******

**FACULTAD DE CIENCIAS EXACTAS INGENIERÍA Y AGRIMENSURA**

**TECNICATURA UNIVERSITARIA EN INTELIGENCIA ARTIFICIAL**

**2023**

**PROBABILIDAD Y ESTADÍSTICA**

**TRABAJO PRÁCTICO**

**N°1**

**Grupo N°:** 10

**Alumnos:**  Farias, Augusto

Demarre, Lucas

Donnarumma Cesar

Vercesi, Patricio

**ACTIVIDAD 1**

1. **Población**

Los alumnos que cursan Probabilidad y Estadística en la TUIA de la FCEIA de la UNR en Rosario, Argentina el primer cuatrimestre de 2023.

1. **Selección de variables y objetivos**
   1. **Formación académica previa**
      1. Tipo de la variable: Cualitativa medida en escala nominal.
      2. Objetivo del análisis: Conocer la proporción de los niveles de estudios que tienen los alumnos del curso.
   2. **¿Realiza ejercicio físico de manera frecuente?**
      1. Tipo de la variable: Cualitativa medida en escala nominal.
      2. Objetivo del análisis: Saber si la mayoría de las personas del curso ejercitan más de dos veces por semana.
   3. **Vitamina D en sangre (ng/mL)**
      1. Tipo de la variable: Cuantitativa continua medida en escala de razón
      2. Objetivo del análisis: Conocer el promedio de vitamina D en sangre de la población.
2. **Recategorización de respuestas:**

Con respecto a la variable del ejercicio físico y teniendo en cuenta nuestro objetivo podemos recategorizar las respuestas para que indique si la persona realiza ejercicio más de dos veces por semana de la siguiente manera:

|  |  |
| --- | --- |
| **Categoría original (Responde a ¿Realiza ejercicio físico de manera frecuente?)** | **Nueva categoría (Responde a ¿Realiza ejercicio físico más de dos veces por semana?)** |
| No | No |
| No, pero uso la calle recreativa los domingos | No |
| Sí, los sábados | No |
| Sí, más de dos veces por semana | Sí |
| Sí, más de 3 veces por semana | Sí |
| Sí, más de 4 veces por semana | Sí |

Para cada variable hay infinitud de objetivos que podrían intentar resolver, de ellos seleccionamos uno sólo.

1. **Análisis Descriptivo**

(Ejercicio resuelto en archivo .r adjunto)

Cuando analizamos la frecuencia de ejercicio tomando como umbral más de dos veces por semana podemos representarlo de la siguiente manera en una tabla:

Tabla, Calendario

Descripción generada automáticamente con confianza media

Y en gráfico de torta:

Gráfico, Gráfico circular

Descripción generada automáticamente

Concluyendo, como el 66% de los alumnos realizan ejercicio más de dos veces por semana, podemos determinar que la mayoría sí lo hacen y dar el objetivo por cumplido.

1. **Variables mal categorizadas y/o medidas**

Mes de nacimiento: la variable fue ingresada de múltiples maneras y la mejor forma de obtenerla hubiese sido proporcionando una lista de opciones con los 12 meses.

Formación académica previa: si el objetivo es obtener la calificación académica más alta obtenida por el estudiante, debería tener ese nombre ya que todo estudiante puede tener múltiples formaciones académicas previas. En caso contrario debería permitir seleccionar múltiples opciones.

Área de estudio previo: debería indicar de ser respondida únicamente si hubo un área de estudio previa, permitir elegir múltiples opciones simultáneamente, incluida la opción “Otra”, para abarcar todas las áreas de estudios y tener el mínimo de gente eligiendo la opción “Otra”.

Hobby: la variable es entendida por la mayoría de gente como “Hobbies que realiza actualmente”. Pero debería ser más explícita acerca de lo que pregunta: si debería completar con sólo uno o muchos, el favorito actualmente, los que ha tenido durante su vida, los que le gustaría hacer, etcétera. Y además de eso debería dar una lista de opciones de la cual elegir (dependiendo el objetivo quizás permitiendo elegir múltiples), porque sino da mucho trabajo luego recategorizar todas las posibles entradas que la gente puede hacer para poder analizarlas.

Grupo Sanguíneo: fue ingresado en múltiples formatos y la mejor forma hubiese sido proporcionando una lista de opciones con todas las posibilidades.

¿Le gusta el color verde?: si el objetivo era simplemente saber si a la persona le gusta el verde está bien. Es algo que se podría pedir con más detalle como con una escala de 3/ 5 niveles de estar de acuerdo con la frase “Me gusta el verde”, siendo el del medio el de “Más o menos”, el mínimo el de “Desacuerdo/ Muy en desacuerdo” y el máximo el de “De acuerdo/ Muy de acuerdo”. Si el objetivo era más complejo entonces la pregunta y opciones seguramente son limitadas para resolverlo.

¿Realiza ejercicio físico de manera frecuente?: En primer lugar es una pregunta muy subjetiva y poco exacta, cuya respuesta puede variar considerablemente según la interpretación de la misma. Y en segundo lugar tiene opciones que no abarcan todas las posibilidades ni se excluyen entre sí, por lo tanto los alumnos pueden no tener una opción que los represente o tener múltiples y no saber cuál elegir. Las maneras más apropiadas de haber tomado las respuestas hubiesen sido permitiendo un ingreso de un número entre 0 y 7 a la pregunta “¿Cuántos días realiza ejercicio físico por semana?” o un ingreso de un entero positivo o el 0 a la pregunta “¿Cuántas horas de ejercicio físico realiza por semana?”. Esas soluciones servirían si el objetivo es determinar con cuanta frecuencia los estudiantes realizan ejercicio. Si el objetivo es determinar sobre qué días de la semana realizan ejercicio, entonces se debería cambiar la variable a algo como “Días de la semana que realiza ejercicio físico”, dar como opciones los días de la semana y permitir elegir múltiples simultáneamente.

**ACTIVIDAD 2**

**Población:** lote de 80 placas de madera.

Aclaración: se considera como población al lote de 80 maderas ya que la distribuidora, desea saber si puede enviar este lote en particular a un cliente estricto.

**Unidad elemental:** cada una de las 80 placas de madera del lote.

**Variable:** número de defectos (variable cuantitativa discreta medida en escala de razón).

**Parámetro de interés:** media (promedio).

**Objetivo:** conocer el promedio de defectos por placa en cada lote.

**Tabla de distribución de frecuencias:**

**Una pantalla de un video juego

Descripción generada automáticamente con confianza media**

**Media:** 1.65 defectos por placa de madera.

**Informe:**

Observando la tabla de distribución de frecuencias podemos decir que:

1. Las placas con ningún defecto (solo 10 unidades) representan el 12.5% del total, lo que invirtiendo el punto de vista nos quiere decir que el 87.5% de las placas tienen defectos.
2. La mayoría de las placas (un total de 29) tienen 1 solo defecto, lo cual representa el 36.25% del lote. Seguidas muy de cerca por las que tienen 2 defectos (un total de 26) que representan el 32.5% del total.
3. Teniendo en cuenta el punto anterior estamos hablando de que el 68.75% (más de la mitad de las placas) tienen entre 1 y 2 defectos.
4. Por otro lado, resulta muy bajo el número de placas con 4 y 5 defectos (4 placas, y 1 respectivamente) ya que solo representan el 6.25% del total, lo que implica que el 93.75% de las placas tienen 3 o menos defectos.
5. También podemos decir que en promedio los lotes tienen 1,65 defectos por placa de madera.

Al tratarse de un estudio de tipo exhaustivo (nos interesaba conocer si este lote en particular cumplía con las especificaciones que un determinado cliente requería) podemos considerar que las conclusiones del informe anterior son definitivas ya que al aplicarse al conjunto de la población no hace falta utilizar herramientas de tipo inferencial.

● ¿Son aptas las maderas para el envío?

Debido a los resultados obtenidos en el estudio podemos decir que las maderas del lote no son aptas para el envío al no cumplir con los requerimientos del cliente. El numero promedio defectos por madera del lote es de 1,65 defectos por placa y el cliente necesitaba que sea menor a 1,2 por unidad.

**Gráfico de bastones:**

**Gráfico

Descripción generada automáticamente**

**Grafico escalonado:**

**Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente**

**Medidas de resumen:**

**De localización:**

**Moda:** 1 defecto por placa.

**Interpretación:** La mayoría de las placas tienen 1 defecto.

**Mediana:** 2 defectos por placa.

**Interpretación:** El 50% de las placas tienen 2 defectos o menos.

**De dispersión:**

**Rango:** 5

**Interpretación:** La mayor diferencia entre la cantidad de defectos que tienen dos placas es 5.

**Desvío estándar**: 1,091

**Interpretación:** En promedio la cantidad de defectos se desvía en 1,091 defectos de la media aritmética.

Para otros objetivos diferentes podría haber sido útil haber registrado las longitudes de las placas de madera (para conocer qué tan uniformes son entre sí) o las fechas de producción (para saber si se están vendiendo primero las más antiguas).

**ACTIVIDAD 3**

**OPCIÓN B (Columna C):**

1. **Problema**: una empresa tiene que saber si los cortes realizados a piezas de plástico en su proceso de producción están dentro de los parámetros especificados en los planos.
2. **Población**: 150 piezas de plástico.
3. **Unidad elemental**: cada pieza de plástico.
4. **Variable**: profundidad del corte en cm. (Cuantitativa continua medida en escala de razón)
5. **Parámetros de interés**: proporción (π) de piezas en los que su corte en profundidad sea inferior a 7 cm.

1. **Objetivo**: detectar que porcentaje de las 150 piezas tiene un corte por debajo de los 7 cm y cuáles son los cortes que se hacen con mayor frecuencia.
2. **Conclusiones preliminares**: tras haber finalizado el análisis de muestreo sobre el proceso de producción llevado a cabo, podemos determinar que, de las 150 piezas, el 79% tiene un corte de profundidad dentro de los parámetros aceptados, es decir, un corte inferior a 7 cm.

Gráfico, Gráfico circular

Descripción generada automáticamente

Un análisis más exhaustivo nos mostró también que los cortes que se repiten con mayor frecuencia son los cortes más cercanos e inferiores a 7 cm, estos rondan  un rango entre 6.5 a 7 cm, teniendo como resultados atípicos cortes no tan profundos y cortes muy profundos.

Gráfico, Histograma

Descripción generada automáticamente

Decimos que nuestras conclusiones son preliminares porque solo aplicamos nuestros métodos a una muestra de nuestra población total. Esto significa que, si hiciéramos lo mismo, pero con el total de nuestra población, las conclusiones pueden cambiar drásticamente dependiendo de los datos.