

7 pecados estadísticos mortales que incluso los expertos cometen

Cometer errores estadísticos es muy fácil. El software estadístico ayuda a eliminar los errores matemáticos, pero interpretar correctamente los resultados de un análisis puede ser incluso más difícil. Hasta los expertos en estadística suelen ser víctimas de errores comunes. Los instructores técnicos de Minitab, todos especialistas en la estadística, con una amplia experiencia en la industria, compilaron esta lista de errores estadísticos comunes con los que se debe tener mucho cuidado.

1: No distinguir entre significancia estadística y práctica

Muestras muy grandes permiten detectar diferencias muy pequeñas. Pero el hecho de que exista una diferencia, no significa que sea importante. "Corregir" una diferencia estadísticamente significativa que no tiene ningún efecto práctico es un desperdicio de tiempo y dinero.

Una empresa de alimentos llena 18,000 cajas de cereal por turno, con un peso objetivo de 360 gramos y una desviación estándar de 2.5 gramos. Un sistema automatizado de medición pesa cada caja al final de la línea de llenado. Con esa cantidad de datos, la empresa puede detectar una diferencia de 0.06 gramos en el peso medio de llenado el 90% de las veces. Pero eso equivale a solo uno o dos pedacitos de cereal, lo cual no es suficiente como para prestarle atención o preocuparse.

El desplazamiento de 0.06 gramos es estadísticamente significativo, pero no tiene significancia práctica. La recopilación automática de datos y las bases de datos de gran tamaño hacen que esto sea más común, así que considere el efecto práctico de una diferencia estadísticamente significativa. Especificar el tamaño de una diferencia significativa cuando calcule el tamaño de la muestra para una prueba de hipótesis le ayudará a evitar este error.

2: Malinterpretar los intervalos de confianza que se superponen

Al comparar las medias, con frecuencia examinamos los intervalos de confianza para ver si se superponen. Cuando los intervalos de confianza de 95% para las medias de dos poblaciones independientes no se superponen, habrá una diferencia estadísticamente significativa entre las medias.

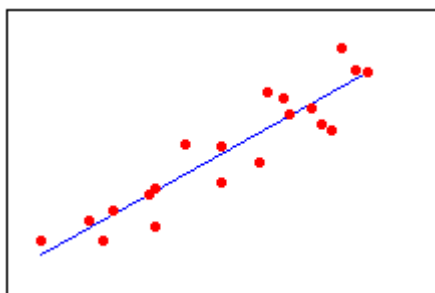
Sin embargo, lo opuesto no es necesariamente cierto. Dos intervalos de confianza de 95% pueden superponerse aun cuando la diferencia entre las medias sea estadísticamente significativa. Para las muestras que se observan en la gráfica de intervalo, el valor p de la prueba t de 2 muestras es menor que 0.05, lo que confirma una diferencia estadística entre las medias; sin embargo, los intervalos se superponen considerablemente. ¡Examine sus gráficas, pero también el resto de la salida antes de concluir que no existe ninguna diferencia!



A pesar de que los niveles de confianza se superpongan, la prueba t de 2 muestras encontró una diferencia estadísticamente significativa entre los tratamientos.

3: Presuponer que correlación = causalidad

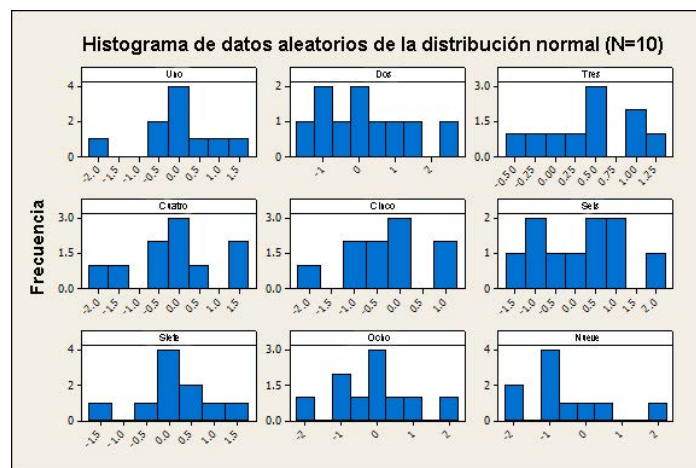
Cuando se observa una correlación, una asociación lineal entre dos variables, es tentador concluir que un cambio en una variable causa un cambio en la otra. Pero la correlación no significa que exista una relación de causa y efecto.



Supongamos que estamos analizando datos que muestran una fuerte correlación entre las ventas de helados y los índices de homicidios. Cuando las ventas de helados son altas en el verano, el índice de homicidios es alto. Eso no significa que las ventas de helados causen homicidios. Dado que las ventas de helados y el índice de homicidios alcanzan su mayor nivel en el verano, los datos más bien sugieren que ambos se ven afectados por otro factor: el clima.

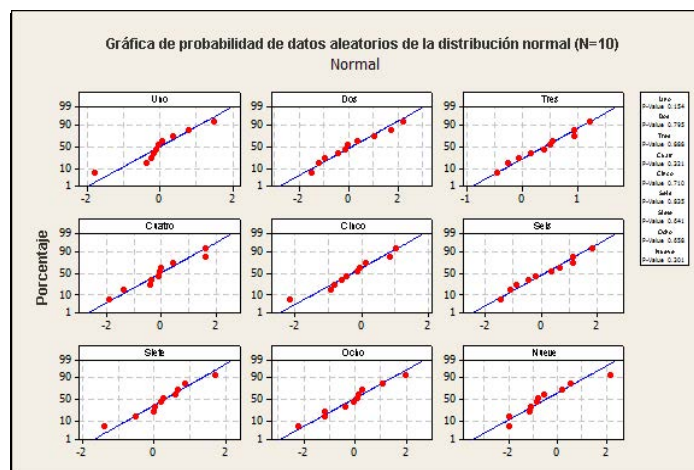
4: Rechazar la normalidad sin razón

Muchos métodos estadísticos se basan en el supuesto de que los datos provengan de una distribución normal, es decir, que sigan una curva en forma de campana. Las personas suelen crear histogramas para confirmar que sus datos siguen la forma de la curva de campana, pero los histogramas, como el que se muestra a continuación, pueden ser engañosos.



Extrajimos estas 9 muestras de la distribución normal, pero ninguno de los histogramas tiene forma de campana.

Al examinar los mismos datos en una gráfica de probabilidad, es más fácil ver que siguen la distribución normal. Cuanto más cerca estén los puntos de datos de la línea azul, mejor seguirán la distribución normal.



Si tiene menos de 50 puntos de datos, una gráfica de probabilidad es una mejor opción que un histograma para evaluar visualmente la normalidad.

7: Hacer inferencias sobre una población que la muestra no representa

Con la estadística, las muestras pequeñas nos permiten hacer inferencias sobre poblaciones enteras. Pero evite hacer inferencias sobre una población que la muestra no represente. Por ejemplo:

- En el análisis de capacidad, los datos de un solo día se utilizan erróneamente para estimar la capacidad de todo el proceso de manufactura.
- En el muestreo de aceptación, se seleccionan muestras de una sección del lote para todo el análisis.
- En un análisis de fiabilidad, solo se incluyen las unidades que no pasaron, pero la población es todas las unidades producidas.

Para evitar estas situaciones, defina cuidadosamente su población antes del muestreo y tome una muestra que realmente la represente.

¿El peor pecado? ¡No buscar ayuda!

La mayoría de los especialistas en estadística tienen de 4 a 8 años de educación en estadística y por lo menos 10 años de experiencia en el mundo real; sin embargo, los empleados que cursan programas de capacitación básica en estadística a veces piensan que saldrán como expertos en la materia. Nadie lo sabe todo acerca de la estadística e incluso los expertos consultan con otras personas cuando se atascan. ¡Si encuentra un problema estadístico que parece ser demasiado difícil, no tenga miedo de pedir ayuda a analistas con más experiencia!