Queen Mary
University of London

MTH4106 Introduction to Statistics

Notes 12 Spring 2012

Point estimation

Let $X_1, X_2, \ldots, X_n$ be random variables. A *statistic* is any real function $g(X_1, X_2, \ldots, X_n)$: for example, the sample mean, the sample variance, or the maximum.

Given any realization $x_1, x_2, \ldots, x_n$ of $X_1, X_2, \ldots, X_n$, then $g(x_1, x_2, \ldots, x_n)$ is a real number. However, $g(X_1, X_2, \ldots, X_n)$ itself is a random variable, so it has an expectation, a variance, etc., as we have already seen for the sample total, sample mean and sample variance.

Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables from the same population (i.e. with the same distribution), and that c is an unknown constant (also called a parameter) associated with that population or distribution, and that we want to find out the value of c.

Example 1: The population consists of all UK households, and c is the proportion with cats. (This is a definite number, which in principle could be measured exactly.)

Example 2: The population consists of all US adults, and c is the proportion who own an iPad.

Example 3: The population consists of all UK adults, and c is the proportion who gamble at least once a week.

Example 4: The population consists of all UK children aged 11–16, and c is the number who have been injured while trying to stop arguments between adults at their home.

Example 5: The population consists of European males aged 20–45, and c is the average decrease in systolic blood pressure (compared to the previous level) when playing football twice per week. (Since we can't force people to play football, this number is theoretical, and we have to estimate it.)

Example 6: The population consists of people aged 18–30, and c is the average increase in blood flow after watching a funny film.

Example 7: The population consists of all diabetic people, and c is the average decrease in blood-sugar levels if they change from diet 1 to diet 2.

Example 8: The population consists of all London buses, and c is the average decrease in emissions of nitrous oxide when a new filter is fitted to the engine.

We are going to use a statistic $g(X_1,X_2,\ldots,X_n)$ to estimate the value of c. Then $g(X_1,X_2,\ldots,X_n)$ is called an *estimator* for c: given any data x_1, x_2, $\ldots$, x_n, the real number $g(x_1,x_2,\ldots,x_n)$ is called the *estimate*, which is sometimes written as $\hat{c}$.

Bias and mean squared error

We now generalize some definitions that we made before in the context of estimating proportions.

Definition Suppose that $g(X_1,X_2,\ldots,X_n)$ is an estimator for c. When we estimate c by using the estimator $g(X_1,X_2,\ldots,X_n)$, the *error* in estimation is $g(X_1,X_2,\ldots,X_n)-c$. The *bias* of $g(X_1,X_2,\ldots,X_n)$ is

$$\mathbb{E}(g(X_1,X_2,\ldots,X_n)-c)=\mathbb{E}(g(X_1,X_2,\ldots,X_n))-c.$$

We say that $g(X_1,X_2,\ldots,X_n)$ is an *unbiased estimator* if its bias is zero; that is,

$$\mathbb{E}(g(X_1,X_2,\ldots,X_n))=c.$$

The *mean squared error* of $g(X_1,X_2,\ldots,X_n)$ is

$$\mathbb{E}\left([g(X_1,X_2,\ldots,X_n)-c]^2\right).$$

We prefer estimators which are unbiased and have small MSE. Just as before,

$$\mathrm{MSE}(g(X_1,X_2,\ldots,X_n))=\mathrm{Var}(g(X_1,X_2,\ldots,X_n))+(\mathrm{Bias}(g(X_1,X_2,\ldots,X_n)))^2.$$

Theorem 19 If X_1, X_2,$\ldots$, X_n are independent random variables from the same distribution with mean μ and variance σ^2, then

(a) the sample mean $\bar{X}$ is an unbiased estimator for μ;

(b) the sample variance S^2 is an unbiased estimator for σ^2.

Proof Use Theorem 18. ■

Definition The standard deviation of an estimator is called its *standard error*, abbreviated to S.E. or s.e.

Estimating a population mean

If the population has mean μ and variance σ^2, then we know that $\bar{X}$ is an unbiased estimator for μ and

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

So

$$\text{“}\texttt{S.E.Mean}\text{”} = \text{S.E.}(\bar{X}) = \text{standard error} = \frac{\sigma}{\sqrt{n}.}$$

Hence

$$\begin{aligned} \text{larger sample} &\equiv \text{larger } n \\ &\rightarrow \text{smaller standard error} \\ &\rightarrow \text{better estimate of } \mu. \end{aligned}$$

The Central Limit Theorem tells us that if n is large then the distribution of $\bar{X}$ is approximately normal, so $\begin{cases} \bar{X} \text{ is within 1 S.E. of } \mu \text{ approximately } 68\% \text{ of the time} \\ \bar{X} \text{ is within 2 S.E. of } \mu \text{ approximately } 95\% \text{ of the time} \\ \bar{X} \text{ is within 3 S.E. of } \mu \text{ approximately } 99\tfrac{3}{4}\% \text{ of the time} \end{cases}$.
This means that if we use $\bar{X}$ to estimate μ then we shall be in error by less than $2\sigma/\sqrt{n}$ in 19 cases out of 20.

When researchers estimate a population mean, they usually report the estimated mean $\hat{\mu}$ (which is usually $\bar{x}$) and its standard error. In general, we do not know σ either, so we have to estimate it too.

$S^2 =$ sample variance.

S^2 is an unbiased estimator for σ^2.

S is not an unbiased estimator for σ (because $\mathbb{E}(\sqrt{X}) \neq \sqrt{\mathbb{E}(X)}$ in general), but it is the most reasonable estimator available, giving the sample standard deviation s as the estimate of σ, so the estimated standard error is

$$\frac{s}{\sqrt{n}}.$$

Because of the normal approximation, for large n,

$$\begin{aligned} \frac{s}{\sqrt{n}} &\quad \text{is called the approximate 68\% error bound} \\ \frac{2s}{\sqrt{n}} &\quad \text{is called the approximate 95\% error bound} \\ \frac{3s}{\sqrt{n}} &\quad \text{is called the approximate 99\% error bound} \end{aligned}$$

In some fields, for example the reporting of opinion polls, the error bound is called the *margin of error*.

Notation (for a sample of size n)

Name	True value	Estimator	Estimate
population mean	μ	$\bar{X}$	$\hat{\mu}$ or $\bar{x}$
population variance	σ^2	S^2	s^2
population standard deviation	σ	$+\sqrt{S^2}$	$\hat{\sigma}$ or s
standard error of $\bar{X}$	$\frac{\sigma}{\sqrt{n}}$	$+\sqrt{\frac{S^2}{n}}$	$\frac{s}{\sqrt{n}}$

Example Twenty-four pieces of cotton yarn were stressed until they broke. The breaking load in ounces was recorded.

15.0, 17.0, 13.8, 15.5, 15.7, 15.6, 17.6, 17.1, 14.8, 15.8, 18.2, 16.0, 14.9,
14.2, 15.0, 12.8, 13.0, 16.2, 16.4, 14.8, 15.9, 15.6, 15.0, 15.5

$$n = 24 \qquad \sum x = 371.4 \qquad \sum x^2 = 5785.98$$

So the estimated mean $= \bar{x} = 15.48$ (given to one more decimal place than the data).

The estimated standard error should be given to the same accuracy as $\bar{x}$, so the sample variance should be calculated to twice as many decimal places. Thus

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \\ &= \frac{1}{23}\left(5785.98 - \frac{(371.4)^2}{24}\right) \\ &= 1.6767, \end{aligned}$$

so $s = \sqrt{1.6767} = 1.29$. The estimated standard error is

$$\sqrt{\frac{s^2}{n}} = \sqrt{\frac{1.6767}{24}} = 0.26.$$

Thus an approximate 95% error bound is $2 \times 0.26 = 0.52$.

Estimating a total

Suppose that we use n data x_1, $x_2\dots$, x_n from some population to estimate its mean μ and variance σ^2. In future, we will take further random observations Y_1, Y_2,..., Y_m from the same population, and put $T = Y_1 + Y_2 + \cdots + Y_m$. How should we estimate $\mathbb{E}(T)$ and $\mathrm{Var}(T)$?

Theorem 18 shows that $\mathbb{E}(T) = m\mu$, so the obvious estimator for $\mathbb{E}(T)$ is $m\bar{X}$. Then $\mathbb{E}(m\bar{X} - \mathbb{E}(T)) = m\,\mathbb{E}(\bar{X}) - m\mu = m\mu - m\mu = 0$, and so this estimator is unbiased. Also,

$$\mathrm{Var}(m\bar{X}) = m^2\,\mathrm{Var}(\bar{X}) = m^2\frac{\sigma^2}{n},$$

so the estimated standard error of this estimator is

$$\sqrt{m^2\frac{s^2}{n}} = \frac{ms}{\sqrt{n}}.$$

Theorem 18 also shows that $\mathrm{Var}(T) = m\sigma^2$, so the obvious estimator of $\mathrm{Var}(T)$ is mS^2.

Estimating a proportion

Suppose that we have a large population with just two values. For example, we might ask people "Have you published a book?" The possible answers are "Yes" and "No". Code the two values as 1 and 0 respectively.

Let p be the proportion of individuals in the population with value 1; that is, the proportion who have published a book. If a several individuals are chosen randomly from the population, and X_i is the value for the i-th person sampled, then

$$\begin{aligned}\mathbb{P}(X_i = 1) &= p \\ \mathbb{P}(X_i = 0) &= 1 - p = q,\end{aligned}$$

where $q = 1 - p$, and so $X_i \sim \mathrm{Bernoulli}(p)$, with $\mathbb{E}(X_i) = p$ and $\mathrm{Var}(X_i) = pq$.

Suppose that we have a random sample of n individuals. Because p is the mean of this Bernoulli distribution, our previous results show that $\bar{X}$ is an unbiased estimator of p. If m individuals in the sample have value 1, then

$$\bar{x} = \frac{m}{n} = \text{sample proportion} = \hat{p}.$$

Because $\mathrm{Var}(X_i)$ is a function of p, we don't need to use S^2 to estimate the variance. We estimate q by $\hat{q} = 1 - \hat{p}$, and then the estimated standard error is $\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$.

These results all agree with our earlier work on estimating proportions.

Estimating a total count

If there are N individuals in the population altogether, then Np of them have value 1. We estimate this total by

$$N\hat{p} = \frac{Nm}{n},$$

so the estimator is $N\bar{X}$. Its variance is

$$\operatorname{Var}(N\bar{X}) = N^2 \operatorname{Var}(\bar{X}) = N^2 \frac{pq}{n},$$

so the estimated standard error is

$$N\sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

Example Let the population be all children in the U.K. aged 11–16. Then $N = 4,412,700$. How many have been hurt in arguments between adults at home?

Let p be the proportion who have been hurt. In a sample (was it random?) of 1075 children, suppose that m reported that they had been hurt in this way. Then

$$\hat{p} = \frac{m}{1075} \quad \text{and} \quad N\hat{p} = \frac{m}{1075} \times 4,412,700.$$

If $m = 75$ then

$$\hat{p} = \frac{75}{1075} = 0.0698 \quad \text{and} \quad N\hat{p} = \frac{75}{1075} \times 4,412,700 = 307,863.$$

If $m = 76$ then

$$\hat{p} = \frac{76}{1075} = 0.0707 \quad \text{and} \quad N\hat{p} = \frac{76}{1075} \times 4,412,700 = 311,968.$$

The figure reported by GfK Custom Research was 308,889, amazingly precise. How did they get it? Possibly m was 75, giving $\hat{p} = 0.0698$, which they rounded to $\hat{p} = 0.07$, giving

$$N\hat{p} = 4,412,700 \times \frac{7}{100} = 308,889.$$

Or possibly m was 76, giving $\hat{p} = 0.0707$, which they also rounded to $\hat{p} = 0.07$.

In fact, changing m by 1 (the smallest possible change) changes the estimated total $N\hat{p}$ by

$$\frac{4,412,700}{1075} \approx 4104.8,$$

so it is ridiculous to give the answer to any closer precision than 2000.

If $m = 76$ then
$$\hat{p} = \frac{76}{1075} \quad \text{so} \quad \hat{q} = \frac{999}{1075}$$
so the standard error in the estimation of p is estimated as
$$\sqrt{\frac{76}{1075} \times \frac{999}{1075} \times \frac{1}{1075}} \approx 0.0078.$$
Multiplying this (unrounded) by $4,412,700$ gives a standard error of $34,497$ for the estimate of the number of hurt children. So the approximate 95% error bound is $2 \times 34,497 = 68,994$.

Why did none of the newspapers report this uncertainty?

Estimating a Poisson mean

If $X_i \sim \text{Poisson}(\lambda)$ for all i in the population then $\mathbb{E}(X_i) = \lambda$ and $\text{Var}(X_i) = \lambda$. So we use $\bar{X}$ to estimate λ.

Because $\text{Var}(X_i) = \lambda$ also, we don't need to use S^2 to estimate the variance. In fact,
$$\text{Var}(\bar{X}) = \frac{1}{n}\text{Var}(X_i) = \frac{\lambda}{n},$$
so the estimated standard error is
$$\sqrt{\frac{\hat{\lambda}}{n}},$$
where $\hat{\lambda} = \bar{x}$.

Estimating a variance

We have already seen that S^2 is an unbiased estimator for σ^2. We use this estimator in general if σ^2 is not a function of μ.

However, $\text{Var}(S^2)$ depends on the distribution of X_i, so we do not cover the estimated standard error for variance in this module.

Assumption of normality

Let n be the sample size. When can $\bar{X}$ be assumed to have a normal distribution?

- if the population is approximately normal—always
  (for example, if the population is large and the variable concerned is made up of a sum of many independent parts, such as:

  height —many different genes contribute;

  measurement of the same quantity —many different sources of error;

  time for a person to do a job —many different causes of delay).

- if the population is Bernoulli(p)—if $np > 9q$ and $nq > 9p$.

- if the population is known to be (roughly) symmetric—if $n \geqslant 20$.

- in general, for most populations—if $n \geqslant 50$.

(But even this isn't big enough for very asymmetric distributions, like those used in finance, or if the random variables are not independent, which happens if many of the bankers are using the same software.)