**Innovación en Tecnologías de la Información para PYMES de Santiago de Cali**

**Metodología y análisis de datos**

El proceso metodológico desarrollado para estudiar el efecto que tiene la innovación en tecnologías de la información sobre el direccionamiento estratégico en las PYMES de Santiago de Cali, abarcó las siguientes etapas:

1. Análisis descriptivo de datos.
2. Identificación de variables que definan procesos de innovación en tecnologías de la información y actividades orientadas hacia el direccionamiento estratégico en la población objetivo. Para este fin se especificaron 4 constructos:

* Innovación tecnológica en Hardware
* Innovación tecnológica en Software
* Innovación tecnológica en Redes
* Direccionamiento estratégico

1. Análisis de patrones e imputación de valores faltantes.
2. Exploración de relaciones entre las variables.
3. Realización de análisis exploratorio intragrupo, para cada uno de los constructos definidos en la Fase 2.
4. Ajuste del modelo estructural.
5. Evaluación del modelo.

**Operacionalización de variables**

La información disponible proviene del cuestionario. . . aplicado en las PYMES de Santiago de Cali en el año 2012, el cual integra ítems que exploran acerca de diversos aspectos relacionados con la innovación tecnológica, direccionamiento estratégico, eficiencia y efectividad de las Tecnologías de la Información y las Comunicaciones (TIC) en dicha población en estudio. Dichos ítems corresponden a variables categóricas que se encuentran medidos en escala nominal y ordinal en su mayoría, así como variables de respuesta múltiple.

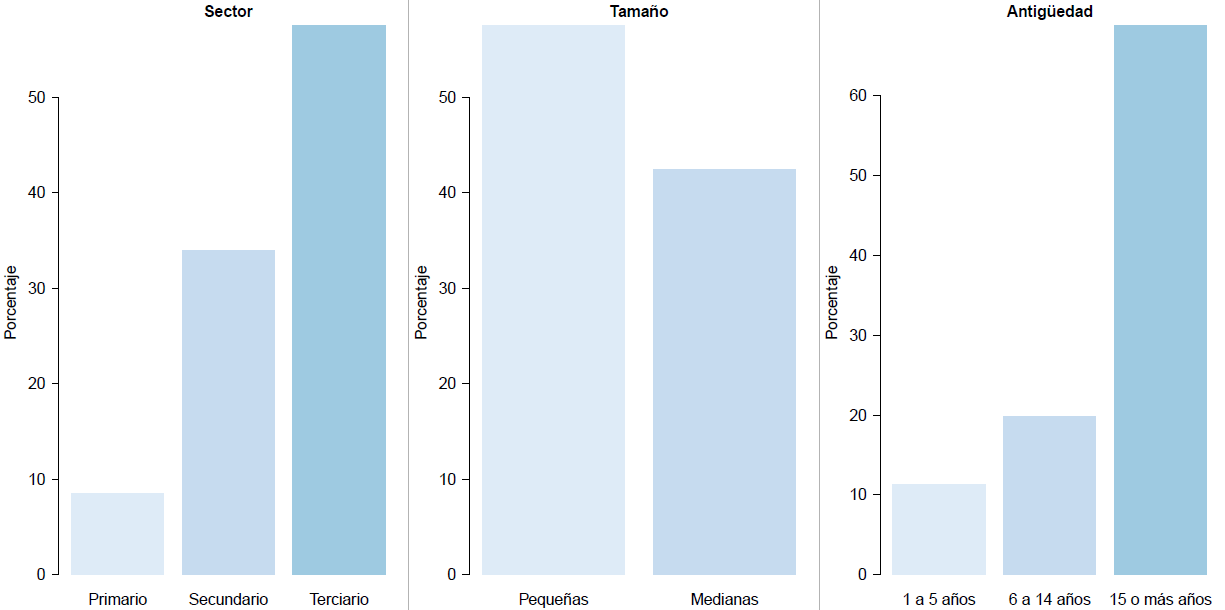
De este modo, para disponer de todas las variables en un nivel de medida adecuado para los análisis posteriores se desarrolló el siguiente procedimiento de depuración:

* Para las variables de respuesta múltiple, recodificación en variables dicotómicas, siendo estas últimas iguales al número máximo de categorías seleccionadas por los encuestados, esto es, si para una variable de respuesta múltiple el número máximo de categorías respondidas es igual a 5, se crearon cinco variables dicotómicas indicando si el encuestado cuenta o no con la categoría respondida; este proceso incrementa de manera considerable el número de variables en el conjunto de datos.
* Para las variables de respuesta múltiple (recodificadas), variables en escala nominal y ordinal, que cuentan con categorías de respuesta ambigua y por ende no tienen un aporte a la explicación de los constructos a definir, son excluidas del conjunto principal de datos.
* En el caso de la pregunta asociada al tipo de conexión usado para el acceso a internet en la empresa (P38), se combinaron las categorías de respuesta WIMAX y USB-Móvil en la categoría OTROS, debido a la baja frecuencia por categorías.

**Análisis descriptivo de datos**

Con variables originales, sin realizar imputación (por el momento). En la fase inicial de nuestro análisis partimos con la exploración univariada de las variables de cada uno de los grupos en estudio:

* Control: charla

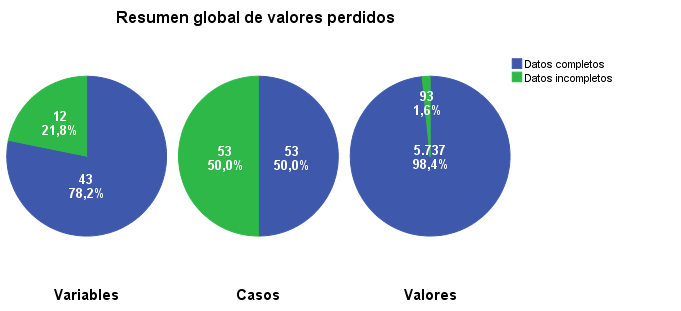


* Innovación tecnológica en Hardware
* Innovación tecnológica en Software
* Innovación tecnológica en Redes

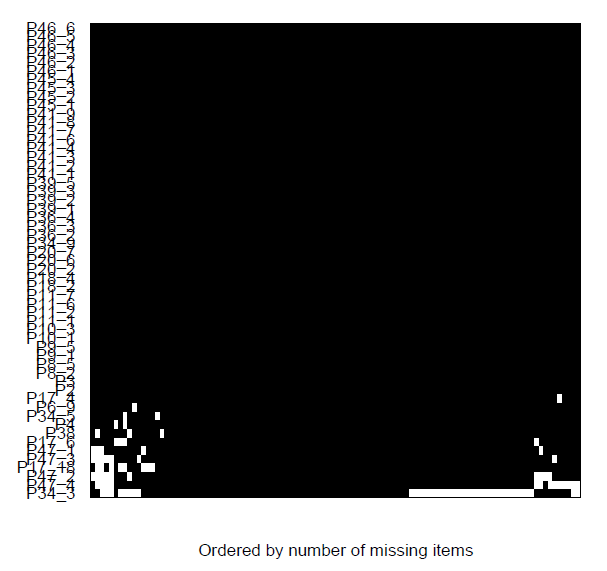
**Análisis de patrones e imputación de valores faltantes**

Resumen de análisis de patrones en datos faltantes del set de datos.

Por variables contamos con un porcentaje de aproximadamente 22% de información faltante bajo el supuesto de exclusión de toda la variable conservando únicamente las variables con datos completos. Por otro lado en términos de información completa por casos contamos con un 50% de esta, lo cual limita mucho los resultados del estudio y de los posteriores modelos.



Patrón de datos faltantes



*Indicadores descriptivos*

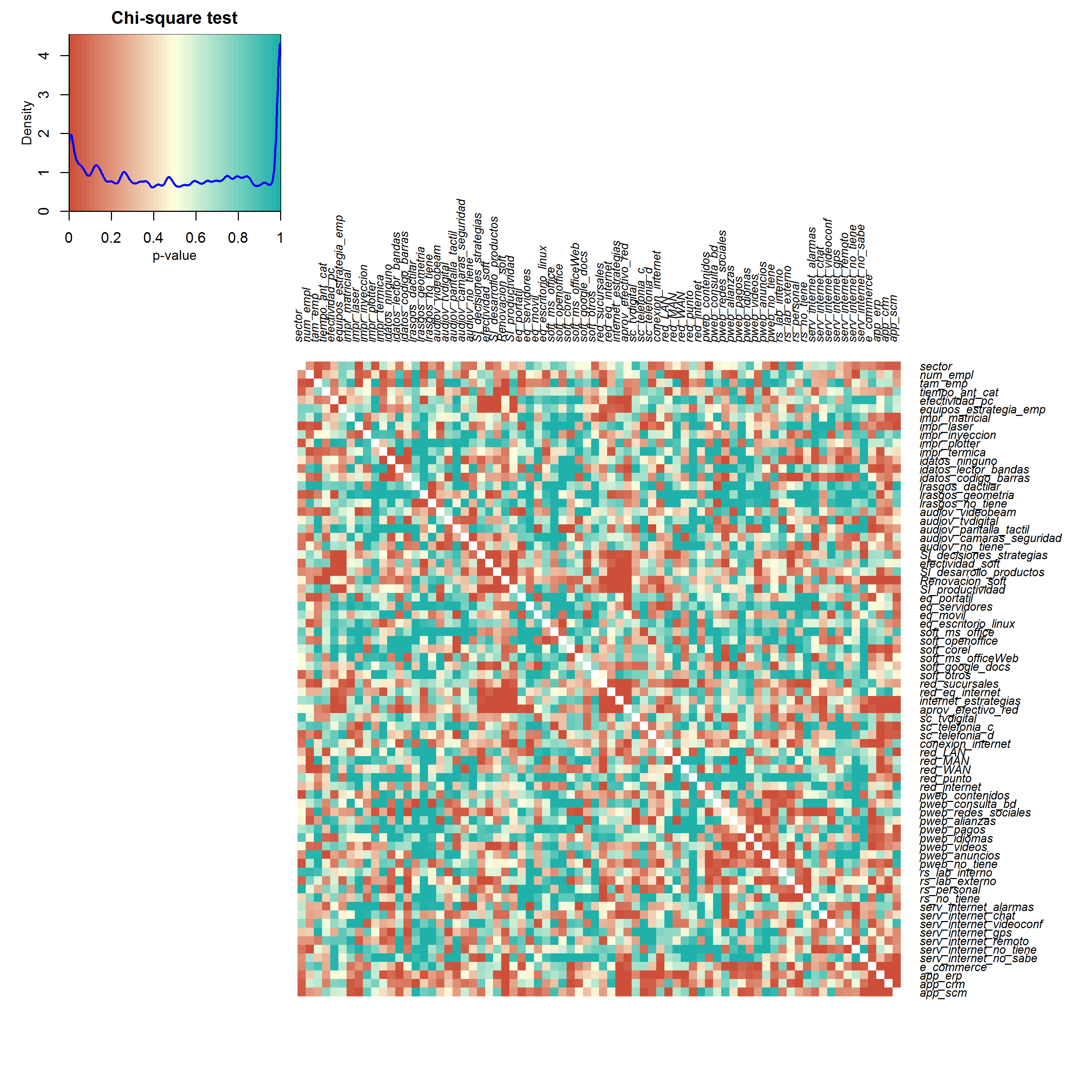
Estadísticos descriptivos y análisis gráfico de las variables seleccionadas.

*Detección de puntos atípicos*

Identificar individuos en la muestra que se destacan del resto por valores extremos en los indicadores medidos.

**Análisis de correlación**

Como primer paso para identificar las posibles relaciones de dependencia entre los grupos de variables, se parte de la realización de un test de independencia Chi-cuadrado entre todos los pares de variables, hallándose un 10% de relaciones estadísticamente significativas entre todos los posibles pares de variables para un nivel de significancia α = 0.05.



*Figura 1. Valores-p correspondiente a test de independencia Chi-cuadrado entre todos los posibles pares de variables, la diagonal superior e inferior contiene los mismos valores*

Las relaciones encontradas en este aparte involucran variables de cada uno de los grupos definidos, indicando de esta manera una posible estructura de relaciones entre más compleja.

*Análisis factorial múltiple*

Aplicación de Análisis de Correspondencias Múltiple para explorar relaciones dentro de cada uno de los sets de datos definidos: Hardware, Software, Redes y Direccionamiento Estratégico. Identificar las variables que tienen un alto peso. Tener en cuenta que este análisis es válido cuando todas las variables a analizar se encuentran medidas en escala nominal. En caso contrario podemos utilizar el Análisis de Componentes Principales Categórico.

Aplicación de Análisis Factorial Múltiple para explorar relaciones entre variables de los constructos de innovación tecnológica y de direccionamiento estratégico. Son 4 sets de datos: Hardware, Software, Redes y Direccionamiento Estratégico.

**Construcción del modelo**

Seleccionar las variables que van a definir cada constructo

Ajustar el modelo (probarlo con: variables que tengan aporte en términos teóricos, sin tener en cuenta su relación con las variables de control; variables que tengan relación con las variables de control; CON BASE EN LOS RESULTADOS DECIDIR).

Verificar resultados

Evaluar el modelo

Equipos de cómputo para ingreso de datos: tabla digitalizadora

Equipos de cómputo para ingreso de datos: no sabe

Equipos audiovisuales interconectados a equipos de cómputo: control de turnos

Equipos audiovisuales interconectados a equipos de cómputo: tableros digitales

Equipos que se tienen en la empresa: PC de escritorio

Sistema operativo usado en equipos de escritorio: mac

Sistema operativo usado en portátiles: mac

Sistema operativo usado en servidores: Solaris

Programas usados en la empresa: Prezi desktop

Programas usados en la empresa: Prezi on line

Programas usados en la empresa: Office 365

Programas usados en la empresa: ninguno

Programas usados en la empresa: no sabe

Existe red de datos

Servicios de comunicación: internet

Red de datos: no sabe

Página web contenido: no sabe

Uso redes sociales: no sabe

====================================================

**FASE 1**: Análisis exploratorio de datos

Dimensiones a evaluar

Innovación tecnológica

Hardware

Software

Redes

Direccionamiento estratégico

**FASE 2**: Exploración de correlaciones

**FASE 3**: Construcción de modelos

Hipótesis

1. La innovación en Hardware apoya las estrategias de la empresa
2. La innovación en Software apoya las estrategias de la empresa
3. La innovación en Redes apoya las estrategias de la empresa

Partial Least Squares Path Modeling (PLS Path Modeling)

**Metodología**

Recolección de datos

La muestra de datos de la cual se dispone corresponde a la caracterización de las pequeñas y medianas empresas de la ciudad de Santiago de Cali en el año 2012 con el objetivo de estimar la efectiva de las tecnologías de la información y las comunicaciones (TIC) en las empresas pequeñas y medianas de la ciudad, en la cual se evaluaron diferentes aspectos concernientes a la efectividad, direccionamiento estratégico de las empresas y aspectos de innovación en tecnologías informáticas. El muestreo aplicado correspondió a un muestro aleatorio simple sobre el marco muestral de pequeñas y medianas empresas de la ciudad con un margen de error del 3% y un nivel de confianza del 95%, escogiendo para esto un total de 107 empresas de la ciudad.

Muestra

Operacionalización

Modelo conceptual

**RECOMENDACIONES PROFESORA SANDRA**

**Resultados**

1. Procesamiento previo de las diversas variables analizadas en el contexto de la innovación en Hardware a través del método….. ; obteniéndose variables con mayor relación con las variables de control establecidas.

Grafica

1. Procesamiento previo de las diversas variables analizadas en el contexto de la innovación en software a través del método….. ; obteniéndose variables con mayor relación con las variables de control establecidas.

Grafica

1. Procesamiento previo de las diversas variables analizadas en el contexto de la innovación en redes a través del método. . . ; obteniéndose variables con mayor relación con las variables de control establecidas.

Gráfica

1. Una vez identificadas las variables relevantes para el estudio en el aspecto de innovación en Hardware, software y redes, se procedió a la construcción del siguiente modelo, donde se definen los constructos de innovación en hardware, software y redes, así como el constructo de direccionamiento estratégico. Y se exploran las siguientes relaciones, para verificar las siguientes hipótesis.

* La innovación en Hardware apoya las estrategias de la empresa
* La innovación en Software apoya las estrategias de la empresa
* La innovación en Redes apoya las estrategias de la empresa

Las variables altamente asociadas con las variables de control, es un conjunto muy reducido que puede no representar asociaciones relevantes entre estas con las variables excluidas. Se recomienda conservar para el análisis que presentan altas correlaciones entre sí, para probar los resultados.

Etapas del análisis de datos

1. Dar formato a los datos:

* Variables de respuesta múltiple pasan a ser variables dicotómicas
* Se eliminan las variables asociadas a categorías de respuesta *no sabe*
* Se eliminan las variables con alto número de datos faltantes
* Se eliminan las variables que provienen de variables filtro, ya que estas constituyen subconjuntos de individuos que cumplen con ciertas propiedades que no son generalizadas para toda la población.
* PARA PENSAR: eliminar variables con un reducido número de conteos por categoría en especial si son menores a 5, como mínimo contar con 10 datos por categoría.

1. Explorar asociaciones entre variables

Se realizan test Chi-cuadrado entre todas las variables restantes para explorar las asociaciones entre variables, a partir de aquí se seleccionan las variables que presentan asociaciones con un nivel de significancia del 5%.

**Organización de datos**

En vista de que una de las empresas estudiadas posee datos faltantes asociados a la pregunta filtro número 32, entonces el número de empresas se reduce a 106 empresas.

Se eliminan las variables con reducidos conteos por categoría de respuesta (variables de baja variabilidad).

Se eliminan las variables que se encuentran poco relacionadas con las demás variables.

Se eliminan las variables que tienen una relación negativa en la definición del primer eje factorial y que además tengan una baja contribución en la definición del primer eje factorial.

**Resultados**

*Modelo 0*: No se tiene en cuenta las variables de control.

Constructos analizados:

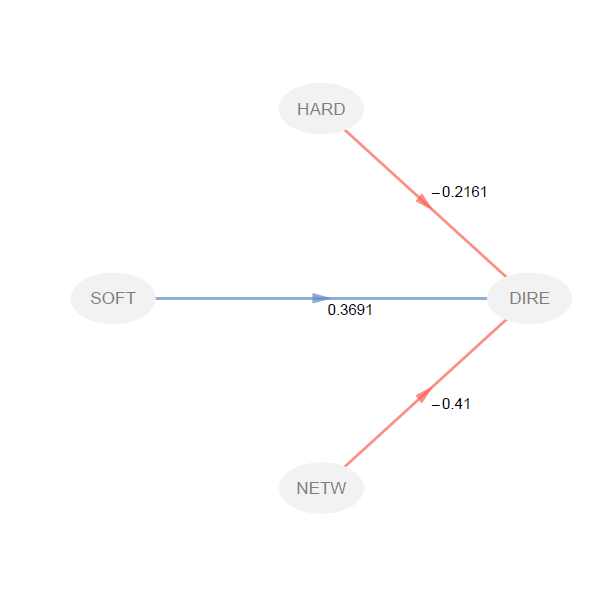
Innovación en hardware

Innovación en software

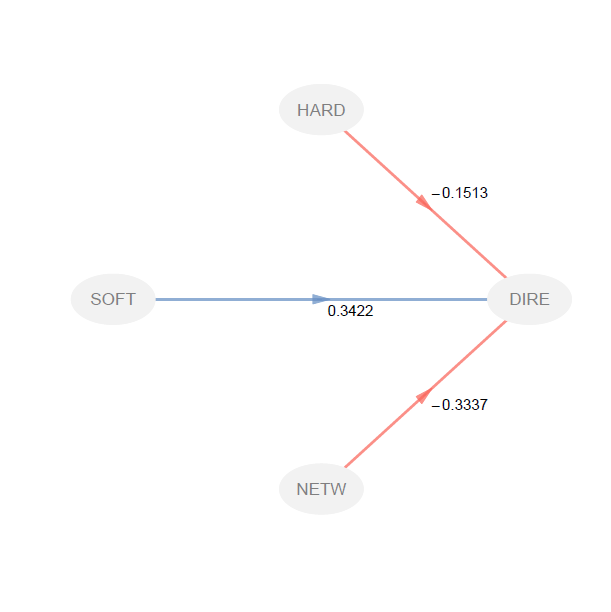
Innovación en redes

Direccionamiento estratégico

Sin hacer selección de variables a partir de las cargas, la validación bootstrap indica que ninguno de los parámetros del modelo es significativo para la interpretación.



Realizando selección de variables en función de las cargas superiores a (0.5) o (-0.5), indicando que existe una relación lineal moderada, se encontró que ninguno de los parámetros del modelo es significativo, con un coeficiente de determinación (R2) del 35% (intervalo de confianza del 95%: [0.316, 0.607])



Modelo 1