

Elementos de diseño

Métodos de investigación cuantitativa

UAH-EMAPE

Enfoques Metodológicos para el Análisis de
Políticas Educativas

Septiembre 13
2024

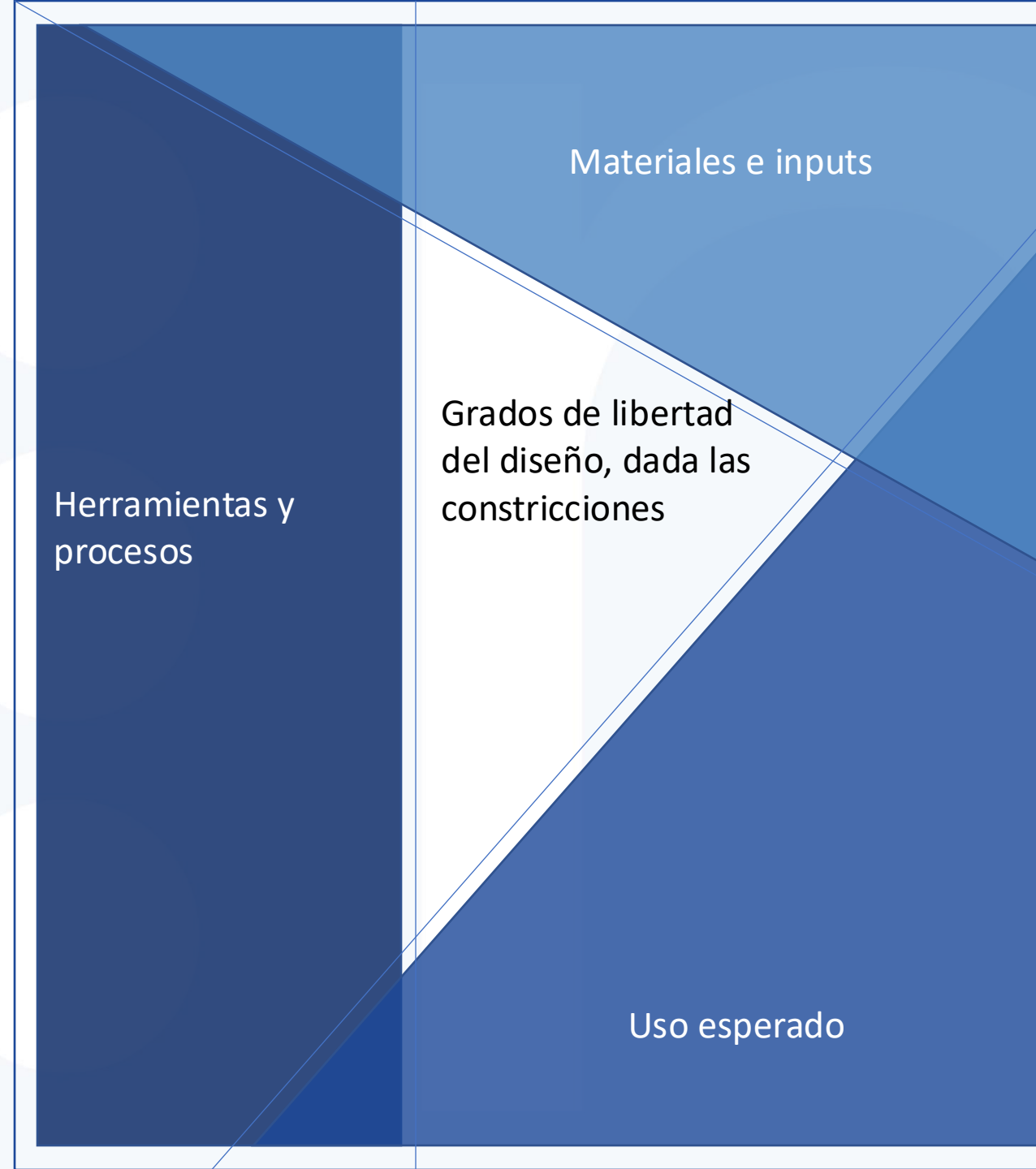
Profesor invitado

Carrasco, D., PhD,
Centro de Medición MIDE UC
Pontificia Universidad Católica de Chile

Diseño e investigación

La producción de resultados producto del ejercicio de una investigación, es raramente reconocida como un oficio (i.e. ***craftmanship***). Siendo el libro de Canales (2006) una notable excepción.

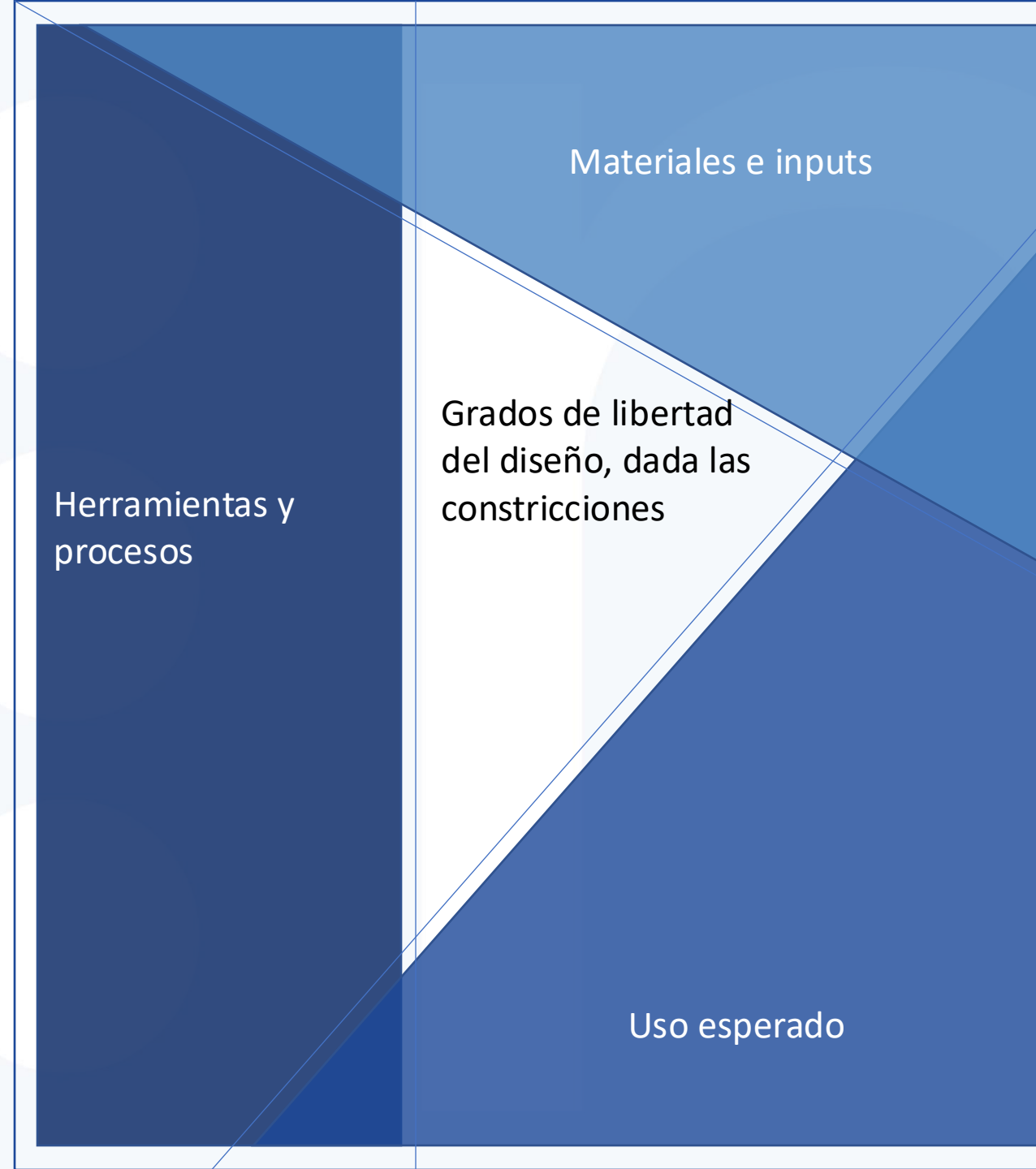
Esta omisión obstaculiza la reflexión que se requiere acerca de sus propias prácticas. Por ejemplo, la mayor parte del entrenamiento en métodos se explaya sobre los modelos estadístico, y menos en como estos modelos son implementados, y como son creados (Parker, 2017).



Constricciones de diseño

Para enmendar esta brecha, en las siguientes sesiones vamos a tener como trasfondo la noción de **constricciones de diseño** del diseño industrial (Broniowski, 1978; en Yau, 2015).

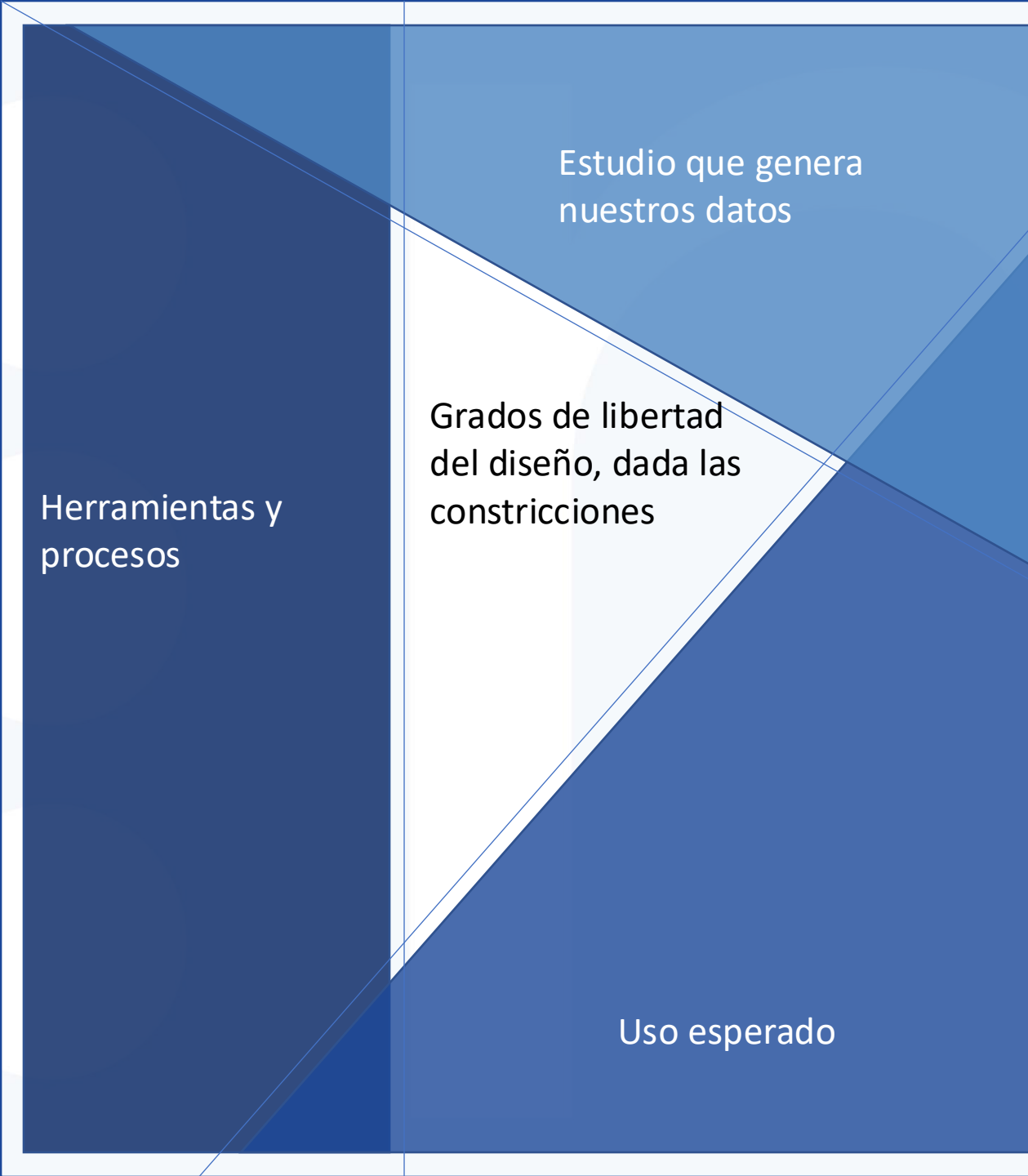
Estas constricciones pueden ser vistas como límites, pero también como **factores condicionantes**, como elementos fundamentales. En vez de verlos como limitaciones que impiden, los veremos como condiciones con las que partimos.



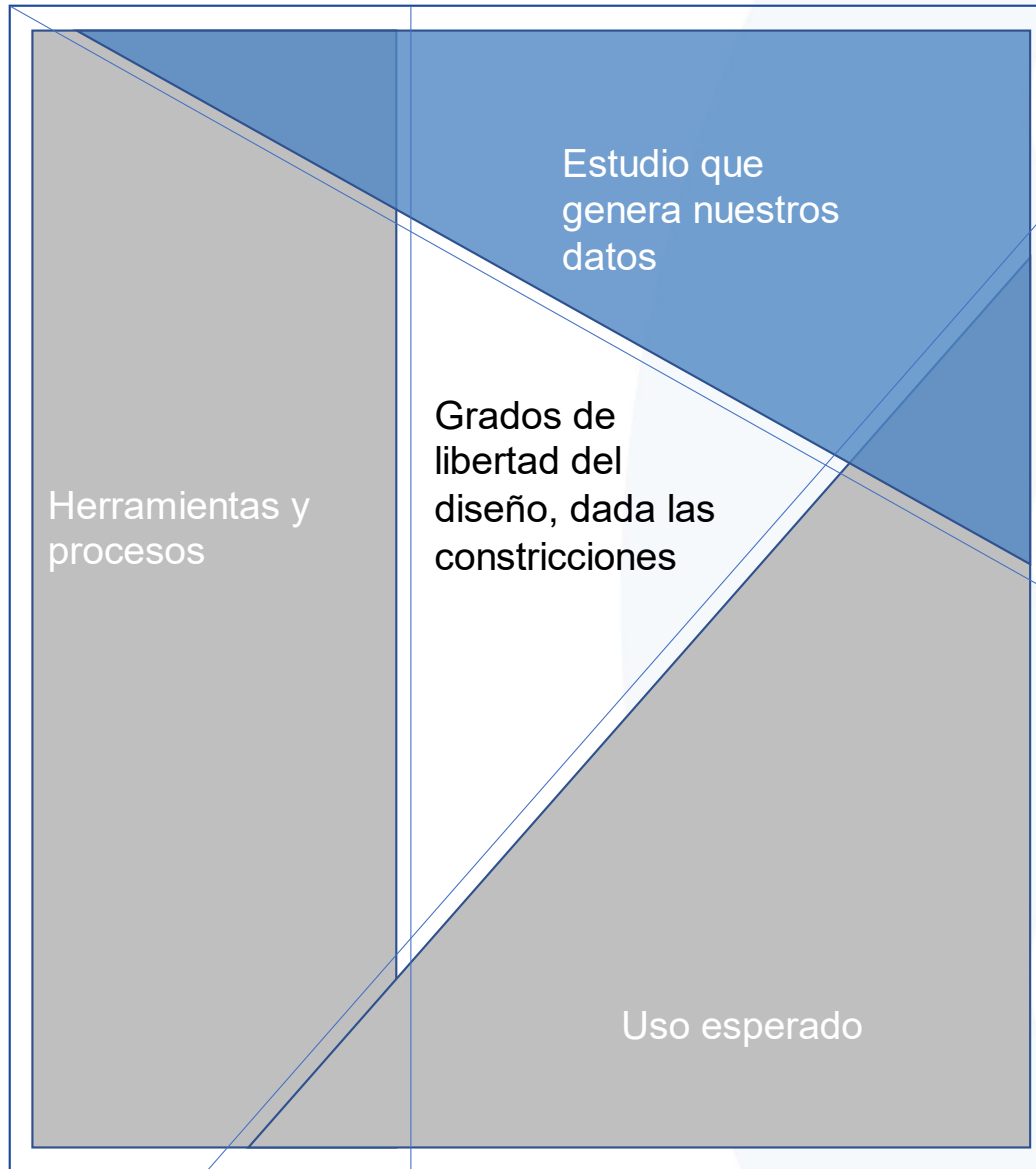
Constricciones en investigación

El proceso de investigación, de generar resultados que serán empleados en un argumento, incluye como elemento al estudio que produce los datos que serán analizados.

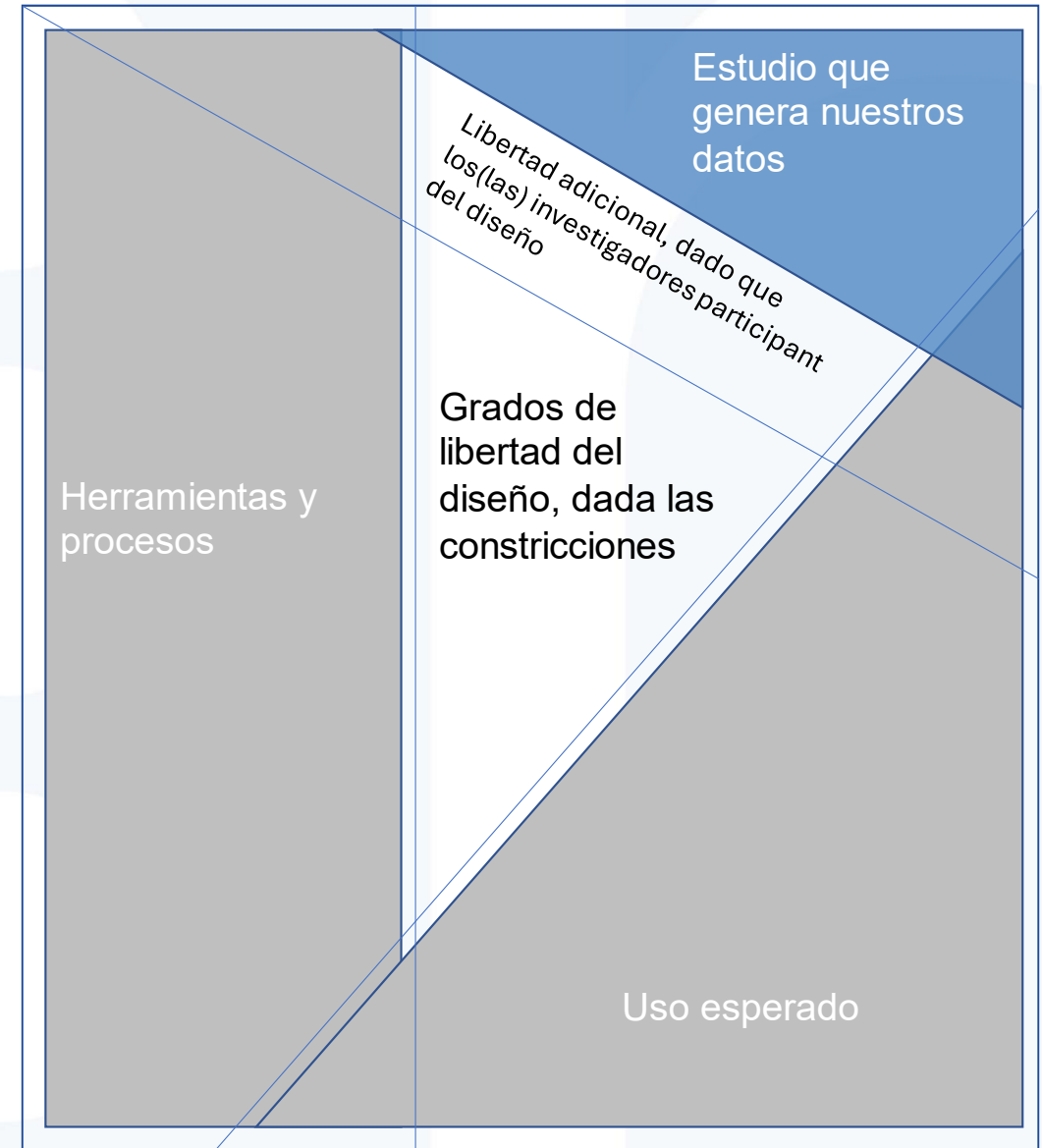
En escenarios de **datos secundarios**, el investigador comienza con un diseño pre establecido. En otros escenarios los y (las) investigadores pueden ampliar el rango de acción y diseñar al estudio que provee de datos y respuestas (**datos primarios**).



Datos Secundarios



Datos Primarios

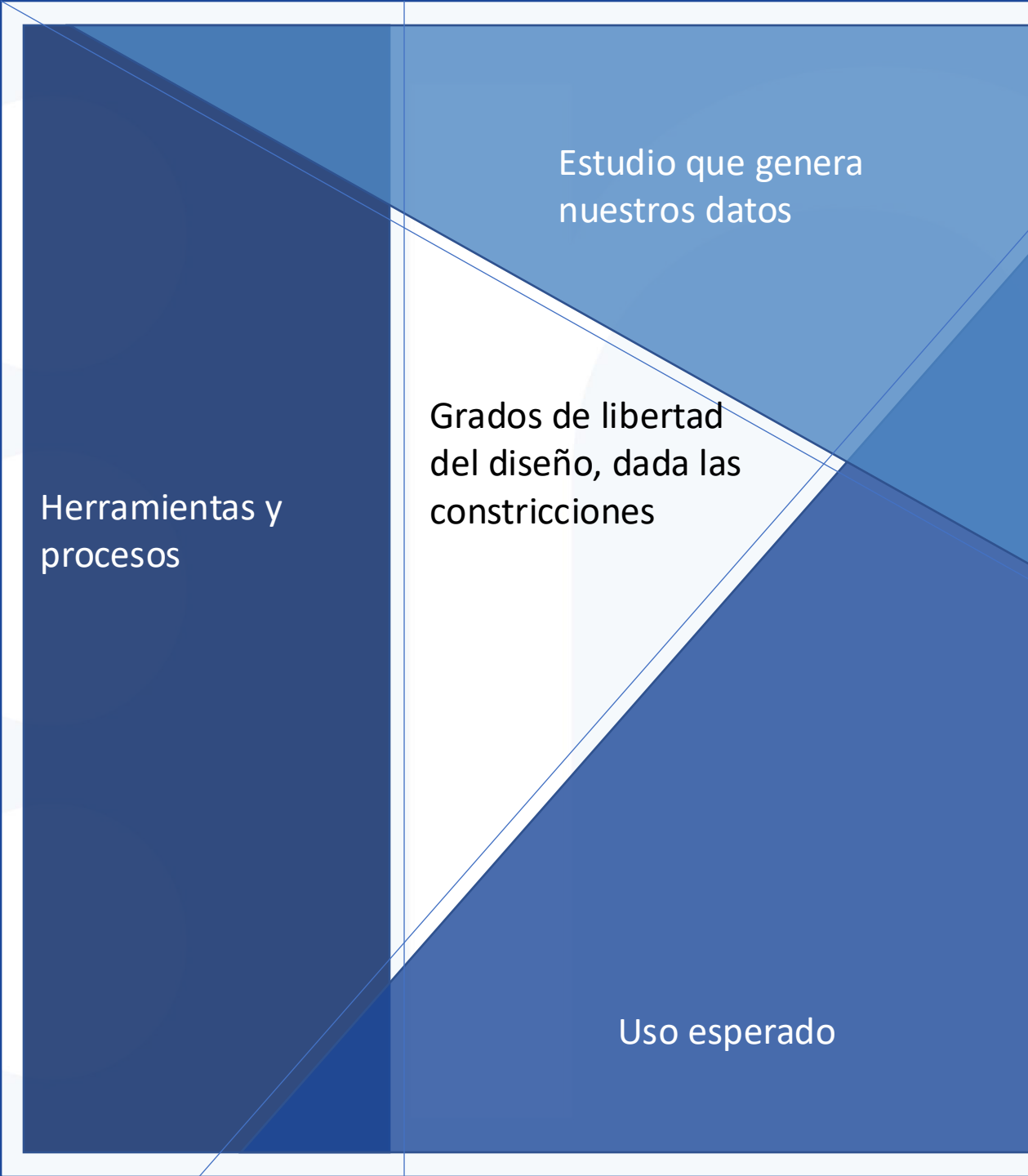


Métodos cuantitativos

Es común que el proceso de aprendizaje de métodos cuantitativos se sirva de datos secundarios. En este escenario, es muy importante que consideremos el diseño que producen los datos.

Las **características que posee el estudio** que genere los datos, son relevantes para:

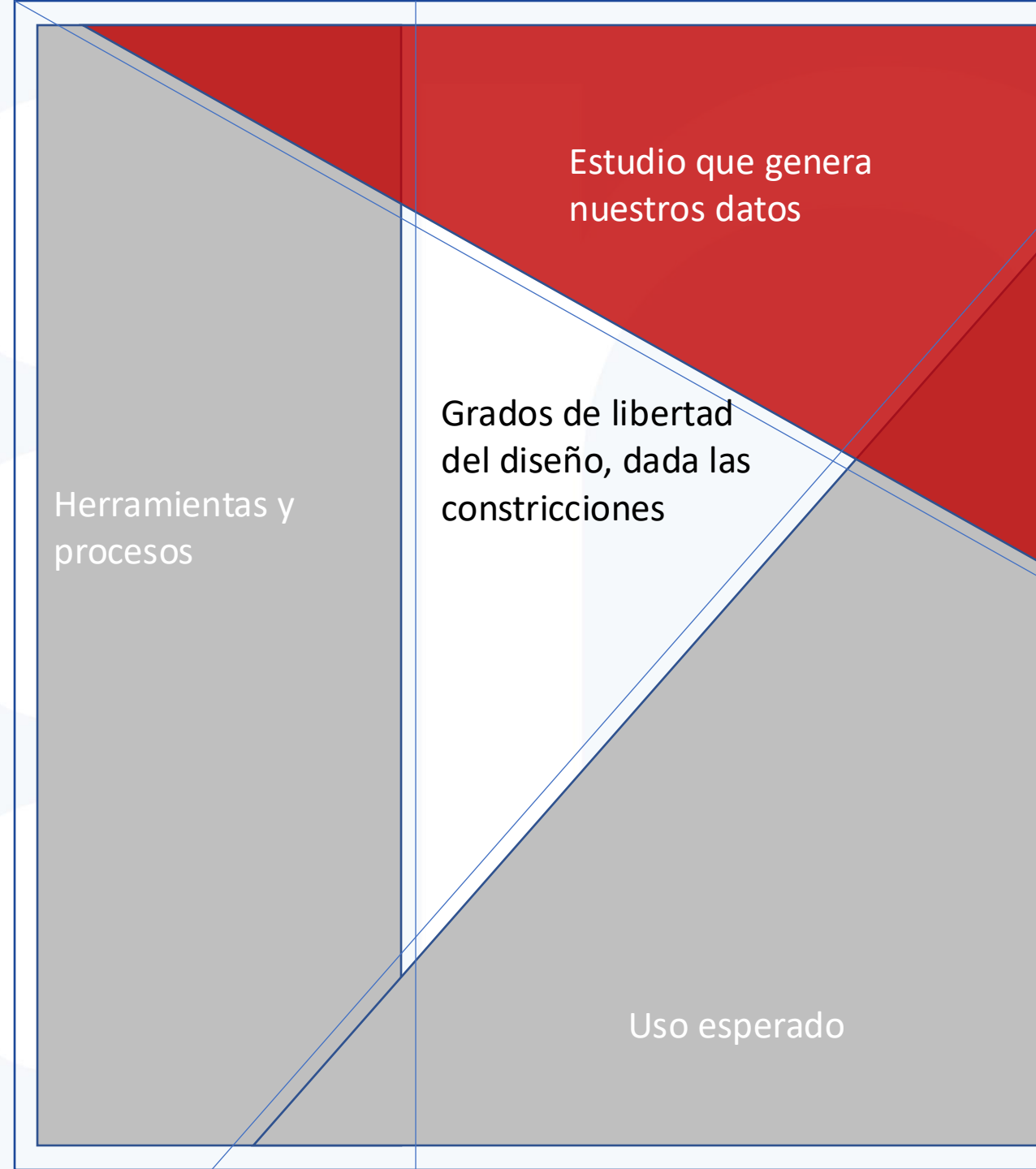
- a) Interpretar los resultados
 - comunicar el significado de los resultados.
- b) Para cualificar las inferencias
 - extensión de nuestras generalizaciones



Generalización

Que tanto podemos generalizar nuestros resultados si el estudio que nos provee de información...

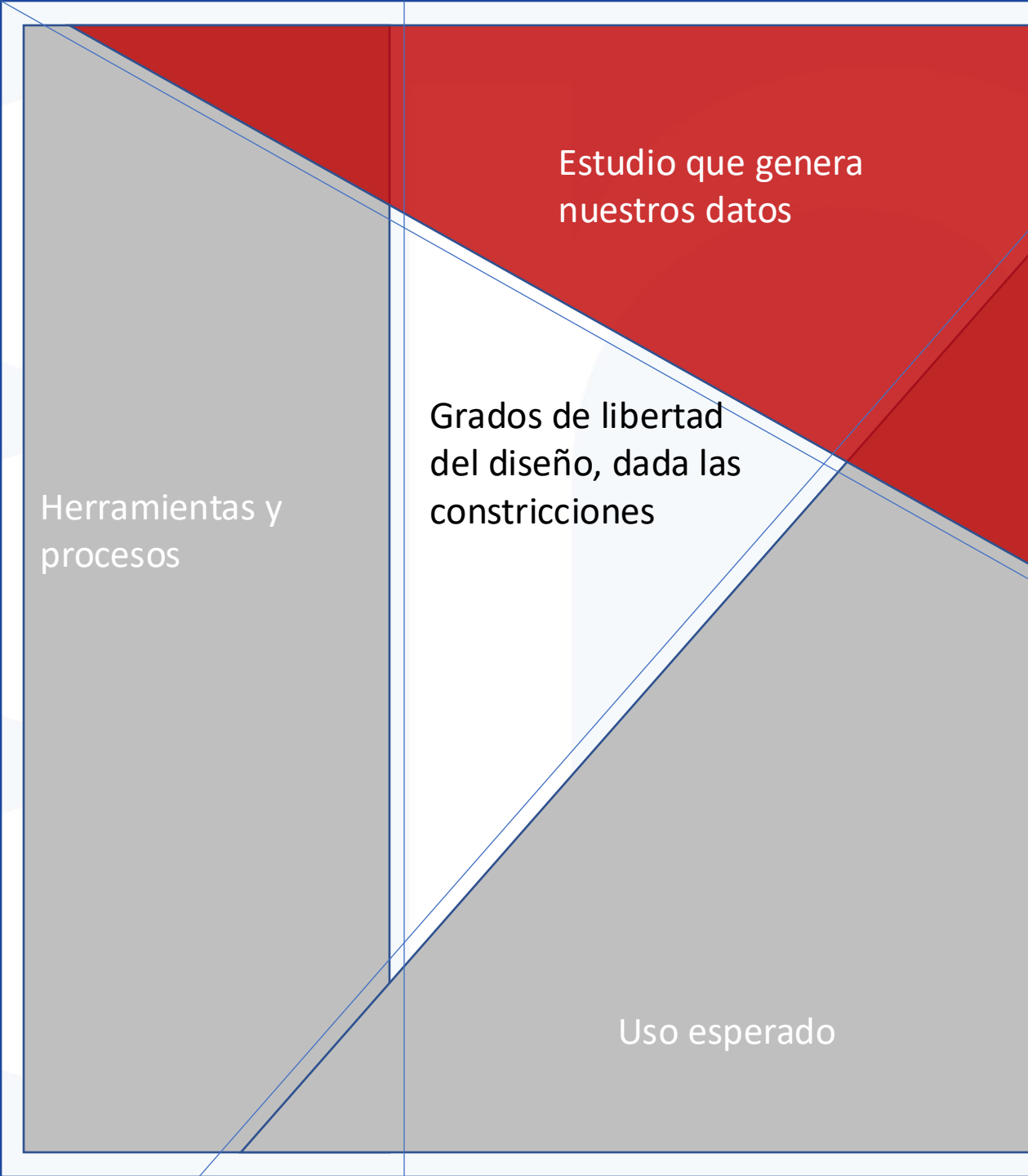
- a) Consiste en una muestra por **conveniencia** de 500 caso.
- b) Proviene de una muestra **probabilística** de 500 caso.
- c) El estudio que nos provee de información tiene características **censales**



Ejemplo: Rotación de profesores

Que tanto podemos generalizar resultados de tasa de rotación de profesores, si el estudio que provee de datos

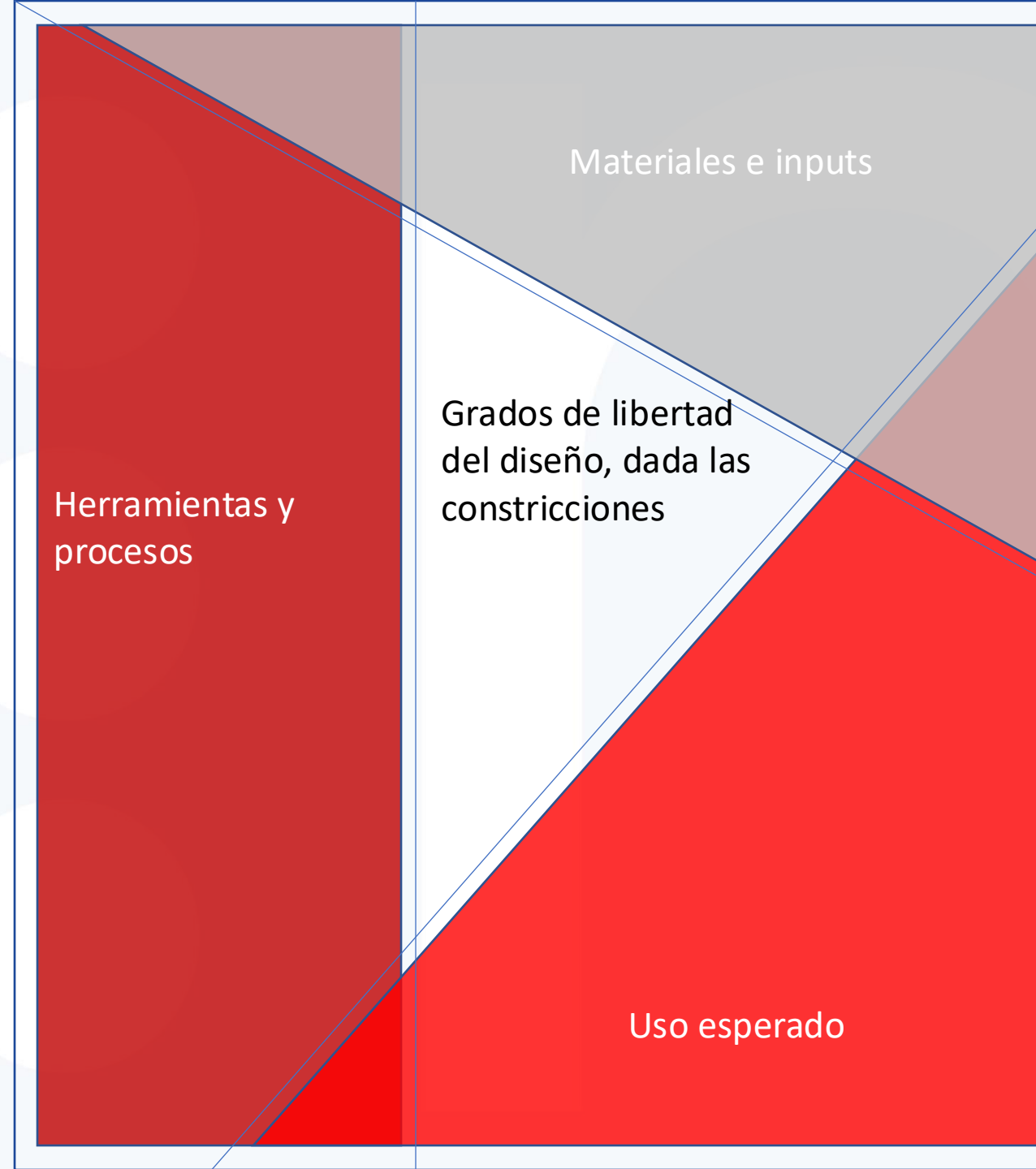
- a) Consiste en una muestra por **conveniencia** de 500 profesores, que trabajan con una plataforma web (e.g., ProfeJobs)
- b) Proviene de una muestra **probabilística** de 500 caso del SLEP Santiago Centro.
- c) El estudio que nos provee de información tiene características **censales**



Constricciones de diseño

Empleando la metáfora de las constricciones de diseño, podemos pensar otros escenarios, en que los investigadores empujan los límites convencionales:

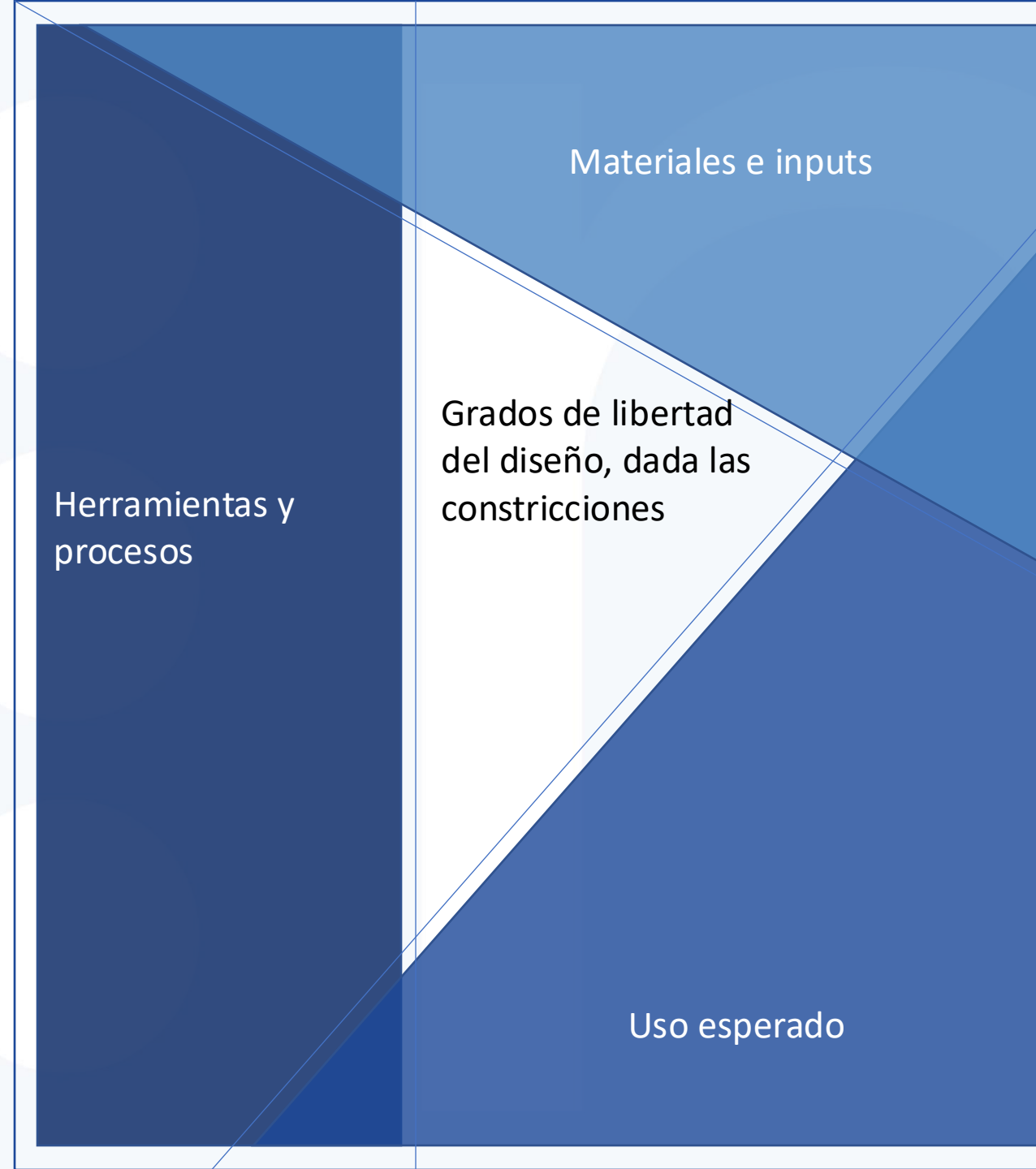
- a) Los investigadores generan herramientas nuevas para producir inferencias (e.g., **innovación metodológica**, estadística, o de instrumentos).
- b) Los investigadores **reformulan** las preguntas de investigación, y el uso esperado de resultados en políticas públicas.



Diseño e investigación

El mensaje principal de la idea de constricciones de diseño es al menos dos:

- a) El que hacer de la investigación, los resultados generados, y los argumentos contruidos requieren de tener en cuenta estos límites.
- b) Estos límites pueden ser ampliados (diseñar, crear herramientas, reorientar usos).



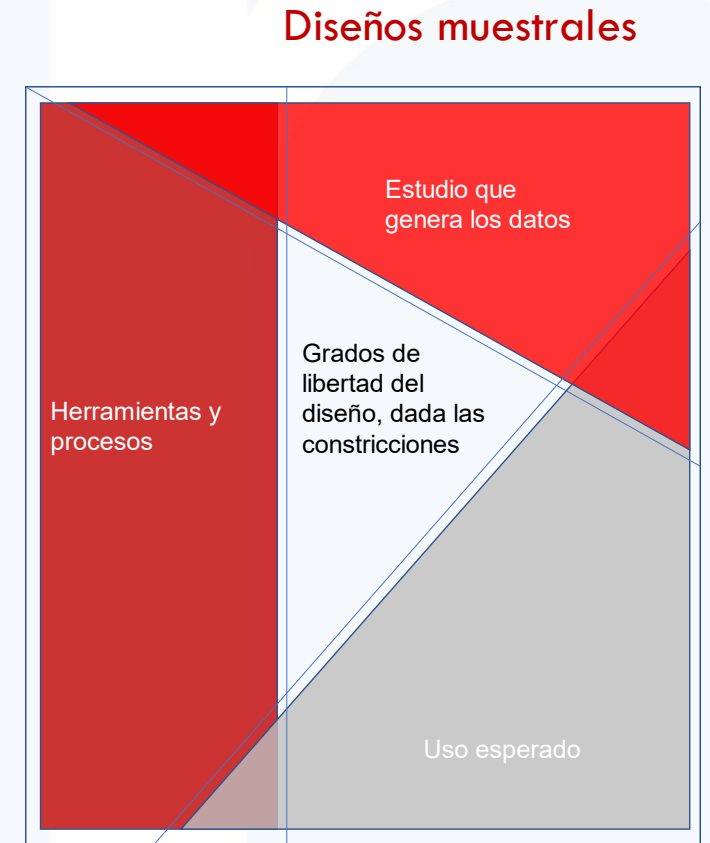
Elementos de diseño

En las próximas dos sesiones vamos a profundizar respecto a dos aspectos:

- a) El tipo de diseño que genera los datos.
- b) Los instrumentos empleados para recoger información.

En la clase de hoy, solo veremos brevemente aspectos de diseño muestral, así como elementos centrales sobre los instrumentos empleados para producir datos.

Instrumentos



Diseños muestrales

Uso esperado

Diseños y tipos de inferencias

Métodos de investigación cuantitativa

UAH-EMAPE

Enfoques Metodológicos para el Análisis de
Políticas Educativas

Septiembre 13
2024

Profesor invitado

Carrasco, D., PhD,
Centro de Medición MIDE UC
Pontificia Universidad Católica de Chile

Métodos cuantitativos como herramientas

Preguntas principales

Qué tipo de preguntas generalmente abordaremos

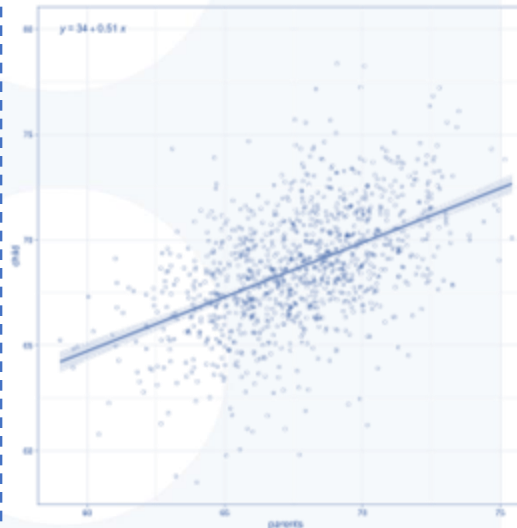
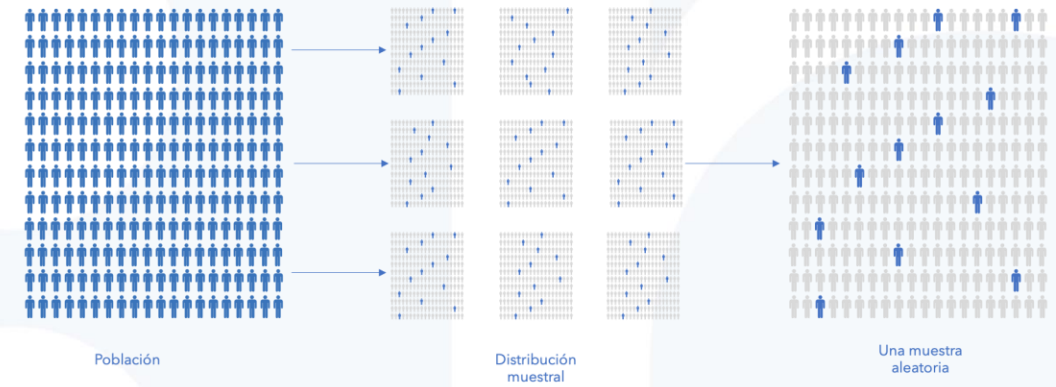
Preguntas fundamentales

Una pregunta propia de la estadística descriptiva, propio de la **inferencia poblacional** (Cobb 2017), es preguntarse por la cantidad o proporción de casos que presenta un atributo o condición.

¿Cuántas observaciones presentan una condición? ¿o qué valor caracteriza a una población?

Desde el punto vista de Vik (2014), desde la **inferencia basada en modelos** (Sterba, 2009), hay al menos tres preguntas fundamentales que orientan al ejercicio de la producción de análisis de datos.

1. ¿Hay una relación entre dos variables?
2. ¿Cuál es la dirección de esta relación?
3. ¿Qué tan fuerte es la relación?



28

Chapter 1 Review of linear regression

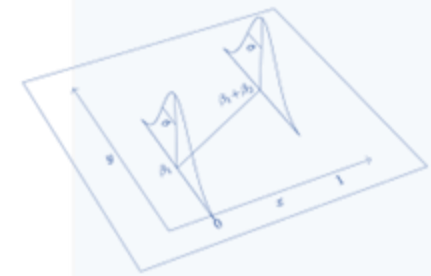


Figure 1.7: Illustration of simple linear regression with a dummy variable

¿Qué es la inferencia?

En términos generales, la **inferencia** es el ejercicio de producir conclusiones a partir de información parcial. La **inferencia estadística**, por su parte, refiere al ejercicio de realizar estas conclusiones, empleando un modelo de probabilidades, que permite vincular las observaciones parciales con nuestras conclusiones.

Existen diferentes formas de inferencia. Sterba (2009) distingue entre las formas de inferencia **basadas en modelos**, y la inferencia basada en **diseños**. Cobb (2017) por su parte distingue entre **inferencia de tipo causal** (via asignación aleatoria), y la **inferencia poblacional** (via muestreo aleatorio).

Primero revisaremos en que consiste la inferencia poblacional, y su relación con el muestreo.

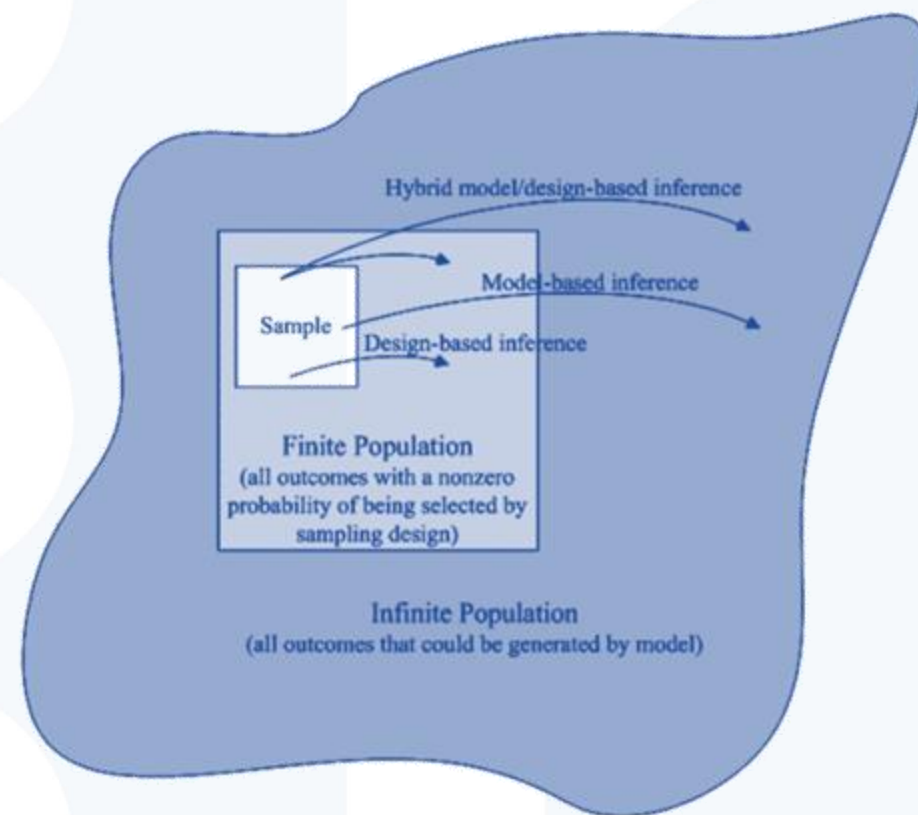


FIGURE 1.
Schematic of alternative populations of inference and mechanisms for inference.

Muestras y probabilidades

Inferencia poblacional

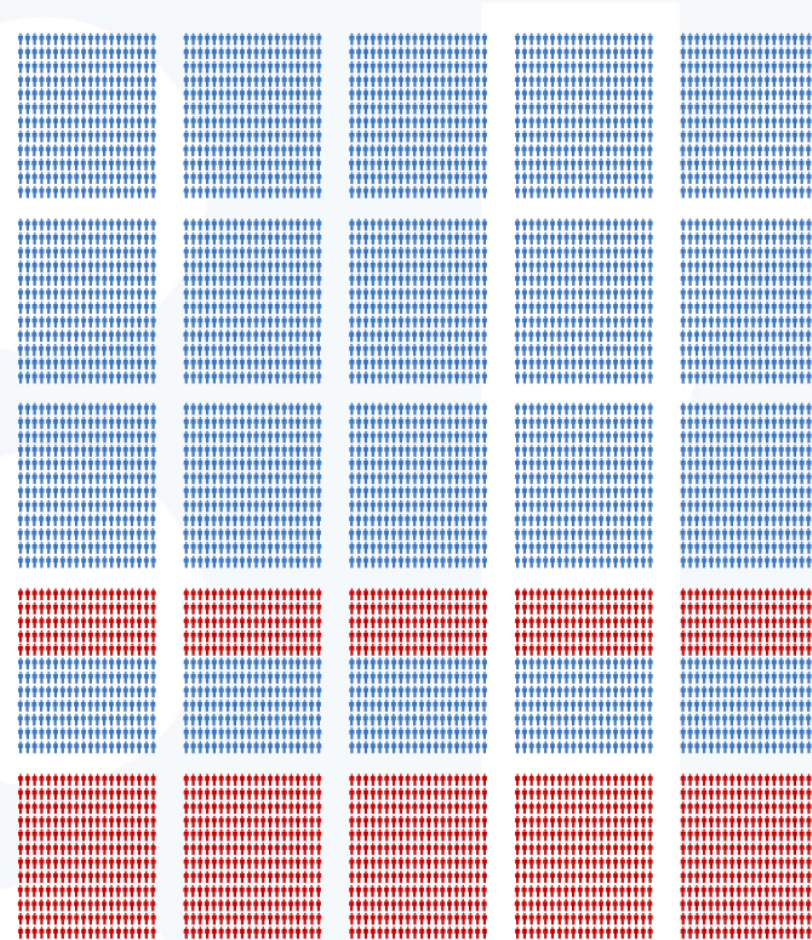
Cómo llegar a conclusiones acerca de eventos o cantidades con información parcial

De la Población finita

Cuando tenemos un conjunto de observaciones totales de una población finita podemos obtener de forma directa un parámetro poblacional. Comencemos con un parámetro sencillo: una proporción.

Por ejemplo, empleando registros administrativos podríamos obtener cual es la cantidad de profesores en Chile, que son mujeres, y que son hombres.

De forma similar, podríamos saber cuál es la cantidad de estudiantes que son hijos de padres con estudios universitarios.



μ

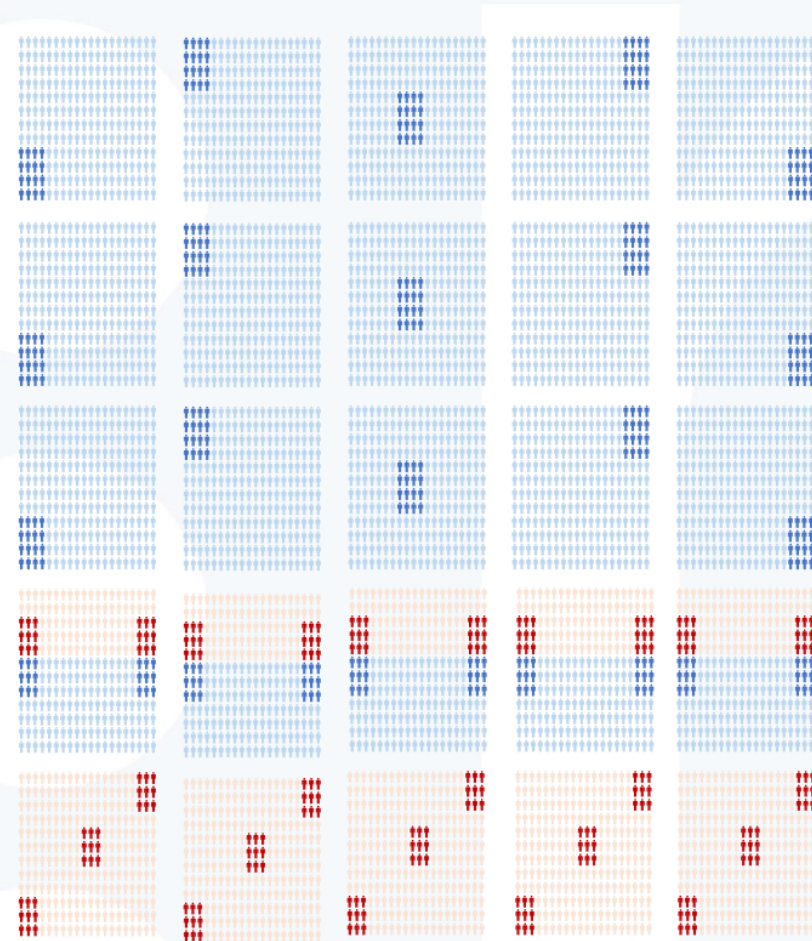
Imaginemos una población finita. Y que tenemos acceso al total de estas observaciones.

Muestra de una población finita

Para ambos ejemplos, si obtuviéramos una muestra de escuelas, y una muestra de profesores en el primer caso, y una muestra de estudiantes al interior de las escuelas, en el segundo caso, podríamos obtener un “estimado” de la proporción de interés en cada caso.

Sin embargo, la cifra de una sola muestra, el estadígrafo, presenta incertidumbre respecto a si se acerca o no al parámetro poblacional.

Pero lo importante, no es lo que sabemos de una sola muestra, sino de a diferentes muestrales posibles.

 μ \bar{x}

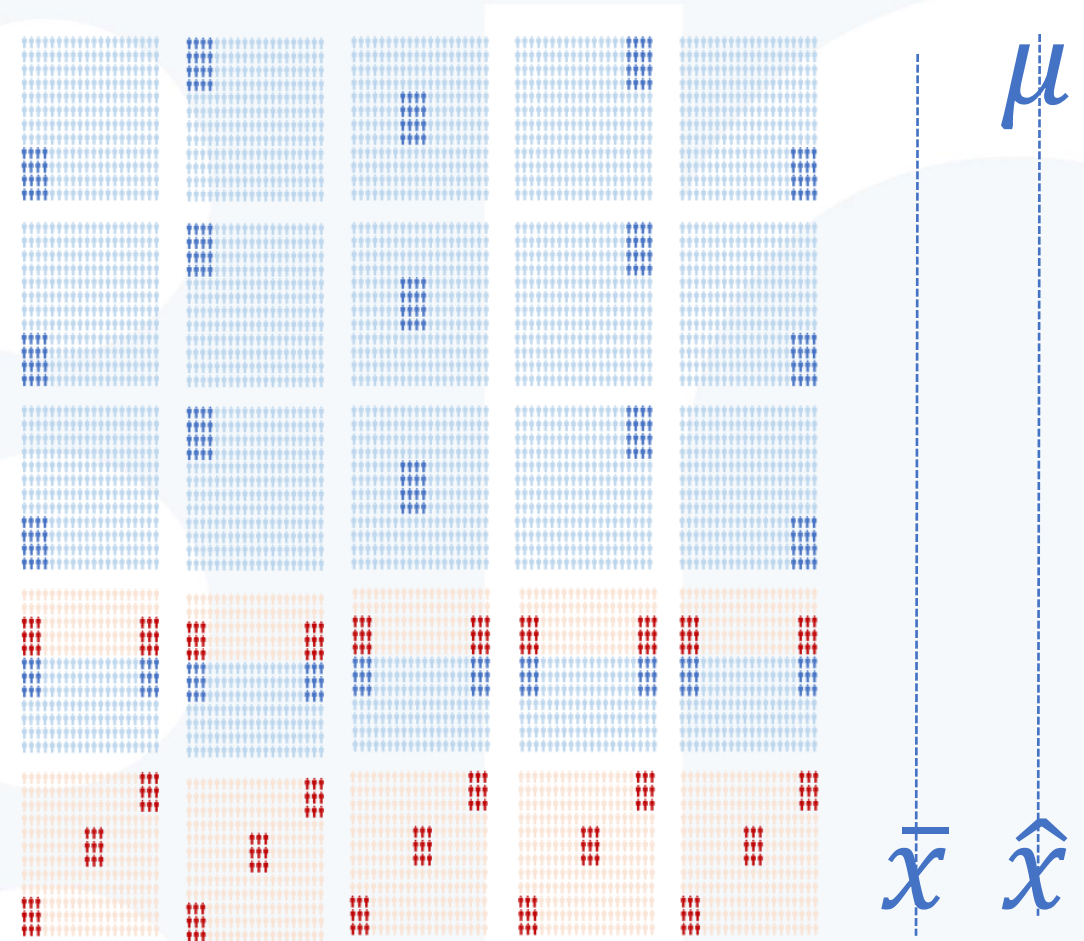
Si tuviéramos una muestra aleatoria, podríamos estimar el parámetro poblacional.

Muestra de una población finita

Para ambos ejemplos, si obtuviéramos una muestra de escuelas, y una muestra de profesores en el primer caso, y una muestra de estudiantes al interior de las escuelas, en el segundo caso, podríamos obtener un “estimado” de la proporción de interés en cada caso.

Sin embargo, la cifra de una sola muestra, el estadígrafo, presenta incertidumbre respecto a si se acerca o no al parámetro poblacional.

Pero lo importante, no es lo que sabemos de una sola muestra, sino de diferentes muestras posible.



Si tuviéramos otra muestra, quizás esta se podría acercar más al parámetro poblacional.

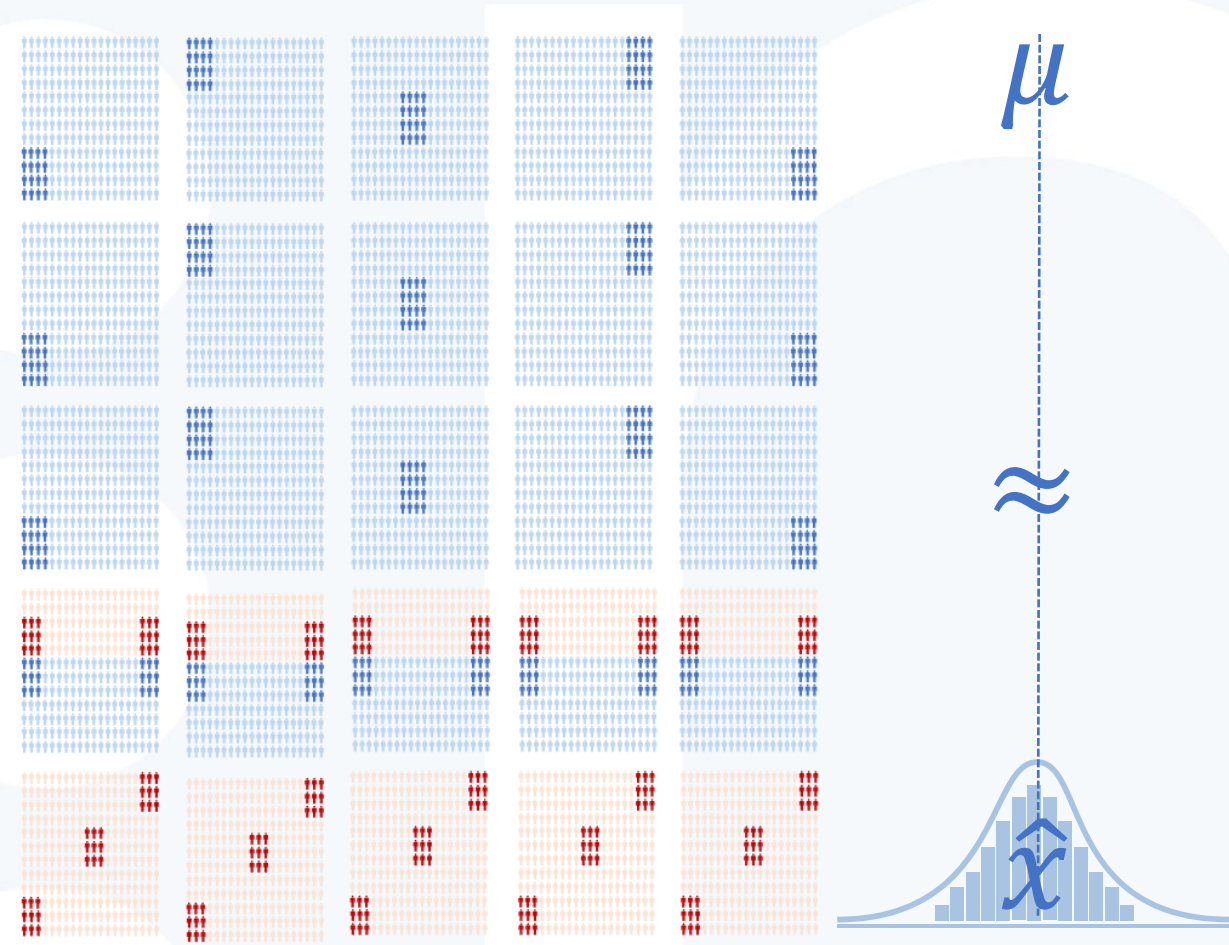
Muestra de una población finita

Para ambos ejemplos, si obtuviéramos una muestra de escuelas, y una muestra de profesores en el primer caso, y una muestra de estudiantes al interior de las escuelas, en el segundo caso, podríamos obtener un “estimado” de la proporción de interés en cada caso.

Sin embargo, la cifra de una sola muestra, el estadígrafo, presenta incertidumbre respecto a si se acerca o no al parámetro poblacional.

Pero lo importante, no es lo que sabemos de una sola muestra, sino de diferentes muestras posibles.

Con una distribución muestral (i.e., muchas muestras aleatorias), de seguro podemos recuperar el parámetro poblacional empleando el promedio de todas las muestras.



De lo que se sabe de las propiedades de una distribución muestral, es que esta nos permite recuperar el parámetro poblacional.

Estadígrafos como estimadores de parámetros poblacionales

Los **estadígrafos son cifras descriptivas** de muestras observadas. Estas cifras pueden ser proporciones, **medias, medianas, desviaciones estándar, o varianzas, u otras cifras**. Estas cifras se pueden obtener tanto para una muestra, como para una población finita.

Gracias a las propiedades que poseen las **distribuciones muestrales** de los estadígrafos, podemos emplear porciones de datos (i.e., muestras aleatorias), para llegar a conclusiones acerca de la población.

En clases posteriores revisaremos el **teorema de limite central, la ley de los grandes números, y diferentes simulaciones**, para ilustrar porque podemos ocupar una muestra aleatoria, y lo que sabemos de la distribución muestral, para realizar inferencias. Es decir, calcular estimados para realizar conclusiones acerca de un parámetro.

Si la forma en que se obtuvieron las observaciones es **aleatoria**, o en un sentido más amplio es **probabilística**, podemos inferir los parámetros poblacionales con los estadígrafos calculados.



Las propiedades de las distribuciones muestrales de los estadígrafos nos permiten realizar inferencias.

Diseños de investigación

Tipos de inferencia

Alcances de los modelos estadísticos condicional a los diseños

Qué inferencias podemos realizar

Las ciencias sociales, y psicología en particular, emplean:

- Experimentos para realizar inferencias acerca de mecanismos
- Se realizan ensayos aleatorios controlados (i.e. *Randomized Control Trials*), para realizar inferencias en la población, respecto a algún tratamiento.
- Se emplean estudios observacionales aleatorios, para describir condiciones en la población (e.g., *encuestas representativas*).

Adicionalmente, existen diferentes estudios, donde el método de obtención de muestra varía en calidad. Desde muestras estratificadas, a muestras por conveniencia. En este último tipo de estudios, las inferencias realizables son muy limitadas.

		Asignación de unidades a condiciones	
		Por aleatorización	Asignación no aleatoria
Selección de observaciones	Aleatoria	Muestras aleatorias son seleccionadas de poblaciones pre - existentes . Estas muestras son asignadas a condiciones experimentales	Muestras aleatorias son seleccionadas de poblaciones pre - existentes .
	No aleatoria	Se obtiene una muestra de casos . Esta muestras es asignadas a condiciones experimentales	Se obtiene una muestra de casos . Esta muestras es examinada.
		Experimento poblacional	Estudio observacional
		Experimento	

¿Es posible realizar inferencias a la población mediante el uso de muestras no probabilísticas? **No.**

Una muestra no probabilística, por azar puede producir resultados razonables, pero carece de fundamento, para evaluar su sesgo y variabilidad de las inferencias realizadas (Heeringa et al. 2009).

Tipos de inferencia



Tipos de inferencia



Tipos de inferencia



Tipos de inferencia

La estadística inferencial nos ofrece herramientas para, dadas ciertas condiciones **realizar afirmaciones acerca de una población, acerca de un mecanismo que explica la generación de datos observados**. Incluso es posible, producir conclusiones de ambos tipos, **inferencias sobre un mecanismo, que sea generalizable a una población**.

Inferencia poblacional

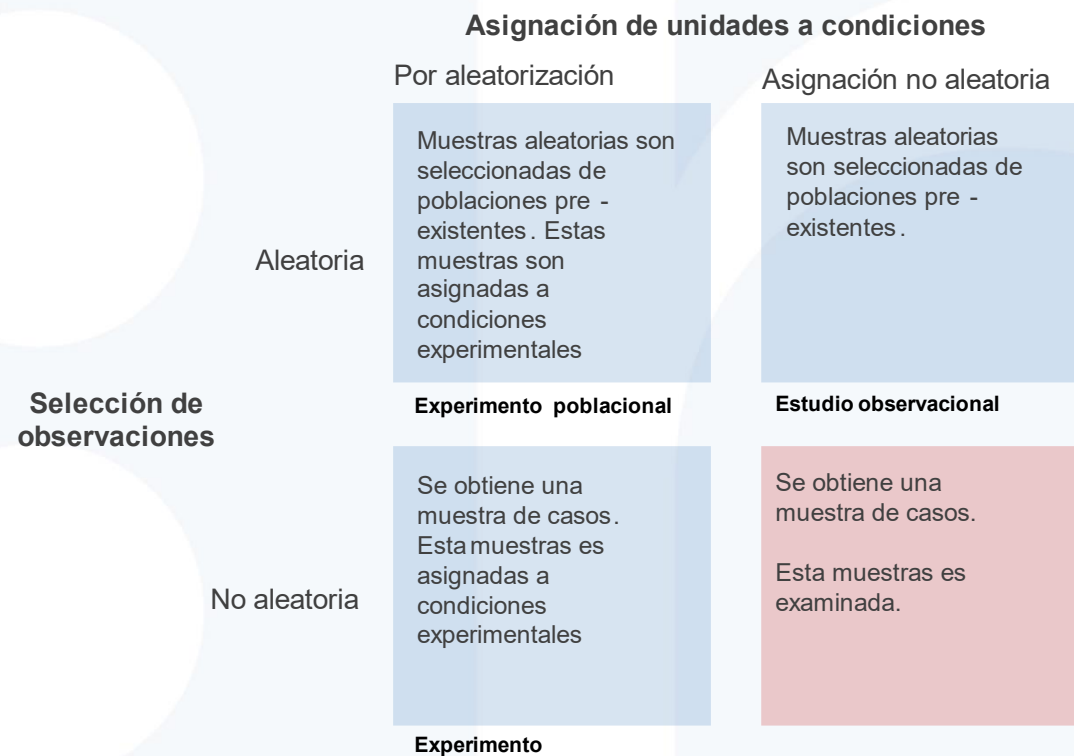
¿Cuánta gente tiene COVID 19?

Inferencia causal

¿Es efectiva la vacuna SINOVAC?

Inferencia poblacional y causal

¿Es efectiva la vacuna SINOVAC para toda la población?



¿Es posible realizar inferencias a la población mediante el uso de muestras no probabilísticas? **No**.

Una muestra no probabilística, por azar puede producir resultados razonables, pero carece de fundamento, para evaluar su sesgo y variabilidad de las inferencias realizadas (Heeringa et al. 2009).

Métodos cuantitativos

Producción e interpretación de cifras

Análisis de datos para abordar preguntas

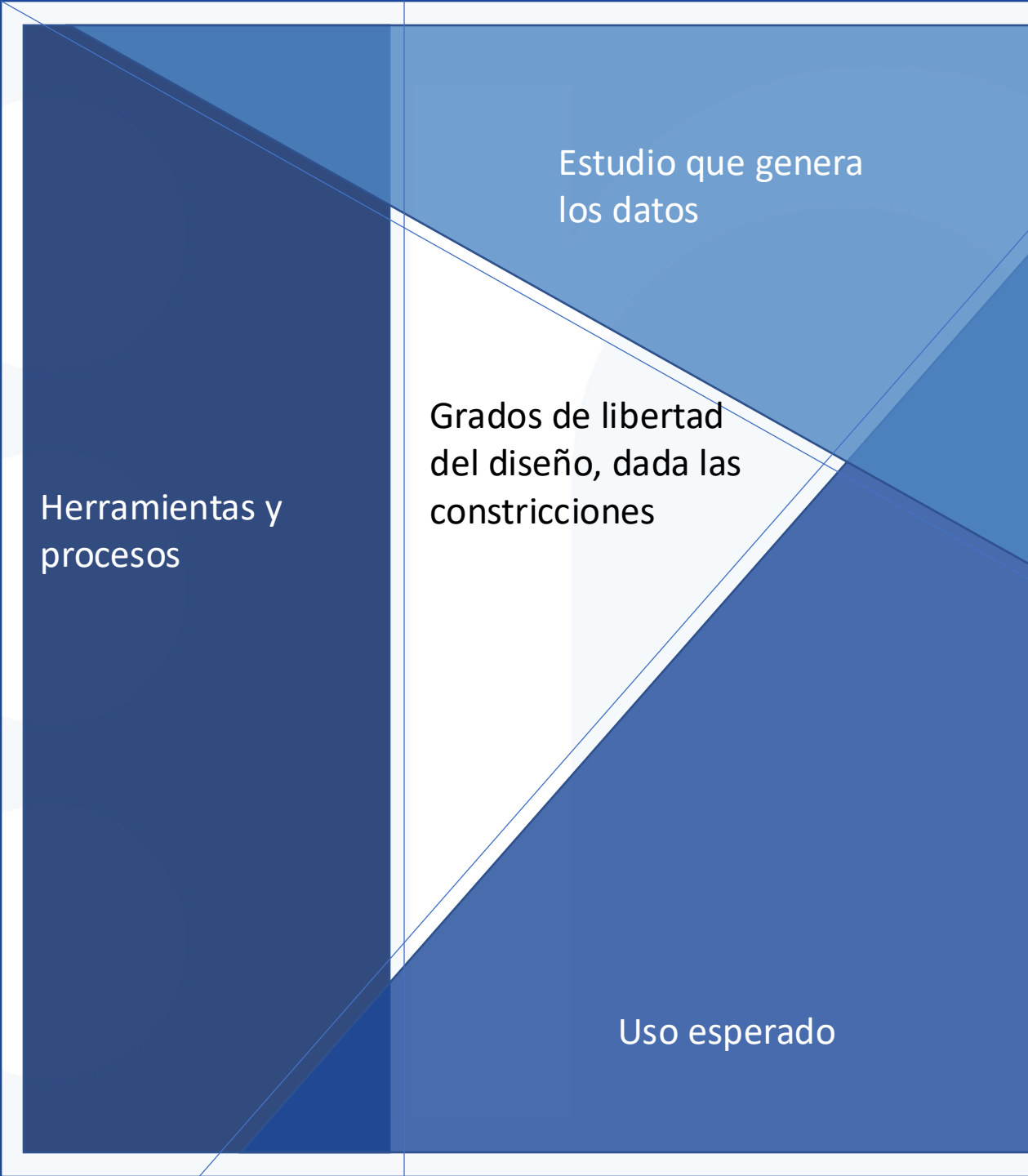
Métodos cuantitativos

Para producir información interpretable se integra conocimiento matemático y estadístico de cómo se producen diferentes cifras.

Se emplea información respecto a las propiedades de las distribuciones muestrales de los estadígrafos, para realizar inferencias a la población, utilizando muestras de datos obtenidas por métodos particulares (i.e., muestras aleatorias, o probabilísticas).

Además, se considera el tipo de escala de medición en el cual se encuentran registradas las variables. Esto último guía nuestras prácticas respecto a que operaciones realizar, cómo interpretar las cifras generadas.

Finalmente, se integra información acerca del diseño general para condicionar los alcances de nuestras conclusiones.



Muchas gracias!

Referencias

- Cobb, G. W. (2007). The Introductory Statistics Course: A Ptolemaic Curriculum. *Technology Innovations in Statistics Education*, 1(1), 1–16.
- Heeringa, S. G., West, B., & Berglund, P. A. (2009). *Applied Survey Data Analysis*. Taylor & Francis Group.
- Parker, H. (2017). Opinionated analysis development (pp. 1–13). <https://doi.org/10.7287/peerj.preprints.3210>
- Ramsey, F., & Schafer, D. (2012). *The Statistical Sleuth: A Course in Methods of Data Analysis* (3rd ed.). Brooks/Cole.
- Sterba, S. K. (2009). Alternative Model-Based and Design-Based Frameworks for Inference From Samples to Populations: From Polarization to Integration. In *Multivariate Behavioral Research* (Vol. 44, Issue 6, pp. 711–740). <https://doi.org/10.1080/00273170903333574>
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684), 677–680.
- Torres Irribarra, D. (2021). *A Pragmatic Perspective of Measurement*. Springer. <https://doi.org/10.1007/978-3-030-74025-2>
- Vik, P. (2014). Introduction. In *Regression, ANOVA, and the general linear model: A statistics primer* (pp. 1–6). Sage.
- Yau, N. (2015) <https://flowingdata.com/2015/02/12/visualization-constraints>

Carrasco, D., PhD
Centro de Medición MIDE UC,
Pontificia Universidad Católica de Chile
<https://dacarras.github.io/>