¿Por qué dividimos por $n-1$ para calcular la varianza muestral?

MAURICIO FUENTES ALBURQUENQUE*

PROGRAMA DE BIOESTADÍSTICA, ESCUELA DE SALUD PÚBLICA
FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE

La respuesta breve: el insesgamiento

Cuando se estudia estadística descriptiva, los textos señalan generalmente (o quizás siempre) que para calcular la varianza s^2 de una muestra de tamaño n se usa como denominador la cantidad $n-1$, es decir,

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}, \tag{1}$$

donde x_i es el valor de la variable aleatoria X observada en el i-ésimo elemento o sujeto de la muestra, y $\bar{x}$ es la media de la muestra para dicha variable.

Si uno mira la suma de cuadrados del numerador en (1), que tiene n términos, resultaría más natural e intuitivo usar como denominador esa misma cantidad n, ya que nos entregaría la diferencia cuadrática media de la muestra. La expresión resultante, que también es válida como estadística muestral y que llamaremos $\tilde{s}^2$ para distinguirla de la anterior, sería[1]

$$\tilde{s}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}. \tag{2}$$

Entonces, ¿por qué se recomienda usar (1) y no (2)? La respuesta es breve y precisa: porque la estadística muestral (1) entrega una estimación insesgada de la varianza poblacional σ^2, mientras que (2) no.

Una (no tan) breve demostración

Con el fin de mantener cierta rigurosidad, denotemos los estimadores representados por (1) y (2) como variables aleatorias, es decir,

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \tag{3}$$

y

$$\tilde{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}, \tag{4}$$

donde $X_1, \ldots, X_n$ es una muestra aleatoria de tamaño n y $\bar{X}$ es la respectiva media muestral[2].

Para demostrar el insesgamiento de S^2, recordemos primero que un estimador $\hat{\theta}$ de un parámetro θ es insesgado si su esperanza es igual al parámetro, es decir, si $\mathrm{E}[\hat{\theta}] = \theta$. En este caso, se cumple que $\mathrm{E}[S^2] = \sigma^2$.

La demostración es bastante directa y la podemos encontrar en textos clásicos de Estadística y Probabilidad, como Canavos (1988), Meyer (1992) y Mood et al. (1974), entre otros. Antes tomemos nota de algunos supuestos y propiedades importantes:

1. Tenemos una muestra aleatoria $X_1, \ldots, X_n$ proveniente de una población con media μ y varianza finita σ^2 (notar que la población no es necesariamente normal).
2. Dado que las variables aleatorias X_i son *iid* (independientes e idénticamente distribuidas), de la definición de varianza tenemos que
$$\mathrm{E}[(X_i - \mu)^2] = \sigma^2.$$
3. La media muestral es $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$. Entonces $\mathrm{E}[\bar{X}] = \mu$, y
$$\mathrm{E}[(\bar{X} - \mu)^2] = \mathrm{Var}[\bar{X}] = \frac{\sigma^2}{n}.$$
4. El numerador de S^2 se puede expresar como (ver Apéndice 1)
$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

Entonces,

$$\begin{aligned}
\mathrm{E}[S^2] &= \mathrm{E}\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}\right] \\
&= \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathrm{E}[(X_i - \mu)^2] - n\mathrm{E}[(\bar{X} - \mu)^2]\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\right) \\
&= \frac{1}{n-1}(n\sigma^2 - \sigma^2) \\
&= \frac{1}{n-1}(n-1)\sigma^2 \\
&= \sigma^2.
\end{aligned}$$

*mauricio.fuentes@uchile.cl.

[1] En efecto, la expresión (2) corresponde a la estimación muestral de la varianza obtenida mediante el método de máxima verosimilitud, para una población distribuida normal.

[2] Es conveniente hacer la distinción entre una variable aleatoria U y su realización u, donde esta última es el valor específico que toma U en una observación particular. Por este motivo vemos que en los textos de Estadística se escribe, por ejemplo, la probabilidad $\mathbb{P}[U = u]$.

El sesgo al dividir por n

En cuanto al sesgo de $\tilde{S}^2$, aplicando los mismos pasos obtenemos

$$\mathrm{E}[\tilde{S}^2] = \frac{n-1}{n}\sigma^2. \tag{5}$$

El resultado (5) muestra que $\tilde{S}^2$ subestima el valor de la varianza poblacional, aunque esta subestimación sería despreciable para tamaños de muestra grandes. Por ejemplo, para un tamaño de muestra de 10 sujetos, la estimación es del orden de un 10 % por debajo del valor de σ^2, mientras que para una muestra de tamaño 100 sería del orden de 1 % inferior. En cualquier caso, la costumbre más «sana» que podemos adoptar es usar siempre la expresión (1).

¿Y la desviación estándar?

Lamentablemente, no podemos ser tan optimistas y pensar que el hecho que S^2 sea un estimador insesgado de σ^2 implique que $\mathrm{E}[S] = \sigma$. La raíz cuadrada, aplicada a S^2, es una función cóncava e introduce un sesgo adicional.

Para el caso en que la variable aleatoria sigue una distribución normal hay una expresión para $\mathrm{E}[S]$, en la que aparece el tamaño de la muestra y la función gamma (ver, por ejemplo, Canavos (1988)). El sesgo también es a la baja y despreciable para muestras grandes. De este tema quizás nos ocupemos en un futuro artículo.

Comentarios y una (breve) comprobación

Si bien fue Gauss quien demostró el insesgamiento de S^2 (Gauss, 1823), y de quien se dice lo habría estado utilizando desde años antes, fue por el trabajo del astrónomo alemán Friedrich Wilhelm Bessel que se hizo más conocido, al punto que la división por $n-1$ se conoce como *corrección de Bessel* (Hald, 2004; Kenney & Keeping, 1951; Nobach, 2017; Weisstein, 2026). Más tarde, en 1876, el geodesta Friedrich Robert Helmert derivó la distribución de S^2, más de tres décadas antes que Student (Hald, 2004; Zabell, 2008).

Aunque el insesgamiento de S^2 es un hecho objetivo y demostrable matemáticamente, resulta contraintuitivo, al menos en principio, que no se deba dividir por n. La explicación más común hace referencia al esquivo y a menudo abstruso concepto de *grados de libertad*. Así, la explicación habitual es que, dado que las n diferencias cuadráticas están calculadas respecto a la media muestral (que además es función de los mismos datos), solo $n-1$ de ellas pueden variar libremente debido a la restricción $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$. Por lo tanto, una vez establecidas $n-1$ diferencias (o residuos), la última ya no puede variar libremente, es decir, pierde su grado de libertad.

Debemos tener en cuenta, eso sí, que la independencia de $n-1$ residuos se cumple para una muestra aleatoria con su respectiva media muestral $\bar{X}$. Si la suma cuadrática se realiza en torno a la media poblacional, es decir, $\sum_{i=1}^{n}(X_i - \mu)^2$, esta tendría n grados de libertad y ese sería el denominador adecuado. Sin embargo, en la práctica esto carece de sentido ya que lo habitual es no conocer μ.

Otra explicación que se puede encontrar, por ejemplo en Snedecor (1948), es que una muestra tiene poca probabilidad de contener los valores extremos de la población, con lo cual la media de la suma cuadrática (dividiendo por n) es menor que la media de la suma cuadrática poblacional (dividiendo por el tamaño N de la población). Por lo tanto, al dividir por $n-1$ se corrige esta subestimación. El carácter *ad hoc* de esta explicación es bastante evidente.

Podemos comprobar el insesgamiento de S^2 y, a la vez, el sesgo de $\tilde{S}^2$ para algunas distribuciones particulares[3]. Para esto, se simularon tres poblaciones de tamaño 5000, con distribuciones normal $N(10, 4)$, exponencial $exp(3)$[4] y binomial $bin(10, 1/5)$. De cada población se extrajeron muestras aleatorias de distintos tamaños, desde 5 hasta 100, con 10000 repeticiones cada una, calculando los valores medios de S^2 y $\tilde{S}^2$ (ver código de R para la población normal en Apéndice 2).

Los resultados de estas simulaciones se muestran en las figuras 1 a 3, donde podemos comprobar que el promedio de S^2 está en torno al valor poblacional para todos los tamaños de muestra, mientras que la media de $\tilde{S}^2$ subestima dicho valor, acercándose a medida que el tamaño muestral aumenta.

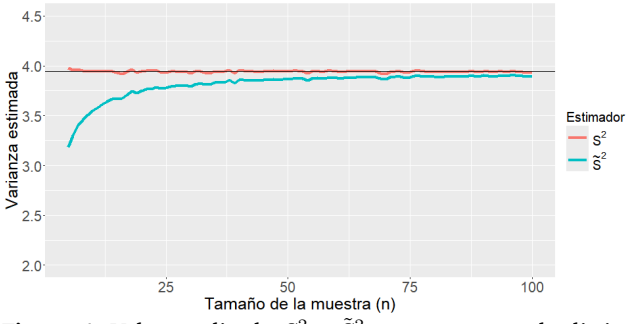


Figura 1. Valor medio de S^2 y $\tilde{S}^2$ para muestras de distinto tamaño (10000 repeticiones), extraídas de una población simulada normal con $\mu = 10$ y $\sigma^2 = 4$ de tamaño 5000. La línea horizontal indica la varianza verdadera de la población simulada.

[3] No podríamos verificar empíricamente la propiedad de insesgamiento ya que tendríamos que comprobarla para las infinitas distribuciones posibles, por lo que solo somos capaces de comprobar casos particulares. No obstante, no deberíamos dudar de la generalidad de la demostración matemática (¿o sí?).

[4] En este caso $\lambda = 3$ es el parámetro de escala (o la media) y no la tasa, que sería $1/3$.

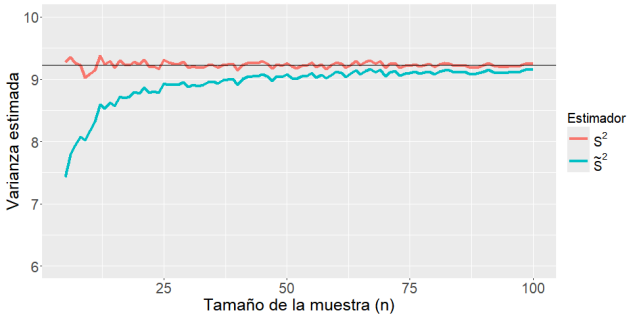


Figura 2. Valor medio de S^2 y $\tilde{S}^2$ para muestras de distinto tamaño (10000 repeticiones), extraídas de una población exponencial con parámetro de escala $\lambda = 3$. La línea horizontal indica la varianza verdadera de la población simulada.

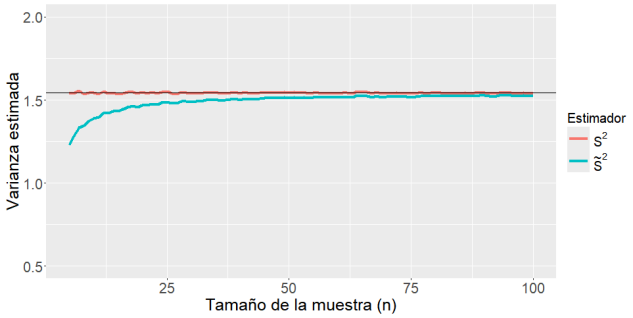


Figura 3. Valor medio de S^2 y $\tilde{S}^2$ para muestras de distinto tamaño (10000 repeticiones), extraídas de una población binomial con $m = 10$ y $\pi = 1/5$. La línea horizontal indica la varianza verdadera de la población simulada.

Agradecimientos: A los profesores Sergio Alvarado y Rodrigo Villegas por sus valiosos comentarios, sugerencias y correcciones.

Referencias

Canavos, G. C. (1988). *Probabilidad y Estadística. Aplicaciones y métodos*. McGraw-Hill.

Gauss, C. F. (1823). Theoria combinationis observationum erroribus minimis obnoxiae. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores*, *5*, 33-90.

Hald, A. (2004). *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713 to 1935*. University of Copenhagen.

Kenney, J. F., & Keeping, E. S. (1951). *Mathematics of Statistics, part II* (2nd edition). Van Nostrand.

Meyer, P. L. (1992). *Probabilidad y aplicaciones estadísticas* (2a edición). Addison-Wesley Iberoamericana.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics* (3rd edition). McGraw-Hill.

Nobach, H. (2017). Practical Realization of Bessel's Correction for a Bias-Free Estimation of the Auto-Covariance and the Cross-Covariance Functions. https://arxiv.org/abs/2303.11047

Snedecor, G. W. (1948). *Métodos de Estadística. Su aplicación a experimentos en Agricultura y Biología* (4a edición). Acme Agency.

Weisstein, E. W. (2026). Bessel's Correction. *MathWorld — A Wolfram Resource*. https://mathworld.wolfram.com/BesselsCorrection.html

Zabell, S. L. (2008). On Student's 1908 Article "The Probable Error of a Mean". *Journal of the American Statistical Association*, *103*(481), 1-7. https://doi.org/10.1198/016214508000000030

Apéndice 1

En la suma de cuadrados del numerador de S^2 sumemos y restemos la media μ, con lo que tenemos

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2 \\
&= \sum_{i=1}^{n}(X_i - \mu - (\bar{X} - \mu))^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) + \sum_{i=1}^{n}(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)(n\bar{X} - n\mu) + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2.
\end{aligned}$$

Apéndice 2

```
library("ggplot2")

N <- 5000 # tamaño de la población
mu <- 10 # media
sigma2 <- 4 # varianza
set.seed(2026)
X <- rnorm(n = N, mean = mu, sd = sqrt(sigma2)) # población simulada
sigma2_pob <- ((N-1)/N)*var(X) # varianza de la población simulada (dividida por N)
n_min <- 5 # mínimo tamaño de muestra
n_max <- 100 # máximo tamaño de muestra

# Cálculo de promedios de S^2 para los distintos tamaños de muestra
s2_bar <- numeric()
set.seed(2026)
for (n in n_min:n_max) {
  s2 <- numeric()
  k <- 10000
  for (i in 1:k) {
    x <- sample(X, n, replace = TRUE)
    s2[i] <- var(x)
  }
  s2_bar[n-4] <- mean(s2)
}

# Cálculo de promedios de Sn^2 (dividida por n) para los distintos tamaños de muestra
sn2_bar <- numeric()
set.seed(2026)
for (n in n_min:n_max) {
  sn2 <- numeric()
  k <- 10000
  for (i in 1:k) {
    x <- sample(X, n, replace = TRUE)
    sn2[i] <- ((n-1)/n)*var(x)
  }
  sn2_bar[n-4] <- mean(sn2)
}

# Base de datos con resultados
datos <- data.frame(n = rep(seq(n_min,n_max),2),
                    grupo = factor(rep(c(1,2), each = n_max-n_min+1)),
                    s = c(s2_bar,sn2_bar))

# Gráfico de resutados
png(filename="grafico.png", width=960, height=480, bg="transparent", pointsize=24)
ggplot(datos, aes(x=n, y=s, colour=grupo, group=grupo)) + geom_line(linewidth=1.5) +
  labs(x="Tamaño de la muestra (n)", y="Varianza estimada", colour="Estimador") +
  theme(axis.title.x=element_text(size=25), axis.title.y=element_text(size=25),
        axis.text.x=element_text(size=22), axis.text.y=element_text(size=22),
        plot.title=element_text(vjust = -0.5,size=26),
        legend.title=element_text(size=20), legend.text=element_text(size=20)) +
  scale_colour_discrete(labels=c(expression(S^2),expression(tilde(S)^2))) +
  ylim(2,4.5) + geom_hline(yintercept=sigma2_pob)
dev.off()
```