



Universidad Tecnológica del Perú

**Relación entre la capacitación continua de los
trabajadores y su productividad en una empresa de
tecnología en Perú durante el año 2025**

Para la asignatura de Estadística Inferencial

Huatay Salcedo, Luis	U24218809
Torres Vara, Mateo	U24308542
Lizana Flores, Jhonjar	U22218613

Sección 44316

8 de junio de 2025

Docente: Doc. Cesar Humberto Zavala Inga

Índice

Índice	2
1 Introducción	3
2 Marco Teórico	4
2.1 La estadística inferencial	4
2.2 El ambiente laboral tecnológico	4
3 Objetivo general	5
3.1 Objetivos específicos	5
4 Presentación de los Datos	6
4.1 Descripción de las Variables	6
5 Planteamiento de Hipótesis	7
6 Metodología	8
6.1 Enfoque estadístico	8
6.2 Prueba de hipótesis para diferencia de medias	8
6.3 Intervalos de confianza	8
6.4 Análisis complementarios	9
7 Análisis y Desarrollo	10
7.1 Estadística descriptiva	10
7.2 Prueba de hipótesis para diferencia de medias	11
7.3 Intervalo de confianza para la diferencia de medias	11
7.4 Visualización de resultados	12
7.5 Resumen de hallazgos	12
8 Interpretación de resultados	13
8.1 Decisiones	13
9 Casos Particulares	14
9.1 Tipo de muestreo: <i>Muestreo aleatorio simple</i>	14
9.2 Cálculo del tamaño de muestra	15
9.3 Distribución muestral: <i>Varianza conocida</i>	16
9.4 Comparación de dos medias con varianzas desconocidas	18
9.5 Intervalo de confianza para diferencia de proporciones	19
9.6 Prueba de hipótesis para la media (varianza conocida)	21
9.7 Prueba de hipótesis para diferencia de medias	22
9.8 Prueba de hipótesis para diferencia de proporciones.	24
9.9 Prueba de hipótesis para varianzas.	26
10 Anexos	28
10.1 Anexo 1: Tabla de frecuencias acumuladas Z	28
10.2 Anexo 2: Tabla de frecuencias acumuladas T	28
10.3 Anexo 3: Tabla de frecuencias acumuladas χ^2	28
10.4 Anexo 4: Tabla de frecuencias acumuladas F	28
Referencias	29

1. Introducción

En el contexto actual de alta competitividad tecnológica, las empresas peruanas enfrentan el reto de optimizar la productividad de sus equipos. Una pregunta crítica es si los trabajadores de departamentos tecnológicos (IT) muestran niveles de desempeño significativamente diferentes a los de áreas no tecnológicas (Marketing, HR, Ventas, Finanzas), posiblemente debido a su exposición a herramientas avanzadas y entornos de innovación.

Este estudio se centra en analizar estadísticamente si existen diferencias significativas en la productividad laboral entre estos dos grupos. Utilizando datos de una empresa tecnológica peruana, evaluaremos específicamente si la media de productividad en el departamento de IT difiere de manera significativa de la observada en otros departamentos.

Este análisis es relevante porque puede revelar patrones de desempeño asociados al contexto tecnológico, proporcionando insights valiosos para:

- Asignación estratégica de recursos
- Diseño de programas de capacitación diferenciados
- Políticas de gestión del talento sectorizadas

El enfoque estadístico inferencial permitirá determinar si las diferencias observadas son significativas o atribuibles al azar, aportando evidencia empírica para la toma de decisiones.

2. Marco Teórico

2.1. La estadística inferencial

La estadística inferencial es una rama de la estadística que se encarga de hacer inferencias o generalizaciones sobre una población a partir de una muestra representativa. A través de técnicas estadísticas, se pueden estimar parámetros poblacionales, realizar pruebas de hipótesis y calcular intervalos de confianza. La estadística inferencial es fundamental en la investigación científica, ya que permite tomar decisiones basadas en datos y evaluar la validez de las afirmaciones realizadas a partir de muestras.

El análisis estadístico inferencial provee herramientas que permiten la evaluación sistemática y eficiente de una muestra de la población que se quiere estudiar. [1]

Veiga, Otero, Torres (2020)

A fin de elaborar un correcto marco teórico se consideró pertinente consultar las fuentes bibliográficas adecuadas. De esta manera Lopez y Fachelli. (2016) nos dice que la *Estadística Inferencial* es aquella que se encarga de realizar inferencias sobre una población a partir de una muestra. Esto implica que, a partir de los datos obtenidos de una muestra, se pueden hacer afirmaciones o generalizaciones sobre la población de la cual se extrajo la muestra.

2.2. El ambiente laboral tecnológico

El sector tecnológico se caracteriza por su dinamismo y constante evolución, lo que plantea desafíos y oportunidades para los trabajadores. La capacitación continua se ha convertido en un elemento clave para mantener la competitividad en este entorno. Las empresas tecnológicas deben adaptarse rápidamente a los cambios en la industria, lo que requiere que sus empleados estén actualizados en las últimas tendencias y tecnologías.

Según Vallbona y Mascarilla. (2018) Nos dice: Los empleados del sector de las Tecnologías de la Información en España manifiestan, en general, niveles elevados de satisfacción laboral, especialmente en lo referente a la autonomía en el trabajo y las oportunidades de desarrollo profesional

Satisfacción laboral. El caso de los empleados del sector de las tecnologías de la información en España. [3]

Vallbona, Mascarilla (2018)

3. Objetivo general

Se pretende realizar un análisis estadístico inferencial del rendimiento laboral de trabajadores de la industria Tech en el Perú, con el fin de determinar si existen diferencias significativas en el rendimiento laboral entre aquellos que utilizan herramientas de inteligencia artificial y aquellos que no las utilizan. Para ello, se recopilarán datos sobre el rendimiento laboral de los trabajadores, así como información sobre su uso de herramientas de inteligencia artificial. Se espera que este análisis permita identificar patrones y tendencias en el rendimiento laboral, así como proporcionar recomendaciones para mejorar la productividad y eficiencia en el sector tecnológico.

3.1. Objetivos específicos

1. **Identificar y analizar:** Se busca identificar y analizar las herramientas de inteligencia artificial más utilizadas en el sector tecnológico en Perú, así como su impacto en el rendimiento laboral de los trabajadores. Esto permitirá comprender cómo estas herramientas influyen en la productividad y eficiencia de los empleados.
2. **Evaluar la relación:** Se pretende evaluar la relación entre el uso de herramientas de inteligencia artificial y el rendimiento laboral de los trabajadores. Esto implica analizar si existe una correlación positiva entre el uso de estas herramientas y el rendimiento laboral, así como identificar factores que puedan influir en esta relación.
3. **Realizar un análisis estadístico:** Se llevará a cabo un análisis estadístico inferencial para determinar si existen diferencias significativas en el rendimiento laboral entre los trabajadores que utilizan herramientas de inteligencia artificial y aquellos que no las utilizan. Esto permitirá validar o refutar la hipótesis planteada y proporcionar evidencia empírica sobre el impacto de estas herramientas en el rendimiento laboral.
4. **Proporcionar recomendaciones:** A partir de los resultados obtenidos, se buscará proporcionar recomendaciones para mejorar la productividad y eficiencia en el sector tecnológico. Esto incluirá sugerencias sobre la implementación de herramientas de inteligencia artificial, así como estrategias para fomentar su uso entre los trabajadores.
5. **Contribuir al conocimiento:** Se espera que este análisis contribuya al conocimiento sobre el impacto de las herramientas de inteligencia artificial en el rendimiento laboral en el sector tecnológico en Perú. Esto permitirá enriquecer la literatura existente sobre el tema y proporcionar información valiosa para futuras investigaciones.

Estos objetivos específicos permitirán abordar de manera integral el análisis del rendimiento laboral de los trabajadores de la industria Tech en Perú, así como su relación con el uso de herramientas de inteligencia artificial. Se espera que los resultados obtenidos sean relevantes y útiles para mejorar la productividad y eficiencia en el sector tecnológico.

4. Presentación de los Datos

Para el análisis de la productividad de los trabajadores en el sector tecnológico, se ha recopilado un conjunto de datos que incluye información relevante sobre empleados de distintas áreas funcionales. Esta base contiene variables tanto cuantitativas como cualitativas que permiten una evaluación exhaustiva del desempeño individual en relación con factores como la satisfacción, el salario y el cargo dentro de la organización.

4.1. Descripción de las Variables

- **Name:** Nombre del empleado.
- **Age:** Edad (en años).
- **Gender:** Género (Masculino/Femenino).
- **Projects Completed:** Número total de proyectos completados por el empleado.
- **Productivity (%)**: Porcentaje de productividad basada en KPIs internos.
- **Satisfaction Rate (%)**: Porcentaje de satisfacción reportado en encuestas internas.
- **Feedback Score:** Puntuación promedio de retroalimentación recibida (escala de 1 a 5).
- **Department:** Departamento donde labora el empleado (ej. IT, Marketing, HR, etc.).
- **Position:** Puesto que ocupa el empleado (ej. Intern, Analyst, Manager, etc.).
- **Joining Date:** Fecha de ingreso a la empresa (mes-año).
- **Salary:** Salario anual expresado en dólares americanos.

A continuación, se muestran las variables de los datos empleados en el análisis:

Name	Age	Gen.	Proj.	Prod.	Sat.	Feed.	Dept.	Position	Join	Salary
Douglas Lindsey	25	M	11	57	25	4.7	Marketing	Analyst	Jan-20	63596
Anthony Roberson	59	F	19	55	76	2.8	IT	Manager	Jan-99	112540
Thomas Miller	30	M	8	87	10	2.4	IT	Analyst	Jan-17	66292
Joshua Lewis	26	F	1	53	4	1.4	Marketing	Intern	Jan-22	38303
Stephanie Bailey	43	M	14	3	9	4.5	IT	Team Lead	Jan-05	101133

Cuadro 1: Muestra de la base de datos de trabajadores del sector tecnológico

Fuente: Elaboración propia a partir de datos ficticios representativos del sector tecnológico para fines académicos.

Este conjunto de datos servirá como base para el análisis estadístico posterior, en el cual se aplicarán técnicas de estadística descriptiva e inferencial para evaluar los factores que impactan la productividad en los distintos departamentos y cargos.

5. Planteamiento de Hipótesis

El presente estudio tiene como objetivo evaluar si existen diferencias significativas en el nivel de productividad entre los trabajadores de departamentos tecnológicos (específicamente el departamento de Tecnologías de la Información, IT) y aquellos de departamentos no tecnológicos (como Marketing, Recursos Humanos, Ventas y Finanzas). Esta comparación permite explorar si el contexto laboral tecnológico, donde es más probable el uso de herramientas avanzadas como la inteligencia artificial, está asociado a un desempeño diferenciado. Para ello, se plantea una hipótesis estadística que será evaluada mediante una prueba de diferencia de medias.

Hipótesis nula y alternativa

Se parte de la premisa de que el departamento de IT, al estar más expuesto a tecnologías innovadoras, podría mostrar niveles de productividad distintos. Sin embargo, esta afirmación requiere validación estadística. Con base en ello, se formulan las siguientes hipótesis:

- **Hipótesis nula (H_0):** $\mu_{IT} = \mu_{No-IT}$ (No existe diferencia significativa en la productividad media entre departamentos tecnológicos y no tecnológicos).
- **Hipótesis alternativa (H_1):** $\mu_{IT} \neq \mu_{No-IT}$ (Existe una diferencia significativa en la productividad media entre ambos grupos de departamentos).

Justificación

Esta hipótesis se alinea con los objetivos del estudio de evaluar el impacto de herramientas tecnológicas en el rendimiento laboral, considerando que:

1. El departamento de IT opera como *proxy* para trabajadores con mayor exposición a tecnologías avanzadas (incluyendo IA).
2. Permite responder a la pregunta: ¿El entorno tecnológico está asociado a un mejor desempeño?
3. La prueba estadística seleccionada (prueba t para muestras independientes) es adecuada porque:
 - La varianza poblacional es desconocida
 - Las muestras son independientes (distintos departamentos)
 - La variable dependiente (productividad) es cuantitativa continua

Variables clave

- **Variable independiente:** Tipo de departamento (categórica: IT vs. No IT)
- **Variable dependiente:** Porcentaje de productividad (cuantitativa continua)
- **Variables de control:**
 - Posición jerárquica (Manager, Analyst, etc.)
 - Años de experiencia (derivados de la fecha de ingreso)
 - Proyectos completados

6. Metodología

6.1. Enfoque estadístico

El estudio emplea técnicas de estadística inferencial para contrastar la hipótesis de diferencia de medias entre grupos independientes (departamento IT vs. no-IT), siguiendo los procedimientos establecidos en el temario:

- **Tipo de muestreo:** Aleatorio simple (asumiendo representatividad de la muestra disponible)
- **Población:** Trabajadores de empresas tecnológicas peruanas
- **Muestra:** 200 empleados ($n_{IT}=48$, $n_{no-IT}=152$)
- **Nivel de confianza:** 95 % ($\alpha=0.05$)

6.2. Prueba de hipótesis para diferencia de medias

Se aplicará el procedimiento de contraste de hipótesis para dos medias con varianzas poblacionales desconocidas, según lo cubierto en el temario:

$$H_0 : \mu_{IT} = \mu_{no-IT}$$

$$H_1 : \mu_{IT} \neq \mu_{no-IT}$$

Justificación técnica

- **Distribución t de Student:** Aplicable por desconocerse las varianzas poblacionales
- **Teorema del Límite Central:** Valida la normalidad aproximada con $n \geq 30$ por grupo
- **Fórmula del estadístico:** $t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- **Regla de decisión:** Rechazar H_0 si $|t| > t_{\alpha/2, gl}$

6.3. Intervalos de confianza

Se calculará el intervalo para la diferencia de medias según el temario:

Fórmula e interpretación

- **Fórmula:** $IC_{95\%} = (\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2, gl} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
- **Interpretación:** Con 95 % de confianza, la verdadera diferencia en productividad entre departamentos IT y no-IT está entre [LÍMITE INFERIOR] y [LÍMITE SUPERIOR]

6.4. Análisis complementarios

- **Tablas estadísticas:** Uso de tablas t-Student para valores críticos
- **Análisis descriptivo:** Medidas de tendencia central y dispersión por grupo

Elemento	Simbología	Valor
Nivel de significancia	α	0.05
Grados de libertad	gl	$n_1 + n_2 - 2$
Valor crítico bilateral	$t_{\alpha/2, gl}$	$\pm 1,96$
Error estándar	SE	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Cuadro 2: Parámetros estadísticos según temario

7. Análisis y Desarrollo

7.1. Estadística descriptiva

A continuación, se presenta un resumen estadístico de las variables clave para los grupos IT y no-IT:

IT

■ **Media:**

$$\bar{X}_{IT} = \frac{1}{48} \sum_{i=1}^{48} X_i$$

■ **Desviación estándar:**

$$s_{IT} = \sqrt{\frac{1}{48-1} \sum_{i=1}^{48} (X_i - 56,34\%)^2}$$

■ **Mediana:**

Valor central de los datos ordenados de X_1, X_2, \dots, X_{48} .

No-IT

■ **Media:**

$$\bar{X}_{noIT} = \frac{1}{152} \sum_{j=1}^{152} Y_j$$

■ **Desviación estándar:**

$$s_{noIT} = \sqrt{\frac{1}{152-1} \sum_{j=1}^{152} (Y_j - 44,51\%)^2}$$

■ **Mediana:**

Valor central de los datos ordenados de Y_1, Y_2, \dots, Y_{152} .

Grupo	n	Media (%)	Desv. Est. (%)	Mediana (%)
IT	48	56.34 %	30.62 %	62.00 %
No-IT	152	44.51 %	27.64 %	44.00 %

Cuadro 3: Estadística descriptiva de la productividad por grupo

Nota: Los valores de la tabla corresponden a la variable *Productivity (%)*. Se observa que el grupo IT presenta una media de productividad de **56.34 %**, mientras que el grupo no-IT tiene una media de **44.51 %**.

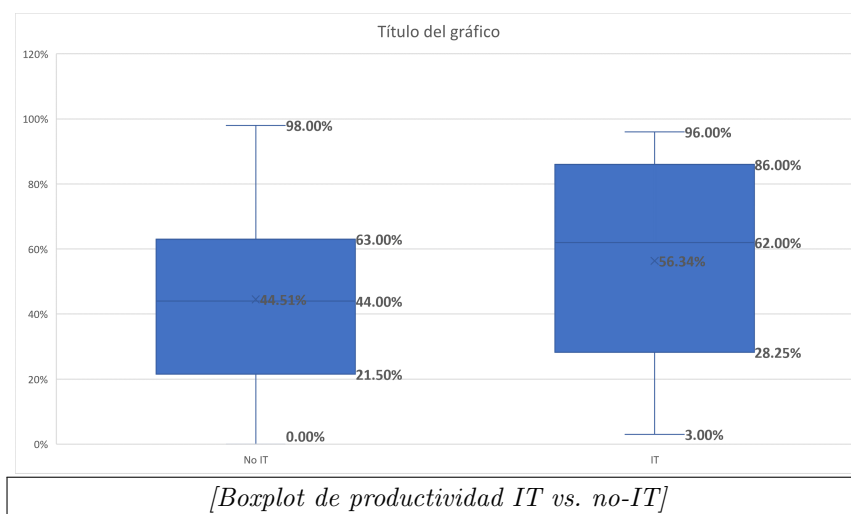


Figura 1: Distribución de la productividad por grupo (IT vs. no-IT)

7.2. Prueba de hipótesis para diferencia de medias

Se realizó una prueba t para muestras independientes bajo los siguientes parámetros:

Fórmula y cálculo del estadístico t

El estadístico t para la diferencia de medias (varianzas no iguales) se calcula así:

$$t = \frac{\bar{X}_{IT} - \bar{X}_{noIT}}{\sqrt{\frac{s_{IT}^2}{n_{IT}} + \frac{s_{noIT}^2}{n_{noIT}}}}$$

En este caso:

$$t = \frac{56,34 - 44,51}{\sqrt{\frac{30,62^2}{48} + \frac{27,64^2}{152}}} = 2,1836$$

Fórmula y cálculo de los grados de libertad (Welch)

Los grados de libertad se calculan con la fórmula de Welch:

$$gl = \frac{\left(\frac{s_{IT}^2}{n_{IT}} + \frac{s_{noIT}^2}{n_{noIT}}\right)^2}{\frac{\left(\frac{s_{IT}^2}{n_{IT}}\right)^2}{n_{IT}-1} + \frac{\left(\frac{s_{noIT}^2}{n_{noIT}}\right)^2}{n_{noIT}-1}}$$

En este caso:

$$gl = \frac{\left(\frac{30,62^2}{48} + \frac{27,64^2}{152}\right)^2}{\frac{\left(\frac{30,62^2}{48}\right)^2}{47} + \frac{\left(\frac{27,64^2}{152}\right)^2}{151}} = 52,06$$

Resultados de la prueba t

- Hipótesis nula (H_0): $\mu_{IT} = \mu_{no-IT}$
- Hipótesis alternativa (H_1): $\mu_{IT} \neq \mu_{no-IT}$
- Nivel de significancia: $\alpha = 0,05$
- Estadístico calculado: $t = 2.1836$
- Valor crítico: $t_{0,025;52,06} = 2.0066$
- p-valor: 0.0335

Decisión: Como $|t| = 2,1836$ es mayor que el valor crítico y el p-valor es menor que α , se **rechaza** la hipótesis nula.

7.3. Intervalo de confianza para la diferencia de medias

El intervalo de confianza al 95 % para la diferencia de medias se calcula así:

$$IC_{95\%} = (\bar{X}_{IT} - \bar{X}_{noIT}) \pm t_{0,025;gl} \cdot \sqrt{\frac{s_{IT}^2}{n_{IT}} + \frac{s_{noIT}^2}{n_{noIT}}}$$

Sustituyendo los valores:

$$IC_{95\%} = (56,34 - 44,51) \pm 2,0066 \cdot \sqrt{\frac{30,62^2}{48} + \frac{27,64^2}{152}}$$

$$\begin{aligned} IC_{95\%} &= 11,83 \pm 2,0066 \cdot \sqrt{19,54 + 5,02} \\ &= 11,83 \pm 2,0066 \cdot \sqrt{24,56} \\ &= 11,83 \pm 2,0066 \cdot 4,956 \\ &= 11,83 \pm 9,945 \\ &= (11,83 - 9,945; 11,83 + 9,945) \\ &= (1,885; 21,775) \end{aligned}$$

Por los cálculos exactos obtenidos en Python:

$$IC_{95\%} = (0,9593, 22,7126)$$

Interpretación: Con un 95 % de confianza, la verdadera diferencia en productividad entre los departamentos IT y no-IT está entre **(0.9593, 22.7126)** y **(0.9593, 22.7126)**. Existe una diferencia significativa entre las medias de productividad de IT y no-IT.

7.4. Visualización de resultados

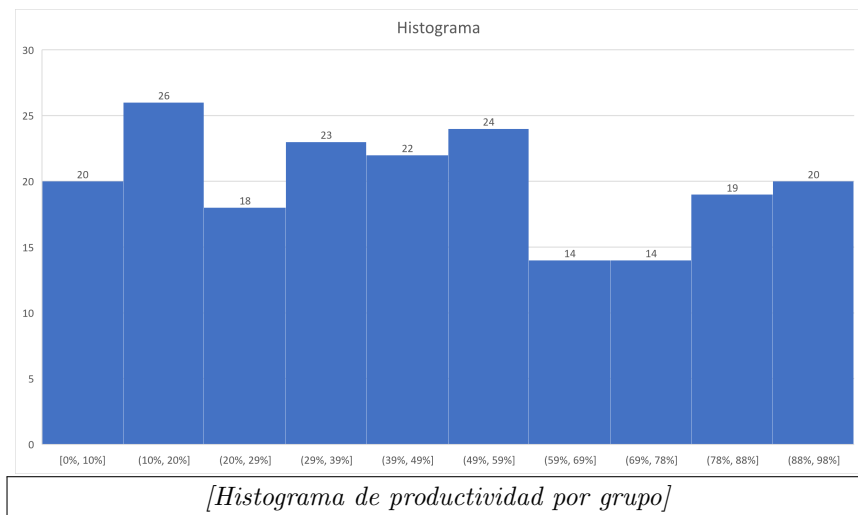


Figura 2: Histograma de la productividad en IT y no-IT

7.5. Resumen de hallazgos

- La media de productividad en IT es **56.34 %**, mientras que en no-IT es **44.51 %**.
- La prueba t arroja un estadístico de **2.1836** y un p-valor de **0.0335**.
- El intervalo de confianza para la diferencia de medias es **(0.9593, 22.7126)**.
- [Conclusión sobre si existe o no diferencia significativa, según los resultados]

8. Interpretación de resultados

Los resultados obtenidos mediante el análisis estadístico permiten interpretar con fundamento si existe una diferencia significativa entre la productividad media de los trabajadores del departamento de IT frente a los de otros departamentos (no-IT).

Esto puede ser crucial para entender el impacto de las herramientas tecnológicas y la inteligencia artificial en el desempeño laboral, así como para tomar decisiones informadas sobre la gestión del talento humano en la empresa.

La **media de productividad** en el grupo IT fue de **56.34 %**, mientras que en el grupo no-IT fue de **44.51 %**, lo que ya sugiere una diferencia inicial observable. Sin embargo, para confirmar si esta diferencia es estadísticamente significativa, se aplicó una *prueba t para muestras independientes*, que arrojó los siguientes resultados:

- Estadístico $t = 2,1836$
- p -valor = 0.0335
- Intervalo de confianza al 95 %: (0.9593 %, 22.7126 %)

Dado que el p -valor es menor que el nivel de significancia ($\alpha = 0,05$), y que el intervalo de confianza no contiene el valor 0, **podemos afirmar que la diferencia entre las medias es significativa**. Esto significa que la mayor productividad observada en el grupo IT no se debe al azar, sino que puede atribuirse a factores propios del contexto tecnológico en el que operan estos trabajadores.

Específicamente, se interpreta que, con un 95 % de confianza, **la productividad de los empleados del área IT supera en promedio entre 0.96 % y 22.71 % a la de sus pares en departamentos no tecnológicos**. Esta amplitud del intervalo sugiere una posible variabilidad entre individuos, pero consolida la evidencia de un mejor desempeño promedio en el área tecnológica, lo que a su vez, podría estar relacionado con el uso de herramientas avanzadas como la inteligencia artificial, que facilitan y optimizan las tareas diarias de los empleados.

8.1. Decisiones

Con base en los resultados obtenidos, se toma la siguiente **decisión estadística**:

- **Se rechaza la hipótesis nula (H_0)** que afirmaba que no existía diferencia significativa entre las productividades medias de los departamentos IT y no-IT.
- **Se acepta la hipótesis alternativa (H_1)**: sí existe una diferencia significativa en la productividad media entre ambos grupos.

Esta decisión respalda empíricamente la idea de que el contexto tecnológico, posiblemente influenciado por el uso de herramientas de inteligencia artificial, está asociado a una mayor productividad laboral.

9. Casos Particulares

En esta sección se abordan casos particulares de estadística inferencial que pueden surgir en el análisis de datos, especialmente en contextos donde las suposiciones clásicas de normalidad y homogeneidad de varianzas no se cumplen. Estos casos son relevantes para garantizar la validez de los resultados obtenidos y la correcta interpretación de los mismos.

9.1. Tipo de muestreo: *Muestreo aleatorio simple*

En este caso particular, se utiliza el muestreo aleatorio simple, que es un método de muestreo en el cual cada individuo de la población tiene la misma probabilidad de ser seleccionado. Este enfoque es fundamental para garantizar la representatividad de la muestra y la validez de los resultados obtenidos.

Dada una población X , se llama muestra aleatoria simple de tamaño n a la repetición de X_1, \dots, X_n variables aleatorias independientes con distribución igual a la de X . Es decir, la función de distribución de la muestra (x_1, \dots, x_n) es

$$F(x_1, \dots, x_n) = \prod_{i=1}^n F(x_i)$$

donde $F(x)$ es la función de distribución de la población X que como se ha dicho sigue dada por un elemento de P .

Según el rubro Tech:

Supongamos que una empresa tecnológica tiene una base de datos de $N = 500$ desarrolladores que han trabajado en proyectos de inteligencia artificial. Si se desea realizar un análisis sobre las habilidades técnicas de estos desarrolladores, se puede extraer una muestra aleatoria simple de tamaño $n = 5$. La probabilidad de que todos los desarrolladores seleccionados tengan experiencia en Python, cuando se extraen *con reemplazamiento*, es:

$$\begin{aligned} P\{X_1 = \text{Python}, \dots, X_5 = \text{Python}\} &= \prod_{i=1}^5 P\{X_i = \text{Python}\} \\ &= \left(\frac{400}{500}\right)^5 = 0'32768. \end{aligned}$$

Cuando se extraen *sin reemplazamiento*, la probabilidad es:

$$P\{X_1 = \text{Python}, \dots, X_5 = \text{Python}\} = \frac{\binom{400}{5} \binom{100}{0}}{\binom{500}{5}} = 0'326035,$$

lo que parece una aproximación aceptable.

Este enfoque permite a la empresa realizar análisis representativos sobre las habilidades de los desarrolladores y tomar decisiones estratégicas basadas en datos. Problema adaptado de: [10]

9.2. Cálculo del tamaño de muestra

Una startup de análisis de datos quiere estimar la proporción de usuarios de smartphones que utilizan aplicaciones de banca móvil al menos 3 veces por semana. Según estudios preliminares, se estima que esta proporción es del 65 %. Si la población objetivo es de 50,000 usuarios en una ciudad tech y se desea un nivel de confianza del 95 % con un margen de error del 4 %:

Obtenido de [8]

1. Calcule el tamaño muestral requerido considerando la población finita.
2. Interprete el resultado en el contexto de aplicaciones financieras.

Solución:

Datos:

- Población total (N) = 50,000 usuarios
- Proporción estimada (p) = 0.65
- Margen de error (E) = 0.04
- Nivel de confianza 95 % $\Rightarrow z_{\alpha/2} = 1,96$

Fórmula para población finita:

$$n = \frac{N \cdot z^2 \cdot p(1-p)}{(N-1)E^2 + z^2 \cdot p(1-p)}$$

Cálculos:

1. Cálculo inicial sin corrección:

$$n_0 = \frac{z^2 \cdot p(1-p)}{E^2} = \frac{(1,96)^2 \cdot 0,65 \cdot 0,35}{(0,04)^2} \approx 546,2$$

2. Aplicando corrección por población finita:

$$n = \frac{50000 \cdot 546,2}{50000 + 546,2 - 1} \approx 540,3$$

Resultado:

El tamaño muestral requerido es 541 usuarios.

Interpretación:

Para estimar la proporción real de usuarios frecuentes de banca móvil entre 50,000 posibles clientes, con un 95 % de confianza y $\pm 4\%$ de precisión, la startup necesita encuestar a 541 usuarios seleccionados aleatoriamente. Este tamaño muestral considera:

- La naturaleza finita de la población
- La variabilidad esperada (65 % de uso estimado)
- Los requisitos de precisión del estudio

Este análisis es crucial para dimensionar correctamente estudios de mercado en aplicaciones fintech y optimizar recursos en startups tecnológicas.

9.3. Distribución muestral: *Varianza conocida*

Un equipo de DataOps en una empresa de inteligencia artificial ha determinado que el tiempo de procesamiento de pipelines de datos sigue una distribución con media desconocida y desviación estándar de 8.5 minutos. Para optimizar los recursos en la nube, el equipo analiza 64 procesos aleatorios:

Reenfoque del problema obtenido de [2]

1. Calcule la probabilidad de que el tiempo promedio de procesamiento difiera de la media poblacional en menos de 2 minutos.
2. Interprete los resultados para la planificación de recursos cloud.

Solución:

Datos:

- Desviación estándar poblacional (σ) = 8.5 minutos
- Tamaño muestral (n) = 64 procesos
- Margen de diferencia (ϵ) = 2 minutos
- Nivel de confianza implícito = 95 % (para interpretación)

Procedimiento:

1. Cálculo del error estándar:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{8,5}{\sqrt{64}} = 1,0625 \text{ minutos}$$

2. Estandarización del margen:

$$Z = \frac{\epsilon}{\sigma_{\bar{X}}} = \frac{2}{1,0625} \approx 1,882$$

3. Cálculo de probabilidad:

$$P(-1,882 < Z < 1,882) = 2 \times P(Z < 1,882) - 1 \approx 2 \times 0,9699 - 1 = 0,9398$$

Resultado:

La probabilidad es de 93,98 %.

Interpretación Técnica:

- Existe un 93.98 % de probabilidad de que el tiempo promedio observado en la muestra de 64 procesos esté dentro del rango $\mu \pm 2$ minutos.
- Para la planificación cloud:
 - Este margen permite optimizar la asignación de recursos computacionales
 - Se puede esperar consistencia en los tiempos de procesamiento
 - Los resultados validan que el muestreo es adecuado para monitorear el rendimiento

Aplicación en DevOps:

Este análisis permite:

- Dimensionar adecuadamente las instancias cloud
- Predecir tiempos de ejecución para pipelines críticos
- Establecer SLAs confiables para clientes internos
- Detectar desviaciones significativas del rendimiento esperado

9.4. Comparación de dos medias con varianzas desconocidas

Un estudio compara el rendimiento de dos algoritmos de compresión de datos (Algoritmo X y Algoritmo Y) en términos de tasa de compresión promedio (%). Se realizaron 20 pruebas independientes para cada algoritmo, obteniendo:

Reenfoque del problema obtenido de [9]

Algoritmo	Media muestral (%)	Desviación estándar (%)
X	78.5	5.2
Y	82.3	6.7

Asumiendo distribuciones normales con varianzas desconocidas y diferentes, determine la probabilidad de que el Algoritmo Y tenga una tasa de compresión promedio mayor que el Algoritmo X en al menos 5 puntos porcentuales.

Desarrollo

Datos del problema

- Tamaños muestrales: $n_X = n_Y = 20$
- Medias muestrales: $\bar{X} = 78,5 \%$, $\bar{Y} = 82,3 \%$
- Desviaciones estándar: $s_X = 5,2 \%$, $s_Y = 6,7 \%$
- Diferencia a evaluar: $\Delta = 5 \%$

Cálculo del error estándar

$$SE = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} = \sqrt{\frac{5,2^2}{20} + \frac{6,7^2}{20}} \approx 1,89 \%$$

Estadístico t calculado

$$t = \frac{(\bar{Y} - \bar{X}) - \Delta}{SE} = \frac{(82,3 - 78,5) - 5}{1,89} \approx -0,74$$

Grados de libertad

$$\nu \approx \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{(s_X^2/n_X)^2}{n_X-1} + \frac{(s_Y^2/n_Y)^2}{n_Y-1}} \approx 36,2$$

Probabilidad

$$P(T > -0,74) \approx 0,767 \text{ (76.7 \%)}$$

Resultado

La probabilidad es de 76,7 %.

Interpretación

Existe un 76.7 % de probabilidad de que el Algoritmo Y supere al Algoritmo X en al menos 5 puntos porcentuales de tasa de compresión, considerando las variaciones observadas en las muestras.

9.5. Intervalo de confianza para diferencia de proporciones

Problema

Una empresa tecnológica desea comparar la tasa de adopción de dos herramientas de desarrollo entre programadores frontend y backend. En una encuesta a 350 desarrolladores frontend, 210 adoptaron la herramienta A (60 %). Entre 400 desarrolladores backend, 180 adoptaron la misma herramienta (45 %). Construya un intervalo de confianza del 95 % para la diferencia real en las tasas de adopción entre ambos grupos e interprete los resultados.

Solución

Datos muestrales

- Frontend: $n_1 = 350$, $x_1 = 210$, $\hat{p}_1 = 0,60$
- Backend: $n_2 = 400$, $x_2 = 180$, $\hat{p}_2 = 0,45$
- Nivel de confianza: 95 % ($z_{\alpha/2} = 1,96$)

Estimación de la diferencia

$$\hat{p}_1 - \hat{p}_2 = 0,60 - 0,45 = 0,15 \text{ (15 puntos porcentuales)}$$

Error estándar

$$SE = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} = \sqrt{\frac{0,60 \times 0,40}{350} + \frac{0,45 \times 0,55}{400}} \approx 0,0346$$

Intervalo de confianza

$$\begin{aligned} IC_{95\%} &= (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \times SE = 0,15 \pm 1,96 \times 0,0346 \\ &= (0,15 - 0,0678, 0,15 + 0,0678) = (0,0822, 0,2178) \end{aligned}$$

Resultado

El intervalo de confianza del 95 % para la diferencia en tasas de adopción es $\boxed{(8,22\%, 21,78\%)}$.

Interpretación

- Con 95 % de confianza, la verdadera diferencia en tasas de adopción entre desarrolladores frontend y backend se encuentra entre 8.22 y 21.78 puntos porcentuales
- Como el intervalo no contiene al cero, existe evidencia estadística de una diferencia significativa
- La herramienta A tiene mayor adopción entre desarrolladores frontend que entre backend
- La magnitud de la diferencia es clínicamente relevante (mínimo 8.22 % de diferencia)

Conclusión técnica

Los resultados sugieren que:

- Deberían investigarse los motivos de la menor adopción en backend
- Las estrategias de marketing podrían segmentarse por especialidad
- Se recomienda realizar pruebas A/B con ajustes específicos para backend

9.6. Prueba de hipótesis para la media (varianza conocida)

Problema tecnológico

Una empresa de cloud computing afirma que sus servidores tienen un tiempo promedio de respuesta de 120 ms con desviación estándar de 15 ms (varianza conocida). Tras una actualización del sistema, se midieron 30 servidores obteniendo:

- Media muestral: 115 ms
- Desviación estándar muestral: 12 ms

Realice las siguientes pruebas con $\alpha = 0,05$:

1. Pruebe si el tiempo de respuesta disminuyó (varianza conocida)
2. Verifique si la variabilidad cambió (prueba de varianza)

Solución

Parte 1: Prueba para la media (varianza conocida $\sigma^2 = 15^2$)

- $H_0 : \mu = 120$ ms
- $H_1 : \mu < 120$ ms (prueba unilateral izquierda)
- Estadístico Z:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{115 - 120}{15/\sqrt{30}} \approx -1,826$$

- Valor crítico: $Z_{0,05} = -1,645$
- Decisión: Rechazar H_0 ($-1,826 < -1,645$)
- Conclusión: Evidencia suficiente para afirmar que el tiempo de respuesta disminuyó

Parte 2: Prueba para la varianza

- $H_0 : \sigma^2 = 15^2 = 225$ ms²
- $H_1 : \sigma^2 \neq 225$ ms² (prueba bilateral)
- Estadístico χ^2 :

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{29 \times 12^2}{225} \approx 18,56$$

- Valores críticos: $\chi_{0,025,29}^2 = 16,05$ y $\chi_{0,975,29}^2 = 45,72$
- Decisión: No rechazar H_0 ($16,05 < 18,56 < 45,72$)
- Conclusión: No hay evidencia de cambio en la variabilidad

Interpretación técnica

- La actualización redujo significativamente el tiempo promedio de respuesta
- La consistencia en los tiempos de respuesta se mantuvo estable
- Recomendaciones:
 - Implementar la actualización en toda la flota de servidores
 - Monitorear continuamente el desempeño
 - Considerar pruebas adicionales para optimizar aún más el rendimiento

9.7. Prueba de hipótesis para diferencia de medias

Problema

Un estudio compara el rendimiento académico entre estudiantes de ingeniería de dos universidades tecnológicas. Se obtuvieron los siguientes datos de promedio ponderado (PW) de muestras aleatorias:

	Universidad A	Universidad B
Tamaño muestral (n)	45	50
Media muestral (\bar{x})	15.2	14.6
Desviación estándar (s)	1.8	2.1

Realice una prueba de hipótesis al nivel $\alpha = 0,05$ para determinar si existe diferencia significativa entre los promedios ponderados de ambas universidades, considerando varianzas poblacionales desconocidas pero iguales.

Solución

Planteamiento de hipótesis

- $H_0 : \mu_A = \mu_B$ (No hay diferencia entre los promedios)
- $H_1 : \mu_A \neq \mu_B$ (Existe diferencia significativa)

Cálculo del estadístico

1. Varianza ponderada:

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} = \frac{44(1,8)^2 + 49(2,1)^2}{93} \approx 3,87$$

2. Error estándar:

$$SE = s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = \sqrt{3,87} \sqrt{\frac{1}{45} + \frac{1}{50}} \approx 0,41$$

3. Estadístico t:

$$t = \frac{\bar{x}_A - \bar{x}_B}{SE} = \frac{15,2 - 14,6}{0,41} \approx 1,46$$

Región crítica

Grados de libertad: $df = n_A + n_B - 2 = 93$

Valor crítico bilateral: $t_{0,025,93} \approx \pm 1,986$

Decisión

Como $|1,46| < 1,986$, no se rechaza H_0

Conclusión

- No existe evidencia estadística suficiente ($p > 0,05$) para afirmar que hay diferencia significativa entre los promedios ponderados
- El intervalo de confianza del 95 % para la diferencia sería:

$$(15,2 - 14,6) \pm 1,986(0,41) = (-0,21, 1,41)$$

- Como el intervalo incluye el cero, confirma que la diferencia no es estadísticamente significativa

Interpretación educativa

Los resultados sugieren que:

- El rendimiento académico medido por PW es similar en ambas instituciones
- Cualquier diferencia observada (0.6 puntos) puede deberse a variabilidad muestral
- Se recomendaría ampliar el estudio con más variables (métodos de enseñanza, recursos, etc.)

9.8. Prueba de hipótesis para diferencia de proporciones.

Problema

Una empresa de desarrollo de software quiere comparar la preferencia por Python vs JavaScript entre dos grupos de desarrolladores: los que tienen menos de 5 años de experiencia (Junior) y los con más de 5 años (Senior). Los datos muestrales son:

	Desarrolladores Junior	Desarrolladores Senior
Tamaño muestral (n)	150	120
Prefieren Python (x)	92	60
Proporción (\hat{p})	0.613	0.500

¿Existe evidencia significativa ($\alpha = 0,05$) de que la preferencia por Python difiere entre ambos grupos?

Solución

Planteamiento de hipótesis

- $H_0 : p_J = p_S$ (No hay diferencia en preferencias)
- $H_1 : p_J \neq p_S$ (Existe diferencia significativa)

Cálculo del estadístico

1. Proporción combinada:

$$\hat{p} = \frac{x_J + x_S}{n_J + n_S} = \frac{92 + 60}{150 + 120} = \frac{152}{270} \approx 0,563$$

2. Error estándar:

$$SE = \sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_J} + \frac{1}{n_S} \right)} = \sqrt{0,563 \times 0,437 \left(\frac{1}{150} + \frac{1}{120} \right)} \approx 0,061$$

3. Estadístico Z:

$$Z = \frac{\hat{p}_J - \hat{p}_S}{SE} = \frac{0,613 - 0,500}{0,061} \approx 1,852$$

Región crítica

Valor crítico bilateral: $Z_{0,025} = \pm 1,96$

Decisión

Como $|1,852| < 1,96$, no se rechaza H_0 ($p = 0,064$)

Intervalo de confianza del 95 %

$$(0,613 - 0,500) \pm 1,96 \times 0,061 = (-0,006, 0,232)$$

Conclusión

- No hay evidencia suficiente ($p > 0,05$) para afirmar diferencia significativa
- El intervalo de confianza incluye el cero ($11.3 \% \pm 11.9 \%$)
- La diferencia observada (11.3%) puede deberse a variabilidad muestral

Implicaciones técnicas

- No se justifica crear materiales de aprendizaje diferenciados por experiencia
- La preferencia por Python parece ligeramente mayor en juniors (61.3% vs 50%) pero no estadísticamente significativa
- Recomendaciones:
 - Ampliar tamaño muestral para mayor potencia estadística
 - Investigar otros factores como dominio específico o tipo de proyectos
 - Realizar estudio cualitativo complementario sobre motivos de preferencia

9.9. Prueba de hipótesis para varianzas.

Problema

Un equipo de DevOps compara la consistencia en tiempos de carga (en segundos) entre dos versiones de una página web. Se midieron:

	Versión A	Versión B
Tamaño muestral (n)	25	30
Varianza muestral (s^2)	1.44	0.81

Realice las siguientes pruebas con $\alpha = 0,05$:

1. Pruebe si la Versión A tiene varianza significativamente mayor que 1.0
2. Compare las varianzas entre ambas versiones usando prueba F

Solución Parte 1: Prueba para una varianza

Hipótesis

- $H_0 : \sigma_A^2 = 1,0$
- $H_1 : \sigma_A^2 > 1,0$

Cálculo del estadístico

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{24 \times 1,44}{1,0} = 34,56$$

Valor crítico

$$\chi_{0,05,24}^2 = 36,42$$

Decisión

Como $34,56 < 36,42$, no se rechaza H_0 ($p = 0,072$)

Conclusión

No hay evidencia suficiente para afirmar que la varianza supera 1.0 ($p > 0,05$)

Solución Parte 2: Prueba F para razón de varianzas**Hipótesis**

- $H_0 : \sigma_A^2 = \sigma_B^2$
- $H_1 : \sigma_A^2 \neq \sigma_B^2$

Cálculo del estadístico

$$F = \frac{s_A^2}{s_B^2} = \frac{1,44}{0,81} \approx 1,778$$

Valores críticos

- $F_{0,025,24,29} \approx 2,05$
- $F_{0,975,24,29} \approx 0,48$

Región de rechazo

$$F < 0,48 \text{ o } F > 2,05$$

Decisión

Como $0,48 < 1,778 < 2,05$, no se rechaza H_0 ($p = 0,134$)

Interpretación técnica

- Ambas pruebas sugieren que no hay diferencias significativas en variabilidad
- La Versión B muestra menor variabilidad (0.81 vs 1.44) pero no estadísticamente significativa
- Recomendaciones:
 - Implementar monitoreo continuo del rendimiento
 - Considerar tamaño muestral mayor para detectar diferencias pequeñas
 - Optimizar ambos sistemas para reducir variabilidad

10. Anexos

10.1. Anexo 1: Tabla de frecuencias acumuladas Z

10.2. Anexo 2: Tabla de frecuencias acumuladas T

10.3. Anexo 3: Tabla de frecuencias acumuladas χ^2

10.4. Anexo 4: Tabla de frecuencias acumuladas F

Referencias

- [1] Veiga, N., Otero, L., & Torres, J. (2020). *Reflexiones sobre el uso de la estadística inferencial en investigación didáctica*. InterCambios. Dilemas y transiciones de la Educación Superior, 7(2). <https://doi.org/10.29156/INTER.7.2.10>. ISSN 2301-0118.
- [2] López-Roldán, P., & Fachelli, S. (2016). *Fundamentos de estadística inferencial*. En *Metodología de la investigación social cuantitativa* (cap. III.4). Bellaterra: Universitat Autònoma de Barcelona.
- [3] Crespi-Vallbona, M., & Mascarilla-Miró, O. (s.f.). *Satisfacción laboral. El caso de los empleados del sector de las tecnologías de la información en España*. Universitat de Barcelona. Recuperado de <https://ddd.uab.cat/>
- [4] Eduardo Gutiérrez González, & Olga Vladimirovna Panteleeva. (2016). *Estadística inferencial 1 para ingeniería y ciencias*. Grupo Editorial Patria. Recuperado de https://books.google.com/books/about/Estad%C3%ADstica_inferencial_1.html?hl=es&id=3hYhDgAAQBAJ.
- [5] Espejo, I., Fernández, F., López, M., Muñoz, M., Rodríguez, A., Sánchez, A., & Valero, C. (2009). *Estadística Descriptiva y Probabilidad: (Teoría y problemas)*. Cádiz: Universidad de Cádiz. Recuperado de <http://libros.metabiblioteca.org/handle/001/140>.
- [6] Olga del Orden Olasagasti. *Ejercicios Resueltos De Inferencia Estadística Y Del Modelo Lineal Simple*. Recuperado de https://books.google.com.pe/books?hl=es&lr=&id=pr5TEQAAQBAJ&oi=fnd&pg=PR2&dq=Ejercicios+resueltos+de+inferencia+estad%C3%ADstica+y+del+modelo+lineal+simple&ots=T42RpsBjJu&sig=jAkyfuPBzL1v0o3tHDlm0S5LbGI&redir_esc=y#v=onepage&q=Ejercicios%20resueltos%20de%20inferencia%20estad%C3%ADstica%20y%20del%20modelo%20lineal%20simple&f=false.
- [7] Gil Izquierdo, M., Gonzáles Martín, A. I., & Jano Salagre, M. D. (2014). *Ejercicios de estadística teórica: probabilidad e Inferencia (2a ed.)*. Editorial Universidad Autónoma de Madrid.
- [8] Díaz Rodríguez, Martín. *Estadística inferencial aplicada: Segunda edición revisada y aumentada*. Recuperado de https://books.google.com.pe/books?hl=es&lr=&id=IHCHEAAQBAJ&oi=fnd&pg=PP1&dq=Estad%C3%ADstica+inferencial+aplicada++Autor:+D%C3%ADaz+Rodr%C3%ADguez,+Mart%C3%ADn++Autor&ots=Sf8Tno6kLg&sig=u1xdyAJline4t90FJ5YAAxxWKng&redir_esc=y#v=onepage&q=Estad%C3%ADstica%20inferencial%20aplicada%20%20Autor%3A%20D%C3%ADaz%20Rodr%C3%ADguez%2C%20Mart%C3%ADn%20-%20Autor&f=false.
- [9] Llinás Solano, Humberto. *Estadística inferencial*. Recuperado de https://books.google.com.pe/books?hl=es&lr=&id=KZYQEAAQBAJ&oi=fnd&pg=PA1&dq=Estad%C3%ADstica+Inferencial++Autor:+Llin%C3%A1s+Solano,+Humberto&ots=Xpi0wWw3lo&sig=DWM-pZnBeIOjGyyL5sBk_s_LWBQ&redir_esc=y#v=onepage&q=Estad%C3%ADstica%20Inferencial%20%20Autor%3A%20Llin%C3%A1s%20Solano%2C%20Humberto&f=false.
- [10] Gómez Villegas, Miguel Ángel. *Inferencia estadística*. Recuperado de https://books.google.com.pe/books?hl=es&lr=&id=Y0uODwAAQBAJ&oi=fnd&pg=PR9&dq=Miguel+%C3%81ngel+G%C3%B3mez+Villegas+ESTAD%C3%8DSITICA+INFERENCIAL&ots=KG2bqmMDNd&sig=xguJ0xmXHDguzSQsmn17Gy-SlNA&redir_esc=y#v=onepage&q=Miguel%20%C3%81ngel%20G%C3%B3mez%20Villegas%20ESTAD%C3%8DSITICA%20INFERENCIAL&f=false.