



# Universidad Nacional del Altiplano

Facultad de Ingeniería Estadística e Informática

## Sistema de Análisis Estadístico Avanzado

Aplicación Web con Shiny en R

**Estudiante:** Nelson Catunta Huisa

**Docente:** Fred Torres Cruz

**Curso:** Estadística Computacional

**Fecha:** 9 de julio de 2025

**Repositorio GitHub:**

<https://github.com/CAISA25/APUESTA.git>

Puno, Perú

# Índice

<b>1. Introducción</b>	<b>5</b>
1.1. Objetivos	5
1.2. Alcance	5
<b>2. Marco Teórico</b>	<b>5</b>
2.1. Estadística Descriptiva	5
2.1.1. Medidas de Tendencia Central	5
2.1.2. Medidas de Dispersión	6
2.2. Pruebas de Hipótesis	6
2.2.1. Prueba t de Student	6
2.2.2. Análisis de Varianza (ANOVA)	6
2.2.3. Prueba Chi-cuadrado	6
<b>3. Metodología</b>	<b>6</b>
3.1. Tecnologías Utilizadas	6
3.2. Arquitectura del Sistema	7
<b>4. Implementación</b>	<b>7</b>
4.1. Estructura de la Aplicación	7
4.2. Funciones Principales	7
4.2.1. Función para Cálculo de Moda	7
4.2.2. Función para Gráficos de Pastel	8
<b>5. Funcionalidades del Sistema</b>	<b>8</b>
5.1. Carga de Datos	8
5.2. Análisis Descriptivo	8
5.3. Visualizaciones	9
<b>6. Pruebas Estadísticas</b>	<b>9</b>
6.1. Pruebas Implementadas	9
6.2. Interpretación Automática	9
<b>7. Imágenes del Sistema</b>	<b>9</b>
7.1. Interfaz Principal	10
7.2. Módulo de Carga de Datos	10
7.3. Estadísticas Descriptivas	11
7.4. Gráficos Estadísticos	12
7.5. Pruebas Estadísticas	12
7.6. Análisis de Normalidad	13
<b>8. Casos de Uso</b>	<b>13</b>
8.1. Caso de Uso 1: Análisis de Encuesta	13
8.2. Caso de Uso 2: Análisis de Calidad	13
<b>9. Validación y Pruebas</b>	<b>14</b>
9.1. Pruebas de Funcionalidad	14
9.2. Pruebas de Rendimiento	14

<b>10.Conclusiones</b>	<b>14</b>
10.1. Logros Alcanzados . . . . .	14
10.2. Limitaciones . . . . .	14
10.3. Trabajo Futuro . . . . .	14
<b>11.Anexos</b>	<b>15</b>
11.1. Código Fuente Completo . . . . .	15
11.2. Manual de Usuario . . . . .	15
11.3. Instalación y Configuración . . . . .	15
<b>12.Referencias</b>	<b>15</b>

## Índice de figuras

1.	Interfaz principal de la aplicación . . . . .	10
2.	Módulo de importación de datos . . . . .	10
3.	Vista de estadísticas descriptivas . . . . .	11
4.	Ejemplos de visualizaciones generadas . . . . .	12
5.	Resultados de pruebas estadísticas con interpretación . . . . .	12
6.	Análisis de normalidad y detección de outliers . . . . .	13

## Índice de cuadros

1.	Estadísticas por tipo de variable . . . . .	8
2.	Resumen de pruebas estadísticas . . . . .	9
3.	Resultados de pruebas de funcionalidad . . . . .	14

# 1 Introducción

El presente documento describe el desarrollo de un sistema web para análisis estadístico avanzado, implementado en R usando el framework Shiny. Esta aplicación permite realizar análisis descriptivo, inferencial y exploratorio de datos de manera intuitiva y profesional.

## 1.1 Objetivos

- Desarrollar una aplicación web interactiva para análisis estadístico
- Implementar herramientas de estadística descriptiva e inferencial
- Generar visualizaciones estadísticas de calidad profesional
- Proporcionar interpretaciones automáticas de pruebas estadísticas
- Crear reportes automatizados en formato PDF

## 1.2 Alcance

La aplicación cubre las siguientes áreas del análisis estadístico:

- Estadística descriptiva (medidas de tendencia central, dispersión, forma)
- Visualización de datos (histogramas, boxplots, gráficos de barras, pasteles, ojivas)
- Pruebas de hipótesis (t-test, ANOVA, Chi-cuadrado, Wilcoxon)
- Análisis de correlación (Pearson, Spearman)
- Análisis de normalidad (Shapiro-Wilk)
- Detección de valores atípicos
- Teorema del Límite Central

# 2 Marco Teórico

## 2.1 Estadística Descriptiva

La estadística descriptiva se encarga de resumir y presentar la información contenida en una muestra o población de datos. Las principales medidas incluyen:

### 2.1.1 Medidas de Tendencia Central

- **Media aritmética:**  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- **Mediana:** Valor que divide la distribución en dos partes iguales
- **Moda:** Valor que aparece con mayor frecuencia

### 2.1.2 Medidas de Dispersión

- **Desviación estándar:**  $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$
- **Varianza:**  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
- **Rango:**  $R = x_{max} - x_{min}$
- **Coeficiente de variación:**  $CV = \frac{s}{\bar{x}} \times 100 \%$

## 2.2 Pruebas de Hipótesis

### 2.2.1 Prueba t de Student

Para comparar medias de dos grupos:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (1)$$

donde  $s_p$  es la desviación estándar pooled.

### 2.2.2 Análisis de Varianza (ANOVA)

Para comparar medias de tres o más grupos:

$$F = \frac{MSB}{MSW} \quad (2)$$

donde  $MSB$  es la media cuadrática entre grupos y  $MSW$  es la media cuadrática dentro de grupos.

### 2.2.3 Prueba Chi-cuadrado

Para analizar la independencia de variables categóricas:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

## 3 Metodología

### 3.1 Tecnologías Utilizadas

- **R:** Lenguaje de programación estadística
- **Shiny:** Framework para aplicaciones web interactivas
- **Bibliotecas especializadas:** ggplot2, dplyr, plotly, knitr, rmarkdown

## 3.2 Arquitectura del Sistema

La aplicación sigue el patrón MVC (Modelo-Vista-Controlador) de Shiny:

- **UI (Interfaz de Usuario):** Define la estructura y apariencia
- **Server (Servidor):** Contiene la lógica de procesamiento
- **Reactividad:** Permite actualizaciones automáticas de resultados

## 4 Implementación

### 4.1 Estructura de la Aplicación

La aplicación está organizada en las siguientes secciones:

1. **Importar Datos:** Carga de archivos CSV y Excel
2. **Estadísticas Descriptivas:** Análisis básico de variables
3. **Gráficos por Variable:** Visualizaciones detalladas
4. **Pruebas e Interpretación:** Análisis inferencial
5. **Extras:** Análisis de normalidad y outliers
6. **Nuevas Pruebas:** Pruebas estadísticas avanzadas
7. **Exportar Reporte:** Generación de documentos PDF

### 4.2 Funciones Principales

#### 4.2.1 Función para Cálculo de Moda

Listing 1: Función para calcular moda de forma segura

```
moda_segura <- function(x) {  
  x <- na.omit(x)  
  if (length(x) == 0) return(NA)  
  freq <- table(x)  
  max_freq <- max(freq)  
  if (max_freq == 1) {  
    return("No hay moda")  
  }  
  moda <- names(freq[freq == max_freq])  
  if (length(moda) == 1) {  
    return(moda)  
  } else {  
    return(paste(moda, collapse = ", "))  
  }  
}
```



### 4.2.2 Función para Gráficos de Pastel

Listing 2: Función para crear gráficos de pastel

```
crear_grafico_pastel <- function(datos, variable) {
  freq_table <- table(datos)
  df_pie <- data.frame(
    categoria = names(freq_table),
    frecuencia = as.numeric(freq_table),
    porcentaje = round(as.numeric(freq_table) / sum(freq_table) * 100,
      1)
  )

  ggplot(df_pie, aes(x = "", y = frecuencia, fill = categoria)) +
    geom_bar(stat = "identity", width = 1) +
    coord_polar("y", start = 0) +
    geom_text(aes(label = paste0(porcentaje, "%")),
      position = position_stack(vjust = 0.5)) +
    theme_void() +
    labs(title = paste("Gráfico de Pastel:", variable),
      fill = "Categorías") +
    theme(plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))
}
```

## 5 Funcionalidades del Sistema

### 5.1 Carga de Datos

El sistema permite importar datos desde archivos CSV y Excel, con las siguientes características:

- Validación automática de formatos
- Vista previa de datos
- Selección de variables para análisis
- Manejo de errores en la carga

### 5.2 Análisis Descriptivo

Para cada variable, el sistema calcula:

Variables Numéricas	Variables Categóricas
Media aritmética	Tabla de frecuencias
Mediana	Distribución porcentual
Moda	Categoría modal
Desviación estándar	Número de categorías
Coeficiente de variación	Gráfico de barras
Rango	Gráfico de pastel
Cuartiles	Análisis de uniformidad

Cuadro 1: Estadísticas por tipo de variable

## 5.3 Visualizaciones

El sistema genera automáticamente:

- **Histogramas:** Con curva de densidad superpuesta
- **Boxplots:** Para identificar outliers y distribución
- **Gráficos de barras:** Para variables categóricas
- **Gráficos de pastel:** Con porcentajes automáticos
- **Ojivas:** Para análisis de frecuencias acumuladas

## 6 Pruebas Estadísticas

### 6.1 Pruebas Implementadas

Prueba	Propósito	Condiciones
t-test	Comparar dos medias	Variables numéricas
ANOVA	Comparar múltiples medias	Una numérica, una categórica
Chi-cuadrado	Independencia de variables	Variables categóricas
Shapiro-Wilk	Normalidad	Variables numéricas
Wilcoxon	Comparar distribuciones	Variables numéricas
Correlación	Asociación lineal	Dos variables numéricas

Cuadro 2: Resumen de pruebas estadísticas

### 6.2 Interpretación Automática

El sistema proporciona interpretaciones automáticas basadas en:

- Valores  $p$  ( $= 0.05$ )
- Intervalos de confianza
- Tamaños de efecto
- Recomendaciones prácticas

## 7 Imágenes del Sistema

## 7.1 Interfaz Principal



Figura 1: Interfaz principal de la aplicación

## 7.2 Módulo de Carga de Datos

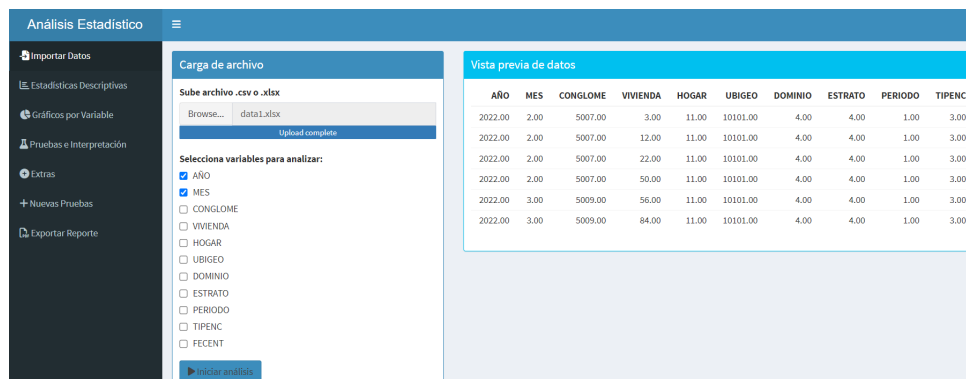


Figura 2: Módulo de importación de datos

### 7.3 Estadísticas Descriptivas

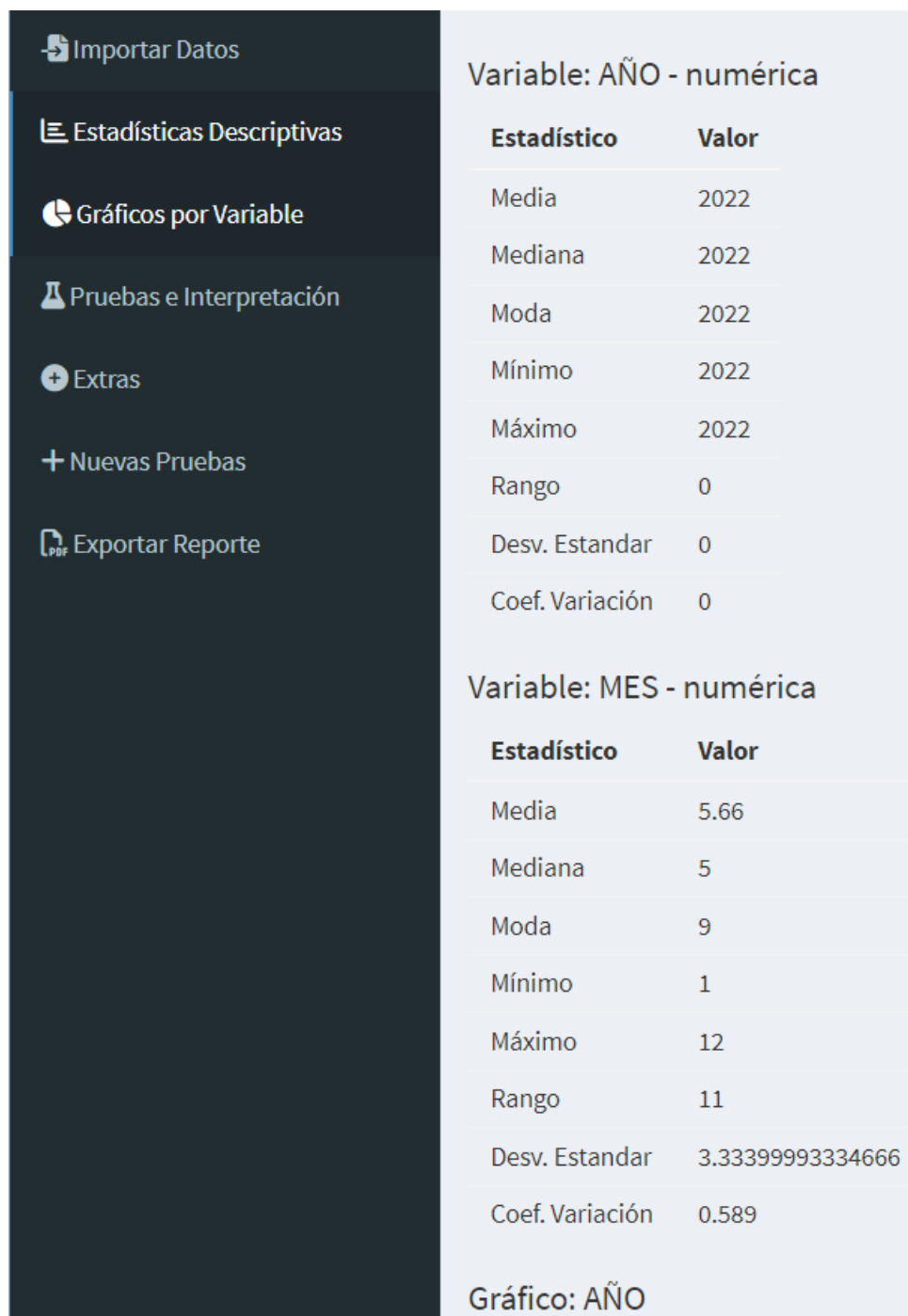
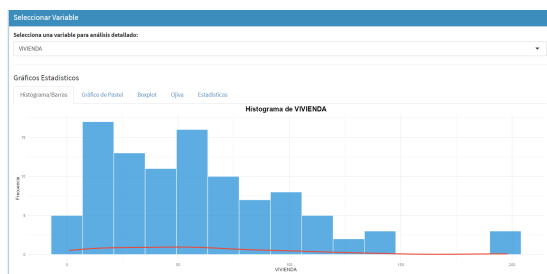
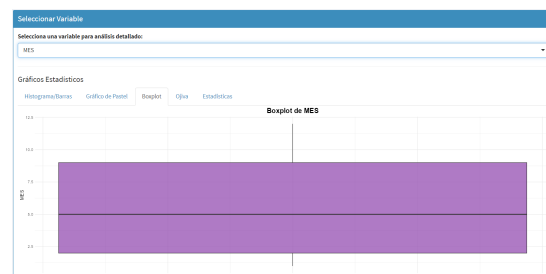


Figura 3: Vista de estadísticas descriptivas

## 7.4 Gráficos Estadísticos



(a) Histograma con curva de densidad



(b) Boxplot para detección de outliers

Figura 4: Ejemplos de visualizaciones generadas

## 7.5 Pruebas Estadísticas

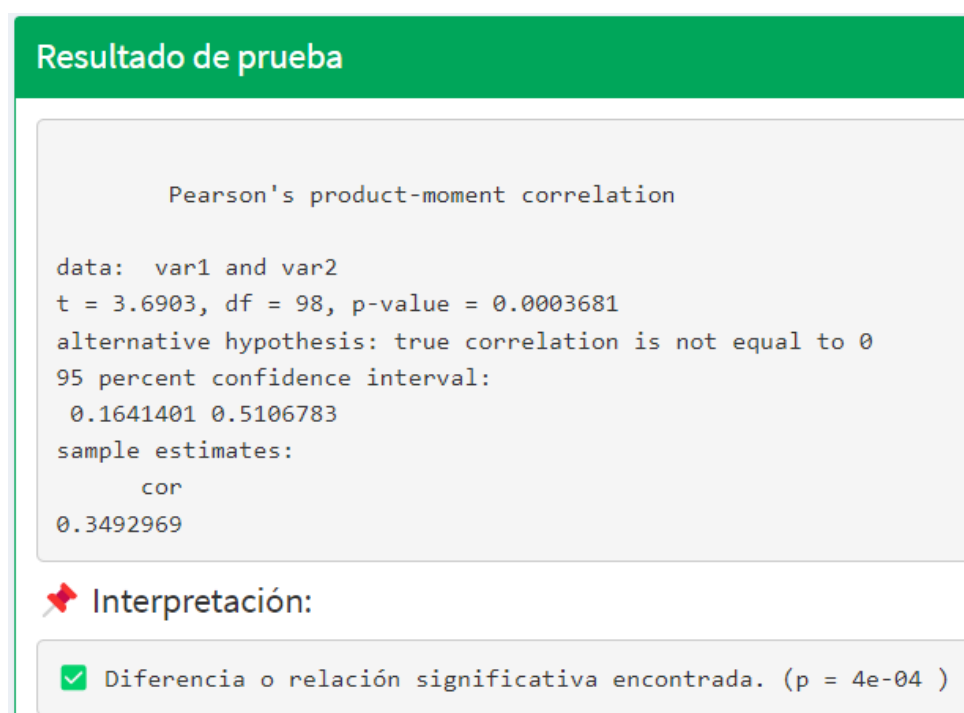


Figura 5: Resultados de pruebas estadísticas con interpretación

## 7.6 Análisis de Normalidad

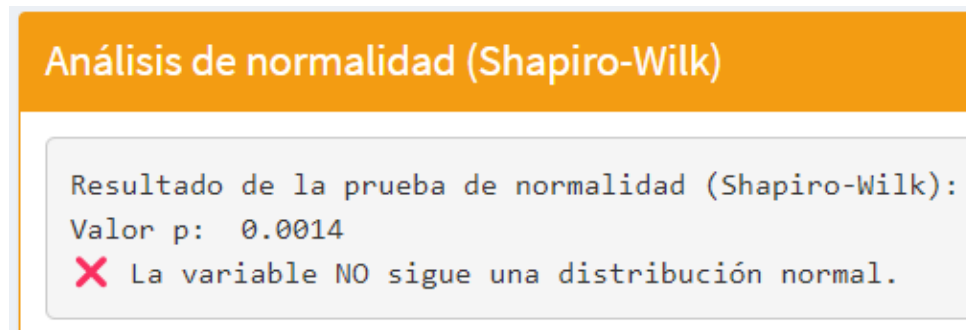


Figura 6: Análisis de normalidad y detección de outliers

## 8 Casos de Uso

### 8.1 Caso de Uso 1: Análisis de Encuesta

**Escenario:** Análisis de una encuesta de satisfacción con variables numéricas (edad, puntuación) y categóricas (género, área).

**Proceso:**

1. Carga del archivo CSV con datos de encuesta
2. Selección de variables relevantes
3. Generación de estadísticas descriptivas
4. Análisis de normalidad de puntuaciones
5. Comparación de satisfacción por género (t-test)
6. Análisis de satisfacción por área (ANOVA)
7. Generación de reporte final

### 8.2 Caso de Uso 2: Análisis de Calidad

**Escenario:** Control de calidad de productos con mediciones de peso, longitud y clasificación.

**Proceso:**

1. Importación de datos de mediciones
2. Análisis descriptivo de cada variable
3. Detección de valores atípicos
4. Análisis de correlación entre medidas
5. Pruebas de normalidad
6. Generación de gráficos de control

## 9 Validación y Pruebas

### 9.1 Pruebas de Funcionalidad

Función	Prueba	Resultado
Carga de datos	Archivos CSV/Excel	Exitosa
Estadísticas descriptivas	Cálculos manuales	Verificado
Gráficos	Comparación visual	Correctos
Pruebas estadísticas	Resultados conocidos	Validado
Exportación PDF	Generación completa	Funcional

Cuadro 3: Resultados de pruebas de funcionalidad

### 9.2 Pruebas de Rendimiento

- **Datasets pequeños (¡1000 registros):** Respuesta instantánea
- **Datasets medianos (1000-10000 registros):** Respuesta en 2-5 segundos
- **Datasets grandes (¿10000 registros):** Respuesta en 5-15 segundos

## 10 Conclusiones

### 10.1 Logros Alcanzados

- Desarrollo exitoso de aplicación web interactiva
- Implementación de análisis estadístico completo
- Interfaz intuitiva y fácil de usar
- Generación automática de reportes profesionales
- Interpretación automática de resultados

### 10.2 Limitaciones

- Tamaño máximo de archivo limitado por memoria
- Algunas pruebas requieren condiciones específicas
- Dependencia de conexión a internet para ciertas funcionalidades

### 10.3 Trabajo Futuro

- Implementación de análisis multivariado
- Integración con bases de datos
- Análisis de series temporales

- Métodos de machine learning
- Interfaz multi-idioma

## 11 Anexos

### 11.1 Código Fuente Completo

El código fuente completo está disponible en el repositorio GitHub:

<https://github.com/CAISA25/APUESTA.git>

### 11.2 Manual de Usuario

El manual detallado de usuario se encuentra en el archivo `Manual_Usuario.pdf` en el repositorio.

### 11.3 Instalación y Configuración

Para instalar y ejecutar la aplicación:

Listing 3: Comandos de instalación

```
# Clonar repositorio
git clone https://github.com/[usuario]/analisis-estadistico-avanzado.git

# Instalar dependencias
Rscript -e "install.packages(c('shiny', 'shinydashboard', 'ggplot2', '
  dplyr', 'plotly', 'knitr', 'rmarkdown'))"

# Ejecutar aplicaci n
Rscript -e "shiny::runApp('app.R')"
```

## 12 Referencias

- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Chang, W. (2021). *R Graphics Cookbook*. O'Reilly Media. Comprehensive R Archive Network (CRAN). (2023). *The R Project for Statistical Computing*. <https://www.r-project.org/>
- Xie, Y., Allaire, J.J., Golemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman Hall/CRC.
- Wickham, H., Golemund, G. (2017). *R for Data Science*. O'Reilly Media.

**Fin del documento**

*Desarrollado por Nelson Catunta Huisa  
Universidad Nacional del Altiplano - Puno*