MS Module 1 Background and data visualization – practice problems

(The attached PDF file has better formatting.)

Exercise 1.1: Bias

An estimator has a standard error of 2 and a mean squared error of 6.

A. What is the variance of the estimator?
B. What is the bias of the estimator?

Part A: The variance is the square of the standard error = 2^2 = 4.

Part B: The mean squared error = the variance + the square of the bias ⇛

the bias = the square root of (the mean squared error – the variance) = $\sqrt{}(6 – 2^2)$ = $\sqrt{}2$,

Exercise 1.2: Bias, variance, and mean squared error

- Let D be the number of deaths in a sample of N = 1,000 men age 40.
- An actuary estimates the mortality rate as (D + 2) / (N + 4).
- The true mortality rate (*p*) is 5%.

A. What is the bias of the estimator?
B. What is the variance of the estimator?
C. What is the mean squared error?

Part A: The bias of the estimator is

$E(x+2 / n+4) – p = (1/ n+4)E(x+2) – p = (np+2) / (n+4) – p = (2/n – 4p/n) / (1 + 4/n) =$

$(2/1{,}000 – 4 \times 5\% / 1{,}000) / (1 + 4/1{,}000) = 0.001793$

Part B: The variance of the estimator is

$var ((x+2) / (n+4)) = 1/(n+4)^2 \times var(x+2) = var(x) / (n+4)^2 = np(1–p)/(n+4)^2 = p(1–p)/(n+8 + 16/n) =$

$(5\% \times (1 – 5\%)) / (1{,}000 + 8 + 16/1{,}000) = 0.000047$

Part C: The mean squared error is

$MSE = p(1–p)/(n+8 + 16/n) + [(2/n – 4p/n) / (1 + 4/n)]^2 =$

$(5\% \times (1 – 5\%)) / (1{,}000 + 8 + 16/1{,}000) + ((2/1{,}000 – 4 \times 5\% / 1{,}000) / (1 + 4/1{,}000))^2 = 0.000050$

Exercise 1.3: Bias, variance, and mean squared error

- Let D be the number of deaths in a sample of N = 1,000 men age 40.
- An actuary estimates the mortality rate as (D + 2) / (N + 4).
- The standard deviation of the estimator is 0.01.

A. What is the variance of the estimator?
B. What is the mortality rate (*p*)?
C. What is the bias of the estimator?
D. What is the mean squared error?

Part A: The variance is the square of the standard deviation = 0.01^2 = 0.0001.

Part B: The variance of the estimator is

$\text{var}((x+2)/(n+4)) = 1/(n+4)^2 \times \text{var}(x+2) = \text{var}(x)/(n+4)^2 = np(1–p)/(n+4)^2 = p(1–p)/(n+8 + 16/n) \Rrightarrow$

p(1–p) = variance × (n+8 + 16/n) = 0.0001 × (1000 + 8 + 16/1000) = 0.1008016

The solutions are $(+1 \pm ((–1)^2 – 4 \times 1 \times 0.1008016)^{0.5}) / (2 \times 1)$:

- $(1 + ((–1)^2 – 4 \times 1 \times 0.1008016)^{0.5}) / (2 \times 1) = 0.886262$
- $(1 – ((–1)^2 – 4 \times 1 \times 0.1008016)^{0.5}) / (2 \times 1) = 0.113738$

The mortality rate is 88.63% or 11.37%.

Part C: The bias of the estimator is

E(x+2 / n+4) – p = (1/ n+4)E(x+2) – p = (np+2) / (n+4) – p = (2/n – 4p/n) / (1 + 4/n) =

(2/1,000 – 4 × 11.3738% / 1,000) / (1 + 4/1,000) = 0.001539

Part D: The mean squared error = the variance + the square of the bias =

$\text{MSE} = p(1–p)/(n+8 + 16/n) + [(2/n – 4p/n) / (1 + 4/n)]^2 = 0.0001 + 0.001539^2 = 0.000102$

Exercise 1.4: Bias, variance, and mean squared error

- Let D be the number of deaths in a sample of N = 1,000 men age 40.
- An actuary estimates the mortality rate as (D + 2) / (N + 4).
- The bias of the estimator is 0.0015.

A. What is the mortality rate (*p*)?
B. What is the variance of the estimator?
C. What is the mean squared error?

Part A: The bias of the estimator is

E(x+2 / n+4) – p = (1/ n+4)E(x+2) – p = (np+2) / (n+4) – p = (2/n – 4p/n) / (1 + 4/n).

Solving for the mortality rate *p* gives

p = (2/n – bias × (1 + 4/n)) / (4/n) = (2 / 1000 – 0.0015 × (1 + 4/1000)) / (4 / 1000) = 0.123500

Part B: The variance of the estimator is

var ((x+2) / (n+4)) = $1/(n+4)^2$ × var(x+2) = var(x) / $(n+4)^2$ = np(1–p)/$(n+4)^2$ = p(1–p)/(n+8 + 16/n) =

(12.35% × (1 – 12.35%)) / (1,000 + 8 + 16/1,000) = 0.000107.

Part C: The mean squared error is

MSE = p(1–p)/(n+8 + 16/n) + $[(2/n – 4p/n) / (1 + 4/n)]^2$ =

(12.35% × (1–12.35%)) / (1,000 + 8 + 16/1,000) + $((2/1,000 – 4 × 12.35\%/1,000) / (1 + 4/1,000))^2$ = 0.000110

Exercise 1.5: Normal distribution

A statistician draws a sample of 8 values from a normal distribution.

The summary statistics are

- $\sum x_i = 7{,}113$
- $\sum x_i^2 = 6{,}991{,}551$

A. What is the estimate of μ?
B. What is the estimated variance of the population?
C. What is the estimated standard deviation of the population?
D. What is the standard error of the sample mean?
E. What is the maximum likelihood estimator of σ, the standard deviation of the population?

Part A: The estimate of μ is 7,113 / 8 = 889.125

Part B: The estimated variance of the population is

$(6{,}991{,}551 – (7{,}113)^2 / 8) / (8 – 1) = 95{,}314.98$

Part C: The estimated standard deviation of the population is $95{,}314.98^{0.5} = 308.73$

Part D: The standard error of the sample mean = $308.73 / 8^{0.5} = 109.152538$

Part E: The maximum likelihood estimator of σ, the standard deviation of the population, is

$((6{,}991{,}551 – (7{,}113)^2 / 8) / 8)^{0.5} = 288.79$

The maximum likelihood estimator uses N instead of N-1 as the denominator, so it is biased for both the variance and the standard deviation.

Exercise 1.6: Mortality rate

An actuary studying mortality rates in Country W finds 20 deaths among 40,000 men age 40. The estimated mortality rate is the number of deaths / the sample size.

A. What is the estimated variance of the mortality rate for men age 40?
B. What is the standard error of the mortality rate for men age 40?

Part A: The mortality rate is a percentage, and the variance of a sample percentage *p* is np(1–p).

The estimated mortality rate is 20 / 40,000 = 0.00050, so the estimated variance is

0.0005 × (1 – 0.0005) / 40,000 = 0.000000012494

Part B: The standard error of the mortality rate for men age 40 is

$(0.0005 \times (1 - 0.0005) / 40{,}000)^{0.5} = 0.000112$