

Análisis de texto

para identificar las modificaciones de profesores universitarios en la enseñanza de la estadística debido al COVID-19

Mtro. José Guadalupe Rivera-Pérez¹,
Dra. Ana Luisa Gómez-Blancarte²

¹ Universidad Autónoma de Tamaulipas. Facultad de Enfermería Tampico

² Instituto Politécnico Nacional. CDMX

Stata Conference. 25 oct 2023.



UAT



Contenido

1. Introducción.
2. Fundamentos básicos.
3. Comando Coin.
4. Investigación.
5. Resultados.
6. Conclusiones.

Introducción

Introducción

- El objetivo de un análisis de coincidencias de texto es detectar caracteres, objetos, atributos, características o eventos que tienden a ocurrir en ciertos escenarios y encontrar la relación entre ellos.

Introducción

- El objetivo de un análisis de coincidencias de texto es detectar caracteres, objetos, atributos, características o eventos que tienden a ocurrir en ciertos escenarios y encontrar la relación entre ellos.
- Para esta investigación se utilizó un análisis de coincidencias de texto para identificar las modificaciones de profesores universitarios en la enseñanza de la estadística debido al COVID-19 en instituciones de nivel superior de México.

Introducción

- El objetivo de un análisis de coincidencias de texto es detectar caracteres, objetos, atributos, características o eventos que tienden a ocurrir en ciertos escenarios y encontrar la relación entre ellos.
- Para esta investigación se utilizó un análisis de coincidencias de texto para identificar las modificaciones de profesores universitarios en la enseñanza de la estadística debido al COVID-19 en instituciones de nivel superior de México.
- Se utilizó la versión 17 del programa [Stata](#), el programa ado [txttool](#) y los programas [PRECOIN](#) y [COIN](#).

Introducción

- El objetivo de un análisis de coincidencias de texto es detectar caracteres, objetos, atributos, características o eventos que tienden a ocurrir en ciertos escenarios y encontrar la relación entre ellos.
- Para esta investigación se utilizó un análisis de coincidencias de texto para identificar las modificaciones de profesores universitarios en la enseñanza de la estadística debido al COVID-19 en instituciones de nivel superior de México.
- Se utilizó la versión 17 del programa [Stata](#), el programa ado [txttool](#) y los programas [PRECOIN](#) y [COIN](#).
- El comando [Coin](#) permite estudiar estructuras de datos y es una combinación de análisis estadístico y redes sociales.

Introducción

- El objetivo de un análisis de coincidencias de texto es detectar caracteres, objetos, atributos, características o eventos que tienden a ocurrir en ciertos escenarios y encontrar la relación entre ellos.
- Para esta investigación se utilizó un análisis de coincidencias de texto para identificar las modificaciones de profesores universitarios en la enseñanza de la estadística debido al COVID-19 en instituciones de nivel superior de México.
- Se utilizó la versión 17 del programa [Stata](#), el programa ado [txttool](#) y los programas [PRECOIN](#) y [COIN](#).
- El comando [Coin](#) permite estudiar estructuras de datos y es una combinación de análisis estadístico y redes sociales.
- Este comando puede ser de utilidad para realizar análisis exploratorio de cuestionarios, estudio de redes de texto o contenidos de bases de datos, entre otras aplicaciones.

Fundamentos básicos

Definición

La estructura de datos llamada co-ocurrencia consta de dos conjuntos finitos de elementos,

$$X = \{x_1, \dots, x_N\}$$

y

$$Y = \{y_1, \dots, y_M\}$$

donde las observaciones elementales son los pares de elementos formados por los conjuntos X e Y junto con su correspondiente frecuencia.

Análisis de Coincidencias

Matriz de Co-ocurrencias

Una matriz de co-ocurrencias es una matriz que consta de un número determinado de filas como expresiones de textos disponibles y de columnas como frecuencias de palabras a analizar. Los elementos de esta matriz serán la frecuencia de cada palabra en cada texto.

Cuadro: 1. Matriz de co-ocurrencias

	Palabra 1	Palabra 2	Palabra 3	Palabra 4
Texto 1	2	4	0	2
Texto 2	4	2	1	0
Texto 3	5	4	3	6
Texto 4	2	2	0	7
Texto 5	1	4	5	3

Análisis de incidencias

Matriz de incidencias

Una matriz de co-ocurrencias puede convertirse en una matriz de incidencias al dividir el texto en pequeñas unidades o convertir las frecuencias en valores de ausencia o presencia del texto.

Cuadro: 2. Matriz de incidencias

	Palabra 1	Palabra 2	Palabra 3	Palabra 4
Texto 1	1	1	0	1
Texto 2	1	1	1	0
Texto 3	1	1	1	1
Texto 4	1	1	0	1
Texto 5	1	1	1	1

Análisis de coincidencias de texto

Fundamentos

- El objetivo de un análisis de coincidencias de texto es detectar que caracteres, objetos, atributos, características o eventos tienden a ocurrir en ciertos límites.

Análisis de coincidencias de texto

Fundamentos

- El objetivo de un análisis de coincidencias de texto es detectar que caracteres, objetos, atributos, características o eventos tienden a ocurrir en ciertos límites.
- Estos límites son llamados escenarios (*/*) y se consideran unidades de análisis.

Análisis de coincidencias de texto

Fundamentos

- El objetivo de un análisis de coincidencias de texto es detectar que caracteres, objetos, atributos, características o eventos tienden a ocurrir en ciertos límites.
- Estos límites son llamados escenarios (I) y se consideran unidades de análisis.
- Cada escenario i puede ocurrir una serie de eventos J . Los eventos pueden ser dependientes o independientes de otro y se representan con variables J dícotómicas (0 o 1) X_j .

Análisis de coincidencias de texto

Fundamentos

- El objetivo de un análisis de coincidencias de texto es detectar que caracteres, objetos, atributos, características o eventos tienden a ocurrir en ciertos límites.
- Estos límites son llamados escenarios (I) y se consideran unidades de análisis.
- Cada escenario i puede ocurrir una serie de eventos J . Los eventos pueden ser dependientes o independientes de otro y se representan con variables J dicitómicas (0 o 1) X_j .
- El objetivo de este análisis es encontrar el subconjunto de pares de categoría que no son independientes en el conjunto de eventos posibles $J(J-1)/2$ de cada escenario. (Escobar, 2015)

Fundamentos

- El análisis se complica cuando crece el número de escenarios en un solo conjunto de datos (n).

Análisis de coincidencias de texto

Fundamentos

- El análisis se complica cuando crece el número de escenarios en un solo conjunto de datos (n).
- Se forma una matriz de dimensiones $n \times J$ (Matriz de incidencias I).

Análisis de coincidencias de texto

Fundamentos

- El análisis se complica cuando crece el número de escenarios en un solo conjunto de datos (n).
- Se forma una matriz de dimensiones $n \times J$ (Matriz de incidencias I).
- Se tendrá una matriz correspondiente de frecuencias F con dimensiones $J \times J$ obtenida por las medias de los productos $I'I$.

Análisis de coincidencias de texto

Fundamentos

- El análisis se complica cuando crece el número de escenarios en un solo conjunto de datos (n).
- Se forma una matriz de dimensiones $n \times J$ (Matriz de incidencias I).
- Se tendrá una matriz correspondiente de frecuencias F con dimensiones $J \times J$ obtenida por las medias de los productos $I'I$.
- Los elementos de esta matriz son las frecuencias (f_{ij}) y las frecuencias bivariadas (f_{jk}) de los eventos en un conjunto de escenarios. (Escobar, 2015)

Comando Coin

Características del comando

Comando Coin

- Puede ser de utilidad para realizar el análisis de coincidencias.

(Escobar, 2015)

Características del comando

Comando Coin

- Puede ser de utilidad para realizar el análisis de coincidencias.
- Facilita un amplia gama de analisis estadísticos y genera gráficas para un conjunto de variables dicotómicas.

(Escobar, 2015)

Características del comando

Comando Coin

- Puede ser de utilidad para realizar el análisis de coincidencias.
- Facilita un amplia gama de analisis estadísticos y genera gráficas para un conjunto de variables dicotómicas.
- Permite el cálculo de diversos coeficientes de correlación y medidas de distancia, con el objetivo de descubrir patrones observables en una serie de eventos.

(Escobar, 2015)

Características del comando

Comando Coin

- Puede ser de utilidad para realizar el análisis de coincidencias.
- Facilita un amplia gama de analisis estadísticos y genera gráficas para un conjunto de variables dicotómicas.
- Permite el cálculo de diversos coeficientes de correlación y medidas de distancia, con el objetivo de descubrir patrones observables en una serie de eventos.
- Se pueden generar gráficas de barras, de coincidencias, valores residuales, dendogramas,escalas multidimensionales y mas.

(Escobar, 2015)

Características del comando

Comando Coin

- Puede ser de utilidad para realizar el análisis de coincidencias.
- Facilita un amplia gama de analisis estadísticos y genera gráficas para un conjunto de variables dicotómicas.
- Permite el cálculo de diversos coeficientes de correlación y medidas de distancia, con el objetivo de descubrir patrones observables en una serie de eventos.
- Se pueden generar gráficas de barras, de coincidencias, valores residuales, dendogramas,escalas multidimensionales y mas.
- La sintaxis de Coin es
`coin varlist [if] [in] [weight] [using filename] [,options]`

(Escobar, 2015)

Investigación

Metodología

Se elaboró un cuestionario basado en las ideas teóricas de las 6 recomendaciones **GAISE** de la **American Statistical Association**:



(GAISE College Report Revision Committee, 2016)

Metodología

El cuestionario se diseñó de la siguiente manera:



Los resultados presentados provienen de una de las dos preguntas abiertas:

A raíz de la contingencia sanitaria del COVID-19, ¿cómo ha modificado su enseñanza?

Aplicación de la encuesta



Procedimiento

Análisis de coincidencias de texto

Se utilizó el análisis de coincidencias para analizar co-ocurrencias de palabras clave en un conjunto de respuestas emitida por profesores para identificar los cambios realizados en su enseñanza en tiempos de pandemia.

Procedimiento

Análisis de coincidencias de texto

Se utilizó el análisis de coincidencias para analizar co-ocurrencias de palabras clave en un conjunto de respuestas emitida por profesores para identificar los cambios realizados en su enseñanza en tiempos de pandemia.

Procedimiento

1. Para esta investigación se utilizó la version 17 de [Stata](#).
2. Se utilizó el comando [txttool](#) para eliminar palabras que no aportaban a la investigación y se realizó una limpieza de datos (subword y stopwords).

Procedimiento

Análisis de coincidencias de texto

Se utilizó el análisis de coincidencias para analizar co-ocurrencias de palabras clave en un conjunto de respuestas emitida por profesores para identificar los cambios realizados en su enseñanza en tiempos de pandemia.

Procedimiento

1. Para esta investigación se utilizó la version 17 de [Stata](#).
2. Se utilizó el comando [txttool](#) para eliminar palabras que no aportaban a la investigación y se realizó una limpieza de datos (subword y stopwords).
3. Se convirtieron las respuestas de los participantes en variables binarias (matriz de incidencias) mediante el comando [precoin](#).

Procedimiento

Análisis de coincidencias de texto

Se utilizó el análisis de coincidencias para analizar co-ocurrencias de palabras clave en un conjunto de respuestas emitida por profesores para identificar los cambios realizados en su enseñanza en tiempos de pandemia.

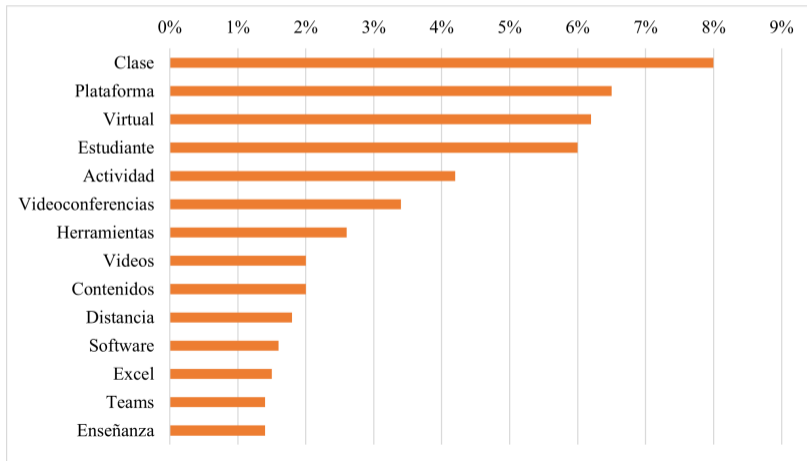
Procedimiento

1. Para esta investigación se utilizó la version 17 de [Stata](#).
2. Se utilizó el comando [txttool](#) para eliminar palabras que no aportaban a la investigación y se realizó una limpieza de datos (subword y stopwords).
3. Se convirtieron las respuestas de los participantes en variables binarias (matriz de incidencias) mediante el comando [precoin](#).
4. Se realiza el análisis de coincidencias a través del comando [coin](#) y se obtiene diferentes reportes para su interpretación.

Resultados

Resultados

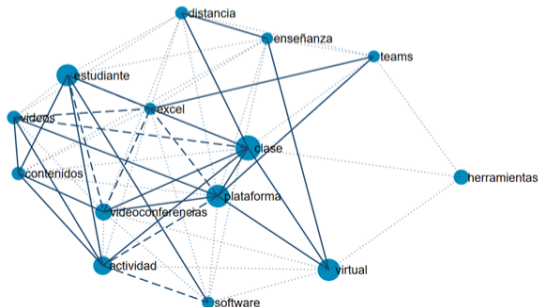
Se obtuvieron 627 respuestas de los 750 participantes. Los resultados obtenidos del análisis de contenido a esta pregunta permitieron conocer los conceptos o palabras más frecuentes que utilizaron los participantes en su respuesta.



Gráfica de escala multidimensional

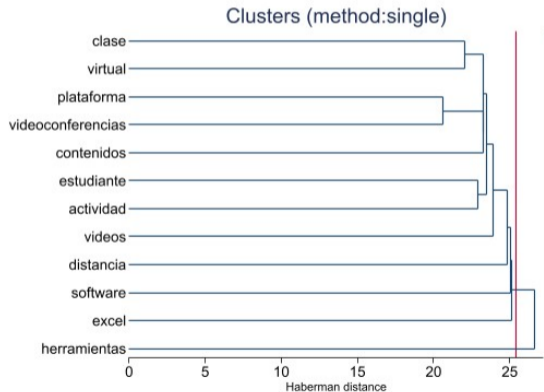
Es una técnica de visualización y reducción de dimensiones. Esta representación tiene dos características:

- Dos eventos correlacionados tienden a estar mas cercanos.
- Los eventos con mayor frecuencia tienden a estar localizados al centro del diagrama.



Dendogramas

Extiende la coincidencia en una serie de eventos agrupando estratos jerárquicos.



Ejemplos de las respuestas de los profesores a la pregunta " **Modificación de la enseñanza a raíz de la pandemia COVID-19**":

- Clases online y diseño de videos interactivos. (Profesor no. 3)
- Usando Microsoft Teams. (Profesor no. 11)
- Imparto clases a través de Microsoft Teams. (Profesor no. 22)
- La clase se imparte de manera sincrónica en TEAMS y asincrónica para las actividades en Eminus (plataforma institucional). (Profesor no. 109)
- La he tenido que realizar a distancia, mediante clases sincrónicas y asincrónicas. (Profesor no. 117)
- A distancia, sesiones virtuales y plataformas educativas. (Profesor no. 206)
- Por medio de recursos tecnológicos y uso de Excel para la elaboración de análisis estadísticos. (Profesor no. 234)

Conclusiones

Resumen

Se analizaron 627 respuestas de modificaciones a la enseñanza de la estadística de 750 profesores participantes, entre las cuales destacaron las siguientes interpretaciones:

- La transición de la modalidad presencial a clases a distancia.

Resumen

Se analizaron 627 respuestas de modificaciones a la enseñanza de la estadística de 750 profesores participantes, entre las cuales destacaron las siguientes interpretaciones:

- La transición de la modalidad presencial a clases a distancia.
- El aprendizaje virtual del estudiante.

Resumen

Se analizaron 627 respuestas de modificaciones a la enseñanza de la estadística de 750 profesores participantes, entre las cuales destacaron las siguientes interpretaciones:

- La transición de la modalidad presencial a clases a distancia.
- El aprendizaje virtual del estudiante.
- El uso de plataformas de aprendizaje.

Resumen

Se analizaron 627 respuestas de modificaciones a la enseñanza de la estadística de 750 profesores participantes, entre las cuales destacaron las siguientes interpretaciones:

- La transición de la modalidad presencial a clases a distancia.
- El aprendizaje virtual del estudiante.
- El uso de plataformas de aprendizaje.
- La impartición de clase por videoconferencia.

Resumen

Se analizaron 627 respuestas de modificaciones a la enseñanza de la estadística de 750 profesores participantes, entre las cuales destacaron las siguientes interpretaciones:

- La transición de la modalidad presencial a clases a distancia.
- El aprendizaje virtual del estudiante.
- El uso de plataformas de aprendizaje.
- La impartición de clase por videoconferencia.
- El uso de contenidos virtuales (e.g., videos, documentos digitales, presentaciones).

Conclusión

Estos cambios coinciden con García (2021) y Chávez-Sánchez et al. (2020), quienes señalaron:

- Múltiples cambios en todos los niveles educativos.

(García, Á. P. , 2021), (Chávez-Sánchez et al., 2020)

Conclusión

Estos cambios coinciden con García (2021) y Chávez-Sánchez et al. (2020), quienes señalaron:

- Múltiples cambios en todos los niveles educativos.
- Fortalecimiento en el uso de las plataformas de aprendizaje.

(García, Á. P. , 2021), (Chávez-Sánchez et al., 2020)

Conclusión

Estos cambios coinciden con García (2021) y Chávez-Sánchez et al. (2020), quienes señalaron:

- Múltiples cambios en todos los niveles educativos.
- Fortalecimiento en el uso de las plataformas de aprendizaje.
- Constante comunicación por videoconferencia.

(García, Á. P. , 2021), (Chávez-Sánchez et al., 2020)

Conclusión

Estos cambios coinciden con García (2021) y Chávez-Sánchez et al. (2020), quienes señalaron:

- Múltiples cambios en todos los niveles educativos.
- Fortalecimiento en el uso de las plataformas de aprendizaje.
- Constante comunicación por videoconferencia.
- El uso de recursos digitales como apoyo a la docencia universitaria.

(García, Á. P. , 2021), (Chávez-Sánchez et al., 2020)

Consideraciones finales

Resumen

- El método de investigación fue realizar un análisis de coincidencias en un conjunto de respuestas de participantes de una encuesta (conurrencia de eventos en un conjunto de escenarios).

Consideraciones finales

Resumen

- El método de investigación fue realizar un análisis de coincidencias en un conjunto de respuestas de participantes de una encuesta (conurrencia de eventos en un conjunto de escenarios).
- El comando **Coin** permitió identificar palabras clave en las respuestas de los cambios en la enseñanza de la estadística en el país.

Consideraciones finales

Resumen


- El método de investigación fue realizar un análisis de coincidencias en un conjunto de respuestas de participantes de una encuesta (conurrencia de eventos en un conjunto de escenarios).
- El comando `Coin` permitió identificar palabras clave en las respuestas de los cambios en la enseñanza de la estadística en el país.
- Existen otros programas que pueden ayudar en el análisis de textos (El comando `Ngram`, por ejemplo).


Consideraciones finales


Resumen


- El método de investigación fue realizar un análisis de coincidencias en un conjunto de respuestas de participantes de una encuesta (conurrencia de eventos en un conjunto de escenarios).
- El comando `Coin` permitió identificar palabras clave en las respuestas de los cambios en la enseñanza de la estadística en el país.
- Existen otros programas que pueden ayudar en el análisis de textos (El comando `Ngram`, por ejemplo).
- Las análisis estadísticos y herramientas gráficas del análisis de coincidencias de texto pueden ser de gran utilidad para futuras investigaciones cuantitativas con algunos ítems cualitativos.


Referencias

 GAISE College Report Revision Committee. (2016). *Guidelines for assessment and instruction in statistics education college report 2016*. American Statistical Association. [https://www.amstat.org/education/guidelines-for-assessment-and-instruction-in-statistics-education-\(gaise\)-reports](https://www.amstat.org/education/guidelines-for-assessment-and-instruction-in-statistics-education-(gaise)-reports).

 Chávez-Sánchez, G., Hernández-García, J., y González-Basilio, S. de J. (2020). Principales retos educativos de los profesores de la unidad académica del norte del estado de Nayarit durante la pandemia de COVID-19. *Revista de Educación Superior 4* (12), 20–30.

 Escobar, M. (2015). Studying coincidences with network analysis and other multivariates tools. *The Stata Journal*, 15(4), 118-1156. <https://doi.org/10.1177/1536867X1501500410>

 García, Á. P. (2021). La enseñanza online post pandemia: Nuevos retos. *HOLOS 37*(2), 1–13.

 StataCorp. (2021). Stata Statistical Software: Release 17. *College Station, TX: StataCorp LLC*.

Muchas gracias



Mtro. José Guadalupe Rivera-Pérez
Universidad Autónoma de Tamaulipas
Facultad de Enfermería Tampico
Tampico, Tamaulipas
jriverap@docentes.uat.edu.mx



Dra. Ana Luisa Gómez-Blancarte
Instituto Politécnico Nacional
CDMX
algomez@ipn.mx