

Ejercicio Unidad 2: Análisis de Resultados de Exámenes en una Universidad

Se analizarán calificaciones de estudiantes de un curso universitario. La población de referencia es de 1000 estudiantes y, para fines didácticos, las calificaciones se asumen como generadas a partir de una distribución normal con media aproximada de 75 y desviación estándar aproximada de 10.



Contexto del Ejercicio

Este ejercicio se enfoca en el análisis de calificaciones de estudiantes universitarios. Trabajaremos con una población de referencia de 1000 estudiantes, cuyas calificaciones siguen una distribución normal con:

- Media aproximada de 75 puntos
- Desviación estándar aproximada de 10 puntos

A través de este análisis, aplicaremos conceptos estadísticos fundamentales y aprenderemos a manipular datos en SPSS para obtener información relevante sobre el rendimiento académico.



Objetivos de Aprendizaje

Aplicación de la Distribución Normal

Calcular probabilidades y percentiles a partir de la variable Calificación utilizando la distribución normal.

Construcción de Variables Derivadas

Crear y recodificar variables en SPSS, distinguiendo correctamente entre variables tipo cadena y tipo numérico.

Representación Gráfica

Visualizar la distribución mediante histogramas y percentiles, y calcular probabilidades acumuladas.

Muestreo y Comparación

Realizar muestras de diferentes tamaños ($n=30, 100, 300$) y comparar estadísticos muestrales con parámetros esperados, aplicando el Teorema del Límite Central.

Documentación de Resultados

Elaborar interpretaciones claras de los resultados obtenidos en el análisis.

Requisitos Técnicos



Herramientas y Archivos Necesarios

- SPSS en español (menús en español)
- Archivo CSV con las siguientes características:
 - Separador: coma (,)
 - Punto decimal: punto (.)
 - Primera fila con nombres de variables (ID, Calificación)
 - Sin celdas vacías en las columnas

Estructura del Archivo Base

El archivo CSV de partida contiene únicamente dos variables:

Variable	Tipo	Descripción
ID	Entero	Valores del 1 al 1000 (sin decimales)
Calificación	Numérico	Valores con punto decimal (.)

Nota: Todas las demás variables se generarán dentro de SPSS como parte del ejercicio.

Bloque 1 — Archivos Base

Usted partirá de un archivo CSV con dos variables únicamente:



ID

Valores 1 a 1000 (entero, sin decimales).



Calificación

Valores numéricos (punto decimal ".").

Nota: No hay variables adicionales. Todas las demás se generarán dentro de SPSS como parte del ejercicio.

Bloque 2 — Importación y Verificación Inicial en SPSS

Paso 1. Abrir el archivo CSV en SPSS

1. Menú Archivo → Abrir → Datos...
2. En "Tipo de archivo" seleccione CSV (*.csv).
3. Busque y seleccione el archivo base calificaciones_1000.csv (o equivalente).
4. Se abre el Asistente de importación de texto:
 - Paso 1: confirme que el archivo es delimitado → Siguiente.
 - Paso 2: marque "Los nombres de las variables están en la primera fila" → Siguiente.
 - Paso 3: seleccione delimitador Coma (,) y punto decimal Punto (.) → Siguiente.
 - Paso 4: revise vista previa, luego Finalizar.

Paso 2. Verificar las variables en SPSS

1. Cambie a Vista de variables.
 - ID debe aparecer como Numérico, 0 decimales.
 - Calificación debe aparecer como Numérico, 2 decimales.
2. Cambie a Vista de datos.
 - Revise que existan 1000 filas con un ID único por cada estudiante.
 - Verifique que los valores de Calificación rondan entre ~40 y ~110 (propios de una normal $\mu=75$, $\sigma=10$).

Paso 3. Guardar archivo de trabajo en SPSS

1. Menú Archivo → Guardar como...
2. Tipo: Archivo de datos de SPSS (*.sav).
3. Nombre sugerido: EjercicioU2_Base.sav.

👉 A partir de este archivo, se desarrollarán todas las variables adicionales en los pasos siguientes.

Bloque 3 — Estadísticos Descriptivos, Detección de Outliers y Reemplazo



Paso 1

Correr estadísticos descriptivos básicos

1. Menú Analizar → Estadísticos descriptivos → Descriptivos...
2. Mueva la variable Calificación a "Variables".
3. Active Media, Desviación estándar, Mínimo y Máximo.
4. Aceptar.

👉 Con esta salida obtiene la media que servirá como valor de reemplazo, además de los indicadores generales.



Paso 2

Validar valores extremos con un diagrama de caja

1. Menú Gráficos → Diagramas de caja...
2. Seleccione Resúmenes de distintas variables y la opción Simple
3. Envíe la variable Calificación al recuadro Variables.
4. Aceptar.
5. Identifique el ID del valor más alto de los outlier inferiores. (ejemplo: 40)



Paso 3

Reemplazar outliers por el promedio

1. Menú Transformar → Calcular variable...
 - Variable de destino: Calif_Limpia
 - Expresión numérica: Calificación
 - → Aceptar
2. Nuevamente Transformar → Calcular variable...
 - Variable de destino: Calif_Limpia (misma variable).
 - Expresión numérica: escriba la media obtenida en el Paso 1
 - Pulse en Si... y escriba la condición: $\text{Calificación} < 40 \text{ OR } \text{Calificación} > 100$
 - → Aceptar.



Paso 4

Verificar resultados

- En Vista de datos, compare Calificación y Calif_Limpia.
- Los valores normales se mantienen, y los extremos fueron reemplazados por la media.
- Repita Gráficos → Diagramas de caja... → Simple → Resúmenes de distintas variables con Calif_Limpia para confirmar que ya no aparecen valores extremos.

Bloque 4 — Creación de Variables Derivadas en SPSS

Paso 1. Crear variable de Clasificación (tipo cadena)

1. Menú Transformar → Recodificar en variables diferentes...
2. Seleccione la variable Calif_Limpia y envíela a "Variable de entrada".
3. En Nombre de variable de salida, escriba Clasificación.
4. Haga clic en Valores antiguos y nuevos:
 - Rango: Lowest – 59.99 → Valor nuevo: **Reprobado**
 - Rango: 60 – 79.99 → Valor nuevo: **Aprobado**
 - Rango: 80 – Highest → Valor nuevo: **Sobresaliente**
 - Agregue cada recodificación y al final Continuar → Aceptar.
5. Vaya a Vista de variables y confirme:
 - Que Clasificación es de tipo Cadena.
 - Importante: cambie el Ancho de 8 a 15 para que los valores largos (Sobresaliente) no se trunquen.

Paso 2. Crear variable numérica para histogramas (CALIF_NUM)

👉 Como los histogramas requieren una variable numérica, se crea una versión codificada de la clasificación.

1. Menú Transformar → Recodificar en variables diferentes...
2. Variable de entrada: Calif_Limpia.
3. Variable de salida: CALIF_NUM.
4. En Valores antiguos y nuevos:
 - Rango: Lowest – 59.99 → Valor nuevo: 1
 - Rango: 60 – 79.99 → Valor nuevo: 2
 - Rango: 80 – Highest → Valor nuevo: 3
5. Aceptar.
6. En Vista de variables, defina etiquetas de valores:
 - 1 = Reprobado
 - 2 = Aprobado
 - 3 = Sobresaliente

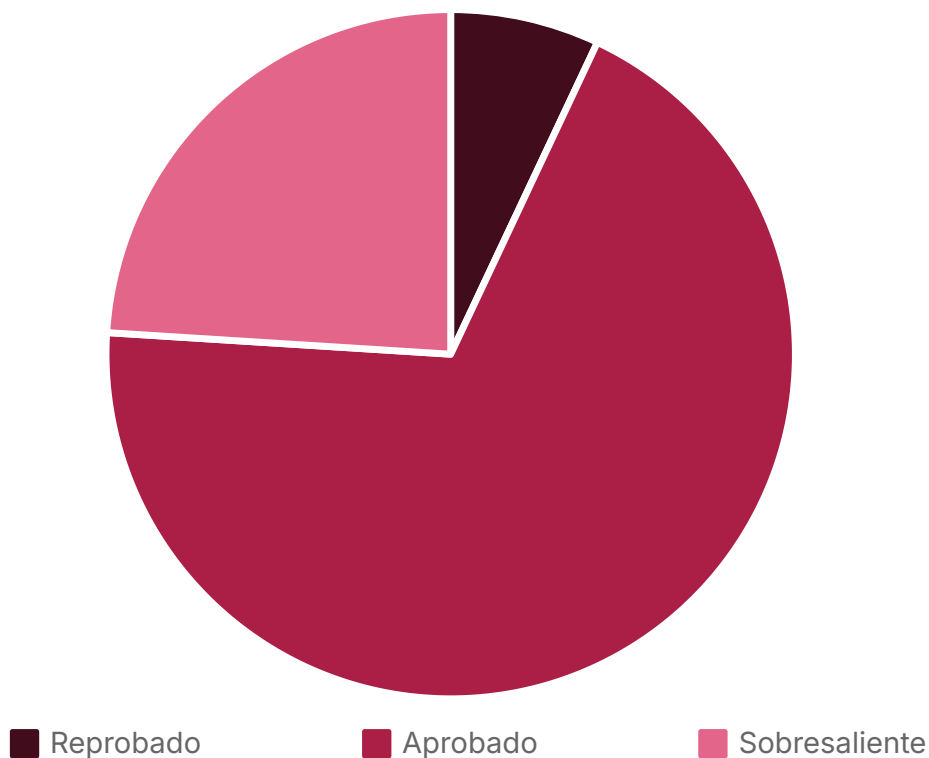
Paso 3. Verificación de variables creadas

- En Vista de datos, verifique que ahora tiene tres columnas relevantes:
 - Calif_Limpia (numérica, base corregida).
 - Clasificación (cadena, ancho 15).
 - CALIF_NUM (numérica).
- Compruebe que cada estudiante está correctamente categorizado según su calificación.

👉 Con esto queda preparada la base para:

- Usar Clasificación en análisis de frecuencias.
- Usar CALIF_NUM en histogramas y gráficos.

Distribución de Calificaciones por Categoría



Este gráfico muestra la distribución esperada de estudiantes según la clasificación de sus calificaciones. Con una distribución normal de media 75 y desviación estándar 10, aproximadamente:

- 7% de los estudiantes obtienen calificaciones por debajo de 60 (Reprobado)
- 69% obtienen entre 60 y 79.99 (Aprobado)
- 24% obtienen 80 o más (Sobresaliente)

Estos porcentajes se derivan de las propiedades de la distribución normal y serán verificados mediante los cálculos de probabilidad en SPSS.

Bloque 5 — Cálculo de Probabilidades y Percentiles en SPSS (Parte 1)

Paso 1. Calcular probabilidades con la distribución normal

👉 Usamos funciones de SPSS en Transformar → Calcular variable....

Probabilidad de aprobar (≥ 60 puntos)

- Variable de destino: P_aprueba
- Expresión numérica:
- $1 - \text{CDF.NORMAL}(60, 75, 10)$

Probabilidad de sobresaliente (≥ 80 puntos)

- Variable de destino: P_sobresaliente
- Expresión numérica:
- $1 - \text{CDF.NORMAL}(80, 75, 10)$

Probabilidad entre 70 y 85 puntos

- Variable de destino: P_70_85
- Expresión numérica:
- $\text{CDF.NORMAL}(85, 75, 10) - \text{CDF.NORMAL}(70, 75, 10)$

Acepte cada cálculo y revise en Vista de datos que aparezcan las nuevas variables con valores constantes (idénticos para todos los casos, porque son probabilidades teóricas).

Paso 2. Calcular percentiles 25, 50 y 75

1. Menú Analizar → Estadísticos descriptivos → Explorar....
2. Variable dependiente: Calif_Limpia.
3. En la pestaña Estadísticos, active la opción Percentiles.
4. En la pestaña Gráficos:
 - Marque la opción Histogramas.
 - No marque "Diagramas de tallo y hojas" (se desactiva para evitar ruido innecesario).
5. Aceptar.

👉 SPSS mostrará en la salida:

- Los percentiles 25, 50 y 75.
- Un histograma de la variable Calif_Limpia, útil para visualizar la forma de la distribución.

Paso 3. Calcular probabilidad acumulada hasta 70 puntos

1. Menú Transformar → Calcular variable....
2. Variable de destino: P_Acumulada_70.
3. Expresión numérica:
4. $\text{CDF.NORMAL}(70, 75, 10)$
5. Aceptar.

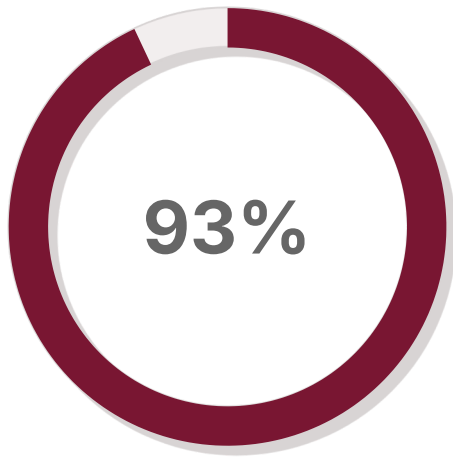
👉 Esta variable indica la probabilidad de que un estudiante tenga ≤ 70 puntos.

Paso 4. Representar gráficamente

1. Menú Gráficos → Histograma....
2. Seleccione la variable Calif_Limpia.
3. Active la casilla Mostrar curva de distribución normal.
4. Aceptar.

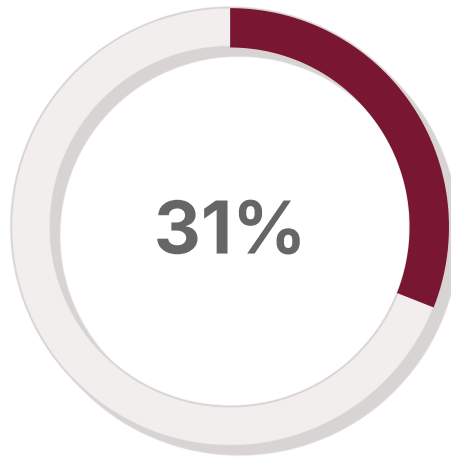
👉 Obtendrá el histograma de calificaciones con la curva normal superpuesta.

Interpretación de Probabilidades Calculadas



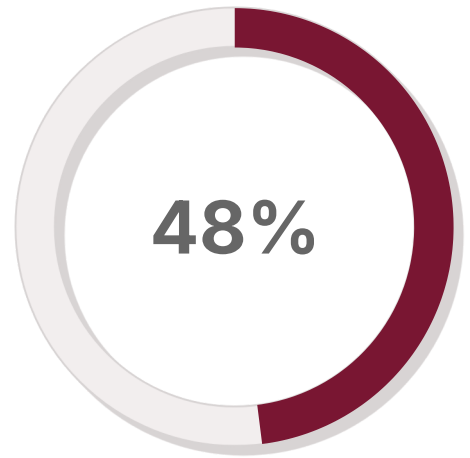
Probabilidad de aprobar

Aproximadamente el 93% de los estudiantes obtienen una calificación ≥ 60 puntos, lo que corresponde a las categorías "Aprobado" y "Sobresaliente".



Probabilidad de sobresaliente

Cerca del 31% de los estudiantes alcanzan una calificación ≥ 80 puntos, correspondiente a la categoría "Sobresaliente".



Probabilidad entre 70-85

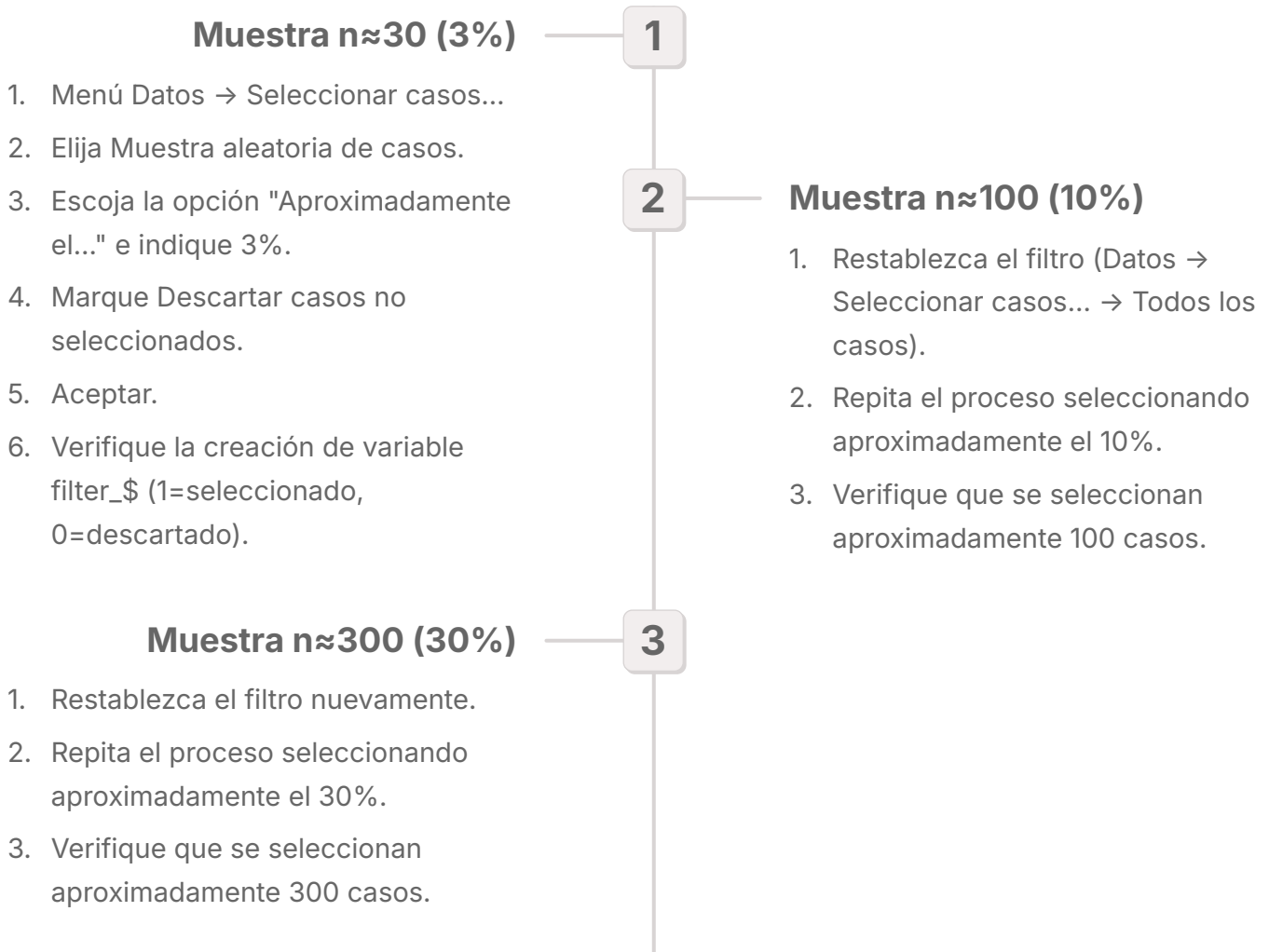
Aproximadamente el 48% de los estudiantes obtienen calificaciones entre 70 y 85 puntos, que abarca parte de "Aprobado" y parte de "Sobresaliente".

Estas probabilidades se calculan utilizando la función de distribución acumulativa normal (CDF.NORMAL) en SPSS, considerando que las calificaciones siguen una distribución normal con media 75 y desviación estándar 10.

La probabilidad acumulada hasta 70 puntos (P_Acumulada_70) nos indica que aproximadamente el **31%** de los estudiantes obtienen calificaciones iguales o inferiores a 70 puntos.

Bloque 6 — Muestreo y Comparación con Parámetros (Parte 1)

Paso 1. Seleccionar muestras aleatorias



👉 Repita este procedimiento tres veces (una para cada tamaño muestral), asegurándose de restablecer el filtro cada vez (Datos → Seleccionar casos... → Todos los casos).

Bloque 6 — Muestreo y Comparación con Parámetros (Parte 2)

Paso 2. Calcular estadísticos muestrales

1. Menú Analizar → Estadísticos descriptivos → Explorar...
2. Variable dependiente: Calif_Limpia.
3. En la pestaña Gráficos, seleccione únicamente Estadísticos.
4. Aceptar.

👉 SPSS mostrará en la salida solamente los estadísticos de interés:

- Media muestral
- Desviación estándar muestral
- Otros indicadores que acompañan la tabla

Paso 3. Comparar con la media poblacional ($\mu=75$)

1. Menú Analizar → Comparar medias → Prueba T para una muestra...
2. Variable de prueba: Calif_Limpia.
3. Valor de prueba: 75.
4. Aceptar.

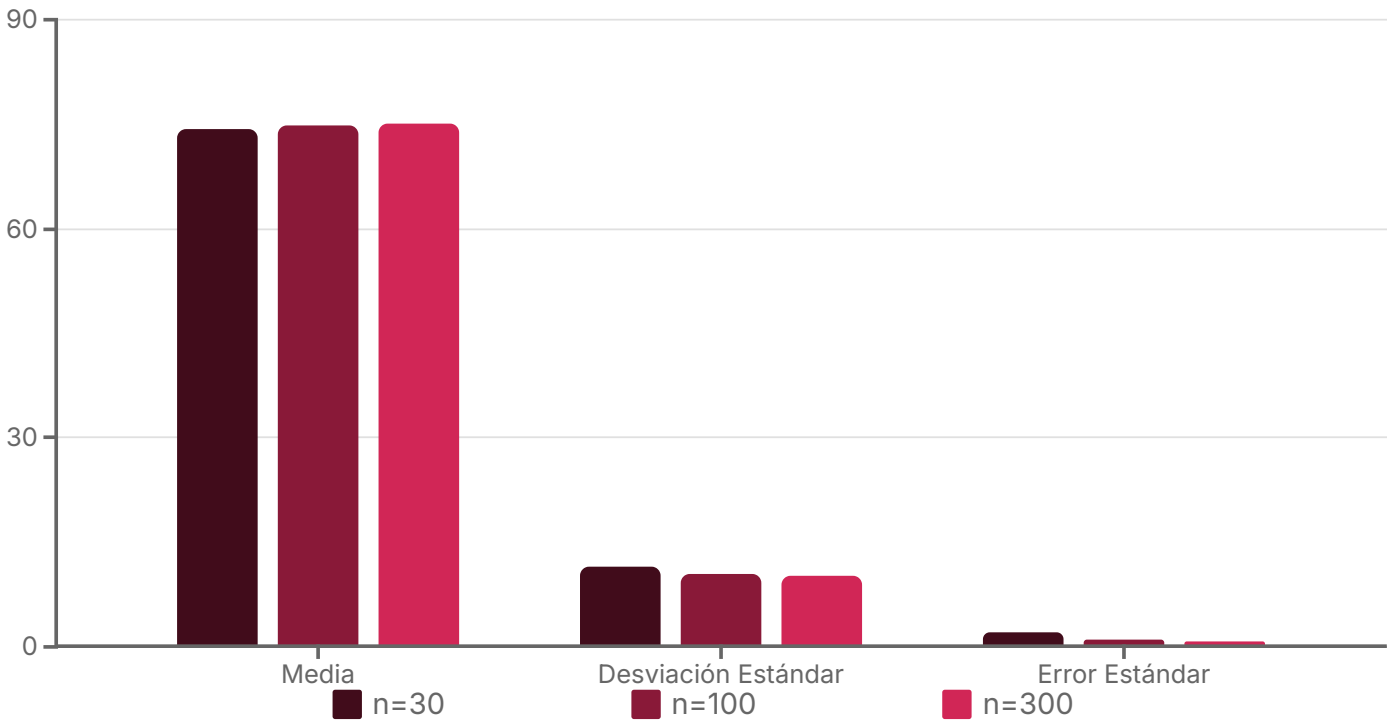
👉 SPSS genera para cada muestra:

- Media y desviación estándar observadas.
- Intervalo de confianza de la media.
- Resultado de la prueba t indicando si la diferencia con 75 es significativa.

Paso 4. Reflexión comparativa

- Observe cómo al aumentar el tamaño de la muestra (30 → 100 → 300), la media muestral se acerca más a 75.
- La desviación estándar muestral se estabiliza alrededor de 10.
- El error estándar de la media disminuye conforme crece n, lo que refleja el Teorema del Límite Central y la relación entre tamaño muestral y precisión.

Comparación de Resultados por Tamaño de Muestra



Esta gráfica ilustra cómo los estadísticos muestrales se aproximan a los parámetros poblacionales ($\mu=75$, $\sigma=10$) a medida que aumenta el tamaño de la muestra. Observe especialmente:

- El **error estándar** disminuye considerablemente al aumentar n , siguiendo la relación σ/\sqrt{n} .
- La **media muestral** se acerca más al valor poblacional de 75 conforme aumenta el tamaño de la muestra.
- La **desviación estándar** muestral se estabiliza cerca del valor poblacional de 10.

Estos resultados demuestran empíricamente el Teorema del Límite Central, que establece que la distribución de las medias muestrales se aproxima a una distribución normal con media igual a la media poblacional y error estándar igual a σ/\sqrt{n} .



Intervalos de Confianza por Tamaño de Muestra

Muestra $n=30$

Intervalo de confianza (95%): [70.0 - 78.4]

Amplitud: ± 4.2 puntos

Mayor incertidumbre debido al tamaño reducido de la muestra.

Muestra $n=100$

Intervalo de confianza (95%): [72.7 - 76.9]

Amplitud: ± 2.1 puntos

Reducción significativa de la incertidumbre al triplicar el tamaño muestral.

Muestra $n=300$

Intervalo de confianza (95%): [74.0 - 76.2]

Amplitud: ± 1.1 puntos

Alta precisión en la estimación, con un intervalo muy cercano al parámetro poblacional.

Observe cómo la amplitud del intervalo de confianza disminuye aproximadamente en proporción a la raíz cuadrada del aumento en el tamaño de la muestra, lo que confirma la relación teórica entre el error estándar y el tamaño muestral.

Bloque 7 — Entregables y Criterios de Evaluación

Entregables de la actividad

1

Archivo .sav con las variables generadas en SPSS

- ID
- Calificación (original)
- Calif_Limpia (con outliers reemplazados)
- Clasificación (cadena, ancho 15)
- CALIF_NUM (numérica 1=Reprobado, 2=Aprobado, 3=Sobresaliente)
- P_aprueba, P_sobresaliente, P_70_85, P_Acumulada_70 (probabilidades)

2

Gráficos exportados desde SPSS

- Diagrama de caja inicial (Calificación) y corregido (Calif_Limpia).
- Histograma con curva normal de Calif_Limpia.
- Histograma/Boxplot con categorías (CALIF_NUM).

1

Tablas de resultados

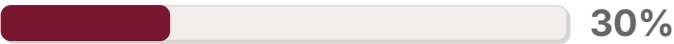
- Estadísticos descriptivos (media, desviación estándar, mínimos, máximos).
- Percentiles 25, 50, 75.
- Resultados de pruebas T para muestras de tamaño 30, 100 y 300.

2

Reporte en Word/PDF con

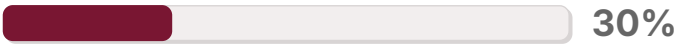
- Explicación de cómo se generaron las variables.
- Comparación entre resultados poblacionales y muestrales.
- Reflexiones sobre:
 - Cómo afecta el tamaño de muestra a la precisión.
 - Diferencias entre p unilateral y bilateral.
 - Manejo de outliers y sus implicaciones.

Criterios de Evaluación



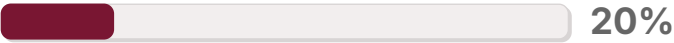
Precisión en los cálculos

Resultados correctos en probabilidades, percentiles y pruebas T.



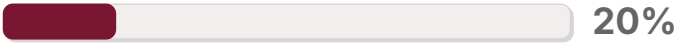
Interpretación de resultados

Conclusiones bien fundamentadas sobre probabilidades, muestreo y significancia estadística.



Calidad de las visualizaciones

Gráficos claros, correctamente etiquetados y relevantes para la interpretación.




Conclusiones y reflexiones

Explicaciones que relacionen teoría (distribución normal, TLC, p-valor) con los resultados prácticos.

La evaluación se enfocará tanto en la precisión técnica como en la capacidad para interpretar y comunicar los resultados estadísticos de manera clara y fundamentada.


Resumen del Ejercicio

En este ejercicio, hemos aplicado conceptos fundamentales de estadística utilizando SPSS para analizar calificaciones de estudiantes universitarios:



Análisis Descriptivo

Calculamos estadísticos básicos, identificamos y tratamos valores atípicos, y creamos variables derivadas para clasificar el rendimiento de los estudiantes.



Probabilidades

Aplicamos la distribución normal para calcular probabilidades de aprobar, obtener sobresaliente, y otras combinaciones relevantes para la evaluación académica.



Muestreo

Extrajimos muestras de diferentes tamaños para observar empíricamente el Teorema del Límite Central y la relación entre tamaño muestral y precisión de las estimaciones.

Este ejercicio ha proporcionado una experiencia práctica en el manejo de datos estadísticos, permitiendo comprender la aplicación de conceptos teóricos en situaciones reales de análisis de rendimiento académico.