

Aplicación web interactiva para el cálculo del tamaño muestral con RShiny

Iñaki Salgado Uralde



Universidad de Oviedo



Escuela de
Ingeniería
Informática
Universidad de Oviedo

Trabajo Fin de Grado - Grado en Ingeniería Informática del Software



Índice

- Motivación y objetivos
- Estudio de la situación actual
- Aspectos teóricos
- Planificación y presupuesto
- Análisis, diseño y pruebas
- Implementación del sistema
- Conclusiones y ampliaciones

Motivación

El cálculo del tamaño de muestra es fundamental en la planificación de estudios estadísticos.



```
graph TD; A[El cálculo del tamaño de muestra es fundamental en la planificación de estudios estadísticos.] --> B[Un tamaño muestral inadecuado puede afectar la validez de los resultados, ya sea por falta de poder estadístico en muestras pequeñas o por el desperdicio de recursos en muestras demasiado grandes.]; B --> C[Sin una herramienta adecuada, determinar el tamaño muestral puede ser complejo y propenso a errores, especialmente para usuarios sin un sólido conocimiento estadístico.];
```

Un tamaño muestral inadecuado puede afectar la validez de los resultados, ya sea por falta de poder estadístico en muestras pequeñas o por el desperdicio de recursos en muestras demasiado grandes.

Sin una herramienta adecuada, determinar el tamaño muestral puede ser complejo y propenso a errores, especialmente para usuarios sin un sólido conocimiento estadístico.

Objetivos I



Desarrollo de Algoritmos:

Implementar algoritmos para calcular el tamaño muestral para medias y proporciones poblacionales.



Diseño de Interfaz:

Crear una interfaz de usuario fácil de usar y accesible.



Validación:

Implementar feedback para informar sobre posibles errores en los datos ingresados.



Compatibilidad y Accesibilidad:

Hacer la aplicación accesible para usuarios con diferentes niveles de conocimiento estadístico.

Objetivos II



Documentación:

Crear documentación técnica y una sección de ayuda dentro de la aplicación.



Aprendizaje y Desarrollo:

Ampliar conocimientos en el lenguaje R aplicado al desarrollo web con Shiny.



Publicación:

Publicar la aplicación en una plataforma accesible para su uso gratuito.

Estudio de la situación actual I

Aplicaciones similares:

- QuestionPro
- SurveyMonkey
- Qualtrics:

Calculadora del tamaño de la muestra

Gracias a nuestra calculadora y, una vez que hayas determinado sus variantes, podrás calcular el tamaño de la muestra en cuestión de segundos:

Nivel de confianza:

95% ▼

Tamaño de la población:

10000

Margen de error:

5% ▼

Tamaño ideal de la muestra:

370

Estudio de la situación actual II

Ventajas:

- Específicamente diseñado para el cálculo muestral.
- Uso de R y Shiny para cálculos precisos.

Desventajas:

- Menos funcionalidades de encuestas comparado con las otras herramientas.

Aspectos teóricos I



Denominamos POBLACIÓN a la colección de elementos sobre la que se desea examinar el comportamiento del experimento considerado. Los elementos de la población se denominan INDIVIDUOS.



Llamamos MUESTRA al conjunto de individuos de la población de los cuales se obtiene información, cuando no es posible obtenerla de todos los individuos que la componen.



Cuando se extrae una muestra, los datos obtenidos nos permiten inferir unos valores aproximados de la población en su totalidad. A estos valores aproximados se les denomina ESTIMACIONES, las cuales estarán afectadas por un error conocido como ERROR DEBIDO AL MUESTREO. Cuanto menor sea este error, mayor será la PRECISIÓN de las estimaciones.

Aspectos teóricos II

Media poblacional

Para calcular el tamaño muestral necesario para estimar una media poblacional con varianza conocida σ^2 se utilizan las siguientes fórmulas:

- Con un error de muestreo fijado e :
 - Con un Muestreo Aleatorio con Reposición (M.A.C.R.):

$$n_0 = \frac{\sigma^2}{e^2}$$

- Con un Muestreo Aleatorio Simple (M.A.S.) sobre una población de N individuos:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

- Con un error máximo admisible e_α y un nivel de confianza $1-\alpha$ fijados:
 - Con un Muestreo Aleatorio con Reposición (M.A.C.R.):
Primero se calculan los valores dados por la aproximación normal (n_{01}) y la desigualdad de Tchebychev (n_{02}):

$$n_{01} = \frac{z_{1-\alpha/2}^2 \cdot \sigma^2}{e_\alpha^2},$$

siendo $z_{1-\alpha/2}^2$ el cuantil de orden $1-\alpha/2$ de la distribución normal estándar.

$$n_{02} = \frac{\sigma^2}{\alpha \cdot e_\alpha^2}$$

Entonces el tamaño muestral requerido (n) viene dado por:

$$n = \begin{cases} n_{01} & \text{si } n_{01} \geq 30 \\ \min\{30, n_{02}\} & \text{en otro caso.} \end{cases}$$

- Con un Muestreo Aleatorio Simple (M.A.S.) sobre una población de N individuos:

Primero se calculan los valores dados por la aproximación normal (n_1) y la desigualdad de Tchebychev (n_2):

$$n_1 = \frac{n_{01}}{1 + \frac{n_{01}}{N}}$$

$$n_2 = \frac{n_{02}}{1 + \frac{n_{02}}{N}}$$

Entonces el tamaño muestral requerido (n) viene dado por:

$$n = \begin{cases} n_1 & \text{si } n_1 \geq 100 \\ \min\{100, n_2\} & \text{en otro caso.} \end{cases}$$

Aspectos teóricos III

Proporción poblacional

Para calcular el tamaño muestral necesario para estimar una proporción poblacional a partir de una estimación conocida P se utilizan las siguientes fórmulas:

- Con un error de muestreo fijado e :
 - Con un Muestreo Aleatorio con Reposición (M.A.C.R.):

$$n_0 = \frac{P(1-P)}{e^2}$$

- Con un Muestreo Aleatorio Simple (M.A.S.) sobre una población de N individuos:

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

- Con un error máximo admisible e_α y un nivel de confianza $1-\alpha$ fijados:
 - Con un Muestreo Aleatorio con Reposición (M.A.C.R.):
Primero se calculan los valores dados por la aproximación normal (n_{01}) y la desigualdad de Tchebychev (n_{02}):

$$n_{01} = \frac{z_{1-\alpha/2}^2 \cdot P(1-P)}{e_\alpha^2},$$

siendo $z_{1-\alpha/2}^2$ el cuantil de orden $1-\alpha/2$ de la distribución normal estándar.

$$n_{02} = \frac{P(1-P)}{\alpha \cdot e_\alpha^2}$$

Entonces el tamaño muestral requerido (n) viene dado por:

$$n = \begin{cases} n_{01} & \text{si } n_{01} \geq 30 \\ \min\{30, n_{02}\} & \text{en otro caso.} \end{cases}$$

- Con un Muestreo Aleatorio Simple (M.A.S.) sobre una población de N individuos:

Primero se calculan los valores dados por la aproximación normal (n_1) y la desigualdad de Tchebychev (n_2):

$$n_1 = \frac{n_{01}}{1 + \frac{n_{01} - 1}{N}}$$

$$n_2 = \frac{n_{02}}{1 + \frac{n_{02} - 1}{N}}$$

Entonces el tamaño muestral requerido (n) viene dado por:

$$n = \begin{cases} n_1 & \text{si } n_1 \geq 100 \\ \min\{100, n_2\} & \text{en otro caso.} \end{cases}$$

Planificación

01

Análisis y diseño: desarrollo de la especificación de requisitos, diseño de los prototipos, planificación del servidor y la interfaz gráfica, así como análisis de la situación actual, alternativas y un estudio de las herramientas a utilizar.

02

Implementación: contiene el desarrollo tanto de la parte de servidor como la parte de la interfaz gráfica. Además de la refactorización de código que se debe realizar a lo largo de todo el trabajo.

03

Pruebas: incluye el diseño del plan de pruebas y la implementación de las pruebas unitarias, usabilidad y accesibilidad.

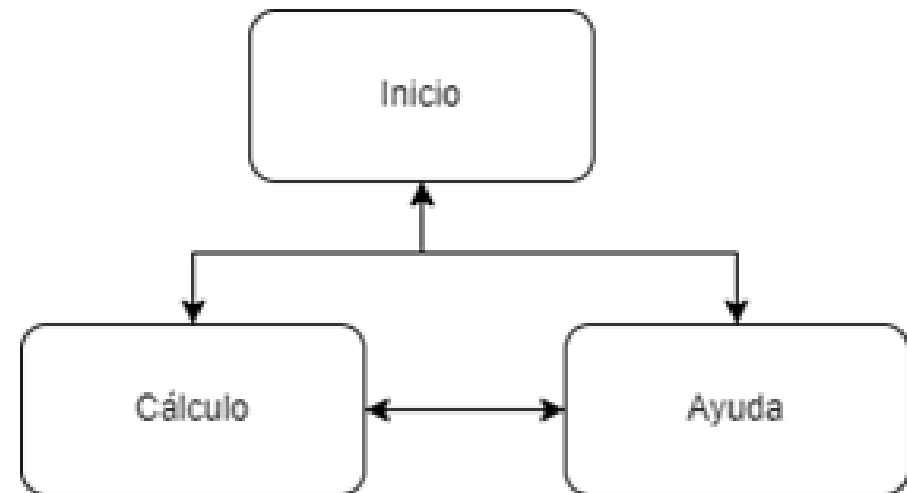
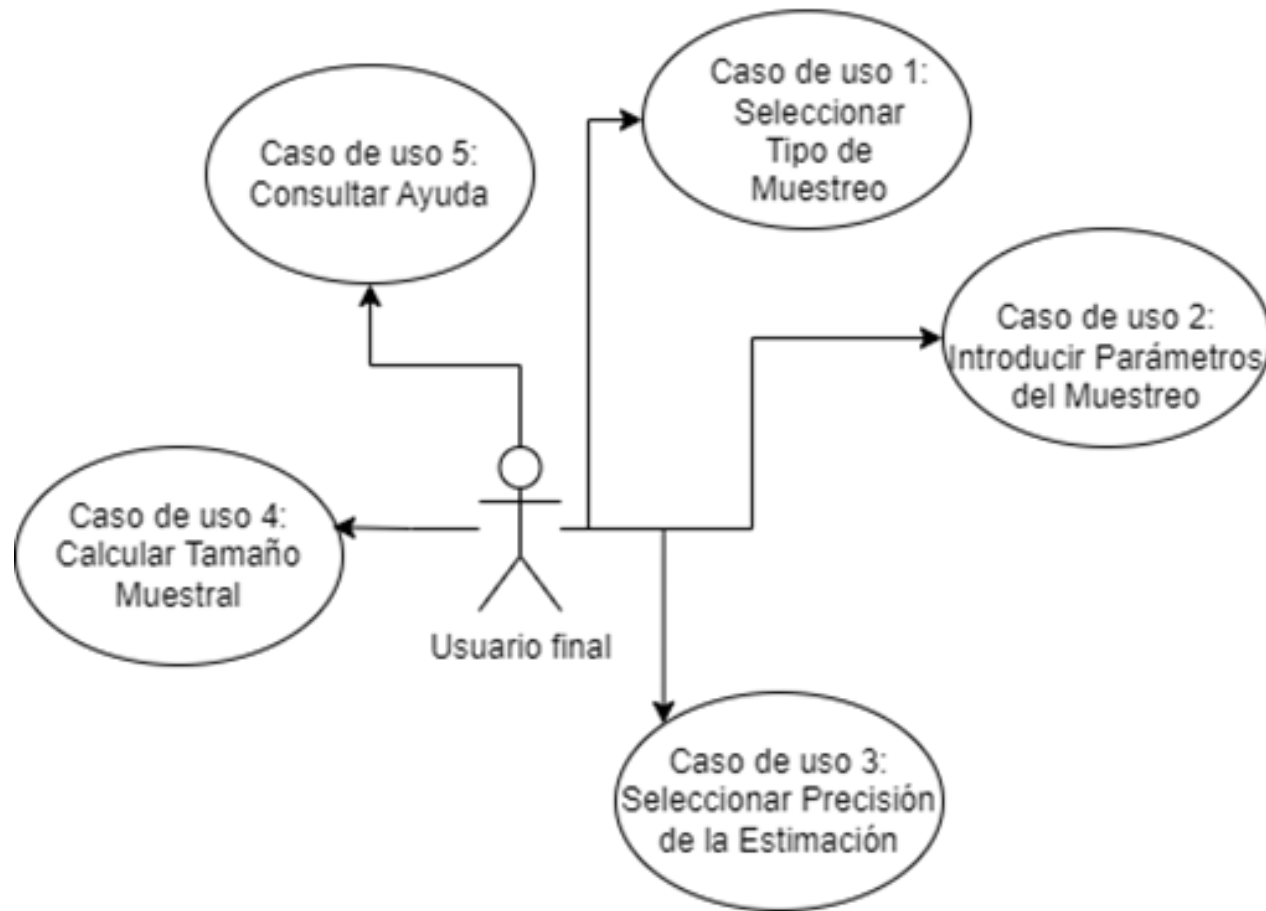
04

Documentación: cada una de las fases anteriores producirá una serie de elementos que formarán parte de la memoria final.

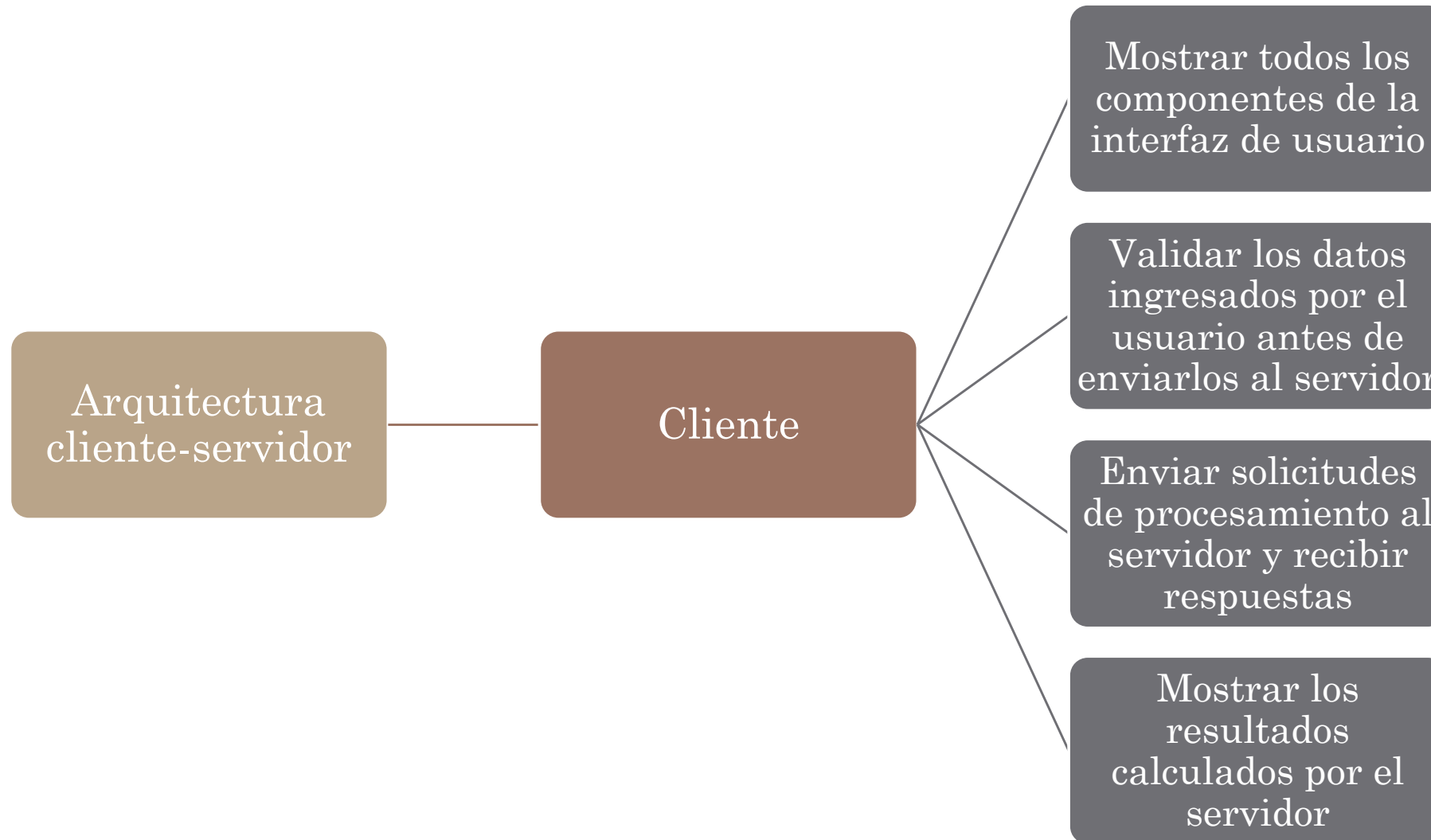
Presupuesto

Item	SubItem	Concepto	Cantidad	Precio Unitario	TOTAL
01		Desarrollo de la Aplicación			6.475,00 €
	001	Análisis y diseño	20,00	35,00 €	700,00 €
	002	Implementación	100,00	35,00 €	3.500,00 €
	003	Pruebas	25,00	35,00 €	875,00 €
	004	Documentación	40,00	35,00 €	1.400,00 €
02		Licencias software			81,25 €
	001	R y paquetes asociados	1,00	0,00 €	0,00 €
	002	Rstudio Desktop Pro	1,00	50,00 €	50,00 €
	003	Microsoft Word	1,00	31,25 €	31,25 €
03		Gastos indirectos			237,50 €
	001	Internet	2,50	30,00 €	75,00 €
	002	Agua	2,50	30,00 €	75,00 €
	003	Luz	2,50	35,00 €	87,50 €
Base					6.793,75 €
Beneficios(10%)					679,38 €
Subtotal					7.473,13 €
IVA (21%)					1.569,36 €
TOTAL					9.042,48 €

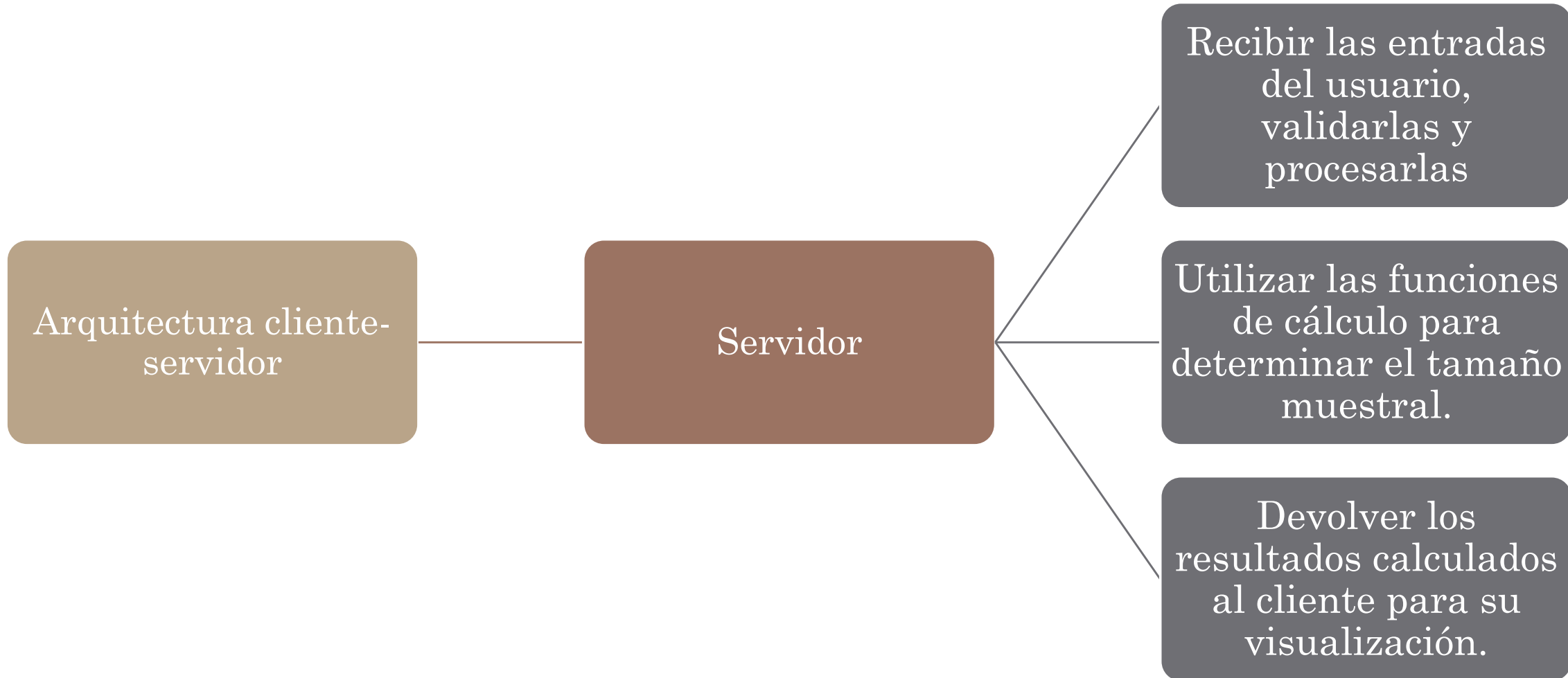
Análisis



Diseño I



Diseño II



Pruebas

- Unitarias
- Integración y del sistema
- Usabilidad
- Accesibilidad

```
> test_dir("test/testthat")
✓ | F W S OK | Context
✓ |          16 | formulas
```


```
== Results ==
[ FAIL 0 | WARN 0 | SKIP 0 | PASS 16 ]
> |
```

Accessibility Review

Accessibility Review (Guidelines: [WCAG 2.0 \(Level AA\)](#))

Export Format: Report to Export: [Get File](#)

Known Problems(0) **Likely Problems (0)** **Potential Problems (19)** **HTML Validation** **CSS Validation**

 **Congratulations! No known problems.**

Implementación del sistema

Tecnologías utilizadas:

R: Lenguaje de programación estadística.

Shiny: Framework para aplicaciones web interactivas en R.

RStudio: Entorno de desarrollo integrado para R.

Componentes del sistema:

Interfaz de usuario: Diseño y funcionalidad.

Servidor: Lógica de cálculo y validación.

Fórmulas: Funciones para calcular el tamaño muestral.

Conclusiones



Herramienta funcional y útil para el cálculo muestral.



Aprendizaje de nuevas tecnologías (R y Shiny).



Diseño intuitivo para facilidad de uso.

Ampliaciones

Añadir nuevos
parámetros de interés
y tipos de error

Internacionalización

Integración con bases
de datos

Gracias



Universidad de Oviedo



Escuela de
Ingeniería
Informática
Universidad de Oviedo