[8.2] Statistics.

DEF: A function of observable random variables, $T = t(X_1, X_2, \cdots, X_n)$ which does not depend on any unknown parameters, is called a statistic

Ex: The sample mean is a statistic since

$$\bar{X} = t(X_1, X_2, \cdots, X_n) = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

does not depend on any parameters

Thm: If $X_1, X_2, \dots, X_n$ denotes a RS from $f(x)$ with $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$

A statistic for which $E(\text{stat}) = \text{parameter}$, we say that stat is an <u>unbiased</u> estimate of the parameter.

EX: $Y \sim HYP(n, M, N)$

Suppose we want to estimate $p = \frac{M}{N}$.

Then $\frac{Y}{n}$ is an unbiased estimate $E\left(\frac{Y}{n}\right) = \frac{\cancel{n}M}{\cancel{n}N} = \frac{M}{N}$

Thus, $\frac{Y}{n}$ is an estimator $\frac{M}{N}$

Note: $\operatorname{Var}\left(\frac{Y}{n}\right) = \frac{1}{n^2}\, n \frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$

$$= \frac{1}{n}\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Note: when $n=N$ (you've sampled everything)

then $\operatorname{Var}\left(\frac{Y}{n}\right) = \frac{1}{N}\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-N}{N-1}\right) = 0$

It follows as $n \to N$, $\operatorname{Var}\left(\frac{Y}{n}\right) \to 0$

Ex: Sample Variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{\sum_{i}^{n} x_i - n\mu}{n-1}$$

Thm: If $x_1, x_2, \ldots$ denotes a RS of size n from $f(x)$ with $E(x)=\mu$, $Var(x)=\sigma^2$, then

$$E(S^2) = \sigma^2$$

$$Var(S^2) = \left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)/n$$

This shows that S^2 is an unbiased estimator of σ^2

When $x_i \sim N(\mu,\sigma^2)$, $Var(S^2) = \frac{2\sigma^4}{n-1}$