[[1]](#footnote-1)

Análisis en SPSS sobre la influencia de diferentes factores en la felicidad de los estudiantes de la Universidad Distrital Francisco José de Caldas

(marzo de 2020)

Alvaro Alejandro Zarabanda Gutierrez.

[aazarabandag@correo.udistrital.edu.co](mailto:aazarabandag@correo.udistrital.edu.co),

Raúl Eduardo Pachón Alarcón.

[repachona@correo.udistrital.edu.co](mailto:repachona@correo.udistrital.edu.co),

Resumen – Este articulo tratara datos de estudiantes de distintas facultades de la Universidad Distrital, con el fin de buscar que variables son determinantes en la felicidad. Para esto se utilizarán diferentes herramientas del software SPSS, haciendo uso del coeficiente alfa de Cronbach para corroborar la fiabilidad de la encuesta utilizada y la regresión logística binaria para revisar la significancia de cada variable con respecto a la felicidad.

**Índice de Términos – Análisis, Correlación, Covarianza, Encuesta, Escala, Ecuación, Fiabilidad, Matriz, Probabilidad, Programa, Regresión, Variable.**

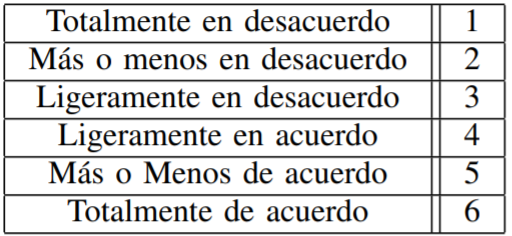
# Introducción

El concepto de felicidad ha desconcertado al ser humano a lo largo de su existencia como especie. Por lo que ha desarrollado una gran variedad de definiciones e investigaciones en torno a este concepto. Estas definiciones han tenido diversos factores sobre los cuales se soportan. Debido a la variación de los parámetros empleados para el desarrollo de estas investigaciones se han precisado algunas perspectivas como: Alarcón [1], basado en la filosofía griega y los recientes estudios, la define como: “un estado de satisfacción, más o menos duradero, que experimenta subjetivamente el individuo en posesión de un bien deseado”, también se han generado perspectivas opuestas como: Fernández D.[2] propone que los factores que influyen en la felicidad, radican en nuestro interior y poco tiene que ver con la acumulación de bienes. Teniendo en cuenta la extensa cantidad de investigaciones desarrolladas en torno a la felicidad el presente trabajo investigativo toma como base los trabajos: The Oxford Happiness Questionnaire desarrollado por los psicólogos Michael Argyle y Peter Hills[3], LA FELICIDAD EN ESTUDIANTES UNIVERSITARIOS DE CIENCIAS ECONÓMICAS: ALGUNOS DETERMINANTES SOCIOECONÓMICOS EN LA CIUDAD DE CARTAGENA DE INDIAS desarrollado por Kevin E. Gamero Tafur, Evelyn M. Medina Martínez y Álvaro A. Escobar Espinoza[4].

El cuestionario empleado en el desarrollo del presente trabajo consto de un total de 42 preguntas de las cuales 29 pertenecen a The Oxford Happiness Questionnaire. Estas 29 preguntas tienen un tipo de respuesta denominada escala Likert. Mientras que las 13 peguntas restantes tienen respuestas de tipo dicotómico que están relacionadas con los factores académicos, sociales, económicos, culturales y personales de los entrevistados.

TABLE I

ESCALA TIPO LIKERT THE OXFORD HAPPINESS QUESTIONNAIRE



## Objetivo General

Analizar y determinar la influencia que tienen diferentes factores en el bienestar subjetivo de los estudiantes de la Universidad Distrital Francisco José de Caldas.

## Objetivos Específicos

* Examinar la incidencia que tienen cada una de las variables propuestas en la encuesta
* Observar las similitudes y correspondencias de este trabajo con investigaciones realizadas con anterioridad y con poblaciones diferentes.

# Marco teórico

## Análisis de Fiabilidad

La fiabilidad una de las principales características que debe cumplir un cuestionario bien estructurado, esta característica hace referencia a la precisión con que el cuestionario mide un determinado rasgo psicológico, independientemente del hecho de si es capaz o no de medirlo. Es decir, se dice que un cuestionario es fiable cuando se mide correctamente aquello que se está midiendo. De manera que el instrumento de medida psicológica no deformará el resultado de una medición debido a cambios o variaciones del instrumento mismo. La fiabilidad tiene dos grandes componentes: La consistencia interna y la estabilidad temporal.[5]

### Consistencia Interna:

Se refiere al grado en que las preguntas de un cuestionario miden un mismo objetivo. Significa la constancia de los ítems para operar sobre un mismo constructo psicológico de un modo análogo.[5]

### Estabilidad Temporal:

Se refiere al grado en que un instrumento de medida arrojará el mismo resultado en diversas mediciones concretas a un objeto o sujeto que ha permanecido invariable.[5]

El coeficiente de fiabilidad es un numero cuyo mínimo valor es 0 y máximo valor es 1. Mientras su valor sea más cercano a 1 la fiabilidad del cuestionario será mejor. Para la estimación empírica del coeficiente de fiabilidad existen diferentes procedimientos como: formas paralelas, test-retest, dos mitades y otros métodos basados en la consistencia interna como el Alfa de Cronbach, Coeficientes de KuderRichardson, Coeficiente beta (β) o los Coeficientes theta (θ) y omega (Ω).[6]

Generalmente se usa el Alfa de Cronbach para hallar el coeficiente de fiabilidad salvo en los casos en los que se desea conocer la consistencia entre dos o más partes de un cuestionario por ej. primera mitad y segunda mitad o cuando queramos conocer otros “subtipos” de fiabilidad (por ejemplo, basados en métodos de dos aplicaciones como el test-retest).[7]

Por otro lado, en el caso de que estemos trabajando con ítems valorados dicotómicamente, se utilizarán las fórmulas de Kuder-Richardson (KR –20 y KR -21). Cuando los ítems tengan diferentes índices de dificultad, se utilizará la fórmula KR –20. En el caso de que el índice de dificultad sea igual, utilizaremos KR –21.[7]

## Correlaciones y covarianzas

La covarianza es el valor que refleja en la manera en que dos variables aleatorias varían de forma conjunta respecto a sus medias. Por lo cual nos permite saber cómo se comporta una variable en función de lo que hace otra variable. la covarianza puede tomar los siguientes valores [8]:

* Covarianza (X,Y) es menor que cero cuando “X” sube e “Y” baja. Hay una relación negativa.[8]
* Covarianza (X,Y) es mayor que cero cuando “X” sube e “Y” sube. Hay una relación positiva.[8]
* Covarianza (X,Y) es igual que cero cuando “X” sube e “Y” baja. No hay relación existente entre las variables “X” e “Y”.[8]

El Cálculo de la covarianza está determinado por ecuación 1, en donde y es la media de la variable Y, x es la media de la variable X. “i” es la posición de la observación y “n” el número total de observaciones.[8]

(1)

La correlación estadística constituye una técnica estadística

que nos indica si dos variables están relacionadas o no. Si

el cambio en una variable está acompañado de un cambio

en la otra, entonces se dice que las variables están correlacionadas.[9]

La correlación estadística es medida por lo que se denomina

coeficiente de correlación (r). Su valor numérico varía de

1,0 a -1,0. Este valor nos indica la fuerza de la relación.

En general, r> 0 indica una relación positiva y r <0 indica

una relación negativa, mientras que r = 0 indica que no

hay relación. Cuando el coeficiente r = 1,0 describe una

correlación positiva perfecta y Cuando el coeficiente r = -1,0

describe una correlación negativa perfecta. De esta manera,

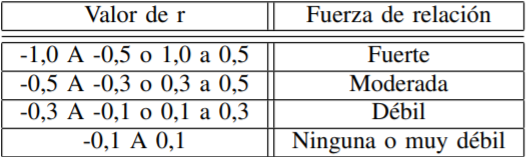
Cuanto más cerca estén los coeficientes de +1,0 y -1,0, mayor

será la fuerza de la relación entre las variables.[9]

Generalmente se establecen las siguientes directrices sobre

la fuerza de la relación:

TABLE II FUERZA DE LA RELACIÓN



Tomada de Explorable.com [9]

La correlación una herramienta útil del análisis estadístico sin embargo se deben tener en cuenta ciertas consideraciones como:

* La correlación es apropiada para examinar la relación entre datos cuantificables significativos en vez de datos categóricos, tales como el sexo, el color favorito, etc.[9]
* Los coeficientes de correlación más utilizados sólo miden una relación lineal. Por lo tanto, es perfectamente posible que, si bien existe una fuerte relación no lineal entre las variables, r está cerca de 0 o igual a 0. En tal caso, un diagrama de dispersión puede indicar aproximadamente la existencia o no de una relación no lineal.[9]
* Hay que tener cuidado al interpretar el valor de ’r’. Por ejemplo, se podría calcular ’r’ entre el número de calzado y la inteligencia de las personas, la altura y los ingresos. Cualquiera sea el valor de ’r’, no tiene sentido y por lo tanto es llamado correlación de oportunidad o sin sentido.[9]
* Al examinar el valor de ’r’ podríamos concluir que las variables X e Y están relacionadas. Sin embargo, el mismo valor de ’r no nos dice si X influencia a Y o al revés.[9]
* La correlación estadística no debe ser la herramienta principal para estudiar la causalidad, por el problema con las terceras variables.[9]

Una de las diferentes variaciones del Coeficiente de correlación es el Coeficiente de correlación múltiple el cual mide la asociación entre varias variables independientes y una dependiente. Este coeficiente se puede definir de manera general como la raíz cuadrada de la suma de los cuadrados explicados por la regresión sobre la suma de los cuadrados totales. Como se muestra en la siguiente ecuación.[10]

##### (2)

Este coeficiente tiene una desventaja, su valor se incrementa cuando se introducen nuevas variables independientes en el modelo, por tanto, resulta engañoso para el análisis.[9]

## Coeficiente de Cronbach

El Alfa de Cronbach se representa mediante y le debe su nombre a Lee Joseph Cronbach, quien bautizó este coeficiente así en 1951. L.J. Cronbach fue un psicólogo estadounidense conocido por sus trabajos en psicometría. Sin embargo, los orígenes de este coeficiente los encontramos en los trabajos de Hoyt y de Guttman. Este coeficiente consiste en la media de las correlaciones entre las variables que forman parte de la escala, y puede calcularse de dos maneras: a partir de las varianzas (Alfa de Cronbach) o de las correlaciones de los ítems (Alfa de Cronbach estandarizado). Como hemos visto cómo la fiabilidad de un cuestionario o instrumento de medida intenta establece la precisión con la éste realiza sus mediciones. Se trata de un concepto muy asociado al error de medida, ya que, a mayor fiabilidad, menos error de medida. El Alfa de Cronbach es un método de cálculo del coeficiente de fiabilidad, que identifica la fiabilidad como consistencia interna. Se denomina así porque analiza hasta qué punto las medidas parciales obtenidas con los diferentes ítems son “consistentes” entre sí y por tanto representativas del universo posible de ítems que podrían medir ese constructo.[7]

El coeficiente Alfa Cronbach se calcula mediante la ecuación 3, en donde n representa el número de ítems la varianza de cada ítem la varianza total del cuestionario.[5]

(3)

## Coeficiente de KR – 20

Es el estimado de homogeneidad usado para instrumentos que tienen formatos de respuestas dicotómicas, (Si - No o Falso - Verdadero), la técnica se establece en una correlación que es basada sobre la consistencia de respuestas a todos los ítems de un test que es administrado una vez. El mínimo aceptable del puntaje de KR-20 es 0.70.[11]

(4)

(5)

Las ecuaciones 4 y 5 pertenecen al cálculo del coeficiente por el método de Kunder-Richardson en donde n representa el número de ítem, T la varianza total de las puntaciones, la proporción de sujetos que aciertan en el ítem, q (q = 1 - p) la proporción de sujetos que no aciertan en el ítem y = suma de las medias de los ítems. [5]

## Regresión Logística

La regresión logística es un grupo de técnicas estadísticas que tienen como objetivo comprobar hipótesis o relaciones causales cuando la variable dependiente es nominal. Existen dos grandes categorías de regresión logística estas la regresión logística binaria y la Regresión logística multinomial.[12]

### Regresión logística binaria

La regresión logística binaria es la técnica estadística que tiene como objetivo comprobar hipótesis o relaciones causales cuando la variable dependiente (resultado) es una variable binaria (dicotómica, dummy), es decir, que tiene solo dos categorías. Aunque su lectura se asemeja a la regresión lineal múltiple, la cual se usa cuando la variable dependiente es ordinal o escalar, la regresión logística está basada en principios diferentes como los odd ratio y las probabilidades. Es decir, se basa en la idea que las variables independientes tratan de predecir la probabilidad que ocurra algo sobre la probabilidad que no ocurra.[12]

Como se ha mencionado anteriormente la regresión logística pretende expresar la probabilidad de que ocurra el evento en cuestión como función de ciertas variables, que se presumen relevantes o influyentes. Si ese hecho que queremos modelizar o predecir lo representamos por Y (la variable dependiente), y las k variables explicativas (independientes y de control) se designan por ,,,…, la ecuación general o función logística es [13]:

(6)

Donde, 1, 2, 3, . . ., k son los parámetros del modelo, y exp denota la función exponencial. Esta función exponencial es una expresión simplificada que corresponde a elevar el número e a la potencia contenida dentro del paréntesis, siendo e el número o constante de Euler, o base de los logaritmos neperianos. [13]

Para el análisis de la regresión logística binaria se tiene en cuenta seis ítems sobre la bondad del modelo:

* Significación de chi-cuadrado del modelo en la prueba ómnibus: Si la significación es menor de 0,05 indica que el modelo ayuda a explicar el evento, es decir, las variables independientes explican la variable dependiente. [12]
* R-cuadrado de Cox y Snell, y R-cuadrado de Nagelkerke: Indica la parte de la varianza de la variable dependiente explicada por el modelo. Hay dos R-cuadrados en la regresión logística, y ambas son válidas. Se acostumbra a decir que la parte de la variable dependiente explicada por el modelo oscila entre la R-cuadrado de Cox y Snell y la R-cuadrado de Nagelkerke. Cuanto más alto es la R-cuadrado más explicativo es el modelo, es decir, las variables independientes explican la variable dependiente. [12]
* Porcentaje global correctamente clasificado: Este porcentaje indica el número de casos que el modelo es capaz de predecir correctamente. Me explico. En base a la ecuación de regresión y los datos observados, se realiza una predicción del valor de la variable dependiente (valor pronosticado). Esta predicción se compara con el valor observado. Si acierta, el caso es correctamente clasificado. Si no acierta, el caso no es correctamente clasificado. Cuantos más casos clasifica correctamente (es decir coincide el valor pronosticado con el valor observado) mejor es el modelo, más explicativo, por tanto, las variables independientes son buenas predictoras del evento o variable dependiente. Si es modelo clasifica correctamente más del 50% de los casos, el modelo se acepta. Sino se deberá realizar un nuevo proceso de selección para nuevas variables independientes. [12]
* Significación de b: si es menor de 0,05 esa variable independiente explica la variable dependiente. [12]
* Signo de b: indica la dirección de la relación. Por ejemplo, a más nivel educativo mayor probabilidad que suceda el evento. [12]
* Exp(b) –exponencial de b–: indica la fortaleza de la relación. Cuanto más alejada de 1 está más fuerte es la relación. Para comparar los exponenciales de b entre sí, aquellos que son menores a 1 deben transformarse en su inverso o recíproco, es decir, debemos dividir 1 entre el exponencial de b (pero solo cuando sean menores a 1). [12]

### Regresión logística multinomial

Una generalización del Modelo de Regresión Logística es el Modelo de Regresión Multinomial, que permite tratar situaciones en las que la variable respuesta tienen un número K, mayor que 2, de alternativas posibles. Una posible forma de abordar este tipo de problemas consistiría en elaborar K modelos de regresión logística, uno para cada una de dichas alternativas.[14]

(7)

Sin embargo, el ajuste independiente de estos K modelos no garantiza que para cualquier conjunto de valores de las variables explicativas se verifique, tal como debería cumplirse, que . [14]

## SPSS

Fue creado en 1968 por Norman H. Nie, C. Hadlai (Tex) Hull y Dale H. Bent. Entre 1969 y 1975 la Universidad de Chicago por medio de su National Opinion Research Center estuvo a cargo del desarrollo, distribución y venta del programa. A partir de 1975 corresponde a SPSS Inc. Originalmente el programa fue creado para grandes computadores. En 1970 se publica el primer manual de usuario del SPSS por Nie y Hall. Este manual populariza el programa entre las instituciones de educación superior en Estados Unidos. En 1984 sale la primera versión para computadores personales. Desde la versión 14, pero más específicamente desde la versión 15 se ha implantado la posibilidad de hacer uso de las librerías de objetos del SPSS desde diversos lenguajes de programación. Aunque principalmente se ha implementado para Python, también existe la posibilidad de trabajar desde Visual Basic, C++ y otros lenguajes. El 28 de junio de 2009 se anuncia que IBM, meses después de ver frustrado su intento de compra de Sun Microsystems, adquiere SPSS, por 1.200 millones de dólares.[15][16]

IBM SPSS: (Statistical Package for the Social Sciences) es un programa destinado al análisis de datos ofrecido por IBM. contiene todas las herramientas necesarias para llevar a cabo completos estudios estadísticos. La base del software estadístico SPSS incluye estadísticas descriptivas como la tabulación y frecuencias de cruce, estadísticas de dos variables, además pruebas T, ANOVA y de correlación. Con SPSS es posible realizar recopilación de datos, crear estadísticas, análisis de decisiones de gestión y graficación de datos obtenidos.[17]

Imagen 1: Icono IBM SPSS



# Procedimiento

## Encuesta

Utilizando esta como método de recolección de datos, se empleó una encuesta que consistió de un total de 42 preguntas de las cuales 29 pertenecen a The Oxford Happiness Questionnaire Estas 29 preguntas tienen un tipo de respuesta denominada escala Likert. La escala Likert de las preguntas tomadas de The Oxford Happiness Questionnaire fueron modificadas de 6 categorías como se aprecia en la tabla 1, a 5 categorías, debido a que la escala de Likert más común contiene 5 categorías y de acuerdo aun análisis interno realizado con anterioridad no se encontró significancia suficiente para manejar dichas categorías. Esta transformación se visualiza en la tabla 3. Mientras que las trece peguntas restantes se definieron con un tipo de respuesta dicotómica con la finalidad de establecer métodos análogos en el cálculo de la fiabilidad como lo son el método Alfa Cronbach y el método KR20 estas preguntas se establecieron con la finalidad de incluir factores académicos, sociales, económicos, culturales y personales de los entrevistados. Las preguntas agregadas fueron:

* Edad(20 o menos/más de 20)
* Género(Hombre/Mujer)
* Semestre(1 a 5/6 a 10)
* Estrato económico(1 y 2/3 o más)
* Estado civil(Soltero/Otro)
* Trabaja(Si/No)
* ¿El nivel de sus ingresos con respecto a sus gastos

mensuales son?(Bajo/Suficiente)

* Promedio académico(Inferior o igual a 3.5/Superior a 3.5)
* Es creyente(Si/No)
* Quién costea sus estudios(Familiar u otro/Usted mismo)
* Tiempo promedio diario de uso del celular(0 – 120 min/Mayor de 120 min)
* ¿En que tipo de casa vive?(Arriendo/Propia)
* Tiempo de traslado(0 - 60 min/más de 60 min)

TABLE III

TRANSFORMACION ESCALA TIPO LIKERT THE OXFORD HAPPINESS QUESTIONNAIRE



Luego de la transformación de la escala Likert las categorías resultantes de la nueva escala Likert con las que se llevó acabo la realización y el posterior análisis de la encuesta son:

TABLE IV

ESCALA TIPO LIKERT



Para el posterior tratamiento de los datos obtenidos en la encuesta se aplica el procedimiento realizado en la encuesta de Oxford, con la finalidad de determinar el valor que tendrá nuestra variable dependiente en cada uno de los registros. Este valor se determina hallando el promedio de las respuestas dadas en el cuestionario este resultado en el trabajo de Oxford es un número entre 1 y 6 sin embargo teniendo en cuenta el cambio en la escala de Likert realizada el resultado en este caso será un numero entre 1 y 5.En cuanto al valor que se debe asignar a la variable principal, para el tipo de análisis que se desea aplicar, es dicótomo, por lo tanto se decidió tomar los valores inferiores a 3.5 como el valor numérico 0 y aquellos que superen o sean iguales a 3.5 su valor será 1.

## Importar Datos

Comenzamos abriendo el software IBM SPSS cuyo icono

se ha mostrado anteriormente. Se desplegará la interfaz del

software creando una hoja de datos automáticamente. La

interfaz del software está estructurada de la siguiente manera:

Imagen 2: Interfaz IBM SPSS



Procederemos a ingresar la información a estudiar, para esto podemos hacerlo ya sea de forma manual, creando variables e ingresando los datos correspondientes, o de manera automática en el caso de poseer los datos almacenados en una base de datos o un archivo compatible con el software SPSS, para esto último solo será necesario importarlos. En este caso las respuestas de la encuesta de la felicidad realizada a los estudiantes se encuentran en formato CSV. Para ello en la barra de menús encontraremos el título archivo al seleccionarlo, se desplegará una serie de opciones, entre ellas importar datos y elegimos Datos CSV, de esta manera el software automáticamente ingresa los datos desde el archivo CSV.

Imagen 3: Importación de datos en IBM SPSS



## Análisis de fiabilidad

### Iremos a la barra de menús en la opción analizar, allí seleccionamos la escala

Imagen: Paso 1 análisis de fiabilidad



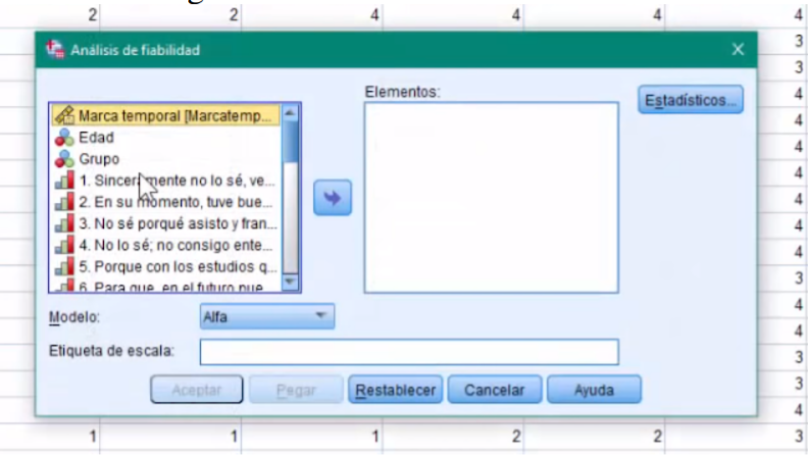
### Una vez seleccionada la opción escala, en el menú desplegado seleccionamos análisis de fiabilidad.

Imagen 5: Paso 2 análisis de fiabilidad



### Se abrirá una ventana, donde tendremos que arrastrar las variables a estudiar con el Alfa de Cronbach, esto para evitar añadir variables como fecha, promedio o estrato que aunque poseen valores numéricos no están estandarizados en la tabla de datos.

Imagen 6: Paso 3 análisis de fiabilidad



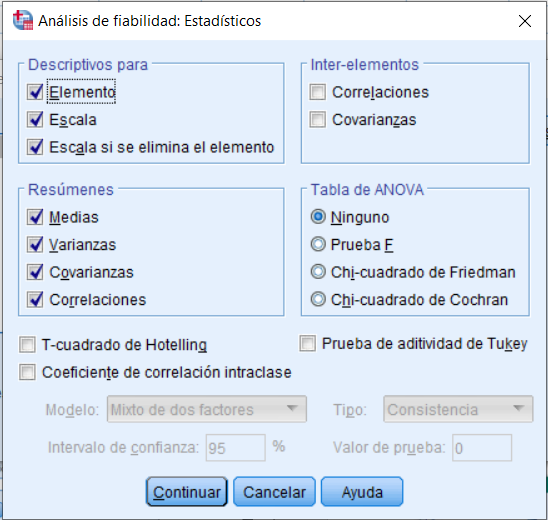
### Se arrastran las preguntas que tomaremos como variables en el estudia al cuadro de elementos, escogeremos en el modelo: la opción Alfa, si se desea encontrar otros valores relacionados al alfa pulsaremos el botón superior derecho con nombre estadísticos.

Para realizar al tiempo otros cálculos con los datos, elegimos aquellos necesarios como las medias, la varianza, las correlaciones y las covarianzas, pulsamos continuar y por último aceptamos para ejecutar.

Finalmente, el software procesa los datos y realiza el análisis de fiabilidad, de que tan confiable son las muestras tomadas respecto a su veracidad.

Por último, se abrirá una nueva ventana con los resultados obtenidos.

Imagen 7: Paso 4 análisis de fiabilidad



## Regresión logística binaria

### Para realizar el proceso de la regresión logista binaria se debe llevar a cabo un procedimiento similar al del analizar de fiabilidad, para empezar, seleccionamos el menú de análisis luego en el meno desplegable de regresión seleccionamos la opción logística binaria.

Imagen 8: Paso 1 regresión logística binaria



### El siguiente paso consiste en asignar los datos requeridos para realizar el proceso, en este tipo de regresión, se debe asignar una variable dependiente de carácter dicótomo y una serie de variables independientes.

Imagen 9: Paso 2 regresión logística binaria



### En caso de requerir resultados con diferentes tipos de evaluación y medida, los ajustes necesarios se pueden realizar en la pestaña que emerge del botón estadísticas Una vez se realicen los ajustes pertinentes se pulsa el botón continuar y posteriormente se abrirá una ventana con los resultados solicitados.

Imagen 10: Paso 3 Regresión Logística Binaria

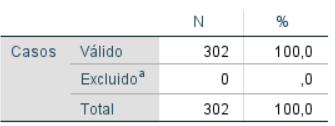


# Resultados

## Análisis de fiabilidad

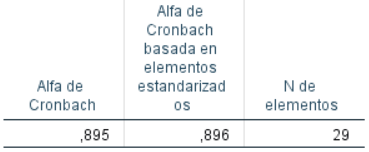
### Oxford Happiness Questionnaire

Imagen 11: Resumen de procesamiento de casos



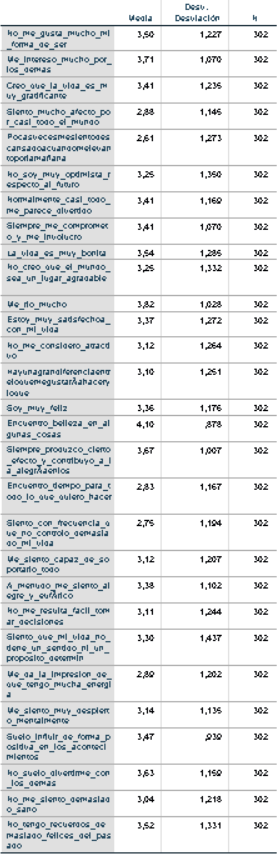
La eliminación por lista se basa en todas las variables del procedimiento

Imagen 12: Estadísticas de fiabilidad



Esta tabla muestra el coeficiente Alfa de Cronbach total del instrumento, el cual fue de 0.895, lo cual es aceptable teniendo en cuenta que el mínimo confiable es de 0.7. El numero de elementos corresponde al numero de preguntas consideradas en el instrumento, siendo tomadas 29 preguntas.

Imagen 13: Estadísticas de elemento



Esta tabla muestra en su primera columna indica la pregunta realizada, la columna 2 indica la media de las respuestas, la columna 3 indica desviación típica o estándar, la cual mide el grado de dispersión de las observaciones individuales alrededor de su media. Por ejemplo, en este caso, la pregunta “Siento que mí no tiene un sentido ni un propósito determinado” tiene la mayor desviación típica de 1.437. Finalmente la ultima columna representa la población estudiada, la cual es de 302 alumnos.

Luego tenemos las matrices de correlación y covarianzas, no serán adjuntadas en el documento debido a su tamaño, sin embargo, se dejarán todos los resultados adjuntos en hojas aparte.

La matriz de correlación es una matriz cuadrada que indica la relación de cada pregunta con las demás, y en sí misma en la diagonal. muestra el coeficiente de correlación de Pearson (r) como: "medida de la fuerza de la relación lineal entre dos variables varía de -1 a 1 cercana a 0 indica poca asociación cercana a 1 indica una asociación directa y cercana a -1 indica una asociación inversa entre las variables"

Teniendo en cuenta los resultados de la matriz de correlación podemos darnos cuenta de los siguientes ejemplos:

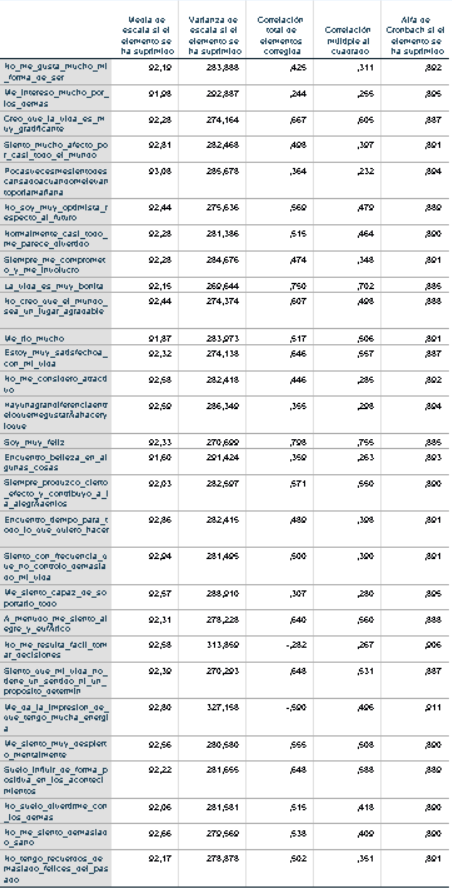
* Valor cercano a 1: La pregunta “Creo que la vida es muy gratificante” tiene un valor de correlación de 0.717 con la pregunta “La vida es muy bonita”, debido a eso tienen una relación directa
* Valor cercano a 0: La pregunta “No me gusta mucho mi forma de ser” tiene un valor de correlación de 0.008 con la pregunta “Me intereso mucho por los demás”. Por ello, estas tienen poca relación entre sí.
* Calor cercano a -1: La pregunta “Me da la impresión que tengo mucha energía” tiene un valor de correlación de -0.574 con la pregunta “Me siento muy despierto mentalmente”, lo que quiere decir que tienen una relación inversa.

Por otro lado, la matriz de covarianzas indica la forma en que las dos variables se mueven juntas, si el valor es positivo las dos variables se encuentran directamente relacionadas, un valor negativo indica que están inversamente relacionadas y un valor de 0 indica que no tiene relación, las variables son independientes.

Teniendo en cuenta los resultados de la tabla anterior se puede tomar como ejemplo:

* Valor positivo: “Siento que mi vida no tiene un sentido ni un propósito determinado” tiene valor de covarianza de 1.021 con la pregunta “no soy muy optimista con respecto al futuro”, por lo tanto, están directamente relacionadas.
* Valor en cero: “No me gusta mucho mi forma de ser” tiene un valor de covarianza de 0.011 con la pregunta “Me intereso mucho por los demás” por ende las variables no tienen relación entre sí.
* Valor negativo: “No soy muy optimista con respecto al futuro” tiene un valor de covarianza de -0.624 con la pregunta “Me da la impresión de que tengo mucha energía”, por lo tanto, están inversamente relacionadas.

Imagen 14: Estadísticas de total de elemento



La media de la escala si se elimina el elemento, indica el valor que tendría la media en el caso de eliminar cada uno de los elementos. La correlación Elemento-Total corregida, es el coeficiente de homogeneidad corregido. Si es 0 o negativo se elimina o se replantea la pregunta. Alfa de Cronbach si se elimina el elemento, equivale al valor de Alfa si eliminamos cada uno de los ítems.

### Variables Independientes

Para estas variables, se hizo uso del análisis de fiabilidad KR-20, eso teniendo en cuenta que son variables dicótomas. Para hacer este procedimiento en SPSS es de manera similar a como se hizo con las variables del Oxford Happiness Questionnaire, y como resultado tenemos:

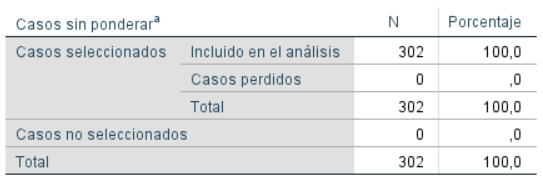
Imagen 15: Análisis KR-20



Como podemos observar, la fiabilidad de esta encuesta no es como se esperaba, sin embargo, gracias a este estudio se están analizando otras alternativas que ya están validadas. Ya en las conclusiones, se hará un análisis más detallado.

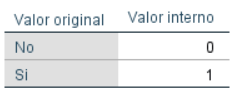
## Regresión Logística Binaria

Imagen 16: Resumen de procesamiento de casos



Esta tabla muestra la cantidad de datos que se incluyeron en el modelo, dende se muestra que tenemos 302 muestras diferentes

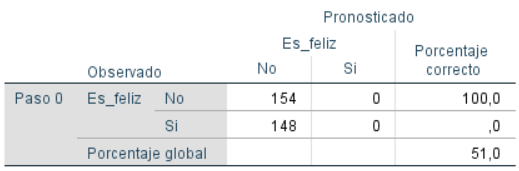
Imagen 17: Codificación de variable dependiente



Aquí se muestra la codificación que va a tener la variable dependiente, cosa que se debe tener en cuenta para la posterior interpretación de resultados.

De la misma forma que la variable dependiente, las variables independientes también se codifican.

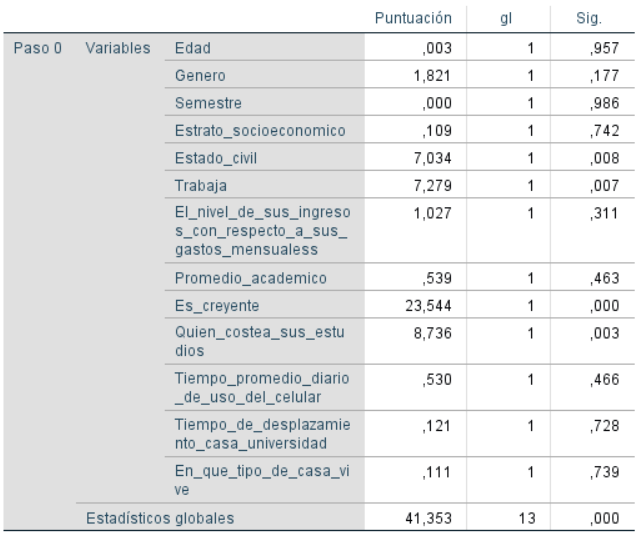
Imagen 18: Tabla de clasificacióna,b



Aquí se define el modelo más básico de predicción de la variable dependiente, donde no se toman las variables independientes. En este caso ese modelo es la categoría donde hay mayor frecuencia, donde se asume que todos los casos están en la categoría de mayor frecuencia, es decir, se asume que todos los estudiantes a los cuales se les hizo el muestreo, no son felices.

Para el análisis de regresión en este bloque, indica que hay un 51.0% de probabilidad de acierto en el resultado de la variable dependiente, donde se asume que los estudiantes de la Universidad Distrital no son felices.

Imagen 19: Variables no están en la ecuación



En esta tabla se encuentran las variables independientes que van a ser incluidas en el próximo paso, donde se puede apreciar el nivel de significancia o repercusión que va a tener cada una de estas en el modelo. En otras palabras, indica si vale la pena o no incluir una variable en el resto del análisis, si la significancia es 0 o aproximado, significa que las variables si van a aportar, en caso de que sea mayor a 0.05 significa que no va a aportar significativamente a la predicción final que nos ofrece el modelo.

Imagen 20: Pruebas ómnibus de coeficientes de modelo



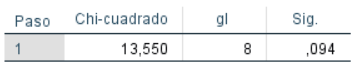
Aquí, nos encontramos una prueba de Chi Cuadrado, la cual es equivalente a la prueba de la ANOVA. Esta prueba nos permite identificar la bondad de ajuste del modelo. En otras palabras, con esta prueba se sabe si las variables que se están incluyendo en el modelo que se está proponiendo mejoran significativamente la predicción de la ocurrencia de la variable dependiente.

Imagen 21: Resumen del modelo



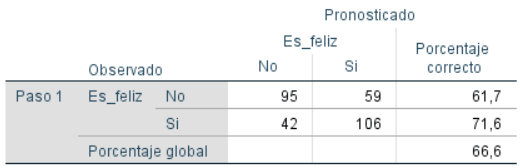
Indica la parte de la varianza de la variable dependiente explicada por el modelo. Hay dos R-cuadrados en la regresión logística, y ambas son válidas. Se acostumbra a decir que la parte de la variable dependiente explicada por el modelo oscila entre la R-cuadrado de Cox y Snell y la R-cuadrado de Nagelkerke. Cuanto más alto es la R-cuadrado más explicativo es el modelo, es decir, las variables independientes explican la variable dependiente.

Imagen 22: Prueba de Hosmer y Lemeshow



Esta prueba indica que la varianza explicada por el modelo explica un porcentaje significativo de la varianza de la variable independiente, es decir si el estudio de la dispersión que se hace con el modelo es significativo con la dispersión poblacional.

Imagen 23: Tabla de clasificacióna



Indica el porcentaje de acierto cuando se tienen en cuenta todas las variables independientes incluidas en el modelo, donde podemos ver que el modelo predictivo aumenta en un pequeño porcentaje.

Imagen 24: Variables en la ecuación



Aquí se indica la probabilidad de ocurrencia de la variable dependiente en función de las variables independientes o la implicación que tiene cada una de estas.

### Variables Significativas

Gracias a la regresión logística binaria podemos hacer un análisis de tal manera que podemos ver cuáles covariables son la que más importancia tienen hacia la variable dependiente, haciendo uso de un estadístico llamado valor p.

Un valor p es una medición estadística entre 0 y 1 que se usa para el contraste de hipótesis. El nivel de significación se debe determinar antes de iniciar la recopilación de datos y se suele establecer en un 5%. [18]

Un resultado es estadísticamente significativo si se corresponde con un valor p igual o inferior al nivel de significación. Se suele expresar como p ≤ 0.05. [18]

Teniendo esto en cuenta, con los resultados de la regresión logística binaria se obtuvieron las siguientes covariables significativas:

* *Estado civil*
* *Trabaja*
* *Es creyente*
* *Quien costea sus estudios*

Posteriormente, se hará un análisis más detallado de los resultados donde se dará la explicación de cada una de las variables independientes y su correspondiente peso en el modelo.

### Análisis Bivariado

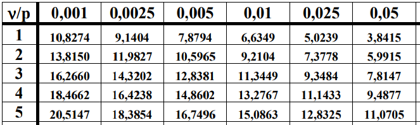
Un análisis Bivariado es una técnica estadística que estudia sucesos que intervienen en dos variables simultáneamente. Sirve para determinar si dos variables tienen relación entre sí, si la relación entre ellas es fuerte, moderada o débil y la dirección de la relación. [19]

Esto no es más que la utilización de la distribución Chi-cuadrado Bivariado, donde todo aparece de forma resumida en la tabla Variables que no están en la ecuación del Software SPSS (Ver imagen 18).

Para este caso de estudio tenemos como hipótesis nula que la variable x no afecta a la felicidad y como hipótesis alternativa que la variable x si afecta a la felicidad. Como vemos en la tabla 1, Tenemos la columna Puntuación y grados y libertad, que son nada más y nada menos que los cálculos correspondientes a Chi-cuadrado.

TABLE V

DISTRIBUCION CHI CUADRADO



El uso de esta distribución nos dice que si el valor chi cuadrado calculado es mayor al de la tabla se rechaza la hipótesis nula, y que si el valor chi cuadrado calculado es menor que el de la tabla se rechaza la hipótesis alternativa.

Para este caso en específico, nuestro valor chi cuadrado de la tabla corresponde a 3.8415, porque ya sabemos que nuestro margen de error es de 0.05(asociado con el p valor) con un solo grado de libertad.

Como podemos ver, tenemos 4 variables que superan dicho valor de Chi Cuadrado (Ver Imagen 19), y son las mismas variables significativas que ya se mencionaron con anterioridad, lo cual nos dice que estas si influyen en la felicidad.

### Análisis Multivariado

El análisis multivariado es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto de estudio.[20]

Para este caso la tabla correspondiente a este análisis aparece en SPSS como variables que están en la ecuación, es decir que han sido incluidas en el modelo de regresión logística binaria. (Ver imagen 24)

La columna B representa el coeficiente del modelo de regresión logística binaria el cual es aplicable en el modelo de regresión y hacer una predicción estadística construyendo la ecuación de regresión logística [21].

Pero antes de realizar este análisis se debe tener en cuenta cuales son las variables significativas en el modelo, lo cual se puede interpretar fácilmente gracias a la prueba de chi cuadrado de Wald para análisis multivariado donde se hace una interpretación muy parecida que con el análisis bivariado. Ya sabiendo el valor de chi cuadrado para un margen de error de 0.05 y un grado de libertad mostrado en la Tabla 5, podemos rechazar la hipótesis nula en las variables “Estado civil” y “Es creyente”, y esto se puede corroborar con el p valor (mostrado en la columna Sig.) donde este es menor a 0.05 en las variables mencionadas.

Ya teniendo estas variables ahora si es posible construir el modelo predictivo, donde hacemos uso de la ecuación (6)

Donde α, β1, β2, β3, …, βk son los parámetros del modelo [21]

**Teniendo esta información, primero aplicaremos la ecuación general en la variable “es creyente” donde la categoría de referencia es *No,* ya que de esta manera quedo codificado en el programa SPSS. Entonces la ecuación sería:**

(8)

Ahora, se calculará reemplazando el valor ES CREYENTE=1, donde 1=No (Cat. de referencia).

(9)

Esto lo podemos interpretar, que con esta probabilidad predicha mayor a 0.50 una persona no creyente se clasificaría como *FELICIDAD=SI*. Cabe recalcar que hay otros factores que influyen en la fiabilidad de los resultados, uno de ellos muy importante es la fiabilidad, factor que se analizara en la concluion.

**A continuación, se hará el análisis con la variable “Estado civil” donde la categoría de referencia es *Soltero*.**

(10)

Ahora, se calculará reemplazando el valor ESTADO CIVIL=1, donde 1=Soltero (Cat. de referencia).

(11)

Esto lo podemos interpretar, que con esta probabilidad predicha mayor a 0.50 una persona soltera se clasificaría como *FELICIDAD=SI*.

IX. CONCLUSIÓN

Referencias

|  |  |
| --- | --- |
| [1] | R. Alarcon, «Desarrollo de una escala factorial para medir la felicidad,» Interamerican Journal of Psychology, 2006, pp. 95-102. |
| [2] | M. R. F. Fernández D., *Construyendo nuestra felicidad para ayudar a construirla,* 2009, pp. 231-239. |
| [3] | D. S. Wright, *Oxford Happiness Questionnaire,* 2008. |
| [4] | E. M. M. A. E. E. K.E. Gamero Tafur, *La felicidad en estudiantes universitarios de ciencias económicas: algunos determinantes socioeconómicos en la ciudad de Cartagena de Indias,* 2017. |
| [5] | P. E. Chiner, *Fiabilidad,* 2011. |
| [6] | J. Oropeza, 2014. |
| [7] | L. R. Mitjana, *Alfa de Cronbach (): qué es y cómo se usa en estadística,* 2019. |
| [8] | J. López, «Covarianza,» [En línea]. Available: https://economipedia.com/definiciones/covarianza.html. |
| [9] | Explorable.com, «La Correlación Estadística. Nov 05, 2019 Obtenido de Snakk Om Mobbing,» 2 Mayo 2009. [En línea]. Available: https://explorable.com/es/la-correlacion-estadistica. |
| [10] | I. L. D. Riobóo, «unidad no.1 Análisis de Regresión Multiple[UNIVERSIDAD NACIONAL DE INGENIERIA-luisd.files.wordpress.com],» Octubre 2008. [En línea]. Available: https://luisdi.files.wordpress.com/2008/09/primera-unidad-regresion-y-correlacion.pdf. |
| [11] | L. B. días, Octubre 2008. [En línea]. Available: CONFIABILIDAD Y VALIDEZ DE CONSTRUCTO DEL INSTRUMENTO “HABILIDAD DE CUIDADO DE CUIDADORES FAMILIARES DE PERSONAS QUE VIVEN UNA SITUACIÓN DE ENFERMEDAD CRÓNICA” [UNIVERSIDAD NACIONAL DE COLOMBIA FACULTAD DE ENFERMERÍA- bdigital.unal.edu.co]. |
| [12] | J. Cardenas, «Qué es la regresión logística binaria y cómo analizarla en 6 pasos.[ networkianos.com],» 21 Febrero 2014. [En línea]. Available: http://networkianos.com/regresion-logistica-binaria/. |
| [13] | M. Aguayo, «Cómo hacer una Regresión Logística con SPSS “paso a paso”.,» [ FUNDACION ANDALUCIA BETURIA PARA LA INVESTIGACION EN SALUD], 2017. [En línea]. Available: http://www.fabis.org/html/archivos/docuweb/Regres\_log\_1r.pdf. |
| [14] | P. R. R. Villafranca, «EL MODELO DE REGRESION LOGISTICA,» [Instituto Valenciano de Investigaciones Agrarias], 2016. [En línea]. Available: http://www.ivia.gva.es/documents/161862582/162456014/MRIM\_mrl. |
| [15] | expansion.com, «expansion.com,» IBM compra SPSS y dispara un 40% sus títulos en Wall Street .[expansion.com], 28 Julio 2009. [En línea]. Available: https://www.expansion.com/2009/07/28/empresas/1248784790.html. |
| [16] | A. Q. G. J.M. Franquet Bernis, Nivelación de terrenos por regresión tridimensional .[ Cadup estudios], 2010. [En línea]. Available: https://books.google.com.co/books?id=qJwn2UqPC28Cprintsec=copyr. |
| [17] | questionpro.com, «Qué es SPSS y cómo utilizarlo,» questionpro.com, [En línea]. Available: https://www.questionpro.com/es/que-es-spss.html. |
| [18] | «EUPATI,» 17 Noviembre 2015. [En línea]. Available: https://www.eupati.eu/es/glossary/valor-p/.. |
| [19] | «SlideShare,» 25 Mayo 2017. [En línea]. Available: https://es.slideshare.net/elisamedinab/anlisis-bivariado-. |
| [20] | S. F. M, «5campus.com,» 2000. [En línea]. Available: http://ciberconta.unizar.es/LECCION/anamul/inicio.html.. |
| [21] | «fabis.org,» [En línea]. Available: http://www.fabis.org/html/archivos/docuweb/Regres\_log\_1r.pdf.. |

1. 1Grupo de Investigación Tratamiento De Historias Clínica Universidad Distrital (TRHISCUD), Facultad de ingeniería Universidad Distrital Francisco José de Caldas, Bogotá Colombia. https://comunidad.udistrital.edu.co/trhiscud/ [↑](#footnote-ref-1)