



UNIVERSIDAD NACIONAL DE FRONTERA

"Año de la recuperación y consolidación de la Economía Peruana"

RESOLUCIÓN DE COMISIÓN ORGANIZADORA

Nº 517-2025-UNF/CO

Sullana, 04 de julio de 2025.

VISTOS:

Informe Nº 100-2025-UNF-VPAC/DGSA-UGA, de fecha 20 de junio del 2025; Oficio Nº 987-2025-UNF-VPAC/DGSA, de fecha 20 de junio de 2025; Oficio Nº 2313-2025-UNF-VPAC, de fecha 26 de junio del 2025; y,

CONSIDERANDO:

Que, el artículo 18º de la Constitución Política del Perú, prescribe que la Universidad es autónoma en su régimen normativo, de gobierno, académico, administrativo y económico: Las Universidades se rigen por sus propios estatutos en el marco de la Constitución y de las leyes.

Que, mediante Ley Nº 29568 del 26 de julio de 2010 se crea la Universidad Nacional de Frontera en el distrito y provincia de Sullana, departamento de Piura, con fines de fomentar el desarrollo sostenible de la Subregión Luciano Castillo Colonna, en armonía con la preservación del medio ambiente y el desarrollo económico sostenible; y, contribuir al crecimiento y desarrollo estratégico de la región fronteriza noroeste del país.

Que, el artículo 8º de la Ley Universitaria, establece que la autonomía inherente a las Universidades se ejerce de conformidad con la Constitución y las Leyes de la República e implica los derechos de aprobar su propio estatuto y gobernarse de acuerdo con él, organizar su sistema académico, económico y administrativo.

Que, mediante Resolución de Comisión Organizadora Nº 461-2021-UNF/CO de fecha 29 de noviembre de 2021, se resuelve aprobar el Estatuto de la Universidad Nacional de Frontera.

Que, en el Estatuto en mención, en su TÍTULO III se establece las DISPOSICIONES TRANSITORIAS, FINALES Y DEROGATORIAS:

A. DISPOSICIONES TRANSITORIAS

PRIMERA. POTESTAD DE LA COMISIÓN ORGANIZADORA

En base al artículo 29 de la Ley Universitaria, la Comisión Organizadora de la UNF tiene a su cargo la aprobación del presente Estatuto, reglamentos y documentos de gestión académica, de investigación y administrativa, formulados en los instrumentos de planeamiento, así como su conducción y dirección hasta que se constituyan los órganos de gobierno que de acuerdo a ley corresponda.

SEGUNDA. PROCESO DE CONSTITUCIÓN

Página | 1



UNIVERSIDAD NACIONAL DE FRONTERA

"Año de la recuperación y consolidación de la Economía Peruana"

RESOLUCIÓN DE COMISIÓN ORGANIZADORA

Durante el proceso de constitución de la Universidad, los artículos del presente Estatuto, que se opongan, contradigan o no puedan implementarse de acuerdo a lo establecido en la normativa de la SUNEDU y MINEDU, respecto a garantizar las condiciones básicas de calidad, quedan en suspenso hasta que se constituyan los órganos de gobierno de la universidad. Encontrándose la Comisión organizadora facultada a emitir resoluciones que permitan el adecuado funcionamiento de la universidad hasta culminar el proceso de constitución.

CUARTA. GOBIERNO DE LA UNF

Durante el proceso de constitución de la Universidad, el gobierno de ésta se ejerce por:

- a) La Comisión Organizadora, tiene atribuciones administrativas que competen a la Asamblea Universitaria, al Consejo Universitario y al Consejo de Facultad.
- b) El Presidente de la Comisión Organizadora de la UNF, tiene atribuciones propias del Rector.
- c) Los Coordinadores de Facultad tiene atribuciones de Decano.

QUINTA. ÓRGANOS DE ALTA DIRECCIÓN

Durante el proceso de constitución de la UNF, los Órganos de Alta Dirección de ésta, lo constituyen:

- a) La Presidencia de Comisión Organizadora, que cumple funciones asignadas al Rectorado.
- b) La Vicepresidencia Académica de Comisión Organizadora, que cumple funciones asignadas al Vicerrectorado Académico.
- c) La Vicepresidencia de Investigación de Comisión Organizadora, que cumple funciones asignadas al Vicerrectorado de Investigación.

Página | 2

Que, en ese sentido el artículo 22° del Estatuto Institucional señala, que es: "Atribución del Consejo Universitario. - f) Concordar y ratificar los planes de estudios y de trabajo propuestos por las unidades académicas".

Que, mediante Resolución de Comisión Organizadora N° 385-2025-UNF/CO, de fecha 16 de mayo de 2025, se **APROBÓ** el Plan de Trabajo del Curso de Capacitación "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON", presentado por la Unidad de Gestión Académica de la Universidad Nacional de Frontera.

Que, mediante Informe N° 100-2025-UNF-VPAC/DGSA-UGA, de fecha 20 de junio del 2025, la Unidad de Gestión Académica comunica a la Dirección de Gestión y Servicios Académicos, que se ha llevado a cabo el Curso de Capacitación "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON", por lo tanto, solicita aprobación del informe final curso en mención.



UNIVERSIDAD NACIONAL DE FRONTERA

"Año de la recuperación y consolidación de la Economía Peruana"

RESOLUCIÓN DE COMISIÓN ORGANIZADORA

Que, con Oficio N° 987-2025-UNF-VPAC/DGSA, de fecha 20 de junio de 2025, la Dirección de Gestión y Servicios Académicos solicita al Vicepresidente Académico, la aprobación del informe final Curso de Capacitación "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON".

Que, con Oficio N° 2313-2025-UNF-VPAC, de fecha 26 de junio del 2025, el Vicepresidente Académico remite al Presidente de la Comisión Organizadora el Informe Final en mención, para efectos de aprobación en Sesión de Comisión Organizadora.

Que, respecto al Artículo IV el Título Preliminar del Texto Único Ordenado de la Ley de Procedimiento Administrativo General, aprobada mediante Decreto Supremo número 004-2019-JUS, recoge como uno de los Principios del Procedimiento Administrativo, el Principio de Legalidad por el cual queda sentado que las autoridades administrativas deben actuar con respeto a la constitución, la ley y al derecho, dentro de las facultades que le estén atribuidas y de acuerdo con los fines para los que les fueron conferidas.

Que, mediante ACTA N° 041-2025-SO-CO, de fecha 03 de julio del 2025, en Sesión Ordinaria de Comisión Organizadora de la Universidad Nacional de Frontera, luego de analizar la documentación presentada y de revisar los informes técnicos y legales indicados en los considerandos de la presente Resolución, por unanimidad se acordó: **APROBAR** el Informe Final del Curso de Capacitación: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON", presentado por la Unidad de Gestión Académica de la Universidad Nacional de Frontera, el mismo que como anexo forma parte integrante de la presente resolución. **AUTORIZAR** la emisión de certificados del Curso de Capacitación: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON", en mérito a lo aprobado en los artículos precedentes.

Página | 3

Estando a lo expuesto y en uso de las atribuciones conferidas por la Ley Universitaria – Ley N° 30220 y por la Resolución Viceministerial N° 045-2023-MINEDU y Acta de Acuerdos de Sesión Ordinaria de Comisión Organizadora N° 041-2025-SO-CO de fecha 03 de julio del 2025.

SE RESUELVE:

ARTÍCULO PRIMERO. - APROBAR el Informe Final del Curso de Capacitación: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON", presentado por la Unidad de Gestión Académica de la Universidad Nacional de Frontera, el mismo que como anexo forma parte integrante de la presente resolución.



UNIVERSIDAD NACIONAL DE FRONTERA

"Año de la recuperación y consolidación de la Economía Peruana"

RESOLUCIÓN DE COMISIÓN ORGANIZADORA

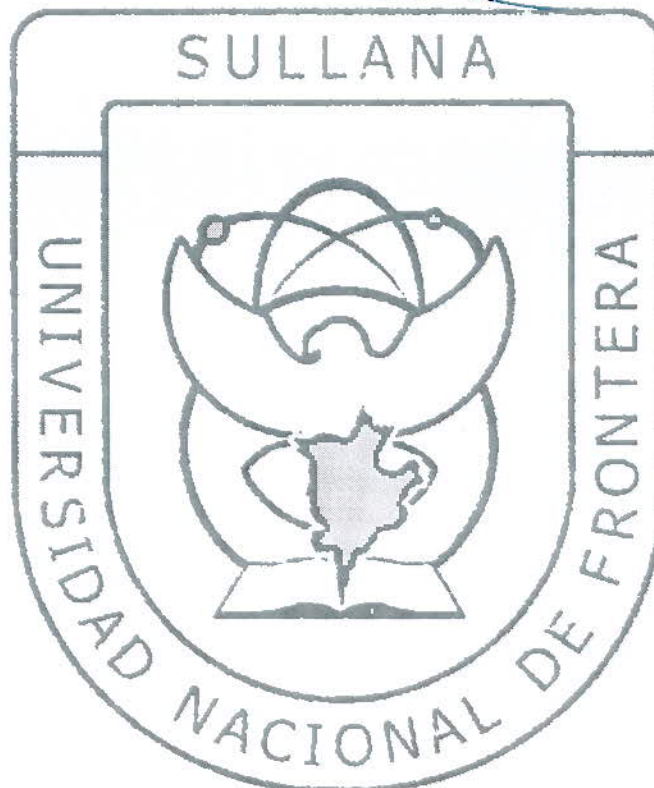
ARTÍCULO SEGUNDO. - AUTORIZAR la emisión de certificados del Curso de Capacitación: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON", en mérito a lo aprobado en los artículos precedentes.

ARTÍCULO TERCERO. - NOTIFICAR a través, de los mecanismos más adecuados y pertinentes, para conocimiento y fines correspondientes.

REGÍSTRESE, COMUNÍQUESE Y EJECÚTESE.

UNIVERSIDAD NACIONAL DE FRONTERA
Dr. José Valentín Molero López
PRESIDENTE DE LA COMISIÓN ORGANIZADORA

UNIVERSIDAD NACIONAL DE FRONTERA
Abd. Jorge Román Galloso Torres
SECRETARIO GENERAL



"Año de la Recuperación y Consolidación de la Economía Peruana"



**UNIVERSIDAD
NACIONAL DE
FRONTERA**

INFORME FINAL

Curso de Capacitación:

**"Estadística Aplicada a la
Investigación Científica con
Ayuda de Software Programación
de "R" y "PYTHON"**

2025

Sullana – Piura

I. INTRODUCCIÓN

El Curso " Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON" se llevó a cabo con éxito y cumplió sus objetivos de capacitar a los docentes de la Universidad Nacional de Frontera, el presente Informe del curso de capacitación "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON" para los docentes de la Universidad Nacional de Frontera , está diseñado para proporcionar a los participantes los conocimientos teóricos y prácticos necesarios para aplicar técnicas estadísticas avanzadas utilizando el lenguaje de programación R, uno de los entornos más poderosos y versátiles para el análisis de datos.

La estadística es una disciplina universalmente aplicada para describir clasificar, inferir, decidir y cuantificar la probabilidad de error en la toma de decisiones científicas. Su aplicación en estudios cuantitativos y cualitativos ya está consolidada, sin embargo, su utilidad en áreas sociales y específicamente en investigación educativa se ha visto impulsada recientemente por la difusión del software open source y el crecimiento exponencial de la ciencia de datos. Con el objetivo de estudiar la contribución diferencial de las técnicas estadísticas multivariadas y del lenguaje R, se han seleccionado los trabajos de investigación educativa del nivel superior de un período reciente. Se analizaron las técnicas multivariadas empleadas en las mencionadas investigaciones y se las comparó con trabajos que analizaban contextos muy similares, pero que no aplicaban estas técnicas. A partir de ello se desprende la evolución de la aplicación de estas técnicas y del lenguaje R en las investigaciones de corte educativo.

El afán desmedido por usar métodos estadísticos para el análisis de datos ha llevado a percibir que la estadística es una "herramienta", o que los estadísticos son técnicos a quienes recurrir para solucionar problemas que se aprecian simplemente como "operativos" (aplicar una determinada prueba estadística, sacar el valor-p, calcular la potencia de la prueba, correr el modelo, cuántas observaciones tomar, etcétera), la mayor parte de las veces sin haberlos vinculado al proceso de formulación, diseño o discusión del estudio. Esto ha llevado a un uso de la estadística sin suficiente reconocimiento de su estatus como disciplina científica independiente, lo que ha devenido en una instrumentalización simple y superficial, que impide el aprovechamiento completo de sus potencialidades, pero que además puede afectar el rigor y validez de los estudios.

La estadística proporciona a los investigadores una alternativa al censo, al indagar solo una fracción de la población con el ánimo de emitir conclusiones a nivel general. Lo anterior resulta cómodo y atractivo para los usuarios, pero es necesario planear cuidadosamente ciertas etapas para disfrutar de un conjunto de resultados que sean relevantes y poder extrapolar las conclusiones a la población. La estadística y el pensamiento estadístico son factores principales en la vida cotidiana y en muchas de las ocupaciones. Casi cada profesión tiene que ver con datos (cuantitativos y cualitativos), y, por consiguiente, una necesidad de investigadores y ciudadanos instruidos estadísticamente que puedan contribuir a generar un razonamiento crítico, fundamental para una sociedad bien informada.

La investigación educativa tiene como fin fundamental, la comprensión e interpretación de los fenómenos que ocurren dentro del ámbito educativo. Asimismo, pretende brindar aportes con el propósito de mejorar las acciones, intervenciones y resolver problemas de ese campo de estudio. Este tipo de investigaciones se caracteriza por proveer una gran variedad de enfoques y metodologías de acuerdo con la complejidad del objeto de estudio. Según la naturaleza de los datos, las investigaciones trabajan con enfoques cuantitativos, cualitativos o mixtos.

La investigación educativa también se vio atravesada por los avances de la estadística y por el florecimiento de la ciencia de datos y hasta por la evolución del aprendizaje automático. En los años recientes la incorporación de estos procedimientos y estrategias ha delineado un cambio en el escenario cultural de la investigación educativa señalando una tendencia hacia visiones integradoras y comprensivas en los contextos educativos.

En la actualidad, se ha ampliado notablemente la disponibilidad de herramientas digitales, facilitando los procesos de relevamiento y almacenamiento de grandes cantidades de datos provenientes de la administración, de la gestión, de la experimentación o de la observación. Simultáneamente, el acceso a una diversidad de lenguajes, plataformas y software estadísticos favorece un análisis más integrador, preciso y efectivo. El uso de software ahorra tiempo, dada su flexibilidad en el trabajo con los datos, facilita las tareas rutinarias, agiliza la gestión y el análisis de datos al mismo tiempo que colabora en el tratamiento conjunto de una multiplicidad de variables vinculadas con un mismo proceso. Sin embargo, estas nuevas facilidades disponibles conllevan en sí mismas nuevas dificultades.



II. PARTICIPANTES, ORGANIZADORES, EJECUTORES Y RESOLUCIÓN DE APROBACIÓN DEL CURSO DE CAPACITACIÓN.

El desarrollo del Curso de Capacitación denominado: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON", estuvo dirigido a los docentes de la Universidad Nacional de Frontera, el cual fue aprobado mediante Resolución de Comisión Organizadora N°385-2025-UNF/CO.

2.1 ORGANIZADORES:

- Mg. Daniel Francisco Castro Navarrete
- Mg. Zoila Beatriz Zeta Eche
- Mg. Greysi Carolina Calva Moscoso
- Ing. Lidman David Gálvez Paucar
- Ing. Pedro Miguel Mauricio Mera
- Lic. Mariela Giovanna Murguía Panta



2.2 PÚBLICO OBJETIVO:

Docentes de la Universidad Nacional de Frontera, orientados a adquirir o actualizar sus conocimientos y competencias en el uso de la Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON".

2.3 OBJETIVO GENERAL:

El objetivo general del curso de capacitación conforme al plan fue: Proporcionar a los participantes las herramientas teóricas y prácticas necesarias para aplicar métodos estadísticos en la investigación científica, utilizando software especializado como R y Python. Esto incluye el desarrollo de habilidades para el manejo de datos, análisis estadístico, visualización de resultados y la interpretación crítica de los mismos, con el fin de apoyar la toma de decisiones basada en evidencia en diversos campos de estudio.

III. Desarrollo del Curso:

3.1 Denominación: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON"

El curso se dividió en cinco sesiones: los días martes se desarrollaron de manera presencial y los días sábados y domingos se realizaron actividades de manera asincrónica.

IV. CRONOGRAMA:

Cuadro 1: Cronograma de ejecución del curso de capacitación denominado "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON".

Curso de Capacitación "Curso: Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON"			Mes de ejecución – Mayo - Junio		
			Día 1	Día 2	Día 3
Módulo 1: Fundamentos de Estadística y su Rol en la Investigación Científica	Módulo 2: Programación y Manejo de Datos en R y Python	Módulo 3: Análisis Descriptivo y Visualización de Datos	20/05	24/05	27/05
Tema 1: martes (presencial - 4 horas) ✓ Conceptos básicos de estadística: Tipos de datos, medidas de tendencia central y dispersión. ✓ Importancia en la investigación: Diseño experimental, muestreo y recolección de datos. ✓ Introducción a R y Python: Breve presentación de ambos entornos como herramientas para el análisis estadístico.	Tema 2: martes (presencial - 4 horas) ✓ Fundamentos de programación: Sintaxis básica en R y Python. ✓ Importación y limpieza de datos: Uso de paquetes como tidyverse en R y pandas en Python. ✓ Transformación y manipulación de datos: Preparación de datasets para el análisis.	Martes (presencial - 4 horas) ✓ Estadística descriptiva: Elaboración de tablas, cálculos de medias, medianas, desviaciones estándar, etc. ✓ Visualización de datos: Creación de gráficos y diagramas usando ggplot2 en R y matplotlib/seaborn en Python. ✓ Interpretación exploratoria: Identificación de patrones y tendencias en los datos			
Módulo 4: Análisis Inferencial y Modelos Estadísticos		Módulo 5: Aplicaciones Avanzadas y Proyectos de Investigación	31/05	03/06	07/06
Martes (presencial - 4 horas) ✓ Pruebas de hipótesis: Conceptos y aplicación de tests estadísticos (t-test, chi-cuadrado, ANOVA). ✓ Regresión y correlación: Análisis de regresión lineal, modelos predictivos básicos y estudios de correlación. ✓ Implementación práctica: Ejecución de análisis inferenciales en R y Python, comparando resultados y enfoques.		Martes (presencial - 4 horas) ✓ Análisis multivariado: Introducción a técnicas como PCA, cluster analysis y análisis discriminante. ✓ Machine Learning básico: Uso de modelos predictivos y algoritmos de clasificación en el contexto de la investigación. ✓ Integración y comunicación de resultados: Desarrollo de proyectos aplicados y presentación de hallazgos para la toma de decisiones en investigaciones científicas.			

V. MÓDULOS DE CAPACITACIÓN DESARROLLADOS

5.1 MODULO N° 01: FUNDAMENTOS DE ESTADÍSTICA Y SU ROL EN LA INVESTIGACIÓN CIENTÍFICA.

5.1.1 Conceptos básicos de estadística: Tipos de datos, medidas de tendencia central y dispersión.

Hablar hoy en día de la definición de la estadística es una disciplina que se encarga de recopilar, organizar, analizar e interpretar datos para tomar decisiones o extraer conclusiones. A continuación, veremos los conceptos fundamentales:

5.1.1.1 Tipos de Datos

Los datos pueden clasificarse en dos grandes categorías:

- a) **Datos Cualitativos** (Nominales u Ordinales)
 - ✓ Nominales: No tienen un orden natural.
 - ✓ Ordinales: Tienen un orden o jerarquía.
- b) **Datos Cuantitativos** (Discretos o Continuos)
 - ✓ Discretos: Toman valores enteros.
 - ✓ Continuos: Pueden tomar cualquier valor dentro de un rango.



5.1.1.2 Medidas de tendencia central y dispersión

A. Medidas de tendencia central:

Las medidas de tendencia central describen el centro de un conjunto de datos, mientras que las medidas de dispersión indican cómo se dispersan los datos alrededor de ese centro.

Medidas de Tendencia Central:

- a) **Media:** Es el promedio aritmético de los datos, sumando todos los valores y dividiendo por el número total de valores.
- b) **Mediana:** Es el valor central de los datos cuando están ordenados de menor a mayor. Si hay un número par de datos, es el promedio de los dos valores centrales.
- c) **Moda:** Es el valor que aparece con mayor frecuencia en los datos.

B. Medidas de tendencia Dispersión:

Las medidas de dispersión indican qué tanto se dispersan o agrupan los datos con respecto a su media aritmética. El rango es una medida de dispersión que, para una serie de datos no agrupados, es igual a la diferencia del dato mayor y el dato menor.

- a) **Rango:** Es la diferencia entre el valor máximo y el valor mínimo de los datos.
- b) **Desviación Estándar:** Mide la dispersión de los datos alrededor de la media, indicando cuánto se alejan los valores individuales de la media.

- c) Varianza: Es el cuadrado de la desviación estándar y también mide la dispersión alrededor de la media.

5.1.1.3 Importancia en la investigación: diseño experimental, muestreo y recolección de datos.

La investigación científica requiere un enfoque sistemático y riguroso para garantizar la validez y confiabilidad de sus resultados. Tres aspectos fundamentales en este proceso son:

A. Diseño Experimental

El diseño experimental es la estructura que organiza cómo se llevará a cabo la investigación. Su importancia radica en:

- ✓ Control de variables: Permite aislar el efecto de la variable independiente sobre la dependiente, minimizando el sesgo.
- ✓ Validez interna y externa: Un buen diseño asegura que los resultados sean aplicables a otras situaciones (validez externa) y que no estén distorsionados por factores externos (validez interna).
- ✓ Tipos de diseños: Experimentales (aleatorización y control), cuasiexperimentales (sin aleatorización) y no experimentales (observacionales).



B. Muestreo

El muestreo es el proceso de seleccionar una muestra representativa de la población de estudio. Su importancia incluye:

- ✓ Reducción de costos y tiempo: Estudiar toda una población puede ser inviable; una muestra bien seleccionada permite generalizar resultados.
- ✓ Precisión estadística: Un muestreo adecuado (aleatorio simple, estratificado, por conglomerados) minimiza errores y sesgos.
- ✓ Representatividad: La muestra debe reflejar las características de la población para que las conclusiones sean válidas.

C. Recolección de Datos

Esta etapa consiste en obtener información relevante para el estudio. Su correcta ejecución es crucial porque:

- ✓ Determina la calidad de los resultados: Datos mal recolectados llevan a conclusiones erróneas.
- ✓ Métodos variados: Encuestas, entrevistas, observación, experimentos o fuentes secundarias. Cada método debe ser válido y confiable.
- ✓ Estandarización: Los instrumentos (como cuestionarios) deben ser claros y consistentes para evitar sesgos.

5.1.1.4 Introducción a R y Python: breve presentación de ambos entornos como herramientas para el análisis estadístico.

Tanto **R** como **Python** son lenguajes de programación ampliamente utilizados en el análisis de datos, la estadística y el machine learning. Cada uno tiene sus fortalezas y

se adapta a diferentes necesidades en el ámbito estadístico y científico.

A. R: Un Entorno Especializado en Estadística

Características Principales:

- Diseñado para estadísticos: Surgió como un lenguaje específico para análisis estadístico y visualización de datos.
- Ecosistema robusto: Cuenta con paquetes especializados (ggplot2, dplyr, tidyr, shiny) para manipulación, visualización y modelado avanzado.
- Excelente para visualización: Librerías como ggplot2 permiten crear gráficos complejos y personalizados con pocas líneas de código.
- Comunidad académica fuerte: Ampliamente adoptado en investigación y ciencias sociales.

Ventajas:

- Ideal para análisis estadísticos tradicionales (regresiones, tests de hipótesis, ANOVA).
- Gran cantidad de paquetes en CRAN (más de 18,000 disponibles).
- Entorno integrado (RStudio) muy amigable para usuarios no programadores.



B. Python: Un Lenguaje Generalista con Potencial Estadístico.

Características Principales:

- Lenguaje de propósito general: No fue creado específicamente para estadística, pero su flexibilidad lo hace poderoso en ciencia de datos.
- Librerías clave: pandas (manipulación de datos), numpy (cálculo numérico), matplotlib/seaborn (visualización), scikit-learn (machine learning).
- Integración con otras áreas: Desarrollo web, automatización, IA, etc.

Ventajas:

- Más versátil para proyectos que combinan análisis de datos con otras tareas.
- Sintaxis más limpia y fácil de aprender para principiantes.
- Dominante en machine learning y big data.



Inicio de la Sesión N° 01. Fundamentos de Estadística y su Rol en la Investigación Científica del Curso: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON", del 20 de mayo del 2025.



5.2 MODULO N° 02: PROGRAMACIÓN Y MANEJO DE DATOS EN R Y PYTHON

5.2.1 FUNDAMENTOS DE PROGRAMACIÓN: SINTAXIS BÁSICA EN R Y PYTHON

La sintaxis básica en R y Python se refiere a las reglas que definen cómo escribir código legible y que el intérprete pueda comprender. En Python, se caracteriza por ser sencilla e intuitiva, utilizando indentación para delimitar bloques de código, mientras que R utiliza operadores como `<-` para asignar valores a variables.

A. Sintaxis básica en Python:

❖ Variables:

Se asignan valores a variables sin necesidad de declarar previamente el tipo de dato. Por ejemplo,
`x = 5, nombre = "Juan".`

❖ Tipos de datos:

Python soporta varios tipos de datos como enteros, flotantes, cadenas, booleanos, listas, tuplas y diccionarios.

❖ Estructuras de control:

Se utilizan `if`, `elif` y `else` para condicionales, y `for` y `while` para bucles.

❖ **Indentación:**

La indentación es crucial, definiendo los bloques de código dentro de estructuras de control.

B. Sintaxis básica en R:

- ❖ **Variables:** Se utilizan el operador `<-` o `=` para asignar valores a variables. Por ejemplo, `x <- 5`, `nombre <- "Juan"`.
- ❖ **Tipos de datos:** R maneja tipos como `numeric`, `character`, `factor` y `logical`.
- ❖ **Estructuras de control:** Se utilizan `if`, `else`, `for` y `while` para condicionales y bucles.
- ❖ **Funciones:** Se definen con la palabra clave `function()`, seguida de los parámetros y el código dentro de la función.
- ❖ **Vectorización:** R es conocido por su vectorización, lo que permite realizar operaciones en vectores enteros sin necesidad de bucles explícitos.

5.2.2 IMPORTACIÓN Y LIMPIEZA DE DATOS: USO DE PAQUETES COMO TIDYVERSE EN R Y PANDAS EN PYTHON.



La importación y limpieza de datos son pasos fundamentales en el análisis de datos. Tanto R como Python ofrecen poderosas herramientas para estas tareas: el paquete `tidyverse` en R y la biblioteca `panda` en Python.

En R (Tidyverse):

A. Importar datos:

- ❖ Usa `read_csv()` para importar datos CSV, `read_excel()` para archivos Excel, y otras funciones similares para otros formatos.
- ❖ Define los tipos de datos de las columnas utilizando el argumento `col_types` para evitar problemas de conversión.
- ❖ Utiliza `read_delim()` para otros delimitadores, como tabuladores.

B. Limpiar datos:

- ❖ Usa `tidyr` para organizar y estructurar los datos.
- ❖ Utiliza funciones como `separate()` para dividir columnas, `unite()` para combinar, `complete()` para manejar valores faltantes, y `fill()` para interpolar valores.
- ❖ `dplyr` te permite filtrar, agregar, y transformar datos con funciones como `filter()`, `mutate()`, y `summarise()`.
- ❖ `stringr` facilita el manejo de cadenas de texto.

En Python (Pandas):

A. Importar datos:

- ❖ `read_csv()` para CSV, `read_excel()` para Excel, y otras funciones para formatos de archivo.
- ❖ Usa el argumento `dtype` para definir el tipo de datos de las columnas al importar.
- ❖ `pd.read_html()` puede ser útil para importar tablas de HTML.

B. Limpiar datos:

- ❖ **Manejo de valores faltantes:** Utiliza `dropna()` para eliminar filas o columnas con valores faltantes, o `fillna()` para rellenarlos con valores específicos.
- ❖ **Tratamiento de tipos de datos:** Con `astype()` puedes cambiar los tipos de datos de las columnas.
- ❖ **Eliminar duplicados:** Usa `drop_duplicates()` para eliminar filas duplicadas.
- ❖ **Eliminar o renombrar columnas:** Usa `drop()` para eliminar columnas y `rename()` para renombrarlas.
- ❖ **Manipulación de datos:** Con `groupby()`, `apply()`, y funciones similares puedes realizar operaciones complejas.
- ❖ **Manejo de datos desordenados:** Con `tidyr` en R, puedes usar `separate()` y `unite()` para organizar tus datos. En Pandas, puedes usar `str.split()` y `str.join()` para tareas similares.



5.2.3 TRANSFORMACIÓN Y MANIPULACIÓN DE DATOS: PREPARACIÓN DE DATASETS PARA EL ANÁLISIS.

La preparación de datasets es una etapa crucial en cualquier proyecto de análisis de datos o ciencia de datos, donde se transforman y manipulan los datos brutos para hacerlos adecuados para el análisis. Este proceso puede consumir hasta el 80% del tiempo en un proyecto de datos.

A. Etapas clave en la preparación de datasets

a) Adquisición y Carga de Datos:

- ❖ Importación desde diversas fuentes (CSV, Excel, bases de datos, APIs)
- ❖ Verificación de la carga correcta
- ❖ Exploración inicial para entender la estructura

b) Limpieza de Datos:

- ❖ Manejo de valores faltantes (NA/null)
- ❖ Identificación y tratamiento de outliers
- ❖ Corrección de errores e inconsistencias
- ❖ Normalización de formatos (fechas, textos, etc.)

c) Transformación de Datos:

- ❖ Normalización y estandarización de variables numéricas.
- ❖ Codificación de variables categóricas (one-hot encoding, label encoding)
- ❖ Ingeniería de características (feature engineering).
- ❖ Discretización de variables continuas cuando sea necesario

d) Reducción y Selección de Datos:

- ❖ Eliminación de variables redundantes o irrelevantes
- ❖ Técnicas de reducción de dimensionalidad (PCA, t-SNE)
- ❖ Muestreo para conjuntos de datos muy grandes

e) Partición de Datos:

- ❖ División en conjuntos de entrenamiento, validación y prueba.
- ❖ Estratificación para mantener distribuciones en problemas de clasificación



Inicio de la Sesión N° 02. Programación y Manejo de Datos en R y Python del Curso: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON", del 24 de mayo del 2025.

5.3 MÓDULO 3: ANÁLISIS DESCRIPTIVO Y VISUALIZACIÓN DE DATOS

5.3.1 ESTADÍSTICA DESCRIPTIVA: ELABORACIÓN DE TABLAS, CÁLCULOS DE MEDIAS, MEDIANAS, DESVIACIONES ESTÁNDAR, ETC.

La estadística descriptiva es una rama de la estadística que se encarga de resumir, organizar y presentar datos de manera significativa. A continuación, te explico los principales conceptos y procedimientos para elaborar tablas, calcular medidas de tendencia central (media, mediana) y de dispersión (desviación estándar, varianza, etc.).

A. Elaboración de Tablas:

❖ **Tablas de Frecuencia:**

Organizan datos en categorías o intervalos, mostrando cuántas veces aparece cada valor o intervalo, según GeeksforGeeks.

❖ **Tablas Cruzadas (o de Contingencia):**

Muestran la relación entre dos o más variables, mostrando la frecuencia conjunta de las categorías de cada variable.

❖ **Tablas Resumen:**

Presentan medidas estadísticas clave (media, mediana, desviación estándar, etc.) de manera concisa.

B. Cálculos de Medidas Descriptivas:

❖ **Medidas de Tendencia Central:**

- **Media:** Promedio de los datos.
- **Mediana:** Valor central cuando los datos están ordenados.
- **Moda:** Valor que más se repite.

❖ **Medidas de Dispersión:**

- **Rango:** Diferencia entre el valor máximo y mínimo.
- **Desviación Estándar:** Medida de la dispersión de los datos alrededor de la media.
- **Varianza:** Medida del cuadrado de la desviación estándar.

❖ **Percentiles:**

Dividen los datos en 100 grupos iguales, indicando el valor por debajo del cual se encuentra un porcentaje específico de los datos.

❖ **Cuartiles:**

Dividen los datos en 4 grupos iguales (percentiles 25, 50 y 75).

C. Representaciones Gráficas:

- ❖ **Histogramas:** Muestran la distribución de frecuencias de los datos en intervalos.
- ❖ **Diagramas de Barras:** Comparan valores de diferentes categorías.
- ❖ **Diagramas Circulares (o de Pastel):** Representan proporciones o porcentajes de un total.
- ❖ **Diagramas de Dispersión:** Muestran la relación entre dos variables.
- ❖ **Gráficos de Cajas (Boxplots):** Resumen la distribución de datos, mostrando la mediana, cuartiles y valores atípicos.

D. Interpretación y Aplicación:

La estadística descriptiva permite:

- ❖ **Describir las características de una muestra de datos:** Por ejemplo, la edad promedio de un grupo de personas, según QuestionPro.



- ❖ **Comparar diferentes grupos o muestras:** Por ejemplo, comparar las notas promedio de dos clases.
- ❖ **Identificar patrones y tendencias en los datos:** Por ejemplo, ver cómo cambia la demanda de un producto a lo largo del tiempo.
- ❖ **Tomar decisiones basadas en la información:** Por ejemplo, decidir qué tipo de marketing utilizar para un producto.

5.3.2 VISUALIZACIÓN DE DATOS: CREACIÓN DE GRÁFICOS Y DIAGRAMAS USANDO GGLOT2 EN R Y MATPLOTLIB/SEABORN EN PYTHON.

Para la creación de gráficos y diagramas en R y Python, se utilizan las bibliotecas ggplot2 en R y Matplotlib y Seaborn en Python. ggplot2 se basa en la "Gramática de los Gráficos" y crea gráficos complejos a través de capas superpuestas, mientras que Matplotlib ofrece una interfaz similar a MATLAB y Seaborn proporciona una interfaz de alto nivel para gráficos estadísticos.

ggplot2 en R:

- ✓ **Concepto principal:**

La "Gramática de los Gráficos", que define cómo construir gráficos superponiendo capas (datos, estética, geometría, facetas, etc.).



- ✓ **Sintaxis:**

Se declara lo que debe contener el gráfico y se superponen capas para construirlo. Por ejemplo: `ggplot (data, aes (x = variable, y = variable_y)) + geom_point() + geom_smooth()`.

- ✓ **Ventajas:**

Sintaxis declarativa y estructurada, facilitando la creación de gráficos complejos.

Matplotlib en Python:

- ✓ **Concepto principal:**

Módulo pyplot que proporciona una interfaz similar a MATLAB para crear gráficos.

- ✓ **Sintaxis:**

Se utilizan funciones para añadir elementos (líneas, texto, imágenes) a los ejes del gráfico. Por ejemplo: `plt.plot (x, y) + plt.xlabel("X") + plt.ylabel("Y")`.

- ✓ **Ventajas:**

Alto grado de personalización, control individual sobre cada elemento del gráfico.

Seaborn en Python:

- ✓ **Concepto principal:**

Construida sobre Matplotlib, proporciona una interfaz de alto nivel para gráficos estadísticos atractivos e informativos.

- ✓ **Sintaxis:**

Utiliza funciones que manejan aspectos estadísticos y visualización de datos de manera más sencilla que Matplotlib. Por ejemplo: `sns.scatterplot(x=x, y=y, hue=hue_variable, data=data)`.

✓ **Ventajas:**

Crea gráficos estadísticos complejos con menos código, estilos y paletas de colores atractivas.

5.3.3 INTERPRETACIÓN EXPLORATORIA: IDENTIFICACIÓN DE PATRONES Y TENDENCIAS EN LOS DATOS

La interpretación exploratoria de datos es una fase crucial en el análisis de datos, donde el objetivo principal es identificar patrones, tendencias, relaciones o anomalías en los datos antes de aplicar técnicas estadísticas más avanzadas. A continuación, te explico cómo abordar este proceso de manera efectiva:

A. Objetivos de la Interpretación Exploratoria

- **Entender la estructura** de los datos (distribuciones, valores atípicos, missing values).
- **Detectar patrones** (correlaciones, agrupamientos, ciclos).
- **Generar hipótesis** para análisis posteriores.
- **Validar la calidad** de los datos (errores, inconsistencias).



B. Técnicas Clave para Identificar Patrones y Tendencias

- **Análisis Univariado**
- **Análisis Bivariado/Multivariado**
- **Serie Temporales**

C. Herramientas Visuales Clave

- **Gráficos básicos:**
- **Gráficos avanzados:**
- **Herramientas interactivas:**

D. Identificación de Anomalías (Outliers)

- **Métodos:**
 - ✓ Boxplots (datos fuera de los "bigotes").
 - ✓ Puntajes Z (valores con $|Z| > 3$).
 - ✓ Algoritmos como Isolation Forest o DBSCAN.
- **Acciones:**
 - ✓ Eliminar, transformar o investigar causas.



Inicio de la Sesión N° 03. Análisis descriptivo y visualización de datos: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON", del 27 de mayo del 2025.

5.4 MODULO N°04 ANÁLISIS INFERENCIAL Y MODELOS ESTADÍSTICOS

5.4.1 PRUEBAS DE HIPÓTESIS: CONCEPTOS Y APLICACIÓN DE TESTS ESTADÍSTICOS (T-TEST, CHI- CUADRADO, ANOVA)

Las pruebas de hipótesis son herramientas estadísticas fundamentales para tomar decisiones basadas en datos. Permiten evaluar afirmaciones sobre parámetros poblacionales (como medias, proporciones o varianzas) utilizando información muestral. A continuación, se explican los conceptos clave y las pruebas más utilizadas:

A. Conceptos Básicos:

- **Hipótesis Nula (H_0/H_0):** Afirmación que se pone a prueba (ej: "no hay diferencia").
- **Hipótesis Alternativa (H_1/H_1 o H_a/H_a):** Afirmación contraria a H_0/H_0 (ej: "existe diferencia").
- **Nivel de significancia (α):** Probabilidad de rechazar H_0/H_0 cuando es verdadera (usualmente 0.05 o 5%).
- **Valor-p (p -value):** Probabilidad de obtener un resultado igual o más extremo que el observado, asumiendo H_0/H_0 cierta. Si $p < \alpha$, se rechaza H_0/H_0 .
- **Error Tipo I:** Rechazar H_0/H_0 cuando es verdadera.
- **Error Tipo II:** No rechazar H_0/H_0 cuando es falsa.

B. t-test (Prueba t)

- **Aplicaciones:**

- ✓ ¿Las notas de dos clases son diferentes?
- ✓ ¿Un fármaco reduce la presión arterial?

- **Tipos:**

- ✓ t-test para una muestra: Compara la media muestral con un valor teórico.
 - Ej: ¿La media de altura de un grupo es 170 cm?
- ✓ t-test independiente: Compara medias de dos grupos distintos.
 - Ej: Notas de estudiantes de dos universidades.
- ✓ t-test pareado: Compara medias del mismo grupo en dos momentos.
 - Ej: Peso antes y después de una dieta.

- **Supuestos:**

- ✓ Datos cuantitativos.
- ✓ Distribución normal (o tamaño muestral grande).
- ✓ Varianzas similares (en t-test independiente).

C. Chi-cuadrado (χ^2)



- **Aplicaciones:**

- ✓ ¿El género está asociado a la preferencia por un producto?
- ✓ ¿La distribución de grupos sanguíneos sigue una proporción conocida?

- **Tipos:**

- ✓ Prueba de bondad de ajuste: Compara frecuencias observadas con esperadas.
 - Ej: ¿La proporción de colores de M&M's coincide con lo anunciado?
- ✓ Prueba de independencia: Evalúa si dos variables categóricas están relacionadas.
 - Ej: ¿Fumar está asociado con cáncer de pulmón?

- **Supuestos:**

- ✓ Datos categóricos.
- ✓ Frecuencias esperadas ≥ 5 (para evitar sesgos).

D. ANOVA (Análisis de Varianza)

- **Aplicaciones:**

- ✓ ¿El rendimiento de estudiantes varía según el método de enseñanza (A, B, C)?
- ✓ ¿El crecimiento de plantas difiere con distintos fertilizantes?

- **Tipos:**

- ✓ ANOVA de un factor: Compara grupos basados en una sola variable categórica.
- ✓ ANOVA factorial: Evalúa múltiples variables categóricas y sus interacciones.

- **Supuestos:**

- ✓ Normalidad de los residuos.
- ✓ Homogeneidad de varianzas (prueba de Levene).
- ✓ Independencia de las observaciones.
- ✓ Post-hoc: Si el ANOVA es significativo, se usan pruebas como Tukey o Bonferroni para identificar qué grupos difieren.

5.4.2 REGRESIÓN Y CORRELACIÓN: ANÁLISIS DE REGRESIÓN LINEAL, MODELOS PREDICTIVOS BÁSICOS Y ESTUDIOS DE CORRELACIÓN.

La regresión y la correlación son técnicas estadísticas fundamentales para analizar la relación entre variables. Mientras que la correlación mide la fuerza y dirección de una asociación, la regresión permite modelar relaciones predictivas entre variables.

A. Regresión Lineal:

- ✓ **Análisis de Regresión Lineal:** Es una técnica que permite modelar la relación entre variables y predecir valores.
- ✓ **Modelo de Regresión Lineal Simple:** Se basa en la ecuación $y = a + bx$, donde 'y' es la variable dependiente, 'x' es la independiente, 'a' es la intersección y 'b' es la pendiente de la recta.
- ✓ **Modelo de Regresión Lineal Múltiple:** Extiende el modelo simple a varias variables explicativas.
- ✓ **Modelos Predictivos:** La regresión lineal proporciona un modelo matemático que permite predecir el valor de una variable en función de otras.



B. Correlación:

- ✓ **Medida Estadística:** Cuantifica la fuerza y dirección de la relación lineal entre dos variables, expresada como un coeficiente entre -1 y +1.
- ✓ **No Implica Causalidad:** Una alta correlación no implica que una variable cause la otra.
- ✓ **Estudios de Correlación:** Evalúan si dos variables tienden a variar juntas y en qué grado.

5.4.3 IMPLEMENTACIÓN PRÁCTICA: EJECUCIÓN DE ANÁLISIS INFERENCIALES EN R Y PYTHON, COMPARANDO RESULTADOS Y ENFOQUES.

R y Python son dos de los lenguajes más populares para análisis estadísticos y ciencia de datos. Ambos pueden realizar análisis inferenciales, pero con enfoques y sintaxis diferentes. A continuación, compararé cómo realizar varios análisis inferenciales comunes en ambos lenguajes.

Sin embargo, Python y R ofrecen herramientas que permiten hacer análisis eficiente de datos y desarrollar algoritmos, lo que contribuye a la innovación educativa en el campo de la enseñanza de las Tecnologías.

La elección entre R y Python depende del contexto:

- ❖ Para investigación estadística pura o informes analíticos, R suele ser más eficiente.
- ❖ Para proyectos que integran análisis inferencial con otros componentes (web apps, pipelines de datos, IA), Python puede ser más conveniente.

Ambos lenguajes son capaces de producir resultados equivalentes, aunque pueden diferir ligeramente en implementaciones numéricas de algunos algoritmos:

A. PRUEBA T DE STUDENT (MUESTRAS INDEPENDIENTES):

La prueba t para muestras independientes (o prueba t no pareada) es una prueba estadística que determina si existe una diferencia entre dos grupos no relacionados.

La prueba t para muestras independientes se utiliza para hacer una afirmación sobre la población basándose en dos muestras independientes. Para hacer esta afirmación se compara el valor medio de las dos muestras. Si la diferencia de medias es lo suficientemente grande, se supone que los dos grupos difieren.

B. ANOVA de una vía:

En la Anova de una vía, se plantea una hipótesis nula en la que se asume que no hay diferencias entre las medias de los grupos, y se busca evidencia estadística para rechazar esta hipótesis en favor de una hipótesis alternativa que sugiere que al menos un grupo tiene una media diferente.

El análisis de varianza se basa en la descomposición de la variabilidad total de los datos en dos componentes: la variabilidad debida a las diferencias entre los grupos (variabilidad entre grupos) y la variabilidad debida a las diferencias dentro de los grupos (variabilidad dentro de grupos). Si la variabilidad entre grupos es significativamente mayor que la variabilidad dentro de grupos, se concluye que existen diferencias entre los grupos.

C. REGRESIÓN LINEAL:

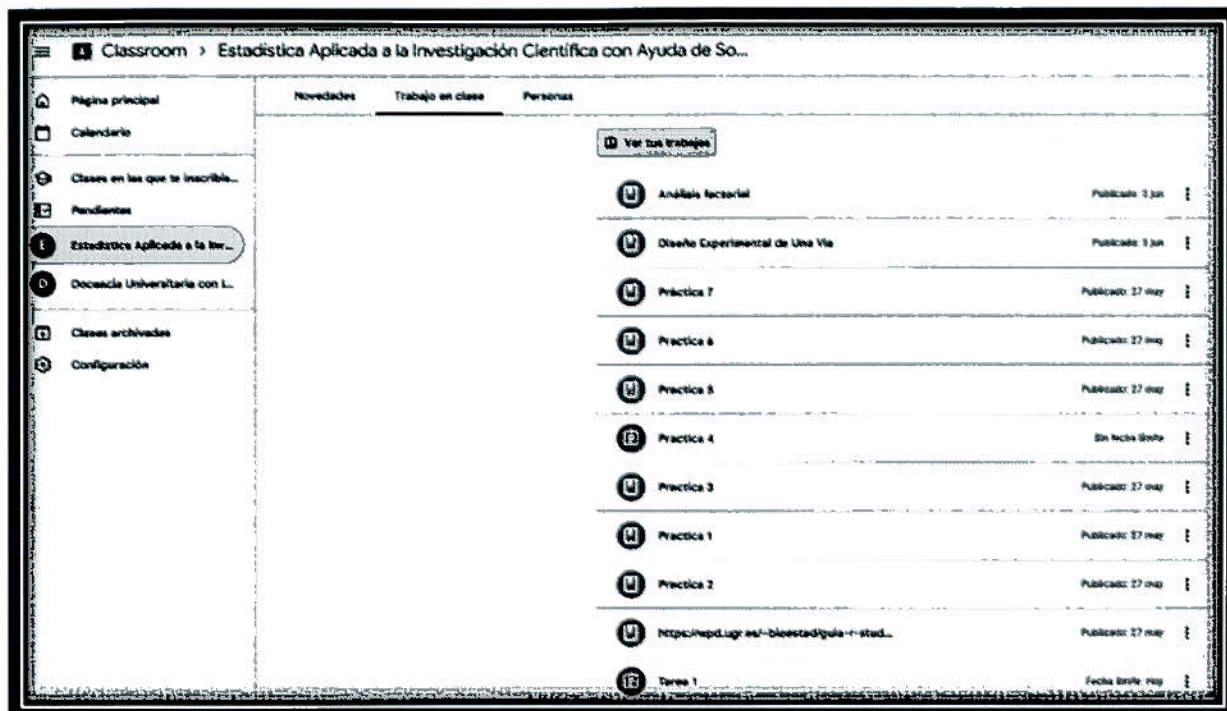
La regresión lineal es un algoritmo de aprendizaje supervisado que se utiliza en Machine Learning y en estadística. En su versión más sencilla, lo que haremos es "dibujar una recta" que nos indicará la tendencia de un conjunto de datos continuos. La regresión lineal utiliza la relación entre los puntos de datos para trazar una línea recta que los atraviese. Esta línea permite predecir valores futuros.

- ✓ R muestra los resultados en un formato estándar bien organizado
- ✓ Python ofrece un resumen similar, pero requiere añadir manualmente la constante
- ✓ Ambos proporcionan coeficientes, errores estándar, valores p y R^2 .

D. PRUEBA DE CHI-CUADRADO PARA INDEPENDENCIA:

Una prueba de independencia de chi-cuadrado (χ^2) es un tipo de prueba de chi-cuadrado de Pearson. Las pruebas de chi-cuadrado de Pearson son pruebas no paramétricas para variables categóricas. Se utilizan para determinar si los datos difieren significativamente de lo esperado. Se puede utilizar una prueba de chi-cuadrado de independencia, también conocida como prueba de chi-cuadrado de asociación, para determinar si dos variables categóricas están relacionadas. Si dos variables están relacionadas, la probabilidad de que una tenga un valor determinado depende del valor de la otra.

- ✓ R muestra más información por defecto (incluyendo frecuencias esperadas).
- ✓ Python requiere desempaquetar los resultados manualmente.



Inicio de la Sesión N° 04. Análisis inferencial y modelos estadísticos: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de R y PYTHON", del 31 de mayo del 2025.

5.5 MÓDULO 5: APLICACIONES AVANZADAS Y PROYECTOS DE INVESTIGACIÓN

5.5.1 ANÁLISIS MULTIVARIADO: INTRODUCCIÓN A TÉCNICAS COMO PCA, CLUSTER ANALYSIS Y ANÁLISIS DISCRIMINANTE.

El análisis multivariado es un conjunto de técnicas estadísticas que permiten analizar datos con múltiples variables simultáneamente. Estas metodologías son fundamentales en campos como la inteligencia artificial, la bioinformática, la economía y las ciencias sociales, donde se busca entender patrones, reducir dimensiones o clasificar observaciones.

A. PCA (Análisis de Componentes Principales):

Objetivo: Reducir la dimensionalidad de un conjunto de datos conservando la mayor variabilidad posible.

¿Cómo funciona?

- Transforma las variables originales en un nuevo conjunto de variables no correlacionadas llamadas **componentes principales**.
- El primer componente captura la mayor varianza, el segundo (ortogonal al primero) captura la siguiente mayor varianza, y así sucesivamente.

Aplicaciones:

- Visualización de datos en 2D/3D.
- Eliminación de ruido o redundancia en variables.
- Preprocesamiento para modelos de machine learning.

B. Cluster Analysis (Análisis de Clusters)

Objetivo: Agrupar observaciones en grupos homogéneos (**clusters**) basados en similitudes.

Enfoques comunes:

- **K-means:** Divide datos en k clusters minimizando la varianza intra-cluster.
- **Jerárquico:** Crea un dendrograma para visualizar relaciones entre clusters.
- **DBSCAN:** Identifica clusters basados en densidad, útil para datos no esféricos.

Aplicaciones:

- Segmentación de clientes en marketing.
- Clasificación de imágenes en visión por computadora.
- Detección de anomalías.

C. Análisis Discriminante (LDA: Linear Discriminant Analysis)

Objetivo: Encontrar combinaciones lineales de variables que **separen mejor las clases predefinidas**.

¿Cómo funciona?



Maximiza la separación entre clases y minimiza la dispersión dentro de cada clase.
Genera **funciones discriminantes** para clasificar nuevas observaciones.

Diferencias con PCA:

- PCA es no supervisado (sin etiquetas); LDA es supervisado (usa etiquetas de clases).
- PCA maximiza varianza total; LDA maximiza separación entre grupos.

Aplicaciones:

- Diagnóstico médico (ej., clasificar tumores benignos/malignos).
- Reconocimiento facial o de voz.

5.5.2 MACHINE LEARNING BÁSICO: USO DE MODELOS PREDICTIVOS Y ALGORITMOS DE CLASIFICACIÓN EN EL CONTEXTO DE LA INVESTIGACIÓN.

El machine learning (aprendizaje automático) se ha convertido en una herramienta fundamental para la investigación científica, permitiendo analizar patrones complejos en datos y hacer predicciones basadas en evidencia empírica.

5.2.2.1 Tipos de Modelos Predictivos Básicos:

A. Modelos de Regresión:

- ✓ **Regresión Lineal:** Predice valores continuos
- ✓ **Regresión Logística:** Aunque se llama "regresión", es un algoritmo de clasificación binaria

B. Algoritmos de Clasificación:

- ✓ **Árboles de Decisión:** Fáciles de interpretar, buenos para datos categóricos.
- ✓ **Random Forest:** Conjunto de árboles de decisión que mejora la precisión.
- ✓ **SVM (Máquinas de Vectores de Soporte):** Efectivos en espacios de alta dimensión.
- ✓ **K-Nearest Neighbors (KNN):** Clasificación basada en similitud.

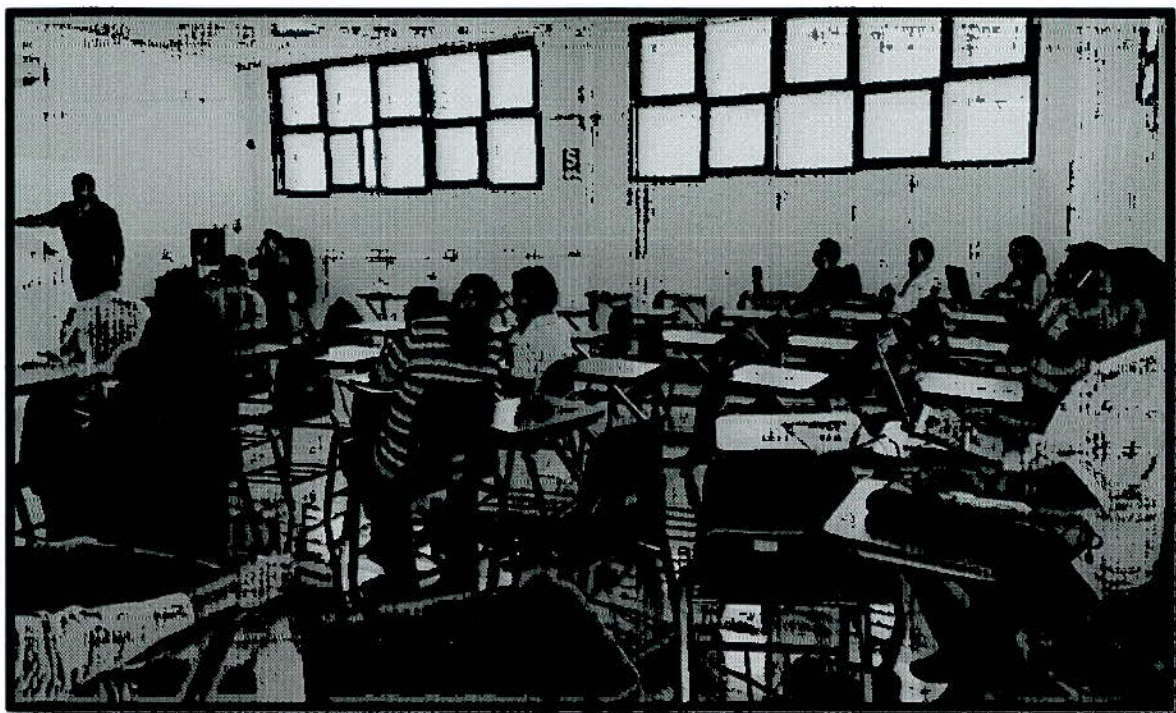
Proceso Básico para Implementación:

- a) **Definición del Problema:** ¿Qué queremos predecir o clasificar?
- b) **Recolección de Datos:** Conjuntos de entrenamiento y prueba
- c) **Preprocesamiento:**
 - Limpieza de datos
 - Normalización/escalado
 - Manejo de valores faltantes
- d) **Selección del Modelo:** Basado en la naturaleza del problema
- e) **Entrenamiento:** Ajuste de parámetros
- f) **Evaluación:** Métricas como precisión, recall, F1-score
- g) **Implementación:** Uso del modelo en nuevos datos



Aplicaciones en Investigación:

- Clasificación de muestras biológicas
- Predicción de resultados clínicos
- Análisis de imágenes médicas
- Segmentación de datos de encuestas
- Detección de patrones en datos experimentales



Inicio de la Sesión N° 05. Aplicaciones avanzadas y proyectos de investigación: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON", del 03 de junio del 2025.



VI. RESULTADOS

Mediante el desarrollo del curso de capacitación, se ha logrado fortalecer las capacidades de aprendizaje en un total de 77 docentes de la Universidad Nacional de Frontera, y poder contribuir en la calidad educativa dentro de la casa Universitaria.

Cuadro 2: Docentes asistentes al curso de capacitación: "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON"

N°	APELLIDOS Y NOMBRES	20-May	24-May	27-May	31-May	03-Jun	07-Jun	PORCENTAJE DE ASISTENCIA
01	ALDANA JUAREZ WILLIAM	X	X	X	X	X	X	100.00
02	ALDANA YARLEQUE CRISTHIAN NICOLAS	X	X	X	X	X	X	100.00
03	AREVALO CARCAMO LILIANA MILAGROS	X	X	X	X	X	X	100.00
04	ATO COBA SALVADOR RAFAEL	X	X	X	X	X	X	100.00
05	ATOCHÉ SILVA LUZ ANGELICA	X	X	X	X	X	X	100.00
06	BARTUREN CABRERA JUAN PABLO FRANCISCO	X	X	X	X	X	X	100.00
07	BRUNO COBEÑAS PRIMITIVO	X	X	X	X	X	X	100.00
08	BUSTAMANTE VILCHEZ DE TAY SHIRLEY TATIANA	X	X	X	X	X	X	100.00
09	CABRERA TORRES ANTERO ALEXANDER	X	X	X	X	X	X	100.00
10	CADENILLAS RICHARD ORDINOLA	X	X	X	X	X	X	100.00
11	DALLE RUIZ SAMANTHA HILDA	X	X	X	X	X	X	100.00
12	CAMINO ORDINOLA ALBA MARINA	X	X	X	X	X	X	100.00
13	CANGO CORDOVA JUAN ISAIAS	X	X	X	X	X	X	100.00
14	CARRASCO CHOQUE FREDDY	X	X	X	X	X	X	100.00
15	CASTILLO CHUNG LUIGGI BRUNO	-	X	X	X	X	X	83.33
16	CASTRO SILUPU WILSON MANUEL	X	X	X	X	X	X	100.00
17	CUNGUIA PIEDRA DEYVI DAVID	X	X	X	X	X	X	100.00
18	CHAVEZ EPIQUEN ABDIAS	X	X	X	X	X	X	100.00
19	COELLO DAVILA EUDOCIA DEL PILAR	X	X	X	X	X	X	100.00
20	FLORES CASTILLO MARCOS MARCELO	X	X	X	X	X	-	83.33
21	FLORES RIVERA, ANTONY STEVE	X	X	X	X	X	X	100.00
22	GARCIA CRIOLLO MARIA LUZ	X	X	X	X	X	X	100.00
23	GARCIA VILELA LUCY MARIELLA	X	X	X	X	X	X	100.00
24	GONZALEZ ESPINO DAVID	X	X	X	X	X	X	100.00
25	GUERRERO CAMPOS FRANKLIN	X	X	X	X	X	X	100.00
26	GUTIERREZ VALVERDE KARINA SILVANA	X	X	X	X	X	X	100.00



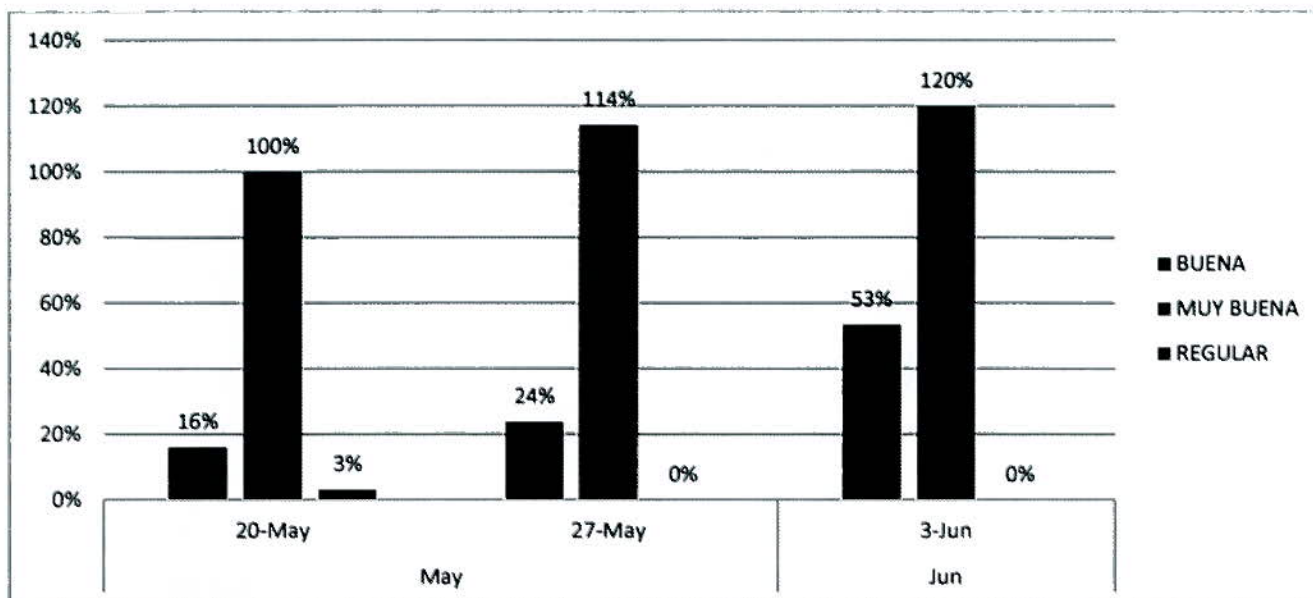
27	HERRERA CALLO ALINA DAMMAR	X	X	X	X	X	X	100.00
28	HINOJOSA NIQUÉN YAHAIRA ANABEL	X	X	X	X	X	X	100.00
29	ICANAQUE ORDINOLA JUAN DAMIAN	X	X	X	X	X	X	100.00
30	ISHUZA TAPULLIMA RAIMUNDO	X	X	-	X	X	X	83.33
31	JARA ALARCON TERESA JULIANA	X	X	X	X	X	X	100.00
32	JUAREZ ORTIZ VLADIMIR	-	X	X	X	X	X	83.33
33	LECARNAQUE AREVALO CARLOS	X	X	X	X	X	X	100.00
34	LOPEZ GUAYANAY REYDER OVIDIO	X	X	X	X	X	X	100.00
35	LOPEZ CORDOVA JENNY	X	X	X	X	X	X	100.00
36	LUJAN VERA PRISCILA ESTELITA	X	X	X	X	X	X	100.00
37	MAMANI CORNEJO JOYCE	X	X	X	X	X	X	100.00
38	MARTINEZ NOLE IRMA VICTORIA	X	X	-	X	X	X	83.33
39	MENDOZA CULQUI EDIN	X	X	X	X	X	X	100.00
40	NUÑEZ ALEJOS LUIS ALBERTO	X	X	X	X	X	X	100.00
41	ORREGO ZAPO SALVADOR	X	X	X	X	X	X	100.00
42	ORTIZ CASTILLO ROSA GISELA	X	X	X	X	X	X	100.00
43	PALACIOS ZAPATA CLAUDIA	X	X	X	X	X	X	100.00
44	PATINO GARCIA LUCIA	X	X	X	X	X	X	100.00
45	PENA ALVARADO EXILDA	X	X	X	X	X	X	100.00
46	PIÑARRETA OLIVARES RONY ALEXANDER	X	X	X	X	X	X	100.00
47	PURIZACA SANTISTEBAN STEFANY	X	X	X	X	X	X	100.00
48	RETO GOMEZ JANNYNA	X	X	X	X	X	-	83.33
49	REYES ARTEAGA LUCY MARISOL	X	X	X	X	X	X	100.00
50	REYES VIDAL MARCO	X	X	X	X	X	X	100.00
51	RISCE REYES DAVID ROBERTO	X	X	X	X	X	X	100.00
52	RODAS PURIZAGA HANDRY MARTIN	X	X	X	X	X	X	100.0
53	ROSALES CUENTAS MIRIAM MARLENI	X	X	X	X	X	-	83.33
54	SAAVEDRA CANO FERMIN MÁXIMO	X	X	-	X	X	X	83.33
55	SAAVEDRA NAVARRO JESSENIA	X	X	X	X	X	X	100.0
56	SALAZAR SANDOVAL CARLOS AUGUSTO	X	X	-	X	X	X	83.33
57	SANCHEZ CASTRO JANET DEL ROCIO	X	X	X	X	X	X	100.00
58	SANCHEZ CHERO JOSE ANTONIO	X	X	-	X	X	X	83.33
59	SANCHEZ PACHECO EDUARDO	X	X	X	X	X	-	83.33
60	SANCHEZ PRIETO MARIA GREGORIA	X	X	X	X	X	X	100.00

61	SANCHO NORIEGA CARLOMAGNO	X	X	X	X	X	X	100.00
62	SANDOVAL CALDERON HUGO ALEXIS	X	X	X	X	X	X	100.00
63	SANTA CRUZ AREVALO JOHANA	X	X	X	X	X	X	100.00
64	SEMINARIO MORALES MARIA VERONICA	X	X	X	X	X	X	100.00
65	SIANCÁS ESCOBAR DARWIN ALEJANDRO	X	X	X	X	X	X	100.00
66	SILVA JUAREZ RAQUEL	X	X	X	X	X	X	100.00
67	SOCOLA JUAREZ ZURY MABELL	X	X	X	X	X	X	100.00
68	TAY LEÓN JULIO CESAR	X	X	X	X	X	X	100.00
69	TIMANA ALVAREZ MARCOS	X	X	X	X	X	X	100.00
70	TIRAVANTTI COSTANTINO JULIO CESAR	X	X	X	X	X	X	100.00
71	TORRES BECERRA PATRICIA MERCEDES	X	X	X	X	X	X	100.00
72	TRELES POZO LUIS RAMÓN	X	X	X	X	X	X	100.00
73	VALLEJOS MORE LEANDRO ALONSO	X	X	X	X	X	X	100.00
74	VEGA PORTALATINO EDWIN JORGE	X	X	X	X	X	X	100.00
75	VILLALTA ARELLANO SEGUNDO RAMOS	X	X	X	X	X	X	100.00
76	ZAPATA MENDOZA PROSPERO CRISTHIAN ONOFRE	X	X	X	X	X	X	100.00
77	ZERGA ROMANI JOSÉ JORGE	X	X	X	X	X	X	100.00

6.1 Encuesta de satisfacción:

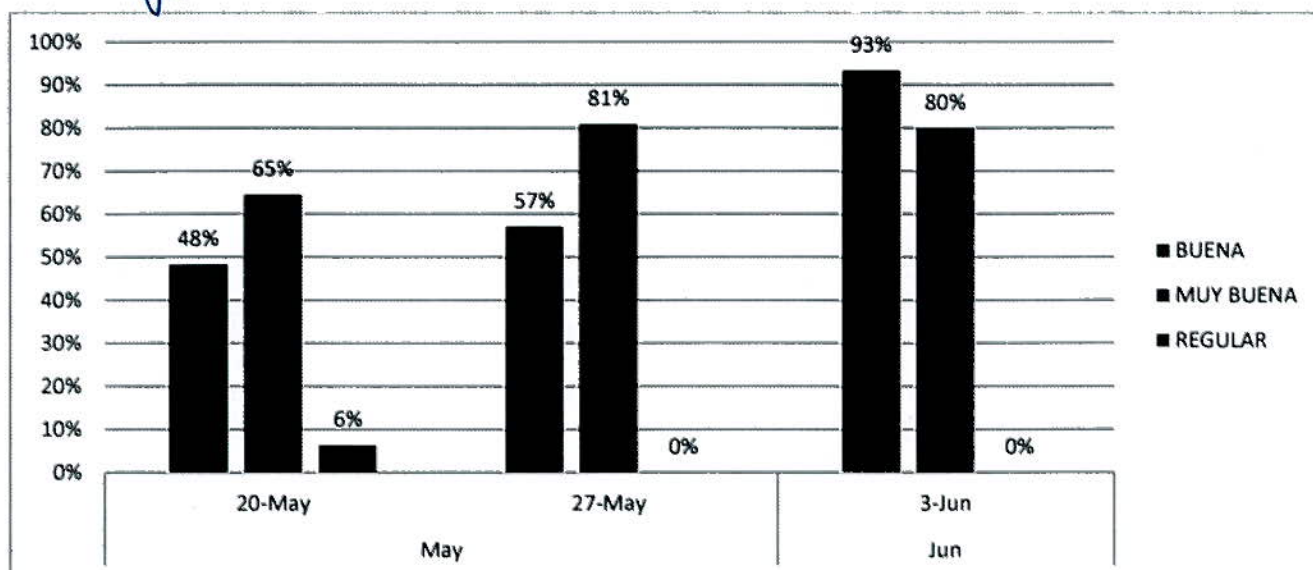
Finalizando la última sesión del curso de capacitación ""Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON""", se le aplicó una encuesta de satisfacción a los docentes participantes, con la finalidad de saber si fue de su agrado este curso, dando como resultado una satisfacción aceptable, la cual se da a conocer el siguiente gráfico:

CRITERIO N°01: PUNTUALIDAD Y ASISTENCIA



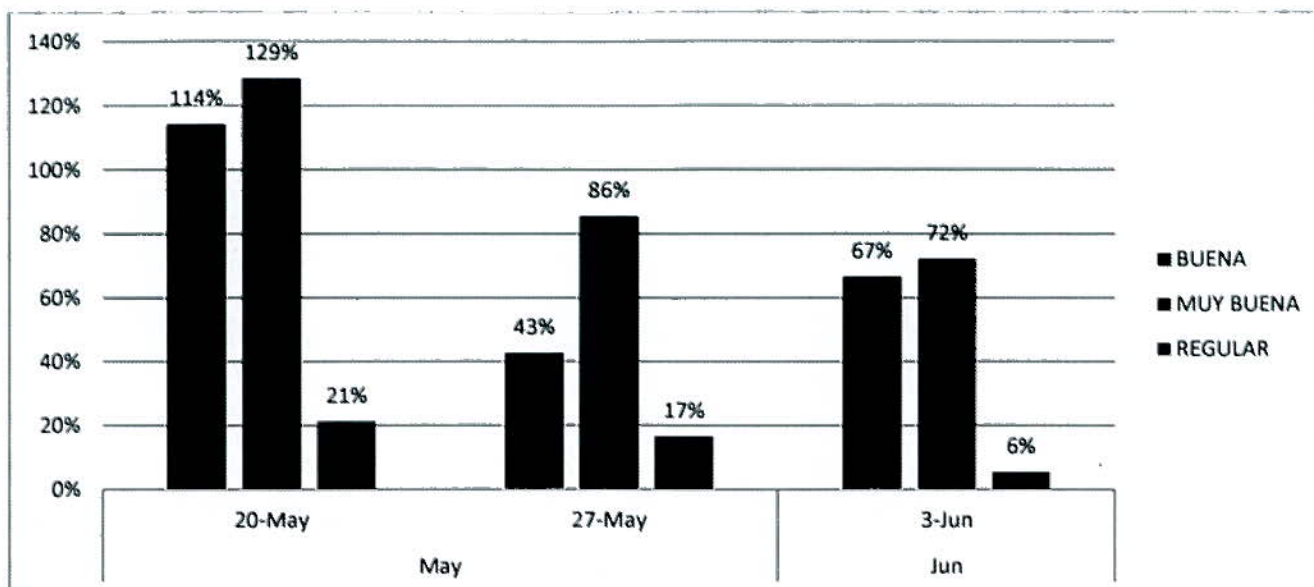
Los criterios propuestos para la capacitación han sido considerados entre buena, muy buena y regular por los docentes participantes así tenemos que para el día 20 de mayo, 16% buena, 100% Muy buena, y Regular 3% el día 27 de mayo 24% buena y 114% Muy buena y el día 03 de junio 53% buena y 120% Muy buena.

CRITERIO N°02: PRESENTACIÓN PERSONAL



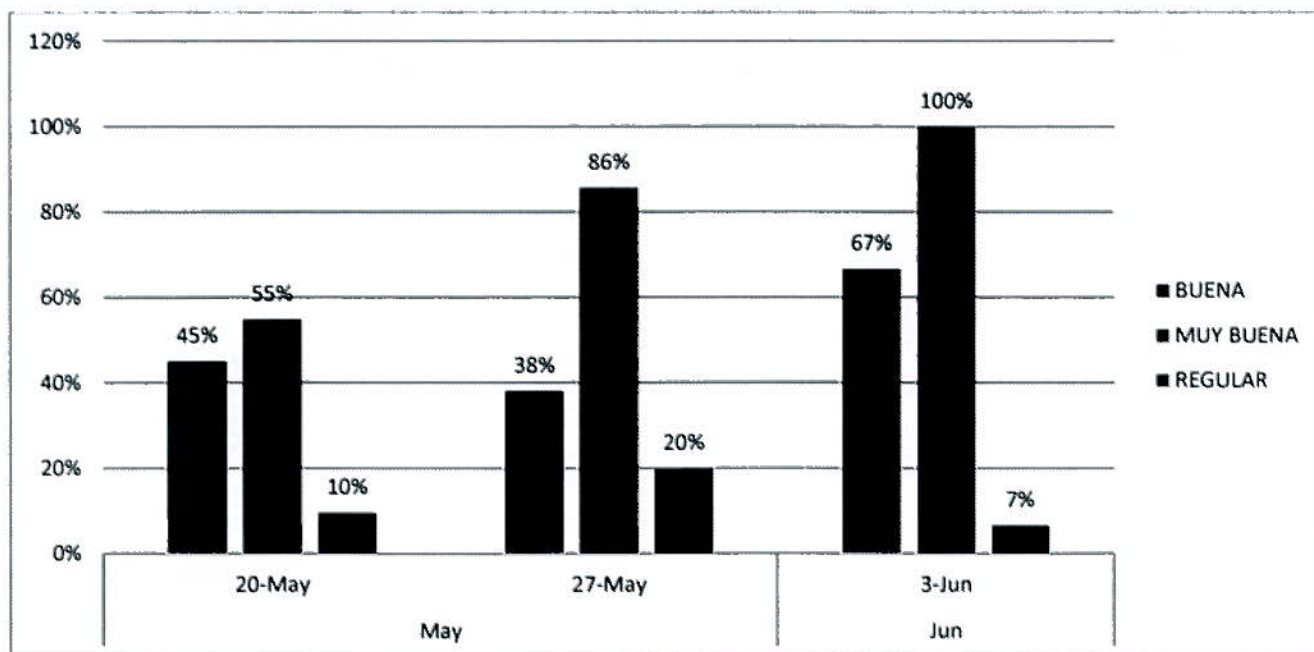
Los criterios propuestos para la capacitación han sido considerados entre buena, muy buena y regular por los docentes participantes así tenemos que para el día 20 de mayo, 48% buena, 65% Muy buena, y Regular 6% el día 27 de mayo 57% buena y 81% Muy buena y el día 03 de junio 93% buena y 80% Muy buena.

CRITERIO N°03: DOMINIO DEL TEMA:



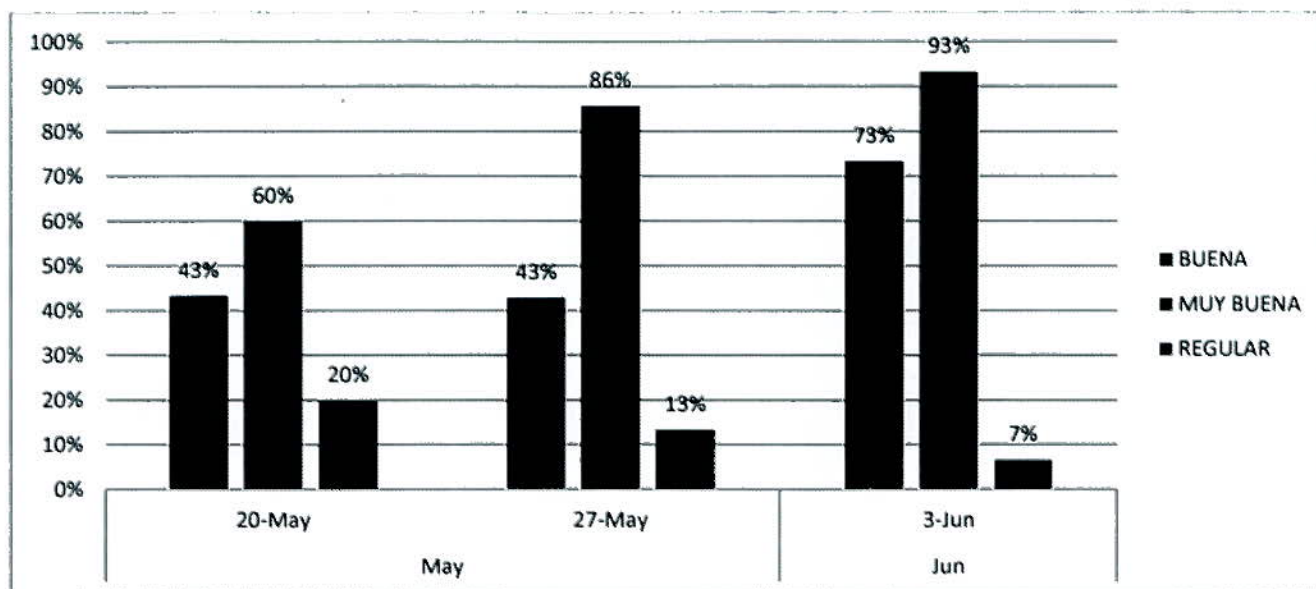
Los criterios propuestos para la capacitación han sido considerados entre buena, muy buena y regular por los docentes participantes así tenemos que para el día 20 de mayo, 114% buena, 129% Muy buena, y Regular 21% el día 27 de mayo 43% buena, 86% Muy buena y Regular 17% y el día 03 de junio 67% buena, 72% Muy buena y Regular 6%.

CRITERIO N°04: CAPACIDAD DE CRITERIO ANALÍTICO



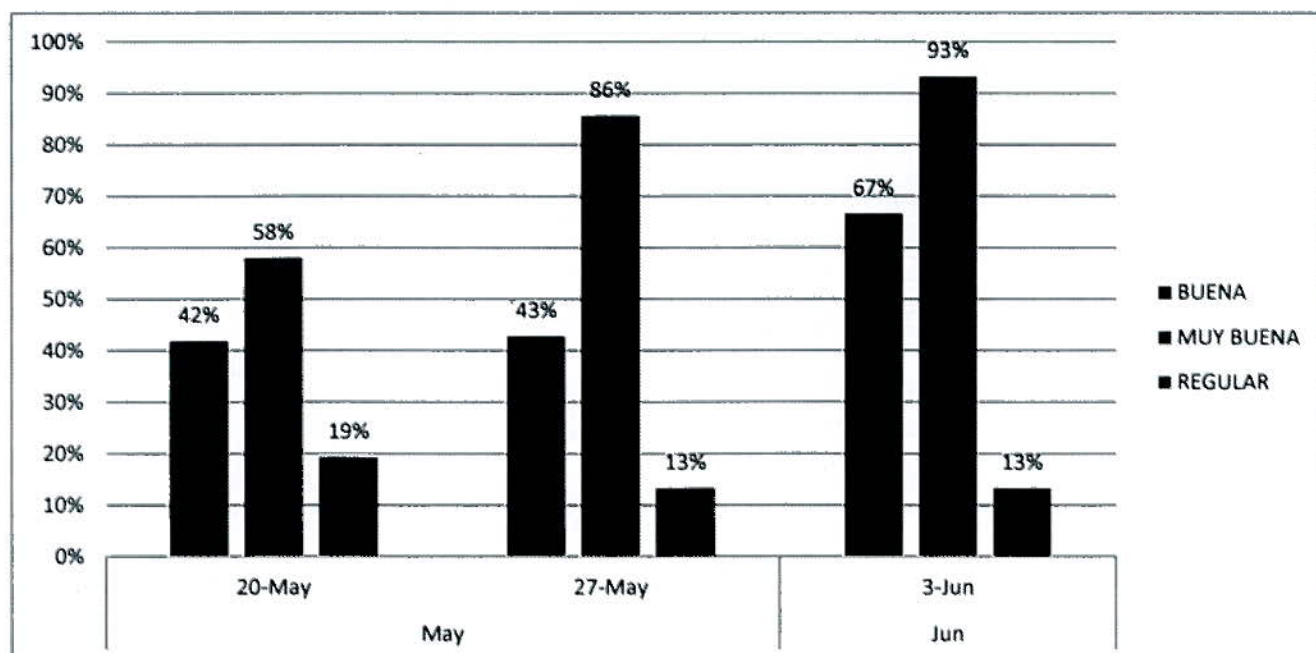
Los criterios propuestos para la capacitación han sido considerados entre buena, muy buena y regular por los docentes participantes así tenemos que para el día 20 de mayo, 45% buena, 55% Muy buena, y Regular 10% el día 27 de mayo 38% buena, 86% Muy buena y Regular 20% y el día 03 de junio 67% buena, 100% Muy buena y Regular 7%.

CRITERIO N°05: PRESENTACIÓN Y DOMINIO DEL MATERIAL DIDÁCTICO



Los criterios propuestos para la capacitación han sido considerados entre buena, muy buena y regular por los docentes participantes así tenemos que para el día 20 de mayo, 16% buena y 100% Muy buena, y Regular 3% el día 27 de mayo 24% buena y 114% Muy buena y el día 13 de mayo 13% buena y 87% Muy buena.

CRITERIO N°06: ÓPTIMA COMUNICACIÓN VERBAL



Los criterios propuestos para la capacitación han sido considerados entre buena, muy buena y regular por los docentes participantes así tenemos que para el día 20 de mayo, 16% buena y 100% Muy buena, y Regular 3% el día 27 de mayo 24% buena y 114% Muy buena y el día 13 de mayo 13% buena y 87% Muy buena.

IV. CONCLUSIONES Y RECOMENDACIONES

7.1 CONCLUSIONES

Con el desarrollo del curso de Capacitación denominado **"Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON"**, se concluye qué:

- ✦ Se ha brindado una sólida base de conocimientos a los docentes participantes de la Universidad Nacional de Frontera.
- ✦ El desarrollo del curso se ha dado con normalidad, cumpliéndose los objetivos propuestos en el plan de trabajo, y despertando gran interés entre los diversos asistentes.
- ✦ El curso de "Estadística Aplicada a la Investigación Científica con Ayuda de Software Programación de "R" y "PYTHON", este curso proporciona las herramientas teórico-prácticas para integrar el análisis estadístico en la investigación científica, aprovechando el poder de R y Python para obtener resultados robustos y confiables.
- ✦ Proporciona una base sólida para el análisis de datos y la interpretación de resultados en el ámbito científico. Al finalizar el curso, los participantes habrán adquirido habilidades en la aplicación de métodos estadísticos, la programación en R y Python para el análisis de datos, y la interpretación de resultados en el contexto de la investigación. Además, habrán desarrollado una comprensión de cómo estas herramientas pueden mejorar la eficiencia y la calidad de los estudios científicos.
- ✦ La intervención de la estadística aplicada en el ámbito educativo es clave para el desarrollo de evaluaciones sobre la efectividad de los programas educativos y de este modo poder tomar decisiones fundamentadas empíricamente de los estudiantes brindando una educación de calidad.

7.2 RECOMENDACIONES

- ✦ Siendo los docentes el principal actor dentro del proceso formativo a través de las capacitaciones, se sugiere a las áreas académicas de la UNF (Vicepresidencia Académica, Coordinación de Facultad, Departamentos Académicos, Direcciones de Escuela, entre otras), mantener una comunicación constante entre sí, facilitando de esta manera la asistencia adecuada de los docentes a los eventos académicos planificados.
- ✦ Gestionar en forma célere y oportuna la aprobación administrativa y legal de los diferentes documentos, permitiendo que el desarrollo de los cursos de capacitación se dé adecuadamente, según la planificación formulada.