Universidad Autónoma del Estado de Hidalgo

Dirección General de Planeación

Metodología de muestreo e inferencia estadística para el Seguimiento de Egresados

Apoyado con recursos
del fondo PIFI 3.0

Act. José de Jesús Martínez Guarneros

Dirección General de Planeación

Directorio

Lic. Juan Manuel Camacho Bertrán
Rector

M. en C. Enrique G. Macedo Órtiz
Secretario General

Lic. Gonzalo Villegas de la Concha
Director General de Planeación

CONTENIDO

INTRODUCCIÓN

I. Introducción

El propósito de un estudio estadístico suele ser el de extraer conclusiones acerca de la naturaleza de una población. Al resultar la población grande y no poder ser estudiada en su integridad, en la mayoría de los casos las conclusiones obtenidas deben basarse solamente de una parte de ella, lo que nos lleva a la justificación y la definición de técnicas de muestreo.

Los primeros términos obligados a los que debemos hacer referencia serán los de **estadístico** y **estimador**. Dentro de este contexto, será necesario asumir un **estadístico** o **estimador** como una variable **aleatoria** con una determinada **distribución**, y que será la pieza clave en las dos amplias categorías de la **inferencia** estadística: la estimación de nuestros parámetros y el contraste de hipótesis o **pruebas de hipótesis** de nuestra información obtenida.
El concepto de **estimador**, como herramienta fundamental, lo caracterizamos mediante una serie de propiedades que nos servirán para elegir el "mejor" para un determinado **parámetro** de una población, así como mediante algunos métodos para la obtención de ellos, tanto en la estimación puntual como por intervalos.[1]

La tarea fundamental de la **estadística inferencial** es hacer inferencias acerca de la población a partir de una **muestra** extraída de la misma.

[1] *Bioestadística: métodos y aplicaciones. U.D. Bioestadística. Facultad de Medicina. Universidad de Málaga, 2002. Versión electrónica: http://ftp.med.prev.uma.es/libro*

Es importante enfatizar que la información que proviene de un **censo** corresponde a la de toda la población, por lo que los estadísticos de interés que se obtienen a partir del mismo son los correspondientes a la población, mientras que, cuando la información proviene de una muestra, los estadísticos que se calculan son los propios de ella, debiendo utilizarse métodos de inferencia estadística apropiados para poder establecer a partir de los resultados muestrales, inferencias sobre los poblacionales. La metodología desarrollada está dirigida al caso en que la información sea recabada con base en el **muestreo aleatorio**.[2]

La teoría del muestreo tiene por objetivo el estudio de las relaciones existentes entre la distribución de una característica en una población y las distribuciones de dicha característica en todas sus muestras.
Las ventajas de estudiar una población a partir de sus muestras son principalmente:

- ·Costo reducido
- ·Mayor rapidez
- ·Más posibilidades

De este modo se ve que al hacer estadística inferencial debemos enfrentarnos con dos problemas:

- ·Elección de la muestra (muestreo).
- ·Generalización de las conclusiones obtenidas con la muestra al resto de la población (inferencia).[3]

[2] "Metodología estadística para la realización de estudios de egresados en una institución de educación superior"; Alberto Castillo Morales y Rosa Obdulia González Robles, 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.
[3] Bioestadística: Métodos y aplicaciones. U.D. Bioestadística. Facultad de Medicina. Universidad de Málaga; 2002. Versión electrónica; http://ftp.med.prev.uma.es/libro

El tipo de muestreo más importante es el **muestreo aleatorio**, en el que todos los elementos de la población tienen la misma probabilidad de ser extraídos y es el que utilizaremos en nuestros estudios de egresados.

La mayor parte de la información de este manual es referida a la metodología estadística para la realización de estudios de egresados en una institución de educación, desarrollada por la Universidad Autónoma Metropolitana-Iztapalapa (Alberto Castillo Morales y Rosa Obdulia González Robles; Departamento de Matemáticas), la cual ha sido tomada como referencia para institucionalizar el seguimiento de los egresados de la UAEH (exaueh).

II. Algunos conceptos preeliminares

La metodología que se describe en este documento es aplicable para hacer un estudio de egresados por **encuesta** en una institución de educación superior.

El cuestionario para llevar a cabo dicha encuesta debe ser elaborado como resultado de una **definición clara de los objetivos de la investigación,** dado que con base en éstos se determinan los indicadores y las variables que le conducen a la construcción del instrumento adecuado para medirlas. Una vez que se conocen los objetivos, las variables y el cuestionario para mediarlas, se establecen las técnicas de **muestreo** y de **inferencia estadística** apropiadas. Las primeras permiten utilizar la estructura de la población y el **marco muestral** para proponer un diseño de la muestra, incluyendo los cálculos del tamaño de la muestra; en tanto que las segundas dependen del tipo de variables y de los objetivos que la investigación involucra.

Es importante destacar que el cuestionario, las técnicas de muestreo y las de inferencia estadística no son partes aisladas de una investigación por encuesta. De hecho están estrechamente relacionadas, por lo que es importante conocer los métodos para la obtención del tamaño óptimo de la muestra, la selección de ésta, el cálculo de estadísticos apropiados y las técnicas de pruebas estadísticas, permitiendo obtener información confiable sobre los egresados, con márgenes de error pequeños en las conclusiones que se hagan a partir del estudio.

2.1 Nuestro marco muestral[4]

Para poder llevar a cabo nuestro estudio de egresados es importante, en primera instancia, identificar la **población**, sobre la que deseamos obtener información de acuerdo con los objetivos ya planteados.

En este caso, la población de egresados es dividida por carrera o área académica, debido a que éste es un criterio natural para subdividir a los egresados en nuestra institución. Por otro lado, conviene también definir a la población por **cohorte** de año de egreso, criterio que es adecuado siempre y cuando los egresados hayan llevado el mismo plan de estudios por carrera. El año de egreso permite ubicar a la población de egresados de una misma carrera en una situación homogénea respecto al entorno académico y a las particularidades de la economía. Se pueden unir dos o más cohortes que tengan al menos cinco años de haber egresado.

Entonces, partiendo del hecho de que la carrera y el año de egreso formarán los estratos de las **subpoblaciones** de las que se va a elegir una muestra, el número total de egresados lo denotaremos con la letra **N** y el número de entrevistas que se realicen en esa subpoblación se identificará con la letra **n**. Habrá una **N** y una **n** para cada subpoblación, es decir, para cada carrera o área académica y su respectivo año de egreso.

[4] "Metodología estadística para la realización de estudios de egresados en una institución de educación superior"; Alberto Castillo Morales y Rosa Obdulia González Robles; 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.

Una vez identificadas las subpoblaciones de estudio, se debe de determinar de dónde se elegirán los elementos de la muestra; esto será el **marco muestral**. Si se tiene una lista de egresados para cada subpoblación, que incluya la identificación de su carrera, el año de egreso y las direcciones, teléfonos y datos de cada uno de ellos, entonces el marco muestral y la subpoblación coinciden.
La selección de la muestra se hará sobre el marco muestral o lista de cada subpoblación, obteniéndose una muestra que permitirá hacer inferencias sobre toda la subpoblación y sobre cada uno de los estratos.

Puede darse el caso de que no se cuente con información completa de la subpoblación de estudio; es decir, que el marco muestral es menor a la población; por ello es importante que, antes de iniciar cualquier estudio de egresados, se verifique la calidad del marco muestral por área académica; esto es, que se identifique a los egresados que no aparecen en este con el fin de completar la lista de nombres y datos para su localización. Si por alguna razón no se pudiera completar la lista, es necesario conocer la **proporción** de los egresados que no están en el marco muestral correspondiente; y si esta proporción es grande, digamos mayor a 20%, no conviene hacer el estudio, debido a que el alcance de las conclusiones que se pudieran obtener se alejarían bastante de la realidad.

Si la lista no está actualizada y completa, pero la proporción de falta de información de los egresados es menor a 20%, la selección de la muestra se hará sobre el marco para el que se tengan registros, a sabiendas de que el alcance de las conclusiones del estudio disminuirá conforme el marco se refiere a una parte cada vez más pequeña de la población, pues los egresados que no están en la lista, o aquellos de los cuales no se poseen sus datos, no podrán aportar información para el estudio.

2.2 Variables estadísticas cualitativas y cuantitativas[5]

Una vez que se ha determinado el objetivo del estudio de egresados, procedemos a elaborar el correspondiente cuestionario para llevar a cabo la encuesta. Para cada egresado, la característica que se desea medir, la manera como se va a preguntar para obtener la información y el tipo de respuestas permitido determinaran el tipo de variable utilizada. Esta variable condicionará la forma de llevar a cabo el análisis y la interpretación de los resultados. Básicamente tendremos dos tipos de variables: las **cualitativas**, llamadas también **variables categóricas** (en escala nominal u ordinal), con dos o más posibles respuestas, y las variables **cuantitativas** (en escala de intervalo o de razón), cuya respuesta es un número.

Ejemplo de variable cualitativa:

En el caso en que se pregunte el régimen jurídico de la empresa donde está empleado el egresado, tendremos dos posibles respuestas: pública o privada. Para facilitar la captura de esta variable en la base de datos, supongamos que decidimos etiquetar con “0” al régimen público y con “1” al privado. No significa que estos dos números que sustituyen al nombre de la variable sean susceptibles de ser operados aritméticamente, es decir, manejados en la suma, el producto o el promedio de los valores obtenidos de la muestra para esta variable, pues los resultados son sólo etiquetas, no números.

[5] “Metodología estadística para la realización de estudios de egresados en una institución de educación superior”; Alberto Castillo Morales y Rosa Obdulia González Robles; 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.

En este caso, interesará conocer la **proporción** de egresados que se clasifican en cada una de las dos posibilidades, “público” o “privado”; y en el caso de que se deseara comparar la variable régimen en dos o más carreras, se utilizaría una prueba estadística que compare proporciones, cuyo método es descrito posteriormente.
El cálculo en general de una proporción es el siguiente:

$$p = \frac{m}{n}$$

Donde:
p = proporción de una población,
m = número de elementos de la población que tienen cierta característica en común,
n = número total de elementos de esa población.

Ejemplo de variable cuantitativa:

Si se pregunta la edad a la que se concluyó la carrera, o el ingreso percibido actualmente, tendremos una respuesta numérica de manera natural, es decir, no hay necesidad de etiquetarla; en este ejemplo, los valores obtenidos de la muestra para esta variable serán susceptibles de ser operados aritméticamente.
En este caso, interesa conocer el **promedio** de la edad de los egresados; y si se desea comparar la variable edad de terminación de estudios entre dos o más carreras, se deberá utilizar una prueba estadística para comparar **medias**, cuyo método también es descrito más adelante.
El cálculo en general para un promedio o media es:

$$\mu = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Donde:
μ = media o promedio de una población,
x_i = elemento i-ésimo de la población,

n = número total de elementos x_i en la población.

Nota: para uniformar el análisis y la presentación de los resultados, no se deberán hacer preguntas abiertas en los cuestionarios.

2.3 Consideraciones para determinar los tamaños de muestra[6]

Para decidir el tamaño de muestra de cada subpoblación se debe tener en cuenta que la muestra misma y los valores muestrales que de ésta se deriven, ya sea una proporción o una media, estimarán las correspondientes proporciones o medias de la subpoblación. Los primeros se denominan **estimadores** y los segundos **parámetros**. El objetivo en el estudio de egresados, al igual que en toda encuesta por muestreo, es utilizar los estimadores hallados en una muestra para hacer inferencias sobre los valores de los parámetros en la población total. Por ejemplo, la proporción de egresados que se encuentran trabajando en empresas del régimen privado, en la muestra de una carrera, estima la proporción de los egresados que trabajan en empresas privadas de toda la población de egresados de esa carrera.

Un punto importante por señalar es que los valores de los estimadores, ya sea una proporción o una media, son diferentes en cada muestra que se elija de una misma población. Por ello debemos tener en cuenta lo siguiente: Un estimador *no es un valor concreto* sino una **variable aleatoria**, ya que, aunque depende unívocamente de los valores de la muestra

[6] Bioestadística: Métodos y aplicaciones. U.D. Bioestadística. Facultad de Medicina. Universidad de Málaga; 2002. Versión electrónica; http://ftp.med.prev.uma.es/libro

observados ($X_i=x_i$), la elección de la muestra es un proceso aleatorio. Una vez que la muestra ha sido elegida, se denomina **estimación** al valor numérico que toma el estimador sobre esa muestra.

Las características que son deseables para esta nueva variable aleatoria (que usaremos para estimar el parámetro desconocido) deben ser:

Consistencia:
cuando el tamaño de la muestra crece arbitrariamente, el valor estimado se aproxima al parámetro desconocido.
Carencia de sesgo:
el valor medio que se obtiene de la estimación para diferentes muestras debe ser el valor del parámetro.
Eficiencia:
al ser variable aleatoria, al estimador no puede exigírsele que para una muestra cualquiera se obtenga como estimación el valor exacto del parámetro. Sin embargo, podemos pedirle que su dispersión con respecto al valor central (varianza) sea tan pequeña como sea posible.
Suficiencia:
el estimador debería aprovechar toda la información existente en la muestra.

Para nuestros estudios de egresados, con el fin de garantizar que las diferencias de los valores de los estimadores de una muestra a otra sean pequeñas y, por lo tanto, que las inferencias sobre los parámetros sean lo más parecido posible, se debe llevar a cabo lo siguiente: el tamaño de muestra se determinará de tal manera que se garantice con una probabilidad alta que el estimador no difiera del parámetro por una cantidad pequeña y establecida de antemano, a la cual denominaremos **precisión** y, una vez que la muestra ha sido obtenida, a dicha probabilidad se le denominará **confianza** de la muestra.

Para un tamaño de muestra determinado, los elementos que la conformarán se deberán obtener por selección al azar, es decir, de manera **aleatoria** y sin reemplazo; esto garantiza que se seleccionarán n nombres de nuestro marco muestral con igual probabilidad de ser extraídos, éste método se conoce como **muestreo aleatorio** y es el que deberá ser aplicado en nuestros estudios de egresados.

Debemos tener en cuenta que en nuestros estudios de egresados, a partir de nuestras variables cualitativas y cuantitativas, para calcular el tamaño de la muestra, vamos a enfocarnos en estimar principalmente dos tipos de parámetros: **proporciones** y **medias**, calculando un tamaño de muestra para cada uno de éstos, y eligiendo el tamaño más grande para aplicar los cuestionarios.

Como se muestra en el apéndice a, la ecuación para el cálculo del tamaño de muestra cambia de acuerdo con el tipo de parámetro que se desea estimar (proporción o media), y de acuerdo con el tipo de inferencias que se desee hacer. En el estudio de seguimiento de egresados se nos presentarán principalmente los siguientes tipos de inferencias:

- ·Intervalo de confianza para una proporción
- ·Intervalo de confianza para una media
- ·Comparación de proporciones
- ·Comparación de medias

En resumen, para determinar el tamaño de muestra, debemos primeramente identificar y tomar las variables más relevantes en el estudio y calcular para éstas el tamaño de la muestra, de acuerdo con el tipo de inferencia que se desea hacer de ellas, eligiendo el mayor de los tamaños de muestra obtenidos. Este tamaño de la muestra deberá dar, con una buena aproximación, la precisión y la confianza requeridas para las inferencias que se deseen realizar.

Propuesta de tamaño de muestra para estudio de egresados:

Se sugiere obtener el tamaño de la muestra a partir de las variables: “proporción de egresados que volverían a elegir la misma universidad, en caso de volver a estudiar una licenciatura”, para el caso de variable cualitativa o de **proporción**; y “promedio final con el que se concluye la licenciatura”, para el caso de variable cuantitativa o de **razón**. Se toma el mayor de los valores muestrales que resulten.

Una vez que se obtiene el tamaño de la muestra, se procede a seleccionar los elementos de manera aleatoria, como se explica más adelante. Lo anterior es utilizado con variables derivadas de preguntas del cuestionario que deben ser aplicadas a todos los elementos de la muestra; es decir, no deben usarse con variables que contemplen sólo una proporción de la muestra, como podrían ser aquellas que se derivan de preguntas hechas únicamente a los que han realizado estudios posteriores a la licenciatura.

Es importante mencionar que, en aquellos casos en que el tamaño de la muestra fuese igual o mayor que N (total del tamaño del marco muestral), se hará censo.

En el apéndice a se presentan los métodos para obtener el tamaño de la muestra para el estudio de egresados de la institución, mostrándolos por separado, para cada uno de los casos de inferencia.

III. Obtención de la muestra aleatoria[7]

Una vez que el tamaño de la muestra ha sido obtenido, se procede a seleccionar los elementos del marco muestral de la siguiente manera:

·En primer término, se numera la lista de egresados de todo nuestro marco muestral; el criterio de numeración no es relevante.

·Se obtienen “n” números aleatorios entre 1 y N, usando el paquete estadístico SPSS.

·Los egresados del marco muestral, cuyo número en la lista corresponda con los seleccionados en los n números obtenidos, son los que conformarán la muestra para nuestro estudio y a los que se deberá aplicar el cuestionario.

·En el proceso de localización de los egresados puede darse el caso de que sea necesario intentar contactarlos más de una vez antes de conseguir la entrevista; en ocasiones no se les localiza, o en caso de ser localizados, no acceden a ser entrevistados. Este grupo de egresados se denotará de “no respuesta”.

·En el caso en que se tengan k “no respuestas” en la muestra, si k es menor o igual que 20% de n, se deberá llevar a cabo una segunda selección aleatoria de k elementos, para sustituir a los elementos del grupo de “no respuesta”. Esta nueva muestra se tomará de los elementos del marcomuestral que no habían sido seleccionados en la primera

[7] *“Metodología estadística para la realización de estudios de egresados en una institución de educación superior”; Alberto Castillo Morales y Rosa Obdulia González Robles; 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.*

muestra.

·Si k es mayor que 20% de n, conviene hacer un estudio de **sesgo;** y en caso de probar que el sesgo es significativo en determinadas variables, se deberán calcular y reportar para éstas las estimaciones ponderadas, ya sea de proporciones o de medias, que serán calculadas a partir de las proporciones o medias de los dos subestratos: de los que se obtuvo “respuesta” y de los “no respuesta”, y cuya metodología se describe más adelante.

IV. Análisis de sesgo[8]

Este análisis tiene como objetivo la búsqueda de una posible diferencia entre el grupo de egresados de tamaño (n-k), que responde al cuestionario y los del grupo de "no respuesta" de tamaño k, en las variables más importantes del estudio; es decir, el investigador debe destacar la posibilidad de un efecto de grupo sobre dichas variables.

El hecho de que el grupo de "no respuesta" no conteste el cuestionario puede deberse a razones que no afectan a las variables de interés, pero también puede depender de alguna causa que influye en ellas. Para conocer su respuesta al cuestionario y decidir si es similar o diferente de la respuesta de los que sí responden, se debe hacer una labor especial de búsqueda y convencimiento para conseguir las respuesta de una cantidad de m egresados del grupo de "no respuesta". El valor de m se obtendrá como:

$$m = k\left(\frac{n}{N}\right)$$

Donde:

k = número del grupo de "no respuesta" en la muestra de los egresados

N = total de la población de egresados o marco muestral del estudio

n = tamaño de muestra para esa población

m = tamaño de muestra para el subgrupo de egresados de "no respuesta"

Una vez que se conoce el valor de m, la

[8] "Metodología estadística para la realización de estudios de egresados en una institución de educación superior"; Alberto Castillo Morales y Rosa Obdulia González Robles; 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.

obtención de éste en la lista del grupo de "no respuesta" debe ser también aleatoria. Una vez que se tienen las entrevistas hechas a esos m egresados, se calculan los estimadores correspondientes de acuerdo con el tipo de variable; es decir; para las variables cualitativas, se calcularán las proporciones y en el caso de las cuantitativas, se calcularan las medias o promedios, en ambos grupos, en el de "respuesta" y en el de "no respuesta".

Por último, se procederá a realizar las **pruebas de hipótesis**[9] de comparación de proporciones o de medias, según sea el caso. Si los datos dan suficiente evidencia de que las proporciones o medias de una variable determinada difieren en los grupos, se desprende que hay **sesgo** en esa variable y, en caso contrario, se concluye que no hay sesgo.

En aquellas variables en las que se ha encontrado sesgo significativo, se deben calcular y reportar los estimadores ponderados.

4.1 Estimadores ponderados para variables cualitativas[10]

Sean:

$\hat{p}_1$ = la proporción estimada de una variable cualitativa en el grupo de "respuesta" en una muestra,

[9] Éstas constituyen la parte de nuestro análisis estadístico y serán explicadas detalladamente más adelante.

[10] "Metodología estadística para la realización de estudios de egresados en una institución de educación superior"; Alberto Castillo Morales y Rosa Obdulia González Robles; 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.

$\hat{p}_2$ = la proporción estimada de una variable cualitativa en el grupo de "no respuesta" en una muestra,

El estimador ponderado se calcula de la siguiente manera:

$$\hat{p}^* = \frac{\hat{p}_1(n-k) + \hat{p}_2(k)}{n}$$

Donde:
k = número del grupo de "no respuesta" en la muestra de los egresados
n = tamaño de muestra para esa población

$\hat{p}^*$ = estimador ponderado para la proporción de una variable cualitativa en la población de estudio

4.2 Estimadores ponderados para variables cuantitativas[11]

Por lo general, una variable cuantitativa depende de las respuestas obtenidas en una cualitativa; esto es que, la media o promedio que se calcula para un grupo dependerá primero de si respondió afirmativamente a alguna pregunta previa; de lo contrario, si la respuesta es negativa, no tendría caso calcular dicho promedio o media. Para este caso, tenemos que nuestro estimador ponderado estará dado de la siguiente manera.

Sean:

$\hat{\mu}_1$ = la media estimada de una variable cuantitativa en el grupo de "respuesta" en una muestra, y

$\hat{\mu}_2$ = la media estimada de una variable

[11] "Metodología estadística para la realización de estudios de egresados en una institución de educación superior"; Alberto Castillo Morales y Rosa Obdulia González Robles; 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.

cuantitativa en el grupo de “no respuesta” en una muestra

El estimador ponderado se calcula de la siguiente manera:

$$\hat{\mu}^{*} = \frac{\hat{\mu}_1 (n - k - u) + \hat{\mu}_2 \left(k * \frac{v}{m} \right)}{n + \left(k * \frac{v}{m} \right)}$$

Donde:
n = tamaño de muestra para la población de estudio
k = número del grupo de “no respuesta” en la muestra de los egresados
m = tamaño de muestra para el subgrupo de egresados de “no respuesta”
u = número de egresados en el grupo de “respuesta” que contestan “no” a una variable categórica de interés
v = número de egresados entrevistados en el subgrupo de “no respuesta” que contestan afirmativamente a una variable categórica de interés

$\hat{\mu}^{*}$ = estimador ponderado para la media de una variable cuantitativa en la población de estudio.

En el caso de que el cálculo del estimador de una media no dependa de una respuesta a una variable cualitativa anterior, el cálculo del estimador ponderado se hará de manera similar al visto para una proporción:

$$\hat{\mu}^{*} = \frac{\hat{\mu}_1 (n - k) + \hat{\mu}_2 (k)}{n}$$

Donde:
k = número del grupo de “no respuesta” en la muestra de los egresados

n = tamaño de muestra para esa población

$\hat{\mu}^{*}$ = estimador ponderado para la media de una variable cualitativa en la población de estudio

Nota: *Para las variables en las cuales el sesgo resulte no significativo, se unen los n-k egresados que sí respondieron con los m del grupo de "no respuesta" que se lograron entrevistar, obteniéndose una muestra de tamaño n-k+m, con la cual se calculan los estimadores que se deben reportar en el estudio.*

Todo lo anterior se puede mostrar de manera más clara en el siguiente ejemplo:

En un estudio de egresados, en la búsqueda de los n=98 seleccionados en la muestra, de un total de N=385 de la carrera de Medicina, se logró entrevistar a 67 de ellos, se obtuvo una "no respuesta" de k=31. Como k>20% de 98, se hace un estudio de sesgo. Con este fin se procede a la localización de m de los 31 egresados del grupo de "no respuesta", donde m = 31*(98/385) = 8. Para las variables "¿volvería estudiar la misma carrera?" y "¿labora actualmente?", se obtienen las frecuencias de las respuestas, tanto en la muestra de los que respondieron como en los ocho casos localizados para el estudio de sesgo (elegidos aleatoriamente del grupo de "no respuesta"), reportadas en las siguientes tablas de frecuencias[12]:

[12] Nótese que para la comparación de proporciones se utiliza la prueba de Ji-Cuadrada; y para la comparación de medias o promedios, se utiliza la prueba F de Fisher. Se sugiere al lector estudiar estas pruebas estadísticas con la finalidad de no tener problemas para interpretar los resultados obtenidos.

Volvería a estudiar la misma carrera. Grupo de respuesta y no respuesta.
**Crosstabulation*

Volvería a estudiar la misma carrera		Respuesta	No respuesta	Total
Si	Grupo de Respuesta y No Respuesta	52 77.8%	6 75%	58 77.3% (= 58/75)
No	Grupo de Respuesta y No Respuesta	15 22.4%	2 25%	17 22.7% (= 17/75)
Total	Grupo de Respuesta y No Respuesta	67 100%	8 100%	75 100%

*Chi-Square Test

	Value	df	Asymp.Sig. 2-sided
Pearson Chi-Square	.028	1	.868
Continuity Correction	.000	1	1.000
Likelihood Radio	.027	1	.869
Linear by Linear association	.027	1	.868
N of valid cases	75		

Esta tabla produce una Ji-cuadrada de 0.028. Se concluye que no hay efecto de grupo sobre esta variable, esto es, no hay diferencia significativa entre las proporciones del nivel "sí" en ambos grupos, y la muestra no da evidencia de que exista sesgo. Uniendo los dos grupos se obtiene entonces una proporción de

$\hat{p} = \frac{58}{75} = 77.3\%$ para el nivel "sí" de la variable "volvería a estudiar la misma carrera".

Para la variable "labora actualmente", se obtuvo:

Trabaja actualmente. Grupo de respuesta y de no respuesta.
**Crosstabulation*

			Respuesta	No Respuesta	Total
Trabaja actualmente	Si	Grupo de Respuesta y de No Respuesta	60 89.6	4 50%	64 85.3%
	No	Grupo de Respuesta y de No Respuesta	7 10.4%	4 50%	11 14.7%
Total		Grupo de Respuesta y de No Respuesta	67 100%	8 100%	75 100%

*Chi-Square Test

	Value	df	Asymp. Sig. (2-Sided)
Pearson Chi-Square	8.933	1	.003
Continuity Correction	6.052	1	.014
Likelihood Radio	6.577	1	.010
Linear by linear Association	8.814	1	.003
N of valid cases	75		

Con una Ji-cuadrada de 8.93, significativa al .003, se concluye que hay efecto de grupo sobre esta variable. Las proporciones de respuesta para el nivel “sí” de “trabaja actualmente” son significativamente diferentes en los dos grupos. Dado que la muestra da evidencia de que hay sesgo, se calcula el estimador ponderado para el nivel “sí” de la variable “trabaja actualmente”:

$$\hat{p}^* = \frac{0.896(98-31)+0.50(31)}{98} = 0.77$$

Para la variable cuantitativa “ingreso”, medida en salarios mínimos, se obtuvo lo siguiente:

**Descriptives*

*Ingreso actual en salarios mínimos	N	Mean	Std. desviation	Std. Error
Grupo de Respuesta	60	8.385	2.600	336
Grupo de No Respuesta	4	4.675	0.780	390
Total	64	8.153	2.679	335

*Ingreso actual en salarios mínimos	95% confidence interval for mean		Minimun	Maximun
	Lower bound	Upper bound		
Grupo de Respuesta	7.713	9.057	5	16
Grupo de No Respuesta	3.433	5.917	4	6
Total	7.484	8.822	4	16

ANOVA

*Ingreso Actual en Salarios Mínimos	Sum of squares	df	Mean Square	F	Sig.
Between groups	51.615	1	51.615	7.989	.006
Within Groups	400.564	62	6.461		
Total	452.179	63			

El análisis de varianza produce un estadístico de F con valor de 7.989, significativo al 0.006. Se concluye, entonces, que hay diferencia entre los ingresos de los grupos de respuesta y de no respuesta de los egresados de la carrera de Medicina, por lo que se da un caso de sesgo, debiéndose utilizar el estimador ponderado.

Para proceder a hacer el cálculo de dicho estimador, se deben tener en cuenta los 60+4 egresados que perciben ingresos y que representan a los subestratos de 60 que respondieron y a los 31*(4/8) que no respondieron; entonces:

$$\hat{\mu}^{*} = \frac{8.385(60) + 4.675\left(31\left(\frac{4}{8}\right)\right)}{60 + 31\left(\frac{4}{8}\right)} = 7.6233$$

V. Análisis estadístico[13]

5.1 Detección de errores en la información

En primer lugar se debe verificar que cada variable tome sólo los valores correspondientes a las respuestas posibles que se dan en el cuestionario y, posteriormente, asegurarse de que el número de egresados que no contesta a cada una de las variables sea pequeño, ya que de lo contrario, debe evitarse hacer inferencias sobre estas variables. Se denotará al número de egresados que no contestan una variable como casos “no validos”, y al número de egresados que si la contestan, como “casos validos”.

La primera parte de la verificación puede hacerse pidiendo al paquete SPSS una tabla de frecuencias para cada variable cualitativa (nominal u ordinal) y un resumen descriptivo de las variables de razón que incluya el mínimo y el máximo valor de cada una de ellas; ambos análisis reportan número de casos válidos. La verificación de las variables ligadas se puede hacer con tablas de frecuencia o usando tablas de clasificación cruzadas entre las variables de liga y las que se derivan de ella. Si se detectan errores, es necesario pedir al paquete que imprima el número de cuestionario para los casos donde se presentan los valores diferentes a los permitidos, detectados a partir de los casos de frecuencia. Los errores se localizan por revisión directa del cuestionario.

[13] *“Metodología estadística para la realización de estudios de egresados en una institución de educación superior”; Alberto Castillo Morales y Rosa Obdulia González Robles; 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.*

5.2 Análisis de variables cualitativas

Para cada variable cualitativa se piden las frecuencias para la muestra total de la población. El paquete SPSS presenta una columna de “porcentajes válidos”, misma que constituye en términos de porcentaje, las proporciones correspondientes a cada valor de la variable, obtenidas para todos los que contestaron, esto es; sin considerar los casos “no validos”, los cuales debieron haber sido previamente definidos como “valores faltantes”. Puede darse el caso de que una DES, para un estudio de egresados en particular; en la mayoría de sus variables también requiera las frecuencias para cada una de las carreras, lo que se obtiene pidiendo los cruces (tablas cruzadas) de cada variable por carrera. El total de casos por carrera en la tabla no necesariamente coincide con el tamaño de la muestra por carrera que se tiene en el estudio, ya que en la tabla no aparecen contabilizados los casos “no validos” correspondientes a los egresados que no contestan alguna pregunta, situación que trae como consecuencia que el número de casos para la variable que se mide con dicha pregunta sea menor que el tamaño de la muestra calculado para la carrera.

Es muy importante tener en cuenta que **el tamaño de la muestra real de una variable** determinada para un grupo de interés (como, por ejemplo, todos los que trabajan o todos los que estudian) se obtiene restando los casos “no validos”. El tamaño de la muestra real de una variable cualquiera es el que debe utilizarse tanto para la **comparación de proporciones**, como para calcular **intervalos de confianza** (cuyos análisis se presentan en los siguientes apartados), y que debe reportarse en cada análisis que se realice con una variable determinada.

5.2.1 Obtención del intervalo de confianza para una proporción (p)

El valor que toma una proporción $\hat{p}$ en la muestra no coincide con el valor de la proporción p en la población, pero, no obstante, es posible establecer entre qué valores se ubica el valor en la población con una confianza establecida. Los paquetes estadísticos no cuentan con rutinas para generar el intervalo de confianza de una proporción, ni para poblaciones finitas ni infinitas. Sin embargo, es muy simple su obtención: solamente necesitamos calcular un valor b, dado por la siguiente ecuación, y posteriormente sumarlo y restarlo al estimador muestral de la proporción $\hat{p}$, para obtener el intervalo:

$$b = Z\sqrt{\frac{(N-n)\hat{p}(1-\hat{p})}{N(n-1)}}$$

Donde:

N = tamaño total de la población
n = tamaño de la muestra
Z = valor de la distribución Normal Estándar asociado al valor de la confianza requerida
$\hat{p}$ = estimador muestral de la proporción

Entonces, los límites del intervalo son $\hat{p}-b$ y $\hat{p}+b$, y se tendrá la confianza especificada de que entre esos límites está el valor de la proporción de la población.

Ejemplo:

En una carrera con N=362 egresados, se obtuvo una muestra real de n=194, calculándose la proporción de respuestas afirmativas a una variable, se obtuvo . El intervalo con 90% de confianza para el valor de la proporción en la población se calcula como:

$$b = 1.645 * \sqrt{\frac{(362-194) * 0.31 * 0.69}{362(193)}} = 0.0373$$

Y el intervalo con 90% de confianza va de 0.272 a 0.347. Esto quiere decir que la proporción de respuestas “sí” en la población está entre 0.272 y 0.347, con 90% de confianza.

5.2.2 Comparación de proporciones

A continuación se puntualizan los distintos casos de comparación entre proporciones (tipos de inferencias) que se pueden realizar para una variable cualitativa determinada.
Para poder declarar que las proporciones p y q=(1-p) de una variable con dos niveles son diferentes, se requiere que el intervalo de confianza para una de ellas no contenga al valor 0.5. El intervalo se construye a partir de una de las proporciones estimadas, es decir, la obtenida en la muestra, y el valor de b que se obtenga.

Ejemplo1:

Para la variable “tipo de estudios posteriores a la licenciatura” se tiene que obtener primero la frecuencia de cada uno de los niveles o clases de la variable. Los pasos por seguir en el SPSS para obtenerlas son:

1. A partir del menú principal de comandos, se activa la opción “statistics”, y a partir del sub-menú correspondiente,
2. Se selecciona la opción “sumaries”, y luego
3. Se elige “frecuencias”:

	N	
	Valid	Missing
Estudios posteriores	474	0

Nótese que no se registran casos “no validos” en esta muestra. Luego tenemos:

Estudios posteriores

		Frecuency	Percent	Valid Percent	Comulative percent
Valid	Educación Continua	370	78.1	78.1	78.1
Cases	Postgrado	104	21.9	21.9	100.0
	Total	474	100.0	100.0	
Total		474	100.0		

Nota: *La variable “tipo de estudios posteriores a la licenciatura” no es dicotómica, pero ha sido* ***recodificada*** *para serlo, uniendo los niveles: especialización, maestría y doctorado, de la variable original en el nivel “postgrado”, usado en este análisis. El resto de las opciones de la variable original (diplomados, cursos cortos y otros) se ha unido y conforma el nivel “educación continua” en este análisis.*

Se puede comparar la proporción poblacional de los que realizan estudios de educación continua con la proporción poblacional de los que hacen postgrado, cuyos estimadores son respectivamente: 78.1% y 21.9%. Para poder declarar que las proporciones poblacionales son diferentes, se requiere que el intervalo de confianza para una de ellas no contenga el valor 0.5.

El intervalo puede obtenerse calculando el valor de b, con los siguientes datos: tamaño de la población N=4800, tamaño de la muestra n=474, proporción estimada en la muestra de los que estudian algún postgrado $\hat{p} = 0.219$, valor de confianza de 95%; con lo que se obtiene un valor de b=0.04. Entonces el intervalo de confianza para la proporción de los que estudian postgrado es {(0.219-0.04; 0.219+0.04)} = (0.179; 0.259), el cual no incluye el valor 0.5; por lo tanto, se declara con el 5% de significación que la proporción de los que estudian postgrado difiere de aquella de los que realizaron otro tipo de estudios.

Ejemplo 2:

Supongamos que el intervalo de confianza para una proporción p de una población al 95% de confianza fuera (0.428; 0.541), entonces la muestra no da suficiente evidencia para declarar que p difiera de q[14], pues el intervalo de confianza contiene el valor 0.5.

5.2.3. Comparación entre las proporciones de dos niveles de una variable con más de dos clases

En este caso, el análisis se reduce a las dos clases de la variable cualitativa que se desea comparar. La muestra real para el análisis será la suma de las frecuencias de dichas clases. Como siguiente paso, se obtienen los intervalos de confianza para las dos proporciones y, de ser ajenos (es decir, que no se traslapan), se concluye que las proporciones de las dos clases no son iguales. En caso contrario, si los intervalos de confianza de las dos proporciones se traslapan, la muestra no da evidencia de que las proporciones en la población sean diferentes.

[14] Recuérdese que p+q=1.

Ejemplo:

La variable “tamaño de empresa” tiene cuatro niveles: grande, mediana, pequeña y micro. Para comparar las proporciones poblacionales de los niveles mediano y pequeño, se toma como tamaño de muestra la suma de las frecuencias de estos dos niveles, n=627. Si los intervalos de confianza al 95% (α_i=0.05), basados en los 627 egresados que pertenecen a esos niveles en una población de 10,000 son (0.082; 0.142) y (0.107; 0.167), se concluye que la muestra no da suficiente evidencia con 10% de significación ($1-\alpha_1-\alpha_2$ = 1-0.05-0.05 = 90%), para declarar diferentes a las proporciones poblacionales de los egresados que laboran en empresas de tamaño mediano y pequeño, debido a que los intervalos obtenidos con base en la suma de frecuencias no son ajenos.

5.2.4 Comparación de la respuesta a los niveles de una variable cualitativa entre dos o más carreras

Supóngase que en una DES se tienen dos o más muestras independientes, una por cada carrera, y se quiere saber si el patrón de respuestas a los niveles de una variable es igual entre las diferentes carreras, excepto por la variación aleatoria que se presenta en las muestras. Con las frecuencias de las respuestas, se puede generar una tabla de doble entrada, donde una de las entradas corresponde a los niveles de la variable (criterio de clasificación 1), y la otra, a las carreras (criterio de clasificación 2).

La prueba de homogeneidad de distribución de respuestas entre las diferentes carreras se basa en la Ji-cuadrada que se obtiene de la tabla de frecuencias (el SPSS lo calcula a través del procedimiento "crosstabs"). Si la Ji-cuadrada es significativa, se concluye que las carreras de la DES tienen diferente distribución de sus respuestas, debiéndose interpretar por separado el resultado de cada carrera. En caso contrario, la muestra no da evidencia de diferencias en los patrones de distribución de las respuestas entre las carreras, lo que significa que los egresados de las carreras se comportan de manera similar en cuanto a la variable analizada, pudiéndose interpretar que las diferentes carreras de la DES presentan el patrón de respuestas dado por los valores de las proporciones correspondientes a los totales de las carreras. El uso de "crosstabs" del SPSS se detalla en el apartado 5.2.5 de tablas cruzadas para variables nominales.

Nótese que la forma de realizar la prueba es similar si la variable cualitativa tiene sólo dos o más de dos clases, y también si se desea comparar dos carreras, varias carreras o todas las carreras de la DES.

Ejemplo:
En un estudio de egresados, se desea comparar la distribución de respuestas de la variable "empleo" para tres carreras. Al pedir la tabla de doble entrada y la Ji-cuadrada, se obtiene lo siguiente:

Carrera/empleo
Crosstabulation
Count

	Empleo		
Carrera	Si	No	Total
Medicina	84	12	96
Biología	60	28	88
Sociología	25	9	34
Total	169	49	218

	Value	df	Asymp. Sig. (2-sided)
Pearson	10.202	2	.006
Chi-Square			
Likehood Ratio	10.611	2	.005
Linear-by Linear Association	6.080	1	.014
N of valid cases	218		

En la tabla de salida del SPSS se lee el valor de Ji-cuadrada = 10.202, el cual es significativo al 0.5%. Se concluye que las proporciones de respuesta “sí” y “no” a la pregunta: “¿trabaja actualmente?”, no es la misma para las tres carreras consideradas en el ejemplo.

5.2.5 Tablas cruzadas para variables nominales

***Nota:** Desde que se elabora el cuestionario, se debe determinar qué variables se van a cruzar con otras.*

Para cada cruce, usando el paquete SPSS:

1. Se activa primero la opción “statistics” del menú principal,

2. Se selecciona del submenú correspondiente la opción “sumaries”,

3. Del submenú correspondiente, se debe seleccionar la opción “crosstabs”,

4. En la ventana de diálogo de “crosstabs” se debe seleccionar el botón “statistics”, que ofrece la opción “Chi-Square”. Esta opción permite probar si las variables son independientes o están relacionadas. De acuerdo con el resultado

de la prueba, se lleva a cabo el análisis de porcentajes de la siguiente manera:
a). Si las dos variables son independientes, lo cual ocurre cuando el nivel de significación muestral que aparece en el renglón de Chi-Square y en la columna Asymp. Sig. (2-sided) es mayor que 0.05, pueden describirse por separado los porcentajes totales para cada variable, además de los porcentajes para cada celda.
b). Si las dos variables están relacionadas, lo cual ocurre cuando el nivel de significación muestral que aparece en el renglón Chi-Square y el la columna Asymp. Sig. (2-sided) es menor o igual a 0.05, sólo se puede hacer referencia a los porcentajes en las celdas, analizando sus tendencias por renglón o por columna, y no es conveniente describir por separado los porcentajes totales de renglones ni de columnas, ya que pueden no reflejar las tendencias de los renglones o de las columnas en las celdas de la tabla.

Para el caso en que se pruebe que los criterios son independientes, conviene que la presentación gráfica se haga por separado para cada variable. Por el contrario, si la muestra da evidencia de que las variables están relacionadas, conviene hacer una sola gráfica donde las barras para una variable correspondan a la suma de las partes de la otra variable.

Ejemplo:

Suponga que se hace el cruce entre “sexo” y la opinión de los egresados sobre cómo les ha ido en el cambio de empleo que tenían al egresar, con el empleo que tienen actualmente (en el momento de la entrevista). Supóngase que el número de egresados con el que se realiza este cruce (967) corresponda a aquellos que trabajaban tanto al egresar como en el momento de la entrevista.

En el siguiente cuadro se presenta el cruce entre las dos variables mencionadas, que se pidió con la prueba de Ji-cuadrada, con la finalidad de probar que hay relación entre dichas variables. La prueba produce un valor de significación muestral de 0.00000, lo cual indica que sí hay asociación entre la opinión y el sexo:

**Opinión sobre el nivel de ingresos del empleo al egresar con el empleo actual por sexo*

Count Row Pct. ColPct	Mejoró	Está igual	Empeoró	Row Total
Masculino	146	204	129	479
	30.5	42.6	28.9	49.5
	3.1	54.7	67.9	
Femenino	258	169	61	488
	52.9	34.6	12.5	50.5
	63.9	45.3	3.1	
Column	404	373	190	967
Total	41.8	38.6	19.6	100.0

Chi-Square	Value	df	Significance
Pearson	8.59185	2	0.00000
Likelihood ratio	59.55022	2	0.00000
Mantel-Haeriszel test For linear association	52.72035		0.00000

En este caso en particular, se puede establecer que la opinión con respecto al cambio de empleo es diferente en los hombres que en las mujeres. De hecho, si se toman los porcentajes por renglón (segundo número en la celda) para hombres y mujeres, la distribución de los porcentajes correspondientes a los distintos niveles de opinión difiere; esto es, mientras que los hombres tienden en mayor proporción a opinar que "están igual" (42.6%), las mujeres opinan en mayor proporción que mejoraron (52.9%). Las razones por las que se está dando esta situación no están al alcance de la estadística, por lo que el nivel explicativo corresponde darlo a algún especialista.

5.2.6 Tablas cruzadas para una variable dicotómica medida antes y después

Si se tiene una variable cualitativa con dos clases, que se mide en dos tiempos distintos (antes y después), y se desean conocer las proporciones de los cambios (por ejemplo, si se pregunta a los egresados si trabajan al terminar la carrera y si lo hacen cinco años después), los cambios de un periodo a otro se miden con la prueba de Mc Nemar, la cual permite probar la hipótesis nula de que los cambios en un sentido tienen igual magnitud que los cambios en el sentido contrario. Si el nivel de significación muestral es menor o igual que 0.05, quiere decir que la proporción de cambios en un sentido difiere de la proporción de los cambios en el sentido contrario. De otra manera, la muestra no permite identificar la diferencia significativa. En el paquete SPSS esta prueba se realiza siguiendo los siguientes pasos:

1. A partir del menú principal, se activa la opción "statistics",
2. Del submenú correspondiente, se selecciona la opción "Non parametric test",
3.- Del submenú correspondiente, se selecciona la opción "2 related samples",
4.- Al desplegarse la ventana de diálogo, se

marca la opción “Mc Nemar” de entre las opciones de “test type”.

Ejemplo 1:
Suponga que se tiene la información sobre el número de egresados que estuvieron empleados al egresar y durante el último año de una carrera. Dado que se trata de un caso típico de una variable medida en dos tiempos distintos, “antes” y “después” a los mismos individuos, se pide la prueba de Mc Nemar. Se desea probar que la proporción de egresados que, sin haber trabajado durante el último año de su carrera, sí lo hacen al egresar; difiere de la proporción de los que habiendo trabajado el último año de su carrera no lo hacen al egresar. Los egresados que trabajan en ambos periodos no cambiaron, como tampoco lo hicieron los que no trabajaron en ninguno de los dos periodos (estos casos se llaman empates). Los empates se toman en cuenta para realizar la prueba de Mc Nemar, en la que se busca la significación entre los dos tipos posibles de cambios.

La siguiente tabla es generada cuando se pide la prueba de Mc Nemar:

*Empleo al egresar & Empleo durante el último año

Empleo al ingresar	Empleo durante el último año	
	1	2
1	408	98
2	118	304

Test Statistics (Mc Nemar test)

	Empleo al egresar & Empleo durante el último año
N	928
Chi-Square (Continuity Corrected)	1.671
Asymp. Sig.	0.196

De los 928 egresados, 712 corresponden a empates, es decir, sujetos que no cambiaron su estatus laboral. Del resto, 98 egresados dejaron de trabajar y 118 se incorporaron a un trabajo, dado que no lo estaban haciendo en el último año de la carrera. Los datos muestrales no dieron suficiente evidencia para concluir que la proporción hacia obtener trabajo difiere de la proporción hacia dejar de trabajar (ya que la significación muestral de 0.196 es muy grande). Es importante señalar que, a pesar de que los números de casos 98 y 118 son evidentemente diferentes, esa diferencia se debe sólo al azar, y no está evidenciando una diferencia en las proporciones poblacionales, que es hacia donde se quiere inferir.

Ejemplo 2:
En forma similar al ejemplo anterior, en éste suponga que se desea probar si la proporción de los cambios entre el tener empleo al egresar y tener empleo actualmente (en el momento de la entrevista) es diferente en un sentido o en otro:
*Empleo al egresar & Empleo actual

Empleo al egresar	Empleo actual 1	2
1	520	6
2	380	22

*Test Statistics (Mc Nemar test)

	Empleo al egresar & Empleo actual
N	928
Chi-Square (Continuity corrected)	360.483
Asymp.Sig.	0.000

La tabla muestra una significación muestral de 0.000, que permite declarar que la proporción de aquellos que de no trabajar al egresar sí lo hacen actualmente, difiere de la proporción de los que trabajando al egresar no lo hacen actualmente.

5.3 Análisis de variables cuantitativas

En la encuesta de egresados hay pocas variables cuantitativas, pero para cada una de ellas se debe solicitar un análisis descriptivo que presente su media, su desviación estándar y los valores mínimo y máximo. Este análisis se lleva a cabo en el paquete SPSS de la siguiente manera:

1. Se marca en el menú de comandos la opción "statistics",
2. En el submenú correspondiente, la opción "summarize",
3. Del submenú se elige "descriptives".

Para una presentación gráfica, se sugiere la gráfica de barras[15] para cada variable de interés. A fin de comparar medias de una variable

[15] En este caso deseamos obtener el histograma de frecuencias.

cuantitativa en los diferentes niveles de una variable categórica, se usa el **análisis de varianza** con un solo factor.

En el paquete SPSS este análisis se realiza:

1. Seleccionando la opción “statistics” del menú principal,
2. Del submenú correspondiente, la opción “compare means”, que lleva a su vez a otro submenú,
3. Se elige la opción “One-way ANOVA”.

Si el valor de significación muestral para la prueba F (la columna sig. en la tabla de ANOVA) es mayor que 0.05, se concluye que no hay diferencia de medias en este caso; los datos no dieron evidencia de que las medias poblacionales (parámetros) son diferentes, aunque sus estimadores lo sean numéricamente, debido a variaciones de aleatoriedad. Si el valor de la significación muestral para la prueba de F (la columna sig. en el ANOVA) es menor o igual a 0.05, entonces se concluye que los datos dan suficiente evidencia para rechazar la hipótesis de igualdad de medias; por lo tanto, al menos una de las medias difiere.

Si se prueba (en la tabla de ANOVA) que al menos una de las medias es diferente, se procede entonces a realizar **la prueba de Tukey para comparaciones múltiples de medias** (esta prueba se elige después de marcar el boton “Post Hoc” en la ventana de diálogo desplegada por la opción “One-way ANOVA” en el SPSS). Este análisis, a partir de la comparación de cada media con las demás, presenta un arreglo por grupos, donde todas aquellas medias que entran en un grupo no difieren significativamente entre sí, pero sí difieren con respecto a los otros grupos.

Ejemplo:

Supóngase que se desea comparar el ingreso actual promedio de los egresados de las carreras de Turismo, Administración y Economía. Dicha comparación se lleva a cabo realizando un análisis de varianza:

*Descriptives

	N	Mean	Std. Desviation	Std. error
Turismo	31	5.2290	1.4542	0.2612
Administración	20	4.8300	1.2140	0.2715
Economía	48	17.3750	5.7223	0.8259
Total	99	11.0374	7.4058	0.7443

	95% confidence interval for mean	
	Lower bound	Upper bound
Turismo	4.6956	5.7624
Administración	4.2618	5.3982
Economía	15.7134	19.0366
Total	9.5603	12.5144

ANOVA

Ingreso actual en salarios mínimos

	Sum of Squares	df	Mean Square	F	Sig.
Between groups	3744.416	2	1872.208	110.234	0.000
Within groups	1630.456	96	16.984		
Total	5374.872	98			

La tabla de análisis de varianza prueba que al menos en una de las carreras el promedio del ingreso es diferente, con una significación muestral de 0.000
En la tabla de análisis descriptivo se presentan las medias, desviaciones estándar, desviaciones estándar para la media, e intervalos de confianza para la media correspondientes a cada una de las carreras comparadas. A partir de estos estimadores, se puede apreciar que las medias muestrales de ingreso para las carreras de Turismo y Administración son más cercanas numéricamente que la media muestral de Economía, lo que resulta más contundente cuando se analizan los intervalos de confianza de las medias, que coinciden para Turismo y Administración y son totalmente ajenos en Economía.

Sin embargo, se debe usar la prueba de Tukey para agrupar las medias, ya que se desea probar la diferencia entre medias poblacionales y no entre medias muestrales, que son las que ofrece el análisis descriptivo.

En las siguientes tablas se presenta el reporte de la prueba de Tukey. En la primera se muestran todas las comparaciones posibles de pares de medias, y en la segunda se resumen los resultados. En ésta se presentan dos grupos; de los que se concluye lo siguiente: los datos no ofrecieron suficiente evidencia de que las medias de ingreso que perciben actualmente los egresados de las carreras de Turismo y Administración difieren; pero dieron suficiente evidencia de que dichas medias, son diferentes de la media de los ingresos que perciben los egresados de la carrera de Economía:

**Multiple Comparisons*
Dependent Variable: Ingreso actual
Tukey HSD

(I) Carrera	(J) Carrera	Mean Difference (I-J)	Std. Error
Turismo	Administración Economía	0.3990 -12.1460*	1.182 0.950
Administración	Turismo Economía	-0.3990 -12.5450*	1.182 1.097
Economía	Turismo Administración	12.1460* 12.5450*	0.950 1.097

(I) Carrera	Sig.	95% Confidence interval	
		Lower Bound	Upper Bound
Turismo	0.939 0.000	-2.4148 -14.4065	3.2128 -9.8854
Administración	0.939 0.000	-3.2128 -15.1561	2.4148 -9.9339
Economía	0.000 0.000	9.8854 9.9339	14.4065 15.1561

**The mean difference is significant at the 0.05 level.*
Ingreso Actual
**Tukey HSD[A,B]*

Carrera	N	Subset for alpha= 0.05	
		1	2
Turismo	20	4.8300	
Administración	31	5.2290	
Economía	48		17.3750
Sig.		0.928	1.000

Means for groups in homogeneous subset are displayed.
**A: Uses Harmonic Mean Samples Size=29.100*
**B: The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.*

VI. GLOSARIO DE TÉRMINOS

Aleatoriamente:
Actividades o métodos producidos o llevados a cabo simulando un comportamiento al azar.

Censo:
Recabación de la información de todos los individuos pertenecientes a un grupo o una población.

Datos:
Son los valores cualitativos o cuantitativos mediante los cuales se miden las características de los elementos por estudiar de un grupo o población.

Egresado:
Persona que haya concluido y aprobado su correspondiente plan de estudios.

Encuesta:
Método de recolección de datos, la cual es llevada a cabo generalmente a través de algún formulario que la persona seleccionada en la muestra, debe responder.

Estadística:
Ciencia que estudia los métodos para recoger, organizar, resumir y analizar, así como para sacar conclusiones y tomar decisiones razonables basadas en dicho análisis.

Estimador:
Medida resumen calculada en una muestra para aproximarse al valor de un parámetro de la población total. Los estimadores se usan para hacer inferencias sobre la población de estudio.

Inferencia estadística:
Aplicación de los resultados de estudios de una muestra a la población, y emisión de juicios o conclusiones sobre esa población en general.

Inferir:
Emisión de juicios o conclusiones basados en algún conocimiento o experiencia sobre un evento o suceso.

Marco muestral:
Base de datos o lista definida por la población que conforma el estudio de interés, y cuyo número de registros puede ser menor o igual al tamaño total de la población o grupo de estudio.

Muestra:
Grupo de individuos u objetos elegidos de un grupo más amplio o población, de acuerdo con un criterio preestablecido. Los métodos estadísticos asumen que las muestras son aleatorias.

Muestra aleatoria:
Muestra elegida de tal modo que todos los individuos u objetos de la población tienen la misma probabilidad de ser elegidos independientemente.

Muestreo:
Selección al azar y sin reemplazo de elementos de una población, que conforman el tamaño de muestra determinado, con igual probabilidad de ser seleccionados.

Población:
Grupo de individuos u objetos también llamado universo.

Proporción:
Porcentaje de una cantidad de datos específicos con respecto al total de esos datos.

Sesgo:

Desviación sistemática entre el resultado obtenido para cualquier estimador en la muestra y el verdadero valor del parámetro en la población, debido a la forma en que se hizo el estudio. Es decir, qué tanto se aleja de la realidad un valor estimado.

Tamaño de la muestra:

Cantidad de datos que serán extraídos de la población para formar parte de la muestra.

Variable estadística cualitativa:

También llamada variable categórica, medida en escala nominal u ordinal, con dos o más posibles respuestas.

Variable estadística cuantitativa.:

Son variables definidas en escala de razón, cuya respuesta es un número.

VII. APÉNDICE a

Cálculo de tamaños de muestra[16]

En este apéndice se presenta la metodología para los siguientes cálculos:

A.1 Tamaño de muestra para un intervalo de confianza de una proporción (p).
A.2 Tamaño de muestra para la comparación de dos proporciones independientes (p_1, p_2).
A.3 Tamaño de muestra para un intervalo de confianza de una variable cualitativa o categórica con más de dos clases.
A.4 Tamaño de muestra para un intervalo de confianza de una variable cuantitativa o de razón (μ).
A.5 Tamaño de muestra para comparación de dos medias (μ_1, μ_2).
A.6 Tamaño de muestra para comparación de más de dos medias (μ_1, μ_2,..., μ_k).

A.1 Tamaño de muestra para un intervalo de confianza de una proporción (p)

Si se desea tener un nivel determinado de confianza tal que, entre una proporción de una variable cualitativa de todos los egresados de una carrera y la proporción de ese mismo nivel de la variable en la muestra obtenida de estos egresados no difiera por más de una cantidad pequeña, la cual debe ser determinada por el investigador y la denotaremos por la letra griega

[16] *"Metodología estadística para la realización de estudios de egresados en una institución de educación superior"; Alberto Castillo Morales y Rosa Obdulia González Robles; 2002. Universidad Autónoma Metropolitana-Iztapalapa. Departamento de Matemáticas.*

α, se dice que se desea hacer una estimación $\hat{p}$ para la proporción p en la población, con una precisión de α o menor.
De esta manera, tenemos que:
La proporción $\hat{p}$ en la muestra es el **estimador** para el **parámetro** p en la población, α es la precisión o el margen de error máximo que deseamos tener para que el estimador $\hat{p}$ no difiera del parámetro p. [17]
Por ejemplo, si se desea que la diferencia de las proporciones de los que opinan que "si volverían a estudiar en la misma institución", de entre todos los egresados de la carrera de Administración y la muestra de egresados de ésta no difiera por más de .05=α, se dice que se desea hacer una estimación de los egresados de la carrera de Administración que volverían a estudiar en la misma institución, con precisión de .05 o menor.

Una vez que se decide la precisión, se debe especificar el nivel de confianza o intervalo de confianza para dicha precisión, esto es, la probabilidad con que se espera que se cumpla que el parámetro y el estimador difieran la precisión deseada o menos. Con el fin de que no crezcan mucho los tamaños de la muestra, se sugiere usar una confianza entre 90% y 95% para cada subpoblación.
De esta manera, se tiene la siguiente fórmula:

$$n = \frac{N\hat{p}(1-\hat{p})}{\frac{(N-1)\,\alpha^2}{Z^2} + \hat{p}(1-\hat{p})}$$

Donde:
n = tamaño de la muestra

[17] *En este caso deseamos establecer que* $|p-\hat{p}| \leq \alpha$, *siendo esta nuestra hipótesis alternativa* (H_a)*; para ello tendríamos que haber rechazado la hipótesis nula* H_o: $|p-\hat{p}| > \alpha$.

$\hat{p}$ = estimador para la proporción en la población total
N = número total de la población
α = precisión o margen de error deseado para el estimador
Z = es el valor asociado para el nivel de confianza o intervalo de confianza requerido (1-α), tomado de la distribución normal estándar

En la práctica se usa un valor de $\hat{p}$ que proviene ya sea de una muestra piloto, siempre y cuando el estudio piloto se haya realizado con el rigor estadístico requerido, o bien, de estudios anteriores. Si no existe información de este tipo, se puede utilizar $\hat{p}$ =0.5, lo cual nos arroja el mayor tamaño de muestra.

A.2 Tamaño de muestra para la comparación de dos proporciones independientes (p_1, p_2)

Es frecuente encontrarse con el hecho de querer comparar dos proporciones. Por ejemplo: se dice que la proporción de egresados de la carrera A, que se encuentran realizando estudios posteriores a la licenciatura y es mayor que la proporción de egresados de la carrera B, que se encuentran realizando ese mismo tipo de estudios. Para hacer la inferencia de que “dos proporciones difieren”, se debe haber rechazado la hipótesis nula de que “son iguales”,[18] ya que, de no ser así, las diferencias entre las proporciones muestrales se deben

[18] *En este caso, nuestra hipótesis nula es* H_o*:* $p_1 = p_2$ *, y nuestra hipótesis alternativa es* H_a*:* $p_1 \neq p_2$ *.*

sólo a variación aleatoria, aunque numéricamente una sea mayor que la otra. Debe notarse que se compara entre carreras o estratos las proporciones de un mismo nivel de variable; en este caso, son las proporciones de la categoría “sí” de la variable “realización de estudios posteriores” las que se están comparando entre las carreras A y B.

Para hacer la hipótesis de igualdad de proporciones, sólo se dispone de métodos aproximados que no utilizan el tamaño de la población y producen tamaños de muestra mayores. En lugar de eso, y para aprovechar la metodología que se vio en el apartado anterior, se sugiere obtener el tamaño de muestra de los dos estratos que se desea comparar, usando las ecuaciones y tablas para intervalos de confianza de una proporción. Dichos tamaños de muestra se calculan usando las precisiones α_1 y α_2, de los dos intervalos, asumiendo que la suma de $\alpha_1+\alpha_2$ es igual a la precisión deseada α; esto es $\alpha_1+\alpha_2=\alpha$.

Por lo tanto, si se desea declarar diferentes a las proporciones en las subpoblaciones, cuando la diferencia entre los estimadores sea mayor que una constante o diferencia mínima α, para un nivel de significación establecido (probabilidad fijada o intervalo de confianza), se buscan los tamaños de muestra para cada una de las proporciones con nivel de confianza o intervalo de confianza igual a uno menos la mitad de esa significación establecida $\left(1-\frac{\alpha}{2}\right)$, para que las precisiones al sumarse den la diferencia mínima α. La modificación al nivel o intervalo de confianza es necesaria para garantizar el nivel de significación de la prueba.

De este modo, debemos usar la fórmula:

$$n = \frac{Np(1-\hat{p}_i)}{\frac{(N-1)\alpha_i^2}{Z^2} + \hat{p}_i(1-\hat{p}_i)}$$

Donde:
n = tamaño de la muestra
N = tamaño de toda la subpoblación

$\hat{p}_i$ = estimador en nuestra muestra para la proporción p_i en la subpoblación; i=1,2
$\alpha = \alpha_1 + \alpha_2$; nivel de precisión deseada o valor máximo que pueden diferir las proporciones p_1 y p_2 en las subpoblaciones.
Para calcular el tamaño de muestra se debe hacer, por ejemplo:

$$\%\text{ de precisión deseada} = 1 - \left(\frac{\alpha_1 + \alpha_2}{2}\right),$$

Z = valor asociado para nuestro nivel de significación o intervalo de confianza en la distribución normal estándar, el cual debe ser modificado de la siguiente manera; por ejemplo: si se desea que las proporciones p_1 y p_2 no difieran por más del valor a de 10%, entonces deberemos usar la confianza del 95%, la cual es hallada como: .95=1-(.10/2).

A.3 Tamaño de muestra para un intervalo de confianza de una variable cualitativa o categórica con más de dos clases

Para este tipo de variables se puede utilizar un tamaño de muestra dirigido hacia la clase o nivel de la variable de mayor interés y reducir el problema al caso de una proporción. De manera alternativa, si la variable es orden o cuantitativa, de manera aproximada se puede hacer referencia a la media y se utilizan los resultados de las variables de razón que se verán en el siguiente apartado.

A.4 Tamaño de muestra para un intervalo de confianza de una variable cuantitativa o de razón (μ)

Para calcular el tamaño de muestra para un intervalo de confianza de una variable de razón, por ejemplo, el ingreso, también es necesario especificar un valor para la precisión α; pero, en este caso, debido a que la variable puede tomar cualquier valor numérico, esta precisión deberá estar definida en términos de la desviación estándar; es decir:

$$\beta = \frac{\alpha}{\sigma}$$

Donde:

β = precisión en términos de la desviación estándar de nuestra variable de razón

α = nivel de precisión deseado

σ = desviación estándar

Cuando se ha determinado un valor de precisión α, del cual obtendremos un valor de precisión , para que la media estimada $\hat{\mu}$ no difiera del parámetro μ en la población de estudio por más de dicho valor β[19], aclarando que, para utilizarla, la desviación estándar debe ser conocida. Entonces la ecuación para calcular el tamaño de la muestra es la siguiente:

[19] *En este caso, nuestra hipótesis nula es* H_o: $|\mu - \hat{\mu}| > \beta$ *y nuestra hipótesis alternativa es* H_a: $|\mu - \hat{\mu}| \leq \beta$.

$$n = \frac{N}{\frac{N\beta^2}{Z^2}+1}$$

Donde:
n = tamaño de muestra
N = tamaño total de la población
β = precisión deseada en términos de la desviación estándar
Z = valor asociado al nivel de significación o intervalo de confianza deseado (igual a 1-α), tomado de la distribución normal estándar

A.5 Tamaño de muestra para comparación de dos medias (μ_1, μ_2)

Con frecuencia surge la necesidad de comparar dos medias independientes o más; por ejemplo, supóngase que un investigador de una DES desea agrupar entre algunas carreras, aquellas cuyas medias poblacionales de alguna **variable cuantitativa**; como podrían ser el ingreso o la edad, no difieran.

Para realizar la prueba de hipótesis de igualdad de dos medias (μ_1 y μ_2), se puede proceder como en el caso de dos proporciones. Se obtiene el tamaño de muestra de los dos estratos que se desea comparar, usando las ecuaciones para intervalos de confianza de una media μ. A partir de las precisiones de los dos intervalos (para β_1 y β_2), se diseña de manera que la suma $\beta_1+\beta_2=\beta$ sea la precisión deseada en la comparación de dos medias (μ_1 y μ_2).

Entonces, si se desea declarar que las medias son diferentes cuando la diferencia entre los parámetros sea mayor que una constante o diferencia mínima β[20], para un nivel de significación o intervalo de confianza establecido,

[20] *En este caso, nuestra hipótesis nula es* H_o: $\mu_1 = \mu_2$, *y nuestra hipótesis alternativa es* H_a: $\mu_1 \neq \mu_2$.

se buscan los tamaños de muestra para cada una de las medias con el nivel de confianza o intervalo de confianza igual a uno menos la mitad de la significación establecida ($1-\alpha/2$), para precisiones que al sumarse den la diferencia mínima α. Debe recordarse que las precisiones deben estar en términos de las desviaciones estándar. Por lo tanto, tenemos:

$$n = \frac{N}{\frac{N\beta_i^2}{Z^2} + 1}$$

Donde:
n = tamaño de muestra
N = tamaño de toda la subpoblación
$\beta = \beta_1 + \beta_2 =$
nivel de precisión deseada o valor máximo que pueden diferir las medias μ_1 y μ_2 en las subpoblaciones
Z = Valor asociado para nuestro nivel de significación o intervalo de confianza en la distribución normal estándar, el cual debe ser modificado de la siguiente manera; por ejemplo:

Si se desea que las medias μ_1 y μ_2 no difieran por más del valor β con un nivel de significación de 10%, entonces deberemos usar una confianza de 95%, la cual es hallada como:

$$1 - \frac{(\alpha_1 + \alpha_2)}{2} = 1 - \frac{.10}{2} = 95\% .$$

A.6 Tamaño de muestra para comparación de más de dos medias ($\mu_1, \mu_2, \ldots, \mu_k$)

Cuando se desee comparar más de dos medias ($\mu_1, \mu_2, \ldots, \mu_k$), en general k medias, se sugiere utilizar el mismo procedimiento que al comparar dos medias.[21]

[21]*En este caso, tenemos la hipótesis nula* H_o: $\mu_1 = \mu_2 = \ldots = \mu_k$ *y la hipótesis alternativa Ha:* $\mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$.

Pero calculando el nivel de confianza como:

Nivel de confianza =
1 - ((significación deseada) / k)=
$1-((á_1 + á_2 + \ldots + á_k)/ k)$

Luego la precisión para calcular el tamaño de la muestra de cada uno de los estratos se tendría que corregir por: $\gamma = \beta_1 + \beta_2 + ... + \beta_k$

Donde:
γ =
precisión corregida para inferir sobre la diferencia de k medias entre k subpoblaciones
β_i =
precisión deseada para la media correspondiente a la i-ésima subpoblación, en términos de la desviación estándar:

$$\beta_i = \frac{\alpha_i}{\sigma_i}; \qquad i=1,2,...,k.$$

Ejemplo:
Si se tuvieran cinco carreras, y se deseara comparar entre éstas el ingreso de los egresados. Ya que el ingreso está medido en moneda nacional o en número de salarios mínimos, la variable es cuantitativa, por lo que la comparación que se desea se tendría que realizar comparando (estadísticamente) las medias de las carreras. El procedimiento para calcular el tamaño de muestra por estrato, en el que se deseara obtener 10% de significación para la comparación de esas cinco medias ($10\%=\alpha=\alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5$), el nivel de confianza o intervalo de confianza que se tendría que usar sería: 1-(.10/5)=.98.

Y la precisión usada para calcular el tamaño de muestra de la i-ésima subpoblación es:

$$\gamma=\beta_1+\beta_2+\beta_3+\beta_4+\beta_5$$

VIII. APÉNDICE b
Responsabilidades

B.1 Del personal que labora en la Dirección de Seguimiento de Egresados

a. Servir de instrumento de apoyo a cada una de las áreas académicas para llevar a cabo los estudios de seguimiento de egresados.
b. Determinar los muestreos correspondientes para cada estudio.
c. Ordenar y clasificar la información obtenida de las encuestas realizadas, llevar a cabo el análisis de resultados y publicar el estudio.

B.2 Del área de Control Escolar

a. Verificar que los datos sean capturados de manera correcta, siguiendo un mismo criterio.
b. Validar los datos que al finalizar cada semestre son enviados para alimentar la base del Sistema Institucional de Seguimiento de Egresados.
c. Enviar la base de datos de los egresados actualizada.
d. Validar, con la documentación que obra en su archivo, la información que respalda los estudios de los exauaeh, generando números de cuenta y NIPS de acceso al sistema.

B.3 Del que origina los procedimientos

a. Los directores de la DES, a través de los responsables de los programas educativos, elaborarán los cuestionarios que serán enviados a la Dirección de Seguimiento de Egresados para su aplicación, indicando cuál es la finalidad de este estudio.
b. Contactar a sus egresados para la actualización de la base de datos

correspondiente y mantener el vínculo para contar con información confiable cada vez que se requiera.

c. Aplicar las encuestas específicas que se elaboren para estudiar los programas educativos y la productividad de los egresados.

d. Es labor del área académica correspondiente la toma de decisiones con base en los resultados obtenidos del estudio.

APENDICE C

IX. Apéndice c

Descripción de los procedimientos de trabajo para el seguimiento de egresados

C1. Procedimiento 1: entre la Dirección de Control Escolar y las Dependencias de Educación Superior

1. El área de Seguimiento de Egresados sustenta su actividad con la base de datos de los egresados que la Dirección de Control Escolar proporciona de manera automática en forma semestral. Se reciben estos datos.

2. Los datos son revisados por las personas designadas por la Dependencia de Educación Superior correspondiente, con la finalidad de verificar que estén completos y sean congruentes.

3. En caso de que haga falta obtener algunos datos faltantes, es labor de las Dependencias de Educación Superior establecer contacto con los egresados para que se pueda validar su información.

4. Las Dependencias de Educación Superior deben registrar y/o modificar los datos de los egresados y alumnos que así lo requieran, por cambio de domicilio, número telefónico, cambio de dirección de correo electrónico, actualización de datos por acreditación de estudios de postgrado, etc. Deben reportar los cambios respectivos a la Dirección de Control Escolar, por medio de una aplicación en línea, con la finalidad de que los datos de egresados y alumnos estén siempre actualizados.

C2. Procedimiento 2: entre las áreas académicas y la Dirección de Seguimiento de Egresados

1. El área académica correspondiente solicita al área de Seguimiento de Egresados un estudio, especificando el cuestionario que desea que sea aplicado a sus egresados.

2. El área de Seguimiento de Egresados revisa el cuestionario, lo aprueba, lo calendariza y lo programa para su aplicación.

3. El área de Seguimiento de Egresados publica en el portal web el cuestionario solicitado y elabora copias impresas del mismo.

4. Se baja la base de datos de Control Escolar de la población de egresados del respectivo plan de estudios.

5. El área de Seguimiento de Egresados establece las cohortes poblacionales de esa base de datos, de acuerdo con las fechas en las que se han llevado a cabo cambios en el plan de estudios y éste ha entrado en vigor.

6. Una vez definido el universo de egresados del correspondiente plan de estudios, el área académica procede a contactar a los egresados para la actualización de los datos (domicilios y teléfonos); al concluir esta labor, el área de Seguimiento de Egresados recibe dicha información.

7. El área de Seguimiento de Egresados revisa esta base actualizada. El objetivo es contar con más del 80% de los datos actualizados.

8. El área de Seguimiento de Egresados hace un muestreo de la base de datos actualizada.

9. El área de Seguimiento de Egresados asigna números de identificación personal (NIP'S) a los egresados seleccionados en la muestra, para que únicamente ellos tengan acceso a responder el cuestionario por medio del portal web.

10. El área académica contacta a los egresados seleccionados en la muestra y los invita a que contesten el cuestionario por medio del portal web en primer lugar y, de no ser posible para el egresado(a) esta opción, se concertará una visita domiciliaria para la entrega de un cuestionario impreso.

11. El área de Seguimiento de Egresados captura las respuestas obtenidas en una base de datos, conforme el área académica le va haciendo llegar los cuestionarios contestados. En caso de existir respuestas incompletas en algunos cuestionarios, se invita nuevamente al egresado a responder sus preguntas faltantes.

12. Se monitorean, cada periodo determinado, las respuestas obtenidas.

13. Se establece una fecha límite para la recepción de cuestionarios contestados.

14. En esa fecha límite, se comienza a depurar la base de datos, y se deja abierta la opción de capturar algunos cuestionarios que pudieran llegar al área de Seguimiento de Egresados durante la depuración de la base de datos.

15. Una vez depurada la base de datos, en caso de que las respuestas obtenidas sean una proporción menor que 80% del total del tamaño de muestra definido, se procede a hacer un análisis de sesgo, y si se hallara que este es significativo, se trabaja nuevamente a partir del punto 8, para lo cual se debe recalendarizar la publicación de los resultados. En caso de que el sesgo no fuese significativo, se procede al punto 17.

16. Si la proporción de respuestas obtenidas es mayor que 80% del total de la muestra definida, se cierra el proceso de recepción de cuestionarios.

17. Se procede al análisis estadístico de los datos obtenidos y se obtienen conclusiones.

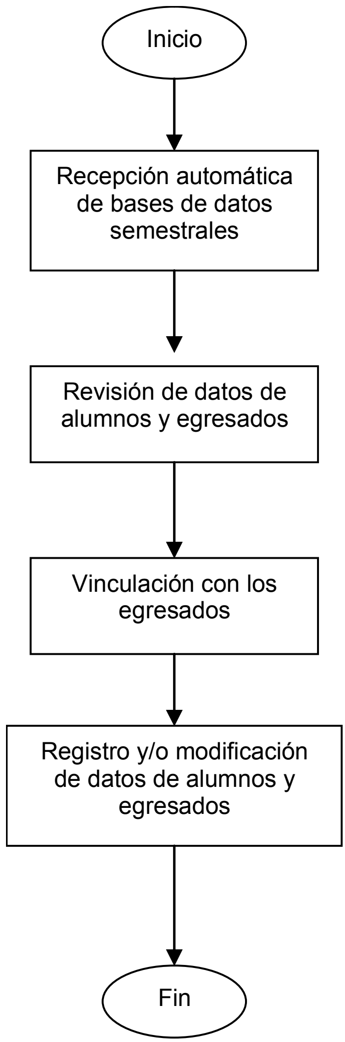
18. Se revisa y se estandariza la presentación de los resultados en un formato para su publicación.

19. Se publican los resultados en forma impresa y en web.

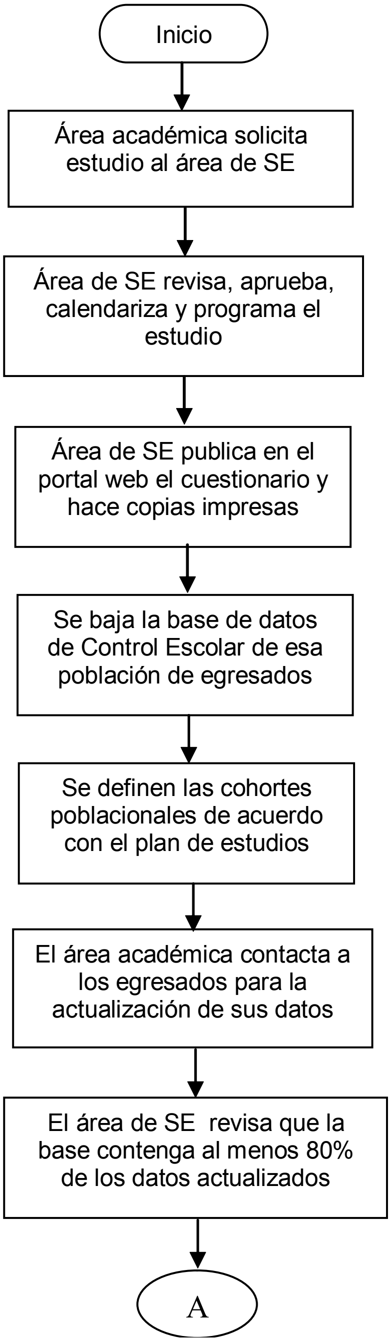
X. APÉNDICE d

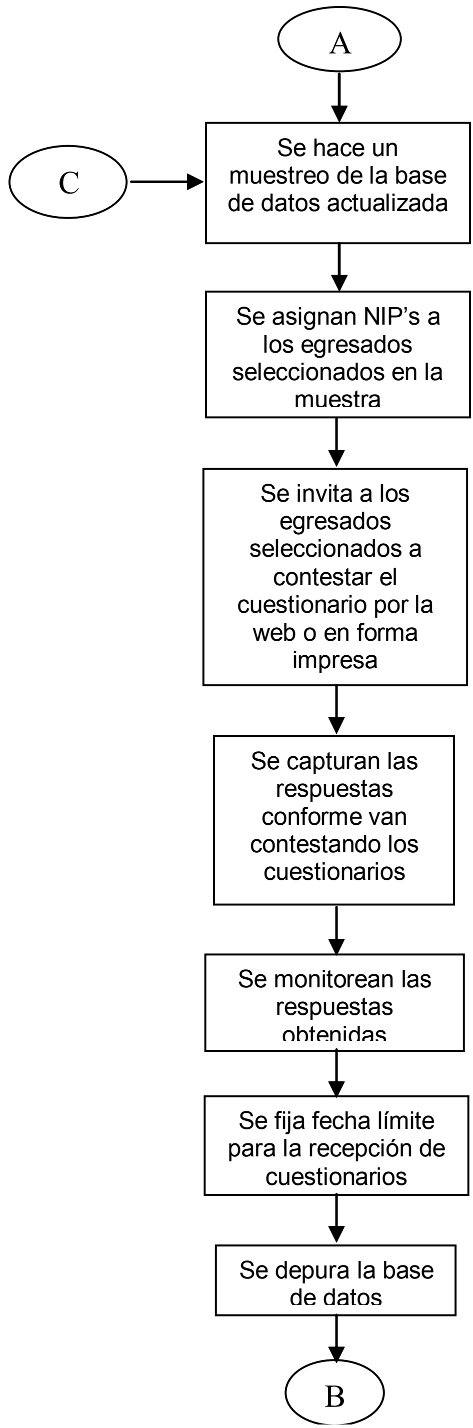
Diagramas de flujo de procedimientos

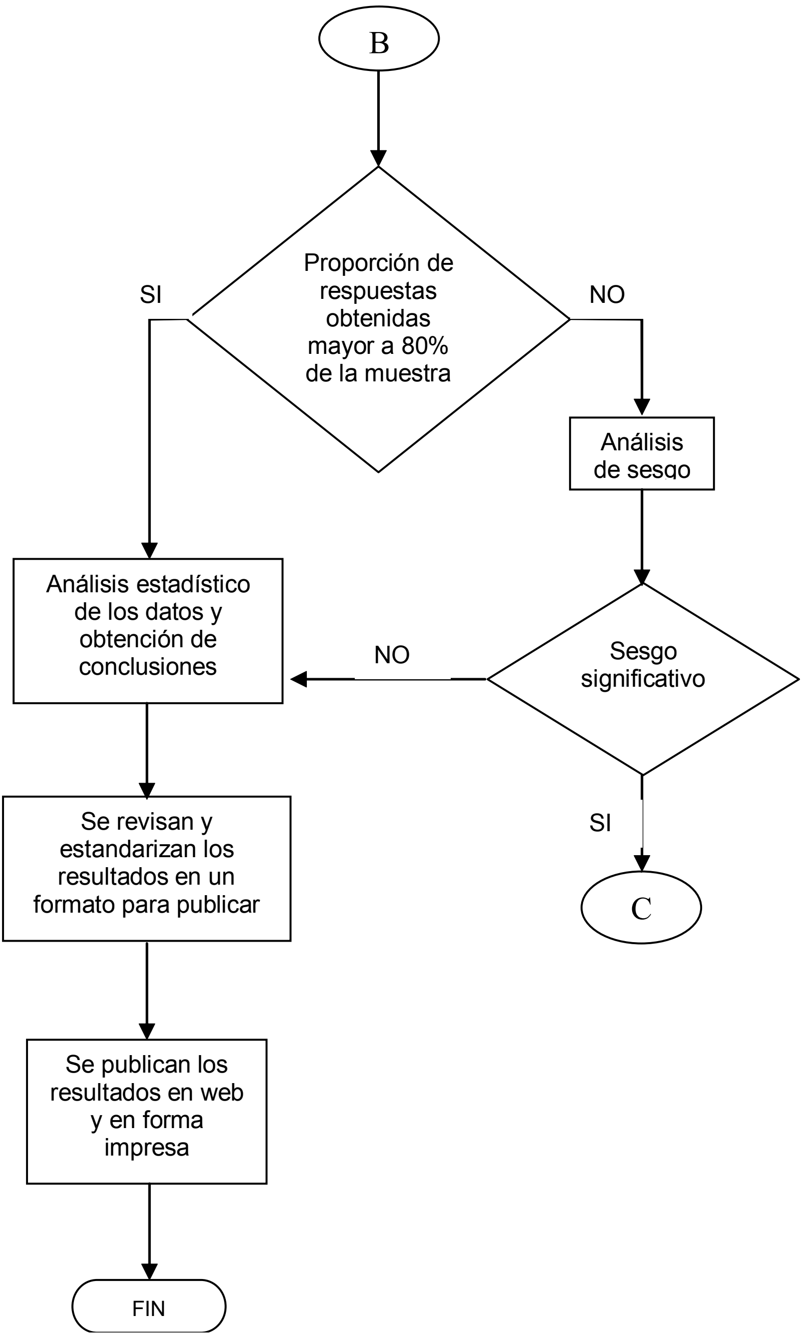
D.1 PROCEDIMIENTO 1: entre la Dirección de Control Escolar y las Dependencias de Educación Superior.



D.2 PROCEDIMIENTO 2: entre las áreas académicas y el área de Seguimiento de Egresados.




A
C
Se hace un muestreo de la base de datos actualizada
Se asignan NIP's a los egresados seleccionados en la muestra
Se invita a los egresados seleccionados a contestar el cuestionario por la web o en forma impresa
Se capturan las respuestas conforme van contestando los cuestionarios
Se monitorean las respuestas obtenidas
Se fija fecha límite para la recepción de cuestionarios
Se depura la base de datos
B


B
Proporción de respuestas obtenidas mayor a 80% de la muestra
SI
NO
Análisis de sesgo
Análisis estadístico de los datos y obtención de conclusiones
NO
Sesgo significativo
Se revisan y estandarizan los resultados en un formato para publicar
SI
C
Se publican los resultados en web y en forma impresa
FIN