MthStat 465, Spring 2005, Lecture Number 23
Unbiased Estimates

A fundamental aim of statistics is to use data to determine an appropriate probability model. For example, if the random variables R_1, $R_2, \ldots, R_N$ are to represent the heights of tomato plants, ideally we would like to know what the distribution function is for each of these random variables, for then we could compute the expected value of each random variable, the variance, and so forth, and then make predictions about the heights of future tomato plants.

If we are less greedy, we could simply try to say what the mean of this distribution is, or perhaps what the variance is.

In general, we say that a random variable R is an **unbiased estimate** of the quantity a if $\mathrm{E[R] = a}$. In what follows, we will assume that the random variables R are independent and all have the same distribution function F. Futhermore, we assume that each random variable has expected value equal to μ (μ for **m**ean) and standard deviation σ (σ for **s**tandard deviation). We would like to find unbiased estimates of μ and σ^2.

0.1. **Some terminology.** In the set up we have here, the theoretical quantities are described as being those of the **population**, and the quantities derived from the data are described as those of the **sample**. Thus the distribution of the R_k is called the **population distribution**, μ is called the population mean, and σ is called the **population standard deviation**, while $\overline{x}$ and s are called the **sample mean** and **sample standard deviation** respectively.

0.2. **The minimum variance linear unbiased estimate of the population mean, μ.** Let us look at μ first. A reasonable idea would be the use the average of the data, that is

$$m(R_1, \ldots, R_N) := N^{-1}(R_1 + \cdots + R_N).$$

Since we know that the expected value of a sum is the sum of the expected values and that expected values scale,

$$\begin{aligned} \mathrm{E[m(R_1, \ldots, R_N)]} &= N^{-1}(\mathrm{E[R_1]} + \cdots + \mathrm{E[R_N]}) \\ &= N^{-1}(N\mu) \\ &= \mu, \end{aligned}$$

so the obvious guess gives us an unbiased estimate. However, suppose that we have some constants a_k and we look at $a_1R_1 + \cdots + a_NR_N$ as a potential unbiased estimate of μ:

$$\begin{aligned} \mathrm{E[a_1R_1 + \cdots + a_NR_N]} &= a_1\mathrm{E[R_1]} + \cdots + \mathrm{a_N E[R_N]} \\ &= a_1\mu + \cdots + a_N\mu \\ &= (a_1 + \cdots + a_N)\mu \end{aligned}$$

We see that it is an unbiased estimate so long as $a_1 + \cdots + a_N = 1$. This is true, for example, when $a_1 = \cdots = a_N = 1/N$. Why, then, do we use $m(R_1, \ldots, R_N)$. The answer is that $m(R_1, \ldots, R_N)$ has the lowest variance. Observe that since the

R_k are independent with common standard deviation σ we have

$$\begin{aligned} \mathrm{Var}[a_1R_1 + \cdots + a_NR_N] &= \mathrm{Var}[a_1R_1] + \cdots + \mathrm{Var}[a_NR_N] \\ &= a_1^2\mathrm{Var}[R_1] + \cdots + a_N^2\mathrm{Var}[R_N] \\ &= (a_1^2 + \cdots + a_N^2)\sigma^2 \end{aligned}$$

It can be shown that if $a_1 + \cdots + a_N = 1$ then $a_1^2 + \cdots + a_N^2 \geq 1/N$, and that equality holds if and only if $a_1 = \cdots = a_N = 1/N$. Here is a simple derivation of this fact. Put $a_k = (1/N) + \Delta_k$. Then

$$\begin{aligned} 1 &= a_1 + \cdots + a_N \\ &= (1/N) + \Delta_1 + \cdots + (1/N) + \Delta_N \\ &= 1 + \Delta_1 + \cdots + \Delta_N \end{aligned}$$

so $\Delta_1 + \cdots + \Delta_N = 0$. Therefore

$$\begin{aligned} a_1^2 + \cdots + a_N^2 &= ((1/N) + \Delta_1)^2 + \cdots + ((1/N) + \Delta_N)^2 \\ &= (1/N)^2 + (2/N)\Delta_1 + \Delta_1^2 + \cdots + (1/N)^2 + (2/N)\Delta_N + \Delta_N^2 \\ &= N(1/N)^2 + (2/N)(\Delta_1 + \cdots + \Delta_N) + \Delta_1^2 + \cdots + \Delta_N^2 \\ &= (1/N) + \Delta_1^2 + \cdots + \Delta_N^2. \end{aligned}$$

Therefore $a_1^2 + \cdots + a_N^2 \geq 1/N$ with equality if and only if $\Delta_1^2 + \cdots + \Delta_N^2 = 0$. The latter occurs if and only if $\Delta_1 = \cdots = \Delta_N = 0$, that is, if and only if $a_1 = \cdots = a_N = 1/N$.

0.3. **Linear unbiased estimates of the population variance, σ^2.** We have seen that the sample standard deviation of the data $(r_1, \ldots, r_N)$ is defined by the expression

$$s(r_1, \ldots, r_N) = \sqrt{\frac{1}{N-1}\left((r_1 - \overline{r})^2 + \cdots + (r_N - \overline{r})^2\right)}$$

where $\overline{r}$ is the sample mean $N^{-1}(r_1 + \cdots + r_N)$. A natural question is "Why should we divide by $N - 1$ and not N?". As we will see, the random variable $(s(R_1, \ldots, R_N))^2$ is an unbiased estimate of the population variance σ^2.

To see this, let $\overline{R} := N^{-1}(R_1 + \cdots + R_N)$. The random variable $(R_k - \overline{R})^2$ measures how far R_k is from $\overline{R}$ is the same way that $(r_k - \overline{r})^2$ measures how far r_k is from $\overline{r}$. In fact, the model says that we should imagine that there is some sample point s in the sample space S so that $r_k = R_k(s)$ for each $k = 1, 2, \ldots, N$.

We now do a little algebraic rearranging (that is what algebra was invented for!):

$$\begin{aligned} R_k - \overline{R} &= R_k - \frac{R_1 + \cdots + R_N}{N} \\ &= \left(1 - \frac{1}{N}\right) R_k + \sum_{j=1,\ j\neq k}^{N} \frac{-1}{N} R_j \end{aligned}$$

and this last expression is a sum of independent random variables. What is more, $\mathrm{E[R_k - \overline{R}]} = \mu - \mu = 0$, so

$$\begin{aligned}
\mathrm{E[(R_k - \overline{R})^2]} &= \mathrm{E[(R_k - \overline{R})^2]} - \left(\mathrm{E[R_k - \overline{R}]}\right)^2 \\
&= \mathrm{Var[R_k - \overline{R}]} \\
&= \mathrm{Var}\left[\left(1 - \frac{1}{\mathrm{N}}\right)\mathrm{R_k} + \sum_{\mathrm{j=1,\ j \neq k}}^{\mathrm{N}} \frac{-1}{\mathrm{N}}\mathrm{R_j}\right] \\
&= \left(1 - \frac{1}{N}\right)^2 \mathrm{Var[R_k]} + \sum_{\mathrm{j=1,\ j \neq k}}^{\mathrm{N}} \left(\frac{-1}{\mathrm{N}}\right)^2 \mathrm{Var[R_j]} \\
&= \left(1 - \frac{1}{N}\right)^2 \sigma^2 + (N-1)\left(\frac{-1}{N}\right)^2 \sigma^2 \\
&= \frac{N-1}{N}\sigma^2.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{E[(s(R_1, \ldots, R_N))^2]} &= \mathrm{E}\left[\frac{1}{\mathrm{N} - 1}\sum_{\mathrm{k=1}}^{\mathrm{N}} \left(\mathrm{R_k - \overline{R}}\right)^2\right] \\
&= \frac{1}{N-1}\sum_{k=1}^{N} \mathrm{E}\left[\left(\mathrm{R_k - \overline{R}}\right)^2\right] \\
&= \frac{1}{N-1} N \frac{N-1}{N}\sigma^2 \\
&= \sigma^2,
\end{aligned}$$

showing that $(s(R_1, \ldots, R_N))^2$ is an unbiased estimate of the population variance σ^2. Why is $s(R_1, \ldots, R_N)$ **NOT** an unbiased estimate of σ?

Of course, $(s(R_1, \ldots, R_N))^2$ is not the only unbiased estimate of σ^2. It is clear that to find any estimator of σ^2 we will have to take products of the random variables. Let A be an $N \times N$ matrix of real numbers, and define the estimator $a(R_1, \ldots R_N)$ by

$$a(R_1, \ldots, R_N) := \sum_{j=1}^{N}\sum_{k=1}^{N} A_{j,k} R_j R_k.$$

Since R_j and R_k are independent when $j \neq k$ we have $\mathrm{E[R_j R_k]} = \mu^2$ when $j \neq k$. On the other hand, $\sigma^2 = \mathrm{Var[R_k]} = \mathrm{E[R_k^2]} - \mu^2$, so $\mathrm{E[R_k^2]} = \sigma^2 + \mu^2$. Therefore

$$\begin{aligned}
\mathrm{E[a(R_1, \ldots, R_N)]} &= \sum_{j=1}^{N}\sum_{k=1}^{N} A_{j,k}\mathrm{E[R_j R_k]} \\
&= \sum_{k=1}^{N} A_{k,k}(\sigma^2 + \mu^2) + \sum_{j=1}^{N}\sum_{k=1,\ k \neq j}^{N} A_{j,k}\mu^2 \\
&= \sigma^2\left(\sum_{k=1}^{N} A_{k,k}\right) + \mu^2\left(\sum_{j=1}^{N}\sum_{k=1}^{N} A_{j,k}\right)
\end{aligned}$$

so $a(R_1, \ldots, R_N)$ is an unbiased estimate of σ^2 so long as

$$\begin{aligned} \sum_{k=1}^{N} A_{k,k} &= 1 \\ \sum_{j=1}^{N} \sum_{k=1}^{N} A_{j,k} &= 0. \end{aligned}$$

What matrix A gives us $(s(R_1, \ldots, R_N))^2$?