

Sistemas de información y base de datos

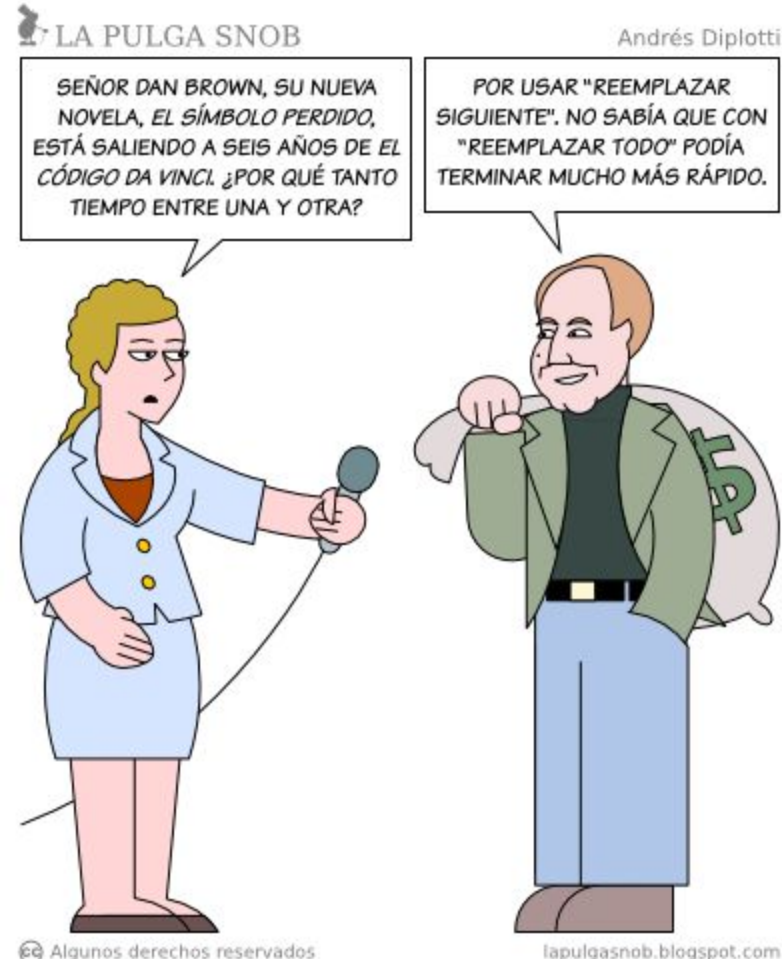
CLASE 4

UNTREF
2018

María Teresa Arriola mteresaarriola@gmail.com

Emilio Platzer emilioplatzer@gmail.com

Estefanía Capece eycapece@gmail.com

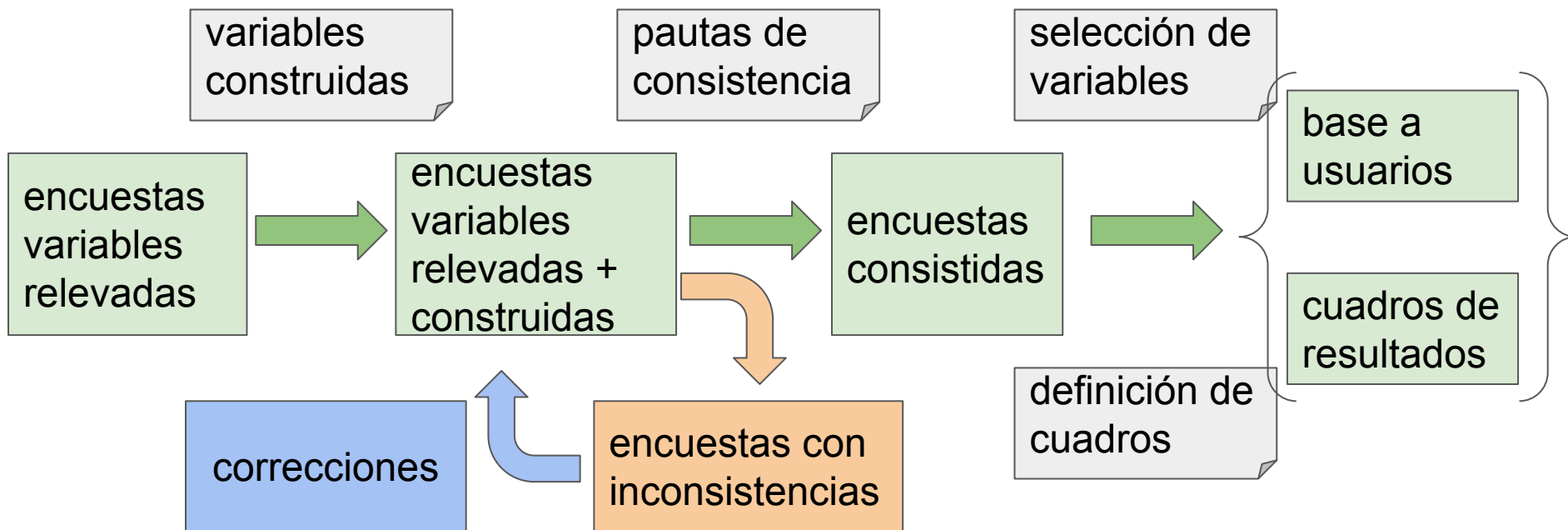


Preguntas de detección

	<i>Identificación</i>
MD 1 ¿Tiene Ud. alguna dificultad de largo plazo para caminar, subir o bajar escalones?	Sí.....1 No.....2
MD 2 ¿Tiene Ud. alguna dificultad de largo plazo para mover uno o los dos brazos o las manos?	Sí.....1 No.....2
MD 3 ¿Tiene Ud. alguna dificultad de largo plazo para agarrar objetos y/o sostener peso con una o las dos manos?	Sí.....1 No.....2
MD 4 ¿Tiene Ud. alguna dificultad de largo plazo para levantarse, acostarse, mantenerse de pie o sentado?	Sí.....1 No.....2
MD 5 ¿Tiene Ud. alguna dificultad de largo plazo para ver, aún con anteojos o lentes puestos?	Si, es ciego.....1 Si, tiene disminución visual....2 No.....3
MD 6 ¿Tiene Ud. alguna dificultad de largo plazo para oír aún usando audífonos o aparatos?	Si, es sordo1 Si, tiene disminución auditiva .2 No.....3

EAH 2011: Preguntas en el cuestionario individual que definirán operativamente la apertura del formulario de discapacidad.

Etapas del procesamiento de las encuestas



Cómo generar los tabulados de las encuestas

Cada tabulado necesita un SQL para obtener los datos

```
SELECT comuna, tipo_viv, SUM(fexp)
      SUM(fexp)*100.0/(SELECT SUM(fexp) FROM hogares),
FROM hogares
WHERE nhogar=1
GROUP BY comuna, tipo_viv
ORDER BY comuna, tipo_viv
```

The diagram illustrates the components of the SQL query. A red box labeled 'Variables de corte' points to the 'SUM(fexp)' and 'SUM(fexp)*100.0/(SELECT SUM(fexp) FROM hogares)' expressions. A purple box labeled 'Filtro' points to the 'nhogar=1' condition in the WHERE clause. Red boxes also highlight the 'comuna, tipo_viv' columns in the SELECT, GROUP BY, and ORDER BY clauses.

Muchos de esos SQL van a ser muy parecidos, solo cambian las variables de corte y el filtro

Metadatos para generar los tabulados de las encuestas

```
SELECT ${var_corte}, SUM(fexp)
      SUM(fexp)*100.0/(SELECT SUM(fexp) FROM ${u_a}),
FROM ${u_a}
WHERE ${filtro}
GROUP BY ${var_corte}
ORDER BY ${var_corte}
```

Se definen plantillas SQL paramétricas y una tabla de metadatos define la composición

tabulado	var_corte	filtro	u_a	plantilla
cuadro 1	comuna, tipo_viv	nhogar=1	hogares	t_2
cuadro 2	zona, sexo	edad>=10	personas	t_2

Consistencias - Origen

Origen de las inconsistencias:

- error de tipeo
- error de llenado u omisión de marca
- error de interpretación de la pregunta
- respuesta inconsistente del entrevistado:
 - voluntaria
 - involuntaria
- error del programa de ingreso o dispositivo móvil

Consistencias - Tipos

auditoría		
validez/rango	saltos	completitud
<ul style="list-style-type: none">• opciones válidas• máximos y mínimos• tipos de datos <div>pueden definirse automáticamente</div> <div>si el sistema está basado en metadatos</div>	<ul style="list-style-type: none">• las preguntas salteadas no deben tener respuestas• las preguntas filtradas, tampoco	<ul style="list-style-type: none">• debe haber tantos formularios de personas como personas declaradas• debe haber formulario de mascota si el hogar declaró que tiene

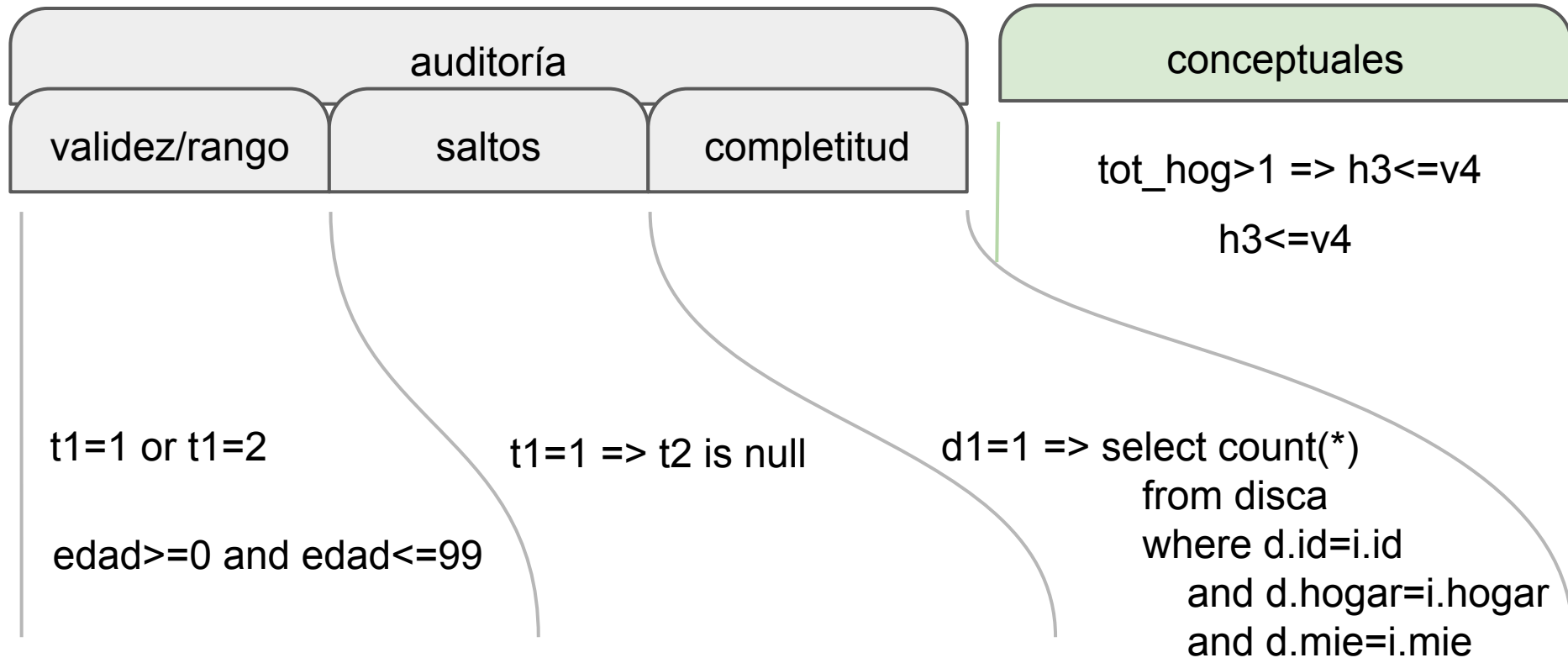
conceptuales

- dadas dos o más respuestas individualmente válidas, su combinación podría no serlo
- las variables a controlar podrían provenir de distintos formularios

deben definirse
manualmente

**no pueden deducirse
automáticamente**

Consistencias - Tipos



Consistencias en dispositivo móvil - Momentos de detección

encuesta	cierre	recepción	procesamiento
<ul style="list-style-type: none">• en cuanto se marca una respuesta inconsistente• deben ser pocas para no quitar fluidez a la encuesta	<ul style="list-style-type: none">• al finalizar la encuesta• pueden ser más• evita que se olviden cuestionarios o personas	<ul style="list-style-type: none">• cuando el encuestador entrega su trabajo• a veces no existe el rol• sirve para recapacitar• podría volver a campo	<ul style="list-style-type: none">• se pueden verificar un conjunto muy complejo y completo de consistencias• se pueden consistir datos agregados• la retroalimentación a campo es lenta y complicada
el encuestado está presente, se puede repreguntar		el encuestador está presente, se puede apelar a su memoria	

Consistencias en sistema clásico - Momentos de detección

recepción

- cuando el encuestador entrega su trabajo
- **las consistencias deben hacerse manualmente**
- **son muy pocas y el proceso es lento**

el encuestador está presente,
se puede apelar a su memoria

ingreso

- se pueden controlar un conjunto importante de consistencias
- el ingresador puede parar y consultar
- puede capacitarse al ingresador para que haga tareas de análisis

procesamiento

- se pueden verificar un conjunto muy complejo y completo de consistencias
- se pueden consistir datos agregados

Consistencias con sistema de ingreso previo - M. de detección

ingreso

- **se ingresan las encuestas en cuanto el encuestador llega y se van liberando a medida que se termina el ingreso**
- **se corren solo las consistencias de valores válidos y rangos**
- **se necesitan más ingresadores**
- **el encuestador debe esperar para ser recepcionado**

recepción

- **a medida que se va ingresando, con el encuestador presente**
- **se pueden hacer las mismas consistencias que tendría el dispositivo móvil**

**el encuestador está presente,
se puede apelar a su memoria**

procesamiento

- **se pueden verificar un conjunto muy complejo y completo de consistencias**
- **se pueden consistir datos agregados**

Metadatos - consistencias

u_a	precondición	postcondición	gravedad	A	descripción
hogar		$h3 \leq v4$	error	s	ambientes exclusivos de
hogar	$tot_hog > 1$	$h3 < v4$	error	s	no todos los ambientes d
persona	$e6 \in (12,13,14)$	$p3b > 17$	advertencia	s	si es terciario, universitari
persona		$i7_2 \leq i7_i$	error	s	el importe parcial no pue
persona		$t53_bis2 \leq 24$	error	s	el día tiene menos de 24h
persona		$t53_bis2 \leq 12$	advertencia	s	la cantidad de horas sem

Metadatos - construcción de variables

variable	condición	valor	etiqueta
cond_ocup	t1=1 and t7=1	1	trabajo con remuneración
cond_ocup	t1=1 and t7=2 and t8 in (1,2)	2	trabajador familiar o ad honorem
cond_ocup	t1=2 and t2=1 and t7=1	3	changa con remuneración
cond_ocup	t1=2 and t2=1 and t7=2 and t8 in (1,2)	4	changa para familiar o ad honorem
cond_ocup	t1=2 and t2=2 and t3=5 and t4 in (1,2,3)	5	con licencia laboral o vacaciones
cond_ocup	t1=2 and t2=2 and t3=5 and t4=4 and t5=1	6	suspensión con pago

Metadatos - definición de tabulados

Cuadro 5 Distribución porcentual de los jefes de hogar por condición de actividad según comuna. Ciudad de Buenos Aires. Año 2014

Comuna	Condición de actividad			
	Total	Ocupado	Desocupado	Inactivo
Total	100,0	69,7	2,7 ^a	27,6
1	100,0	71,5	3,5 ^b	25,0
2	100,0	67,8	3,0 ^b	29,2
3	100,0	73,6	2,3 ^b	24,0 ^a
4	100,0	67,5	3,3 ^b	29,2
5	100,0	74,5	1,1 ^b	24,4
6	100,0	69,3	0,9 ^b	29,8

cuadro	fila	columna	tipo	filtro	título
5	comuna	cond_activ	persona	p4=1	Distribución porcentual de los jefes de

Datos abiertos

Datos públicos

- son los datos sobre los que el público en general tiene derecho a conocer

Datos abiertos

- es la forma en que se dan a conocer esos datos públicos:
 - características tecnológicas de fácil accesibilidad (on line, formatos libres)
 - características legales (se puede copiar, usar, procesar; citando fuente)
 - procesos (permanente, oportuno)

Datos abiertos - principios básicos

- **Gratuito**, no tiene costo de adquisición
- **De libre uso**, puede utilizarse con cualquier fin (pueden haber condiciones de citar fuente, no engañar, etc)
- **Integral**, debe acompañarse con la información que permita la correcta interpretación de los datos (contexto y metadatos)
- **Oportuno**, están disponibles tan rápido como sea necesario para preservar el valor de los datos
- **No discriminatorio**, no hay restricción de acceso
- **Primario**, directos de la fuente original, con el máximo nivel de desagregación
- **Permanente**, se conservan las versiones previamente publicadas
- **Legibles por máquinas**, estructurados para permitir el procesamiento automatizado de los datos.

Datos abiertos - nivel de apertura

