**METODOLOGÍA PARA EL ANÁLISIS DE ENCUESTA DE CAPACIDADES DE INNOVACIÓN EMPRESARIALES**

1. **Introducción**

Al aplicar herramientas estadísticas, es fundamental establecer la naturaleza de las variables involucradas. Esta distinción resulta crucial porque no todas las variables pueden analizarse mediante los mismos métodos estadísticos. En particular, se debe determinar si una variable es cuantitativa o cualitativa, dado que esta clasificación condiciona las técnicas apropiadas para su análisis.

Una variable cualitativa representa una característica o atributo que clasifica a cada unidad de análisis en una categoría. En el contexto de este estudio, todas las variables recolectadas mediante el instrumento de medición corresponden a respuestas en escalas tipo Likert, con opciones como “Nunca”, “Raramente”, “A veces”, “Muy a menudo” y “Siempre”.

Las variables cualitativas se dividen habitualmente en nominales y ordinales. Las nominales no tienen un orden inherente entre sus categorías (por ejemplo, el color de ojos: azul, marrón, verde). En cambio, las ordinales sí presentan una jerarquía clara entre categorías, como ocurre con escalas de satisfacción o desempeño (por ejemplo: Malo, Regular, Bueno, Muy Bueno, Excelente). Aunque estas escalas asignan valores numéricos a las categorías, los números son meramente indicativos del orden y no implican intervalos iguales entre respuestas.

A pesar de que en la práctica es común codificar las respuestas ordinales con valores numéricos consecutivos (por ejemplo, 1 a 5), esta codificación no convierte a la variable en cuantitativa. Así, calcular promedios o aplicar métodos estadísticos convencionales como regresión lineal o análisis de correlación puede conducir a interpretaciones erróneas, ya que estos métodos asumen intervalos constantes y una escala continua.

En este contexto, resulta necesario utilizar modelos diseñados explícitamente para respuestas categóricas ordenadas, como los modelos de la Teoría de Respuesta al Ítem (IRT). Esta familia de modelos permite estimar una variable latente no observable —como la propensión de una empresa a adoptar prácticas asociadas al sensing, seizing o transformation— a partir de patrones de respuesta a ítems observables con escalas ordinales.

En particular, el Modelo de Respuesta Graduada (GRM), propuesto por Samejima (1969), es uno de los más adecuados para este tipo de datos. El GRM permite modelar la probabilidad de que una entidad seleccione una determinada categoría (o una superior) en función de su nivel en la variable latente y de las características del ítem (como su dificultad y capacidad de discriminación).

Este modelo ha sido ampliamente adoptado no solo en psicometría, sino también en áreas como educación, salud pública, análisis organizacional y marketing (Embretson & Reise, 2000; De Ayala, 2009). En nuestro caso, se utiliza el GRM para estimar una variable continua latente que resume la propensión de cada empresa a desarrollar prácticas específicas en una dimensión dada. Esta transformación permite aplicar posteriormente métodos estadísticos convencionales, como pruebas estadísticas de comparación de medias y algoritmos supervisados o no supervisados.

En resumen, el uso del GRM constituye una alternativa teórica y empíricamente sólida frente a la práctica incorrecta de promediar ítems ordinales o aplicar directamente técnicas multivariadas a variables categóricas. Al reconocer la naturaleza ordinal de los datos y modelar explícitamente la estructura latente subyacente, el GRM produce mediciones más precisas, comparables y psicométricamente válidas.

1. **Metodología**

El GRM asume que cada ítem tiene un conjunto de categorías de respuesta ordenadas y que la probabilidad de responder en una categoría determinada o superior depende de un rasgo latente θ. Para un ítem con categorías ordenadas, el modelo especifica umbrales y un parámetro de discriminación .

**Formulación Matemática:**

Para el ítem , la probabilidad de que una empresa con rasgo latente θ responda en la categoría o superior se expresa como:

La probabilidad de responder exactamente en la categoría es:

**Componentes del Modelo:**

* θ: variable latente — representa el nivel subyacente del rasgo que se desea medir (por ejemplo, la capacidad de *sensing* de una empresa).
* : parámetro de discriminación del ítem — indica qué tan bien el ítem distingue entre entidades con diferentes niveles de θ.
* : umbral de dificultad para el ítem y la categoría — representa el nivel de θ requerido para que una entidad tenga al menos una probabilidad del 50 % de seleccionar la categoría o una superior.

Este enfoque permite usar ítems con diferentes niveles de discriminación y dificultad, generando una medición más flexible e informativa.

Con el fin de facilitar la comprensión de los parámetros de discriminación y dificultad, a continuación, se plantean dos ejemplos representativos:

Considérese una encuesta aplicada a empresas que contiene las siguientes preguntas:

* **Ítem A**: "¿Con qué frecuencia su empresa revisa las tendencias del mercado?"
* **Ítem B**: "¿Con qué frecuencia su empresa aplica modelos predictivos avanzados de inteligencia artificial para anticiparse a la demanda?"

Ambos ítems utilizan una escala ordinal de cinco puntos, donde 1 representa "Nunca" y 5 representa "Siempre".

Desde el punto de vista del parámetro de dificultad:

* El Ítem A tiende a recibir respuestas afirmativas (puntuaciones altas) por parte de la mayoría de las empresas, incluso aquellas con un nivel incipiente en capacidades de sensing. Esto sugiere que los umbrales asociados a este ítem se ubican en valores bajos del continuo , por lo que se clasifica como un ítem de baja dificultad.
* El Ítem B, en contraste, exige un mayor nivel de desarrollo organizacional para ser respondido afirmativamente. Es esperable que solo las empresas más avanzadas en prácticas analíticas puedan seleccionar categorías superiores en este ítem, lo que implica umbrales ubicados en niveles altos de . Por tanto, se clasifica como un ítem de alta dificultad.

Respecto al parámetro de discriminación:

* Si el Ítem B permite diferenciar de manera precisa a las empresas con altos niveles de aquellas con bajos niveles —es decir, si pequeñas variaciones en implican cambios significativos en la probabilidad de seleccionar categorías altas—, entonces se le atribuye una alta discriminación.
* Por el contrario, si dicha probabilidad cambia de forma paulatina y empresas con distintos niveles de tienden a responder de manera similar, el ítem presenta una baja capacidad de discriminación.

Este ejemplo ilustra cómo el GRM permite capturar no solo la "exigencia" asociada a cada ítem (su dificultad), sino también su capacidad para aportar información útil en la medición de la variable latente (su discriminación).

1. **Nota metodológica**

Una de las principales limitaciones metodológicas del presente estudio radica en el tamaño muestral empleado para la estimación del modelo de Respuesta Graduada (GRM, por sus siglas en inglés). Con una muestra de 56 participantes, se reconoce que las estimaciones obtenidas pueden estar sujetas a restricciones inherentes al ajuste de modelos IRT (Item Response Theory), en particular cuando el número de observaciones es limitado. La literatura especializada ha documentado que tamaños muestrales pequeños pueden comprometer la estabilidad y precisión de los parámetros estimados en modelos IRT, incrementando la variabilidad de los errores estándar, afectando la convergencia del algoritmo de estimación y dificultando la generalización de los resultados (de Ayala, 2009; Baker & Kim, 2017).

No obstante, con el fin de mitigar y evaluar este posible efecto, se realizó un análisis complementario de los errores estándar asociados a las puntuaciones latentes obtenidas en cada una de las dimensiones estimadas por el GRM (Sensing, Seizing y Transformation). Los resultados mostraron que los errores estándar promedio se mantuvieron en niveles bajos (entre 0.120 y 0.132), con intervalos de confianza del 95% relativamente estrechos. En particular, la dimensión *Transformation* presentó el menor coeficiente de variación del error estándar (8.2%), lo que sugiere una alta estabilidad en la estimación de los puntajes latentes. Aunque en la dimensión *Seizing* se identificó una mayor dispersión relativa (coeficiente de variación del 31.1%), los niveles absolutos de error estándar no resultaron preocupantes en la mayoría de los casos.

En conjunto, si bien el tamaño muestral constituye una limitación relevante, los indicadores analizados respaldan una aceptable precisión de las estimaciones obtenidas a través del modelo GRM. Se recomienda, no obstante, replicar este análisis con muestras más amplias que permitan validar la robustez de los hallazgos y reducir potenciales sesgos derivados de la limitada cantidad de observaciones.

1. **Resultados**

En esta sección se presentarán distintos análisis bivariados entre las variables latentes calculadas para cada una de las dimensiones —sensing, seizing y transformation— y las variables de caracterización de las empresas (tamaño de la empresa, ventas del último año, industria a la que pertenece la compañía, etc.), lo anterior, con el fin de obtener algunas inferencias.

* 1. **Análisis Univariados**

**Tamaño de la compañía**

Se presenta a continuación la distribución porcentual de los tamaños de empresa en la muestra analizada.

Gráfico, Gráfico circular

El contenido generado por IA puede ser incorrecto.

En la gráfica anterior se observa que, dentro de la muestra recolectada, la mayoría de las empresas corresponde al grupo de tamaño mediano, representando aproximadamente el 54 % (n = 15). Por su parte, el 18 % (n = 5) corresponde a empresas pequeñas, mientras que tanto las empresas grandes como las microempresas representan cada una el 14 % (n = 4).

* 1. **Análisis bivariados para sensing**

**Tamaño de la compañía**

Para llevar a cabo este análisis, se plantea la siguiente pregunta central, que orienta el desarrollo del estudio: ¿existen diferencias estadísticamente significativas en las variables latentes de sensing entre empresas grandes, medianas, pequeñas y microempresas? En otros términos, **¿la propensión al sensing podría variar en función del tamaño de la empresa?**

Esta pregunta puede abordarse mediante un ejercicio estadístico de comparación de medias de las variables latentes de sensing entre los distintos grupos empresariales definidos según su tamaño.

Antes de responder a la pregunta central formulada previamente, es necesario presentar los resultados de una serie de pruebas estadísticas preliminares. En particular, se evalúan dos supuestos fundamentales para la comparación de medias mediante ANOVA: normalidad (es decir, que las variables latentes de sensing sigan una distribución normal dentro de cada grupo) y homocedasticidad (igualdad de varianzas entre los grupos).

Para la verificación del supuesto de normalidad se emplea la prueba de Shapiro-Wilk. A continuación, se reportan los resultados de la prueba de Shapiro-Wilk (SW) aplicada a cada grupo de empresas, con el objetivo de evaluar el cumplimiento del supuesto de normalidad en las variables latentes de sensing:

| **GRUPO** | **ESTADÍSTICO SW** | **P-VALOR** | **DECISIÓN** |
| --- | --- | --- | --- |
| **Grande** | 0,9166 | 0,5178 | No se rechaza la hipótesis nula. Los datos siguen una distribución normal. |
| **Mediana** | 0,9561 | 0,6249 | No se rechaza la hipótesis nula. Los datos siguen una distribución normal. |
| **Pequeña** | 0,9295 | 0,5929 | No se rechaza la hipótesis nula. Los datos siguen una distribución normal. |
| **Micro** | 0,9628 | 0,7965 | No se rechaza la hipótesis nula. Los datos siguen una distribución normal. |

A partir de los resultados reportados en la tabla anterior, es posible observar que, en cada uno de los grupos, el valor p asociado a la prueba de Shapiro-Wilk supera el umbral de significancia convencional (p > 0.05), por lo que no se rechaza la hipótesis nula. Esto indica que los datos de las variables latentes de *sensing* siguen una distribución normal en todos los grupos analizados.

Posteriormente, para evaluar el supuesto de homocedasticidad —es decir, la igualdad de varianzas entre los grupos definidos por tamaño de empresa— se emplea la prueba de Levene (en adelante, LV).

A continuación, se reportan los resultados de la prueba de Levene (LV) aplicada a los grupos definidos según el tamaño de empresa, con el propósito de evaluar la igualdad de varianzas en las variables latentes de *sensing*:

|  |  |  |
| --- | --- | --- |
| **ESTADÍSTICO LV** | **P-VALOR** | **DECISIÓN** |
| 0,9367 | 0,4383 | No se rechaza la hipótesis nula. Las varianzas de los grupos poblaciones son iguales. |

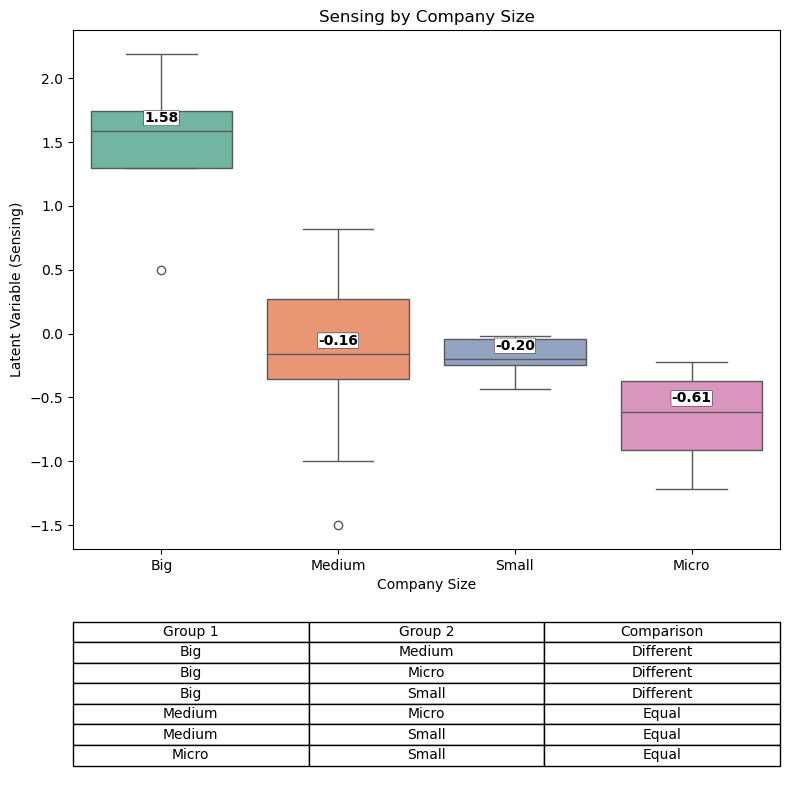
A partir de los resultados reportados en la tabla anterior, es posible observar que, conforme al valor p obtenido, no se rechaza la hipótesis nula de homocedasticidad. Esto permite concluir que las varianzas entre los grupos de empresas, definidos según su tamaño, pueden considerarse iguales, cumpliéndose así el supuesto de igualdad de varianzas.

Una vez verificados los supuestos de normalidad y homocedasticidad, se procede con la aplicación del análisis de varianza (ANOVA) para determinar si existen diferencias significativas entre las medias de los grupos. A continuación, se reportan los resultados de la prueba de ANOVA aplicada a las variables latentes de *sensing*, con el fin de evaluar si existen diferencias estadísticamente significativas entre los grupos de empresas según su tamaño:

|  |  |  |
| --- | --- | --- |
| **ESTADÍSTICO F** | **P-VALOR** | **DECISIÓN** |
| 12,5298 | 0 | Se rechaza la hipótesis nula. Al menos una de las medias poblaciones difiere |

A partir de los resultados reportados en la tabla anterior, es posible observar que, conforme al valor p obtenido, se rechaza la hipótesis nula. Esto permite concluir que al menos una de las medias poblacionales de la variable latente de *sensing* difiere significativamente entre los grupos de empresas definidos según su tamaño.

Como siguiente paso, se aplica la prueba post-hoc de Tukey HSD, la cual permite identificar específicamente entre qué pares de grupos se presentan diferencias estadísticamente significativas en las medias de *sensing*.



En la gráfica anterior se observa que las empresas grandes de la muestra recolectada presentan los valores más altos en la variable latente de sensing, lo que sugiere una mayor propensión al desarrollo de capacidades asociadas al sensing en comparación con los demás grupos: medianas, pequeñas y microempresas. No obstante, el objetivo central del análisis no es únicamente identificar si esta diferencia está presente en la muestra, sino determinar si dicha diferencia puede generalizarse a la población. Es decir, interesa establecer si, a nivel poblacional, las empresas grandes presentan una mayor propensión al sensing en comparación con los otros grupos definidos por tamaño.

La prueba post-hoc de Tukey permite responder a esta cuestión, ya que evalúa las diferencias entre las medias poblacionales de cada par de grupos, controlando el error tipo I asociado a las comparaciones múltiples. Según los resultados mostrados en la tabla anexa al gráfico, se evidencia que la población de empresas grandes presenta valores significativamente más altos en la variable latente de sensing en comparación con los grupos de empresas medianas, pequeñas y micro. En contraste, no se encontraron diferencias estadísticamente significativas entre las medias de las variables latentes de sensing de las empresas medianas, pequeñas y micro, lo que sugiere un comportamiento similar entre estos tres grupos en términos de propensión al sensing.

**Ventas del último año**

Inicialmente podemos revisar la correlación existente entre las variables latentes calculadas para *sensing* y las ventas de la empresa en el último año. Al calcular la correlación siguiendo el método de Pearson encontramos lo siguiente:

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

Como primera medida, encontramos que el coeficiente de correlación de Pearson es igual a 0,58 (r=0,58), lo que indica una relación positiva y positiva entre las ventas del último año de las empresas incluidas en la muestra y la variable sensing.

1. **Anexos (pruebas estadísticas)**

**Prueba de normalidad (Shapiro-Wilk)**

La prueba de Shapiro-Wilk es una de las pruebas estadísticas más utilizadas para evaluar si una muestra proviene de una distribución normal. Fue propuesta por Samuel Shapiro y Martin Wilk en 1965 y es especialmente adecuada para tamaños muestrales pequeños o moderados (n < 50), aunque también se puede aplicar a tamaños mayores. Esta prueba tiene las siguientes hipótesis:

Para decidir si los datos siguen una distribución normal, se aplica la siguiente regla de decisión. Se selecciona un nivel de significancia α (0,05). Luego, se compara el valor-p obtenido en la prueba con el valor α:

* Si p-valor > α, no se rechaza , por lo que se asume normalidad (los datos siguen una distribución normal).
* Si p-valor < α, se rechaza , por lo que se rechaza normalidad (los datos no siguen una distribución normal).

**Prueba de Levene**

La prueba de Levene es una prueba estadística utilizada para evaluar la igualdad de varianzas (homocedasticidad) entre dos o más grupos. Es una condición importante para aplicar pruebas paramétricas como ANOVA. A diferencia de otras pruebas como la de Bartlett, la prueba de Levene es más robusta frente a desviaciones de la normalidad, por lo que se recomienda cuando no se puede garantizar que los datos sean normales. Esta prueba tiene las siguientes hipótesis:

Para decidir si existe igualdad de varianzas entre dos o más grupos, se aplica la siguiente regla de decisión. Se selecciona un nivel de significancia α (0,05). Luego, se compara el valor-p obtenido en la prueba con el valor α:

* Si p-valor > α, no se rechaza , por lo que se asume homogeneidad de varianzas (existe igualdad de varianzas entre dos o más grupos).
* Si p-valor < α, se rechaza , por lo que se rechaza homogeneidad de varianzas (no existe igualdad de varianzas entre dos o más grupos).

**Prueba ANOVA**

La prueba ANOVA (*Analysis of Variance*) de un factor es una técnica estadística paramétrica utilizada para comparar las medias de tres o más grupos independientes. Evalúa si las diferencias observadas entre las medias grupales son estadísticamente significativas o si pueden explicarse por la variabilidad interna del error aleatorio. Fue desarrollada por Ronald Fisher y es ampliamente utilizada en contextos experimentales y comparativos. Esta prueba tiene las siguientes hipótesis:

Las hipótesis de esta prueba se plantean de la siguiente manera:

Para decidir si al menos una de las medias poblaciones difiere, se aplica la siguiente regla de decisión. Se selecciona un nivel de significancia α (0,05). Luego, se compara el valor-p obtenido en la prueba con el valor α:

* Si p-valor > α, no se rechaza , por lo que no hay evidencia de diferencias significativas en las medias poblaciones entre grupos.
* Si p-valor < α, se rechaza , por lo que existe evidencia de diferencias significativas en las medias poblacionales entre grupos.

**Prueba Tukey HSD**

La prueba de Tukey HSD (*Honest Significant Difference*) es un procedimiento post hoc utilizado después de obtener un resultado significativo en un ANOVA. Su propósito es identificar cuáles pares de grupos presentan diferencias significativas entre sus medias, controlando el error tipo I asociado a realizar múltiples comparaciones. Fue propuesta por John Tukey y es una de las pruebas más utilizadas cuando se cumple la homogeneidad de varianzas y se tiene un diseño balanceado o moderadamente desbalanceado. Las hipótesis para esta prueba son las siguientes (se testea cada hipótesis en cada par de grupos de ; en nuestro caso, sería equivalente a comparar cada tamaño de empresa entre sí)

Para decidir si la media poblaciones difiere en cada par de grupo se aplica la siguiente regla de decisión. Se selecciona un nivel de significancia α (0,05). Luego, se compara el valor-p obtenido en la prueba con el valor α:

* Si p-valor > α, no se rechaza , por lo que no hay evidencia de diferencias significativas en las medias poblaciones entre el grupo .
* Si p-valor < α, se rechaza , por lo que existe evidencia de diferencias significativas en las medias poblacionales entre el grupo .

**Referencias**

* Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Psychometrika Monograph Supplement.
* Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates.
* De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. Guilford Press.
* Baker, F. B., & Kim, S.-H. (2017). *The Basics of Item Response Theory Using R* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-319-54205-8>
* Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3/4), 591–611.
* Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (pp. 278–292). Stanford University Press.
* Fisher, R. A. (1925). Statistical Methods for Research Workers. Oliver and Boyd.
* Tukey, J. W. (1949). Comparing individual means in the analysis of variance. Biometrics, 5(2), 99–114.