para las variables latentes exógenas ξ

$$X = \Lambda_X \xi + \delta_X,$$

X es el vector de variables observables que operacionalizan a las variables latentes exógenas. La matriz Λ_X tiene a los coeficientes de regresión. El vector δ_X es un término de error que satisface $\mathrm{E}(\delta_\mathrm{X}) = 0$, $\mathrm{Var}(\delta_\mathrm{X}) = \Sigma_{\delta_\mathrm{X}}$ y $\mathrm{Cov}(\xi, \delta_\mathrm{X}) = 0$.

niversidad Veracruzana (2016).

VARIABLES DEL MODELO DE MEDICIÓN

Variables Latentes	Variables de Medición
ξ_1: Imagen Institucional	Imagen global de la UV como institución de educación superior. Responsabilidad y compromiso social de la institución. Credibilidad y ética de la institución.
ξ_2: Expectativa de la calidad	Calidad del contenido de los programas de las experiencias educativas. Calidad de las aulas. Calidad de las bibliotecas. Calidad de los centros de cómputo. Calidad del nivel académico de los profesores. Calidad del servicio proporcionado por las autoridades administrativas (directores, secretarios académicos) y secretarias.
ξ_3: Calidad percibida en los elementos no humanos (hardware)	Calidad global del contenido de los programas de las experiencias educativas. Calidad global de la oferta educativa. Calidad global de los horarios. Calidad global de la tutoría. Calidad global de las aulas. Calidad global de las bibliotecas. Calidad global de los centros de cómputo. Calidad global de la información recibida respecto al Modelo Educativo. Calidad total de las estrategias innovadoras para el aprendizaje que utilizan tus profesores.
ξ_4: Calidad percibida en los elementos humanos (software)	Calidad global del proceso de enseñanza-aprendizaje que desarrolla el profesor. Calidad global de tu tutor en su quehacer tutorial. Calidad global del servicio prestado por las autoridades académicas (director, secretario, jefe de carrera y/o departamento). Calidad global del servicio prestado por el personal administrativo (administradores, secretarias, encargados de bibliotecas, encargados de centros de cómputo).
η_1: Valor Percibido	Valor relativo de la educación que estás adquiriendo en términos de lo que te está costando (en tiempo, dinero y esfuerzo). Valor relativo de la educación que estás adquiriendo en términos al beneficio esperado (Acceso al empleo). Valor relativo de la educación que estás adquiriendo en términos al beneficio esperado (Acceso al posgrado).
$S = \eta_2$: Satisfacción	**1. Satisfacción global con la UV.** **2. El grado en que se han llenado las expectativas que tenías de la UV** **3. La UV en comparación con la institución de educación superior ideal para ti.**
η_3: Lealtad	Continuar con el estudio de un posgrado en la UV. Recomendar la UV a otros estudiantes. Recomendar la carrera que estudias a otros estudiantes.

EL ECSI

Supongamos que la variable latente Satisfacción tiene soporte compacto dado por $[\min(S), \max(S)]$, lo cual garantiza la existencia del valor esperado $\mathrm{E}(S)$. El ECSI se define por

$$\mathrm{ECSI} = \frac{\mathrm{E(S)} - \min(\mathrm{S})}{\max(\mathrm{S}) - \min(\mathrm{S})} \times 100.$$

El ECSI es un indicador compuesto del tipo min-max que toma valores entre 0 y Un ECSI entre 0 y 55 es totalmente inaceptables, entre 55 y 60 es muy pobre, entre 60 y 65 es pobre/bajo, entre 65 y 75 es promedio, entre 75 y 80 es fuerte/bueno, entre 75 y 85 es muy fuerte y entre 85 y 100 extraordinario/único.

ESTIMACIÓN DEL ECSI

- Sean $Y_1, Y_2, \dots, Y_q$ las variables que operacionalizan a la Satisfacción S.
- Estas variables se observan en una escala tipo Likert de 1 a L, generalmente $L = 7$ o $L = 10$.
- Las variables de medición se observan en n individuos, lo que resulta en los datos $y_{i1}, \dots, y_{iq}$, $(i = 1, \dots, n)$.
- El Método de estimación **Mínimos Cuadrados Parciales** produce unas ponderaciones $w_1, \dots, w_q$ con las que se construye una variable $\hat{\text{S}} = (\hat{s}_i, \dots, \hat{s}_n)$ donde $\hat{s}_i = \sum_{j=1}^{q} w_j y_{ij}$, $(i = 1, \dots, n)$.
- Esta variable se llama score de la Satisfacción y se puede conceptualizar como una aproximación de la variable latente Satisfacción. El ECSI se estima con las contrapartes muestrales del score $\hat{\text{S}}$.
- Estimación con el paquete `plspm` de R, Sánchez (2013).

ESTIMACIÓN DEL ECSI

Contrapartes muestrales

- $\bar{\mathrm{S}} = \frac{1}{n}\sum_{i=1}^{n} \hat{s}_i = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{q} w_j y_{ij} = \sum_{j=1}^{q} w_j \bar{y}_j$
- $\min(\hat{\mathrm{S}}) = \min_i \left\{\sum_{j=1}^{q} w_j y_{ij}\right\} = \sum_{j=1}^{q} \min_i (w_j y_{ij}) = \sum_{j=1}^{q} w_j$
- $\max(\hat{\mathrm{S}}) = \max_i \left\{\sum_{j=1}^{q} w_j y_{ij}\right\} = \sum_{j=1}^{q} \max_i (w_j y_{ij}) = L\sum_{j=1}^{q} w_j$

donde $\bar{y}_j = \sum_{i=1}^{n} y_{ij} /n.$ El estimador del ECSI es

$$\widehat{\mathrm{ECSI}} = \frac{\sum_{j=1}^{q} w_j \bar{y}_j - \sum_{j=1}^{q} w_j}{(L-1)\sum_{j=1}^{q} w_j} \times 100.$$

Tabla 4. Valores del ECSI observados en las regiones.

RESULTADOS

Globalmente para toda el área se observó ECSI = 80.2

Región	Tamaño de muestra	ECSI
Xalapa	910	77.16
Veracruz	427	83.56
Córdoba-Orizaba	320	80.15
Poza Rica-Tuxpan	113	78.21
Coatzacoalcos-Minatitlán	299	84.71

CUESTIONES DE INTERÉS

- Intervalos de confianza para los ECSI
- Comparación de dos ECSIs
- Si dos ECSI son diferentes, ¿podemos estimar el tamaño de la diferencia?
- Comparación de k ECSIs
- Si k ECSIs son diferentes, ¿podemos hacer un ranking?

PROBLEMA

- Mínimos Cuadrados Parciales (MCP) no hace supuestos distribucionales en las variables latentes y en las de medición más allá de la existencia de segundos momentos.
- Las propiedades distribucionales para tamaños de muestra finitos de los estimadores de MCP son desconocidas.
- Los problemas de inferencia estadística para MCP en modelos de ecuaciones estructurales, por ejemplo cálculo de errores estándar, se han abordado con procedimientos no-paramétricos de remuestreo tales como el jackkniffe y el bootstrap; y las pruebas de permutaciones.
- **No hay procedimientos estadísticos explícitamente desarrollados para el ECSI y poder hacer lo anterior.**

PROBLEMA DE INFERENCIA ESTADÍSTICA

Desarrollar procedimientos de inferencia estadística para:

- 1. Intervalo de confianza para un ECSI.
- 2. Prueba de la hipótesis de igualdad de dos ECSIs.
- 3. Intervalo de confianza para la diferencia de dos ECSIs.
- 4. Pruebas de la hipótesis de igualdad de k ECSIs.
- 5. Ranking de k ECSIs.

INTERVALO DE CONFIANZA BOOTSTRAP PARA UN ECSI

Sea $Z = \left(X_1, \ldots, X_p, Y_1, \ldots, Y_q\right)$ la matriz de datos $n \times (p+q)$. Las hileras de Z son observaciones i.i.d. de una distribución F. El problema de inferencia es estimar a la distribución muestral de

$$\mathrm{ECSI} = \frac{\mathrm{E(S)} - \min(\mathrm{S})}{\max(\mathrm{S}) - \min(\mathrm{S})} \times 100,$$

$$G_{\mathrm{ECSI}}(x) = P(\mathrm{ECSI(Z)} \leq x).$$

Se estima a G_{ECSI} con la distribución $\hat{G}_{\mathrm{ECSI}}$ producida por el algoritmo bootstrap no paramétrico presentado a continuación.

¿POR QUÉ FUNCIONA EL BOOTSTRAP PARA EL ECSI?

MAMMEN, E. (1992): **When Does Bootstrap Work?** Springer: New York.

Supongamos que las variables aleatorias $X_1, \dots, X_n$, son i.i.d. con distribución $F \in \mathcal{F}$ y espacio muestral $\mathcal{X}$.

Sea θ una función que va de $\mathcal{X} \times \mathcal{F}$ a $\mathbb{R}$. Interesa estimar a la distribución de $\theta(X_1, \dots, X_n; F)$, $G_{F,n}(x) = P(\theta(X_1, \dots, X_n; F) \leq x)$.

Un estimador de $G_{F,n}$ es la distribución bootstrap $G_{F_n}(x) = P(\theta(X_1, \dots, X_n; F_n) \leq x)$, donde F_n es la función de distribución empírica de $X_1, \dots, X_n$.

Se dice que G_{F_n} es consistente si para cada $\epsilon > 0$ y cada $F \in \mathcal{F}$ se tiene que

$$\lim_{n \to \infty} P[\sup_x |G_{F_n}(x) - G_{F,\infty}(x)| \leq \epsilon] = 1,$$

donde $G_{F,\infty}$ es la distribución asintótica de $\theta(X_1, \dots, X_n; F)$.

¿POR QUÉ FUNCIONA EL BOOTSTRAP PARA EL ECSI?

MAMMEN, E. (1992): **When Does Bootstrap Work?** Springer: New York.

Si $\theta(X_1, \dots, X_n; F)$ es una funcional lineal,

$$\theta(X_1, \dots, X_n; F) = \int g_n(x) dF(x),$$

donde g_n son funciones conocidas, entonces G_{F_n} es consistente si y sólo si existen secuencias de números t_n y σ_n tales que

$$T_n = \sum_{i=1}^{n} \frac{g_n(X_i) - t_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

EL ECSI ES UNA FUNCIONAL LINEAL

S es una variable aleatoria absolutamente continua con soporte compacto $[\min(\mathrm{S}), \max(\mathrm{S})]$ y distribución G_{S}, de modo que el valor esperado

$$\mathrm{E}(\mathrm{S}) = \int_{\min(\mathrm{S})}^{\max(\mathrm{S})} s dG_{\mathrm{S}}(s)$$

es finito y tenemos que $\mathrm{ECSI} = a - b\mathrm{E}(\mathrm{S})$, donde $a = 100 \min(\mathrm{S}) / (\max(\mathrm{S}) - \min(\mathrm{S}))$ y $b = 100/(\max(\mathrm{S}) - \min(\mathrm{S}))$ son simplemente constantes normalizadoras.

Por lo que el ECSI es un parámetro que es funcional lineal de G_{S}. Además, también por el Teorema de Mammen citado arriba, se tiene que la distribución asintótica de la distribución bootstrap $\hat{G}_{\mathrm{ECSI}}$ es normal.

ALGORITMO I: BOOTSTRAP NO PARAMÉTRICO PARA LA DISTRIBUCIÓN DE UN ECSI

```
Dada la matriz de datos n × (p + q)  Z = (X_1, …, X_p, Y_1, …, Y_q)
Para  i = 1, …, B
{
Selecciona una muestra bootstrap de las hileras de Z,  Z_i^* = (X_{i1}^*, …, X_{ip}^*, Y_{i1}^*, …, X_q^*)
Ajusta el modelo ECSI a la muestra bootstrap Z_i^*
Calcula al valor del índice ECSI_i^* = ECSI(Z_i^*)
}
Calcula la distribución bootstrap  Ĝ_ECSI(x) = (1/B) Σ_{i=1}^B 1_(−∞,ECSI_i^*](x)
```

Dada la matriz de datos $n \times (p+q)$ $Z = (X_1, \dots, X_p, Y_1, \dots, Y_q)$

Para $i = 1, \dots, B$

{

Selecciona una muestra bootstrap de las hileras de Z, $Z_i^* = (X_{i1}^*, \dots, X_{ip}^*, Y_{i1}^*, \dots, X_q^*)$

Ajusta el modelo ECSI a la muestra bootstrap Z_i^*

Calcula al valor del índice $\mathrm{ECSI}_i^* = \mathrm{ECSI}(Z_i^*)$

}

Calcula la distribución bootstrap $\hat{G}_{\mathrm{ECSI}}(x) = \frac{1}{B}\sum_{i=1}^{B} 1_{(-\infty,\mathrm{ECSI}_i^*]}(x)$

Una vez que se tiene a $\hat{G}_{\mathrm{ECSI}}$, un intervalo de confianza está dado por $\left[\widehat{\mathrm{ECSI}}_{\alpha}, \widehat{\mathrm{ECSI}}_{1-\alpha}\right]$ donde $\hat{G}^{-1}{}_{\mathrm{ECSI}}(\alpha) = \widehat{\mathrm{ECSI}}_{\alpha}$ y $\hat{G}^{-1}{}_{\mathrm{ECSI}}(1-\alpha) = \widehat{\mathrm{ECSI}}_{1-\alpha}$ son los cuantiles de la distribución bootstrap $\hat{G}_{\mathrm{ECSI}}$. Si la distribución bootstrap $\hat{G}_{\mathrm{ECSI}}$ es aproximadamente normal e insesgada, entonces otro intervalo de confianza es $[\widehat{\mathrm{ECSI}} - z_{\alpha}\hat{\sigma}, \widehat{\mathrm{ECSI}} + z_{\alpha}\hat{\sigma}]$ donde $\widehat{\mathrm{ECSI}}$ es el ECSI estimado en los datos, z_{α} es el cuantil $1-\alpha$ de la normal estándar y $\hat{\sigma}$ es la desviación estándar de $(\mathrm{ECSI}_1^*, \dots, \mathrm{ECSI}_B^*)$.

PRUEBA DE PERMUTACIONES PARA PROBAR LA HIPÓTESIS DE IGUALDAD DE DOS ECSIS

Sean $\mathrm{Z}_1 = (\mathrm{X}_{1,1}, \dots, \mathrm{X}_{1,\mathrm{p}}, \mathrm{Y}_{1,1}, \dots, \mathrm{Y}_{1,\mathrm{q}})$ y $\mathrm{Z}_2 = \left(\mathrm{X}_{2,1}, \dots, \mathrm{X}_{2,\mathrm{p}}, \mathrm{Y}_{2,1}, \dots, \mathrm{Y}_{2,\mathrm{q}}\right)$, las respectivas matrices de datos de tamaño $n_1 \times (p+q)$ y $n_2 \times (p+q)$ y cuyas hileras son observaciones i.i.d. de las distribuciones F_1 y F_2, respectivamente. Se desea probar la hipótesis de que los ECSI de las poblaciones son iguales

$$\mathrm{H}_0\colon\ \mathrm{ECSI}(\mathrm{Z}_1) = \mathrm{ECSI}(\mathrm{Z}_2).$$

Bajo el supuesto de que $F_1 = F_2 = F$, la etiqueta de la población no es informativa.

DISTRIBUCIÓN DE PERMUTACIONES

La distribución conjunta de las hileras de Z, denotadas por $z_1, \ldots, z_{n_1+n_2}$, es la misma que la de $z_{\pi(1)}, \ldots, z_{\pi(n_1+n_2)}$, donde $(\pi(1), \ldots, \pi(n_1+n_2))$ es cualquier permutación de $\{1, \ldots, n_1+n_2\}$.

Sea $\mathcal{Z}$ el conjunto de todas las matrices resultantes de permutar las hileras de Z y sea $D(Z)$, un estadístico de prueba apropiado para probar H_0 en el sentido que valores grandes de D indican que H_0 es falsa, mientras que valores pequeños de D no dan evidencia para rechazar H_0. La distribución de permutaciones de D es

$$P(D \leq d) = \frac{1}{\operatorname{card}(\mathcal{Z})} \sum_{Z \in \mathcal{Z}} 1_{(-\infty, D(z_{\pi(1)}, \ldots, z_{\pi(n_1+n_2)})]}(d)\,.$$

ALGORITMO II: PRUEBA DE PERMUTACIONES PARA COMPARAR DOS ECSIS

Dadas las matrices de datos $Z_1 = (X_{1,1}, \dots, X_{1,p}, Y_{1,1}, \dots, Y_{1,q})$ y $Z_2 = \left(X_{2,1}, \dots, X_{2,p}, Y_{2,1}, \dots, Y_{2,q}\right)$ de dimensión $n_1 \times (p+q)$ y $n_2 \times (p+q)$
Ajusta el modelo ECSI a Z_1 y Z_2, calcula $\mathrm{ECSI}(Z_1)$ y $\mathrm{ECSI}(Z_2)$ y $\widehat{D} = \mathrm{ECSI}(Z_2) - \mathrm{ECSI}(Z_1)$
Sea p = 0
Para $i = 1, \dots, M$
{
Determina una matriz de permutación de hileras de $Z = (Z_1, Z_2)'$, $Z_i^* = (Z_{i1}^*, Z_{i2}^*)' \in \mathcal{Z}$
Calcula el estadístico de prueba $D_i^*(Z_i^*) = \mathrm{ECSI}(Z_{i2}^*) - \mathrm{ECSI}(Z_{i1}^*)$
Si $D_i^* \geq \widehat{D}$, p = p + 1
}
p-valor = p/(M+1)

ALGORITMO III: INTERVALO DE CONFIANZA BOOTSTRAP PARA ESTIMAR LA DIFERENCIA DE DOS ECSIS

Dadas las matrices de datos $Z_1 = (X_{1,1}, \dots, X_{1,p}, Y_{1,1}, \dots, Y_{1,q})$ y $Z_2 = (X_{2,1}, \dots, X_{2,p}, Y_{2,1}, \dots, Y_{2,q})$ de dimensión $n_1 \times (p+q)$ y $n_2 \times (p+q)$
Para $i = 1, \dots, B$
{
Selecciona muestras bootstrap Z_{i1}^* y Z_{i2}^* de cada matriz Z_1 y Z_2
Ajusta el modelo ECSI a las muestras bootstrap Z_{i1}^* y Zi_2^*
Calcula la diferencia $D_i(Z_{i1}^*, Z_{i2}^*) = \mathrm{ECSI}(Z_{i2}^*) - \mathrm{ECSI}(Z_{i1}^*)$
}
Calcula la distribución bootstrap aproximada

$$\hat{G}_D(x) = \frac{1}{B}\sum_{i=1}^{B} 1_{(-\infty, D_i(Z_{i1}^*, Z_{i2}^*)]}(x)$$

PRUEBAS DE HIPÓTESIS DE IGUALDAD DE k ECSIS

Probar la hipótesis nula de igualdad de k ECSIs

$$H_0\text{: ECSI}_1 = \text{ECSI}_2 = \cdots = \text{ECSI}_k$$

Contra la alterna de que al menos un par es diferente.

No idea de cómo hacerlo…

RANKING DE k ECSIS

Supongamos que se tienen k ECSIs calculados de k muestras independientes, ahora es de interés construir un ranking de estos:

$$\mathrm{ECSI}_{(1)} \leq \mathrm{ECSI}_{(2)} \leq \cdots \leq \mathrm{ECSI}_{(k)}.$$

No idea de cómo hacerlo…

DOS ARTÍCULOS

- CHUNG, E.Y., and ROMANO, J.P. (2013). Exact and asymptotically robust permutation test. **The Annals of Statistics** 41 (2): 484-507.

- HALL, P. and MILLER, H. (2009): Using the bootstrap to quantify the authority of an empirical ranking. **The Annals of Statistics**, 37, 6B, 3929-3959.

PRUEBA DE PERMUTACIONES PARA PROBAR LA HIPÓTESIS DE IGUALDAD DE k PARÁMETROS

Tenemos k muestras iid de tamaño n_j de las distribuciones P_j. Sea θ el parámetro de interés (escalar). La hipótesis nula que se desea probar es

$$H_0\colon\ \theta(P_1) = \cdots \theta(P_k)$$

vs

$$H_1\colon\ \theta(P_i) \neq \theta\left(P_j\right) \text{ para algún } i \neq j$$

Bajo el supuesto de $P_1 = \cdots P_k$.

ESTADÍSTICO DE PRUEBA

Lema Chung y Romano (2013): $0 < \sigma_i^2 = E_{P_i} f_{P_i}^2(X_{i,j}) < \infty, n_i \to \infty,$
$n_i / \sum_i n_i \to p_i > 0$

$$T_{n0} = \sum_{i=1}^{k} \frac{n_i}{\sigma_i^2} \left[\hat{\theta}_{n,i} - \frac{\sum_{i=1}^{k} n_i \hat{\theta}_{n,i} / \sigma_i^2}{\sum_{i=1}^{k} n_i / \sigma_i^2} \right]^2$$

$$T_{n1} = \sum_{i=1}^{k} \frac{n_i}{\hat{\sigma}_i^2} \left[\hat{\theta}_{n,i} - \frac{\sum_{i=1}^{k} n_i \hat{\theta}_{n,i} / \hat{\sigma}_i^2}{\sum_{i=1}^{k} n_i / \hat{\sigma}_i^2} \right]^2$$

Entonces, bajo H_0, T_{n0} y T_{n1}convergen en distribución a una ji-cuadrada con $k - 1$ grados de libertad.

PRUEBA DE PERMUTACIONES PARA PROBAR LA HIPÓTESIS DE IGUALDAD DE k ECSIS

Para probar la hipótesis nula de igualdad de k ECSIs

$$H_0\colon \mathrm{ECSI}_1 = \mathrm{ECSI}_2 = \cdots = \mathrm{ECSI}_k$$

usamos el estadístico de Chung y Romano, el cual para el ECSI toma la forma

$$T = \sum_{j=1}^{k} \frac{n_j}{\hat{\sigma}_j^2} \left[\widehat{\mathrm{ECSI}}_j - \frac{\sum_{j=1}^{k} n_j \widehat{\mathrm{ECSI}}_j / \hat{\sigma}_j^2}{\sum_{j=1}^{k} n_j / \hat{\sigma}_j^2} \right]^2$$

donde $\widehat{\mathrm{ECSI}}_{ij}$ es el ECSI estimado en Z_j en la iteración i y además

$$\hat{\sigma}_j^2 = \frac{1}{M-1} \sum_{i=1}^{M} (\widehat{\mathrm{ECSI}}_{ij} - \overline{\mathrm{ECSI}}_j)^2,$$

$$\overline{\mathrm{ECSI}}_j = \frac{1}{M} \sum_{i=1}^{M} \widehat{\mathrm{ECSI}}_{ij}\,.$$

¿POR QUÉ FUNCIONA ESTE ESTADÍSTICO DE PRUEBA?

Si demostramos que existen estimadores $\widehat{\mathrm{ECSI}}(Z_j)$ del parámetro de interés $\mathrm{ECSI}(Z_j)$, que satisfagan la siguiente aproximación lineal

$$n_j^{1/2}\left[\widehat{\mathrm{ECSI}}(Z_j) - \mathrm{ECSI}(Z_j)\right] = \frac{1}{\sqrt{n_j}}\sum_{i=1}^{n_j} f_{F_j}(\mathbf{z}_{ij}) + o_p(1),$$

donde f_{F_j} es una función que caracteriza a la aproximación lineal, $\mathbf{z}_{ij}$ es la hilera i de Z_j, y $n_j \to \infty$ con $n_j/N \to p_j$ para $j = 1, \dots, k$, y $N = \sum_{j=1}^{k} n_j$, y además $\hat{\sigma}_j^2$ es un estimador consiste de la varianza, entonces el estadístico de prueba T converge en distribución a una ji-cuadrada con $k-1$ grados de libertad.

¿POR QUÉ FUNCIONA ESTE ESTADÍSTICO DE PRUEBA?

La aproximación lineal se cumple trivialmente para el ECSI ya que

$$n_j^{1/2}\left[\widehat{\mathrm{ECSI}}(Z_j) - \mathrm{ECSI}(Z_j)\right] = \frac{1}{\sqrt{n_j}}\sum_{i=1}^{n_j} f_{F_j}(\mathbf{z}_{ij})$$

donde $f_{F_j}(\mathbf{z}_{ij}) = \hat{s}_i - \mathrm{E(S)}/\{\max(\mathrm{S}) - \min(\mathrm{S})\}$, y $\hat{s}_i = \sum_{l=1}^{q} w_j y_{ilj}$, $(i = 1, \dots, n, j = 1, \dots, k)$, y por lo tanto la distribución de permutaciones de T que usamos converge a una ji-cuadrada con $k-1$ grados de libertad.

ALGORITMO IV: DE PRUEBA DE PERMUTACIONES PARA COMPARAR k ECSIS

```
Dadas las matrices de datos  Z_j = (X_{j,1}, …, X_{j,p}, Y_{j,1}, …, Y_{j,q}) de dimensión
n_j × (p + q),  j = 1, …, k
Ajusta el modelo ECSI a cada Z_j y calcula ECSI_j,  j = 1, …, k
Calcula el estadístico de prueba T̂ = T̂(Z_1, …, Z_k)
Sea p = 0
Para i = 1, …, M
{
  Determina las matrices Z'_{ij},  j = 1, …, k de permutaciones
 Calcula el estadístico de prueba T*_i = T*_i(Z'_{ij}, …, Z'_{ik})
  Si T*_i > T̂, p = p + 1
}
p-valor = p/(M+1)
```

RANKING BOOTSTRAP DE k ECSIS

Supongamos que se tienen k ECSIs calculados de k muestras independientes, ahora es de interés construir un ranking de estos:

$$\mathrm{ECSI}_{(1)} \leq \mathrm{ECSI}_{(2)} \leq \cdots \leq \mathrm{ECSI}_{(k)}.$$

Este ranking lo construimos adaptando el siguiente procedimiento bootstrap propuesto por Hall y Miller (2009). Sean $F_1, F_2, \dots, F_k$ k distribuciones con sus respectivos parámetros funcionales de interés $\theta_1 = \theta(F_1), \theta_2 = \theta(F_2), \dots, \theta_k = \theta(F_k)$. Consideremos a estos parámetros ordenados $\theta_{(1)} \leq \theta_{(2)} \leq \cdots \leq \theta_{(k)}$. Asignamos a $\theta_{(j)}$ el rango r_j. Formalmente los rangos $r_1, r_2, \dots, r_k$ se definen por

$$r_j = \sum_{i=1}^{k} 1_{\{\theta_i \leq \theta_j\}}(\theta_i) = 1 + \sum_{i \neq j} 1_{\{\theta_i \leq \theta_j\}}(\theta_i), \qquad j = 1, \dots, k.$$

RANKING BOOTSTRAP DE k ECSIS

Los rangos r_j se estiman a partir de estimaciones de los parámetros $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$

$$\hat{r}_j = \sum_{i=1}^{k} 1_{\{\hat{\theta}_i \leq \hat{\theta}_j\}}(\hat{\theta}_i) = 1 + \sum_{i \neq j} 1_{\{\hat{\theta}_i \leq \hat{\theta}_j\}}(\hat{\theta}_i), \qquad j = 1, \dots, k.$$

Sea $y_j = (y_{j,1}, \dots, y_{j,n_j})$ una muestra aleatoria de F_j y sea $F_{j,n}$ la función de distribución empírica de esta muestra. De cada una de las funciones de distribución empíricas $F_{1,n}, F_{2,n}, \dots, F_{k,n}$ se selecciona una respectiva muestra bootstrap $Y_j^* = (Y_{j,1}^*, \dots, Y_{j,n_j}^*)$ y se calculan $\theta_j^* = \theta(Y_j^*)$; con estos valores se calculan los rangos bootstrap

$$r_j^* = \sum_{i=1}^{k} 1_{\{\theta_i^* \leq \theta_j^*\}}(\theta_i^*) = 1 + \sum_{i \neq j} 1_{\{\theta_i^* \leq \theta_j^*\}}(\theta_i^*), \qquad j = 1, \dots, k.$$

RANKING BOOTSTRAP DE k ECSIS

Este proceso se itera B veces para obtener un conjunto de B vectores de rangos bootstrap $r_b^* = (r_{b,1}^*, r_{b,2}^*, \dots, r_{b,k}^*), (b = 1, \dots, B)$

Rangos bootstrap

$F_{1,n}$	$F_{2,n}$	$\cdots$	$F_{k,n}$
$r_{1,1}^*$	$r_{1,2}^*$	$\cdots$	$r_{1,k}^*$
$r_{2,1}^*$	$r_{2,2}^*$	$\cdots$	$r_{2,k}^*$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$r_{B,1}^*$	$r_{B,2}^*$	$\cdots$	$r_{B,k}^*$

RANKING BOOTSTRAP DE k ECSIS

Con estos rangos se calculan intervalos confianza con el método percentil

$$\left[\hat{r}_j^{*(\alpha/2)}, \hat{r}_j^{*(1-\alpha/2)}\right], \qquad j = 1,2,\dots,k,$$

donde $\hat{r}_j^{*(\alpha/2)}$ y $\hat{r}_j^{*(1-\alpha/2)}$ son los cuantiles empíricos $\alpha/2$ y $1-\alpha/2$ de la distribución bootstrap de los rangos (columnas de la Tabla 1), es decir, $F_{j,n}(\hat{r}_j^{*(\alpha/2)}) = \alpha/2$ y $F_{j,n}(\hat{r}_j^{*(1-\alpha/2)}) = 1-\alpha/2$.

ALGORITMO V: RANKING BOOTSTRAP DE k ECSIS

Dadas las matrices de datos $Z_j = \left(X_{j,1}, \dots, X_{j,p}, Y_{j,1}, \dots, Y_{j,q}\right)$ de dimensión $n_j \times (p+q)$, $j = 1, \dots, k$
Para i=1,…,B
{
Selecciona una muestra bootstrap $Z_j^* = (X_{j,1}^*, \dots, X_{j,p}^*, Y_{j,1}^*, \dots, Y_{j,q}^*)$ de cada una de las k matrices Z_j
Calcula los k ECSIs bootstrap $\mathrm{ECSI}_1^* = \mathrm{ECSI}(Z_1^*)$, $\mathrm{ECSI}_2^* = \mathrm{ECSI}(Z_2^*), \dots,$ $\mathrm{ECSI}_k^* = \mathrm{ECSI}(Z_k^*)$
Construye el vector de rangos bootstrap $r^* = (r_1^*, r_2^*, \dots, r_k^*)$, donde

$$r_j^* = 1 + \sum\nolimits_{i \neq j} 1_{\left\{\mathrm{ECSI}_i^* \leq \mathrm{ECSI}_j^*\right\}}(\mathrm{ECSI}_i^*), \qquad j = 1, \dots, k.$$

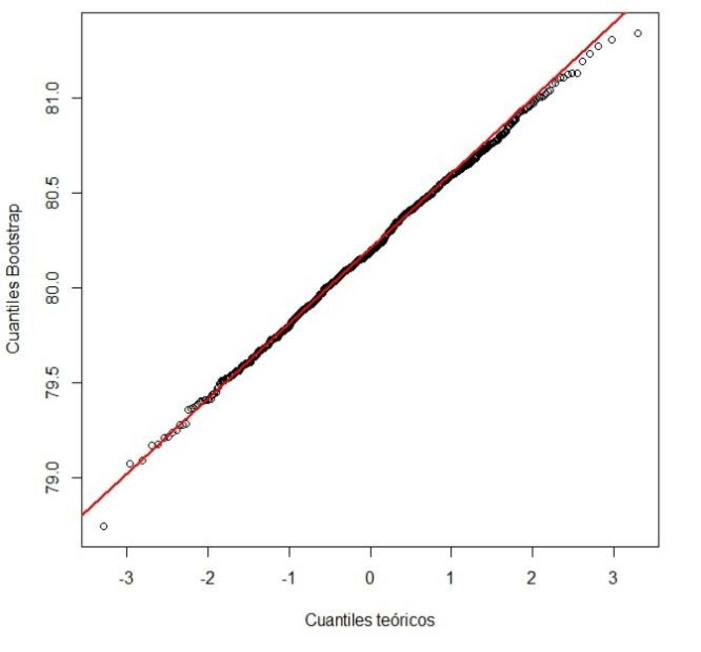
}
Calcula el promedio y la desviación estándar de los rangos para cada muestra j

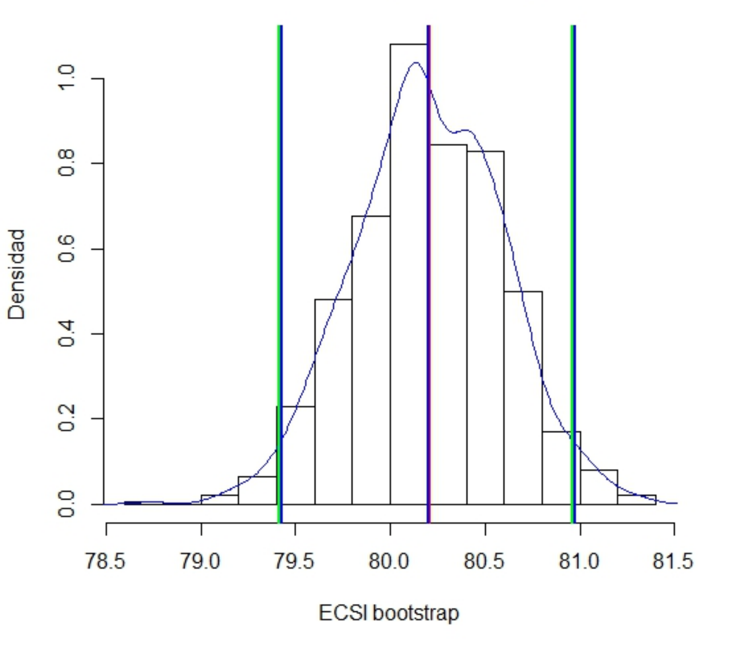
$$\bar{r}_j^* = \frac{1}{B}\sum\nolimits_b^B r_{b,j}^*, \qquad s_j^* = \sqrt{\frac{1}{B-1}\sum\nolimits_b^B (r_{b,j}^* - \bar{r}_j^*)^2}$$

Construye el intervalo de confianza bootstrap para cada rango $\bar{r}_j^* \pm 2s_j^*$

RESULTADOS DEL ESTUDIO DE SATISFACCIÓN EN LA UV

DISTRIBUCIÓN BOOTSTRAP DEL ECSI PARA LA UV





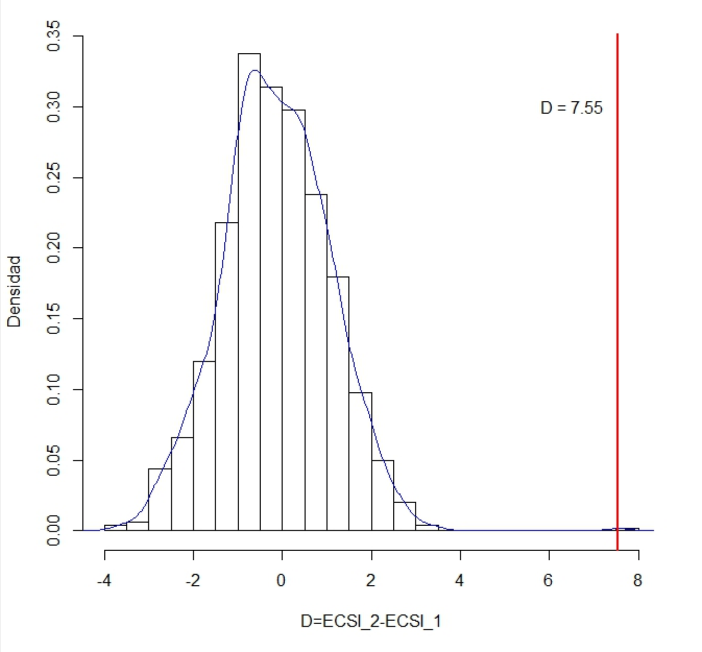
COMPARACIÓN DE DOS ECSI

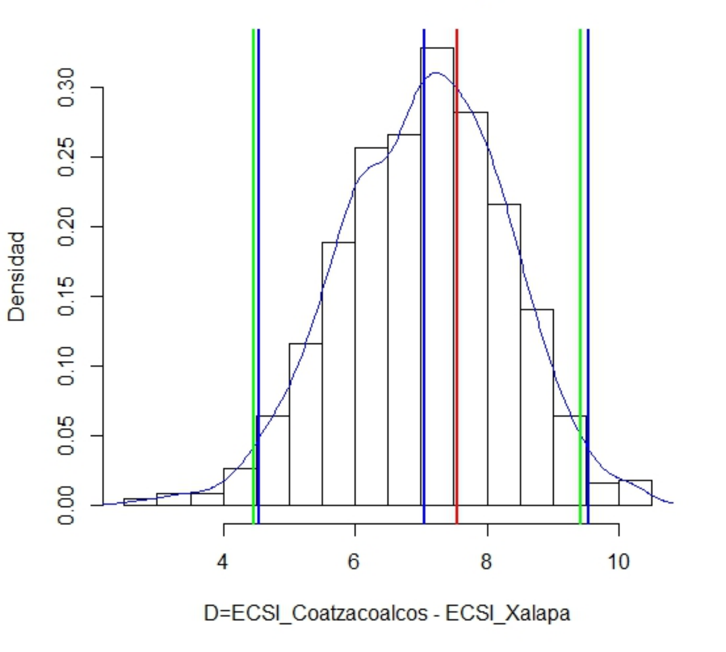
Los niveles más altos de satisfacción estudiantil se observan en Coatzacoalcos-Minatitlán, $\mathrm{ECSI}_{\mathrm{C-M}} = 84.71$, mientras que los más bajos están en Xalapa, $\mathrm{ECSI}_{\mathrm{Xal}} = 77.16$. Para probar si esta diferencia entre los ECSI es significativa, probamos la siguiente hipótesis

$$H_o\colon \mathrm{ECSI}_{\mathrm{Xal}} = \mathrm{ECSI}_{\mathrm{C-M}}$$

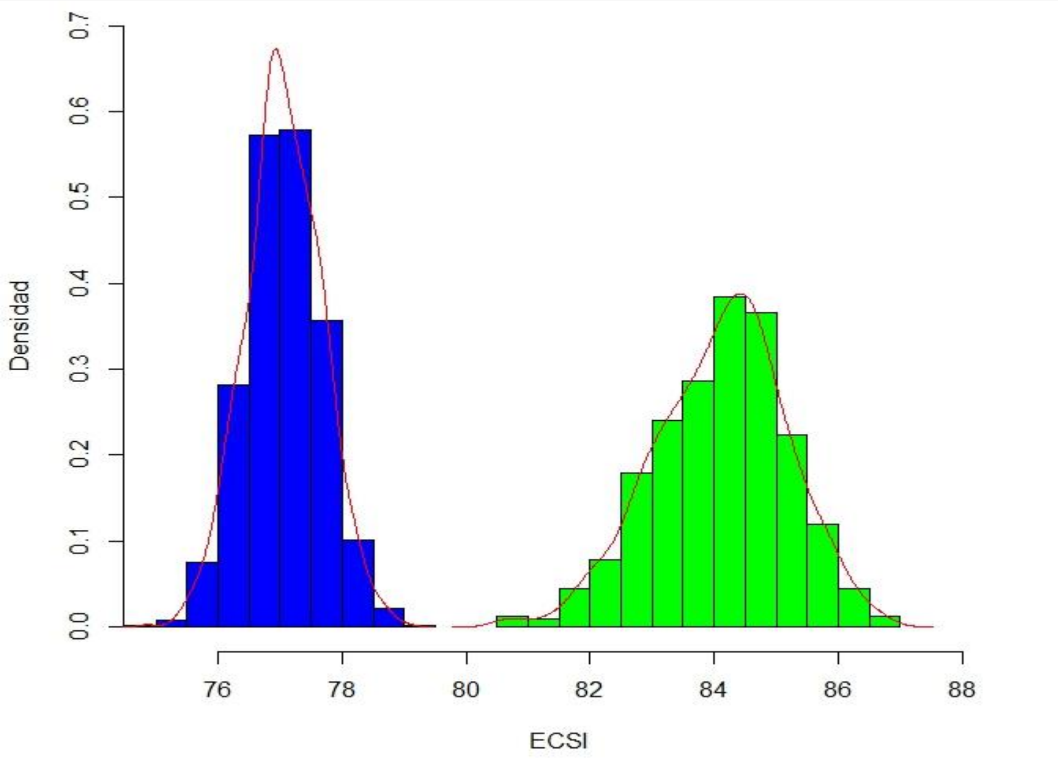
Los resultados de aplicar a prueba de permutaciones para dos ECSIs son $D = \mathrm{ECSI}_{\mathrm{C-M}} - \mathrm{ECSI}_{\mathrm{Xal}} = 7.55$ y en la Figura 8 se muestra la distribución de permutaciones de D basada en 1000 permutaciones. El p-valor es 1/1001.

COMPARACIÓN DE DOS ECSI Y ESTIMACIÓN DE SU DIFERENCIA





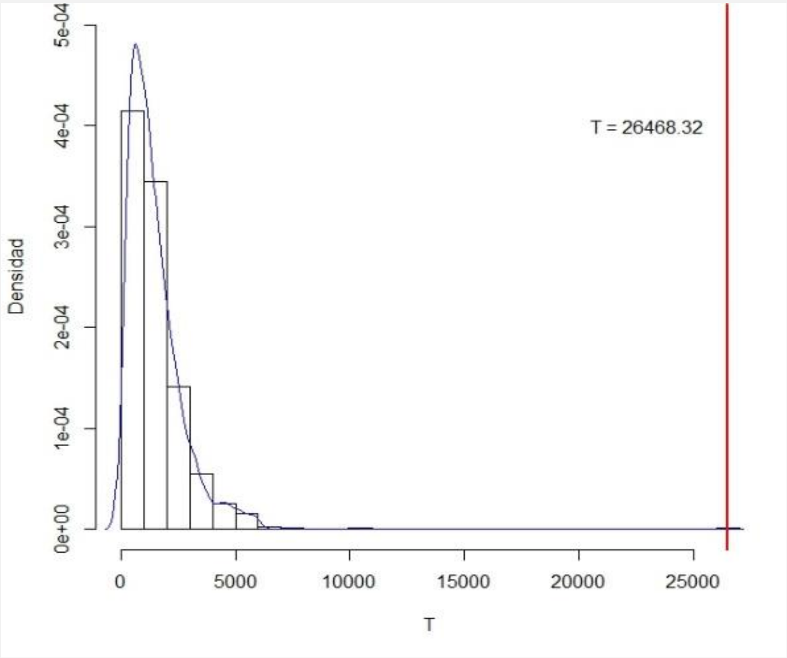
DISTRIBUCIONES BOOTSTRAP DEL ECSI DE XALAPA (AZUL) Y COATZACOALCOS-MINATITLÁN (VERDE)



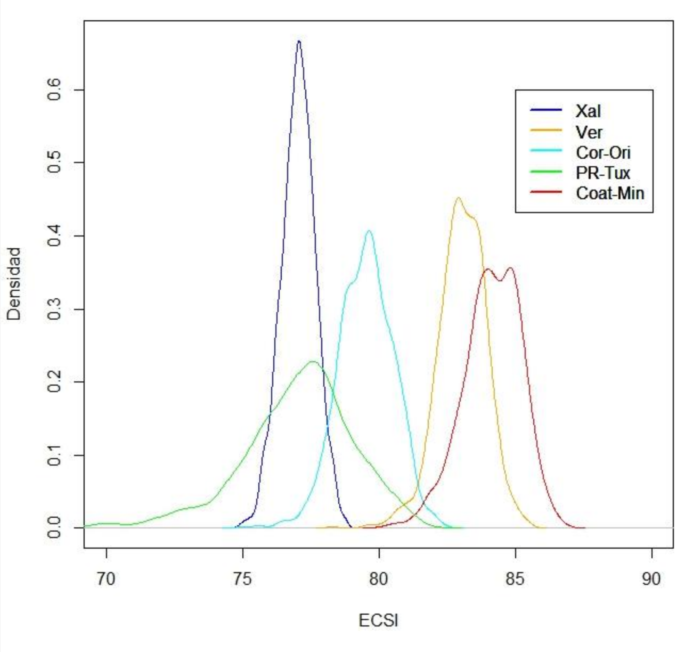
COMPARACIÓN DE LOS ECSI DE LAS REGIONES

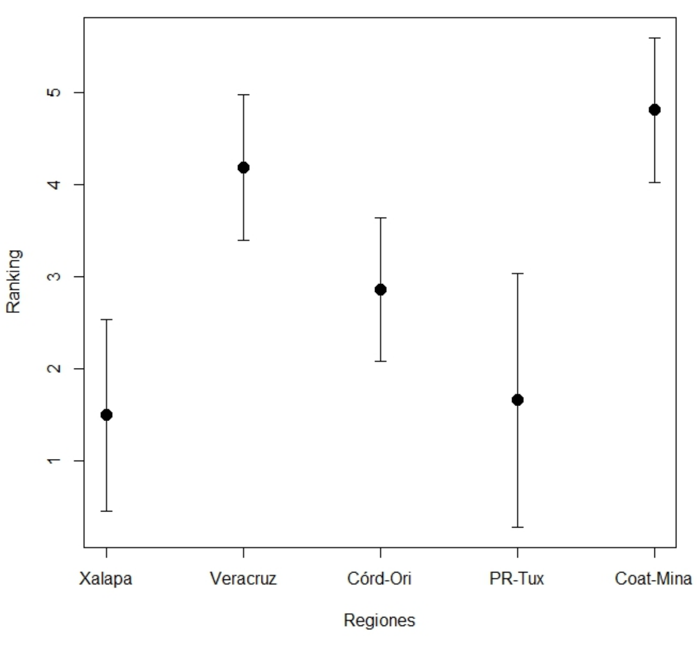
Deseamos probar la hipótesis

$$H_o\text{: } \mathrm{ECSI}_{\mathrm{Xal}} = \mathrm{ECSI}_{\mathrm{Ver}} = \mathrm{ECSI}_{\mathrm{C-O}} = \mathrm{ECSI}_{\mathrm{P-T}} = \mathrm{ECSI}_{\mathrm{C-M}}$$



RANKING DE LOS ECSI DE LAS REGIONES





COMENTARIOS FINALES

- Resultados publicados en: VIÑÁN ANDINO, A.B. y JUÁREZ CERRILLO, S.F. (2018). Inferencia estadística para el índice europeo de satisfacción de clientes basada en bootstrap y pruebas de permutación. **Revista Investigación Operacional.**
- La propuesta es directa y fácil de implementar con R.
- Se llena un vacío en la metodología estadística para hacer estimación por intervalo para el ECSI, para comparar dos o más ECSI y para construir rankings de ECSIs
- El bootstrap y las pruebas de permutaciones proporcionan alternativas para realizar inferencia estadística cuando otros enfoques, por ejemplo, aquellos basados en teoría distribucional asintótica, fallan.
- La metodología propuesta es aplicable directamente a otros índices de satisfacción de clientes similares al ECSI, por ejemplo, al American Customer Satisfaction Index (ACSI) de Fornell et al. (1996).

REFERENCIAS

- [1] ÁLVAREZ, R. y VERNAZZA, E. (2013): Aplicación de los modelos de ecuaciones estructurales para el estudio de la satisfacción estudiantil en los cursos superiores de FCCEEyA. *Serie DT (13/02)*, **Instituto de Estadística, Universidad de la República, Facultad de Ciencias Económicas y Administración, Uruguay.**
- [2] ALVES, H., y RAPOSO, M. (2004): La medición de la satisfacción en la enseñanza universitaria: El ejemplo de la universidad de da Beira Interior. **Revista Internacional de Marketing Público y No Lucrativo**, 1 (1): 73-88.
- [3] BALZANO, S., and TRINCHERA, L. (2011): Structural equation models and student evaluation of teaching: A PLS path modeling study. Attanasio, M., and Capursi, V., (eds.) **Statistical Methods for the Evaluation of University Systems, Contributions to Statistics.** Berlin Heidelberg: Springer-Verlag, 55-66.
- [4] BROWN, R.M., and MAZZAROL, T.W. (2009): The importance of institutional image to student satisfaction and loyalty within higher education. **Higher Education**, 58: 81-95.
- [5] CHIANDOTTO, B., BINI, M., and BERTACCINI, B. (2007): Quality assessment of the educational university processes: An application of the ECSI model. In Fabbris, L. (ed.), **Effectiveness of University Education in Italy. Employability, Competences, Human Capital** (p. 43-54). Physica Verlag, Padua.

REFERENCIAS

[6] CHITTY, B., and SOUTAR, G.N. (2004): Is the european customer satisfaction index model applicable to tertiary education? **Proceedings of the Australian and New Zealand Marketing Academy Conference**, Wellington, New Zealand.

[7] CHUNG, E.Y., and ROMANO, J.P. (2013). Exact and asymptotically robust permutation test. **The Annals of Statistics** 41 (2): 484-507.

[8] DUARTE, P.O., RAPOSO, M.B., and ALVES, H.B. (2012): Using a satisfaction index to compare students' satisfaction during and after higher education service consumption. **Tertiary Education and Management**, 18 (1), 17-40.

[9] DUQUE ZULUAGA, L.C. (2003): La satisfacción del usuario del servicio "Formación Educativa Universitaria". Zorrilla Torras R., and Gómez de Cadiñanos, S.S. (eds.). **Economía de la Educación AEDE XII**, Madrid, España.

[10] ECSI TECHNICAL COMMITTEE (1998): *European customer satisfaction index: foundation and structure for harmonized national pilot projects*. **Report prepared by ECSI Technical Committee.** ECSI document no. 005 ed. 1, 20-11-98.

REFERENCIAS

- [11] FORNELL C., JOHNSON, M. D., ANDERSON, E.W., CHA, J., and BRYANT, B. E. (1996): The american customer satisfaction index: Nature, purpose and findings. **Journal of Marketing**, 60 (4): 7-18.
- [12] HALL, P. and MILLER, H. (2009): Using the bootstrap to quantify the authority of an empirical ranking. **The Annals of Statistics**, 37, 6B, 3929-3959.
- [13] MAMMEN, E. (1992): **When Does Bootstrap Work?** Springer: New York.
- [14] MARTENSEN, A., GRØNHOLDT, L., ESKILDSEN, J.K., and KRISTENSEN, K. (2000): Measuring student oriented quality in higher education: Application of the ECSI methodology. **Sinergie Rapporti di Ricerca**, 9, 371-383.
- [15] Sánchez, G. (2013): **PLS Path Modeling with R**. Trowchez Editions. Berkeley, 2013. http://www.gastonsanchez.com/PLS Path Modeling with R.pdf

REFERENCIAS

- [16] SOSA GALINDO, I., JUÁREZ CERRILLO, S.F., y CRUZ KURI, L. (2014): La universidad Veracruzana desde la satisfacción de sus estudiantes. Hernández González, S.; Núñez-Antonio, G., y Méndez Gómez Humarán, I., (Eds.), **Aportaciones a la Estadística de los XXVII y XXVIII Foros Nacionales de Estadística** (pp. 87-91). INEGI, Aguascalientes, Ags., México
- [17] UNIVERSIDAD VERACRUZANA, DIRECCIÓN GENERAL DEL ÁREA ACADÉMICA ECONÓMICO ADMINISTRATIVA. (2016): **Encuesta de satisfacción de los estudiantes del Área Académica Económico Administrativa, referente a los servicios académicos y administrativos que reciben. Reporte de análisis y resultados estadísticos.**
- [18] VIÑÁN ANDINO, A.B. (2015): **Medición de la Satisfacción de los Estudiantes con la Calidad Educativa: El Caso de la Escuela Superior Politécnica de Chimborazo**. Tesis de la Especialización en Métodos Estadísticos, Facultad de Estadística e Informática, Universidad Veracruzana.
- [19] ZHANG, L., HAN, Z., and GAO, Q. (2008): Empirical study on the student satisfaction index in higher education. **International Journal of Business and Management**, 3 (9): 46-51.

FIN

Gracias por su atención