



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

---

# **POSGRADO EN CIENCIAS MATEMÁTICAS**

## **FACULTAD DE CIENCIAS**

**Un análisis de clases latentes de la  
“Primera Encuesta sobre Percepción  
de la Ciencia y Conocimiento de la  
UNAM”**

**QUE PARA OBTENER EL GRADO ACADÉMICO DE**

**MAESTRA EN CIENCIAS**

**P R E S E N T A**

**NADIA SONIA TORRES GONZÁLEZ**

**DIRECTORA DE LA TESIS: Dra. Silvia Ruiz-Velasco Acosta**

**MÉXICO, D.F.**

**Noviembre, 2011**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Índice	
Introducción .....	1
Capítulo 1 .....	2
Percepción Social de la Ciencia .....	2
Encuestas sobre Percepción de la Ciencia en México.....	7
Capítulo 2 .....	12
Primera Encuesta sobre Percepción de la Ciencia y Conocimiento de la UNAM en la Zona Metropolitana de la Ciudad de México.....	12
Descripción del cuestionario .....	13
Diseño muestral .....	14
Capítulo 3 .....	18
Objetivos .....	18
Hipótesis.....	18
Metodología .....	19
Modelos de clases latentes .....	19
El modelo de clases latentes .....	21
El modelo de clases latentes para una variable latente $Y$ con $C$ categorías o clases latentes.	22
Estimación de los parámetros.....	24
Bondad de ajuste.....	28
Tablas de contingencia.....	30
Prueba de independencia $\chi^2$ .....	34
Regresión logística multinomial .....	35
Modelo de regresión logística binaria.....	35
Estimación de los parámetros.....	35
Bondad de Ajuste .....	39
Razón de verosimilitud.....	41
Prueba de Wald.....	42
Interpretación del modelo logístico.....	43
Observaciones .....	46
Capítulo 4 .....	50
Análisis y Resultados .....	50
Variables.....	50

Resultados de frecuencias simples .....	52
Resultados de los modelos.....	55
Modelo I .....	56
Modelo II .....	58
Modelo III .....	62
Modelo IV .....	64
Modelo V .....	67
Modelo VI .....	69
Conclusiones .....	73
Bibliografía .....	76
Anexos .....	78
La función polCA de R.....	78
Modelos de clases latentes .....	79
Modelos de regresión logística multinomial .....	88
Cuestionario .....	104

## Introducción

La ciencia es de vital importancia para el crecimiento económico, el desarrollo sostenible y el mejoramiento de la calidad de vida de una sociedad, por lo que es fundamental que los ciudadanos se involucren con los temas científicos y participen en la formulación de políticas científicas y tecnológicas.

Con el transcurso del tiempo, se ha reflexionado sobre la percepción de la ciencia entre la población, tanto en los países desarrollados como en los países en vías de desarrollo. Se han realizado estudios con el respaldo de instituciones públicas responsables de la elaboración de políticas científicas y tecnológicas, empleando metodologías de encuestas y de análisis de datos. Estos estudios han reflejado el interés de los ciudadanos ante los diferentes aspectos de la ciencia y la tecnología. Además, se ha detectado que una opinión pública informada puede contribuir a aumentar la tolerancia hacia los científicos y sus investigaciones y asegurar la continuidad de los fondos públicos. Un país que valora la ciencia se encuentra en mejores condiciones de competencia tecnológica, las personas involucradas con el conocimiento científico están en mejores condiciones dentro del mundo social y económico.

Asimismo se ha observado que entre mayor es la formación educativa de los individuos, se tiene una actitud más optimista respecto al desarrollo científico y tecnológico, aunque es cierto que observan tanto las fortalezas como las debilidades del mismo.

El gobierno de cada país debe considerar las bondades que ofrece una cultura científica de todos sus habitantes, por lo que se requiere revisar continuamente las políticas en materia de ciencia y tecnología orientadas a elevar la cultura de todos los sectores de la sociedad, basadas en una mayor difusión y divulgación de estos temas, así como otorgar mayores recursos e incentivos a las organizaciones privadas y públicas para involucrarse en el uso, la adquisición y sobre todo, el desarrollo de nuevas tecnologías y nuevos conocimientos científicos.

La Unidad de Investigación Social Aplicada y de Estudios de Opinión (UDESOS) del Instituto de Investigaciones Sociales de la UNAM, en un trabajo conjunto con la Dirección General de Divulgación de la Ciencia (DGDC), realizó la “Primera Encuesta sobre Percepción de la Ciencia y Conocimiento de la UNAM” con el propósito de obtener información con respecto a la percepción que tienen los habitantes de la zona metropolitana de la Ciudad de México, sobre la ciencia, y el conocimiento que tienen de la UNAM.

Con la finalidad de analizar la información obtenida, en este estudio se desarrolla un análisis de clases latentes de la “Primera Encuesta sobre Percepción de la Ciencia y Conocimiento de la UNAM” para clasificar las respuestas de los individuos, así como un modelo logístico multinomial para determinar los perfiles de la población de acuerdo con sus respuestas.

En el capítulo 1, se define la percepción social de la ciencia, se muestran los principales resultados que se obtuvieron en el "Proyecto Iberoamericano de Indicadores de Percepción Pública, Cultura Científica y Participación Ciudadana", así como un panorama general de los estudios realizados en México sobre esta temática.

En el capítulo 2, se describe la metodología empleada en el proyecto de la “Primera Encuesta sobre Percepción de la Ciencia y Conocimiento de la UNAM en la Zona Metropolitana de la Ciudad de México, 2008”.

En el capítulo 3, se establecen los objetivos de este estudio, se plantean las hipótesis y la metodología a seguir, la cual incluye la descripción conceptual de los modelos de clases latentes, tablas de contingencia, la prueba de independencia  $\chi^2$  y los modelos de regresión logística multinomial.

En el capítulo 4, se muestran y se comentan todos los resultados obtenidos en el análisis.

# Capítulo 1

---

## Percepción Social de la Ciencia

En los últimos años se ha desarrollado un fuerte interés por saber qué piensa la sociedad de la aplicación de los resultados que genera la ciencia y cómo concibe los múltiples impactos de dichos resultados; cómo la sociedad asume los riesgos que entraña el desarrollo de ciertas tecnologías y cómo se apropia del conocimiento generado por la ciencia; cuánta confianza tiene en los científicos y los especialistas; cuánta información sobre la ciencia fluye socialmente y qué actitud se adopta frente a la ciencia, entre otras cuestiones. La preocupación por dar respuesta a estas y otras muchas preguntas se ha convertido en un campo de estudio que ha ido cobrando forma y se ha denominado como “percepción pública de la ciencia” o “cultura científica” (Polino, Fazio y Vaccarezza, 2003: 1), entendiendo como percepción social de la ciencia o cultura científica “el conjunto de aspectos simbólicos, valorativos, cognitivos y actitudinales de los miembros de la sociedad sobre la función de la ciencia y la tecnología, la importancia y el beneficio de su actividad (...) y el manejo de contenidos básicos del conocimiento científico” (Secyt, 2004)<sup>1</sup>.

Conocer cuál es la percepción que tiene la sociedad de la ciencia y la tecnología se ha vuelto imprescindible para las sociedades actuales, tanto como una herramienta de evaluación de la educación como para conocer el grado de alfabetismo científico de la sociedad; la actividad científica es una producción socio-cultural y en ese sentido resulta de gran importancia conocer en qué medida los sujetos pertenecientes a la sociedad valoran y tienen determinadas actitudes hacia la ciencia y la tecnología, acerca de su importancia y efectos.

Por lo que a mediados de 2001, la Organización de Estados Iberoamericanos (OEI) y la Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICYT/CYTED) pusieron en marcha el "Proyecto Iberoamericano de Indicadores de Percepción Pública, Cultura Científica y Participación Ciudadana", con la finalidad de generar indicadores que permitan evaluar la evolución de la

---

<sup>1</sup> Márquez N. E., Tirado S. F., Percepción Social de la ciencia y la tecnología de adolescentes mexicanos, Portafolio CTS No. 2, 2009, pp. 18

percepción pública, la cultura científica y la participación ciudadana, así como aportar nuevos elementos para la definición de políticas públicas; y conformar una red de grupos de investigación e instituciones iberoamericanas para la cooperación en esta temática.

A partir del proyecto conjunto se revisaron los conceptos y las metodologías e indicadores tradicionales, así como se realizaron una serie de estudios exploratorios de carácter empírico:

- a. Estudio sobre desarrollo institucional de la cultura científica;
- b. Estudio sobre relevamiento de experiencias de participación ciudadana en temas de ciencia y tecnología;
- c. Estudio sobre percepción y consumo de fuentes de información científica;
- d. Estudio sobre percepción del riesgo asociado a la ciencia y la tecnología; y
- e. Estudio sobre imaginario social de la ciencia y la tecnología.

Sobre la base de los estudios mencionados se diseñó una encuesta de percepción pública de la ciencia que se aplicó a fines de 2002 en Argentina y, posteriormente, a principios de 2003, se replicó en Brasil, España y Uruguay. De las cuales se obtuvieron los siguientes resultados dentro de las temáticas definidas<sup>2</sup>:

### **Imaginario social de la ciencia y la tecnología**

**Representación social de la ciencia.** La imagen que prevalece entre los entrevistados de todos los países tiene una componente tripartita: la ciencia como una epopeya de "grandes descubrimientos" (35.3% en promedio), la ciencia como condición de "avance tecnológico" (46.4% en promedio) y, por último, la ciencia como fuente bienhechora para la vida del ser humano (45.4% en promedio). Otras imágenes que implican una valoración negativa ("peligro de descontrol", "concentración de poder" o "ideas que pocos entienden") reciben una adhesión secundaria.

---

<sup>2</sup> Proyecto Iberoamericano de Indicadores de Percepción Pública, Cultura Científica y Participación Ciudadana. 2003.



**Utilidad de la ciencia.** La mayoría de los entrevistados (el 72% en promedio) de los cuatro países acuerdan en señalar que el desarrollo de la ciencia y la tecnología es el principal motivo de mejora en la calidad de vida de la sociedad. No obstante, esta imagen positiva no compromete totalmente el imaginario social: la gran mayoría también rechaza que la ciencia y la tecnología pueden solucionar todos los problemas (85.9% en promedio).

**La imagen de la ciencia como conocimiento legítimo.** Los resultados señalan, como tendencia general, que la sociedad moderna enfatiza la racionalidad científica y deposita su confianza en la verdad de la ciencia en desmedro de la fe religiosa. No obstante, cabe aclarar que las respuestas de Argentina, España y Uruguay son equilibradas; sólo en Brasil el acuerdo supera ampliamente (70.4%) al desacuerdo (27.2%).

**La ciencia en la vida cotidiana.** La afirmación "el mundo de la ciencia no puede ser comprendido por el común de la gente" produce respuestas que tienden a equilibrar el acuerdo y el desacuerdo. Si se consideran los resultados globales para los cuatro países, se advierte que el desacuerdo asciende al 53.4% y el acuerdo al 45.7% en promedio. En línea con otras imágenes positivas de la ciencia, también la mayoría de los entrevistados en los cuatro países -en el orden del 60%- considera que la ciencia funciona como factor de racionalidad de la cultura humana, ya que si se descuidara la ciencia "nuestra sociedad sería cada vez más irracional".

**La ciencia y la tecnología como fuentes de riesgo.** La gran mayoría de los entrevistados de los cuatro países (74.3% en promedio) considera que "los beneficios de la ciencia y la tecnología son mayores que los efectos negativos". No obstante, ante la afirmación "el desarrollo de la ciencia trae problemas para la humanidad", en los cuatro países se asumen posturas encontradas. En Argentina, las respuestas están muy equilibradas, aunque al igual que en Brasil prima el "desacuerdo" (50% en promedio). En España y Uruguay las respuestas se decantan por el "acuerdo" (57% en promedio). En este sentido, y pese a una tendencia general de imagen favorable hacia la ciencia, se percibe que la misma no está exenta de generar consecuencias negativas. Entre los principales problemas se mencionan "los peligros de aplicar algunos conocimientos" y "la utilización del conocimiento para la guerra". Asimismo, las controversias científicas son percibidas por la gran mayoría como un factor que alimenta la incertidumbre de la sociedad e impide valorar las consecuencias del desarrollo de ciertos conocimientos.

**La imagen de los científicos y de la actividad científico -tecnológica.** En los cuatro países la vocación por el conocimiento es el principal motivo que se considera moviliza a los científicos en su trabajo cotidiano, seguido por la necesidad de solucionar problemas a la gente. Conquistar poder o un premio importante, reciben adhesiones marginales. Por otra parte, las capacidades de los científicos no son suficientes para la toma de decisiones políticas: en efecto, la mitad de la muestra de los cuatro países (51.6% en promedio) está en desacuerdo con que "los científicos son quienes mejor saben lo que conviene investigar para el desarrollo del país". Aún así, la mayoría (57.7% en promedio) coincide en que "el gobierno no debe intervenir en el trabajo de los científicos aún cuando sea el mismo gobierno quien les pague".

**Percepción de la ciencia y la tecnología local.** En los cuatro países predomina una imagen del desarrollo científico- tecnológico local según la cual existe "un poco de ciencia y tecnología en algunas áreas (temáticas)". En los casos de Argentina, Brasil y España, esta categoría oscila entre el 55% y el 64% de las adscripciones. En Uruguay la orientación es todavía más notoria (80%). Por otra parte, la imagen de una actividad insuficientemente financiada por el estado está muy extendida entre la población entrevistada en todos los países. En Argentina, España y Uruguay, la estimación alcanza el 87% de las respuestas. Sin embargo, Brasil presenta nuevamente un comportamiento diferente, en la medida en que un porcentaje marcadamente superior (27.8%) al resto de los países opina que el estado financia de manera "razonablemente suficiente" la investigación en ese país. Asimismo, el 82% de la muestra en Argentina, el 62.3% en Brasil y el 78.9% en España señala que el "poco apoyo estatal" es el factor principal que limita el desarrollo de la ciencia y la tecnología, descartando en ello la responsabilidad de otros sectores. Por otra parte, en Uruguay (66%), Argentina (59.4%) y, en menor medida, España (43.2%), los entrevistados opinan que los conocimientos generados en sus países "sirven, pero no se difunden".

#### **Procesos de comunicación social de la ciencia**

**Información científica incorporada.** En Argentina (80%), Brasil (71%) y España (67%) los entrevistados se consideran "poco informados" en lo que refiere a la ciencia y la tecnología.

**Consumo de información científica.** El consumo de información científica en diarios (53.4%) y televisión (64%) es mayoritariamente ocasional para Argentina. En Brasil las características de

consumo son similares. También en España el comportamiento es parecido en lo que refiere a diarios -58% del consumo es ocasional- sin embargo se acentúa una tendencia de escaso consumo de contenido científico televisado (81%). A diferencia de Argentina, Brasil y España, los datos de Uruguay presentan un perfil más equilibrado en las mismas categorías. Para las revistas de divulgación, en todos los países el consumo tiene características fundamentalmente esporádicas.

**Valoraciones acerca de científicos y periodistas.** En los cuatro países las respuestas tienden a considerar que sólo en algunas ocasiones la comunicación de los científicos hacia la sociedad es de difícil comprensión. Los entrevistados asumen con ello que la eventual incapacidad de comunicación de los científicos no es una condición estructural de sus competencias profesionales sino que, fundamentalmente, responde a otros factores.

### **Participación ciudadana en temas de ciencia y tecnología**

En los cuatro países la gran mayoría de los entrevistados (94.5% en promedio) remarca la importancia de participar pero, al mismo tiempo, esta opinión es asimétrica respecto al minoritario porcentaje (7.3% en promedio) de personas que manifiesta haber tenido experiencias de participación concretas. Asimismo se observa que en el caso de España, pese a su carácter minoritario, el nivel de la participación efectiva prácticamente duplica al de los otros países. Además, se observa que para la amplia mayoría de los entrevistados de los cuatro países "el cuidado de la vida y de la salud" constituye el principal motivo que justifica la utilidad de la participación. Otras opciones, tales como "controlar el funcionamiento de las empresas" o "controlar la actividad de los científicos" reciben adhesiones que no superan el 25% en ninguno de los casos. Asimismo, uno de los principales obstáculos que coincide en señalar la mayoría en los cuatro países -siempre con una frecuencia superior al 50% de la muestra- es que las personas no tienen conocimientos suficientes para ejercer dicha práctica. En los casos de Brasil, España y Uruguay, este motivo es el principal entre los señalados. Distinto es para el caso de Argentina, donde ocupa el segundo lugar, antecedido por la categoría "la gente tiene problemas más importantes por los cuales reclamar y participar". Sin embargo, esta elección, prioritaria en Argentina y Uruguay -donde obtiene el segundo puesto- ocupa el último lugar en Brasil y España.

## Encuestas sobre Percepción de la Ciencia en México

El Consejo Nacional de Ciencia y Tecnología (CONACYT) realizó en 1997 la primera encuesta sobre la Percepción Pública de la Ciencia y la Tecnología, con la finalidad de detectar el nivel de conocimiento que tienen las personas en el país en torno a nuevos descubrimientos científicos y desarrollos tecnológicos, su actitud frente a sus posibles impactos, las fuentes de información que les hacen llegar tales conocimientos y la percepción que tienen respecto a su propio conocimiento y entendimiento de estos tópicos.

Esta encuesta se levantó directamente en viviendas, considerando una muestra de 2,568 personas mayores de 18 años residentes en zonas urbanas de la República Mexicana y tomó como base la metodología de estudios realizados por la Unión Europea, por la National Science Foundation de los Estados Unidos de América, y recomendaciones de la Organización para la Cooperación y el Desarrollo Económico (OCDE)<sup>3</sup>.

La información recabada en esta encuesta permite tener conocimiento de aspectos como el nivel de entendimiento del público en torno a la ciencia y la tecnología; la actitud de las personas frente a los cambios y efectos originados en dichas actividades; y los hábitos y canales más usuales para obtener información sobre estos temas. Además los resultados permiten caracterizar a los grupos de la población con respecto a su relación con temas de ciencia y tecnología.

El análisis que se llevó a cabo en el reporte de esta encuesta, está basado en el estudio que presentó el Profesor Jon D. Miller, Vicepresidente de la Academia de Ciencias de Chicago, en la sesión 1 del Simposium sobre Percepción Pública de la Ciencia y la Tecnología, que se realizó en la ciudad de Tokio en noviembre de 1996.

Se construyeron los siguientes indicadores:

- Cultura Científica de la Sociedad
- Involucramiento en la Ciencia y la Tecnología

---

<sup>3</sup> Informe de la primera encuesta sobre Percepción Pública de la Ciencia y la Tecnología, CONACYT, 1997.

- Interés por la Ciencia y la Tecnología en dos dimensiones: promesas de la ciencia e inquietud ante el avance científico y tecnológico.

### **Cultura Científica de la Sociedad**

La Cultura Científica de la Sociedad permite calificar a las personas de acuerdo con su nivel de entendimiento de dos dimensiones (vocabulario científico básico y entendimiento de procesos o métodos científicos, estos últimos como forma de prueba de la teoría en la realidad) en tres categorías:

1. Bien Informados (BI): Son los individuos que muestran un alto nivel de entendimiento de conceptos y procesos, por lo que resultan ser los más capacitados para adquirir y comprender información científica o tecnológica y tienen potencial para discernir y opinar sobre controversias en temas relacionados.
2. Moderadamente Informado (MI): Se refiere a los individuos que muestran un nivel aceptable en las dos dimensiones de entendimiento.
3. Individuos con Información Escasa (IE): Son las personas con entendimiento limitado en ambas dimensiones.

### **Involucramiento en la Ciencia y la Tecnología**

Público Involucrado en la Ciencia (PIC): Es una medida de la persistencia en el consumo de información relevante y se calcula mediante las preguntas de interés y conocimiento de tópicos referentes a nuevos descubrimientos y de temas referentes al uso de nuevos inventos y tecnologías, así como de sus hábitos de consumo de programas o artículos de ciencia y tecnología, y noticieros (información relevante). Hay tres categorías:

1. Público Involucrado (PV)
2. Público Interesado (PI)
3. Público no interesado (PNI)

### **Interés por la Ciencia y la Tecnología**

Promesas de la ciencia y la tecnología: refleja el juicio de las personas de que la ciencia y la tecnología son elementos esenciales para mejorar la calidad de vida y, al mismo tiempo, el supuesto implícito de que esta mejora será continua o, en su caso, será posible en un futuro no lejano.

Inquietud ante el avance científico y tecnológico: las preguntas relativas a esta actitud expresan la inquietud de la gente ante el rápido cambio experimentado en la vida moderna, así como una sensación de que la ciencia podría a veces contraponerse a los valores tradicionales o de fe de las personas.

Los resultados de la encuesta muestran que 11.4% de las personas encuestadas se pueden considerar como bien informadas (BI), mientras que el 33.6% tiene información moderada (MI) y el 54.8% cuenta con información escasa (IE).

El público involucrado en la ciencia y la tecnología en México representa 8.9% del público bien informado, lo que equivale a 1% de la muestra total. Por su parte, el público interesado representa el 26.0% del público bien informado y el público no interesado el 65.1% restante.

El 77.3% de los entrevistados presenta un esquema positivo hacia el avance científico y tecnológico. Por su parte, respecto a los esquemas negativos el 56.5% canalizaron sus inquietudes ante el avance científico hacia planteamientos positivos.

Posteriormente, a partir del 2001, el CONACyT en colaboración con el Instituto Nacional de Estadística, Geografía e Informática (INEGI), ha realizado cada dos años la Encuesta Nacional Sobre la Percepción Pública de la Ciencia y la Tecnología en México (ENPECyT), la cual se ha llevado a cabo bajo un esquema de muestreo probabilístico, polietápico, estratificado y por conglomerados, siendo la última unidad de selección una persona de 18 años cumplidos o más al momento de la entrevista, con un tamaño de muestra de 3,200 viviendas.

El cuestionario para estas encuestas se ha diseñado de tal forma que se puedan obtener indicadores sobre el consumo de medios; el nivel de cultura científica de la sociedad; la percepción pública de la ciencia y la tecnología; el apoyo económico del gobierno a la ciencia; entre otros.

El cuadro 1, muestra los porcentajes de la evaluación referente a la cultura científica de la sociedad, para cada una de las encuestas.

Cuadro 1

Evaluación	2001	2003	2005	2007
Bien informados	5.7	7.9	8.9	8.5
Moderadamente informados	28.2	36.3	26	30.1
Información escasa	66.1	55.8	65.1	61.4

Fuente: Informe de la ENPECyT 2001, 2003, 2005 y 2007.

Se observa que en todos los años, la mayor parte de la población cuenta con información escasa, y menos del 10% están bien informados.

En el cuadro 2, se observan los porcentajes de las personas que están de acuerdo con las afirmaciones referentes a las actitudes frente al avance científico y tecnológico.

Cuadro 2

Afirmación	2001	2003	2005	2007
El progreso científico y tecnológico ayudará a encontrar la cura para enfermedades como el cáncer y el sida	90.2	89.9	97.2	94
En general, la automatización de las fábricas y la computación crearán más empleos de los que se eliminarán	43.6	46.5	51.1	38.7
Gracias a los avances científicos y tecnológicos, los recursos naturales de la tierra serán inagotables	39.2	41.5	41.5	27.9

Fuente: Informe de la ENPECyT 2001, 2003, 2005 y 2007.

Se observa, en general, que las personas reportaron sus preferencias del avance científico y tecnológico hacia aspectos relacionados con la salud y el bienestar personal, la mayoría de la población considera que el progreso científico y tecnológico ayudará a encontrar la cura de enfermedades.

Sobre la percepción ante el avance científico, tecnológico y su relación con la fe y tradiciones, en el cuadro 3, se observan los porcentajes de las personas que están de acuerdo con las afirmaciones cuestionadas.

Cuadro 3

Afirmación	2001	2003	2005	2007
Confiamos demasiado en la fe y muy poco en la ciencia	75.2	81	83.3	78.4
Algunos números son de la suerte	51.7	49.9	41.1	35.3
Debido a su conocimiento, los investigadores y científicos tienen un poder que los hace peligrosos	80.7	70.3	55	49.8

Fuente: Informe de la ENPECyT 2001, 2003, 2005 y 2007.

Las personas confían más en la fe que en la ciencia, en cuanto a su percepción sobre el peligro que representan los científicos, se observa que este porcentaje ha ido disminuyendo con el transcurso del tiempo.

En el cuadro 4, se observan los porcentajes de las personas que están de acuerdo con que el gobierno debe apoyar a la ciencia.

Cuadro 4

Afirmación	2001	2003	2005	2007
La investigación científica debe ser apoyada por el gobierno federal, aun cuando los beneficios no sean inmediatos	92	86.2	90	84.3

Fuente: Informe de la ENPECyT 2001, 2003, 2005 y 2007.

Se observa que la población está de acuerdo en que el gobierno debe apoyar a la ciencia.



# Capítulo 2

---

## **Primera Encuesta sobre Percepción de la Ciencia y Conocimiento de la UNAM en la Zona Metropolitana de la Ciudad de México**

La Unidad de Estudios Sobre la Opinión (UDESOS) del Instituto de Investigaciones Sociales de la UNAM, en un trabajo conjunto con la Dirección General de Divulgación de la Ciencia (DGDC), realizó la “Primera Encuesta sobre Percepción de la Ciencia y Conocimiento de la UNAM” en agosto del 2008, la cual tuvo como propósito recabar información concerniente a la percepción que tiene la gente sobre la ciencia, y de manera particular, el conocimiento que tiene de la UNAM, en cuanto a sus instalaciones, actividades y aportes a la ciencia.

La información en cuanto a la percepción y el conocimiento que tiene la población acerca de la UNAM y las actividades que realiza, es importante para la DGDC que tiene entre sus principales objetivos fomentar la ciencia y contribuir a profesionalizar su divulgación. Los divulgadores de la ciencia de la DGDC son los intérpretes del trabajo de investigación y enseñanza que realiza la UNAM y, a través del trabajo en equipo, se encargan de realizar actividades, crear espacios y emplear todos los medios a su alcance para llevar el conocimiento científico y tecnológico a grandes sectores de la población, no sólo nacionales sino de todo el mundo.

En este sentido, la información derivada de este estudio es relevante, pues es una herramienta para la reflexión y el debate sobre la función y el quehacer de las instituciones académicas en general y, de la UNAM en particular, en cuanto a la realización y divulgación de la investigación científica que realiza.

Definidas las temáticas, el tamaño de la muestra y las unidades de análisis de la encuesta, se diseñaron los procedimientos a utilizar en la captación de la información requerida.

## Descripción del cuestionario

Se elaboró un cuestionario orientado a recoger las opiniones, actitudes y valores de los entrevistados en cuanto al tema. Éste incluyó preguntas de dos tipos: 33 preguntas cerradas, diseñadas para recoger percepciones y conocimiento de instituciones, medios informativos, entre otros; y 3 preguntas abiertas enfocadas a indagar las representaciones sociales de la ciencia y el conocimiento de actividades relacionadas con el desarrollo científico.

El cuestionario se conformó de tal manera que permitiera relacionar variables sociodemográficas con indicadores sobre la percepción de la ciencia y conocimiento de la UNAM. Se dividió en tres secciones:

1. Percepción de la Ciencia: A través de esta sección se buscó obtener información de la percepción que tiene la población sobre la ciencia.

Objetivos específicos:

- Conocer la representación social de la ciencia.
- Conocer la percepción de la población en cuanto a las motivaciones del investigador científico y la valoración de la profesión de científico.
- Conocer la percepción de la población sobre la investigación científica en México.

2. Percepción de la UNAM: esta sección buscó registrar el conocimiento que la población tenía de la UNAM, así como la percepción de las actividades que realiza, entre ellas, las relacionadas con la investigación científica.

Objetivos específicos:

- Saber si la población conoce el significado de la abreviatura UNAM.
- Saber cuáles son los lugares de la UNAM que más conoce la población.
- Indagar sobre el conocimiento, la percepción y valoración que las personas tienen de las actividades que se desarrollan en la UNAM.
- Conocer si las personas se informan sobre la ciencia que se desarrolla en la UNAM, qué tanto lo hacen y a través de qué medios.

3. Datos del encuestado: Esta sección se destinó para recabar los datos generales de los entrevistados: edad, sexo, nivel educativo, actividad económica e ingreso.

## **Diseño muestral**

### **Cobertura**

La presente encuesta generará información con precisión y confianza a nivel de la Zona Metropolitana de la Ciudad de México, que comprende las 16 delegaciones del Distrito Federal, 40 municipios del estado de México y un municipio del estado de Hidalgo.

### **Marco muestral**

El marco de muestreo que se empleó, se construyó a partir de la información cartográfica y demográfica que se obtuvo del Censo de Población y Vivienda 2005<sup>4</sup>. De manera que se contó con información a nivel de localidad y por AGEB's (Áreas Geoestadísticas Básicas)<sup>5</sup>.

### **Población Objetivo**

La población objetivo estuvo conformada por personas de 15 años y más, que residen permanentemente en viviendas particulares dentro de la Zona Metropolitana de la Ciudad de México.

---

<sup>4</sup> INEGI, IRIS-SCINCE II Censo de población y Vivienda 2005.

<sup>5</sup> Son subdivisiones de una localidad urbana en conjuntos que generalmente son de 25 a 50 manzanas, perfectamente delimitadas por calles, avenidas, andadores, arroyos, cercas, veredas, límites prediales, cuyo uso de suelo es principalmente habitacional, industrial, de servicios, comercial, etc.

## **Esquema de muestreo**

El esquema de muestreo de la encuesta fue probabilístico, es decir, que las unidades de selección tienen una probabilidad conocida y distinta de cero de ser seleccionadas y polietápico, porque se definieron cuatro etapas de selección.

## **Tamaño de la muestra**

El tamaño de la muestra se calculó para estimar proporciones a nivel Zona Metropolitana de la Ciudad de México. Como desconocemos la proporción de personas que perciben de manera adecuada el concepto de ciencia o conocimiento de la UNAM, se estima que la proporción esté entre 0.2 y 0.8, por lo que se recomienda el uso de la siguiente expresión:

$$n = \frac{1}{\delta^2}, \text{ donde } n = \text{tamaño de muestra y } \delta = \text{error absoluto.}$$

Con lo anterior tenemos que para un nivel de confianza del 95% y un error absoluto de 0.028 se estima un tamaño de muestra aproximado de 1200. Asimismo, se estima una tasa de no respuesta del 5%, por lo que el tamaño final de la muestra es de 1260, de los cuales resultaron efectivos 1256 cuestionarios.

## **Procedimiento de selección**

Etapas de muestreo

1. Las unidades primarias de muestreo (UPM), estuvieron representadas por las AGEB's, se seleccionaron de manera aleatoria 105 AGEB's del total que comprende la Zona Metropolitana de la Ciudad de México.
2. Las unidades secundarias de muestreo (USM), fueron las manzanas, de las cuales se seleccionaron dos al azar, por cada AGEB seleccionada previamente.
3. Las unidades terciarias de muestreo (UTM), fueron las viviendas, se seleccionaron de manera aleatoria, 6 viviendas por manzana seleccionada previamente.

4. Las unidades de cuarto orden (UCM) fueron los individuos, se seleccionó a un individuo de 15 años o más.

### Distribución de la muestra

Se aplicaron un total de 1256 entrevistas distribuidas en tres entidades federativas, como sigue:

Total de cuestionarios por entidad federativa

Clave	Entidad Federativa	Total
13	Hidalgo	12
9	Distrito Federal	609
15	Estado de México	635

### Factores de expansión

Para realizar de manera correcta el análisis de la información de la encuesta es necesario considerar el esquema de muestreo, lo cual incide en la aplicación de ponderadores. El uso de los ponderadores permite asignar un peso a cada uno de los individuos de la muestra de acuerdo al diseño muestral aplicado, y por lo tanto, de acuerdo a su probabilidad de selección.

### Probabilidades de selección para la muestra

Notación:

Sea  $P_i$  la probabilidad de elegir una vivienda en la manzana,

$P_j$  la probabilidad de elegir una manzana dentro de la AGEB y

$P_k$  la probabilidad de elegir una AGEB.

$P_l$  la probabilidad de elegir un individuo.

Entonces la probabilidad de seleccionar un individuo dentro de la vivienda  $i$ , la manzana  $j$  y la AGEB  $k$ , esta dada por:

$$P_{ijkl} = P_i P_j P_k P_l$$

Por tanto los factores de expansión quedan de la siguiente manera:

$$F_{ijkl} = \frac{1}{P_{ijkl}}$$

# Capítulo 3

---

## Objetivos

El objetivo general de este estudio es analizar a través de métodos estadísticos adecuados los resultados de la “Primera Encuesta sobre Percepción de la Ciencia y Conocimiento de la UNAM en la Zona Metropolitana de la Ciudad de México”.

Este análisis se centrará en determinar si existen asociaciones entre las respuestas de los individuos, para los diferentes grupos de variables que se incluyen en el estudio, aplicando modelos de clases latentes.

Asimismo, se desea identificar los grupos de la población que se distinguen por tener mejor percepción de la ciencia, aplicando la prueba  $X^2$  y el modelo de regresión logística multinomial.

## Hipótesis

Dados los antecedentes se establecen las siguientes hipótesis:

- Existe asociación entre las respuestas en las variables de percepción de la ciencia y el perfil sociodemográfico de los individuos (sexo, edad, nivel de escolaridad y condición de actividad). Se espera que entre mayor nivel de escolaridad tengan los individuos, tengan mejor concepto y percepción de la ciencia.
- Existe asociación entre las variables de percepción de la UNAM y el perfil sociodemográfico de la población. Se espera que entre mayor escolaridad tengan los individuos, tengan mejor percepción y conocimiento de la UNAM.

## Metodología

Para cumplir los objetivos de este estudio se realizó un modelo de clases latentes, para cada conjunto de preguntas que involucran los distintos tópicos de esta encuesta. Posteriormente se obtuvieron pruebas de independencia, para verificar si existían asociaciones entre las clases obtenidas con el modelo y cada una de las variables sociodemográficas de los individuos. Finalmente se aplicó el modelo de regresión logística multinomial para confirmar de manera conjunta la existencia de estas asociaciones.

## Modelos de clases latentes

Un modelo de variables latentes, es un modelo que especifica la distribución conjunta de  $(p + q)$  variables aleatorias en el cual alguna de ellas no es observable, por lo tanto el modelo se compone de dos tipos de variables, las que se pueden observar directamente, conocidas como variables manifiestas, las cuales conforman un vector de  $p$  componentes,  $X' = (X_1, \dots, X_p)$  y las variables latentes, que no son observables, denotadas por  $Y$  que se expresan mediante el vector  $Y' = (Y_1, \dots, Y_q)$  con  $q < p$ .

Los modelos de variables latentes pueden clasificarse de acuerdo a la escala de medida tanto de las variables manifiestas como de las variables latentes, en una doble clasificación en variables métricas y categóricas. Se consideran variables métricas a las variables que toman valores en el conjunto de los números reales, pueden ser discretas o continuas, y las variables categóricas a las variables que están formadas por un conjunto de categorías, pueden ser nominales u ordinales.

En la siguiente tabla se muestra la clasificación de las variables y el tipo de análisis:

Variables latentes		Variables manifiestas	
		Métricas	Categóricas
Métricas		Análisis factorial	Análisis de rasgos latentes
Categóricas		Análisis de perfiles latentes	Análisis de clases latentes



Las relaciones entre ellas deben ser expresadas en términos de las distribuciones de probabilidad, después de observar  $x$ , mientras que  $y$  está dada por la distribución condicional dado  $x$ .

Como sólo  $x$  puede ser observada, cualquier inferencia debe estar basada en su distribución conjunta, cuya densidad se puede expresar como

$$P(X = x) = f(x) = \int g(x|y)h(y)dy,$$

donde  $g(x|y)$  es la distribución condicional conjunta de las variables manifiestas  $X_1, \dots, X_p$  dadas las variables latentes, es decir,

$$g(x|y) = P(X = x|Y = y)$$

y  $h(y)$  representa la función de densidad de  $Y$ .

El supuesto en los modelos de variables latentes es el principio de independencia condicional o local, el cual establece que dado un valor fijo de la variable latente, las variables manifiestas son mutuamente independientes, por lo que

$$g(x|y) = P(X = x|Y = y)$$

se descompone como un producto de probabilidades condicionales marginales

$$g(x|y) = \prod_{i=1}^p g_i(x_i|y).$$

Este supuesto implica que las variables latentes son las que explican la relación entre las variables manifiestas, es decir, la asociación observada entre variables manifiestas está explicada por las variables latentes.

El modelo toma la forma

$$f(x) = \int h(y)g(x|y)dy = \int h(y) \prod_{i=1}^p g_i(x_i|y)dy.$$

Dado que el principal interés es estudiar las variables latentes a partir de la información contenida en las variables manifiestas, una de las prioridades es la densidad condicional:

$$h(y|x) = \frac{g(x|y)h(y)}{f(x)},$$

conocida como densidad a posteriori.

## El modelo de clases latentes

El modelo de clases latentes supone que la población de individuos se divide en un determinado número de clases latentes excluyentes y exhaustivas, tantas como categorías tenga la variable latente, de tal forma que cada individuo pertenezca sólo a una clase latente. En este sentido se puede ver el análisis de clases latentes como un procedimiento de segmentación.

El supuesto de independencia local implica que dentro de cada clase latente las variables manifiestas son estadísticamente independientes, por lo que la relación entre estas variables viene dada exclusivamente por la pertenencia de un individuo a una clase en particular, ya que se espera que si la variable latente permanece constante, cualquier relación existente entre las variables manifiestas desaparece.

Otro de los supuestos para el modelo de clases latentes, es que las clases latentes son internamente homogéneas, es decir, todos los miembros de una clase latente tienen la misma distribución de probabilidades con respecto a la variable latente, y esta será distinta a la distribución de probabilidades para los individuos pertenecientes a otra clase, por lo que los individuos de diferentes clases presentan características diferentes. Este hecho sirve para diferenciar a los individuos pertenecientes a diferentes clases y poder caracterizar tanto la variable como las clases latentes.

El modelo se puede definir como un modelo estadístico que estudia las relaciones entre las variables que conforman una tabla de contingencia de  $p$ -vías, considerando un conjunto de  $q$  variables latentes, por lo que la tabla de  $p$ -vías se considera incompleta, pues la tabla completa está conformada por  $p + q$  vías.

Para el desarrollo teórico del modelo es suficiente considerar sólo una variable latente, ya que modelos con más de una variable latente,  $q > 1$ , pueden ser desarrollados considerando  $q = 1$ , bajo adecuadas restricciones a los parámetros del modelo.

## El modelo de clases latentes para una variable latente $Y$ con $C$ categorías o clases latentes

Supongamos que se tiene un conjunto de  $p$  variables manifiestas  $X_1, \dots, X_p$  las cuales se consideran indicadoras de una variable latente  $Y$  y estas variables conforman un modelo de clases latentes con  $C$  clases.

Se denota a  $\pi_x(x)$  como la densidad conjunta de las variables manifiestas  $X' = (X_1, \dots, X_p)$  y  $x' = (x_1, \dots, x_p)$  denota un determinado patrón de respuesta en el cual cada una de las  $x_i$  toma distintos valores dependiendo de las categorías de la correspondiente variable manifiesta. Cada  $X_i$  contiene  $I_i$  categorías, para  $i = 1, \dots, p$ , por lo que las variables manifiestas conforman una tabla de contingencia de  $p$ -vías con  $\prod_{i=1}^p I_i$  patrones de respuesta. Si estas variables son medidas sobre una muestra de  $n$  individuos,  $x_{ih}$  denotará el valor observado de la variable  $X_i$  para el individuo  $h$ ,  $x_h$  será el vector de respuestas para el individuo  $h$  y  $x$  denotará un vector o patrón de respuestas cualquiera. Por el principio de independencia local, la densidad condicional

$$P(X = x|Y = c),$$

está dada por

$$\pi_{X|Y(c)}(x) = \prod_{i=1}^p \pi_{X_i|Y(c)}(x_i),$$

donde

$$\pi_{X_i|Y(c)}(x_i) = P(X_i = x_i|Y = c), \quad x_i = 1, \dots, I_i; \quad c = 1, \dots, C.$$

La distribución conjunta de  $X$  e  $Y$  denotada por

$$\pi_{X,Y}(x, c),$$

es obtenida como

$$\pi_{X,Y}(x, c) = \pi_Y(c)\pi_{X|Y(c)}(x),$$

donde  $\pi_Y(c) = P(Y = c)$  representa la proporción de individuos que pertenecen a la clase latente  $c$ ,  $c = 1, \dots, C$ , también conocida como probabilidad a priori.

Utilizando las expresiones anteriores el modelo de clases latentes puede escribirse como sigue:

$$\pi_X(x) = \sum_{c=1}^C \pi_{X,Y}(x, c) = \sum_{c=1}^C \pi_Y(c) \pi_{X|Y(c)}(x) = \sum_{c=1}^C \pi_Y(c) \prod_{i=1}^p \pi_{X_i|Y(c)}(x_i)$$

donde los parámetros están sujetos a las siguientes restricciones,

$$\sum_{c=1}^C \pi_Y(c) = 1$$

$$\sum_{x_i=1}^{I_i} \pi_{X_i|Y(c)}(x_i) = 1; \quad i = 1, \dots, p, \quad c = 1, \dots, C.$$

El modelo de clases latentes es paramétrico y utiliza los datos observados para estimar los parámetros del modelo que son:

La probabilidad de cada una de las clases latentes

$$\pi_Y(c), \quad c = 1, \dots, C$$

y la probabilidad de respuesta condicional de cada una de las variables manifiestas dentro de cada clase latente  $\pi_{X_i|Y(c)}(x_i), i = 1, \dots, p; c = 1, \dots, C; x_i = 1, \dots, I_i$ .

La probabilidad

$$\pi_X(x) = \sum_{c=1}^C \pi_{X,Y}(x, c),$$

implica que la población puede ser dividida en  $C$  clases latentes exhaustivas y exclusivas, por lo tanto la probabilidad conjunta de las variables manifiestas se puede obtener sumando sobre la dimensión latente. En este sentido, la expresión implica la existencia de una variable latente.

El análisis posterior del modelo de clases latentes se relaciona con lo que se puede decir acerca de los individuos que pertenecen a una clase determinada.

Consideremos

$$g(x|y) = \prod_{i=1}^p g_i(x_i|y) \quad \forall$$

$$\pi_{Y|X(x)}(c) = P(Y = c|X = x),$$

la distribución condicional de  $Y$  dado  $X$ , también conocida como distribución a posteriori. Los individuos se clasifican dentro de la clase latente más probable.

Aplicando la definición de probabilidad condicional se tiene

$$\pi_{Y|X(x)}(c) = \frac{\pi_{X,Y}(x,c)}{\pi_X(x)} = \frac{\pi_Y(c)\pi_{X|Y(c)}(x)}{\pi_X(x)}$$

para cada patrón de respuesta  $x$  se asigna a la clase latente en la cual la probabilidad es mayor.

Debido a que la regla de asignación de los individuos se basa en probabilidades, se pueden tener ciertos errores en las asignaciones. Por lo que se ha propuesto un índice que permite cuantificar la magnitud relativa de este error de asignación. La proporción de individuos correctamente clasificados está dada por,

$$P = \frac{\sum_X f_X \pi_{Y|X(x)}(c^*)}{n}$$

donde para cada patrón de respuesta  $x$ ,  $\pi_{Y|X(x)}(c^*)$  es la probabilidad modal,  $\pi_{Y|X(x)}(c^*) = \max_{1 \leq c \leq C} \pi_{Y|X(x)}(c)$  y  $f_X$  la frecuencia observada.

Mientras el valor de  $P$  esté más cerca de uno, más fuerte es la relación entre la variable latente y las variables manifiestas, lo que indica que a medida que esta cantidad aumenta, la asignación de los individuos a las clases latentes será más adecuada.

## Estimación de los parámetros

En el modelo de clases latentes, lo usual es obtener los estimadores máximo verosímiles de los parámetros del modelo, lo cual conduce a obtener la solución para las ecuaciones de verosimilitud. Para resolver estas ecuaciones, se propone utilizar un procedimiento iterativo, conocido como el algoritmo EM (Expectation Maximization), desarrollado por Dempster, Laird y Rubin, 1977. Este método nos permite encontrar los estimadores máximo verosímiles de los parámetros de la distribución subyacente de un conjunto de datos, cuando los datos son incompletos o existen datos faltantes, o bien para variables latentes.

Bajo el modelo

$$\pi_X(x) = \sum_{c=1}^C \pi_Y(c) \pi_{X|Y(c)}(x) = \sum_{c=1}^C \pi_Y(c) \prod_{i=1}^p \pi_{X_i|Y(c)}(x_i),$$

la función de verosimilitud para una muestra de  $n$  individuos está dada por

$$\mathfrak{S} = \prod_{h=1}^n \pi_X(x_h, \pi_Y(c), \pi_{X_i|Y(c)}(x_{ih})) = \prod_{h=1}^n \sum_{c=1}^C \pi_Y(c) \prod_{i=1}^p \pi_{X_i|Y(c)}(x_{ih}).$$

Equivalentemente se puede maximizar el logaritmo de la función de verosimilitud

$$L = \sum_{h=1}^n \log(\sum_{c=1}^C \pi_Y(c) \prod_{i=1}^p \pi_{X_i|Y(c)}(x_{ih}))$$

bajo las restricciones

$$\sum_{c=1}^C \pi_Y(c) = 1$$

$$\sum_{x_i=1}^{I_i} \pi_{X_i|Y(c)}(x_i) = 1 \text{ para } i = 1, \dots, p; \quad c = 1, \dots, C$$

Para obtener los estimadores máximo verosímiles, las variables manifiestas se reemplazan por un vector  $Z_i$  con  $I_i$  elementos,  $z_i(s)$  definido por

$$z_i(s) = \begin{cases} 1 & \text{si la respuesta está en la categoría } s \text{ de la variable } i \\ 0 & \text{en otro caso} \end{cases}$$

$$s = 1, \dots, I_i$$

donde  $I_i$  es el número de categorías de la variable  $i$ -ésima, así  $\sum_s z_i(s) = 1$ .

El patrón de respuestas para cada individuo está dado por  $z' = (z_1', \dots, z_p')$ .

Si las probabilidades condicionales se definen como

$$\pi_{ic}(s) = P(z_i(s) = 1 | Y = c)$$

$$= P(\text{un individuo de la clase latente } c, \text{ esté en la categoría de la variable } i)$$

La función de probabilidad conjunta de  $X$  está dada por

$$\pi_X(x) = \sum_{c=1}^C \pi_Y(c) \prod_{i=1}^p \prod_{s=1}^{I_i} \{\pi_{ic}(s)\}^{z_i(s)},$$

y la distribución a posteriori toma la forma

$$\pi_{Y|X}(c) = \frac{\pi_Y(c) \prod_{i=1}^p \prod_{s=1}^{I_i} \{\pi_{ic}(s)\}^{z_i(s)}}{\pi_X(x)}.$$

Lo que implica que el logaritmo de la función de verosimilitud, el cual está dado por

$$L = \sum_{h=1}^n \log\left(\sum_{c=1}^C \pi_Y(c) \prod_{i=1}^p \pi_{X_i|Y(c)}(x_{ih})\right),$$

puede escribirse como

$$L = \sum_{h=1}^n \log \pi_X(x_h) = \sum_{h=1}^n \log\left(\sum_{c=1}^C \pi_Y(c) \prod_{i=1}^p \prod_{s=1}^{I_i} \{\pi_{ic}(s)\}^{z_i(s)}\right),$$

es decir,

$$\pi_{X_i|Y(c)}(x_{ih}) = \prod_{s=1}^{I_i} \{\pi_{ic}(s)\}^{z_i(s)}$$

considerando las restricciones  $\sum_{c=1}^C \pi_Y(c) = 1$ ;

$$\sum_{z_i=1}^{I_i} \pi_{X_i|Y(c)}(z_i) = 1 \text{ para } i = 1, \dots, p; \quad c = 1, \dots, C \text{ y } \sum_s z_i(s) = 1$$

La función a maximizar es

$$\Phi = L + \theta \sum_{c=1}^C \pi_Y(c) + \sum_{c=1}^C \sum_{i=1}^p \beta_{ic} \sum_{s=1}^{I_i} \pi_{ic}(s)$$

donde  $\theta$  y  $\{\beta_{ic}\}$  son multiplicadores de Lagrange.

Las derivadas parciales con respecto a las probabilidades de cada clase latente están dadas por

$$\frac{\partial \Phi}{\partial \pi_Y(c)} = \frac{\sum_{h=1}^n \pi_{X|Y(c)}(x_h)}{\pi_X(x_h)} + \theta$$

además se tiene

$$\frac{\partial L}{\partial \pi_{ic}(s)} = \frac{\sum_{h=1}^n \pi_Y(c) \frac{\partial \pi_{X|Y(c)}(x_h)}{\partial \pi_{ic}(s)}}{\pi_X(x_h)},$$

donde

$$\frac{\partial \pi_{X|Y(c)}(x_h)}{\partial \pi_{ic}(s)} = \frac{\partial}{\partial \pi_{ic}(s)} \exp\left[\sum_{i=1}^p \sum_{s=1}^{I_i} x_{ih}(s) \ln \pi_{ic}(s)\right] = \frac{\pi_{X|Y(c)}(x_h) x_{ih}(s)}{\pi_{ic}(s)}$$

luego

$$\frac{\partial \Phi}{\partial \pi_{ic}(s)} = \pi_Y(c) \sum_{h=1}^n \frac{\pi_{X|Y(c)}(x_h) x_{ih}(s)}{\pi_X(x_h) \pi_{ic}(s)} + \beta_{ic} = \sum_{h=1}^n \frac{\pi_{Y|X(x_h)}(c) x_{ih}(s)}{\pi_{ic}(s)} + \beta_{ic}$$

Igualando a cero y sumando sobre  $s$  se obtiene

$$\beta_{ic} = - \sum_{h=1}^n \pi_{Y|X(x_h)}(c),$$

de donde se obtiene

$$\hat{\pi}_Y(c) = \frac{1}{n} \sum_{h=1}^n \pi_{Y|X(x_h)}(c) \quad c = 1, \dots, C,$$

$$\hat{\pi}_{ic}(s) = \sum_{h=1}^n \frac{\pi_{Y|X(x_h)}(c) x_{ih}(s)}{n \pi_Y(c)} \quad i = 1, \dots, p, \quad c = 1, \dots, C.$$

Sustituyendo los estimadores anteriores en

$$\pi_{Y|X}(c) = \frac{\pi_Y(c) \prod_{i=1}^p \prod_{s=1}^{I_i} \{\pi_{ic}(s)\}^{z_i(s)}}{\pi_X(x_h)},$$



se obtienen los estimadores para las probabilidades a posteriori,

$$\hat{\pi}_{Y|X(x)}(c) = \frac{\hat{\pi}_Y(c) \prod_{i=1}^p \prod_{s=1}^{I_i} \{\hat{\pi}_{ic}(s)\}^{z_i(s)}}{\sum_{c=1}^C \hat{\pi}_Y(c) \prod_{i=1}^p \prod_{s=1}^{I_i} \{\hat{\pi}_{ic}(s)\}^{z_i(s)}}$$

Se puede observar que estas expresiones no proporcionan los parámetros estimados en forma explícita, pues dependen de parámetros desconocidos. Sin embargo, se sugiere un esquema iterativo: asignar valores iniciales a un conjunto de parámetros e insertar estos valores en las ecuaciones correspondientes para obtener los estimadores de los parámetros restantes, estos valores se sustituyen en las expresiones para obtener estimadores mejorados y se repite el proceso hasta que se alcance algún criterio de convergencia. Esta aproximación general se conoce como algoritmo EM.

Cada iteración del algoritmo consiste de dos pasos:

1- El cálculo del valor esperado en el cual los datos faltantes, corresponden a las clases a las cuales pertenecen los individuos, los estimadores de estos valores son las probabilidades a posteriori  $\pi_{Y|X(x_h)}(c)$  son estimados mediante valores actuales de los parámetros.

2- Un proceso de maximización en el que la totalidad de los datos son usados como si ellos fueran directamente observados, para calcular los estimadores mejorados.

La estimación de las probabilidades  $\pi_{Y|X(x_h)}(c)$  constituye el paso *E* del algoritmo y la inserción de estos valores para obtener los estimadores mejorados constituyen el paso *M*

## Bondad de ajuste

Para conocer la bondad de ajuste de los modelos para tablas de contingencia, se comparan las frecuencias observadas para cada patrón de respuesta, con las frecuencias estimadas por el modelo, utilizando la prueba ji-cuadrada ( $X^2$ ) o la prueba del cociente de verosimilitudes ( $G^2$ ).

Las probabilidades esperadas para el patrón de respuesta  $x' = (x_1, \dots, x_p)$  estan dados por

$$\hat{\pi}_X(x) = \sum_{c=1}^C \hat{\pi}_Y(c) \prod_{i=1}^p \hat{\pi}_{X_i|Y(c)}(x_i)$$

y las frecuencias esperadas son

$$\hat{f}_X = n \sum_{c=1}^C \hat{\pi}_Y(c) \prod_{i=1}^p \hat{\pi}_{X_i/Y(c)}(x_i).$$

Luego, se tienen las estadísticas de bondad de ajuste:

$$X^2 = \sum_X \frac{(f_X - \hat{f}_X)^2}{\hat{f}_X} \quad \text{y}$$

$$G^2 = 2 \sum_X f_X \ln \left( \frac{f_X}{\hat{f}_X} \right),$$

donde  $f_X$  representa la frecuencia observada para el patrón de respuesta  $X$ .

Los dos estadísticos siguen la distribución asintótica  $X^2$  con  $\prod_{i=1}^p I_i - m - 1$  grados de libertad, donde  $m$  es el número de parámetros independientes a estimar en el modelo.

Se estiman  $C - 1$  probabilidades para las clases latentes y  $\sum_{i=1}^p (I_i - 1)$  probabilidades condicionales para cada clase latente, así

$$m = (C - 1) + C \sum_{i=1}^p (I_i - 1) = C(\sum_{i=1}^p (I_i - 1) + 1) - 1$$

y los grados de libertad están dados por  $v = \prod_{i=1}^p I_i - C(\sum_{i=1}^p (I_i - 1) + 1)$ .

Si  $X^2 < X_{v;\alpha}^2$  (o  $G^2$ ) el modelo con  $C$  clases proporciona un ajuste adecuado, donde  $X_{v;\alpha}^2$  es el cuantil  $1 - \alpha$  de la  $X^2$  con  $v$  grados de libertad.

El procedimiento descrito supone que el número de clases  $C$  es conocido, pero hay aplicaciones en las cuales no hay valores obvios para el número de clases latentes a considerar, un método propuesto para determinar  $C$ , es ajustar secuencialmente 2,3,... clases, y detener el proceso cuando se alcance un buen ajuste del modelo, el máximo número de clases a considerar está dado por,

$$C_{max} < \frac{\prod_{i=1}^p I_i}{\sum_{i=1}^p (I_i - 1) + 1}.$$

## Tablas de contingencia

Una tabla de contingencia de dos dimensiones se obtiene al clasificar a los individuos de una población con respecto a dos rasgos o características, esta clasificación debe ser exhaustiva y mutuamente exclusiva, es decir, que a cada individuo solamente se le puede asignar una de las  $I \times J$  categorías.

Consideremos una muestra de tamaño  $n$  clasificada según dos características  $A$  y  $B$  con  $I$  y  $J$  niveles, respectivamente. Si representamos por  $n_{ij}$   $i = 1, \dots, I$  y  $j = 1, \dots, J$ , el número de individuos de la muestra que se clasifican simultáneamente en el nivel  $A_i$  de  $A$  y  $B_j$  de  $B$ , la tabla bidimensional que contiene en cada una de sus  $I \times J$  casillas las frecuencias observadas  $n_{ij}$  se llama tabla de contingencia bidimensional y su nombre es debido a Pearson (1904).

La notación general de una tabla de contingencia es la siguiente:

A/B	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	...	B <sub>J</sub>	Totales
A <sub>1</sub>	n <sub>11</sub>	n <sub>12</sub>	n <sub>13</sub>	...	n <sub>1J</sub>	n <sub>1.</sub>
A <sub>2</sub>	n <sub>21</sub>	n <sub>22</sub>	n <sub>23</sub>	...	n <sub>2J</sub>	n <sub>2.</sub>
⋮	⋮	...	...	...	⋮	⋮
A <sub>I</sub>	n <sub>I1</sub>	n <sub>I2</sub>	n <sub>I3</sub>	...	n <sub>IJ</sub>	n <sub>I.</sub>
Totales	n <sub>.1</sub>	n <sub>.2</sub>	n <sub>.3</sub>	...	n <sub>.J</sub>	n

A partir de la tabla de contingencia se obtienen las distribuciones de frecuencias marginales y condicionadas como

Característica  $A$ :

$$n_{i.} = \sum_{j=1}^J n_{ij}, \quad i = 1, \dots, I$$

Característica  $B$ :

$$n_{.j} = \sum_{i=1}^I n_{ij}, \quad j = 1, \dots, J$$

Distribuciones condicionadas:

Característica  $A$  condicionada a  $B = B_j$  para  $j = 1, \dots, J$ .

$$n_{i|j} = \frac{n_{ij}}{n_{.j}}, \quad i = 1, \dots, I$$

Característica  $B$  condicionada a  $A = A_i$  para  $i = 1, \dots, I$ .

$$n_{j|i} = \frac{n_{ij}}{n_{i.}}, \quad j = 1, \dots, J$$

Observemos que para cada característica hay tantas distribuciones de frecuencias condicionadas como niveles tiene la otra característica, es decir se han definido en total  $I + J$  distribuciones condicionadas de frecuencias observadas.

Recordemos que cualquier individuo de la población debe clasificarse solamente en una de las  $I \times J$  casillas de la tabla de contingencia. Si denotamos por  $p_{ij}$  la probabilidad de que un individuo elegido aleatoriamente en la población se clasifique en el nivel  $A_i$  de  $A$  y en el nivel  $B_j$  de  $B$ , se obtiene la distribución de probabilidad discreta que debe cumplir

$$\sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1$$

Para representarla se usa la siguiente tabla que tiene la misma estructura que la primera

A/B	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	...	B <sub>J</sub>	Totales
A <sub>1</sub>	p <sub>11</sub>	p <sub>12</sub>	p <sub>13</sub>	...	n <sub>1J</sub>	p <sub>1.</sub>
A <sub>2</sub>	p <sub>21</sub>	p <sub>22</sub>	p <sub>23</sub>	...	n <sub>2J</sub>	p <sub>2.</sub>
⋮	⋮	...	...	...	⋮	⋮
A <sub>I</sub>	p <sub>I1</sub>	p <sub>I2</sub>	p <sub>I3</sub>	...	n <sub>IJ</sub>	p <sub>I.</sub>
Totales	p <sub>.1</sub>	p <sub>.2</sub>	p <sub>.3</sub>	...	p <sub>.J</sub>	1

De igual forma que para las frecuencias observadas, se definen las distribuciones de probabilidad marginales y condicionales asociadas:

Distribuciones de probabilidad marginales:

Característica A:

$$p_{i.} = \sum_{j=1}^J p_{ij}, \quad i = 1, \dots, I.$$

Característica B:

$$p_{.j} = \sum_{i=1}^I p_{ij}, \quad j = 1, \dots, J.$$

Las distribuciones marginales dan información unidimensional sobre cada variable y no dicen algo sobre la asociación entre las dos variables.

Distribuciones condicionales

Característica  $A$  condicionada a  $B = B_j$  para  $j = 1, \dots, J$

$$p_{i|j} = \frac{p_{ij}}{p_{.j}}, \quad i = 1, \dots, I$$

Característica  $B$  condicionada a  $A = A_i$  para  $i = 1, \dots, I$

$$p_{j|i} = \frac{p_{ij}}{p_{i.}}, \quad j = 1, \dots, J.$$

### Independencia

Dos variables cualitativas son estadísticamente independientes si se verifica:

$$p_{j|i} = p_{.j}, \quad i = 1, \dots, I.$$

Esto significa que las distribuciones condicionales de  $B$  son iguales a su distribución marginal. Es decir, dos variables son independientes cuando la probabilidad de clasificarse en la columna  $j$  es igual en todas las filas.

Intuitivamente la independencia significa que los valores de una variable no vienen influidos por el valor que adopte la otra variable.

De igual forma las variables son independientes si las distribuciones condicionales de  $A$  son iguales que su distribución marginal.

Además, la independencia se caracteriza en términos de la distribución conjunta de probabilidades poblacionales como sigue:  $A$  y  $B$  son independientes si y sólo si las probabilidades conjuntas son igual al producto de las probabilidades marginales, es decir:

$$p_{ij} = p_{i.}p_{.j} \quad \text{para todo } i = 1, \dots, I \text{ y } j = 1, \dots, J.$$

Bajo la independencia se tiene que el estimador máximo verosímil de  $p_{ij}$ ,  $\hat{p}_{ij} = \frac{n_{ij}}{n}$  es la proporción de individuos muestrales clasificados en la casilla  $(i, j)$ .

Entonces, la proporción de veces que un individuo de la fila  $i$  se clasifica en la columna  $j$  es  $\hat{p}_{j|i} = \frac{\hat{p}_{ij}}{\hat{p}_{i.}} = \frac{n_{ij}}{n_{i.}}$ , siendo  $\hat{p}_{i.}$  el estimador máximo verosímil de la proporción muestral de individuos en la fila  $i$ .

Las variables  $A$  y  $B$  son independientes en la muestra si se verifica  $\hat{p}_{j|i} = \hat{p}_{.j}$

o el equivalente  $\hat{p}_{ij} = \hat{p}_{i.}\hat{p}_{.j}$

## Prueba de independencia $X^2$

La prueba  $X^2$  de Pearson se basa en que bajo la hipótesis de independencia las discrepancias entre las frecuencias observadas en la tabla y las esperadas son pequeñas.

Consideremos una tabla de contingencia  $I \times J$  y supongamos que se quiere contrastar la hipótesis nula de independencia:

$H_0: p_{ij} = p_{i.}p_{.j}$  para toda  $i$  y  $j$

Sean  $\hat{m}_{ij} = \frac{n_{i.}n_{.j}}{n}$  para toda  $i$  y  $j$ , los estimadores de máxima verosimilitud de las frecuencias esperadas bajo la hipótesis de independencia. Para llevar a cabo este contraste, Pearson definió el estadístico

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

que, bajo la hipótesis nula de independencia, tiene distribución de probabilidad asintótica

$$X^2_{(I-1)(J-1)}.$$

Se rechaza la hipótesis de independencia al nivel  $\alpha$  cuando se verifica

$$X^2_{OBS} > X^2_{(I-1)(J-1), \alpha}.$$

## Regresión logística multinomial

La regresión logística multinomial permite estimar la relación existente entre una variable dependiente discreta (categórica), que toma dos o más valores y un conjunto de variables independientes continuas o discretas, cuantifica la importancia de la relación que hay entre cada covariable y la variable dependiente. Puede ser utilizada para clasificar a los individuos dentro de las categorías de respuesta, según la probabilidad que tengan de pertenecer a cada una de ellas dado el valor de las covariables.

El modelo consiste en una descripción de la relación entre la variable dependiente  $Y$  y las variables explicativas  $X$ .

## Modelo de regresión logística binaria

Sea  $Y$  una variable dependiente binaria que toma dos valores posibles: 0 y 1 y  $X = (X_1, \dots, X_p)$  un conjunto de variables independientes observadas. Entonces el modelo toma la forma:

$$\pi(X) = \frac{\exp(\beta_0 + \sum_{i=1}^p \beta_i X_i)}{1 + \exp(\beta_0 + \sum_{i=1}^p \beta_i X_i)}$$

Donde  $\pi(X) = E(Y|X)$  la media condicional de una variable  $Y$  dado  $X$ .

Una transformación de  $\pi(X)$  que será central en la regresión logística es la transformación logit. Esta transformación se define en términos de  $\pi(X)$  como sigue:

$$g(X) = \ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \sum_{i=1}^p \beta_i X_i$$

La importancia de esta transformación es que  $g(X)$  tiene muchas de las propiedades del modelo lineal.

## Estimación de los parámetros



Supóngase que se tiene una muestra de  $n$  observaciones independientes  $(X_i, Y_i) \ i = 1, \dots, n$

Cuando  $Y_i = 1$  la contribución para la función de verosimilitud es  $\pi(X_i)$  y cuando  $Y_i = 0$  la contribución es  $1 - \pi(X_i)$ , entonces la función de verosimilitud para la observación  $Y_i$  queda como sigue

$$\pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i}.$$

Si las observaciones son independientes, la función de verosimilitud se obtiene con el producto de estos términos, es decir,

$$l(\beta) = \prod_{i=1}^n \pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i}.$$

Tomando el logaritmo se tiene:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{Y_i \ln[\pi(X_i)] + (1 - Y_i) \ln[1 - \pi(X_i)]\}.$$

Para encontrar los valores de  $\beta$  que maximizan  $l(\beta)$  se deriva la ecuación anterior con respecto a cada  $\beta_i$ , y el conjunto de las expresiones resultantes se iguala a cero.

Así se obtienen las ecuaciones de verosimilitud

$$\sum_{i=1}^n [Y_i - \pi(X_i)] = 0$$

$$\sum_{i=1}^n X_{ik} [Y_i - \pi(X_i)] = 0,$$

$$k = 1, 2, \dots, p$$

La solución para las ecuaciones requiere de métodos iterativos que pueden encontrarse en varios paquetes estadísticos.

Sea  $\hat{\beta}$  la solución de estas ecuaciones, entonces el valor ajustado para el modelo de regresión es  $\hat{\pi}(X_i)$ , esto es, el valor de la expresión en la ecuación

$$\pi(X) = \frac{\exp(\beta_0 + \sum_{i=1}^p \beta_i X_i)}{1 + \exp(\beta_0 + \sum_{i=1}^p \beta_i X_i)} \text{ obtenido con } \hat{\beta} \text{ y } X_i.$$

En el modelo de regresión logística multinomial la variable respuesta tiene más de dos categorías, es decir,  $Y$  puede ser codificada como  $0, 1, 2, \dots, l$ , para un modelo multinomial tendríamos  $l - 1$  funciones logit, para el desarrollo de este modelo asumiremos que se tienen  $p$  covariables y un término constante, denotado como  $X = (X_0, X_1, \dots, X_p)$  donde  $X_0 = 1$ , y denotamos las funciones logit como sigue:

$$g_j(X) = \ln \left[ \frac{P(Y=j|X)}{P(Y=0|X)} \right] = \beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2 + \dots + \beta_{jp}X_p = X\beta'_j, \text{ para } j = 0, 1, \dots, l,$$

donde  $X$  es un vector de covariables de dimensión  $p$  y  $\beta_j$  es un vector de dimensión  $p$ . La categoría etiquetada como 0 se denomina categoría base o categoría de referencia.

Entonces, se tiene que las probabilidades condicionales para cada categoría de respuesta dado el vector de covariables  $X$  están dadas por

$$\pi_j(X) = P(Y = j|X) = \frac{\exp\{g_j(X)\}}{1 + \sum_{j=1}^l \exp\{g_j(X)\}} = \frac{\exp\{g_j(X)\}}{\sum_{j=0}^l \exp\{g_j(X)\}}$$

para cada  $j = 1, \dots, l$ ,  $\beta_0 = 0$ , y  $g_0(X) = 0$ .

De tal manera que la función de máxima verosimilitud es construída de  $l$  variables indicadoras  $Y$  codificadas como cero o uno, para indicar el grupo de miembros de una observación. Estas variables se introducen únicamente para aclarar la función de verosimilitud y no son usadas en el análisis de regresión logística multinomial. Además  $\sum_{j=0}^l Y_j = 1$ , de tal manera que la función de verosimilitud condicional para una muestra de  $n$  valores independientes queda como

$$l(\beta) = \prod_{i=1}^n \left[ \prod_{j=0}^l \pi_j(X)^{Y_{ji}} \right],$$

tomando el logaritmo de  $l(\beta)$  y usando la restricción  $\sum Y_{ji} = 1$  para cada  $i$ , la función log-  
verosimilitud condicional está dada por,

$$L(\beta) = \sum_{i=1}^n \sum_{j=1}^l Y_{ji} \left( \sum_{k=0}^p \beta_{jk}(X_{ki}) \right) - \ln \left\{ 1 + \sum_{j=1}^l \exp \left\{ \sum_{k=0}^p \beta_{jk}(X_{ki}) \right\} \right\}$$

donde  $X_{0i} = 1$ .

Los parámetros se obtienen calculando la primera derivada parcial de  $L(\beta)$  con respecto a cada uno de los parámetros desconocidos.

Para simplificar la notación se escribe  $\pi_{ji} = \pi_j(X_i)$ .

Entonces se tiene que:

$$\frac{\partial L(\beta)}{\partial \beta_{jk}} = \sum_{i=1}^n X_{ki} (Y_{ji} - \pi_{ji})$$

para  $j = 1, \dots, l$  y  $k = 0, 1, 2, \dots, p$  con  $X_{0i} = 1$ .

Los estimadores de máxima verosimilitud  $\hat{\beta}$  se obtienen igualando la ecuación anterior a cero y resolviendo el conjunto de ecuaciones resultante mediante algún procedimiento iterativo.

Los estimadores de las varianzas y covarianzas de los parámetros son obtenidos de la matriz de segundas derivadas parciales del logaritmo de la función de verosimilitud. Esas derivadas parciales tienen la siguiente forma general:

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial \beta_{j'k'}} = - \sum_{i=1}^n X_{k'i} X_{ki} \pi_{ji} (1 - \pi_{ji}) \quad \forall$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial \beta_{j'k'}} = - \sum_{i=1}^n X_{k'i} X_{ki} \pi_{ji} \pi_{j'i}$$

Sea  $I(\hat{\beta})$  la matriz de  $(r + 1) \times (r + 1)$  que contiene los términos negativos de las ecuaciones anteriores, entonces se tiene que la matriz de información de varianzas y covarianzas está dada por la inversa, es decir,  $\Sigma(\hat{\beta}) = I^{-1}(\hat{\beta})$

$$\text{Sea } X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1r} \\ 1 & X_{21} & \cdots & X_{2r} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & X_{nl} & \cdots & X_{nr} \end{bmatrix} \text{ y}$$

$$V = \begin{bmatrix} \hat{\pi}_{11}(1 - \hat{\pi}_{11}) & 0 & \cdots & 0 \\ 0 & \hat{\pi}_{11}(1 - \hat{\pi}_{11}) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \hat{\pi}_{11}(1 - \hat{\pi}_{11}) \end{bmatrix},$$

entonces se puede estimar la matriz de información del modelo estimado como  $\hat{I}(\hat{\beta}) = X'VX$ .

Una vez que se han estimado los coeficientes del modelo, se verifica que el modelo predice de manera adecuada a la variable dependiente en un nuevo individuo relacionado con la muestra, donde los valores de las variables explicativas son las probabilidades estimadas para los distintos valores de  $Y$ . Para lo cual se formula y se prueba la hipótesis estadística que determina si las variables independientes influyen significativamente en la probabilidad del suceso del modelo relacionado a la variable de respuesta.

### Bondad de Ajuste

Para enfatizar el hecho de que los valores ajustados en la regresión logística son calculados para cada patrón de covariables, y dependen de la probabilidad estimada para cada uno de ellos, se denota al valor ajustado  $\hat{Y}_j$  como

$$\hat{Y}_j = m_j \hat{\pi}_j = m_j \left( \frac{\exp[\hat{g}(x_j)]}{1 + \exp[\hat{g}(x_j)]} \right), \text{ donde } m_j \text{ es el número de patrones con } X = X_j \text{ y } \hat{g}(x_j) \text{ es el logit estimado.}$$

Se iniciará considerando dos medidas de la diferencia entre los valores ajustados y los observados: los residuales de Pearson y los residuales de la devianza. Para un particular patrón de covariables el residual de Pearson se define como

$$r(Y_j, \hat{\pi}_j) = \frac{(Y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}}$$

La estadística resumen basada en esos residuales es la estadística  $X^2$  de Pearson, la cual se calcula como se muestra a continuación

$$X^2 = \sum_{j=1}^J r(Y_j, \hat{\pi}_j)^2.$$

Por su parte, los residuales de la devianza se definen como

$$d(Y_j, \hat{\pi}_j) = \pm \left\{ 2 \left[ Y_j \ln \left( \frac{Y_j}{m_j \hat{\pi}_j} \right) + (m_j - Y_j) \ln \left( \frac{m_j - Y_j}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{1/2}$$

donde el signo + o - es el mismo signo de  $(Y_j - m_j \hat{\pi}_j)$ . Para las covariables con  $Y_j = 0$  el residual de la devianza es

$$d(Y_j, \hat{\pi}_j) = -\sqrt{2m_j |\log(1 - \hat{\pi}_j)|}$$

y cuando  $Y_j = m_j$ , el residual es

$$d(Y_j, \hat{\pi}_j) = \sqrt{2m_j |\log(\hat{\pi}_j)|}.$$

La estadística resumen basada en los residuales de la devianza se conoce como devianza y se calcula de la siguiente forma

$$D = \sum_{j=1}^J d(Y_j, \hat{\pi}_j)^2.$$

El valor de la devianza decrece conforme el número de parámetros aumenta, esto no sucede necesariamente con la  $X^2$  de Pearson.

Cuando los datos están agrupados y el modelo logístico es correcto, la distribución asintótica de la  $X^2$  y la devianza, son una  $X^2$  con  $n - m$  grados de libertad. Empíricamente, se considera que el modelo logístico se ajusta a un conjunto de datos no agrupados si la  $X^2$  de Pearson generalizada y la devianza son aproximadamente iguales a  $n - m$ .

### Razón de verosimilitud

Para verificar la significancia del coeficiente estimado para algunas variables en cuestión, se comparan los valores observados de la variable respuesta y los valores obtenidos de los modelos con y sin las variables, una manera de hacerlo consiste en comparar los valores observados y los predichos en base al logaritmo de la función de verosimilitud.

Para asegurar la significancia de una variable independiente se utilizará la siguiente estadística:

$$G = -2 \ln \left( \frac{\text{verosimilitud sin la variable}}{\text{verosimilitud con la variable}} \right)$$

Para el caso específico de una sola variable independiente es fácil mostrar que cuando la variable no está en el modelo, el estimador de máxima verosimilitud de  $\beta_0$  es:

$$\ln \left( \frac{n_1}{n_0} \right)$$

donde  $n_1 = \sum Y_i$  y  $n_0 = \sum (1 - Y_i)$  y el valor predicho es constante  $\frac{n_1}{n}$ .

En este caso el valor de la verosimilitud sin la variable es:

$$\left( \frac{n_1}{n} \right)^{n_1} \left( \frac{n_0}{n} \right)^{n_0}$$

y el valor de  $G$  queda determinado como sigue:

$$G = -2 \ln \left[ \frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{Y_i} (1 - \hat{\pi}_i)^{(1-Y_i)}} \right]$$

$$o \quad G = 2 \{ \sum_{i=1}^n [Y_i \ln(\hat{\pi}_i) + (1 - Y_i) \ln(1 - \hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \}$$

Bajo la hipótesis de que  $\beta_1 = 0$  la estadística  $G$  sigue una distribución  $X^2$  con un grado de libertad.

La regla de decisión es rechazar  $H_0: \beta_1 = 0$  con un nivel de significancia  $\alpha$  si  $G > X_{(1-\alpha,1)}^2$

Si se rechaza la hipótesis nula se contará con evidencia convincente de que la variable independiente es significativa en la predicción del resultado. Otra prueba equivalente, es la prueba de Wald.

### Prueba de Wald

Evalúa la significancia de los coeficientes se define como el vector de los coeficientes estimados del siguiente modo según las hipótesis:

$$H_0: \beta_j = 0 \text{ para toda } j = 1, \dots, l$$

$$H_1: \beta_j \neq 0 \text{ para alguna } j = 1, \dots, l$$

Estadístico de prueba:

$$W = \hat{\beta}' [I(\hat{\beta})]^{-1} \hat{\beta} = \hat{\beta}' (X' V X) \hat{\beta}: X_{\alpha, k+1}^2.$$

Si  $W > X_{\alpha, k}^2$  rechazamos  $H_0$  con un nivel de significancia fijo  $\alpha$  y se concluye que la variable independiente influye en la probabilidad del suceso.

## Interpretación del modelo logístico

La interpretación de un modelo ajustado requiere la capacidad de extraer inferencias prácticas de los coeficientes estimados en el mismo. La cuestión fundamental es ¿qué indican los coeficientes estimados en el modelo acerca de las preguntas de investigación que motivaron el estudio? La respuesta debe basarse en el hecho de que los coeficientes estimados para las variables independientes representan la intensidad de cambio de la función de la variable dependiente por unidad de cambio en la variable independiente. Así, la interpretación involucra dos aspectos: determinar la relación funcional entre la variable dependiente y las variables independientes y una definición apropiada de unidad de cambio para la variable independiente.

En el caso del modelo logístico dado por

$$g(X) = \ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \sum_{i=1}^p \beta_i X_i.$$

El coeficiente  $\beta_j$  representa el cambio en el logit resultante del cambio de una unidad en la variable independiente  $X$  así que una interpretación conveniente del modelo de regresión logística depende de la capacidad de dar significado a la diferencia entre los logit, para dos valores diferentes de la variable  $X_j$  manteniendo todas las demás constantes.

Se iniciarán las consideraciones de interpretación de los coeficientes de regresión logística con la situación donde la variable independiente es dicotómica.

Para el caso de una covariable con más de dos categorías, se expande el número de cocientes de momios para hacer comparaciones de cada categoría de la covariable para cada posible función logit.

Supóngase que  $X$  es codificada como 0 ó 1. Bajo este modelo hay sólo dos valores de  $\pi(X)$  y de forma equivalente dos valores de  $1 - \pi(X)$ , suponiendo que todas las demás variables permanecen constantes, esos valores pueden situarse en forma conveniente en una tabla de  $2 \times 2$  como se muestra a continuación:



	$X = 1$	$X = 0$
$Y = 1$	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
$Y = 0$	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
<i>Total</i>	1	1

Los momios del resultado para los individuos con  $X = 1$  se definen como  $\frac{\pi(1)}{1 - \pi(1)}$  y de forma similar, los momios del resultado para los individuos con  $X = 0$  se definen como  $\frac{\pi(0)}{1 - \pi(0)}$ .

El logaritmo de los momios, como se definieron previamente, es llamado el logit y en este caso se tiene que

$$g(1) = \ln \left[ \frac{\pi(1)}{1 - \pi(1)} \right] \text{ y } g(0) = \ln \left[ \frac{\pi(0)}{1 - \pi(0)} \right].$$

La razón de momios se define como el cociente entre los dos momios para  $X = 1$  y  $X = 0$  como se muestra a continuación:

$$\Psi = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}}$$

Ahora, usando las expresiones para el modelo de regresión logística mostrados en el cuadro se tiene que

$$\Psi = \frac{\left( \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \right) \left( \frac{1}{1 + e^{\beta_0}} \right)}{\left( \frac{e^{\beta_0}}{1 + e^{\beta_0}} \right) \left( \frac{1}{1 + e^{\beta_0 + \beta_1}} \right)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}.$$

Es decir, que para la regresión logística con una variable independiente dicotómica  $\Psi = e^{\beta_1}$  y la diferencia de logits es  $\ln(\Psi) = \beta_1$ .

La razón de momios es una medida de asociación, la cual ha encontrado un uso amplio, indica qué tanto más probable es que el resultado de interés se encuentre presente entre los individuos para los cuales  $X = 1$  que entre aquellos para quienes  $X = 0$ .

La interpretación dada para la razón de momios esta basada en el hecho de que en muchas ocasiones se aproxima a una cantidad llamada riesgo relativo o razón de oportunidades.

Para simplificar la discusión de la interpretación de los cocientes de momios en el modelo multinomial, se generaliza la notación usada en el caso binario. Asumiremos que la respuesta  $Y = 0$  es la respuesta de referencia. El subíndice sobre el cociente de momios indica cual es la respuesta de referencia.

Los cocientes de momios de la respuesta  $Y = j$  versus  $Y = 0$  para los valores de  $X = a$  versus  $X = b$  son:

$$OR_j(a, b) = \frac{P(Y=j|X=a)/P(Y=0|X=a)}{P(Y=j|X=b)/P(Y=0|X=b)}.$$

Cuando la variable respuesta tiene  $l$  categorías, se tienen  $l - 1$  funciones logit, y se definen estas funciones de tal forma que los  $l - 1$  coeficientes estimados son iguales a los logaritmos de los cocientes de momios de los pares de las tablas de  $2 \times 2$  obtenidas de cruzar  $Y = j$  y  $Y = 0$ , con  $Y = 0$  como la respuesta de referencia.

La interpretación para cada uno de los cocientes de momios es igual a la interpretación del caso binario.

Para covariables continuas, la interpretación depende de como se introducen en el modelo y de las unidades particulares de la variable, el método para interpretar el coeficiente para una variable continua es como el de un modelo lineal. Bajo la suposición de que el logit es lineal en la covariable continua  $X$ , la ecuación para el logit es

$$g(X) = \beta_0 + \beta_1 X,$$

de aquí, se sigue que  $\beta_1$  da el cambio en el logaritmo del momio para un incremento de una unidad en  $X$ , es decir,  $\beta_1 = g(X + 1) - g(X)$  para cualquier valor de  $X$ .

El logaritmo del momio para un cambio de  $c$  unidades en  $X$  se obtiene de la diferencia

$c\beta_1 = g(X + c) - g(X)$  y el cociente de momios es

$$OR(c) = OR(X + c, X) = \exp(c\beta_1).$$

## Observaciones

Se ha visto que el modelo de clases latentes consiste de dos partes: la distribución a priori de las variables latentes, representada como la función de densidad  $h(y)$  y el conjunto de las distribuciones condicionales de las variables manifiestas  $\{x_i\}$  dadas las variables latentes, denotada por  $g_i(x_i|y)$  para  $i = 1, \dots, p$ .

Si suponemos que las variables latentes se combinan para determinar el valor de un parámetro de la familia exponencial, las distribuciones condicionales se pueden escribir como:

$$g_i(x_i|\theta_i) = F_i(x_i)G_i(\theta_i)\exp(\theta_i u_i(x_i)) \text{ para } i = 1, \dots, p.$$

Donde  $\theta_i$  es alguna función de  $y$  (puede ser una función lineal  $\theta_i = \alpha_{i0} + \alpha_{i1}y_1 + \dots + \alpha_{iq}y_q$ ) para  $i = 1, \dots, p$ .

Ejemplo: supongamos que  $x_i$  es una variable aleatoria Bernoulli, tal que se codifica al 1 como éxito y 0 como fracaso. Entonces la distribución condicional de  $x_i$  está dada por:

$$g(x_i|y) = \pi_i^{x_i}(1 - \pi_i)^{1-x_i}, \quad (x_i = 0,1)$$

$g(x_i|y) = (1 - \pi_i)\exp\{x_i \text{logit}(\pi_i)\}$ , donde  $\pi_i$  es la probabilidad de éxito.

Si  $g_i(x_i|\theta_i) = F_i(x_i)G_i(\theta_i)\exp(\theta_i u_i(x_i))$  entonces  $\theta_i = \text{logit}(\pi_i) = \ln\{\pi_i/(1 - \pi_i)\}$

$$G_i(\theta_i) = 1 - \pi_i, \quad u_i(x_i) = 1, \quad F_i(x_i) = 1,$$

para este caso el modelo de clases latentes, es por lo tanto

$$\text{logit } \pi_i(y) = \alpha_{i0} + \sum_{j=1}^q \alpha_{ij}y_j$$

y los componentes están dados por  $X_j = \sum_{i=1}^p \alpha_{ij} x_i$  el logit aquí es conocido como la función liga, en el caso especial de  $q = 1$  se reduce al modelo logístico  $\pi_i(y) = \frac{1}{1 + \exp(-\alpha_{i0} - \alpha_{i1}y)}$ .

Para el caso general, se denota a  $c_i$  como el número de categorías de una variable  $i$  las cuales se clasifican como  $0, 1, \dots, c_i - 1$  para  $i = 1, \dots, p$  e indexadas por  $s$ , y  $z_i(s)$  la función indicadora tal que:

$$z_i(s) = \begin{cases} 1 & \text{si la respuesta pertenece a la categoría } s \\ 0 & \text{en otro caso} \end{cases}, \text{ con } \sum_s z_i(s) = 1$$

y sea  $x' = (x_1, \dots, x_p)$  un patrón de respuesta en el cual cada una de las  $x_i$  toma distintos valores dependiendo de las categorías de la correspondiente variable manifiesta.

Se denota  $P\{x_i(s) = 1\} = \pi_{is}(y)$ , ( $s = 0, 1, \dots, c_i - 1$ ;  $i = 1, \dots, p$ ).

Por lo que la función de probabilidad condicional de  $x_i$  dado  $y$  está dada por

$$g_i(x_i|y) = \prod_{s=0}^{c_i-1} \{\pi_{is}(y)\}^{z_i(s)}$$

y la función de densidad de  $y$  esta dada por:

$$h(y|x) = \frac{g(x|y)h(y)}{f(x)} = \frac{h(y) \prod_{i=1}^p \prod_{s=0}^{c_i-1} \{\pi_{is}(y)\}^{z_i(s)}}{f(x)} = \frac{h(y) \exp\left[\sum_{i=1}^p \sum_{s=1}^{c_i-1} z_i(s) \ln \pi_{is}(y)\right]}{f(x)}.$$

Sea  $\theta_i = \text{logit } \pi_i(y)$ , entonces

$$g_i(z_i|y) = \prod_{s=0}^{c_i-1} \{\pi_{i0}(y)\}^{z_i(s)} \{\pi_{is}(y)/\pi_{i0}(y)\}^{z_i(s)}$$

$$g_i(z_i|y) = \pi_{i0}(y) \exp \sum_{s=0}^{c_i-1} z_i(s) \ln \{\pi_{is}(y)/\pi_{i0}(y)\}$$

$$g_i(z_i|y) = \pi_{i0}(y) \exp \theta_i' x_i$$

$$\text{Donde } \theta_i' = \left\{ 0, \ln \frac{\pi_{i1}(y)}{\pi_{i0}(y)}, \ln \frac{\pi_{i2}(y)}{\pi_{i0}(y)}, \dots, \ln \frac{\pi_{ic_{i-1}}(y)}{\pi_{i0}(y)} \right\}$$

se puede escribir como  $\theta_i = \alpha_{i0} + \alpha_{i1}y_1 + \dots + \alpha_{iq}y_q$ ; donde  $\alpha'_{ij} = (\alpha_{ij}(0) = 0, \alpha_{ij}(1), \dots, \alpha_{ij}(c_i - 1))$  para  $j = 1, \dots, q$ .

Los componentes están dados por  $X_j = \sum_{i=1}^p \sum_{s=0}^{c_i-1} \alpha_{ij}(s) z_i(s)$

Como  $\sum_s \pi_{is}(y) = 1$  y  $g_i(z_i|y) = \pi_{i0}(y) \exp \theta_i' x_i$ , entonces se tiene que:

$$\pi_{is}(y) = \frac{\exp\{\alpha_{i0}(s) + \sum_{j=1}^q \alpha_{ij}(s)y_j\}}{\sum_{r=0}^{c_i-1} \exp\{\alpha_{i0}(r) + \sum_{j=1}^q \alpha_{ij}(r)y_j\}} \quad \text{para } i = 1, 2, \dots, p; \quad s = 0, 1, \dots, c_i - 1. \text{ El cual corresponde}$$

a un modelo logístico multinomial, de esta forma las clases de una variable latente, se pueden ver como categorías de la variable y puede ser explicada a través de otras variables manifiestas.

Por otro lado, es posible definir un modelo de clases latentes con covariables, sea  $X_i$  las covariables observadas de un individuo  $i$ , tomando la primera clase como referencia, asumiendo que el cociente momios de la clase de pertenencia con respecto a esa clase es una combinación lineal de las covariables y sea  $\beta_r$  el vector de coeficientes correspondientes a la  $r$  -ésima clase latente, con  $r = 1, \dots, R$ , para  $S$  covariables, este vector tendría una dimensión de  $S + 1$ ; es decir, un coeficiente para cada covariable más una constante, porque la primera clase se toma como referencia y  $\beta_1 = 0$ .

Entonces se tiene que:

$$\ln \left( \frac{\pi_{2i}}{\pi_{1i}} \right) = X_i \beta_2$$

$$\ln \left( \frac{\pi_{3i}}{\pi_{1i}} \right) = X_i \beta_3$$

⋮

$$\ln \left( \frac{\pi_{Ri}}{\pi_{1i}} \right) = X_i \beta_R$$

Lo que implica que  $\pi(X_i|\beta) = \frac{\exp(X_i \beta_r)}{\sum_{q=1}^R \exp(X_i \beta_q)}$  lo cual corresponde a un modelo logístico multinomial.

Otra observación importante, es que dentro de los análisis de clases latentes no se han considerado los diseños muestrales, o los factores de expansión para muestras no autoponderadas, donde a cada individuo se le asigna un peso diferente. Patterson, Dayton y Graubard<sup>6</sup> proponen incluir los factores en el modelo al momento de obtener los parámetros, incluyéndolos en la función log-verosimilitud, como sigue:

$$L = \sum_{h=1}^n w_h \log \pi_X(x_h) = \sum_{h=1}^n \log \left( \sum_{c=1}^C w_h \pi_Y(c) \prod_{i=1}^p \prod_{s=1}^{I_i} \{\pi_{ic}(s)\}^{z_i(s)} \right)$$

Donde cada  $w_h$  es el correspondiente factor de expansión.

Para estimar la varianza proponen el método de jackknife:

Sea  $\hat{\gamma}$ , un estimador de  $\gamma$  basado en una muestra de tamaño  $n$ , para un diseño estratificado por conglomerados, con  $H$  estratos y  $k_h$  el número de conglomerados en cada estrato  $h$ ,  $h = 1, \dots, H$ . La aplicación del método jackknife consiste en ir eliminando conglomerados  $s$ , primero se calcula  $\hat{\gamma}_{(sh)}$  pero se excluye el conglomerado  $s$  del estrato  $h$ , y se obtiene el estimador del estrato promediando sobre todos los  $\hat{\gamma}_{(sh)}$ ,  $\hat{\gamma}_{.h} = \sum_{s=1}^{k_h} \frac{\hat{\gamma}_{(sh)}}{k_h}$ , por lo que el estimador global está dado por  $\hat{\gamma}_{..} = \sum_{h=1}^H \sum_{s=1}^{k_h} \frac{\hat{\gamma}_{(sh)}}{n}$ , así el promedio de las estimaciones de los estratos es  $\bar{\gamma}_{..} = \sum_{h=1}^H \frac{\hat{\gamma}_{.h}}{H}$ , posteriormente se obtienen los pseudovalores:

$$\hat{\gamma}_{sh} = (L(k_h/(k_h - 1)) + 1)\hat{\gamma} - L(k_h/(k_h - 1))\hat{\gamma}_{(sh)}, \text{ quedando el estimador jackknife como:}$$

$$\hat{\gamma}' = \sum_{h=1}^H \sum_{s=1}^{k_h} \frac{\hat{\gamma}_{sh}}{Hk_h}$$

$$\text{y el estimador de la varianza como } \widehat{var}(\hat{\gamma}') = \sum_{h=1}^H \left[ \sum_{s=1}^{k_h} \frac{k_h-1}{k_h} (\hat{\gamma}_{sh} - \hat{\gamma}_{.h})^2 \right].$$

Cabe mencionar, que se requieren tantas iteraciones jackknife como sea el total de conglomerados.

---

<sup>6</sup> Patterson, Blossom H., Dayton Mitchell, Graubard Barry I., Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data, Journal of the American Statistical Association, September 2002, Vol. 97, No. 459, Applications and Case Studies.

# Capítulo 4

---

## Análisis y Resultados

La encuesta sobre Percepción de la Ciencia y Conocimiento de la UNAM, se realizó en el mes de julio de 2008, de forma domiciliaria, se aplicaron 1260 cuestionarios en viviendas a personas de 15 años y más que residían en la Zona Metropolitana de la Ciudad de México.

## Variables

Las variables consideradas en la encuesta se dividen en tres grupos:

### 1. Variables de percepción de la ciencia

#### **Representación social de la ciencia:**

¿Qué cree usted que es la ciencia?

¿Cuáles de las siguientes frases considera que expresa mejor la idea de ciencia?

#### **Imagen del científico:**

¿Cuál es la imagen que tiene de la profesión de científico en México? Diría que es una profesión...

- Atractiva para los jóvenes
- Gratificante en lo personal
- Remunerada económicamente
- Reconocimiento social

**Motivaciones del investigador, lugar e importancia del trabajo del científico:**

¿Cuáles cree usted que son, en general, las principales motivaciones que tiene un investigador para dedicarse a la ciencia en México?

¿Cuál cree que es el lugar de trabajo de la mayoría de los científicos en nuestro país?

¿Usted cree que el trabajo científico para el país es: muy importante, importante, poco importante o nada importante?

**Financiamiento y valoración del trabajo del científico:**

¿Quién piensa que aporta más dinero para la investigación científica en el país?

¿Usted cree que el trabajo del científico en México es apoyado por el gobierno?

Dígame por favor, si está totalmente de acuerdo, de acuerdo, en desacuerdo, o totalmente en desacuerdo con la siguiente afirmación: “El gobierno debe aumentar los recursos que destina a la investigación científica en nuestro país”.

¿Usted cree que en México el trabajo del científico es valorado por la sociedad?

**2. Variables de percepción de la UNAM**

**Conocimiento y vínculos con la UNAM:**

¿Sabe usted qué significa UNAM?

¿Usted estudia (o estudió) en la UNAM?, Trabaja (o ha trabajado) en la UNAM?

¿Usted conoce a alguien que estudie (o haya estudiado) en la UNAM? o trabaje (o haya trabajado) en la UNAM?

**El consumo de información sobre ciencia que se desarrolla en la UNAM:**

La UNAM tiene diversos medios de comunicación a través de los cuales divulga la investigación científica que ahí se realiza. De la siguiente lista ¿conoce alguno de ellos?

Radio UNAM

TV UNAM

Revista digital universitaria

Gaceta UNAM



3. **Variables sociodemográficas de la población objeto de estudio:** sexo, edad, nivel de escolaridad, condición de actividad.

## Resultados de frecuencias simples

### Perfil Sociodemográfico y socioeconómico de la población

Del total de personas encuestadas, 51% son hombres y 49% mujeres. Se contó con la siguiente distribución para los grupos de edad: entre 15 y 24 años (39.8%), entre 25 y 34 años (25.8%), entre 35 y 44 años (16.9%) y de 45 años y más (17.5%). En relación al alfabetismo se encontró que 98.7% de la población sabe leer y escribir y el porcentaje restante manifestó que no sabía. Por otro lado, se contó con la siguiente distribución en los niveles de escolaridad: sin escolaridad (6.2%), primaria (15.8%), secundaria (44.1%), preparatoria o bachillerato (14.7%), licenciatura o más (19.1%). Por último, se encontró que el 50.7% de los encuestados trabajan.

### Percepción de la ciencia

Las concepciones recabadas sobre lo que significa la ciencia, dan una noción del desconocimiento que existe alrededor del tema, pues pocos son los casos en los que se maneja una definición en la que se planteen conceptos como el método científico, la observación, u otros que podrían ser definiciones de la ciencia; las personas generalmente asocian el concepto con el estudio (28.1%), con la investigación (15%), y con el conocimiento en todas sus ramas (10.2%). Además de que el 8.4% de la población declaró desconocer por completo lo que significa la ciencia.

La imagen con la que más se asocia a la ciencia es la de “grandes descubrimientos” (39.3%), seguida de “mejora de la vida humana” (22.5%), y “avance técnico” (18.3%).

### **Principales motivadores del científico**

El principal motivador del científico es la “búsqueda de nuevos conocimientos” (54.5%). En segundo lugar se tiene la idea de “ayudar a solucionar problemas de la gente y de la sociedad” (20.6%), y en tercer plano “vocación por el conocimiento” (12%). Esto refleja que el conocimiento es lo que lleva a los individuos a dedicarse a la ciencia.

### **Valoración de la profesión del científico**

La profesión del científico se percibe como “poco atractiva para los jóvenes” (57.9%), “mal remunerada económicamente” (55.6%), con escaso reconocimiento social (64.5%), pero “muy gratificante en lo personal” (68.3%).

Se reconoce el trabajo del científico como “Muy importante” (63.5%) e “Importante” (28.5%)

### **La investigación científica en México**

Sólo el 47.5% de la población señaló conocer alguna institución científica, entre las que destacaron la Universidad Nacional Autónoma de México (UNAM) y el Instituto Politécnico Nacional (IPN).

Se reconoce a los laboratorios como el lugar de trabajo de los científicos (33.1%), seguido de las universidades (30.3%), y las empresas (7.6%). El (3.3%) cree que los científicos trabajan en el extranjero, considerando la “fuga de talentos” por las pocas opciones que existen en el país.

Las Universidades son consideradas como el principal aporte de recursos financieros para el trabajo científico (28.9%), seguidas del gobierno (22.1%) y de las instituciones extranjeras (19.2%).

El 64.5% de la población respondió que el trabajo del científico en México no es apoyado por el gobierno, además de considerar que el gobierno debe aumentar los recursos que destina a la investigación científica en nuestro país (97.4%).

La población reconoce que el trabajo del científico en México no es valorado por la sociedad (71.7%).

### **Percepción de la UNAM**

Las universidades públicas que más se conocen entre la población son: Universidad Nacional Autónoma de México (40.4%), Instituto Politécnico Nacional (27.7%) y la Universidad Autónoma Metropolitana (16.9%).

A pesar de señalar a la UNAM como una de las universidades más reconocidas del país y como una de las instituciones dedicadas al quehacer científico más sobresalientes, la mitad de la población desconoce el significado correcto de sus siglas (51.1%).

16.6% de la población reconoce tener un vínculo directo con la Universidad (de estudio o trabajo) y 69.1% mencionó tener un vínculo indirecto, es decir que conoce a alguien que tiene un vínculo directo con la Universidad.

Se percibe a la Universidad Nacional Autónoma de México como centro de enseñanza (76.4%), dejando en segundo plano su relevancia en investigación (35.9%), en ciencia (27.7%), y en la cultura (22.1%).

Entre las actividades científicas que más se reconocen por la población están las Ciencias Físico-Matemáticas y las Ingenierías (44%) y las ciencias biológicas y de la salud (38.8%), mientras que las Ciencias Sociales, Humanidades y de las Artes fueron mencionadas sólo por el 4.9% de la población.

La población reconoció que la UNAM destaca mucho en Educación (80.4%), Investigación científica (76.3%), Salud (73.3%), Desarrollo de tecnologías (66.1%).

En general la UNAM se reconoce como un factor positivo para la sociedad y se asevera que aporta más ventajas que desventajas en “El desarrollo económico” (90.7%), “La calidad de vida en la sociedad” (90.3%), “La seguridad y protección de la vida humana” (85.5%), “La conservación del medio ambiente y la naturaleza” (88.1%), “Hacer frente a las enfermedades y epidemias” (94%),

“Los productos de alimentación y producción agrícola” (81.9%), “La generación de puestos de trabajo” (72.2), y “El incremento y mejora de las relaciones entre las personas” (73%).

El 66.1% de la población se informa sobre temas de ciencia que se desarrollan en la UNAM, si consideramos que de ese porcentaje el 56.9% lo hace a través de la televisión y primordialmente por noticieros y canales de televisión abierta. La población no busca temas relacionados con la ciencia, por lo que los trabajos científicos tienen poca resonancia en la sociedad.

Los medios de comunicación más reconocidos de la UNAM son: Radio UNAM (44.8%), TV UNAM (41.3%) y Gaceta UNAM (39.1%).

## Resultados de los modelos

Con la finalidad de clasificar a los individuos de acuerdo con las preguntas de las variables correspondientes a la percepción de la ciencia y conocimiento de la UNAM, se realiza un modelo de clases latentes para cada grupo de preguntas, utilizando la función polCA del programa R<sup>7</sup>. Una vez que se han clasificado en un número de clases obtenido por cada grupo, se busca determinar si existe relación entre la clase de pertenencia y las variables sociodemográficas, empleando un análisis de tablas de contingencia con la prueba  $X^2$  para verificar independencia y posteriormente se realiza el modelo logístico multinomial, utilizando el paquete SPSS.

Cabe mencionar que los modelos de clases latentes se realizaron sólo a nivel muestra, es decir, que no se consideró el esquema de muestreo, ya que la función polCA, no tiene la opción para incluir los factores de expansión, por lo que estos resultados son sólo a nivel muestra. No obstante, los modelos de regresión logística multinomial, se realizaron tanto a nivel muestra, como con factores de expansión.

---

<sup>7</sup> La descripción de la función polCA y los resultados de los modelos, se pueden consultar en el anexo.

## Modelo I

### Representación social de la ciencia:

¿Qué cree usted que es la ciencia?

¿Cuáles de las siguientes frases considera que expresa mejor la idea de ciencia?

Principales respuestas por cada clase obtenidas del modelo de clases latentes:

Clase 1:

En la clase 1, las personas definen la ciencia como investigar, estudiar, conocimiento y otros, estas respuestas acumulan 0.7168 de probabilidad.

En la pregunta sobre las frases que asocian a la idea de ciencia, las principales respuestas para esta clase son grandes descubrimientos y mejora de la vida humana, estas respuestas acumulan 0.7957 de probabilidad.

Clase 2:

En la clase 2, las personas definen ciencia como tecnología, otros y no responde, acumulando 0.7911 de probabilidad.

En la pregunta sobre las frases asociadas a la idea de ciencia, las principales respuestas en esta clase son grandes descubrimientos y avance técnico, acumulando 0.8675 de probabilidad.

Clase 3:

En la clase 3, las personas definen la ciencia como estudiar y otros, acumulando 0.6681 de probabilidad.

En la pregunta sobre las frases asociadas a la idea de ciencia, respondieron comprensión del mundo natural y avance técnico, acumulando 0.5707 de probabilidad.

Para este modelo, se obtuvo la estadística  $X^2 = 54.03$  la cual es menor que  $X_{66,05}^2 = 85.96$ , por lo que se considera que el modelo se ajusta adecuadamente.

La distribución de la población por cada clase se muestra en el siguiente cuadro:

	Frecuencia	%
Clase 1	929	74
Clase 2	155	12.3
Clase 3	172	13.7
Total	1256	100

Tablas de contingencia y prueba  $X^2$  para el modelo I

Clase vs Sexo

	Hombre	Mujer	Total
Clase 1	434	495	929
Clase 2	85	70	155
Clase 3	95	77	172
Total	614	642	1256

La prueba  $X^2$  muestra que existe asociación entre el sexo de las personas y la clase asignada, se tiene una estadística de prueba de 6.72, con 2 grados de libertad y un nivel de significancia observado de 0.034.

Clase vs nivel de escolaridad

	Sin escolaridad	Primaria	Secundaria	Bachillerato	Licenciatura o más	Total
Clase 1	66	154	416	127	166	929
Clase 2	5	25	75	32	18	155
Clase 3	7	20	63	26	56	172
Total	78	199	554	185	240	1256

La prueba indica que existe asociación entre el nivel de escolaridad y la clase asignada, se tiene una estadística de prueba de 35.85, con 8 grados de libertad y un nivel de significancia observado muy cercano al cero.

Clase vs condición de actividad

	Trabaja	No trabaja	Total
Clase 1	472	457	929
Clase 2	78	77	155
Clase 3	88	84	172
Total	638	618	1256

La prueba muestra que no hay asociación entre la condición de actividad y la clase, se tiene una estadística de prueba de 0.023, con dos grados de libertad y nivel de significancia observado de 0.988.

Por lo tanto se incluyen en el modelo de regresión logística multinomial la variable sexo, el nivel de escolaridad y la edad como variable continua.

En el modelo de regresión logística multinomial, se observa que la probabilidad de pertenecer a la clase 2, con respecto a la clase 1, aumenta conforme aumentan los niveles de escolaridad, es decir, que el riesgo de pertenecer a la clase 2 es mayor cuando se tienen estudios de licenciatura o bachillerato, por ejemplo el riesgo de pertenecer a la clase 2 es dos veces mayor cuando se tiene bachillerato.

Asimismo, se observa que la probabilidad de pertenecer a la clase 3, con respecto a la clase 1, aumenta también conforme aumentan los niveles de escolaridad, en la clase 3 las personas con nivel de escolaridad de bachillerato tienen 63% más riesgo de estar en esta clase que en la clase 1.

Por lo que las personas con mayor nivel de escolaridad asocian el concepto de ciencia con avances técnicos y comprensión del mundo natural, mientras que las personas con niveles como primaria o secundaria lo asocian con descubrimientos y mejora de la vida humana. La edad no resulta significativa.

## Modelo II

### Imagen del científico:

¿Cuál es la imagen que tiene de la profesión de científico en México? Diría que es una profesión...

Atractiva para los jóvenes

Gratificante en lo personal

Remunerada económicamente

Reconocimiento social

Principales respuestas por cada clase obtenidas del modelo de clases latentes:

Clase 1:

Las personas que se clasifican en la clase 1, opinan que la profesión del científico en México es gratificante en lo personal, con probabilidad 1, mal remunerada económicamente, con probabilidad 0.78, y con poco reconocimiento social, con probabilidad 0.71.

Clase 2:

Las personas que se ubican en la clase 2, opinan que la profesión del científico en México es gratificante en lo personal, con probabilidad 0.78, bien remunerada económicamente, con probabilidad 0.90, y con mucho reconocimiento social, con probabilidad 0.69.

Clase 3:

Las personas que se ubican en la clase 3, no responden si es gratificante en lo personal, con probabilidad 1, mal remunerada económicamente, con probabilidad 0.76, y con poco reconocimiento social, con probabilidad 0.74.

Clase 4:

Las personas que se clasifican en la clase 4, opinan que la profesión del científico en México es poco atractiva para los jóvenes, con probabilidad 0.78, poco gratificante en lo personal, con probabilidad de 0.68, mal remunerada económicamente, con probabilidad 0.82, y con poco reconocimiento social, con probabilidad 0.87.

En este modelo se tiene que la estadística es  $X^2 = 40.25$  la cual es menor que  $X^2_{45,0.05} = 61.65$ , por lo que se considera que el modelo se ajusta adecuadamente.

La distribución de la población por cada clase se muestra en el siguiente cuadro:

	Frecuencia	%
Clase 1	467	37.2
Clase 2	481	38.3
Clase 3	8	0.6
Clase 4	300	23.9
Total	1256	100



Tablas de contingencia y prueba  $X^2$  para el modelo II

Clase vs Sexo

	Hombre	Mujer	Total
Clase 1	216	251	467
Clase 2	236	245	481
Clase 3	6	2	8
Clase 4	156	144	300
Total	614	642	1256

Los resultados de la prueba indican que no hay asociación entre el sexo de las personas y la clase asignada, se tiene una estadística de prueba de 4.65, con 3 grados de libertad y un nivel de significancia observado de 0.199.

Clase vs nivel de escolaridad

	Sin escolaridad	Primaria	Secundaria	Bachillerato	Licenciatura o más	Total
Clase 1	29	54	178	65	141	467
Clase 2	25	95	247	68	46	481
Clase 3	3	0	3	1	1	8
Clase 4	21	50	126	51	52	300
Total	78	199	554	185	240	1256

La prueba muestra que existe asociación entre el nivel de escolaridad y la clase asignada, se tiene una estadística de prueba de 90.08, con 12 grados de libertad y un nivel de significancia observado menor a 0.0001.

Clase vs condición de actividad

	Trabaja	No trabaja	Total
Clase 1	255	212	467
Clase 2	215	266	481
Clase 3	3	5	8
Clase 4	165	135	300
Total	638	618	1256

La prueba indica que existe asociación entre la condición de actividad y la clase asignada, se tiene una estadística de prueba de 12.55, con 3 grados de libertad y un nivel de significancia observado muy cercano al cero.

Por lo que se incluyen en el modelo de regresión logística multinomial las variables escolaridad, condición de actividad y la edad como variable continua.

En el modelo de regresión logística se observa que en la clase 3, los coeficientes de las variables no son significativos, por lo que la opinión de las personas es similar a la clase 2. Además de estar conformada solamente por 8 individuos, lo cual no nos permite estimar todos los parámetros, pues se tienen categorías vacías, por lo que se realizó nuevamente el modelo sin incluir a los individuos en esta clase.

Este nuevo modelo no muestra cambios en los coeficientes de las demás clases al omitir la clase 3. Se observa que la probabilidad de pertenecer a la clase 1 con respecto a la clase 2, es mayor para las personas con niveles de escolaridad de secundaria y bachillerato, es decir, que en estos niveles de escolaridad existe más riesgo de pertenecer a la clase 1, que a la clase 2.

Asimismo, se observa que las personas con secundaria y bachillerato, tienen más riesgo de pertenecer a la clase 4, en un 58% y 82% respectivamente, que a la clase 2.

En general se observa que las personas con mayores niveles de escolaridad tienen más riesgo de pertenecer a la clase 1 o a la clase 4, con respecto a la clase 2 y se observa que la edad también es significativa, conforme aumenta la edad existe más riesgo de ser clasificado en la clase 1 o la 4, con respecto a la clase 2.

Por lo que las personas con mayor nivel de escolaridad y más edad, opinan que la profesión del científico es mal remunerada económicamente y con poco reconocimiento social.

### Modelo III

#### Motivaciones del investigador, lugar e importancia del trabajo del científico:

¿Cuáles cree usted que son, en general, las principales motivaciones que tiene un investigador para dedicarse a la ciencia en México?

¿Cuál cree que es el lugar de trabajo de la mayoría de los científicos en nuestro país?

¿Usted cree que el trabajo científico para el país es: muy importante, importante, poco importante o nada importante?

Principales respuestas por cada clase obtenidas del modelo de clases latentes:

Clase 1:

Las personas que se ubican en la clase 1, creen que las principales motivaciones de un investigador son la búsqueda de nuevos conocimientos y ayudar a solucionar problemas de la gente y de la sociedad, acumulando estas respuestas una probabilidad de 0.7562. Piensan que el lugar de trabajo de los científicos son las universidades y los laboratorios, acumulando estas respuestas una probabilidad de 0.6587, y creen que el trabajo del científico es muy importante, con probabilidad de 0.69.

Clase 2:

Las personas que se ubican en la clase 2, creen que la principal motivación de un investigador son la búsqueda de nuevos conocimientos, con probabilidad de 0.68. Piensan que el lugar de trabajo de los científicos son las empresas, otros y los laboratorios, acumulando estas respuestas una probabilidad de 0.7615, y creen que el trabajo del científico es importante, con probabilidad 0.71.

En este modelo se obtuvo la estadística  $X^2 = 477.5$  la cual es menor que  $X_{502,05}^2 = 555.23$ , por lo que se considera que el ajuste es adecuado.

La distribución de la población por cada clase se muestra en el siguiente cuadro:

	Frecuencia	%
Clase 1	1168	93
Clase 2	88	7
Total	1256	100

Tablas de contingencia y prueba  $X^2$  para el modelo III

Clase vs Sexo

	Hombre	Mujer	Total
Clase 1	572	596	1168
Clase 2	42	46	88
Total	614	642	1256

Los resultados de la prueba muestran que no hay asociación entre el sexo de las personas y la clase asignada, se tiene una estadística de prueba de 0.5, con 1 grado de libertad y un nivel de significancia observado de 0.82.

Clase vs nivel de escolaridad

	Sin escolaridad	Primaria	Secundaria	Bachillerato	Licenciatura o más	Total
Clase 1	70	171	524	168	235	1168
Clase 2	8	28	30	17	5	88
Total	78	199	554	185	240	1256

La prueba muestra que existe asociación entre el nivel de escolaridad y la clase asignada, se tiene una estadística de prueba de 28.93, con 4 grados de libertad y un nivel de significancia observado muy cercano a cero.

Clase vs condición de actividad

	Trabaja	No trabaja	Total
Clase 1	593	575	1168
Clase 2	45	43	88
Total	638	618	1256

La prueba indica que no existe asociación entre la condición de actividad y la clase asignada, se tiene una estadística de prueba de 0.004, con 1 grado de libertad y un nivel de significancia observado de 0.947.

Por lo que en el modelo de regresión logística se incluyen las variables escolaridad y edad como variable continua.

En este modelo se observa que todas las categorías de la variable escolaridad resultan significativas, el riesgo de pertenecer a la clase 2 respecto a la clase 1 es mayor para las personas con nivel primaria con un incremento de 7 unidades, seguida de las personas sin escolaridad con un incremento casi de 6 unidades, es decir, que las personas sin escolaridad o con nivel de primaria opinan que la principal motivación de un investigador es la búsqueda de nuevos conocimientos, y que su lugar de trabajo son las empresas, con respecto a las personas que opinan que la motivación del investigador además de la búsqueda de nuevos conocimientos, es ayudar a solucionar problemas de la gente y de la sociedad, y que el lugar de trabajo también son las universidades.

#### **Modelo IV**

##### **Financiamiento y valoración del trabajo del científico:**

¿Quién piensa que aporta más dinero para la investigación científica en el país?

¿Usted cree que el trabajo del científico en México es apoyado por el gobierno?

Dígame por favor, si está totalmente de acuerdo, de acuerdo, en desacuerdo, o totalmente en desacuerdo con la siguiente afirmación: "El gobierno debe aumentar los recursos que destina a la investigación científica en nuestro país".

¿Usted cree que en México el trabajo del científico es valorado por la sociedad?

Principales respuestas por cada clase obtenidas del modelo de clases latentes:

Clase 1:

Las personas en la clase 1, piensan que las universidades, las fundaciones privadas y las instituciones extranjeras son quienes aportan más dinero para la investigación científica en el país,

con probabilidad acumulada de 0.7972. Creen que el trabajo del científico no es apoyado por el gobierno, con probabilidad 0.89, están de acuerdo con que el gobierno debe aumentar los recursos que destina a la investigación científica, con probabilidad 0.95, y piensan que el trabajo del científico no es valorado por la sociedad, con probabilidad 0.78.

#### Clase 2:

Las personas ubicadas en la clase 2, piensan que las universidades, las fundaciones privadas y las instituciones extranjeras son quienes aportan más dinero para la investigación científica en el país, con probabilidad acumulada de 0.7587. Creen que el trabajo del científico no es apoyado por el gobierno, con probabilidad 0.92, están totalmente de acuerdo con que el gobierno debe aumentar los recursos que destina a la investigación científica, con probabilidad 1, y piensan que el trabajo del científico no es valorado por la sociedad, con probabilidad 0.83.

#### Clase 3:

Las personas en la clase 3, piensan que el gobierno es quien aporta más dinero para la investigación científica en el país, con probabilidad de 0.58. Creen que el trabajo del científico sí es apoyado por el gobierno, con probabilidad de 0.98, están de acuerdo con que el gobierno debe aumentar los recursos que destina a la investigación científica, con probabilidad 0.61.

Para este modelo, se obtuvo la estadística  $X^2 = 299.95$  la cual es menor que  $X_{270,05}^2 = 309.32$ , por lo que se considera que el modelo se ajusta adecuadamente.

La distribución de la población por cada clase se muestra en el siguiente cuadro:

	Frecuencia	%
Clase 1	422	33.6
Clase 2	432	34.4
Clase 3	402	32
Total	1256	100

Tablas de contingencia y prueba  $X^2$  para el modelo IV

Clase vs Sexo

	Hombre	Mujer	Total
Clase 1	208	214	422
Clase 2	212	220	432
Clase 3	194	208	402
Total	614	642	1256

Los resultados de la prueba muestran que no hay asociación entre el sexo de las personas y la clase asignada, se tiene una estadística de prueba de 0.09, con 2 grados de libertad y un nivel de significancia observado de 0.95.

Clase vs nivel de escolaridad

	Sin escolaridad	Primaria	Secundaria	Bachillerato	Licenciatura o más	Total
Clase 1	21	62	186	61	92	422
Clase 2	22	53	189	74	94	432
Clase 3	35	84	179	50	54	402
Total	78	199	554	185	240	1256

La prueba muestra que hay asociación entre el nivel de escolaridad y la clase asignada, se tiene una estadística de prueba de 29.39, con 8 grados de libertad y un nivel de significancia observado muy cercano a cero.

Clase vs condición de actividad

	Trabaja	No trabaja	Total
Clase 1	212	210	422
Clase 2	232	200	432
Clase 3	194	208	402
Total	638	618	1256

La prueba indica que no existe asociación entre la condición de actividad y la clase asignada, se tiene una estadística de prueba de 2.55, con 2 grados de libertad y un nivel de significancia observado de 0.279.

Por lo que en el modelo de regresión logística multinomial se incluyen sólo las variables escolaridad y edad como variable continua.

En el modelo se observa que las personas con menores niveles de escolaridad tienen mayor riesgo de pertenecer a la clase 1 que a la clase 2, por ejemplo las personas con nivel primaria tienen un 20% más posibilidad de pertenecer a la clase 1 que a la clase 2, lo cual indica que las personas con mayor nivel de escolaridad están totalmente de acuerdo con que el gobierno debe aumentar los recursos que destina a la investigación científica.

Se observa además que las personas con menor nivel de escolaridad tienen más riesgo de pertenecer a la clase 3 que a la clase 2, las personas sin escolaridad tienen tres veces mayor riesgo de estar en la clase 3 que en la clase 2, por lo que las personas con menor escolaridad piensan que el gobierno es quien aporta más dinero para la investigación científica en el país, que el trabajo del científico sí es apoyado por el gobierno, y que el trabajo del científico sí es valorado por la sociedad.

La edad es significativa para la clase 3 con respecto a la clase 2, por lo que las personas con mayor edad, tienen más riesgo de pertenecer a la clase 3, es decir que las personas con mayor edad opinan que el gobierno es quien aporta más dinero para la investigación científica en el país y están de acuerdo con que el gobierno debe aumentar los recursos que destina a la investigación científica.

## **Modelo V**

### **Conocimiento y vínculos con la UNAM:**

¿Sabe usted qué significa UNAM?

¿Usted estudia (o estudió) en la UNAM?, Trabaja (o ha trabajado) en la UNAM?

¿Usted conoce a alguien que estudie (o haya estudiado) en la UNAM? o trabaje (o haya trabajado) en la UNAM?



Principales respuestas por cada clase obtenidas del modelo de clases latentes:

Clase 1:

Sí sabe que significa UNAM, con probabilidad 0.9130, trabaja o estudia en la UNAM, con probabilidad 0.4090, conoce a alguien que trabaje o estudie en la UNAM, con probabilidad 1.

Clase 2:

No sabe que significa UNAM, con probabilidad 0.7526, no trabaja o estudia en la UNAM, con probabilidad 0.9716.

En este modelo se tiene que la estadística  $X^2 = 0.26$  es menor que  $X^2_{0,05}$ , por lo que se considera que el modelo tiene buen ajuste.

La distribución de la población por cada clase se muestra en el siguiente cuadro:

	Frecuencia	%
Clase 1	543	43.2
Clase 2	713	56.8
Total	1256	100

Tablas de contingencia y prueba  $X^2$  para el modelo V

Clase vs Sexo

	Hombre	Mujer	Total
Clase 1	288	255	543
Clase 2	326	387	713
Total	614	642	1256

La prueba muestra que existe asociación entre el sexo y la clase, con una estadística de prueba de 6.6, un grado de libertad y nivel de significancia observado de 0.01.

Clase vs nivel de escolaridad

	Sin escolaridad	Primaria	Secundaria	Bachillerato	Licenciatura o más	Total
Clase 1	4	23	212	103	201	543
Clase 2	74	176	342	82	39	713
Total	78	199	554	185	240	1256

La prueba muestra que existe asociación entre el nivel de escolaridad y la clase, con una estadística de prueba de 305.27, 4 grados de libertad y nivel de significancia observado muy cercano a cero.

Clase vs condición de actividad

	Trabaja	No trabaja	Total
Clase 1	311	232	543
Clase 2	327	386	713
Total	638	618	1256

La prueba muestra que existe asociación entre la condición de actividad y la clase asignada, con estadística de prueba de 16.06, un grado de libertad y nivel de significancia observado de cero.

Por lo tanto se incluyen en el modelo de regresión logística multinomial las tres variables y también la edad como variable continua.

En este modelo todas las categorías de la variable escolaridad, resultaron significativas, se observa que el riesgo de pertenecer a la clase 2 en las personas de bachillerato se incrementa 3 unidades con respecto a la clase 1, y esta probabilidad se incrementa casi 7 unidades en las personas con nivel de secundaria. Por lo que las personas con mayor nivel de escolaridad tienen algún vínculo con la UNAM. Las variables edad y condición de actividad no son significativas para el modelo.

## Modelo VI

### **El consumo de información sobre ciencia que se desarrolla en la UNAM:**

La UNAM tiene diversos medios de comunicación a través de los cuales divulga la investigación científica que ahí se realiza. De la siguiente lista ¿conoce alguno de ellos?

Radio UNAM

TV UNAM

Revista digital universitaria

Gaceta UNAM

Revista ¿Cómo Ves?

Cartel UNAMirada a la Ciencia

Principales respuestas por cada clase obtenidas del modelo de clases latentes:

Clase 1:

Conoce radio UNAM, con probabilidad 0.4918, conoce TV UNAM, con probabilidad 0.6883, no conoce la revista digital universitaria, con probabilidad 1, no conoce la gaceta, con probabilidad 0.7242, no conoce la revista ¿Cómo Ves?, con probabilidad 0.9386, no conoce el Cartel UNAMirada a la Ciencia, con probabilidad 0.9678.

Clase 2:

Conoce radio UNAM, con probabilidad 0.9893, conoce TV UNAM, con probabilidad 0.9352, no conoce la revista digital universitaria, con probabilidad 0.8971, conoce la gaceta, con probabilidad 0.7943, no conoce la revista ¿Cómo Ves?, con probabilidad 0.7363, no conoce el Cartel UNAMirada a la Ciencia, con probabilidad 0.7308.

Clase 3:

No conoce radio UNAM, con probabilidad 0.5834, no conoce TV UNAM, con probabilidad 0.7309, no conoce la revista digital universitaria, con probabilidad 0.6290, conoce la gaceta, con probabilidad 0.5386, no conoce la revista ¿Cómo Ves?, con probabilidad 0.6579, no conoce el Cartel UNAMirada a la Ciencia, con probabilidad 0.7334.

Clase 4:

Conoce radio UNAM, con probabilidad 0.9106, conoce TV UNAM, con probabilidad 0.9238, conoce la revista digital universitaria, con probabilidad 0.6822, conoce la gaceta, con probabilidad 0.9919, conoce la revista ¿Cómo Ves?, con probabilidad 0.7099, conoce el Cartel UNAMirada a la Ciencia, con probabilidad 0.8862.

Clase 5:

No conoce radio UNAM, con probabilidad 0.9353, no conoce TV UNAM, con probabilidad 1, no conoce la revista digital universitaria, con probabilidad 0.9777, no conoce la gaceta, con probabilidad 0.9766, no conoce la revista ¿Cómo Ves?, con probabilidad 0.9751, no conoce el Cartel UNAMirada a la Ciencia, con probabilidad 0.9937.

Para este modelo, la estadística es  $X^2 = 25.24$  que es menor que  $X_{29,05}^2 = 42.55$ , por lo que se considera que el modelo se ajusta adecuadamente.

Las personas en la clase 1, conocen radio UNAM y TV UNAM.

Las personas en la clase 2, conocen radio UNAM, TV UNAM y la gaceta.

Las personas en la clase 3, conocen la gaceta.

Las personas en la clase 4, conocen radio UNAM, TV UNAM, la revista digital universitaria, la gaceta, la revista ¿Cómo Ves?, el Cartel UNAMirada a la Ciencia.

Las personas en la clase 5, no conocen los medios de comunicación de la UNAM.

La distribución de la población por cada clase se muestra en el siguiente cuadro:

	Frecuencia	%
Clase 1	161	12.8
Clase 2	267	21.3
Clase 3	323	25.7
Clase 4	36	2.9
Clase 5	469	37.3
Total	1256	100

Tablas de contingencia y prueba  $X^2$  para el modelo VI

Clase vs Sexo

	Hombre	Mujer	Total
Clase 1	84	77	161
Clase 2	150	117	267
Clase 3	158	165	323
Clase 4	22	14	36
Clase 5	200	269	469
Total	614	642	1256

La prueba muestra que existe asociación entre el sexo y la clase, con estadística de prueba de 15.84, 4 grados de libertad y nivel de significancia de 0.003.

Clase vs nivel de escolaridad

	Sin escolaridad	Primaria	Secundaria	Bachillerato	Licenciatura o más	Total
Clase 1	9	15	68	38	31	161
Clase 2	6	12	83	47	119	267
Clase 3	9	47	174	45	48	323
Clase 4	2	2	7	5	20	36
Clase 5	52	123	222	50	22	469
Total	78	199	554	185	240	1256

La prueba indica asociación entre el nivel de escolaridad y la clase, con estadística de prueba de 300.27, 16 grados de libertad y nivel de significancia observado menor a 0.0001.

Clase vs condición de actividad

	Trabaja	No trabaja	Total
Clase 1	89	72	161
Clase 2	166	101	267
Clase 3	150	173	323
Clase 4	20	16	36
Clase 5	213	256	469
Total	638	618	1256

La prueba indica que existe asociación entre la condición de actividad y la clase asignada, con una estadística de prueba de 23.33, 4 grados de libertad y nivel de significancia observado menor a 0.0001.

Por lo que en el modelo de regresión logística se incluyen las variables sexo, escolaridad, condición de actividad y la edad como variable continua.

En este modelo, se observa que la escolaridad es una variable significativa, en general se observa que las probabilidades en cada clase tienen un incremento, y van aumentando conforme aumenta el nivel de escolaridad.

Las personas con mayor nivel de escolaridad tienen más riesgo de pertenecer a la clase 1 que a la clase 5, por ejemplo, las personas con bachillerato tienen el 60% más de riesgo de estar en la clase 1, que en la 5.

Las personas con bachillerato tienen un 18% más riesgo de estar en la clase 2, que en la clase 5.

Las personas con nivel de escolaridad de bachillerato, tienen un 37% más probabilidad de pertenecer a la clase 3 que a la clase 5.

Por lo que las personas con mayores niveles de escolaridad tienen más riesgo de estar en alguna clase donde se ubican los individuos que conocen algún medio de comunicación de la UNAM, con respecto a los que no conocen ninguno.

Cabe mencionar, que se realizaron los modelos de regresión logística multinomial considerando el ponderador, observándose que el comportamiento de los datos no varía, excepto en el modelo VI, donde se obtuvo que para la clase 2, la variable condición de actividad se vuelve significativa, lo cual no ocurría sin considerar el factor de expansión.

Todas las salidas de estos modelos se muestran en el anexo.

## Conclusiones

Con los modelos de clases latentes se ha clasificado a la población de acuerdo a las respuestas manifestadas, y con la prueba  $X^2$  y el modelo de regresión logística se ha determinado si las variables demográficas influyen en la clase de pertenencia de los individuos.

Cabe señalar, que el modelo de clases latentes sólo se realizó con una variable dependiente debido a que los paquetes estadísticos disponibles que realizan modelos latentes sólo tienen la opción para hacerlo con una variable.

Asimismo, es importante mencionar, que la función polCA, para el modelo de clases latentes, tiene la opción de incluir covariables en el modelo, inicialmente las variables sociodemográficas se incluyeron como covariables, pero no se obtuvo un ajuste adecuado en estos modelos, se observaron errores estándar muy grandes, por lo que se optó por utilizar la prueba  $X^2$  y aplicar el modelo de regresión logística multinomial.

En el modelo I, referente a la representación social de la ciencia, se observó que las variables significativas fueron el sexo y la escolaridad, por lo que podemos decir que los hombres y las personas con niveles más altos de escolaridad, tienen mejor concepción de la ciencia.

En el modelo II, el cual involucra la imagen del científico en México, las variables significativas fueron la edad y la escolaridad, las personas con más escolaridad opinan que la ciencia es atractiva y gratificante pero mal remunerada económicamente y con poco reconocimiento social.

En el modelo III, el cual aborda el tema de las motivaciones de los investigadores para dedicarse a la ciencia, así como el lugar y la importancia del trabajo de los científicos, la variable significativa fue la escolaridad, las personas con mayores niveles de escolaridad respondieron que las principales motivaciones de los investigadores son la búsqueda de nuevos conocimientos y ayudar a solucionar problemas de la sociedad. Asimismo, refirieron que su lugar de trabajo son las universidades y laboratorios, a diferencia de las personas con menos escolaridad que respondieron como lugar de trabajo las empresas. Además, se observó que las personas con mayor escolaridad, le dan más importancia al trabajo del científico.

En el modelo IV, el cual se refiere al financiamiento y valoración del trabajo del científico, la variable significativa fue la escolaridad, se observó que las personas con más escolaridad piensan que quienes aportan más dinero a la investigación científica en el país son las universidades, las fundaciones privadas y las instituciones extranjeras, perciben poco apoyo por parte del gobierno y están de acuerdo con que el gobierno debe aumentar los recursos destinados a la investigación. Asimismo, opinan que el trabajo del científico no es valorado por la sociedad.

En el modelo V, el cual comprende el conocimiento y vínculos con la UNAM, la variable significativa fue la escolaridad, las personas con mayor nivel de escolaridad saben el significado de UNAM, y tienen algún vínculo (estudia o trabaja) con la UNAM.

En el modelo VI, el cual se refiere al consumo de información sobre ciencia que se desarrolla en la UNAM, las variables significativas fueron el sexo, la edad y la escolaridad. Se observó que los hombres conocen un poco más de los medios de información de la UNAM, las personas más jóvenes conocen la Gaceta u otras publicaciones como revistas o carteles, mientras que las

personas mayores, conocen más sobre radio y TV UNAM, respecto a la escolaridad, se observó que entre más escolaridad tienen, consumen más información sobre ciencia, por alguno los medios en cuestión.

Se observa que la variable que resultó significativa en todos los modelos fue la escolaridad, por lo que se puede concluir que las personas con más escolaridad, tienen mejor concepción y percepción de la ciencia.

Con estos resultados, se muestra que la población no percibe de forma positiva la ciencia, pues el nivel educativo del 66.4% de los entrevistados corresponde a secundaria o menos. La población en general, tiene ideas vagas de lo que es la ciencia y aprecia poco el trabajo de los científicos.

Para lograr el desarrollo socioeconómico del país, el gobierno debe apoyar programas y actividades que fortalezcan los conocimientos de la sociedad, difundir el trabajo que realizan los científicos y su importancia para la población, así como fomentar la formación de recursos humanos en áreas científicas. En general debería existir una política integral de educación para producir estudiantes conscientes de la importancia de la ciencia y la tecnología.



# Bibliografía

---

- Agresti A (2002). *Categorical Data Analysis*. John Wiley & Sons, Hoboken.
- Bartholomew, D. & Knott, M. *Latent variable models and factor analysis*. London: Arnold, 1999.
- Carullo, J. C., *La percepción social de la ciencia y la tecnología: conceptos, metodologías de medición y ejemplos significativos*, Universidad Nacional de Quilmes, Argentina.
- Clogg, C. C., *Latent class models, Handbook of statistical modeling for the social and behavioral sciences*. New York: Plenum Press, 1995.
- Collins, L.M., *Latent Class and Latent Transition Analysis with Applications in the Social, Behavioral, and Health Sciences*. Wiley, 2010.
- Conacyt (1997). *Encuesta sobre la percepción pública de la ciencia y la tecnología en México*.
- Conacyt (2001). *Encuesta sobre la percepción pública de la ciencia y la tecnología en México*.
- Conacyt (2003). *Encuesta sobre la percepción pública de la ciencia y la tecnología en México*.
- Conacyt (2005). *Encuesta sobre la percepción pública de la ciencia y la tecnología en México*.
- Conacyt (2007). *Encuesta sobre la percepción pública de la ciencia y la tecnología en México*.
- Goodman, L.A., *The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. Part I-A Modified Latent Structure Approach*. *American Journal of Sociology* Vol. 79, No. 5, 1974, pp. 1179-1259.
- Dempster, A. P.; Laird, N. M.; Rubin D.B., *Maximum Likelihood from Incomplete Data via the EM Algorithm*, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, 1977, pp. 1-38.
- Hosmer, David and Lemeshow, Stanley. *Applied logistic regression*. Wiley-Interscience publications. New York, 1989.
- Linzer, D. A., Lewis, J. *poLCA: Polytomous Variable Latent Class Analysis*.
- Márquez N. E., Tirado S. F., *Percepción Social de la ciencia y la tecnología de adolescentes mexicanos*, Portafolio CTS No. 2, 2009.
- Miller, J., Pardo R. y Niwa F. *Percepciones del Público ante la Ciencia y la Tecnología. Estudio comparativo de la Unión Europea, Estados Unidos, Japón y Canadá*. Fundación BBV, Bilbao, España, 1999.
- Organización de las Naciones Unidas para la Educación: *Declaración de Santo Domingo. La Ciencia para el Siglo XXI: Una Nueva Visión y un Marco para la Acción*, UNESCO, 1999.

Patterson, Blossom H., Dayton Mitchell, Graubard Barry I., Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data, Journal of the American Statistical Association, September 2002, Vol. 97, No. 459, Applications and Case Studies.

Proyecto Iberoamericano de Indicadores de Percepción Pública, Cultura Científica y Participación Ciudadana. 2003.

Skrondal, A., Rabe-Hesketh S., Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Chapman & Hall/CRC, interdisciplinary statistics series. United States of America, 2004.

# Anexos

---

## La función polCA de R

polCA es la primera función en R, capaz de estimar modelos de clases latentes para variables manifiestas con cualquier número de respuestas.

polCA estima el modelo de clases latentes maximizando la función:

$$\log L = \sum_{i=1}^N \ln \sum_{r=1}^R p_r \prod_{j=1}^J \prod_{k=1}^{K_k} (\pi_{jrk})^{Y_{ijk}}$$

Con respecto a  $p_r$  y  $\pi_{jrk}$  usando el algoritmo EM.

La salida de la función polCA se constituye por los siguientes elementos:

Y: un conjunto de datos de variables manifiestas

X: un conjunto de datos de covariables, si es especificado

N: el número de casos usados en el modelo

Nobs: Número de casos de observaciones completas (menor o igual que N)

Probs: Una lista de matrices que contienen las probabilidades condicionales, cada lista representa una variable manifiesta, las columnas corresponden a las posibles respuestas de cada variable y los renglones corresponden a las clases latentes.

Probs.se: Errores estándar de las probabilidades de las respuestas condicionales estimadas

P: El respectivo tamaño de cada clase latente, igual a las proporciones estimadas en el modelo básico de clases latentes, o la media a priori en el modelo de regresión de clases latentes.

P.se: Los errores estándar de P

Link: el valor máximo del estimador del modelo máximo verosímil

Numiter: el número de iteraciones requeridas para que el algoritmo converja

Maxiter: El máximo número de iteraciones que realizó el algoritmo.

Gsq: Radio de probabilidad/devianza

Chisq: Ji-cuadrada de Pearson para bondad de ajuste.

Npar: El número de grados de libertad usados para el modelo, es decir el número de parámetros estimados.

Resid.df: El número de grados de libertad residual.

## Modelos de clases latentes

Salidas correspondientes a los modelos de clases latentes:

### Modelo I

P1 Me podría decir ¿Qué cree usted que es la ciencia?

Respuestas:

PR1 No responde

PR2 Investigación, Investigar

PR3 Estudio, Estudiar

PR4 Conocimiento, Saber

PR5 Experimentos

PR6 Descubrimientos

PR7 Tecnología

PR8 Avance, Avanzado, Progreso

PR9 Novedad, Novedoso, Nuevo

PR10 Mejora, Beneficio, Ayuda

PR11 Otros

P2 ¿Cuáles de las siguientes frases considera que expresa mejor la idea de ciencia?

Respuestas:

PR1 Grandes descubrimientos

PR2 Avance técnico

PR3 Mejora de la vida humana

PR4 Comprensión del mundo natural

PR5 Dominio de la naturaleza

PR6 Transformación acelerada

PR7 Peligro de descontrol

PR8 Concentración de poder

PR9 Ideas que pocos entienden

PR10 Beneficio social

PR11 Otros

PR12 No responde

\$P1

	Pr(1)	Pr(2)	Pr(3)	Pr(4)	Pr(5)	Pr(6)	Pr(7)	Pr(8)	Pr(9)	Pr(10)	Pr(11)
class 1:	0.0833	0.1905	0.2857	0.1103	0.0157	0.0673	0.0000	0.0591	0.0197	0.0381	0.1303
class 2:	0.1239	0.0903	0.1629	0.0000	0.0073	0.0000	0.2583	0.0825	0.0000	0.0084	0.2664
class 3:	0.0530	0.0871	0.3707	0.1665	0.0127	0.0064	0.0061	0.0000	0.0000	0.0000	0.2974

\$P2

	Pr(1)	Pr(2)	Pr(3)	Pr(4)	Pr(5)	Pr(6)	Pr(7)	Pr(8)	Pr(9)	Pr(10)	Pr(11)	Pr(12)
class 1:	0.4708	0.1193	0.3249	0.0000	0.0000	0.0089	0.0106	0.0000	0.0152	0.0503	0.0000	0.0000
class 2:	0.4172	0.4503	0.0505	0.0000	0.0000	0.0000	0.0131	0.0267	0.0000	0.0378	0.0000	0.0044
class 3:	0.1447	0.1346	0.0859	0.4361	0.0872	0.0500	0.0073	0.0109	0.0357	0.0000	0.0076	0.0000

Estimated class population shares

0.607 0.183 0.21

Predicted class memberships (by modal posterior prob.)

0.7396 0.1234 0.1369

=====

Fit for 3 latent classes:

=====

number of observations: 1256

number of estimated parameters: 65

residual degrees of freedom: 66

maximum log-likelihood: -4580.597

G^2(3): 56.97306 (Likelihood ratio/deviance statistic)

X^2(3): 54.03229 (Chi-square goodness of fit)

## Modelo II

Cuál es la imagen que tiene de la profesión de científico en México? Diría que es una profesión...

P4\_1 Atractiva para los jóvenes

P4\_2 Gratificante en lo personal

P4\_3 Remunerada económicamente

P4\_4 Reconocimiento social

1.Mucho

2.Poco

3.No responde

\$P41

	Pr(1)	Pr(2)	Pr(3)
class 1:	0.5064	0.4911	0.0024
class 2:	0.5652	0.4348	0.0000
class 3:	0.5002	0.3719	0.1279
class 4:	0.2130	0.7870	0.0000

\$P42

	Pr(1)	Pr(2)	Pr(3)
class 1:	1.0000	0.0000	0.0000
class 2:	0.7836	0.2057	0.0108
class 3:	0.0000	0.0000	1.0000
class 4:	0.3188	0.6812	0.0000

\$P43

	Pr(1)	Pr(2)	Pr(3)
class 1:	0.1700	0.7859	0.0441
class 2:	0.9063	0.0000	0.0937
class 3:	0.0000	0.7675	0.2325
class 4:	0.1407	0.8223	0.0371

\$P44

	Pr(1)	Pr(2)	Pr(3)
class 1:	0.2700	0.7168	0.0133
class 2:	0.6985	0.3015	0.0000
class 3:	0.0000	0.7442	0.2558
class 4:	0.1222	0.8744	0.0035

Estimated class population shares  
0.3267 0.3093 0.0062 0.3578

Predicted class memberships (by modal posterior prob.)  
0.3718 0.383 0.0064 0.2389

=====

Fit for 4 latent classes:

=====

number of observations: 1256  
number of estimated parameters: 35  
residual degrees of freedom: 45  
maximum log-likelihood: -3481.837

G<sup>2</sup>(4): 38.46312 (Likelihood ratio/deviance statistic)  
X<sup>2</sup>(4): 40.25683 (Chi-square goodness of fit)

### Modelo III

P3 ¿Cuáles cree usted que son, en general, las principales motivaciones que tiene un investigador para dedicarse a la ciencia en México?

1. La búsqueda de nuevos conocimientos
2. Ayudar a solucionar problemas de la gente y de la sociedad
3. La búsqueda de prestigio
4. Ganar dinero
5. Conquistar un premio importante
6. La posibilidad de organizar su propio trabajo
7. Las condiciones laborales
8. Vocación por el conocimiento
9. Otros
10. No responde

P6 ¿Cuál cree que es el lugar de trabajo de la mayoría de los científicos en nuestro país?

1. Organismos públicos
2. Universidades
3. Empresas
4. Fundaciones privadas
5. Hospitales
6. Otros
7. Laboratorios
8. En el campo
9. En el extranjero
10. Centros de investigación
11. No responde

P8 ¿Usted cree que el trabajo científico para el país es: Muy importante, importante, poco importante o nada importante?

1. Muy importante
2. Importante
3. Poco importante
4. Nada importante
5. No responde

\$P3

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5) Pr(6) Pr(7) Pr(8) Pr(9) Pr(10)  
class 1: 0.5317 0.2262 0.0211 0.0405 0.0120 0.0143 0.0183 0.1306 0.0017 0.0035  
class 2: 0.6868 0.0000 0.0975 0.0775 0.0203 0.1053 0.0000 0.0127 0.0000 0.0000

\$P6

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5) Pr(6) Pr(7) Pr(8) Pr(9) Pr(10) Pr(11)  
class 1: 0.0549 0.3316 0.0672 0.0148 0.0288 0.0585 0.3271 0.0297 0.0367 0.0061 0.0444  
class 2: 0.0458 0.0024 0.1718 0.0000 0.0000 0.2163 0.3734 0.0085 0.0000 0.0000 0.1818

\$P8R

	Pr(1)	Pr(2)	Pr(3)	Pr(4)	Pr(5)
class 1:	0.6909	0.2435	0.0577	0.0079	0.000
class 2:	0.0528	0.7140	0.2242	0.0000	0.009

Estimated class population shares  
0.9116 0.0884

Predicted class memberships (by modal posterior prob.)  
0.9299 0.0701

=====  
Fit for 2 latent classes:  
=====

number of observations: 1256  
number of estimated parameters: 47  
residual degrees of freedom: 502  
maximum log-likelihood: -5113.797

G<sup>2</sup>(2): 239.5885 (Likelihood ratio/deviance statistic)  
X<sup>2</sup>(2): 477.5011 (Chi-square goodness of fit)

#### Modelo IV

P9 ¿Quién piensa que aporta más dinero para la investigación científica en el país?

1. El gobierno
2. Las empresas
3. Las universidades
4. Las fundaciones privadas
5. Las instituciones extranjeras
6. Otros
7. No responde

P10 ¿Usted cree que el trabajo del científico en México es apoyado por el gobierno?

1. Si
2. No
3. No responde

P11 Dígame por favor, si está totalmente de acuerdo, de acuerdo, en desacuerdo, o totalmente en desacuerdo con la siguiente afirmación: "El gobierno debe aumentar los recursos que destina a la investigación científica en nuestro país".

1. Totalmente de acuerdo
2. De acuerdo
3. En desacuerdo
4. Totalmente en desacuerdo
5. No responde



P12 ¿Usted cree que en México el trabajo del científico es valorado por la sociedad?

1. Si
2. No
3. No responde

\$P9R

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5) Pr(6) Pr(7)  
class 1: 0.0889 0.1062 0.3336 0.2118 0.2518 0.0025 0.0053  
class 2: 0.0645 0.1438 0.3544 0.1913 0.2130 0.0162 0.0169  
class 3: 0.5806 0.1076 0.1513 0.0486 0.0957 0.0060 0.0102

\$P10R

Pr(1) Pr(2) Pr(3)  
class 1: 0.1005 0.8901 0.0094  
class 2: 0.0798 0.9202 0.0000  
class 3: 0.9804 0.0000 0.0196

\$P11R

Pr(1) Pr(2) Pr(3) Pr(4) Pr(5)  
class 1: 0.0000 0.9516 0.0384 0.0051 0.0049  
class 2: 1.0000 0.0000 0.0000 0.0000 0.0000  
class 3: 0.3519 0.6120 0.0282 0.0052 0.0027

\$P12R

Pr(1) Pr(2) Pr(3)  
class 1: 0.2039 0.7839 0.0122  
class 2: 0.1605 0.8395 0.0000  
class 3: 0.5237 0.4763 0.0000

Estimated class population shares  
0.3274 0.3842 0.2884

Predicted class memberships (by modal posterior prob.)  
0.3057 0.3893 0.3049

=====  
Fit for 3 latent classes:

=====  
number of observations: 1256  
number of estimated parameters: 44  
residual degrees of freedom: 270  
maximum log-likelihood: -4502.966

G^2(3): 127.7895 (Likelihood ratio/deviance statistic)  
X^2(3): 299.9595 (Chi-square goodness of fit)

**Modelo V**

P14. ¿Sabe usted qué significa UNAM?

- 1. Si
- 2. No

P15. ¿Usted estudia (o estudió) en la UNAM?

- 1. Si
- 2. No

P16. ¿Usted conoce a alguien que estudie (o haya estudiado) en la UNAM? o trabaje (o haya trabajado) en la UNAM?

- 1. Si
- 2. No

\$P14R

Pr(1) Pr(2)  
class 1: 0.9130 0.0870  
class 2: 0.2474 0.7526

\$P15R

Pr(1) Pr(2)  
class 1: 0.4090 0.5910  
class 2: 0.0284 0.9716

\$P16R

Pr(1) Pr(2)  
class 1: 1.0000 0.0000  
class 2: 0.5153 0.4847

Estimated class population shares  
0.3627 0.6373

Predicted class memberships (by modal posterior prob.)  
0.4323 0.5677

=====

Fit for 2 latent classes:

=====

number of observations: 1256  
number of estimated parameters: 7  
residual degrees of freedom: 0  
maximum log-likelihood: -2048.913

G^2(2): 0.2794972 (Likelihood ratio/deviance statistic)  
X^2(2): 0.2616735 (Chi-square goodness of fit)

## Modelo VI

29. La UNAM tiene diversos medios de comunicación a través de los cuales divulga la investigación científica que ahí se realiza. De la siguiente lista ¿conoce alguno de ellos?

a) Radio UNAM

1. Sí
2. No

b) TV UNAM

1. Sí
2. No

c) Revista digital universitaria

1. Sí
2. No

e) Gaceta UNAM

1. Sí
2. No

f) Revista ¿Cómo Ves?

1. Sí
2. No

h) Cartel UNAMirada a la Ciencia

1. Sí
2. No

\$P29AR

Pr(1) Pr(2)

class 1: 0.4918 0.5082

class 2: 0.9893 0.0107

class 3: 0.4166 0.5834

class 4: 0.9106 0.0894

class 5: 0.0647 0.9353

\$P29BR

Pr(1) Pr(2)

class 1: 0.6883 0.3117

class 2: 0.9352 0.0648

class 3: 0.2691 0.7309

class 4: 0.9238 0.0762

class 5: 0.0000 1.0000

\$P29CR

	Pr(1)	Pr(2)
class 1:	0.0000	1.0000
class 2:	0.1029	0.8971
class 3:	0.3710	0.6290
class 4:	0.6822	0.3178
class 5:	0.0223	0.9777

\$P29ER

	Pr(1)	Pr(2)
class 1:	0.2758	0.7242
class 2:	0.7943	0.2057
class 3:	0.5386	0.4614
class 4:	0.9919	0.0081
class 5:	0.0234	0.9766

\$P29FR

	Pr(1)	Pr(2)
class 1:	0.0614	0.9386
class 2:	0.2637	0.7363
class 3:	0.3421	0.6579
class 4:	0.7099	0.2901
class 5:	0.0249	0.9751

\$P29HR

	Pr(1)	Pr(2)
class 1:	0.0322	0.9678
class 2:	0.2692	0.7308
class 3:	0.2666	0.7334
class 4:	0.8862	0.1138
class 5:	0.0063	0.9937

Estimated class population shares  
0.1713 0.2042 0.2467 0.0411 0.3367

Predicted class memberships (by modal posterior prob.)  
0.1282 0.2126 0.2572 0.0287 0.3734

=====  
Fit for 5 latent classes:  
=====

number of observations: 1256  
number of estimated parameters: 34  
residual degrees of freedom: 29  
maximum log-likelihood: -3738.428

G^2(5): 26.48131 (Likelihood ratio/deviance statistic)  
X^2(5): 25.24128 (Chi-square goodness of fit)

## Modelos de regresión logística multinomial

### Modelo de regresión logística multinomial I

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 2	Intercept	-2.292	0.348	43.5	1	0			
	Edad	0	0.007	0.015	1	0.903	0.999	0.986	1.013
	Hombre	0.363	0.174	4.333	1	0.037	1.437	1.021	2.022
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-0.413	0.542	0.58	1	0.446	0.662	0.229	1.914
	Primaria	0.298	0.325	0.841	1	0.359	1.348	0.712	2.549
	Secundaria	0.401	0.274	2.144	1	0.143	1.494	0.873	2.556
	Bachillerato	0.829	0.312	7.082	1	0.008	2.291	1.244	4.219
	Licenciatura o más	0b	.	.	0	.	.	.	.
Clase 3	Intercept	-1.598	0.29	30.321	1	0			
	Edad	0.007	0.007	1.113	1	0.291	1.007	0.994	1.02
	Hombre	0.488	0.173	7.951	1	0.005	1.628	1.16	2.285
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-1.24	0.45	7.588	1	0.006	0.289	0.12	0.699
	Primaria	-1.078	0.298	13.05	1	0	0.34	0.19	0.611
	Secundaria	-0.723	0.21	11.825	1	0.001	0.485	0.321	0.733
	Bachillerato	-0.625	0.287	4.755	1	0.029	0.535	0.305	0.939
	Licenciatura o más	0b	.	.	0	.	.	.	.

Prueba de bondad de ajuste del modelo

$X^2$	G.L	Sig.
47.289	12	0.000

**Modelo de regresión logística multinomial I (con factores de expansión)**

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 2	Intercept	-2.292	0.348	43.5	1	0			
	Edad	0	0.007	0.015	1	0.903	0.999	0.986	1.013
	Hombre	0.363	0.174	4.333	1	0.037	1.437	1.021	2.022
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-0.413	0.542	0.58	1	0.446	0.662	0.229	1.914
	Primaria	0.298	0.325	0.841	1	0.359	1.348	0.712	2.549
	Secundaria	0.401	0.274	2.144	1	0.143	1.494	0.873	2.556
	Bachillerato	0.829	0.312	7.082	1	0.008	2.291	1.244	4.219
	Licenciatura o más	0b	.	.	0	.	.	.	.
Clase 3	Intercept	-1.598	0.29	30.321	1	0			
	Edad	0.007	0.007	1.113	1	0.291	1.007	0.994	1.02
	Hombre	0.488	0.173	7.951	1	0.005	1.628	1.16	2.285
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-1.24	0.45	7.588	1	0.006	0.289	0.12	0.699
	Primaria	-1.078	0.298	13.05	1	0	0.34	0.19	0.611
	Secundaria	-0.723	0.21	11.825	1	0.001	0.485	0.321	0.733
	Bachillerato	-0.625	0.287	4.755	1	0.029	0.535	0.305	0.939
	Licenciatura o más	0b	.	.	0	.	.	.	.

**Prueba de bondad de ajuste del modelo**

$X^2$	G.L	Sig.
47.289	12	0.000

## Modelo de regresión logística multinomial II

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 1	Intercept	-0.261	0.255	1.05	1	0.305			
	Edad	0.042	0.006	48.087	1	0	1.043	1.03	1.055
	Sin escolaridad	-1.699	0.35	23.513	1	0	0.183	0.092	0.363
	Primaria	-1.888	0.251	56.479	1	0	0.151	0.092	0.248
	Secundaria	-1.209	0.201	35.99	1	0	0.299	0.201	0.443
	Bachillerato	-0.954	0.247	14.89	1	0	0.385	0.237	0.625
	Licenciatura o más	Ob	.	.	0	.	.	.	.
	Trabaja	0.09	0.14	0.409	1	0.522	1.094	0.831	1.441
	No trabaja	Ob	.	.	0	.	.	.	.
Clase 3	Intercept	-5.419	1.444	14.085	1	0			
	Edad	0.056	0.026	4.632	1	0.031	1.058	1.005	1.113
	Sin escolaridad	0.578	1.298	0.198	1	0.656	1.783	0.14	22.701
	Primaria	-19.67	0	.	1	.	2.86E-09	2.86E-09	2.86E-09
	Secundaria	-0.339	1.18	0.083	1	0.774	0.712	0.071	7.188
	Bachillerato	-0.114	1.443	0.006	1	0.937	0.892	0.053	15.095
	Licenciatura o más	Ob	.	.	0	.	.	.	.
	Trabaja	-0.453	0.752	0.363	1	0.547	0.636	0.146	2.776
	No trabaja	Ob	.	.	0	.	.	.	.
Clase 4	Intercept	-1.319	0.294	20.129	1	0			
	Edad	0.042	0.006	42.078	1	0	1.043	1.03	1.056
	Sin escolaridad	-1.007	0.387	6.784	1	0.009	0.365	0.171	0.779
	Primaria	-0.953	0.277	11.833	1	0.001	0.385	0.224	0.664
	Secundaria	-0.544	0.236	5.33	1	0.021	0.58	0.366	0.921
	Bachillerato	-0.192	0.279	0.474	1	0.491	0.825	0.477	1.426
	Licenciatura o más	Ob	.	.	0	.	.	.	.
	Trabaja	0.178	0.155	1.332	1	0.248	1.195	0.883	1.618
	No trabaja	Ob	.	.	0	.	.	.	.

Prueba de bondad de ajuste del modelo

$\chi^2$	G.L	Sig.
159.152	18	0.000

### Modelo de regresión logística multinomial II excluyendo la clase 3

		B	Std. Error	Wald	df	Sig.	Exp(B)	Lower Bound	Upper Bound
Clase 1	Intercept	-0.259	0.254	1.033	1	0.309			
	Edad	0.042	0.006	48.276	1	0	1.043	1.03	1.055
	Sin escolaridad	-1.714	0.351	23.804	1	0	0.18	0.09	0.359
	Primaria	-1.892	0.251	56.675	1	0	0.151	0.092	0.247
	Secundaria	-1.212	0.201	36.19	1	0	0.298	0.201	0.442
	Bachillerato	-0.956	0.247	14.964	1	0	0.384	0.237	0.624
	Licenciatura o más	Ob	.	.	0	.	.	.	.
	Trabaja	0.087	0.141	0.382	1	0.537	1.091	0.828	1.437
	No trabaja	Ob	.	.	0	.	.	.	.
Clase 4	Intercept	-1.314	0.294	20.031	1	0			
	Edad	0.042	0.006	42.047	1	0	1.043	1.03	1.056
	Sin escolaridad	-1.019	0.387	6.916	1	0.009	0.361	0.169	0.771
	Primaria	-0.956	0.277	11.893	1	0.001	0.384	0.223	0.662
	Secundaria	-0.547	0.236	5.394	1	0.02	0.579	0.365	0.918
	Bachillerato	-0.195	0.279	0.489	1	0.485	0.823	0.476	1.422
	Licenciatura o más	Ob	.	.	0	.	.	.	.
	Trabaja	0.177	0.155	1.316	1	0.251	1.194	0.882	1.617
	No trabaja	Ob	.	.	0	.	.	.	.

#### Prueba de bondad de ajuste del modelo

$\chi^2$	G.L	Sig.
149.024	12	0.000



### Modelo de regresión logística multinomial II excluyendo la clase 3 (con factores de expansión)

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 1	Intercept	-0.204	0.258	0.624	1	0.43			
	Edad	0.045	0.006	54.622	1	0	1.046	1.033	1.058
	Sin escolaridad	-1.957	0.361	29.453	1	0	0.141	0.07	0.286
	Primaria	-2.039	0.258	62.449	1	0	0.13	0.078	0.216
	Secundaria	-1.351	0.206	42.952	1	0	0.259	0.173	0.388
	Bachillerato	-0.954	0.255	14.008	1	0	0.385	0.234	0.635
	Licenciatura o más	Ob	.	.	0	.	.	.	.
	Trabaja	-0.017	0.143	0.014	1	0.905	0.983	0.743	1.3
No trabaja	Ob	.	.	0	.	.	.	.	
Clase 4	Intercept	-1.158	0.29	15.884	1	0			
	Edad	0.043	0.006	45.68	1	0	1.044	1.031	1.057
	Sin escolaridad	-1.275	0.393	10.551	1	0.001	0.279	0.129	0.603
	Primaria	-1.087	0.278	15.315	1	0	0.337	0.196	0.581
	Secundaria	-0.659	0.235	7.887	1	0.005	0.517	0.326	0.819
	Bachillerato	-0.24	0.282	0.721	1	0.396	0.787	0.452	1.369
	Licenciatura o más	Ob	.	.	0	.	.	.	.
	Trabaja	0.132	0.153	0.745	1	0.388	1.141	0.845	1.54
No trabaja	Ob	.	.	0	.	.	.	.	

Prueba de bondad de ajuste del modelo

$X^2$	G.L	Sig.
161.601	12	0.000

### Modelo de regresión logística multinomial III

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 2	Intercept	-3.384	0.536	39.905	1	0			
	Edad	-0.014	0.009	2.491	1	0.115	0.986	0.969	1.003
	Sin escolaridad	1.927	0.606	10.098	1	0.001	6.87	2.093	22.549
	Primaria	2.09	0.497	17.688	1	0	8.086	3.053	21.418
	Secundaria	0.896	0.493	3.3	1	0.069	2.449	0.932	6.439
	Bachillerato	1.479	0.521	8.057	1	0.005	4.388	1.58	12.184
	Licenciatura o más	Ob	.	.	0	.	.	.	.

Prueba de bondad de ajuste del modelo

$X^2$	G.L	Sig.
31.412	5	0.000

### Modelo de regresión logística multinomial III (con factores de expansión)

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 2	Intercept	-3.072	0.499	37.835	1	0			
	Edad	-0.019	0.009	4.742	1	0.029	0.981	0.964	0.998
	Sin escolaridad	1.871	0.582	10.328	1	0.001	6.497	2.075	20.34
	Primaria	2.127	0.462	21.166	1	0	8.391	3.39	20.767
	Secundaria	0.776	0.46	2.847	1	0.092	2.172	0.882	5.349
	Bachillerato	1.315	0.494	7.083	1	0.008	3.724	1.414	9.808
	Licenciatura o más	0b	.	.	0	.	.	.	.

Prueba de bondad de ajuste del modelo

$\chi^2$	G.L	Sig.
38.532	5	0.000

### Modelo de regresión logística multinomial IV

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 1	Intercept	0.142	0.235	0.364	1	0.546			
	Edad	-0.005	0.005	0.792	1	0.373	0.995	0.985	1.006
	Sin escolaridad	0.062	0.353	0.031	1	0.86	1.064	0.533	2.124
	Primaria	0.2	0.239	0.703	1	0.402	1.222	0.765	1.952
	Secundaria	-0.025	0.183	0.019	1	0.89	0.975	0.681	1.395
	Bachillerato	-0.201	0.229	0.772	1	0.38	0.818	0.522	1.281
	Licenciatura o más	0b	.	.	0	.	.	.	.
Clase 3	Intercept	-0.013	0.253	0.003	1	0.959			
	Edad	-0.016	0.006	8.297	1	0.004	0.984	0.973	0.995
	Sin escolaridad	1.311	0.339	14.949	1	0	3.709	1.908	7.208
	Primaria	1.082	0.247	19.17	1	0	2.951	1.818	4.79
	Secundaria	0.395	0.204	3.757	1	0.053	1.484	0.996	2.213
	Bachillerato	0.066	0.253	0.069	1	0.793	1.069	0.651	1.755
	Licenciatura o más	0b	.	.	0	.	.	.	.

Prueba de bondad de ajuste del modelo

$\chi^2$	G.L	Sig.
38.235	10	0.000

**Modelo de regresión logística multinomial IV (con factores de expansión)**

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 1	Intercept	0.253	0.23	1.209	1	0.272			
	Edad	-0.006	0.005	1.5	1	0.221	0.994	0.983	1.004
	Sin escolaridad	-0.107	0.345	0.097	1	0.756	0.898	0.456	1.767
	Primaria	0.125	0.238	0.277	1	0.598	1.133	0.711	1.805
	Secundaria	-0.057	0.18	0.102	1	0.75	0.944	0.664	1.343
	Bachillerato	-0.3	0.23	1.699	1	0.192	0.741	0.472	1.163
	Licenciatura o más	0b	.	.	0	.	.	.	.
Clase 3	Intercept	-0.164	0.256	0.411	1	0.521			
	Edad	-0.015	0.006	7.54	1	0.006	0.985	0.974	0.996
	Sin escolaridad	1.176	0.34	11.995	1	0.001	3.243	1.666	6.311
	Primaria	1.156	0.252	21.1	1	0	3.178	1.94	5.205
	Secundaria	0.51	0.209	5.968	1	0.015	1.666	1.106	2.509
	Bachillerato	0.143	0.261	0.303	1	0.582	1.154	0.693	1.924
	Licenciatura o más	0b	.	.	0	.	.	.	.

Prueba de bondad de ajuste del modelo

X <sup>2</sup>	G.L	Sig.
39.677	10	0.000

**Modelo de regresión logística multinomial V**

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 2	Intercept	-1.245	0.26	23.013	1	0			
	Edad	-0.006	0.006	1.086	1	0.297	0.994	0.984	1.005
	Hombre	-0.166	0.138	1.444	1	0.229	0.847	0.647	1.11
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	4.621	0.554	69.592	1	0	101.551	34.294	300.71
	Primaria	3.669	0.285	165.58	1	0	39.195	22.416	68.533
	Secundaria	2.055	0.199	107.12	1	0	7.809	5.291	11.524
	Bachillerato	1.368	0.232	34.869	1	0	3.928	2.494	6.186
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	-0.194	0.142	1.868	1	0.172	0.824	0.624	1.088
	No trabaja	0b	.	.	0	.	.	.	.

Prueba de bondad de ajuste del modelo

X <sup>2</sup>	G.L	Sig.
346.339	7	0.000

**Modelo de regresión logística multinomial V (con factores de expansión)**

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 2	Intercept	1.333	0.264	25.445	1	0			
	Edad	0.006	0.006	1.053	1	0.305	1.006	0.995	1.017
	Hombre	0.266	0.139	3.646	1	0.056	1.305	0.993	1.715
	Mujer	Ob	.	.	0	.	.	.	.
	Sin escolaridad	-4.824	0.579	69.407	1	0	0.008	0.003	0.025
	Primaria	-3.892	0.297	171.184	1	0	0.02	0.011	0.037
	Secundaria	-2.232	0.205	118.224	1	0	0.107	0.072	0.16
	Bachillerato	-1.456	0.24	36.792	1	0	0.233	0.146	0.373
	Licenciatura o más	Ob	.	.	0	.	.	.	.
	Trabaja	0.181	0.142	1.622	1	0.203	1.199	0.907	1.584
	No trabaja	Ob	.	.	0	.	.	.	.

Prueba de bondad de ajuste del modelo

$\chi^2$	G.L	Sig.
372.931	7	0.000

## Modelo de regresión logística multinomial VI

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 1	Intercept	-0.508	0.39	1.693	1	0.193			
	Edad	0.018	0.007	6.448	1	0.011	1.019	1.004	1.033
	Hombre	0.308	0.199	2.397	1	0.122	1.36	0.921	2.008
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-2.4	0.482	24.83	1	0	0.091	0.035	0.233
	Primaria	-2.527	0.396	40.684	1	0	0.08	0.037	0.174
	Secundaria	-1.408	0.315	19.936	1	0	0.245	0.132	0.454
	Bachillerato	-0.504	0.357	1.994	1	0.158	0.604	0.3	1.216
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	0.115	0.202	0.323	1	0.57	1.122	0.755	1.666
No trabaja	0b	.	.	0	.	.	.	.	
Clase 2	Intercept	1.003	0.34	8.688	1	0.003			
	Edad	0.009	0.007	1.506	1	0.22	1.009	0.995	1.022
	Hombre	0.434	0.18	5.836	1	0.016	1.544	1.085	2.197
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-3.933	0.511	59.363	1	0	0.02	0.007	0.053
	Primaria	-4.016	0.386	108.52	1	0	0.018	0.008	0.038
	Secundaria	-2.595	0.269	92.723	1	0	0.075	0.044	0.127
	Bachillerato	-1.698	0.313	29.473	1	0	0.183	0.099	0.338
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	0.295	0.184	2.573	1	0.109	1.343	0.937	1.927
No trabaja	0b	.	.	0	.	.	.	.	
Clase 3	Intercept	1.238	0.335	13.667	1	0			
	Edad	-0.017	0.006	7.137	1	0.008	0.983	0.971	0.995
	Hombre	0.195	0.157	1.537	1	0.215	1.215	0.893	1.654
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-2.225	0.458	23.589	1	0	0.108	0.044	0.265
	Primaria	-1.673	0.312	28.748	1	0	0.188	0.102	0.346
	Secundaria	-1.131	0.28	16.305	1	0	0.323	0.186	0.559
	Bachillerato	-0.986	0.332	8.831	1	0.003	0.373	0.195	0.715
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	-0.002	0.161	0	1	0.988	0.998	0.728	1.367
No trabaja	0b	.	.	0	.	.	.	.	
Clase 4	Intercept	0.06	0.627	0.009	1	0.923			
	Edad	-0.017	0.016	1.105	1	0.293	0.983	0.952	1.015
	Hombre	0.749	0.377	3.941	1	0.047	2.114	1.01	4.427
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-2.813	0.836	11.325	1	0.001	0.06	0.012	0.309
	Primaria	-3.954	0.781	25.655	1	0	0.019	0.004	0.089
	Secundaria	-3.488	0.507	47.397	1	0	0.031	0.011	0.082
	Bachillerato	-2.325	0.569	16.671	1	0	0.098	0.032	0.299
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	0.014	0.381	0.001	1	0.97	1.014	0.481	2.14
No trabaja	0b	.	.	0	.	.	.	.	

### Prueba de bondad de ajuste del modelo

$X^2$	G.L	Sig.
338.716	28	0.000

### Modelo de regresión logística multinomial VI (con factores de expansión)

		B	Error estándar	Wald	Gl.	Sig.	Exp(B)	Cota inferior	Cota superior
Clase 1	Intercept	-0.277	0.399	0.482	1	0.488			
	Edad	0.016	0.007	5.237	1	0.022	1.017	1.002	1.031
	Hombre	0.354	0.199	3.149	1	0.076	1.424	0.964	2.105
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-2.714	0.513	28.006	1	0	0.066	0.024	0.181
	Primaria	-2.587	0.401	41.538	1	0	0.075	0.034	0.165
	Secundaria	-1.599	0.328	23.8	1	0	0.202	0.106	0.384
	Bachillerato	-0.727	0.372	3.823	1	0.051	0.483	0.233	1.002
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	0.2	0.202	0.985	1	0.321	1.222	0.823	1.815
No trabaja	0b	.	.	0	.	.	.	.	
Clase 2	Intercept	1.227	0.352	12.139	1	0			
	Edad	0.006	0.007	0.742	1	0.389	1.006	0.992	1.02
	Hombre	0.502	0.181	7.718	1	0.005	1.652	1.159	2.353
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-3.949	0.506	61.036	1	0	0.019	0.007	0.052
	Primaria	-4.124	0.396	108.372	1	0	0.016	0.007	0.035
	Secundaria	-2.741	0.285	92.396	1	0	0.065	0.037	0.113
	Bachillerato	-1.965	0.333	34.827	1	0	0.14	0.073	0.269
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	0.386	0.184	4.419	1	0.036	1.472	1.026	2.11
No trabaja	0b	.	.	0	.	.	.	.	
Clase 3	Intercept	1.515	0.347	19.117	1	0			
	Edad	-0.019	0.006	9.403	1	0.002	0.981	0.969	0.993
	Hombre	0.19	0.157	1.456	1	0.228	1.209	0.888	1.645
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-2.148	0.453	22.537	1	0	0.117	0.048	0.283
	Primaria	-1.627	0.325	25.02	1	0	0.197	0.104	0.372
	Secundaria	-1.249	0.297	17.628	1	0	0.287	0.16	0.514
	Bachillerato	-1.076	0.348	9.545	1	0.002	0.341	0.172	0.675
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	-0.093	0.16	0.335	1	0.563	0.911	0.666	1.248
No trabaja	0b	.	.	0	.	.	.	.	
Clase 4	Intercept	0.192	0.607	0.1	1	0.752			
	Edad	-0.012	0.015	0.653	1	0.419	0.988	0.959	1.018
	Hombre	0.502	0.359	1.958	1	0.162	1.652	0.818	3.337
	Mujer	0b	.	.	0	.	.	.	.
	Sin escolaridad	-2.962	0.814	13.227	1	0	0.052	0.01	0.255
	Primaria	-4.076	0.786	26.869	1	0	0.017	0.004	0.079
	Secundaria	-3.356	0.48	48.786	1	0	0.035	0.014	0.089
	Bachillerato	-2.312	0.56	17.009	1	0	0.099	0.033	0.297
	Licenciatura o más	0b	.	.	0	.	.	.	.
	Trabaja	0.109	0.365	0.089	1	0.766	1.115	0.545	2.282
No trabaja	0b	.	.	0	.	.	.	.	

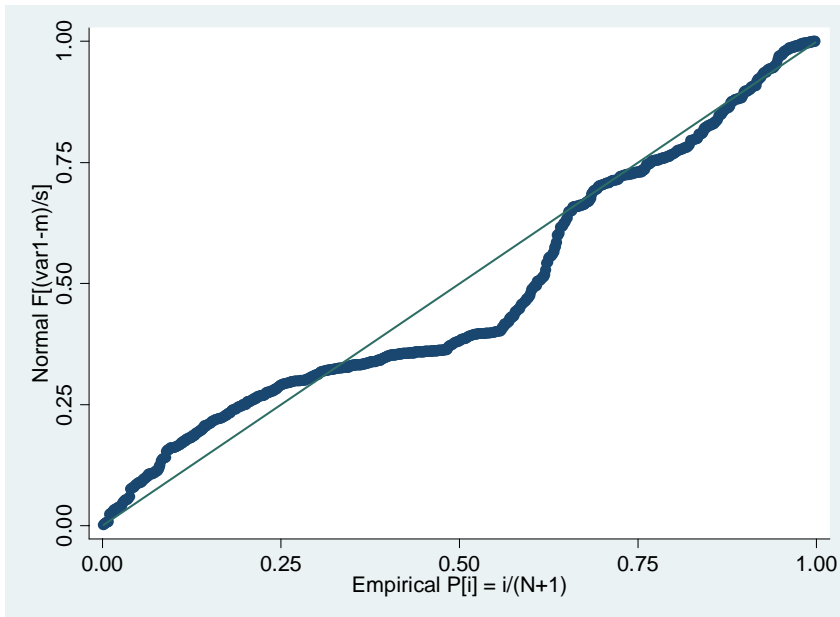
Prueba de bondad de ajuste del modelo

$X^2$	G.L	Sig.
340.008	28	0.000

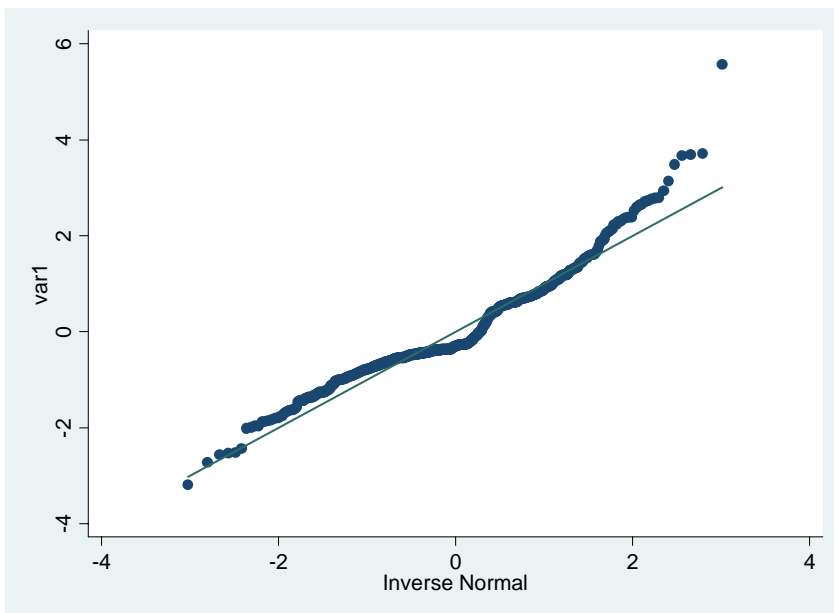
**Gráficas de residuales para los modelos:**

**Modelo I**

Normal probability plot

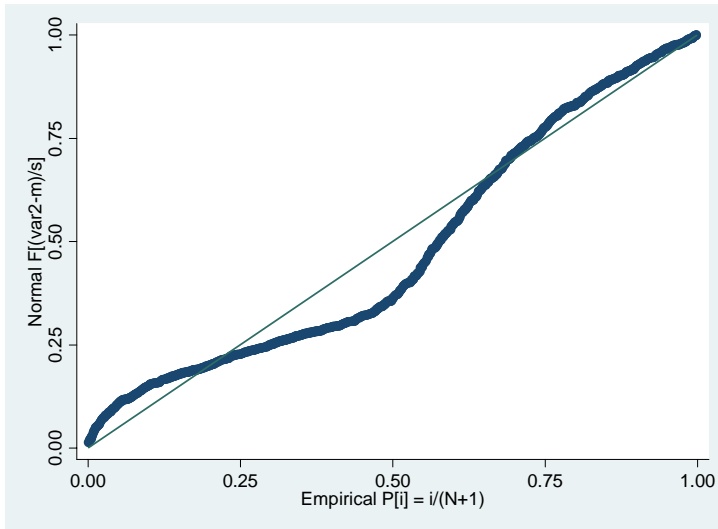


Normal quantile plot

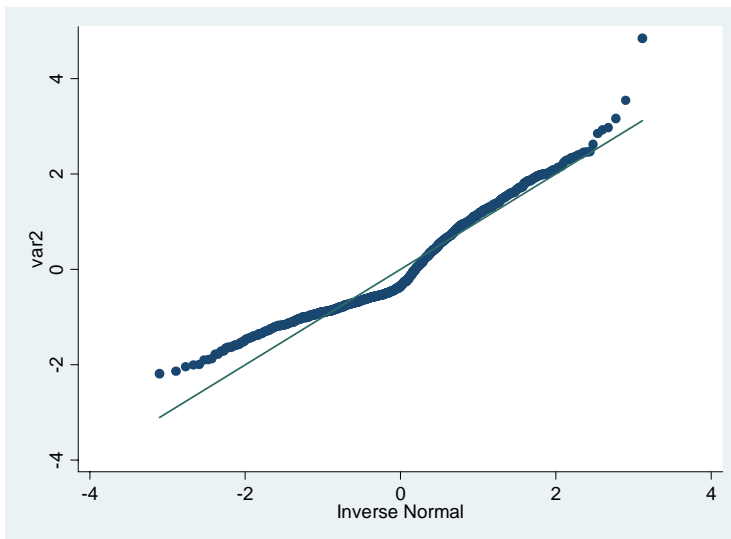


## Modelo II

Normal probability plot



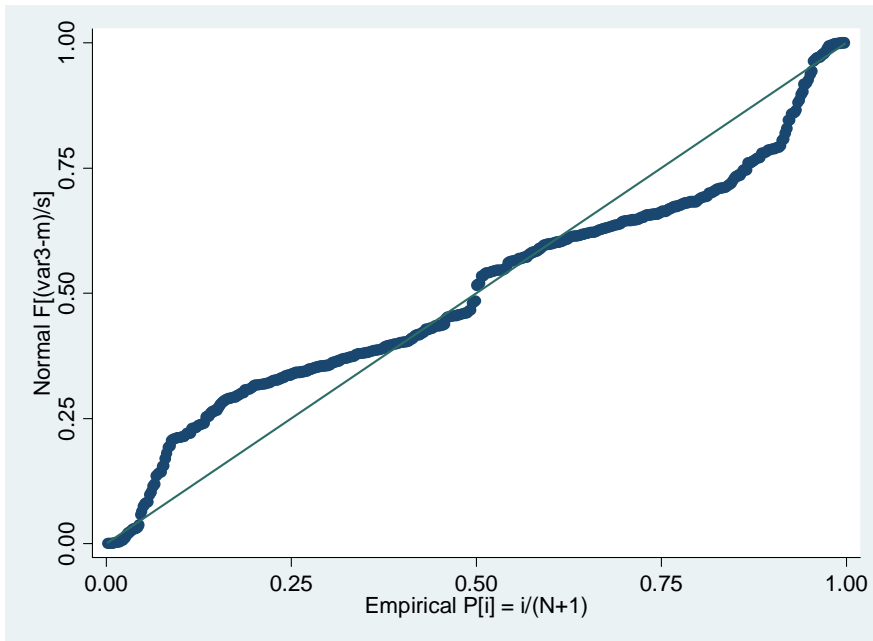
Normal quantile plot



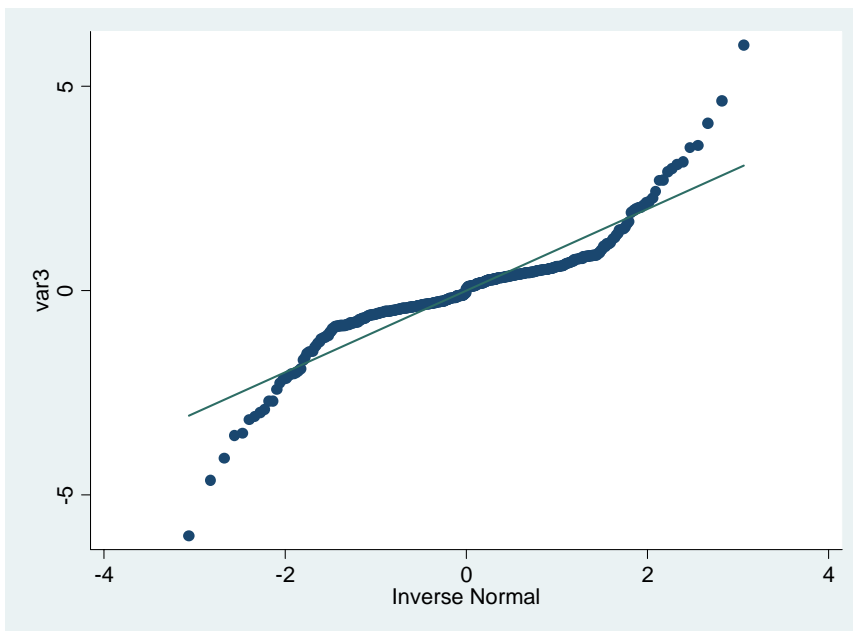


### Modelo III

Normal probability plot

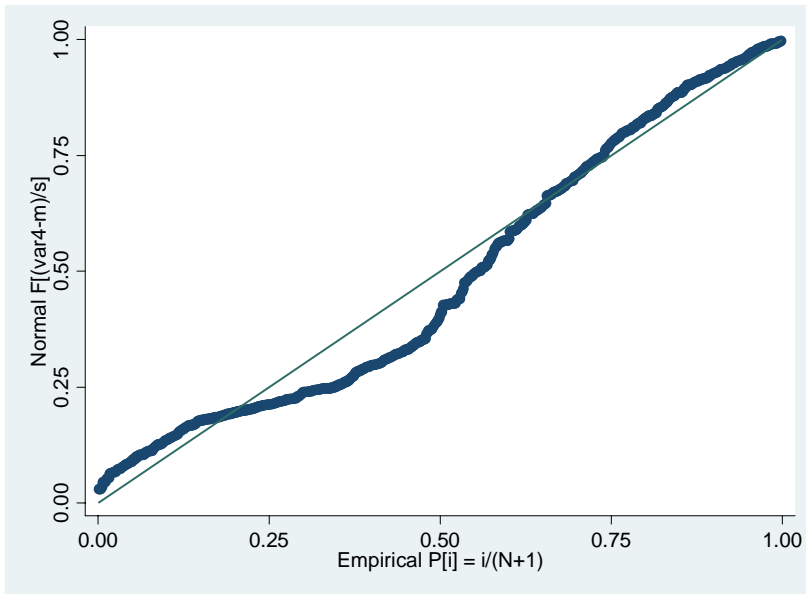


Normal quantile plot

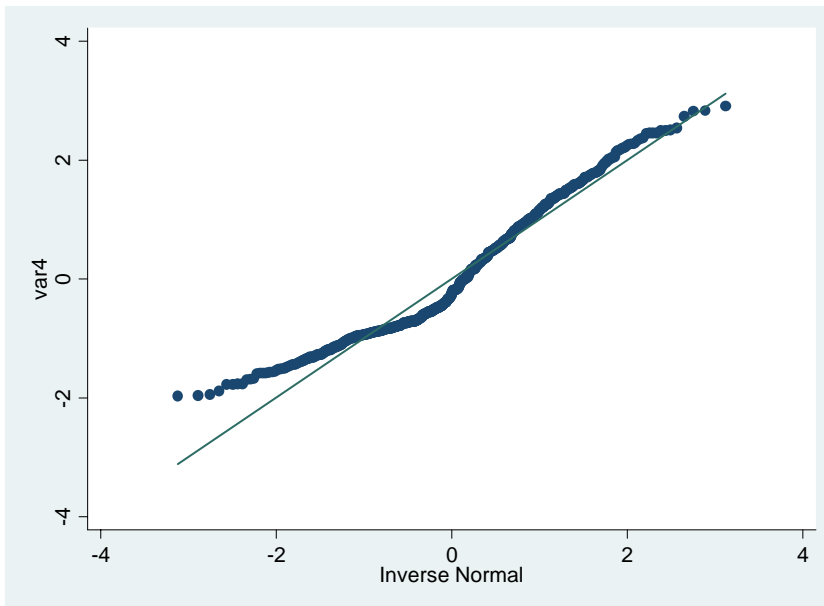


## Modelo IV

Normal probability plot

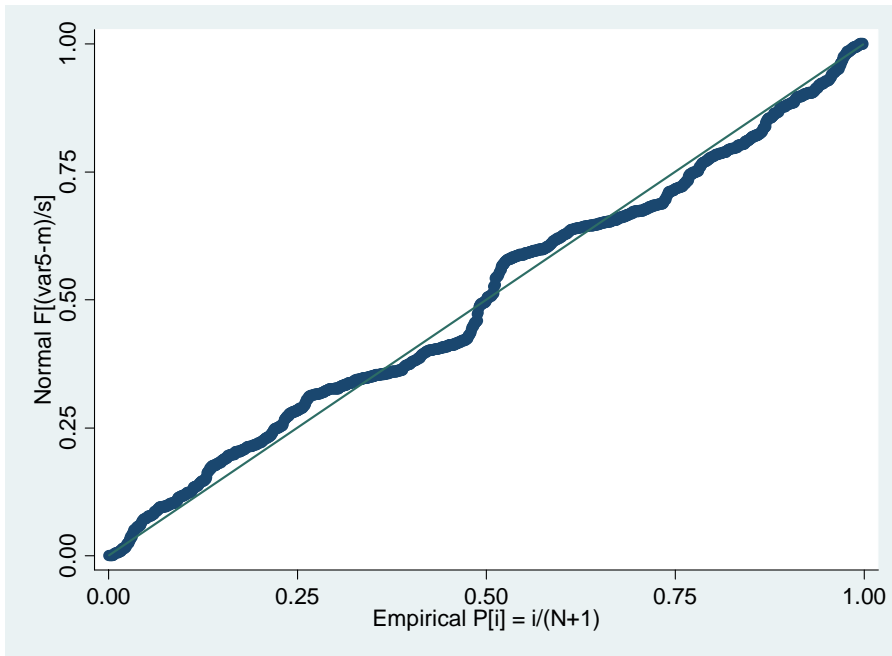


Normal quantile plot

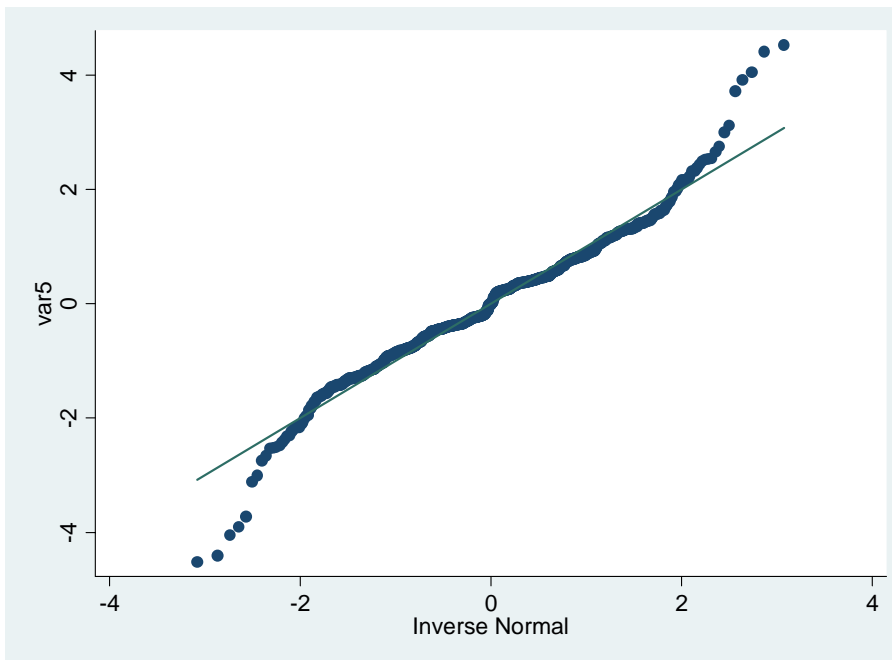


## Modelo V

Normal probability plot

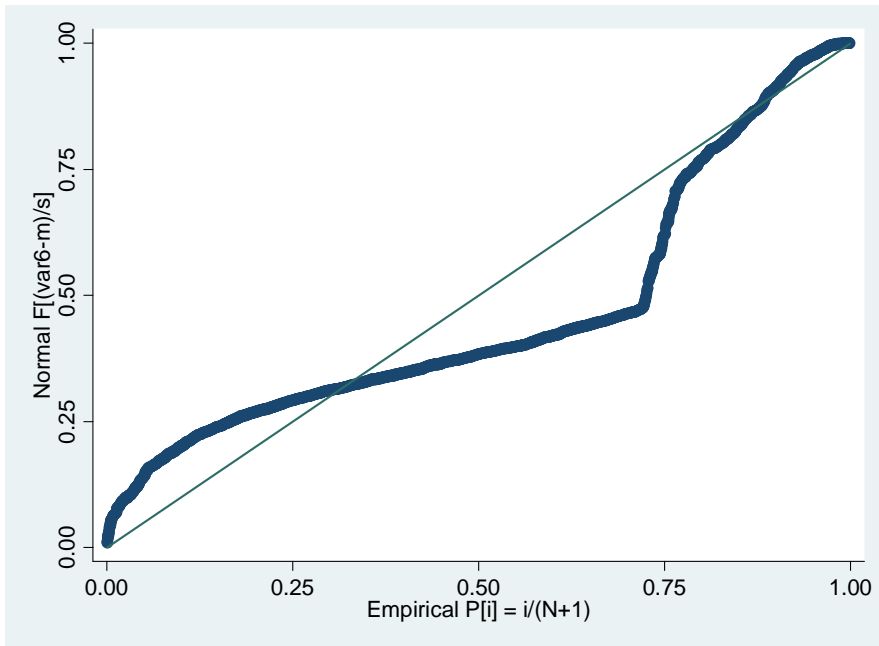


Normal quantile plot

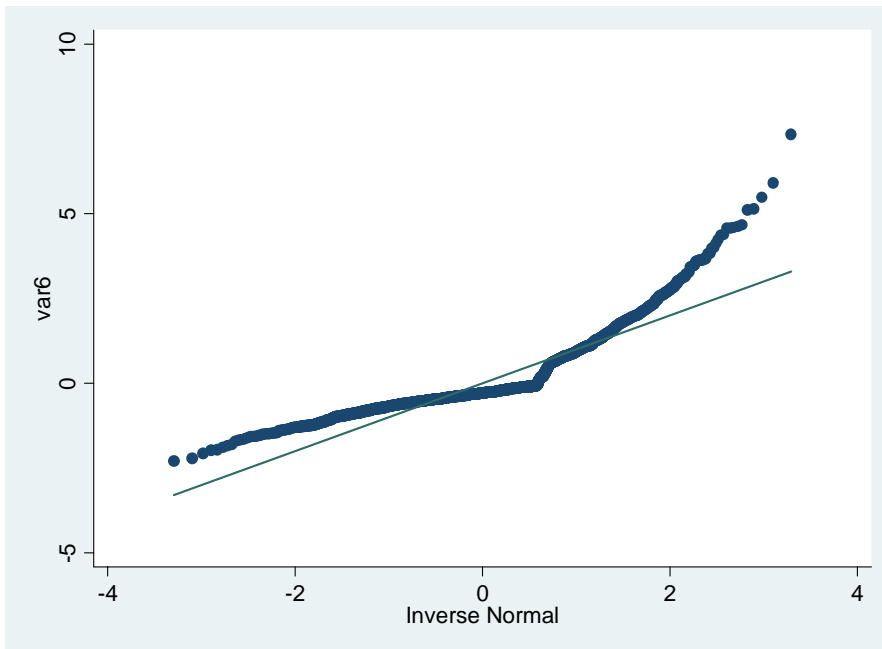


## Modelo VI

Normal probability plot



Normal quantile plot



## Cuestionario

### I. PERCEPCIÓN SOBRE LA CIENCIA

1. Me podría decir ¿Qué cree usted que es la ciencia?

2. ¿Cuáles de las siguientes frases considera que expresa mejor la idea de ciencia? (Acepte hasta dos menciones en orden de importancia)

1. Grandes descubrimientos
2. Avance técnico
3. Mejora de la vida humana
4. Comprensión del mundo natural
5. Dominio de la naturaleza
6. Transformación acelerada
7. Peligro de descontrol
8. Concentración de poder
9. Ideas que pocos entienden
10. Beneficio social
11. Otros
98. No sabe
99. No responde

3. ¿Cuáles cree usted que son, en general, las principales motivaciones que tiene un investigador para dedicarse a la ciencia en México? (Acepte hasta dos menciones en orden de importancia)

1. La búsqueda de nuevos conocimientos
2. Ayudar a solucionar problemas de la gente y de la sociedad
3. La búsqueda de prestigio
4. Ganar dinero
5. Conquistar un premio importante
6. La posibilidad de organizar su propio trabajo
7. Las condiciones laborales
8. Vocación por el conocimiento
9. Otros
98. No sabe
99. No responde

4. ¿Cuál es la imagen que tiene de la profesión de científico en México? Diría que es una profesión...

1. Muy atractiva para los jóvenes
2. Poco atractiva para los jóvenes
98. No sabe
99. No responde

1. Muy gratificante en lo personal
2. Poco gratificante en lo personal
98. No sabe
99. No responde

1. Bien remunerada económicamente
2. Mal remunerada económicamente
98. No sabe
99. No responde

1. Con un alto reconocimiento social
2. Con escaso reconocimiento social
98. No sabe
99. No responde

5. ¿Conoce alguna institución que se dedique a hacer investigación científica en nuestro país?

1. Sí
  2. No
  98. No sabe
  99. No responde
- 5a. ¿Cuál o cuáles? Acepte hasta tres menciones

6. ¿Cuál cree que es el lugar de trabajo de la mayoría de los científicos en nuestro país? (Acepte hasta dos menciones en orden de importancia)

1. Organismos públicos
2. Universidades
3. Empresas
4. Fundaciones privadas
5. Hospitales
6. Otros
98. No sabe
99. No responde

7. ¿Dónde cree usted que se desarrolla la ciencia en México?

1. Organismos públicos
2. Universidades públicas
3. Universidades privadas
4. Empresas
5. Fundaciones
6. Hospitales
7. Otros
98. No sabe
99. No responde

8. ¿Usted cree que el trabajo científico para el país es: Muy importante, importante, poco importante o nada importante?

- 1. Muy importante
- 2. Importante
- 3. Poco importante
- 4. Nada importante
- 98. No sabe
- 99. No responde

9. ¿Quién piensa que aporta más dinero para la investigación científica en el país? (Acepte sólo una mención)

- 1. El gobierno
- 2. Las empresas
- 3. Las universidades
- 4. Las fundaciones privadas
- 5. Las instituciones extranjeras
- 6. Otros
- 98. No sabe
- 99. No responde

10. ¿Usted cree que el trabajo del científico en México es apoyado por el gobierno?

- 1. Si
- 2. No
- 98. No sabe
- 99. No responde

11. Dígame por favor, si está totalmente de acuerdo, de acuerdo, en desacuerdo, o totalmente en desacuerdo con la siguiente afirmación: “El gobierno debe aumentar los recursos que destina a la investigación científica en nuestro país”.

- 1. Totalmente de acuerdo
- 2. De acuerdo
- 3. En desacuerdo
- 4. Totalmente en desacuerdo
- 98. No sabe
- 99. No responde

12. ¿Usted cree que en México el trabajo del científico es valorado por la sociedad?

- 1. Si
- 2. No
- 98. No sabe
- 99. No responde

## II. PERCEPCIÓN DE LA UNAM

13. ¿Podría decirme qué universidades públicas del país conoce? (Acepte hasta tres menciones en orden de importancia)

14. ¿Sabe usted qué significa UNAM?

1. Universidad Nacional Autónoma de México
2. Universidad Nacional de México
3. Universidad Autónoma de México
4. Universidad Autónoma de la Ciudad de México
5. Otros
98. No sabe
99. No responde

15. ¿Usted estudia (o estudió) en la UNAM?, Trabaja (o ha trabajado) en la UNAM?

1. Si
2. No
98. No sabe
99. No responde

16. ¿Usted conoce a alguien que estudie (o haya estudiado) en la UNAM? o trabaje (o haya trabajado) en la UNAM?

1. Si
2. No
99. No responde
17. ¿Podría decirme quien? (Especificar parentesco)

18. ¿Qué conoce de la UNAM?

1. Escuelas
2. Facultades
3. Institutos
4. Centros
5. Bibliotecas
6. Museos
7. Lugares de entretenimiento
8. Salas de cine
9. Teatros
10. Tienda UNAM
11. Estadio UNAM
12. Equipo Pumas
13. Jardín Botánico
13. Otros



19. ¿Qué actividades cree usted que se desarrollan en la UNAM?  
(Señale las que el entrevistado(a) mencione)

1. Enseñanza (docencia)
2. Investigación
3. Ciencia
4. Culturales
5. Deportivas
6. De entretenimiento
7. Otros
98. No sabe
99. No responde

20. ¿Cree que en la UNAM se realiza ciencia?

1. Si
2. No
98. No sabe
99. No responde

21. ¿En qué áreas del conocimiento? (Acepte hasta tres menciones)

1. Ciencias Físico - Matemáticas y las Ingenierías
2. Ciencias Biológicas y de la Salud
3. Ciencias Sociales
4. Humanidades y las Artes
98. No sabe
99. No responde

22. ¿Qué actividades cree usted que se realizan en esa o esas áreas del conocimiento?

23. ¿Sabe usted de alguna persona que se dedique a la investigación científica en la UNAM?

1. Si
2. No
99. No responde

24. ¿Cómo se llama y en qué trabaja?

25. Cree usted que la ciencia que se desarrolla en la UNAM aporta ventajas o desventajas para...

- a) El desarrollo económico
- b) La calidad de vida en la sociedad
- c) La seguridad y la protección de la vida humana
- d) La conservación del medio ambiente y la naturaleza
- e) Hacer frente a las enfermedades y epidemias
- f) Los productos de alimentación y producción agrícola
- g) La generación de nuevos puestos de trabajo

h) El incremento y mejora de las relaciones entre las personas

26. ¿Para usted que tanto destaca la UNAM en las siguientes áreas?

- a) Educación
- b) Deportes
- c) Desarrollo industrial
- d) Salud
- e) Desarrollo de tecnologías
- f) Arte y cultura
- g) Investigación científica
- h) Difusión
- i) Participación en discusiones políticas, económicas y sociales

27. ¿Usted se informa a través de algún medio de comunicación sobre la ciencia que se desarrolla en la UNAM?

- 1. Si
- 2. No
- 99. No responde

28. Nos gustaría saber a través de qué medios de comunicación se informa usted sobre temas de ciencia que se desarrollan en la UNAM.

- 1. Prensa
- 2. Internet
- 3. Libros
- 4. Radio
- 5. Revistas
- 6. Televisión
- 7. Otros
- 98. No sabe
- 99. No responde

29. La UNAM tiene diversos medios de comunicación a través de los cuales divulga la investigación científica que ahí se realiza. De la siguiente lista ¿conoce alguno de ellos?

- a) Radio UNAM
- b) TV UNAM
- c) Revista digital universitaria
- d) E-journal
- e) Gaceta UNAM
- f) Revista ¿Cómo Ves?
- g) El Faro. La luz de la ciencia (Boletín)
- h) Cartel UNAMirada a la Ciencia

### III.DATOS DEL ENCUESTADO

30. Sexo

1. Hombre
2. Mujer

31. ¿Cuántos años cumplidos tiene?

32. ¿Sabe usted leer y escribir un recado?

1. Si
2. No
99. No responde

33. ¿Cuál fue el último año o grado que aprobó en la escuela? (Registrar año y nivel)

Ahora le voy a preguntar sobre su situación laboral...

34. ¿La semana pasada usted:

1. ¿Trabajó?
2. ¿Tenía trabajo pero no trabajó?
3. ¿Trabajó ayudando en algún negocio familiar (con pago)?
4. ¿Trabajó ayudando en algún negocio familiar (sin pago)?
5. ¿buscó trabajo?
6. ¿era estudiante?
7. ¿se dedicó a los quehaceres del hogar?
8. ¿está jubilado(a) o pensionado(a)?
9. ¿está incapacitado (a) permanentemente para trabajar?
10. No trabajó
99. No responde

35. ¿En su trabajo (actividad) principal usted es:

1. ¿Jornalero o peón?
2. ¿Empleado u obrero?
3. ¿Patrón o empresario?
4. ¿Trabajador por su cuenta?
5. ¿Profesionista independiente?
6. ¿Empleado domestico?
7. Trabajador sin pago
9. Otros
99. No responde

36. ¿Cuánto gana usted por su trabajo principal o a cuánto ascienden sus ingresos para sostenerse económicamente?

37. ¿Cuántas personas mayores de 15 años viven en esta vivienda?