

Simulación social basada en redes bayesianas: Aproximación a un enfoque lógico-estadístico

Lic. Jean Christian Díaz Preciado

Dr. Alejandro Guerra Hernández

Dr. Nicandro Cruz Ramírez



Universidad Veracruzana



Contenido

- Introducción
 - Agentes y simulación social
 - Redes bayesianas
 - Estado del arte
 - Propuesta
- Metodología
 - Protocolo ODD
 - Caso de estudio
- Resultados y conclusiones
- Trabajo a futuro



Universidad Veracruzana

Inteligencia Artificial

Cognición Individual



I
N
T
E
R
N
E
T

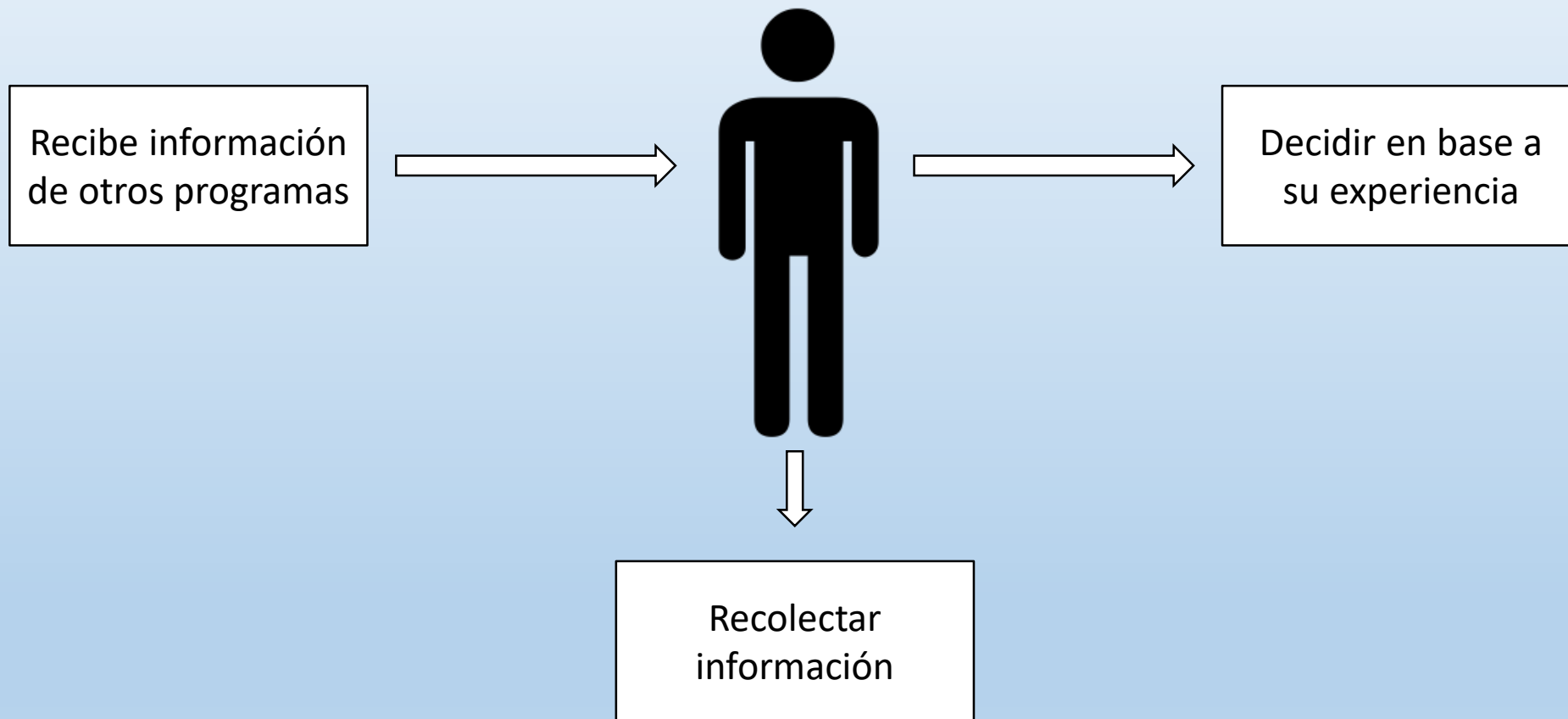
Agente



1940 ----- 1950 ----- 1960 ----- 1970 ----- 1980



Agente





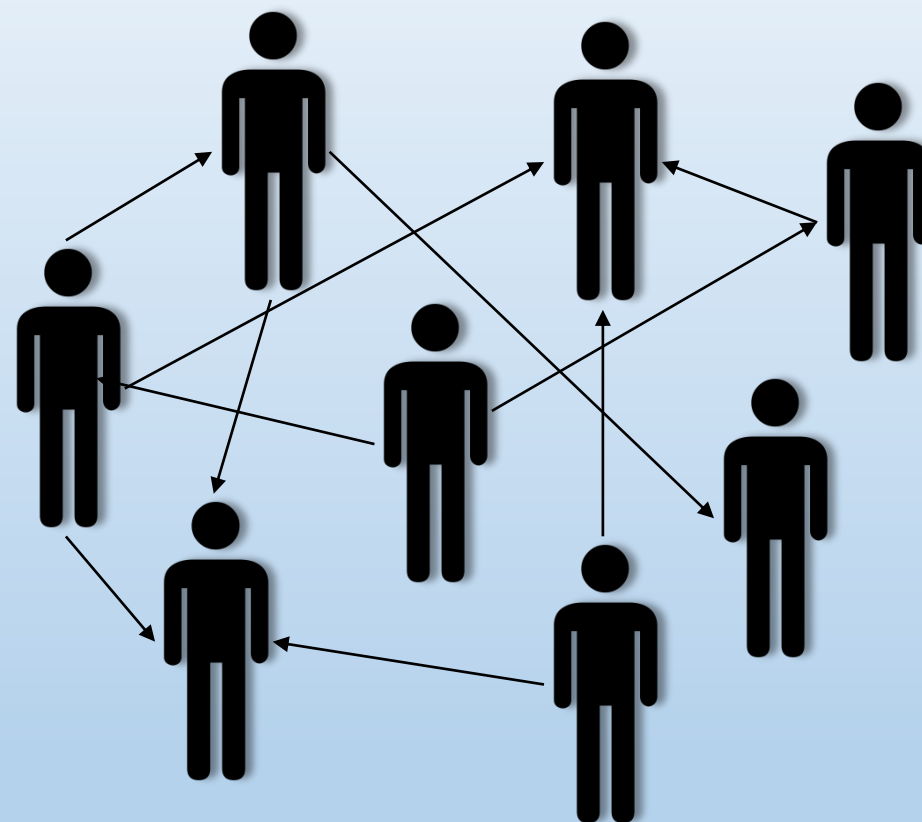
Universidad Veracruzana

Inteligencia Artificial Distribuida

Cognición Individual



I
N
T
E
R
N
E
T



Simulaciones de sociedades humanas

1940 ----- 1950 ----- 1960 ----- 1970 ----- 1980

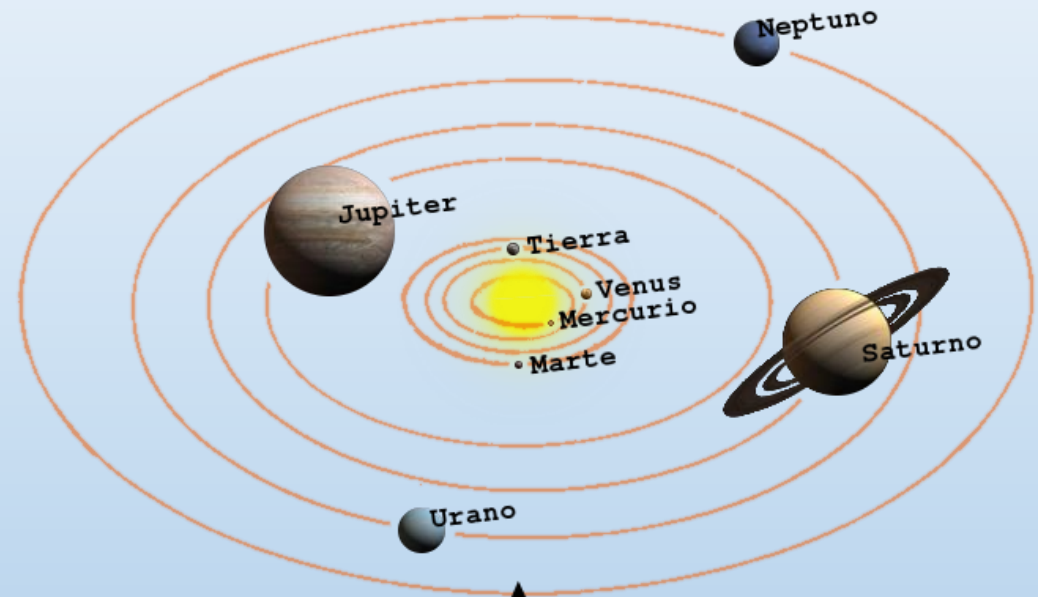
Objetivo ciencias sociales y modelos



Universidad Veracruzana



Modelo

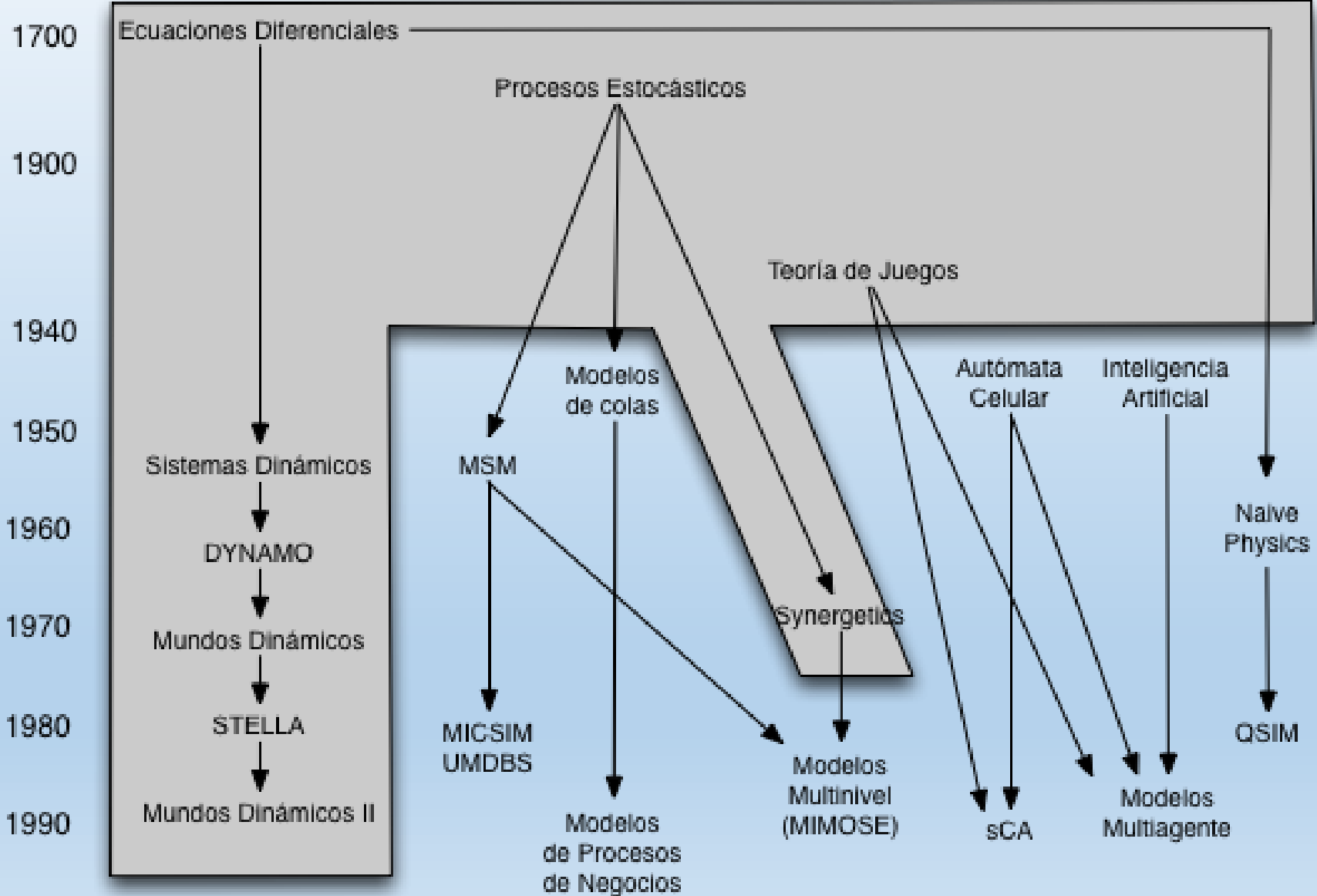


Observación

Abstracción



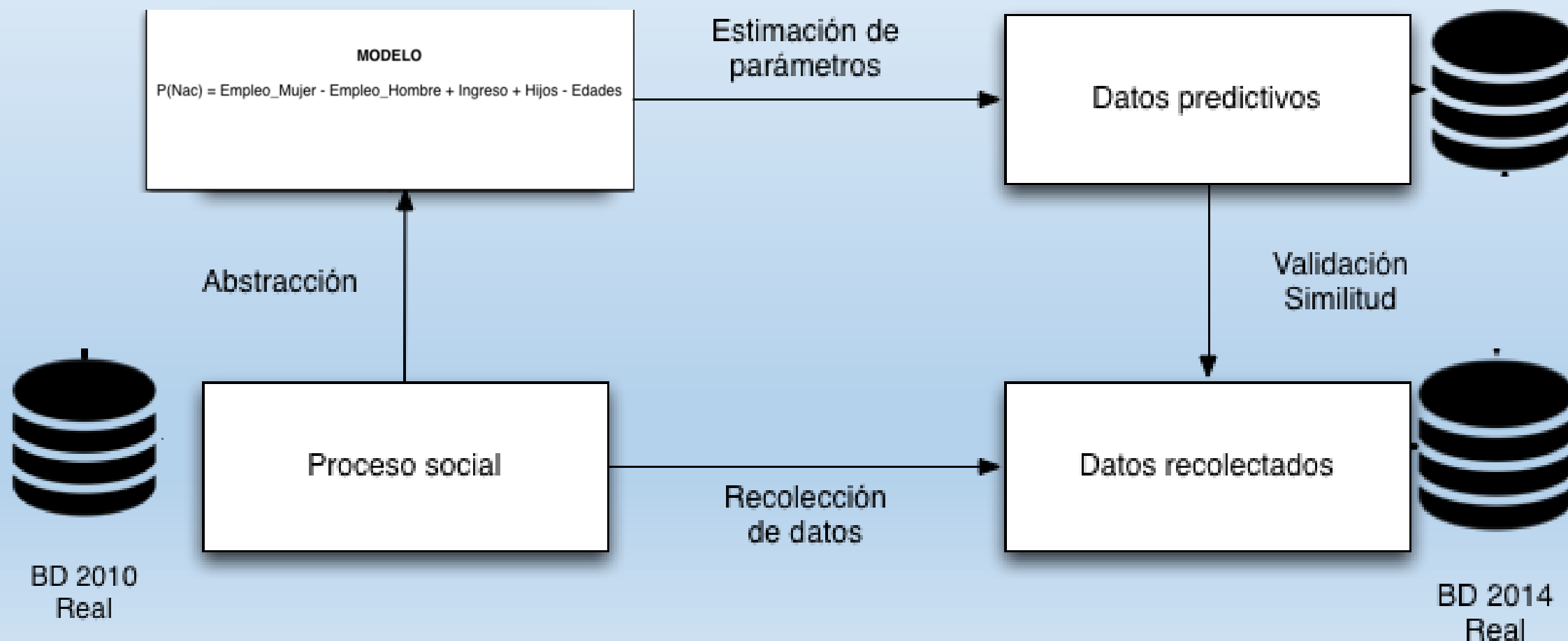
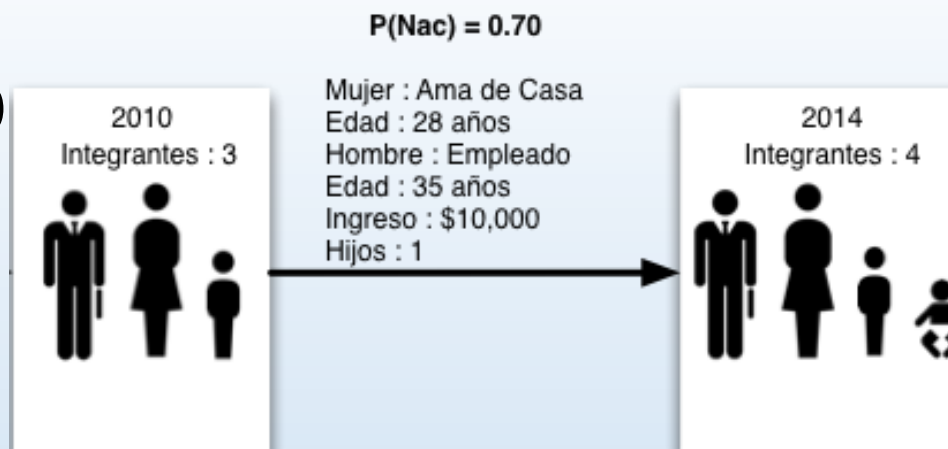
Antecedentes





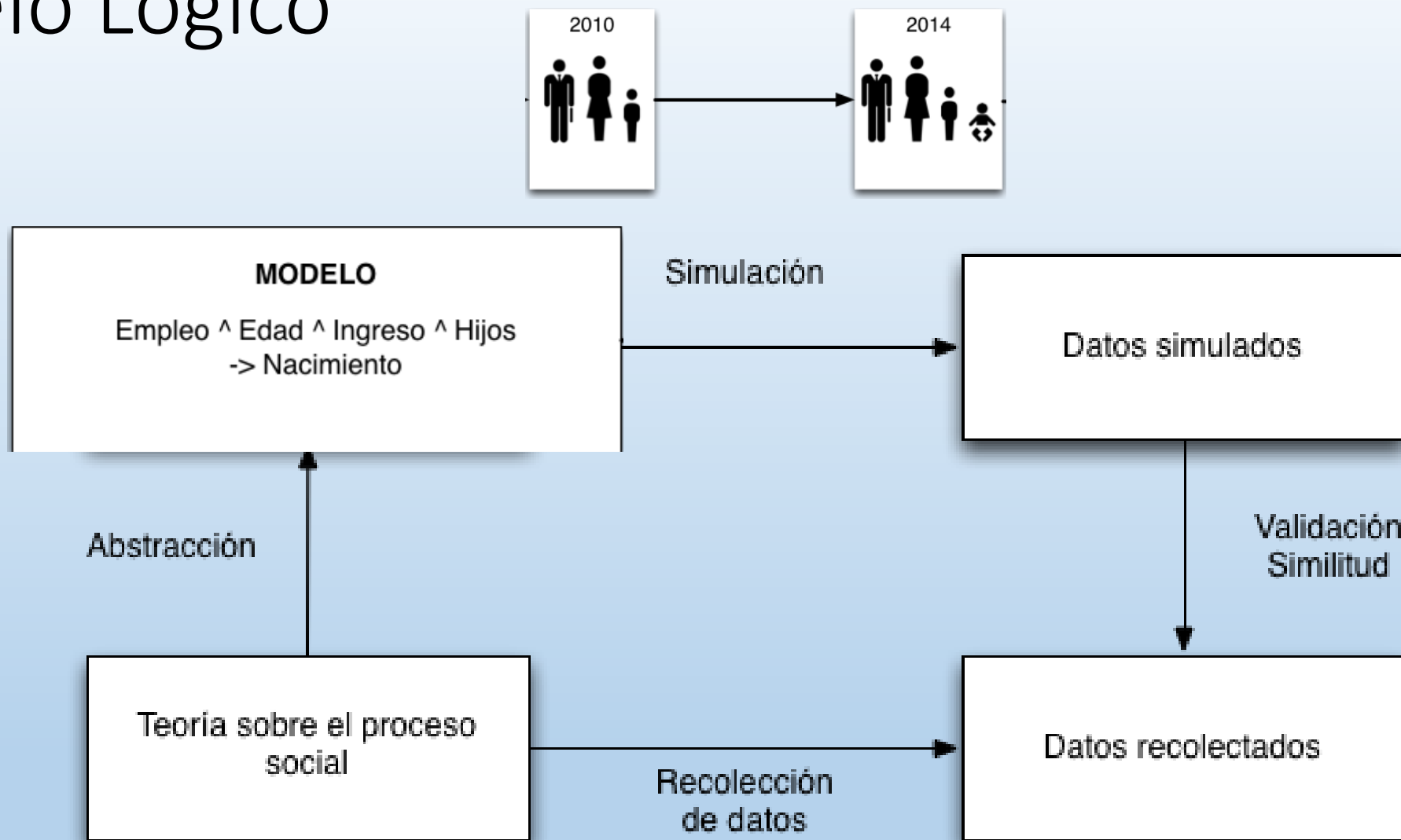
Universidad Veracruzana

Modelo Estadístico



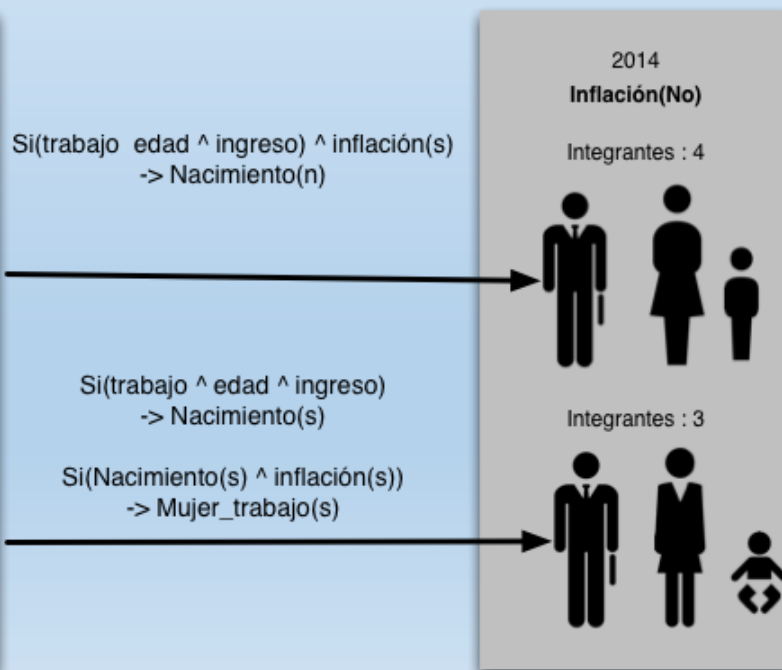
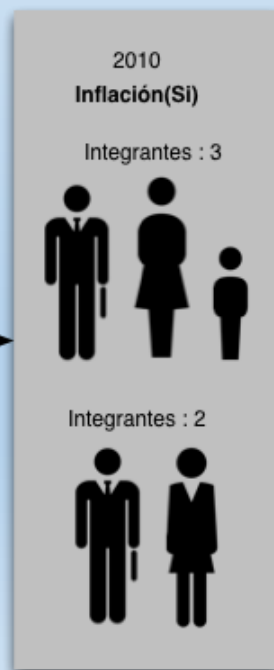
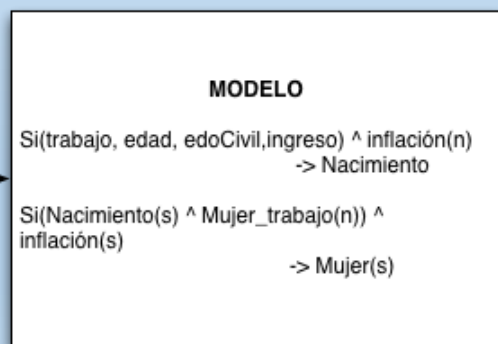
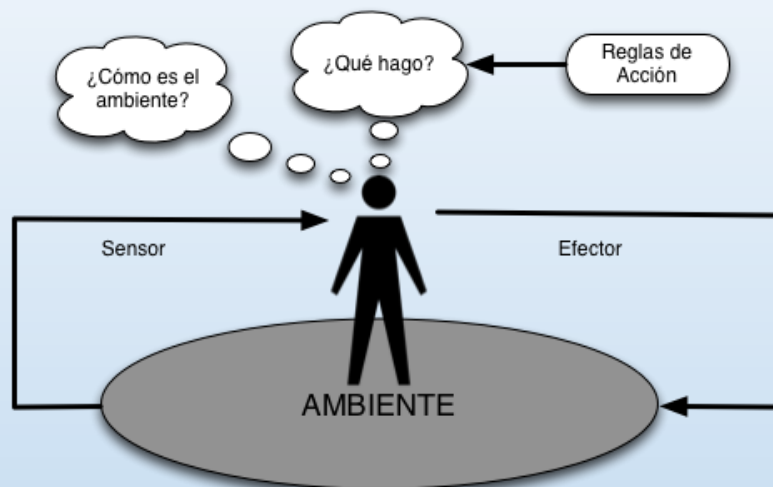


Modelo Lógico





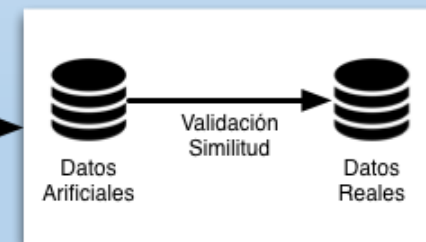
Modelo Basado en Agentes



$Si(\text{trabajo edad} \wedge \text{ingreso}) \wedge \text{inflación}(s) \rightarrow \text{Nacimiento}(n)$

$Si(\text{trabajo} \wedge \text{edad} \wedge \text{ingreso}) \rightarrow \text{Nacimiento}(s)$

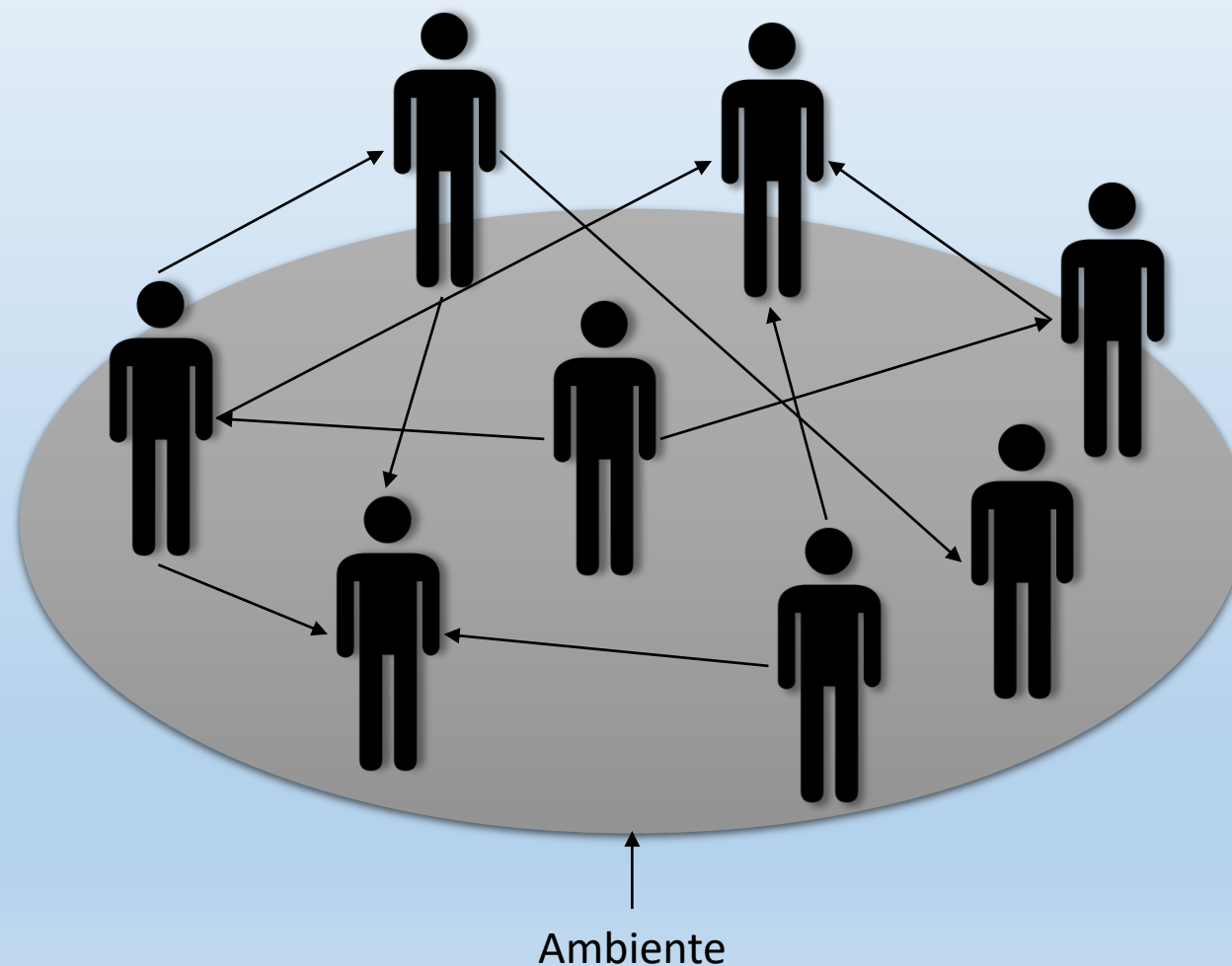
$Si(\text{Nacimiento}(s) \wedge \text{inflación}(s)) \rightarrow \text{Mujer_trabajo}(s)$





Arquitecturas de Agentes

- Con Estado
- Lógicos
- Basados en Metas
- Basados en Utilidad





Universidad Veracruzana

Validación de los modelos



Objetivo General de la Simulación

Enfoque Estadístico

- Predecir futuras distribuciones

Enfoque Lógico / Multiagente

- Similitud de comportamientos



Tipos Generales de Validación

Predicción

- Suponer futuros estados

Retrodicción

- Reproducir aspectos ya observados

Semejanza Estructural

- Comportamiento, relaciones y estructuras sociales

Técnicas de Validación



Universidad Veracruzana

Validez Aparente

- Modelo conceptual, componentes y comportamiento

Validez Histórica

- Comparación con resultados de datos con los que se cuenta

Validez de Eventos

- Nivel de trayectoria individual o en conjunto

Prueba de condiciones extremas

- Ingresar factores o combinaciones poco probables

Análisis de sensibilidad

- Variación sistemática de parámetros del modelo

De Sección Transversal

- Comparación con datos sociales reales

Comparación con otros modelos

- Diferentes modelos, mismo proceso

Elementos cruzados

- Misma arquitectura de agentes, mecanismos diferentes

Enfoques Participativos

- Participación de expertos en áreas sociales



Universidad Veracruzana

Planteamiento del Problema

Modelos Estadísticos



Validación



Explicación

Modelos lógicos



Validación



Explicación

Modelos Basados en Agentes



Validación



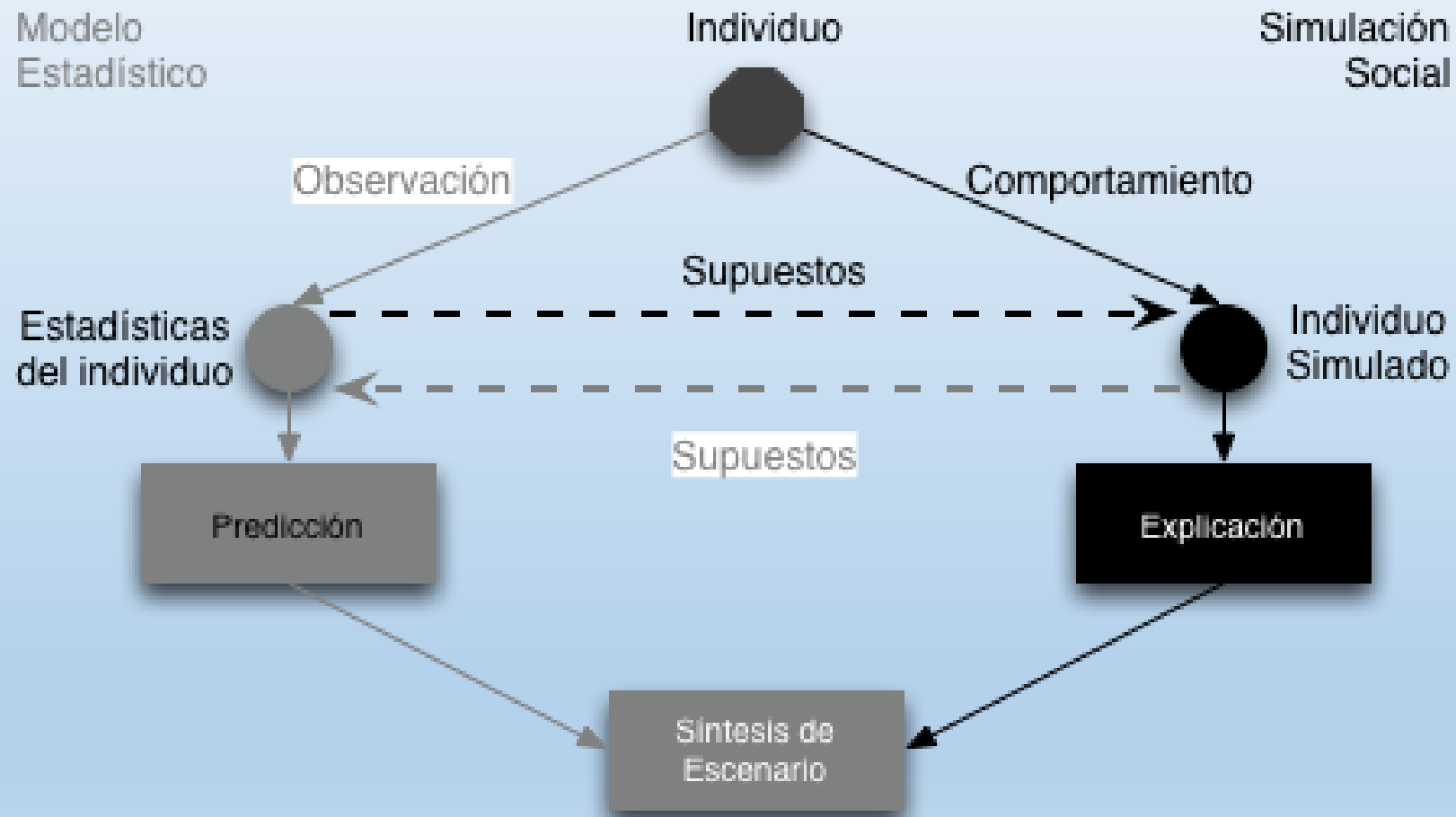
Explicación

Estado del Arte

Propuesta de Silverman



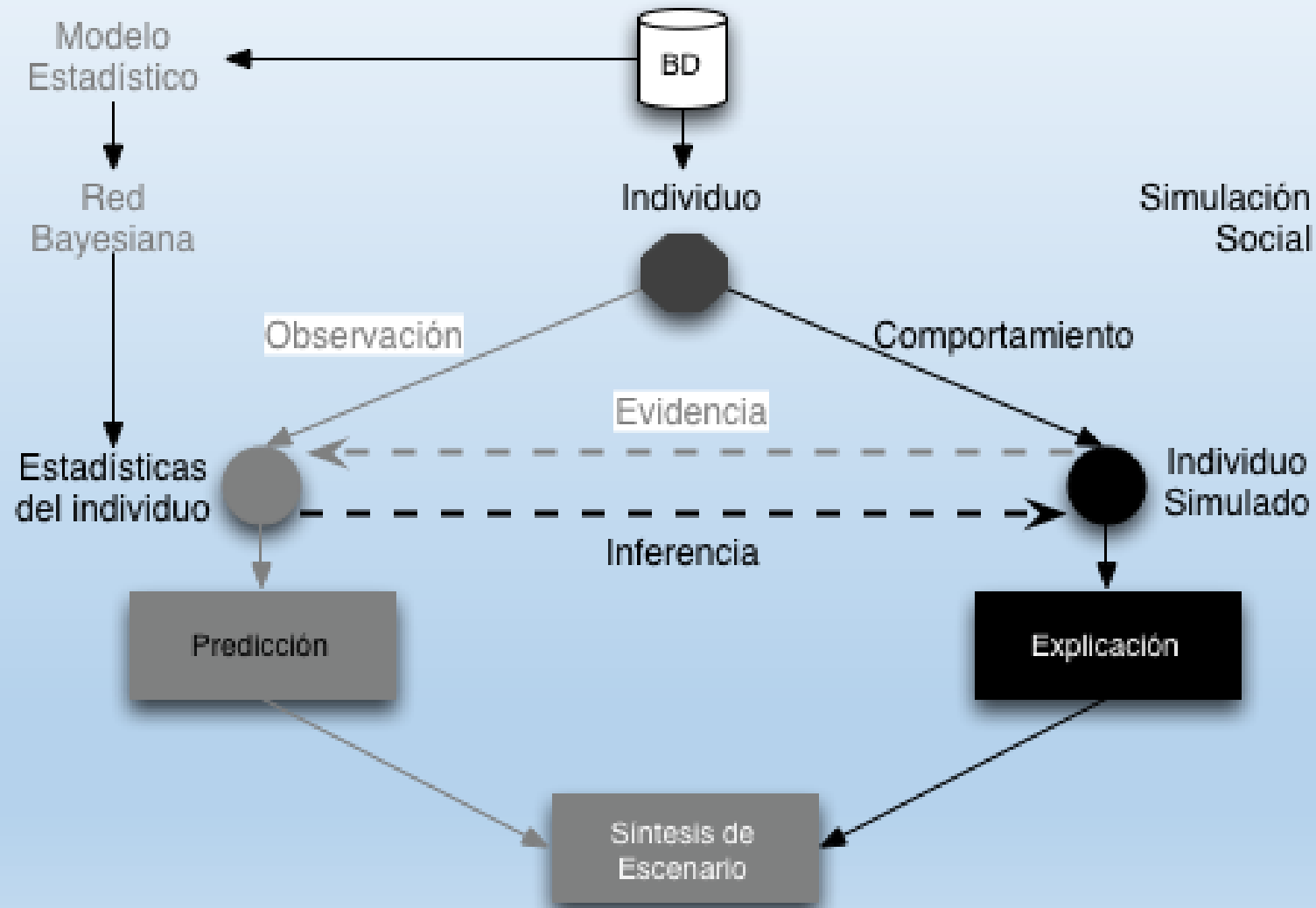
Universidad Veracruzana



Metodología Propuesta



Universidad Veracruzana



Nuevas Estadísticas : INEGI



Universidad Veracruzana

Clasificación según el tipo de la unidad económica empleadora	Clasificación según la posición en la ocupación y condición de informalidad										Totales por perspectiva de la unidad económica y/o laboral	
	Trabajadores subordinados remunerados				Empleadores		Trabajadores por cuenta propia		Trabajadores no remunerados			
	Asalariados		Con percepciones no salariales									
	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal
Sector Informal	3.844		0.792		0.904		7.444		1.192		14.177	
Trabajo doméstico remunerado	2.128	0.059	0.020	0.000							2.148	0.059
Empresas, Gobierno e Instituciones	5.373	17.122	0.913	0.211		0.988		1.272	0.598		6.884	19.593
Ámbito agropecuario	2.169	0.289	0.219	0.017		0.313	2.553		1.129		6.070	0.618
Subtotal	13.514	17.470	1.944	0.227	0.904	1.301	9.997	1.272	2.920		29.279	20.270
Total	30.984		2.172		2.205		11.269		2.920		49.549	

Unidad de medida: millones de personas

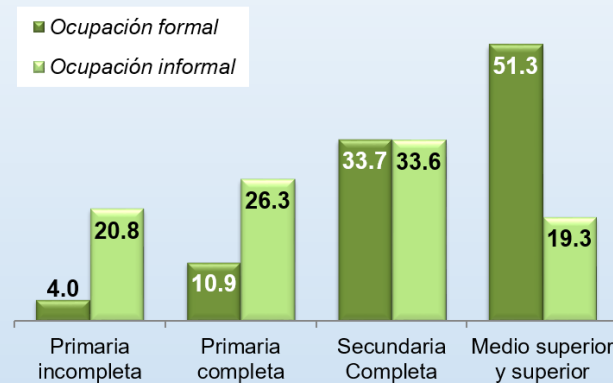
Nota: La suma de los componentes puede no coincidir con los totales debido al redondeo.

Composición porcentual por sexo



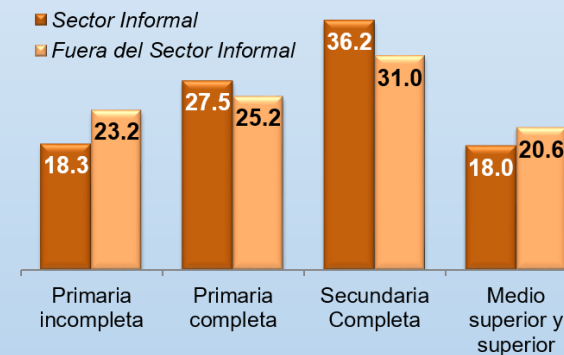
La proporción de mujeres y hombres en empleos formales e informales es similar

Fuente: INEGI, Encuesta Nacional de Ocupación y Empleo (ENOE), segundo trimestre de 2013.



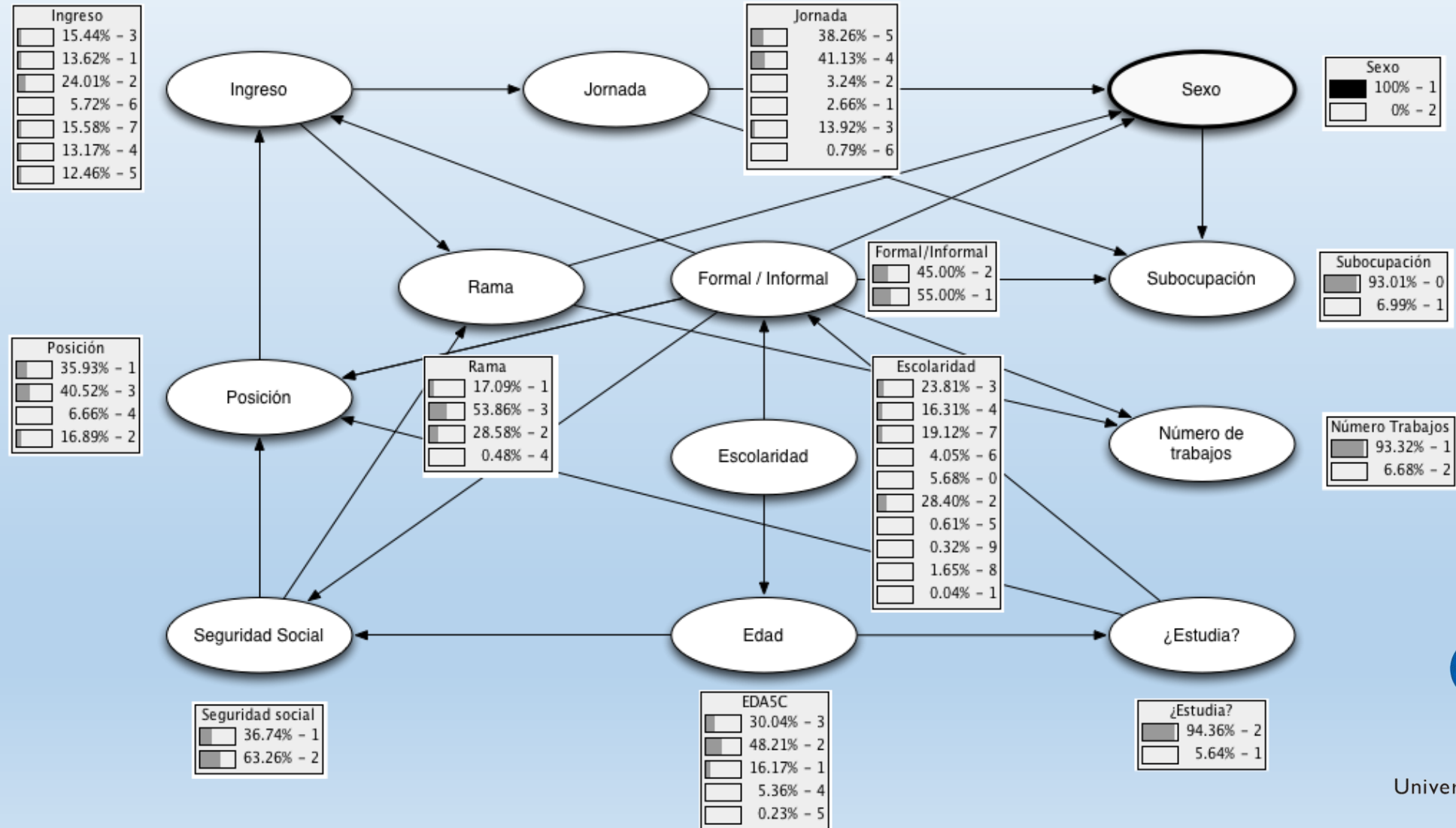
A menor nivel de preparación escolar, mayor informalidad laboral

La incidencia de informalidad por fuera del sector clásico o típico es mayor tanto entre los menos como entre los más escolarizados



Fuente: INEGI, Encuesta Nacional de Ocupación y Empleo (ENOE), segundo trimestre de 2013.

Análisis del fenómeno a través un modelo de red bayesiana





Universidad Veracruzana

Hipótesis

Una simulación social en donde la abstracción del fenómeno tome una forma de red bayesiana para el manejo la incertidumbre, percepción del ambiente o ambas, proporciona datos predictivos muy similares a la realidad



Universidad Veracruzana

Justificación

El desarrollo de una simulación social bajo un enfoque estadístico implementada en un sistema multiagente nos permite agregar modelos de comportamiento de las entidades a simular. Con esto podemos validar teorías de comportamiento en procesos sociales mediante la comparación estadística de los datos predictivos generados por el sistema con datos reales.



Objetivos

- Generar datos artificiales, mediante un sistema multiagente, que nos permitan predecir las características de las entidades a observar en periodos de tiempo determinados.
- Validar los datos generados implementando comparaciones estadísticas con datos reales.



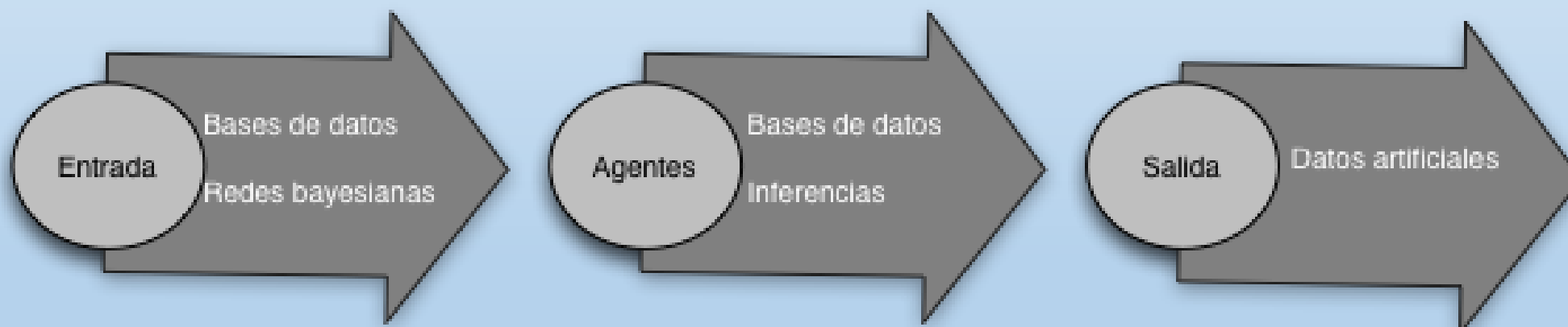
Universidad Veracruzana

Metodología



Sistema multiagente

- Plataforma Jason-CArtAgO
 - Extensión de JaCa-DDM a redes bayesianas
 - WEKA
 - Samlam





Universidad Veracruzana

Protocolo ODD

- Propósito
- Entidades, estados y escalas
- Visión general y planificación de procesos
- Submodelos
- Datos de entrada
- Inicialización
- Conceptos de diseño



Universidad Veracruzana

Caso de Estudio

- Instituto Nacional de Estadística y Geografía
 - Encuesta Nacional de Ocupación y Empleo
 - 800,000 personas al año

Tasa de desempleo (TD)

Tasa de ocupación en el sector informal (TOSI)

Tasa de subocupación (TS)



Protocolo ODD

- **Propósito**
- Entidades, estados y escalas
- Visión general y planificación de procesos
- Submodelos
- Datos de entrada
- Inicialización
- Conceptos de diseño



Universidad Veracruzana

Propósito

El propósito del sistema es observar las **tasas de subocupación y ocupación en el sector informal** en el Estado de Veracruz, en un periodo de **cuatro años y medio**, considerando las condiciones laborales que tienen las personas (**prestaciones, jornada laboral, ingreso y sector en el que se desarrolla**) dadas sus propiedades individuales como **sexo, edad, escolaridad y si se encuentran estudiando**.



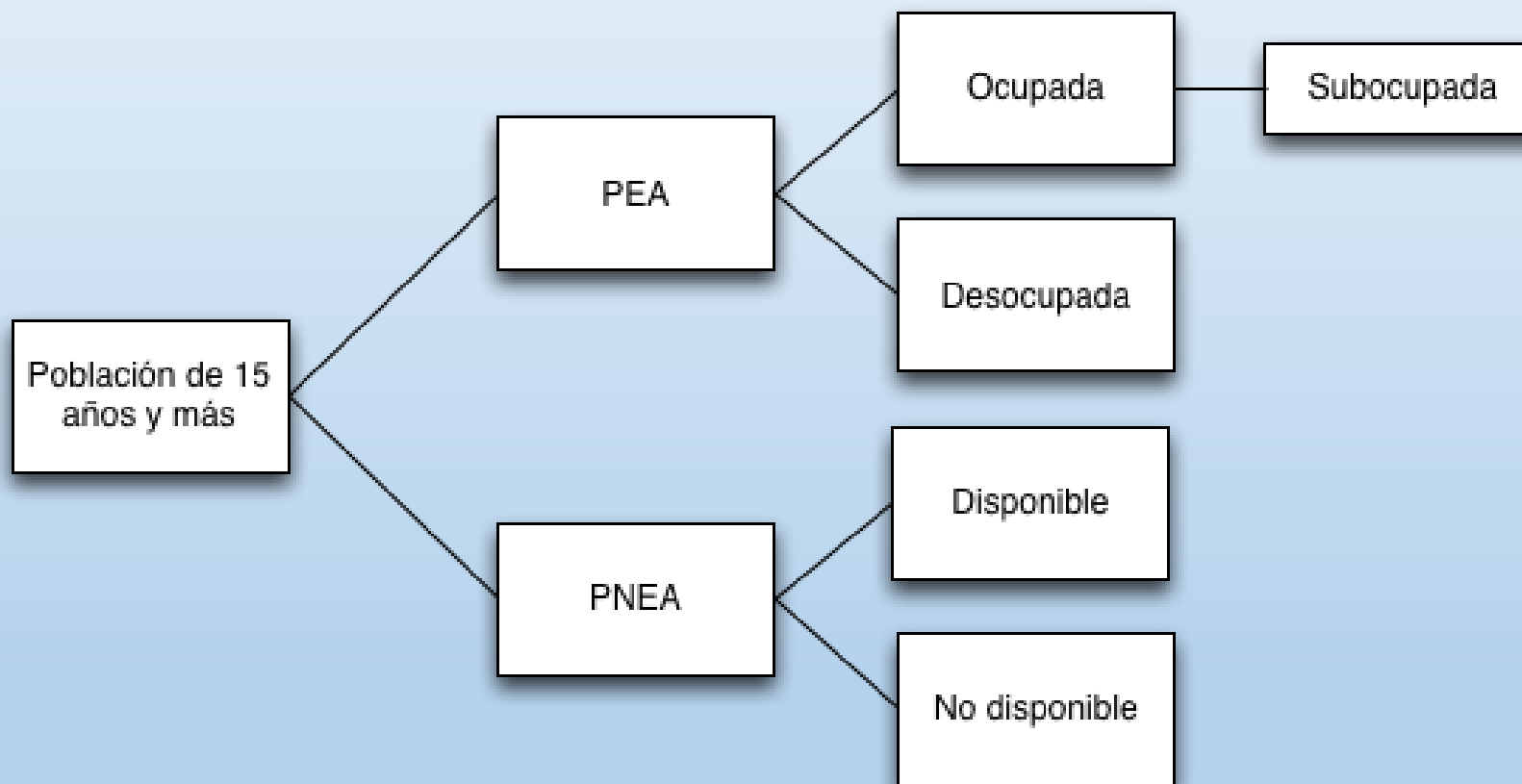
Universidad Veracruzana

Protocolo ODD

- Propósito
- **Entidades, estados y escalas**
- Visión general y planificación de procesos
- Submodelos
- Datos de entrada
- Inicialización
- Conceptos de diseño



Entidades, estados y escalas





Propiedades obtenidas de la base de datos

Atributos generales

Atributo	Variable	Descripción
Sexo	1	Hombre
	2	Mujer
Edad	1	15 a 24 años
	2	25 a 44 años
	3	45 a 64 años
	4	65 años y más
Escolaridad	1	Preescolar
	2	Primaria
	3	Secundaria
	4	Bachillerato
	5	Normal
	6	Carrera técnica
	7	Profesional
	8	Maestría
	9	Doctorado
¿Estudia?	1	Sí
	2	No

Atributos laborales

Atributo	Variable	Descripción
Jornada	1	Ausentes temporales
	2	Menos de 15 horas
	3	De 15 a 34 horas
	4	De 35 a 48 horas
	5	Más de 48 horas
Ingreso	1	Hasta un salario mínimo
	2	Más de 1 hasta 2
	3	Más de 2 hasta 3
	4	Más de 3 hasta 5
	5	Májs de 5
	6	No recibe
Posición en la ocupación	1	Subordinados y remunerados
	2	Empleadores
	3	Cuenta propia
	4	Sin pago

Atributo	Variable	Descripción
Rama	1	Primario
	2	Secundario
	3	Terciario
Seguridad social	1	Con acceso
	2	Sin acceso
Subocupación	0	No
	1	Sí
Número de empleos	1	1 empleo
	2	2 empleos
Formalidad/ Informalidad	1	Informal
	2	Formal



Protocolo ODD

- Propósito
- Entidades, estados y escalas
- **Visión general y planificación de procesos**
- Submodelos
- Datos de entrada
- Inicialización
- Conceptos de diseño

Cambios de estado de los agentes



Universidad Veracruzana

t_n	t_{n+1}	t_n	t_{n+1}
Ocupado	Desocupado	Menor	Ocupado
Desocupado	Ocupado		Desocupado
Disponibile	Ocupado		Disponibile
Disponibile	Desocupado		No disponible

2010/1	2010/2	2010/3	2010/4	2011/1	2011/2	2011/3	2011/4	2012/1	2012/2	2012/3	2012/4
3.7%	4.7%	4.7%	4.6%	5.1%	4.3%	6.1%	5.3%	3.9%	3.8%	5.9%	4.3%

	t_n	t_{n+1}
Tasa de desempleo (TD)	3.7%	4.7%
Ocupados	96	95
Desocupados	4	5

Desocupado a ocupado



Universidad Veracruzana

	Propiedades	t_n	t_{n+1}
Propiedades generales	Sexo	Mujer	Mujer
	Edad	24	24
	Escolaridad	Bachillerato	Bachillerato
	¿Estudia?	No	No
Condición		Desocupado	Ocupado
Propiedades laborales	Jornada		15 a 34 horas
	Ingreso		\$3,000
	Rama		Terciario
	Posición		Empleado
	Subocupación		Sí
	Seguridad social		No
	Clasificación empleo		Informal



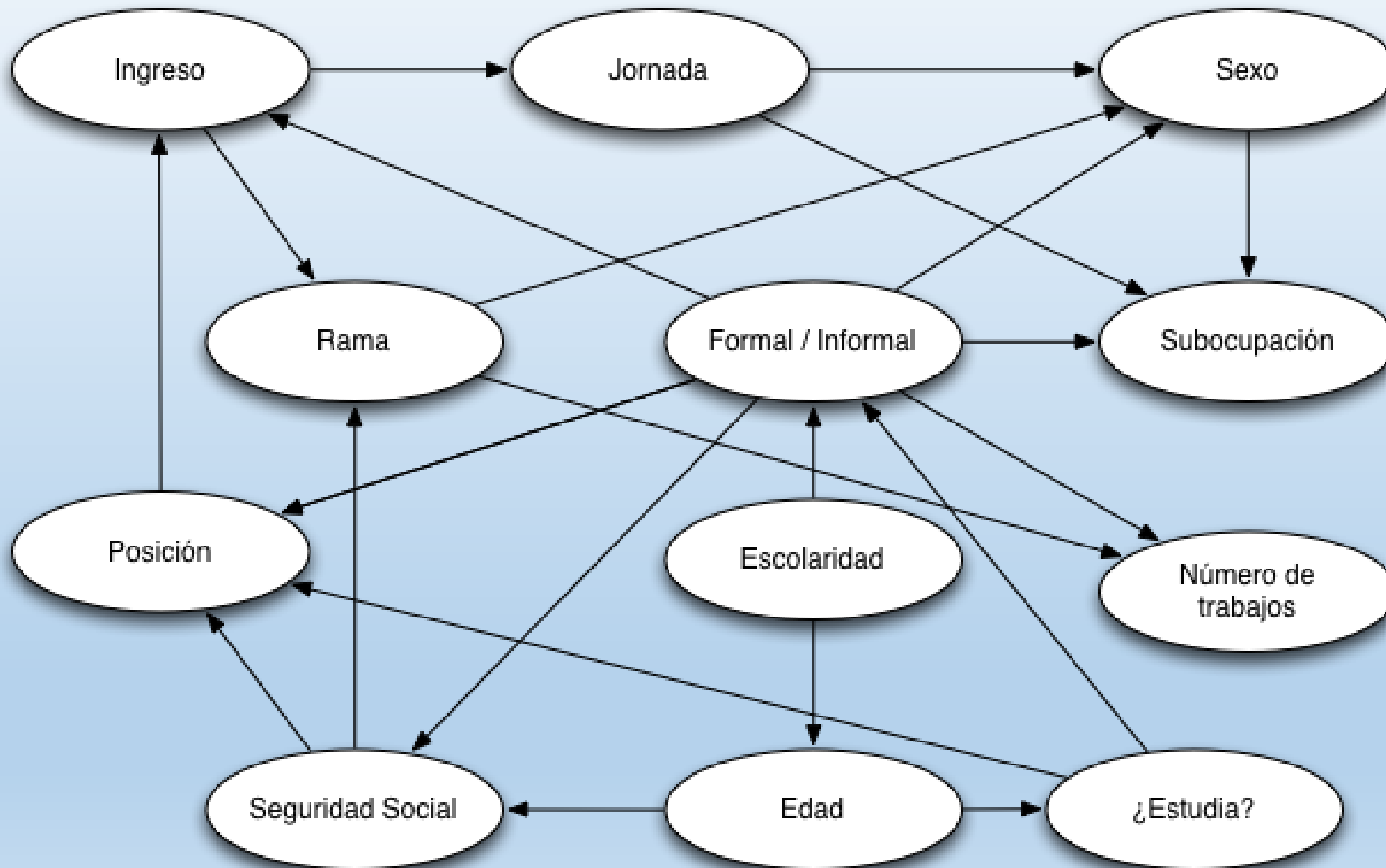
Protocolo ODD

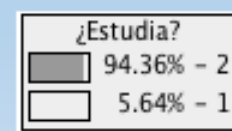
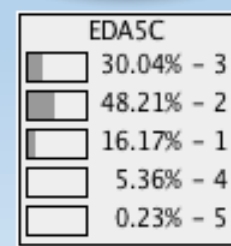
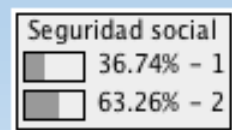
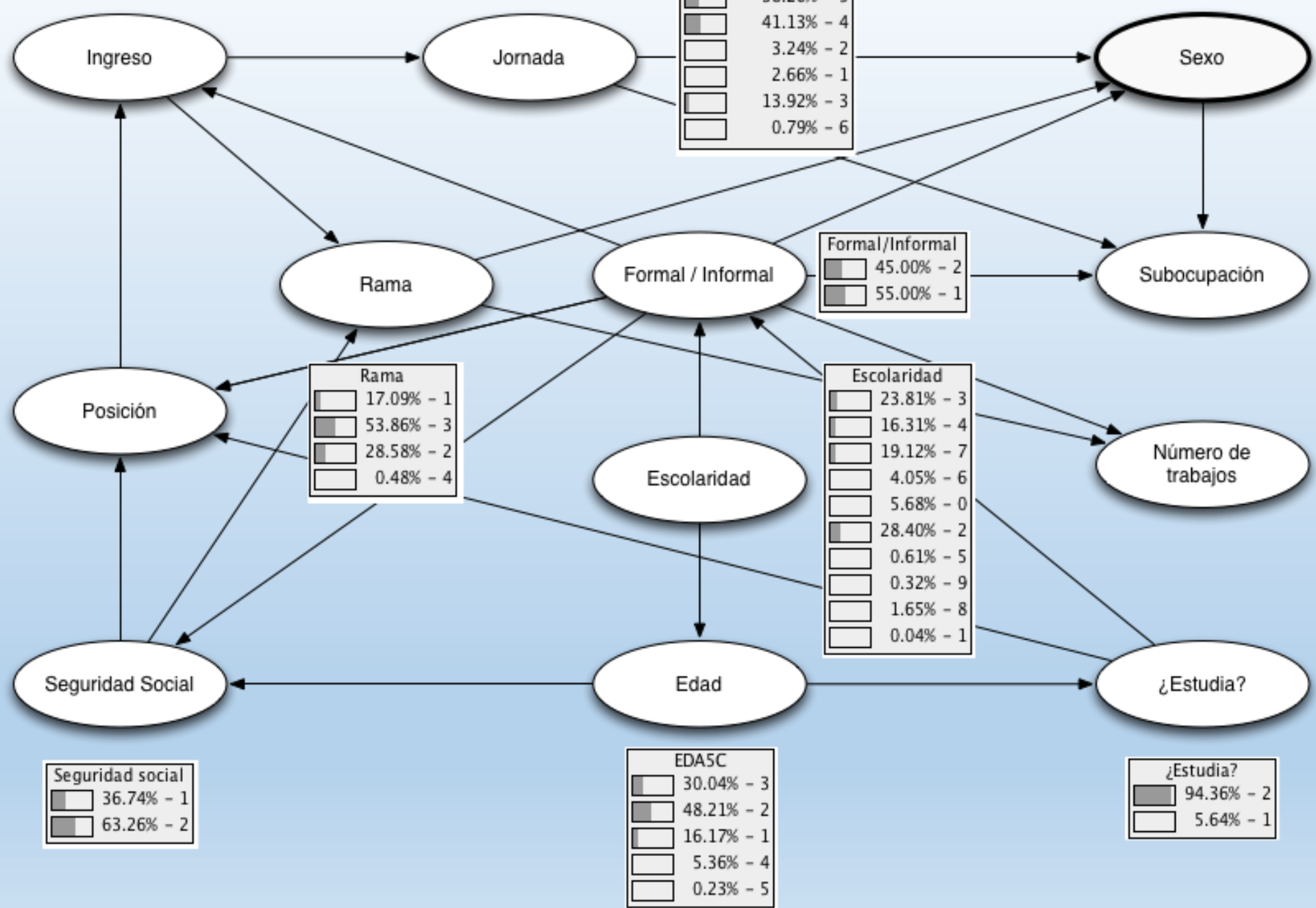
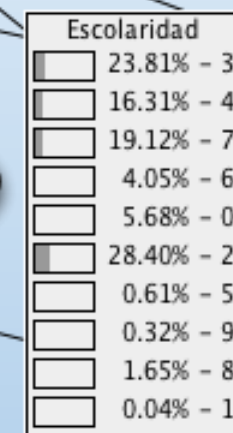
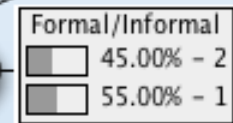
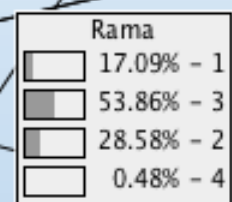
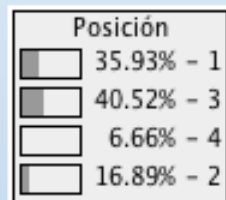
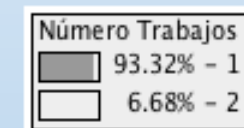
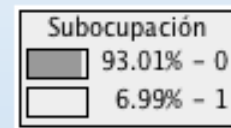
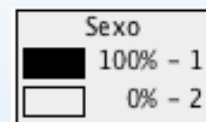
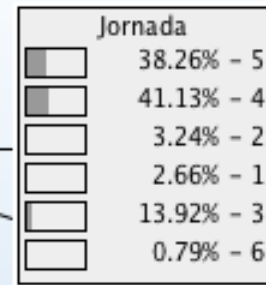
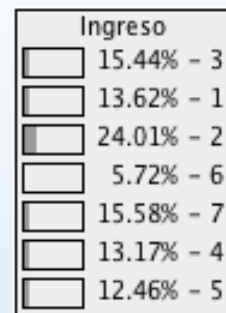
- Propósito
- Entidades, estados y escalas
- Visión general y planificación de procesos
- **Submodelos**
- Datos de entrada
- Inicialización
- Conceptos de diseño

Red Bayesiana de condiciones laborales



Universidad Veracruzana

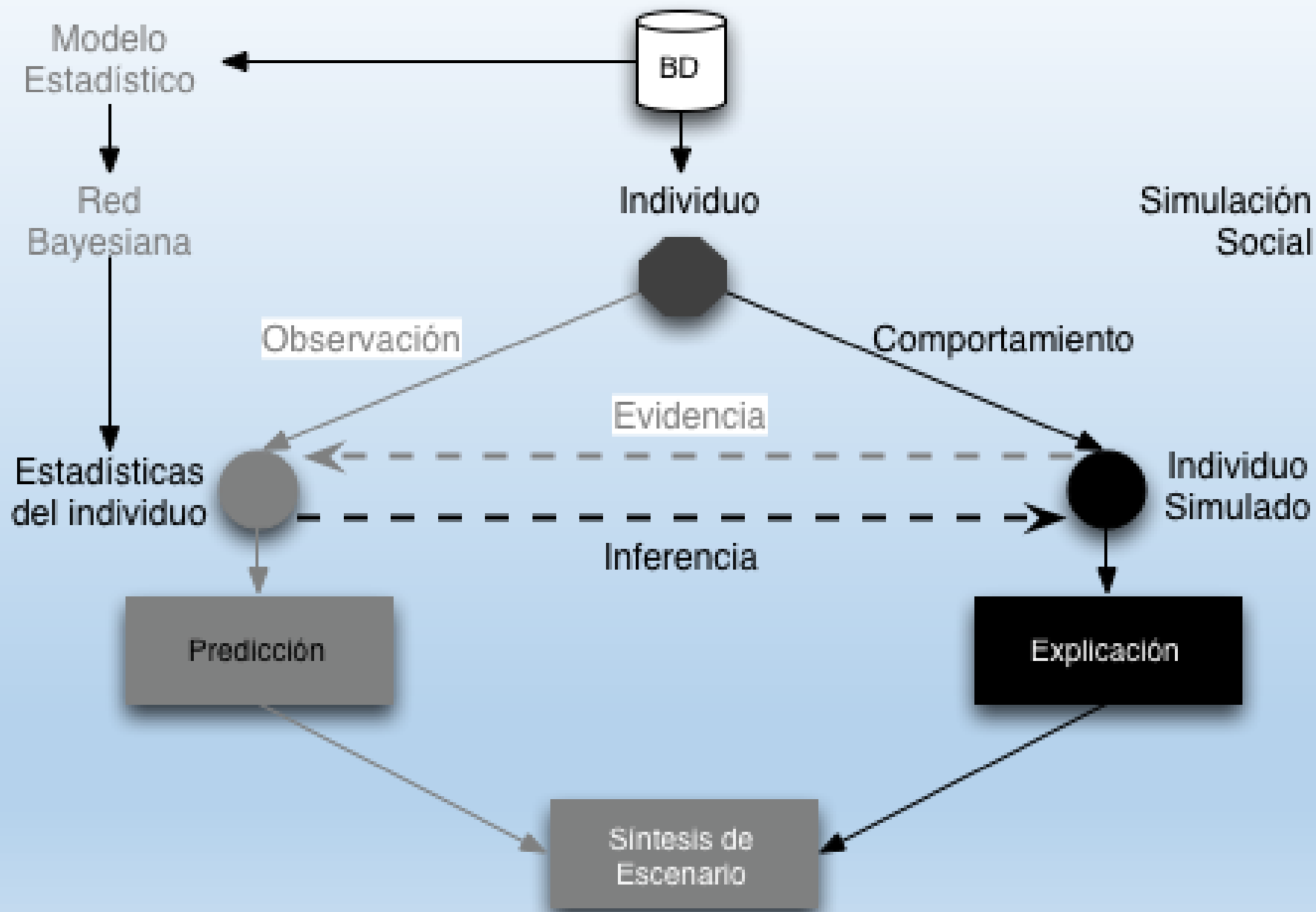


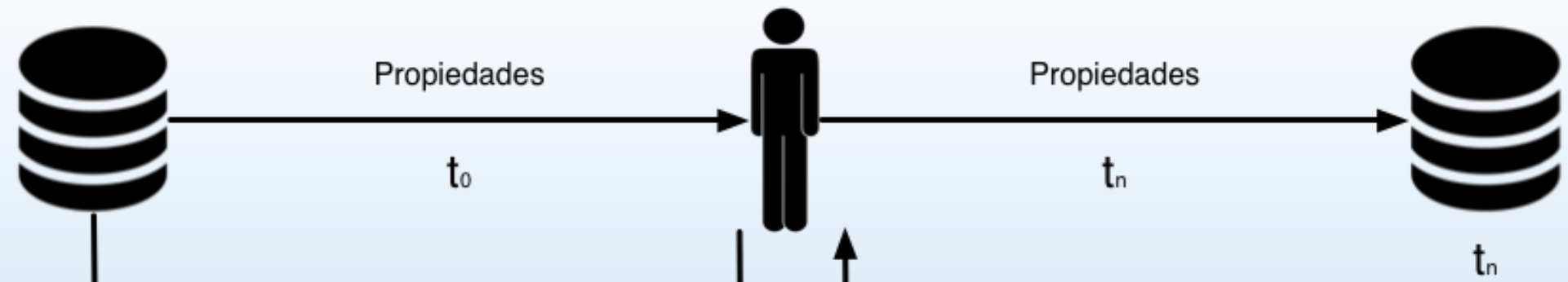




Protocolo ODD

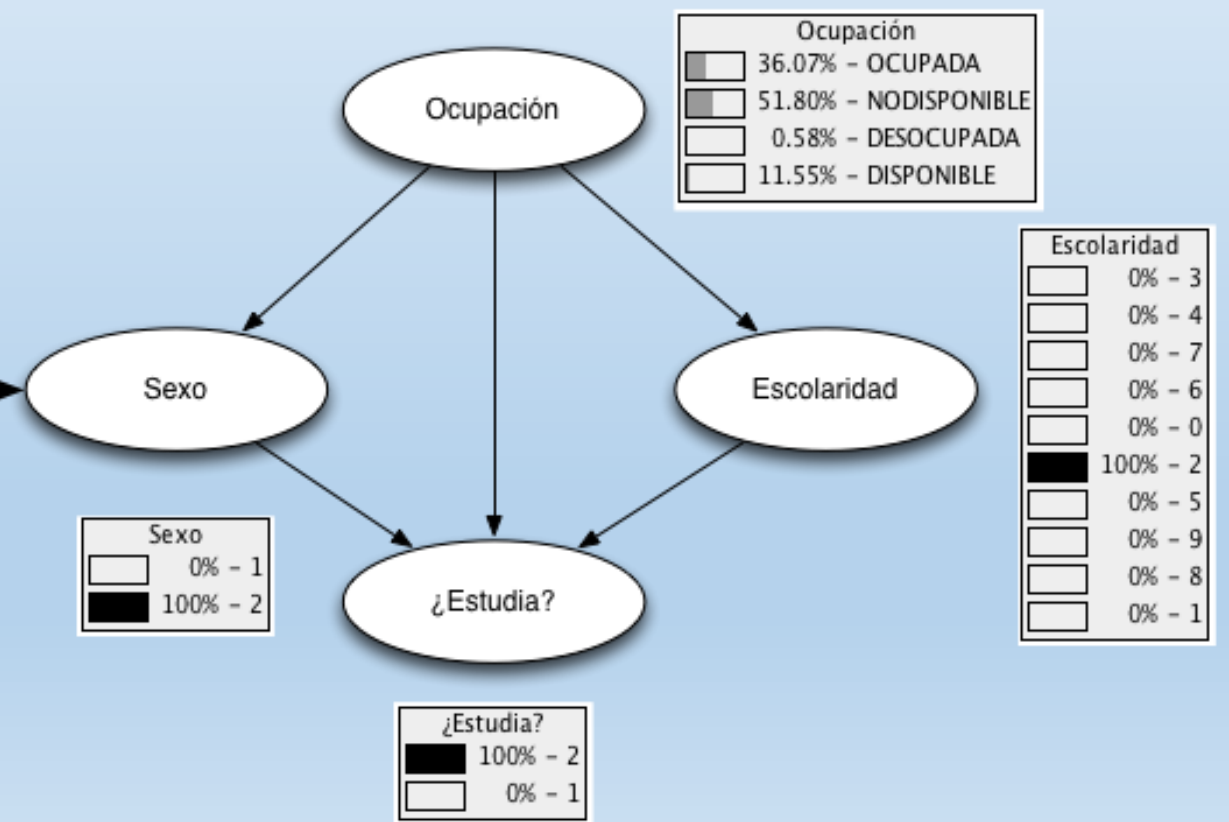
- Propósito
- Entidades, estados y escalas
- Visión general y planificación de procesos
- Submodelos
- Datos de entrada
- **Inicialización**
- **Conceptos de diseño**





Evidencia:
Propiedades
Generales

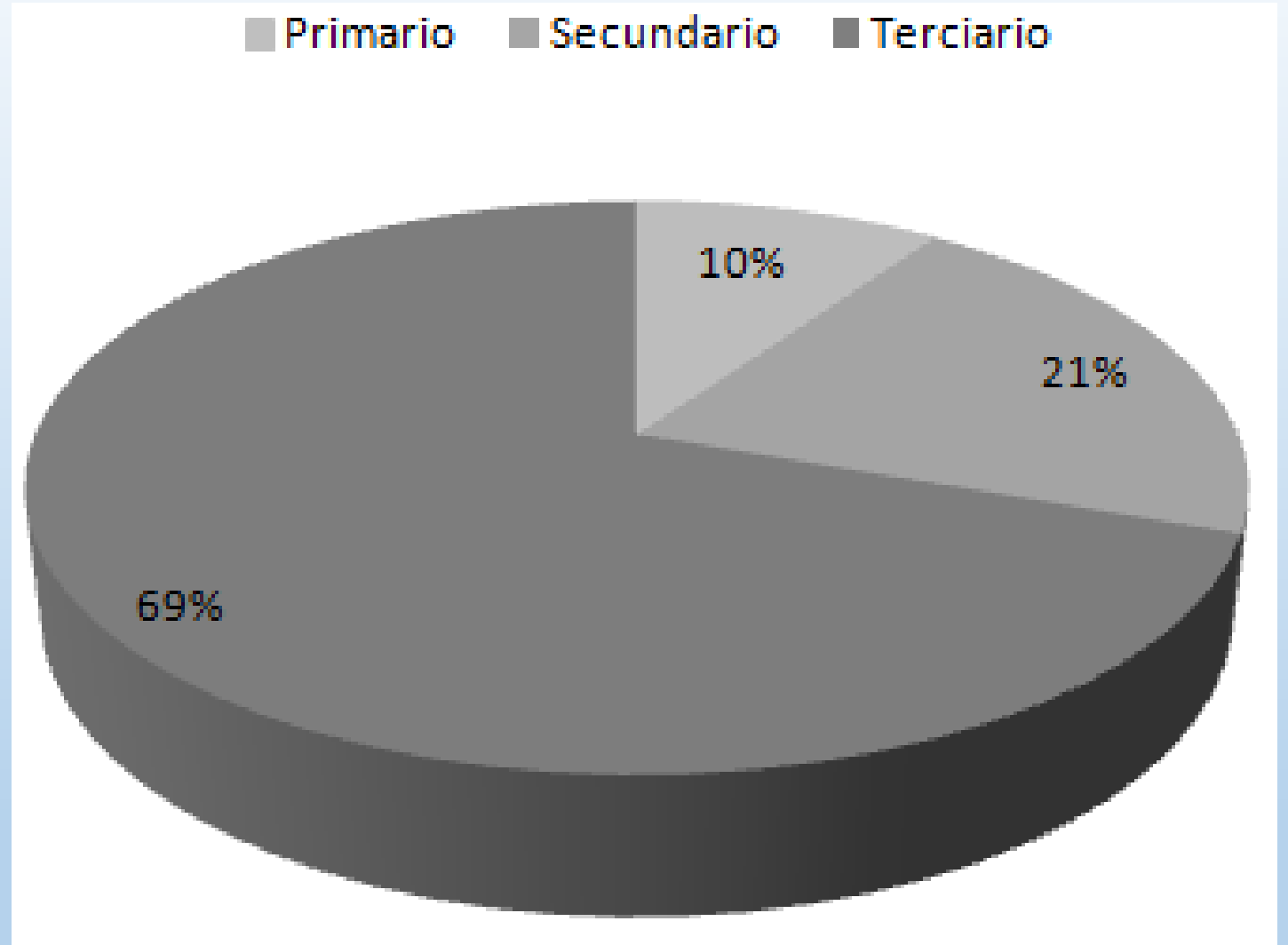
TCP:
Ocupación



Ruleta (DeJong)

TCP : Rama

Primario	10%
Secundario	21%
Terciario	69%





Universidad Veracruzana

Resultados

Red Bayesiana



Universidad Veracruzana

Estimador	Estimador Simple	A 0.5
Algoritmo de búsqueda	HillClimber	
	initNaiveBayes	False
	MarkovBlanket	True
	mxNrOfParents	10,000
	scoreTYPE	MDL
	useArcReversal	True

Parámetros para generar el modelo (WEKA)

Red bayesiana ocupación	Desempeño	65.73%
	DesvEst	1.42
Red bayesiana condición laboral	Desempeño	94.11%
	DesvEst	0.74

Porcentaje de clasificación y desviación estándar



Validación de resultados

Objetivo General
de la Simulación

- Predicción

Tipo de
Validación

- Retrodicción

Técnica de
Validación

- Sección transversal

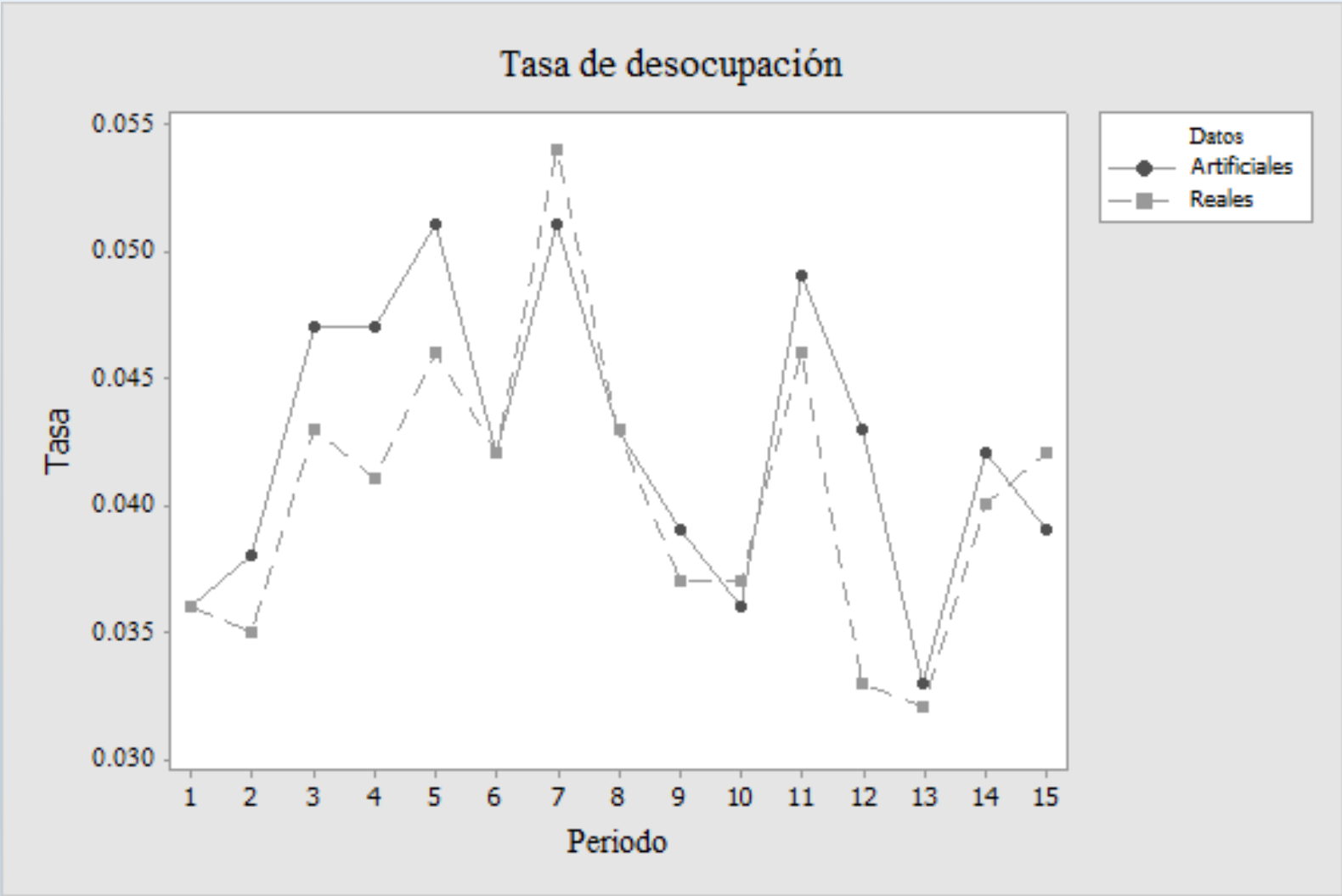
Comparación a través de tasas



Universidad Veracruzana

Periodo	TD		TS		TOSI	
	Artificial	Real	Artificial	Real	Artificial	Real
1	0.036	0.036	0.043	0.043	0.583	0.577
2	0.038	0.035	0.043	0.048	0.581	0.580
3	0.047	0.043	0.048	0.049	0.567	0.567
4	0.047	0.041	0.047	0.044	0.565	0.574
5	0.051	0.046	0.043	0.059	0.575	0.582
6	0.042	0.042	0.049	0.046	0.573	0.581
7	0.051	0.054	0.055	0.050	0.563	0.588
8	0.043	0.043	0.054	0.055	0.575	0.593
9	0.039	0.037	0.059	0.058	0.559	0.586
10	0.036	0.037	0.055	0.046	0.565	0.586
11	0.049	0.046	0.058	0.047	0.569	0.590
12	0.043	0.033	0.058	0.036	0.594	0.575
13	0.033	0.032	0.061	0.038	0.588	0.579
14	0.042	0.040	0.066	0.042	0.578	0.594
15	0.039	0.042	0.063	0.059	0.582	0.595

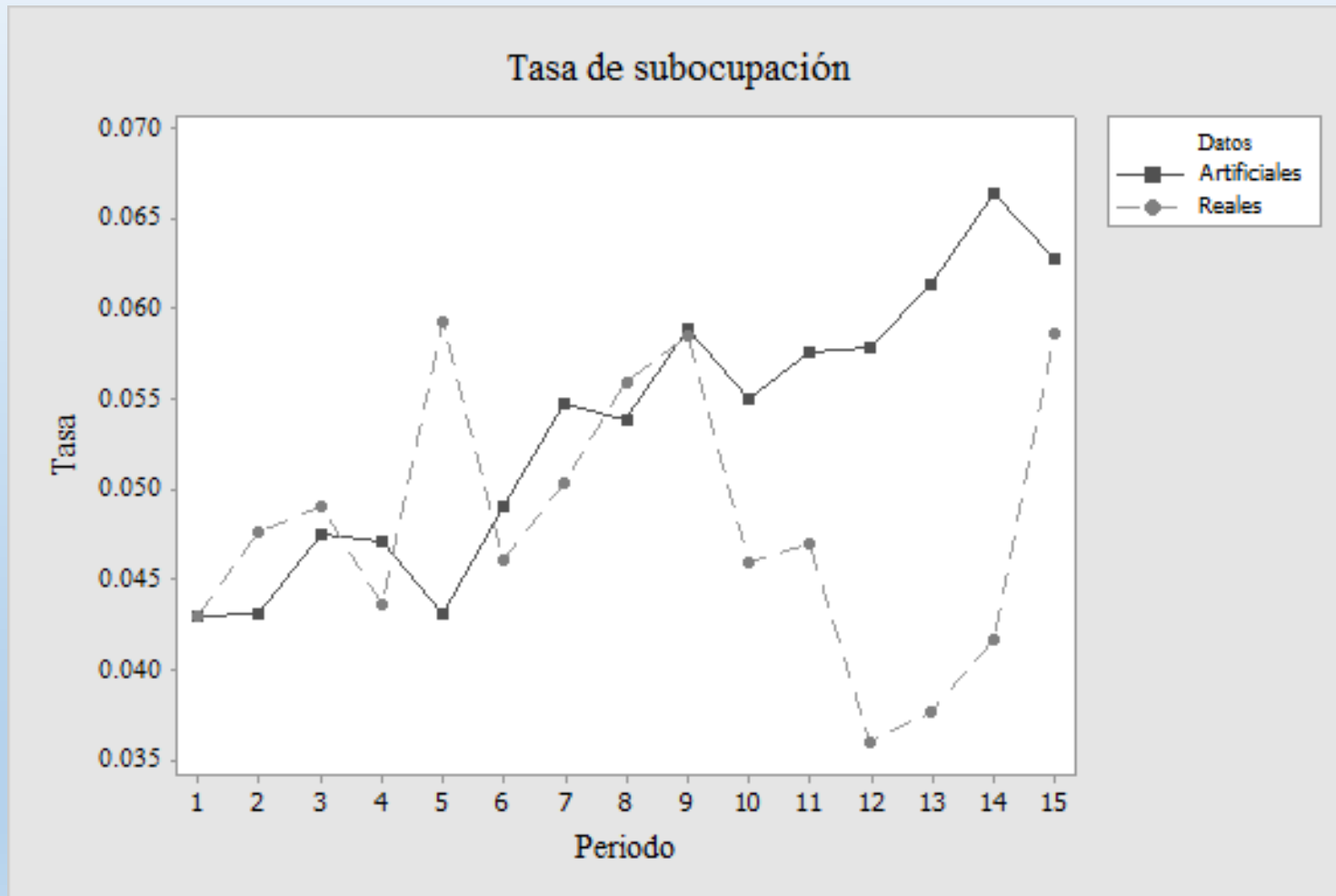
Comparación gráfica TD



Comparación gráfica TS



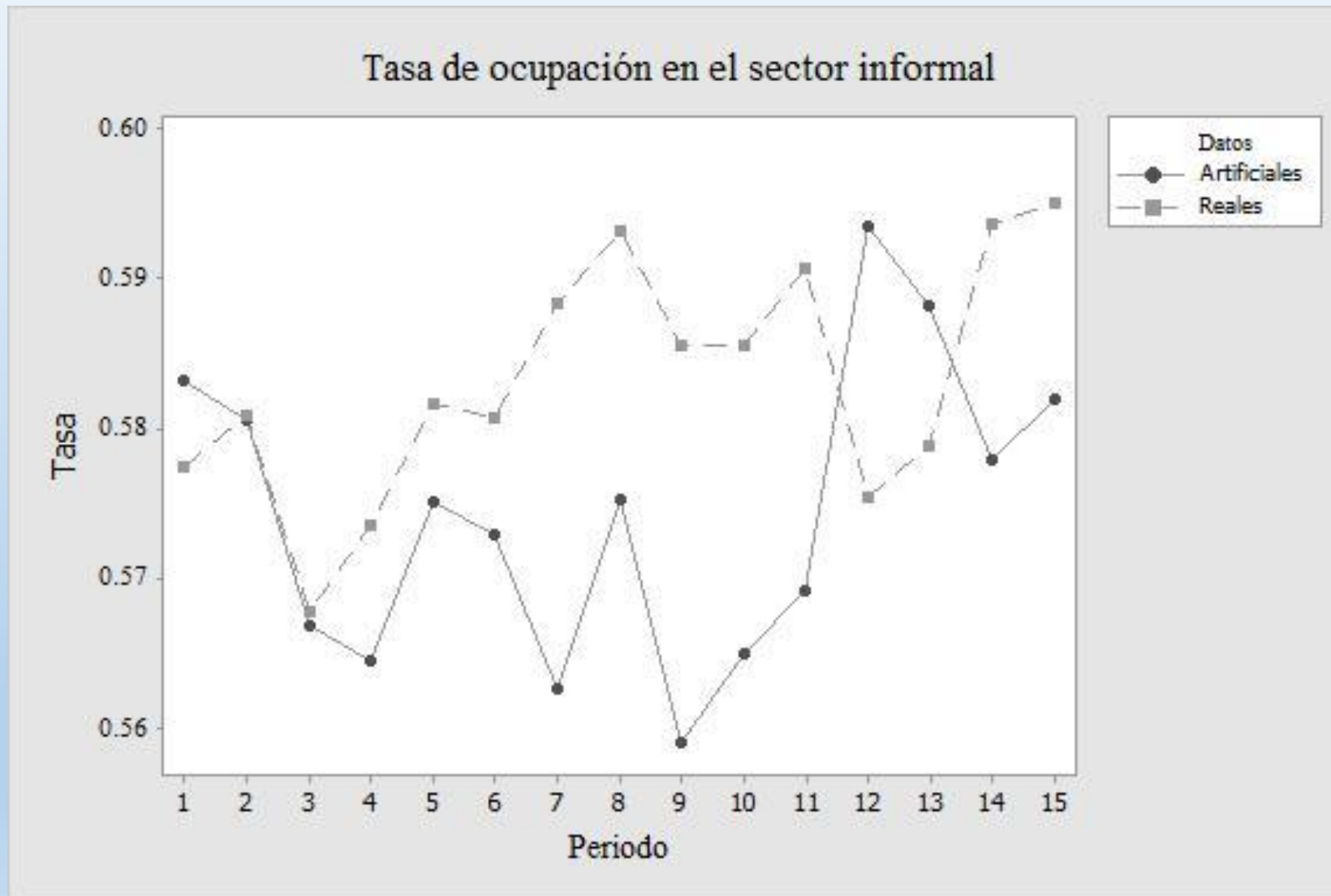
Universidad Veracruzana



Comparaciones gráfica TOSI



Universidad Veracruzana





Error Cuadrático Medio

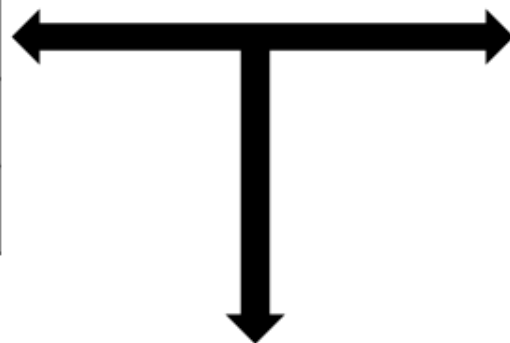
Tasas generadas de datos artificiales

Periodo	TD	TOSI	TS
1	0.047	0.583	0.048
2	0.051	0.559	0.042
N	0.045	0.531	0.038

$$EMC = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

Tasas generadas de la ENOE

Periodo	TD	TOSI	TS
1	0.047	0.577	0.043
2	0.051	0.586	0.066
N	0.045	0.550	0.061

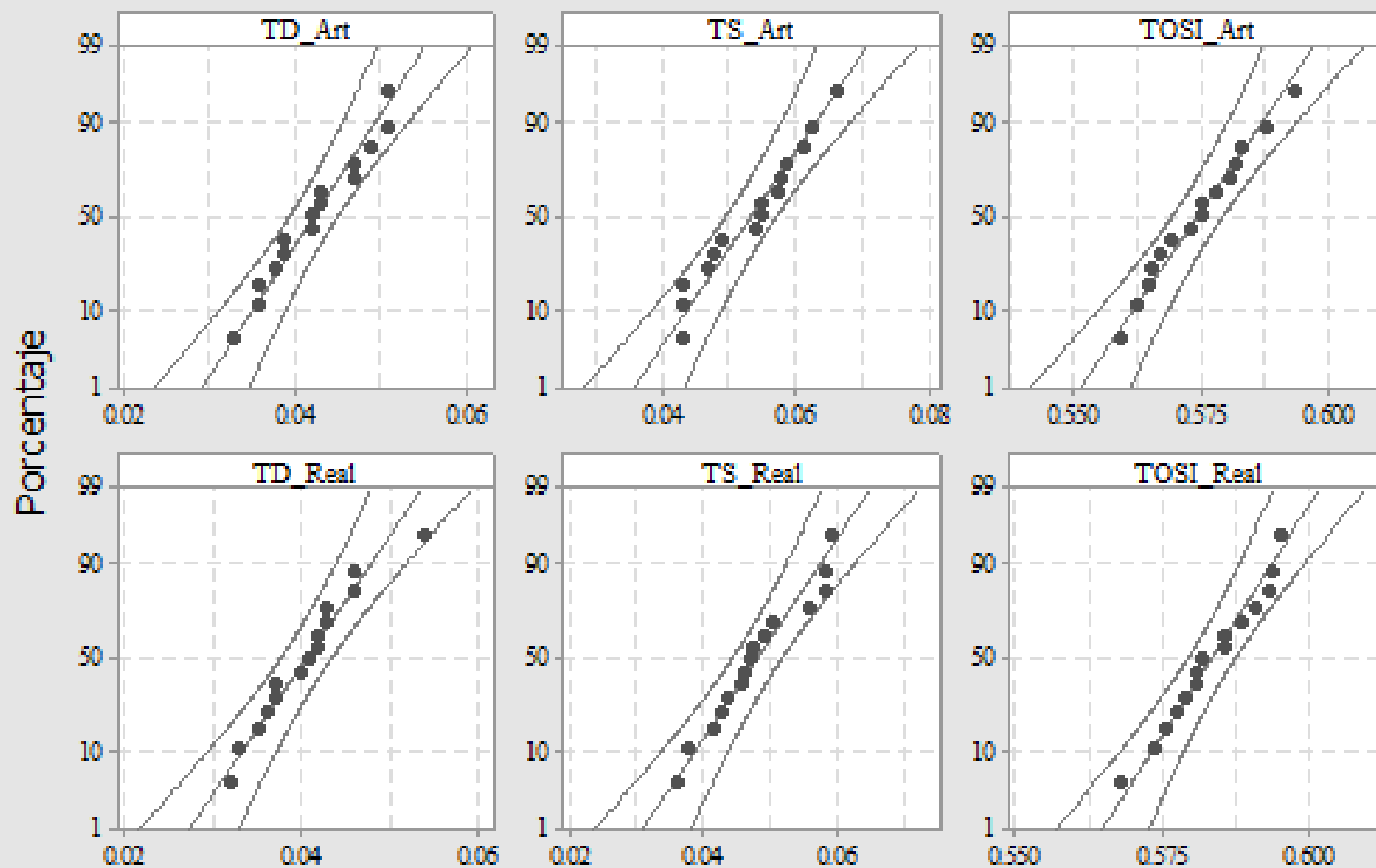


	TD	TS	TOSI
EMC	0.000014866	0.000147335	0.000243216



Universidad Veracruzana

Prueba de Normalidad Normal - 95% CI



Kolmogorov-Smirnov

Tasa	Origen	P-Value
TD	Artificial	0.60
	Real	0.62
TS	Artificial	0.48
	Real	0.42
TOSI	Artificial	0.91
	Real	0.87

Pruebas paramétricas de comparación

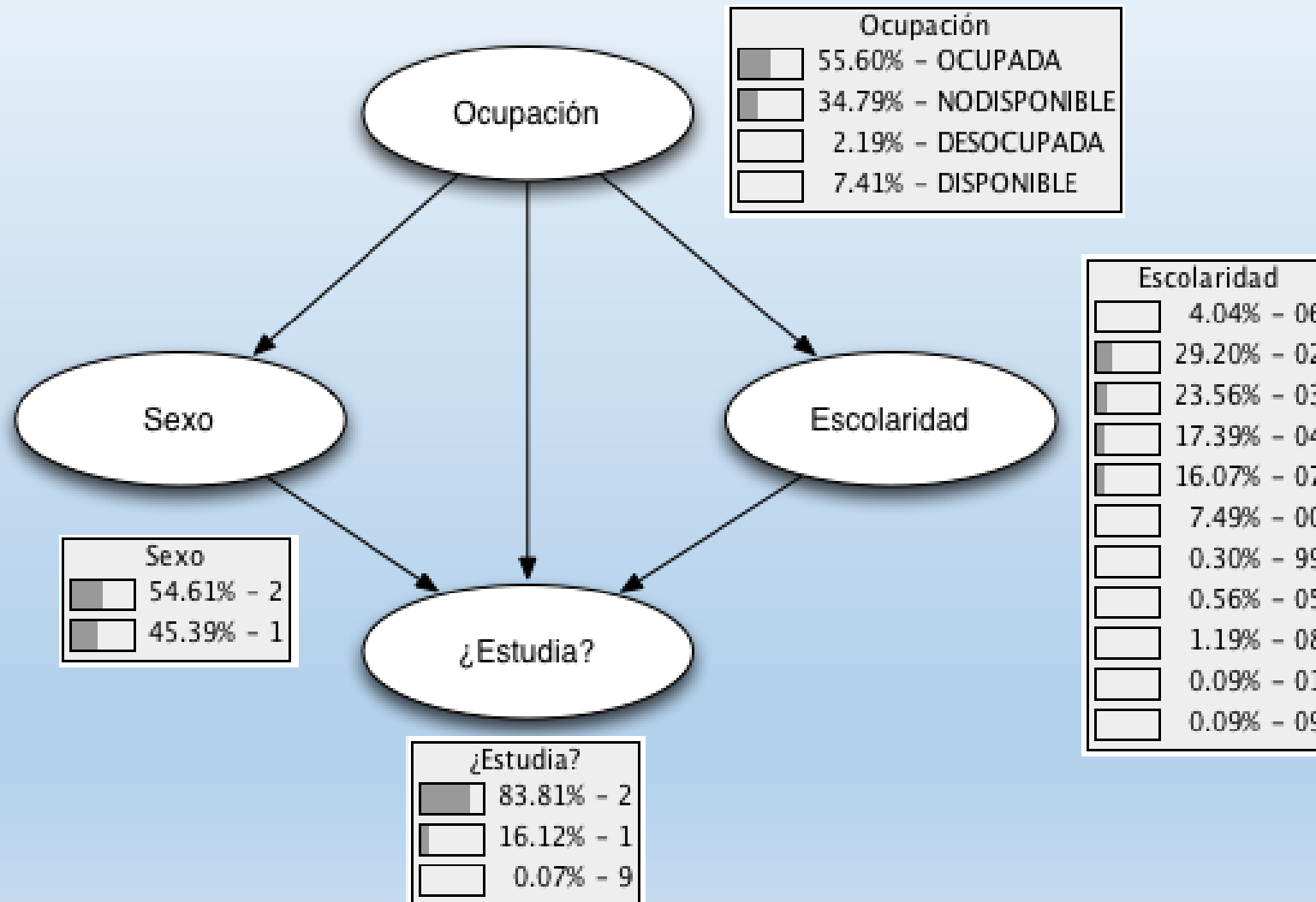
T-Test



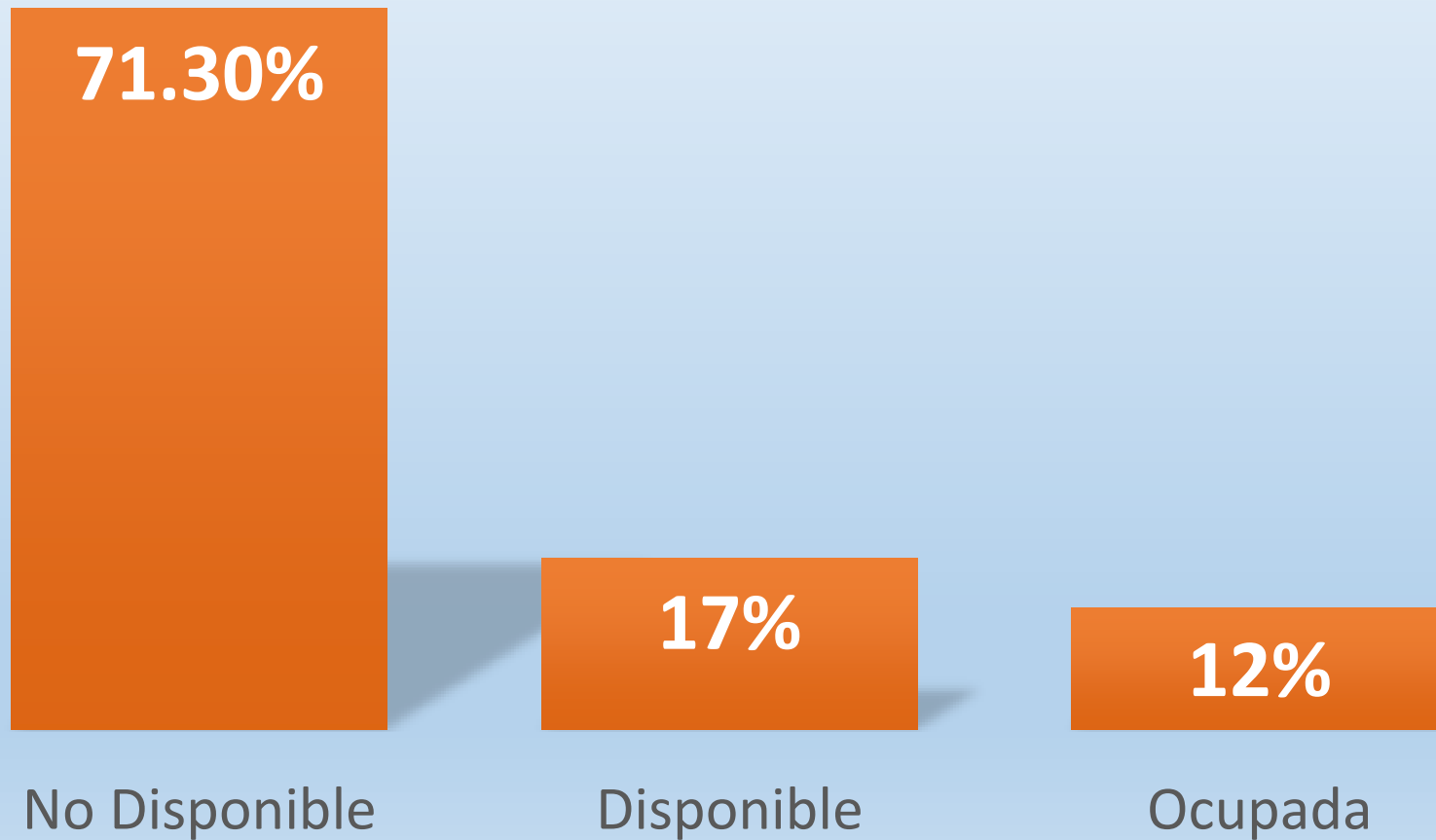
Universidad Veracruzana

Tasa	P-value
TD	0.717666
TOSI	0.717666
TS	0.380299

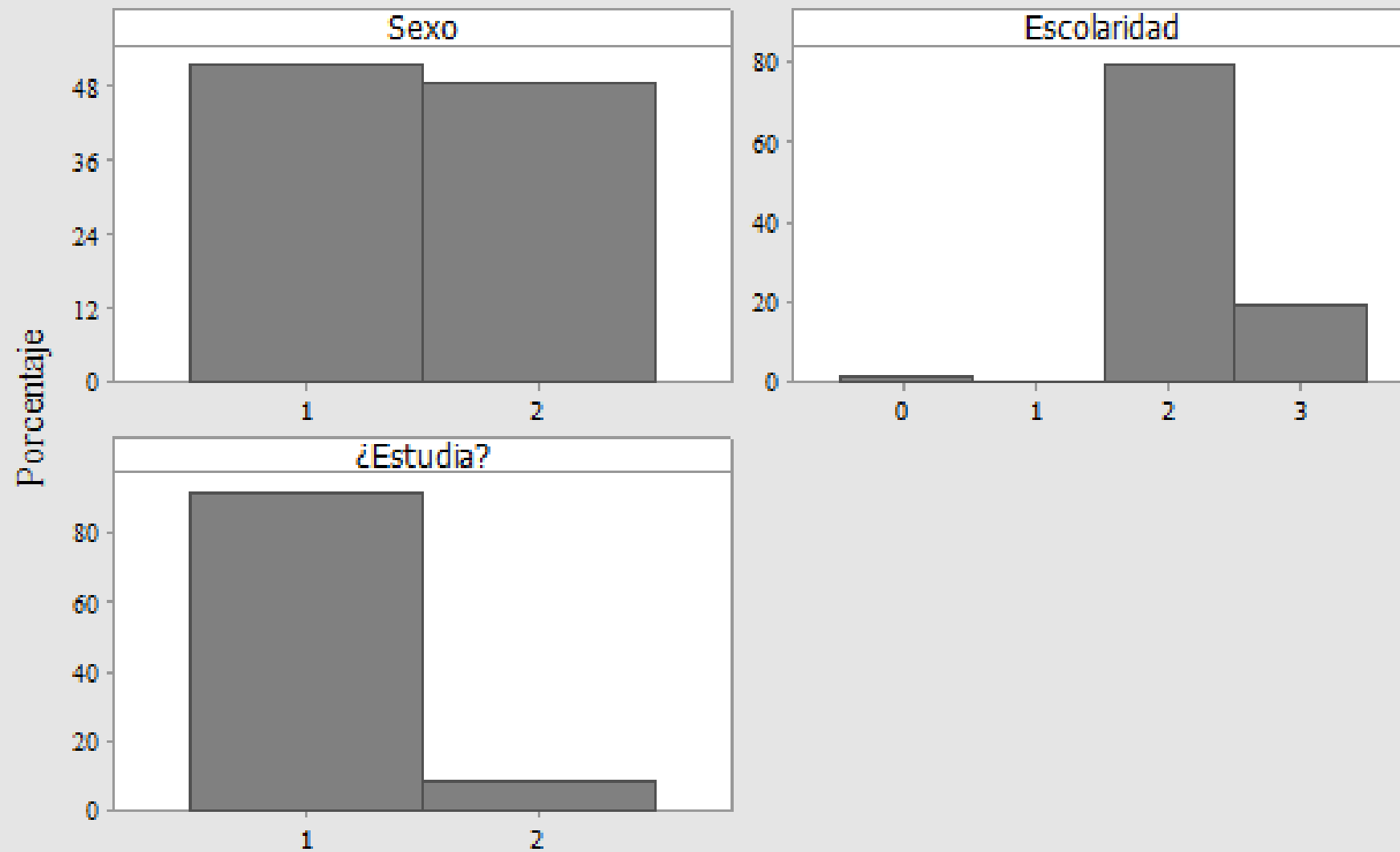
Análisis de personas que cumplen 15 años

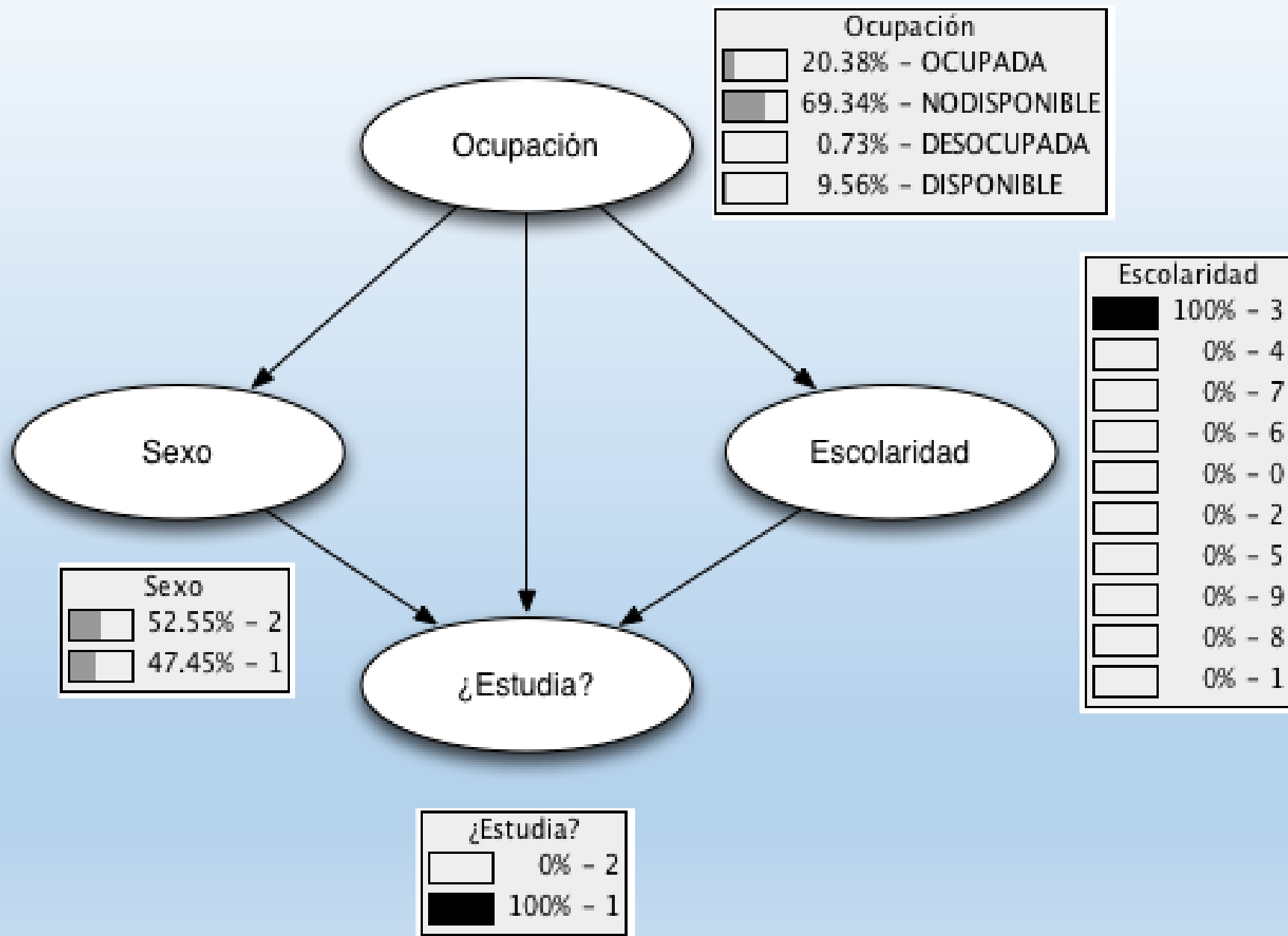


Ocupación

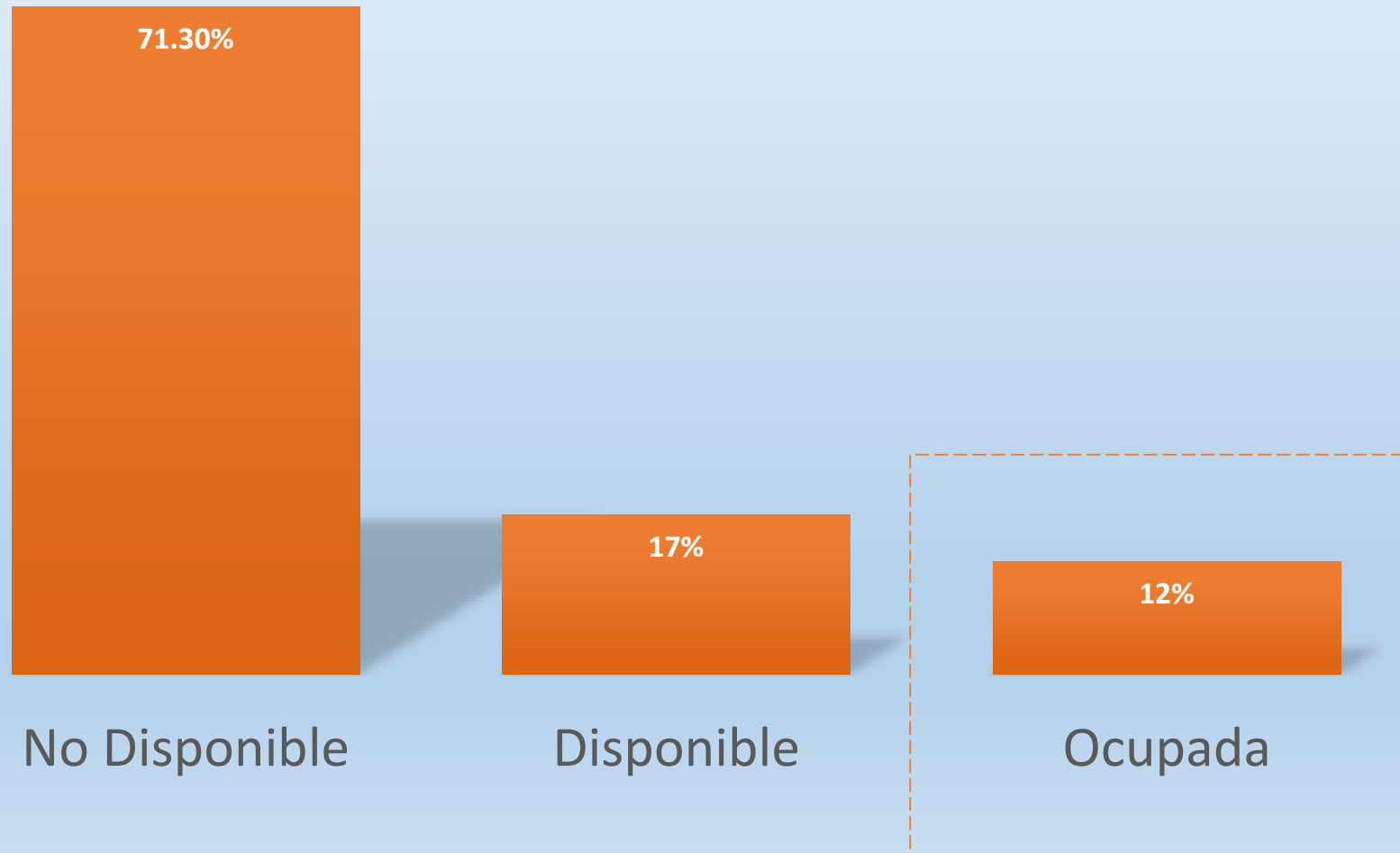


Distribución propiedades generales

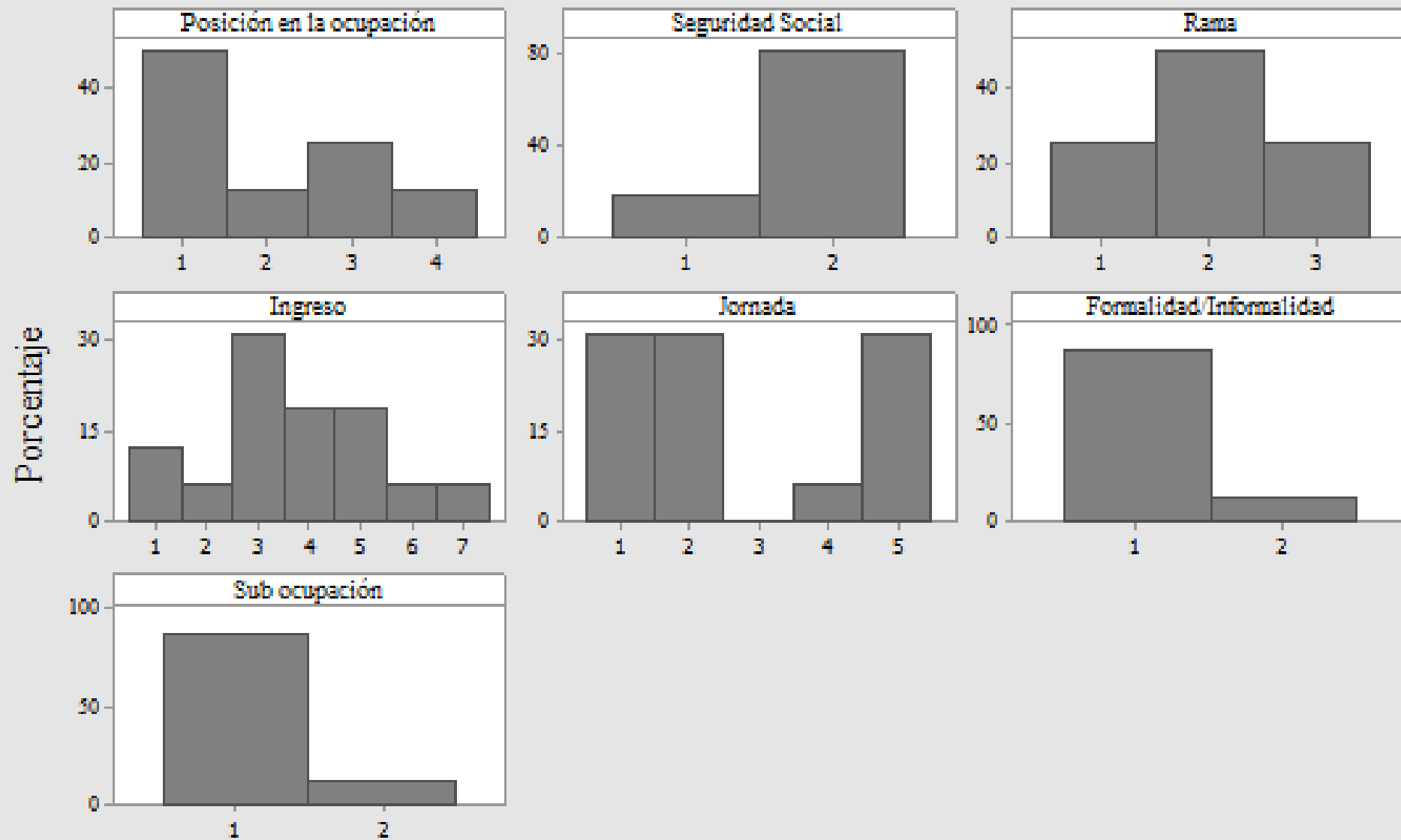




Ocupación



Distribución propiedades laborales



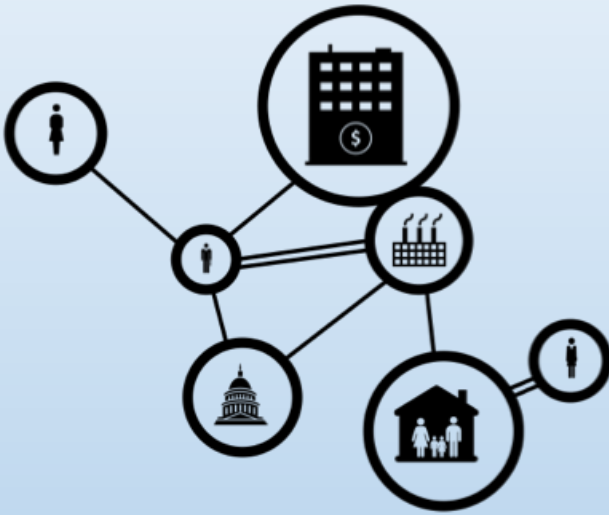
Análisis del fenómeno a través
del modelo

Modelo

Mundo real



Modelo mundo real

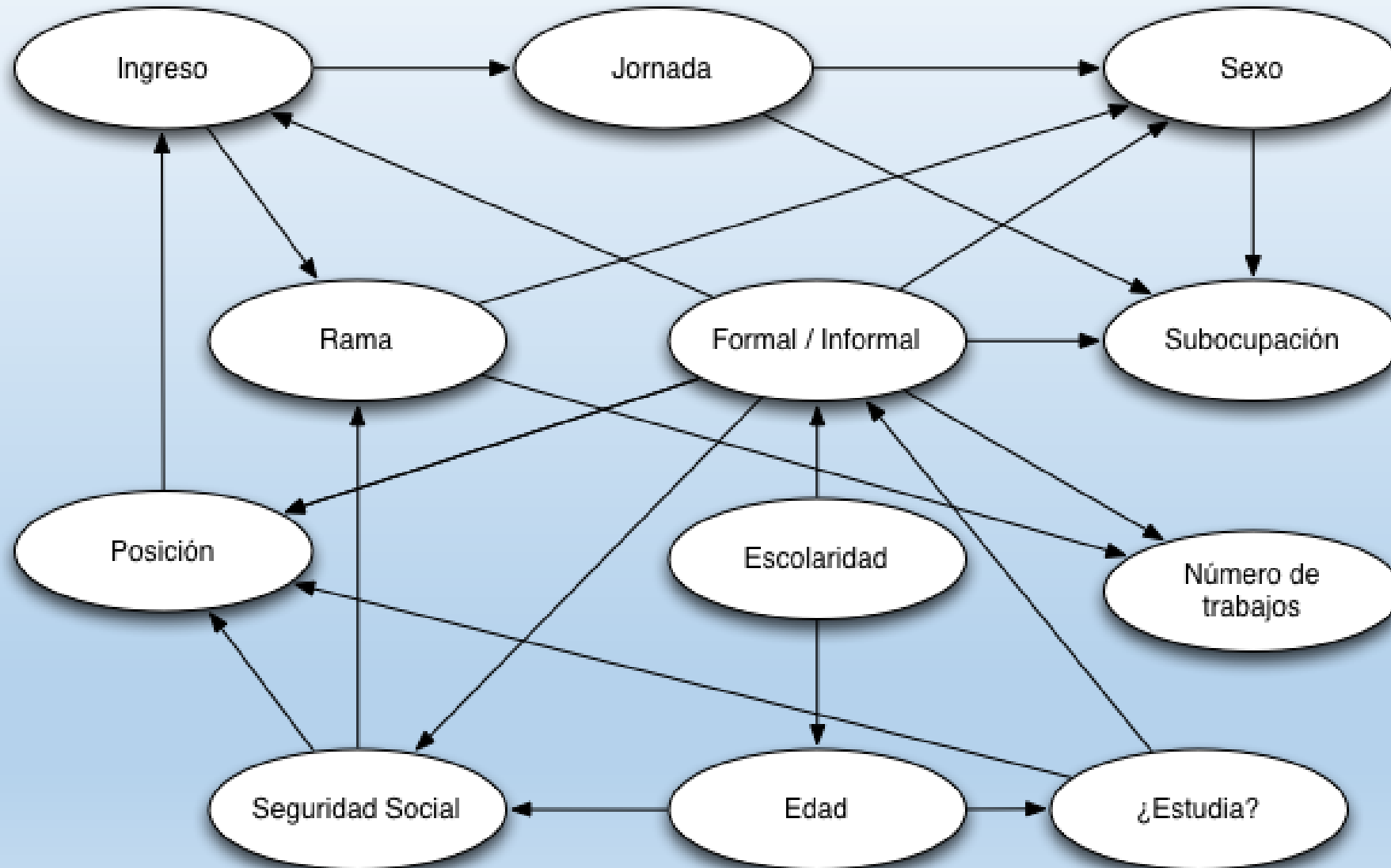


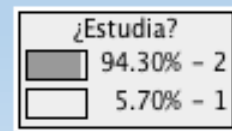
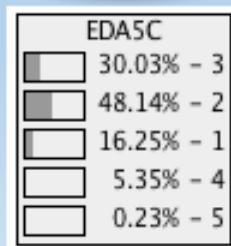
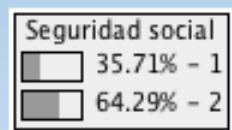
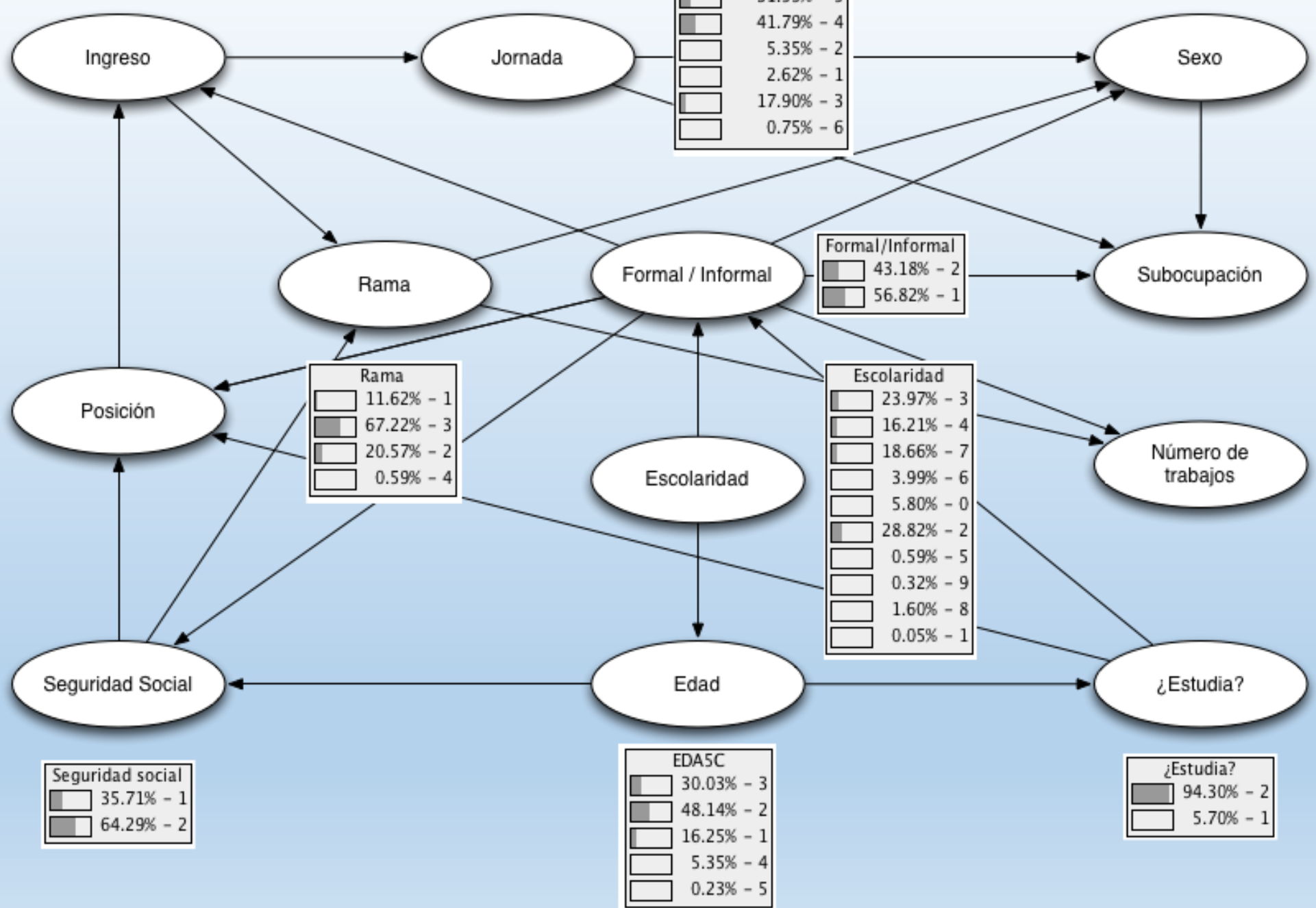
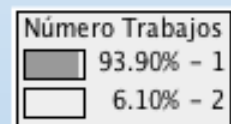
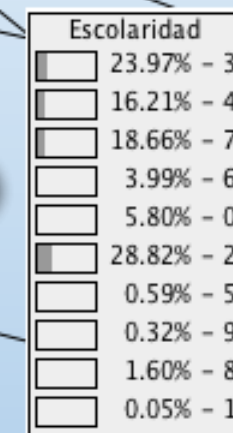
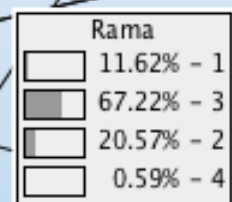
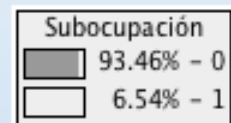
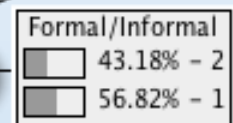
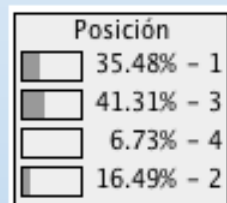
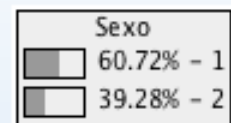
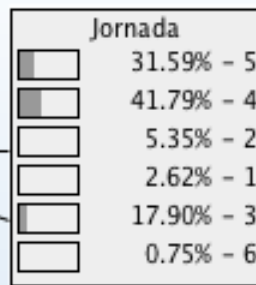
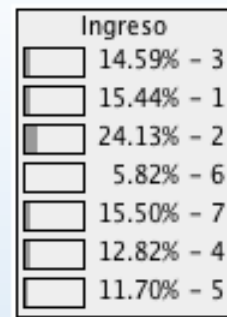
Observación

Abstracción



Análisis del fenómeno a través del modelo







Universidad Veracruzana

Conclusiones

- Datos artificiales generados en una simulación social basada en redes bayesianas muestran una gran similitud con datos reales
- Validación estadística
- Análisis del fenómeno
- Implementación con agentes y artefactos (Jason-CArtAgO)



Trabajo futuro

- Implementar computación evolutiva para optimizar los resultados de la simulación probando la generación del modelo con diferentes parámetros.
- Representar hogares a través de artefactos en donde los agentes conozcan las propiedades de los miembros de su familia
- Implementar por medio artefactos, una abstracción de políticas públicas
- Implementar estrategias colaborativas de los agentes
- Generalizar el sistema para que el caso de estudio sea determinado por la base de datos



Universidad Veracruzana

Muchas gracias