

Curso de métodos estadísticos aplicados



a la investigación

Ejercicios utilizando el módulo estadístico de Excel

1. Estadísticas descriptivas para datos cuantitativos

En la hoja de cálculo "Una población", se tienen datos hematológicos, química sanguínea y concentración de mercurio en orina, de una muestra de 320 estudiantes de la Facultad de Odontología; calcule las estadísticas descriptivas para todas las variables. Vaya al menú Datos, Análisis de datos y seleccione el diálogo "Estadística descriptiva" y dele Aceptar. En Rango de entrada seleccione desde la celda B1 hasta la celda J321 (se hace la selección de todos los datos de todas las variables incluyendo los rótulos de la primera fila), luego haga clic en el recuadro "Rótulos en la primera fila"; en Opciones de salida, seleccione "En una hoja nueva" y póngale un nombre (Análisis descriptivo, por ejemplo), luego seleccione el recuadro "Resumen de estadísticas" y de aceptar.

Ubíquese en la celda A16 de la hoja de cálculo recién creada con el análisis y escriba "Coeficiente de variación", luego en la celda B16 introduzca la fórmula del coeficiente de variación (desviación estándar*100/media), luego copie la fórmula en las demás celdas con datos numéricos.

Establezca qué variables cumplen con los principios básicos de normalidad que se basan en la simetría y forma de la distribución (media \approx mediana, curtosis y coeficiente de asimetría dentro de -1.0 y +1.0). Luego establezca qué variables tienen una dispersión muy grande (coeficientes de variación > 30.0%).

2. Estimación de un promedio

Con la misma hoja de cálculo "Una población", se va a estimar con un 90, 95 y 99% de confianza el valor medio de la concentración de Hematocrito (asumiremos que la muestra es representativa solo para fines de la práctica).

Ir al menú Análisis de datos y seleccionar nuevamente Estadística descriptiva, borrar todas las selecciones que se hicieron anteriormente. Luego en Rango de entrada seleccionar los datos de la columna B (Hematocrito) incluyendo el rótulo, luego dar clic en el cuadro de Rótulos en la primera fila; seleccionar en Opciones de salida "En una hoja nueva" y póngale un nuevo nombre (Estimación, por ejemplo) y por último, seleccione los cuadros Resumen de estadísticas y Nivel de confianza para la media, escriba a la derecha de este último 90, dar aceptar. Repita el procedimiento cambiando el nivel de confianza a 95 y 99%, cuidando de cambiar el nombre cada vez, pero ya no seleccione Resumen de estadísticas.

Copie en la hoja de salida que tiene el resumen de estadísticas los dos valores calculados, elimine las hojas de cálculo repetidas y proceda a construir los intervalos de confianza. En la celda C15 escriba "Límite inferior" y en la celda D15 "Límite superior", ubíquese a la derecha de la primera celda con el valor numérico del primer nivel de confianza, y escriba la fórmula: =\$B\$3-B16 y en la siguiente del nivel superior: =\$B\$3+B16. Copie las dos fórmulas hacia abajo y discuta los resultados.

3. Prueba de hipótesis sobre la media de dos poblaciones o grupos

Se presentan las concentraciones de mercurio en orina de 320 estudiantes de la Facultad de Odontología, para determinar si existe diferencia con relación al tiempo de estudios considerando dos grupos: Estudiantes de primero y segundo año (Uno) y estudiantes de tercero a sexto año (Tres).

Vaya al menú de Análisis de datos y seleccione "Prueba F para varianza de dos muestras" y de Aceptar; seleccione los datos del grupo Uno como Rango para la variable 1 y los datos del grupo Tres como Rango para la variable 2, incluya los rótulos y haga clic en el recuadro correspondiente; luego, en Rango de salida indique la celda D3 y dar Aceptar. Interprete el resultado.

Las varianzas de ambos grupos presentan diferencia significativa, por lo que la prueba de hipótesis sobre la media de las dos poblaciones se hará con este criterio. Vaya nuevamente al menú de Análisis de datos y seleccione "Prueba t para dos muestras suponiendo varianzas desiguales" y de aceptar; luego seleccione los rangos correspondientes a cada grupo con sus rótulos y en la casilla que dice "Diferencia hipotética entre las medias" coloque el valor cero (0), luego en Rango de salida indique la celda D14 y dar aceptar. Interprete el resultado.

4. Prueba de hipótesis sobre la diferencia media de dos mediciones en un diseño pareado

En la hoja de cálculo denominada "Diseño pareado", se tienen datos sobre la efectividad de un método de enseñanza audiovisual (en formato digital) para retroalimentar el aprendizaje de los procedimientos quirúrgicos desarrollados en los cursos de grado y realizados en la Clínica de Cirugía de la Facultad de Odontología. Para el efecto, se escogieron 30 estudiantes de quinto año consistentes en 15 mujeres y 15 hombres, a cada uno de ellos se les evaluó inicialmente con un cuestionario ponderado de 0 a 100 puntos, se les presentó el método de enseñanza audiovisual y se les evaluó nuevamente (diseño pareado pre-test/post-test).

En el menú de Análisis de datos seleccione "Prueba t para medias de dos muestras emparejadas", selecciones los rangos para cada una de las variables y en la casilla que dice "Diferencia hipotética entre las medias" coloque el valor cero (0), luego en Rango de salida indique "En una hoja nueva" y póngale un nombre, de Aceptar e interprete el resultado. Si la prueba es bilateral, la selección de los rangos de entrada para cada variable puede

hacerse indistintamente en cualquier orden, sin embargo, dependiendo del sentido en el que se espere que se dé el efecto, la prueba puede interpretarse de manera unilateral, pero para ello debe tenerse en cuenta dicho sentido para seleccionar el orden del rango de las dos variables. En tal caso, como se esperaba que el método audiovisual mejorara el nivel de conocimientos en al menos 10 puntos, realice el procedimiento con las correcciones necesarias.

5. Prueba de hipótesis sobre la media de tres (o más) poblaciones o grupos: Análisis de varianza

En la hoja de cálculo llamada "Tres poblaciones", se tienen los datos de un experimento realizado en tres grupos de 25 pacientes con gingivitis, con el objeto de determinar la efectividad del digluconato de clorhexidina al 0.2% en gel y solución que se aplicaron tópica o colutorios una vez al día por un minuto por 30 días, el tercer grupo consistió en un control al que no se le aplicó ningún tratamiento. En cada paciente se midió la reducción de niveles de placa dentobacteriana (%), aunque en el grupo control no se completaron los datos.

Vaya al menú de Análisis de datos y seleccione "Análisis de varianza de un factor" y dar Aceptar; seleccione todos los datos de la variable % de reducción de la placa dentobacteriana de los tres tratamientos en un solo bloque incluyendo los rótulos; luego, en Rango de salida indique la celda E1 y dar aceptar. Interprete el resultado.

5. Asociación entre variables

El módulo de estadísticas de Excel permite realizar un análisis de asociación bivariada (dos variables a la vez) por medio del cálculo del coeficiente de correlación de Pearson, por lo que se puede hacer un análisis simultáneo de múltiples variables.

En la hoja de cálculo correspondiente a una población, vaya al menú de Análisis de datos y seleccione "Coeficiente de correlación" y dar Aceptar; en el Rango de entrada seleccione todos los datos de todas las variables incluyendo los rótulos, seleccione la opción de salida en una hoja nueva y póngale nombre (correlación, por ejemplo) y de aceptar. Establezca qué variables están asociadas considerando a priori un coeficiente de correlación menor de -0.60 o mayor de 0.60. Sin embargo, también analice qué otras variables presentan asociaciones que podrían ser de interés.

6. Análisis de regresión lineal simple

Las mediciones hematológicas hemoglobina y hematocrito están correlacionadas y la relación se debe a que a la cantidad de glóbulos rojos reflejada en el hematocrito le corresponde una concentración de hemoglobina, supongamos además, que en el laboratorio el hematocrito es más fácil de medir y presenta mayor confiabilidad. Dada esta situación, se puede desarrollar un modelo matemático (ecuación) que represente la relación de estas variables y que sirva para predecir valores. La variable independiente (x)

será por lo indicado, el hematocrito y la variable dependiente (y), la concentración de hemoglobina..

En esta hoja de cálculo, vaya al menú de análisis de datos y seleccione "Regresión" y dar aceptar; seleccione los rangos de entrada para cada variable con sus respectivos rótulos y seleccione la opción de salida en una hoja nueva con un nombre (regresión, por ejemplo) y de aceptar. Interprete los resultados.

Para visualizar esta relación, seleccione los datos de ambas variables con sus respectivos rótulos, vaya al menú "Insertar" y seleccione el primer gráfico en el menú de "Dispersión", elimine rótulos y líneas en exceso. Excel entiende que la primera variable seleccionada es "x" y la segunda "y", pero en la regresión las variables correspondientes están al revés, por lo tanto debemos corregir los ejes; para ello haga clic sobre la gráfica con el botón derecho del ratón y seleccione "Seleccionar datos", en el cuadro de diálogo que se despliega seleccione "Editar", luego cambie la selección de las variables "x" y "y", de "Aceptar" las dos veces que se indica y observe el cambio en la gráfica.

Coloque el cursor sobre cualquiera de los datos y de clic derecho al ratón, seleccione "Agregar línea de tendencia", en el menú desplegado donde dice extrapolar coloque el valor 30 en "Hacia atrás", seleccione los recuadros de "Presentar ecuación en el gráfico" y "Presentar el valor R cuadrado en el gráfico", de aceptar y coloque en mejor posición los diálogos con ecuación y R cuadrado. Interprete.