

Sample TSP Program for Task 1

? This program will calculate population means and variances
? for task1 in ECON221

? Notice that the use of the ? (question mark) will allow for the introduction
? of user defined comments in any TSP program. Using the ?, as I will be doing,
? you could then write small comments to yourself on what it is that you are doing
? within the body of the program.

? One very important thing to notice is that the end of any command line terminates
? with a ; (semi-colon). This lets the program know to start reading the next line.
? So if you are getting error messages, I would suggest first checking to make sure that
? you are not missing a ; .

options crt;

? Will set the carriage return to 80 characters. Important since
the line printer is set to print at 132 characters per line. This will
give regular 8.5" by 11" printing.

freq n;

? Sets the frequency of the data. In this case none. Other
options include monthly (M), quarterly (Q), and annually (A)

smpl 1 741;

? Sets the sample size for the data

read (file=task1.txt) r x y ;

? Reads in the data from a file called task1.txt, in columns

? Just to make sure that the data was read in correctly it is always a good idea to take a look at it
? using the print command

print r x y;

genr e=c;

? The genr (generate) command is used for performing
mathematical operations on n-tuples. Notice from the data that
we were not given a column of ones. This proves not to be a
problem since TSP reserves the letter c (for constant) as a n-tuple
of ones with n being equivalent to the sample size. In this case
741 ones.

? First to calculate all of these using the e n-tuple as shown in class

mat sumr=e'r;

? the mat command is used to get the inner products of the n-
tuples e and r. Here we get the sum of the column r.

mat sumee= e'e;

? Note that a single quote is used to convert columns to rows.
Notice that e'e=N.

mat sumx=e'x;

? This is the sum of the elements of x.
mat sumxx=x'e;
? This too is the sum of the elements of x
mat sumy=e'y;
? Similar to that of x, this is the sum of the elements of y.

? We could also find the sum of r using the shortcut
set N=741;
set sumr1=(N/2)*(N+1);
? the set command is used for performing mathematical
calculations on scalars.

? Since we now have all of these things calculated we would then want to see if
? All is as we think it should be by printing them.

```
print sumee sumx sumxx;  
print sumy sumr sumr1;
```

? Now to find the population mean for x and y
mat mux=(e'e)"e'x;
? The double quote is used for inversion (in this case a scalar).
Mux is then the population mean of x.
mat muy=(e'e)"e'y;
? Similarly for the population mean of y.

? then printing both population means.
print mux muy;

? Now to get the variances of x and y

```
mat sigmax=(1/N)*(x-e*mux)'(x-e*mux);  
? The population variance of x  
mat sigmay=(1/N)*(y-e*muy)'(y-e*muy);  
? Similarly, the population variance of y  
print sigmax sigmay;
```

? Now to find these using the summation method

? to sum the r
msd r;
? The msd command in TSP provides the user with means sums
and deviations for a particular series or n-tulpe, and stores
certain key calculations such as the mean, sample variance, and
sums into variables. Note that when you use the msd command
again, the previous values will be replace by the current.

```
set sumr2=@sum;  
? This will give us the sum of r and call it sumr2.  
print sumr2;
```

```
? To sum the x
msd x;
set sumx1=@sum;
print sumx1;
```

```
? To sum the y
msd (noprnt) y;
    ? The reason why I put noprnt in parenthesis is to suppress the
    output. In doing this for all of these variables we would get a
    tremendous amount of output that is really not necessary.
set sumy1=@sum;
print sumy1;
```

```
? For the mean of x and y
set mux1=(1/N)*sumx1;
set muy1=(1/N)*sumy1;
print mux1 muy1;
```

```
? Now for the variances
```

```
genr xsumsq=(x-mux1)**2 ;
    ? The ** is used to raise to the power. What is happening here is
    that xsumsq will be an n-tuple with each element equal to x
    minus the population mean of x all squared. Now all we have to
    do is get the sum of xsumsq and divide by N to get the
    population variance of x.
msd (noprnt) xsumsq;
set sumsqx=@sum;
    ? This is the sum of  $(x_i - \mu_x)^2$ 
set sigmax1=(1/N)*sumsqx;
    ? Then dividing by N yields the population variance given by
    sigmax1.
print sigmax1;
```

```
? Similar calculations for the population variance of y
genr ysumsq=(y-muy1)**2;
msd (noprnt) ysumsq;
set sumsqy=@sum;
set sigmay1=(1/N)*sumsqy;
print sigmay1;
```

```
end;
```

Introduction

Errors are inevitable in scientific life. Any attempt to measure a distance between two points or the weight or height of an object depends, to some extent, on the accuracy of the instruments involved, the care with which the instruments are operated and read and even the context in which the instruments are used. In economics, the measurement of a variable like GNP involves estimates of many complex economic transactions which are naturally subject to errors of various sorts. Any attempt to quantify, that is to affix a number to an object or to a sequence of events such as in the measurement of GNP, runs the risk of errors; it is an essential and unavoidable part of the process of quantification. By the same token, to say that quantification runs the risk of error is not to say that every measurement contains an error. Rather, the errors will sometimes be large and sometimes small and may even be so small as to be regarded as no error at all.

Once the possibility of errors is recognized as part of scientific reality, it is useful to have some formal recognition of the properties of the errors. This may be done by formulating a model of how the errors are generated. Such a model, in turn, will form an important part of the methodology to be developed below and generally in Econ 221 and Econ 222. To begin with, several intuitive models of errors will be presented, as a first step towards a more comprehensive theory.

Equally Likely Errors

The first basic intuition common to the models presented below is that errors can sometimes be positive, sometimes zero, or at least very close to zero, and sometimes negative. For the moment the issue of the relative size of errors will be ignored and it will be assumed that errors can be either -1 or zero or $+1$; that is, an error may be positive with numerical value 1, an overestimate, or zero, neither an overestimate nor an underestimate, or negative with numerical value 1, an underestimate. The next intuition concerns the chance that an error of a specific size and direction will take place in the model. To proceed with this simple example, it can be assumed that each specific error, $-1, 0, +1$, is equally likely and that one of these three possible errors must occur. In this case, each error occurs with a chance of occurrence of $\frac{1}{3}$. Table 1 summarizes the possibilities,

Table 1 : Three Equally Likely Errors

| <i>Error(ϵ)</i> | <i>Identification (i)</i> | <i>Chance of Occurrence(p)</i> |
|-------------------------------------|--|---|
| $\epsilon_1 = -1$ | $i = 1$ | $p_1 = \frac{1}{3}$ |
| $\epsilon_2 = 0$ | $i = 2$ | $p_2 = \frac{1}{3}$ |
| $\epsilon_3 = +1$ | $i = 3$ | $p_3 = \frac{1}{3}$ |

Within this framework, it is straightforward to construct other models of errors, simply by increasing the numerical range of equally likely errors. An example is given in Table 2.

Table 2 : Five Equally Likely Errors

| <i>Error(ϵ)</i> | <i>Identification (i)</i> | <i>Chance of Occurrence(p)</i> |
|-------------------------------------|--|---|
|-------------------------------------|--|---|

| | | |
|----------------------|---------|---------------------|
| $\varepsilon_1 = -2$ | $i = 1$ | $p_1 = \frac{1}{5}$ |
| $\varepsilon_2 = -1$ | $i = 2$ | $p_2 = \frac{1}{5}$ |
| $\varepsilon_3 = 0$ | $i = 3$ | $p_3 = \frac{1}{5}$ |
| $\varepsilon_4 = +1$ | $i = 4$ | $p_4 = \frac{1}{5}$ |
| $\varepsilon_5 = +2$ | $i = 5$ | $p_5 = \frac{1}{5}$ |

There are five potential errors in Table 2: two positive, two negative and one zero, each with an equal chance or possibility of occurring.

An important feature of Tables 1 and 2 is that the errors are equally spaced above and below zero so that, with equally likely errors, $\sum \varepsilon_i p_i = 0$ in both cases; in Table 1,

$$\sum \varepsilon_i p_i = (-1 \times \frac{1}{3}) + (0 \times \frac{1}{3}) + (1 \times \frac{1}{3}) = 0; \text{ in Table 2,}$$

$\sum_i \varepsilon_i p_i = (-2 \times \frac{1}{5}) + (-1 \times \frac{1}{5}) + (0 \times \frac{1}{5}) + (1 \times \frac{1}{5}) + (2 \times \frac{1}{5}) = 0$. Because the errors are equally spaced above and below zero with equal chances or possibilities of occurring, the negative errors just balanced the positive ones. While individual error may be negative, zero or positive, on average the errors will have zero effect. This is a common feature of scientific enquiry: although it is ultimately impossible to avoid making errors, provided the errors and their corresponding chances of occurring are symmetrical around zero, then on average the results will not be distorted by the errors made. In the experimental sciences, errors are called experimental errors and experiments must be repeated many times to ensure that on average the errors do not distort the pattern of results.

Different and Symmetrical Probabilities

The first model of errors may now be enriched in a new direction by reducing the chances or probabilities of numerically large errors relative to the chances or probability of numerically small ones. After all, it is more likely that a skillful golfer will land an approach shot close to the pin than not, since this is precisely where he is aiming. Such intuition may be said to motivate the next model in Table 3.

Table 3 : *Five Errors with Different but Symmetrical Probabilities*

| <i>Error(ε)</i> | <i>Identification (i)</i> | <i>Chance of Occurrence(p)</i> |
|--|--|---|
| $\varepsilon_1 = -2$ | $i = 1$ | $p_1 = \frac{1}{9}$ |
| $\varepsilon_2 = -1$ | $i = 2$ | $p_2 = \frac{2}{9}$ |
| $\varepsilon_3 = 0$ | $i = 3$ | $p_3 = \frac{3}{9}$ |
| $\varepsilon_4 = +1$ | $i = 4$ | $p_4 = \frac{2}{9}$ |
| $\varepsilon_5 = +2$ | $i = 5$ | $p_5 = \frac{1}{9}$ |

These are five distinct errors in Table 3. Again it will be noticed that $\sum_i \varepsilon_i p_i = 0$. Thus an important feature of Table 1 and 2 is preserved in Table 3 while also introducing different relative chances of occurrence. This is done by balancing the probability of occurrence for every separate positive value with the probability of the corresponding negative one. Thus $p_1 = p_5 = \frac{1}{9}$ while $\varepsilon_1 = -2$ and $\varepsilon_5 = +2$; $p_2 = p_4 = \frac{2}{9}$ while $\varepsilon_2 = -1$ and $\varepsilon_4 = +1$. Thus the ε_i 's and p_i 's form a

symmetrical distribution.

Table 4 : Nine Errors with Different but Symmetrical Probabilities

| <i>Error(ε)</i> | <i>Identification(i)</i> | <i>Chance of Occurrence(p)</i> | <i>Cumulative Probability*</i> |
|-------------------------|--------------------------|----------------------------------|--|
| $\varepsilon_1 = -1.00$ | 1 | $p_1 = \frac{1}{177} = 0.00565$ | $P(\varepsilon \leq -1.00) = 0.00565 = p(1)$ |
| $\varepsilon_2 = -0.75$ | 2 | $p_2 = \frac{7}{177} = 0.03955$ | $P(\varepsilon \leq -0.75) = 0.04520 = p(2)$ |
| $\varepsilon_3 = -0.50$ | 3 | $p_3 = \frac{21}{177} = 0.11864$ | $P(\varepsilon \leq -0.50) = 0.16384 = p(3)$ |
| $\varepsilon_4 = -0.25$ | 4 | $p_4 = \frac{35}{177} = 0.19774$ | $P(\varepsilon \leq -0.25) = 0.36158 = p(4)$ |
| $\varepsilon_5 = 0$ | 5 | $p_5 = \frac{49}{177} = 0.27684$ | $P(\varepsilon \leq 0) = 0.63842 = p(5)$ |
| $\varepsilon_6 = +0.25$ | 6 | $p_6 = \frac{35}{177} = 0.19774$ | $P(\varepsilon \leq 0.25) = 0.83616 = p(6)$ |
| $\varepsilon_7 = +0.50$ | 7 | $p_7 = \frac{21}{177} = 0.11864$ | $P(\varepsilon \leq 0.50) = 0.95480 = p(7)$ |
| $\varepsilon_8 = +0.75$ | 8 | $p_8 = \frac{7}{177} = 0.03955$ | $P(\varepsilon \leq 0.75) = 0.99435 = p(8)$ |
| $\varepsilon_9 = +1.00$ | 9 | $p_9 = \frac{1}{177} = 0.00565$ | $P(\varepsilon \leq 1.00) = 1.00000 = p(9)$ |

$$*p(j) = \sum_{i=1}^{i=j} p_i, \text{ e.g. } p(4) = p_1 + p_2 + p_3 + p_4.$$

The same principle is applied to the equally spaced errors in Table 4 but here there are nine different error values; numerically equal errors of different signs have equal probabilities of occurring, and hence the balancing of positive and negative errors is maintained.

According to Table 4, the errors can assume four evenly-spaced values less than zero (-1.0, -0.75, -0.50, -0.25), zero and four evenly-spaced values above zero (0.25, 0.50, 0.75, 1.00). The corresponding probabilities of occurrence of the negative values are matched exactly by the probabilities of occurrence of the positive values. The pattern of chances of occurrence is such that the probability that an error occurs decreases with the distance from zero, leaving a 27.7% chance that a zero error will occur. A diagram representing the first column of Table 4, ε , on the horizontal axis and the third column, p , on the vertical axis is given below in Figure 1. Figure 1 reveals the underlying symmetry of the probability distribution of errors, where probabilities are represented by the height of the vertical lines in the diagram.

The same information is presented in a different way in Figure 2. This diagram again has ε on the horizontal axis, but on the vertical axis the cumulative probability $p(j) = \sum_{i=1}^{i=j} p_i, j = 1, 2, \dots, 9$

corresponding to, say, $\varepsilon_4 = -0.25$ is the probability that an error of less than or equal to $\varepsilon_4 = -0.25$ will occur. This is precisely $p_1 + p_2 + p_3 + p_4 = p(4) = \sum_{i=1}^{i=4} p_i = 0.36158 = (64/177)$. At the point $\varepsilon_7 = 0.50$, the chance that an error of $\varepsilon_7 = 0.50$ or less will occur is

$p(7) = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 = \sum_{i=1}^{i=7} p_i = 0.95480 = (169/177)$. It is easy to recover, the

probability of a particular error occurring by subtraction, because $p_j = p(j) - p(j - 1), j = 2, 3, \dots, 9$.

For example, if there is interest in recovering p_5 from the list of cumulative probabilities in Table 4, then $p(5) - p(4) = (113/177) - (64/177) = (49/177) = p_5$.

The Normal Law of Error

The points marked \odot in Figure 2 represent the cumulative probabilities for specific points on the ε -axis. For example, at the point $\varepsilon_4 = -0.25$ the probability $p(4)$ is 0.36158 that one of the errors in the set $\{-1.00, -0.75, -0.5, -0.25\}$ will occur; this may be written $P(\varepsilon \leq -0.25) = p(4) = 0.36158$. Points in between -1.00 and -0.75 or -0.75 and -0.5 or -0.5 and -0.25 are defined as impossible according to the model specification of Table 4. From a practical viewpoint, this feature of the model is obviously unsatisfactory: why should an error of -0.75 be possible while an error of -0.60 is impossible? A more practical model of errors would surely not constrain the size of an error to a specific point, but would allow for the possibility that any error within a certain range of values can occur: for example within the range of greater than or equal to -0.75 and less than -0.5 the probability of occurrence might be 0.0689 or the probability of an error greater than or equal to -0.5 and less than $+0.25$ might be 0.6585, and so on. A model of errors which allows for the possibility of any numerical size within a specific numerical range is called continuous within that range, as opposed to discrete in which only specific errors of certain magnitudes are permitted within the range.

In Figure 2 the points marked \odot represent the discrete cumulative probabilities of Table 4 column 3 and these trace out a shape that is like an elongated S. This shape is characteristic of the cumulative probabilities arising from the famous bell-curve that students like to associate with the determination of final examination marks. The bell-curve is known technically as the normal law of error or just the normal distribution. The cumulative normal law of error corresponding to the cumulative law of error of Table 4 column 3 is traced out with a dotted line joining points marked +. Strictly speaking, the points marked \odot and + are not comparable, since the former apply to the cumulated height of lines above the points ε_i , $i = 1, 2, \dots, 9$, in Figure 1, while the dotted line and the points marked + refer to cumulated areas under the bell-curve. Moreover the probability of errors larger than $\varepsilon_1 = +1.0$ or smaller than $\varepsilon_1 = -1.0$ are impossible in Table 4 column 3, while under the bell-curve an error of less than -1.00 is possible, but with a very small probability (0.37%). In effect, the bell-curve allows for the possibility of any error within the finite range $\{-1.0, +1.0\}$ while permitting a small chance of error outside this range, while Table 4 column 3 allows for only 9 specific numerical errors within the range $\{-1.0, +1.0\}$ and makes it impossible to achieve errors outside this range. For this reason, the + points lie consistently below the \odot points, and the \odot points reach 1.00 (a certainty) while the + points only approach 1.00 asymptotically.

In Table 5 a comparison is made of the two sets of cumulative probabilities, showing that the range of possible error for the normal curve is much more comprehensive (and hence practicable) than the errors in column 3 of Table 4.

Table 5 : Cumulative Probabilities from Table 4 and Corresponding
Cumulative Probabilities for the Normal Law of Error

| <i>Discrete Errors</i> | <i>Error set</i> | <i>Cumulative Probability $p(\bullet)$</i> | <i>Range of Normal Errors</i> | <i>Cumulative Normal Probability</i> |
|-------------------------|--|---|-----------------------------------|--|
| <i>Less than</i> | <i>-1.00</i> | 0 | | |
| $\varepsilon_1 = -1.00$ | $\{\varepsilon_1\}$ | 0.00565 | $\varepsilon \leq -1.00$ | 0.0037 |
| $\varepsilon_2 = -0.75$ | $\{\varepsilon_1, \varepsilon_2\}$ | 0.04520 | $\varepsilon \leq -0.75$ | 0.0222 |
| $\varepsilon_3 = -0.50$ | $\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ | 0.16384 | $\varepsilon \leq -0.50$ | 0.0901 |
| $\varepsilon_4 = -0.25$ | $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\}$ | 0.36158 | $\varepsilon \leq -0.2$ | 0.2514 |
| $\varepsilon_5 = 0$ | $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_5\}$ | 0.63842 | $\varepsilon \leq 0$ | 0.5000 |
| $\varepsilon_6 = +0.25$ | $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_6\}$ | 0.83616 | $\varepsilon \leq 0.25$ | 0.7486 |
| $\varepsilon_7 = +0.50$ | $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_7\}$ | 0.95480 | $\varepsilon \leq 0.50$ | 0.9099 |
| $\varepsilon_8 = +0.75$ | $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_8\}$ | 0.99435 | $\varepsilon \leq 0.75$ | 0.9778 |
| $\varepsilon_9 = +1.00$ | $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_9\}$ | 1.00000 | $\varepsilon \leq 1.00$ | 0.9963 |
| <i>More than</i> | <i>1.00</i> | 0 | <i>More than 1</i> | 0.0037 |

Errors in an Econometric Model

The first econometric model that you have been introduced to takes the form

$$y_i = \mu_y + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad \#$$

where n represents the size of the sample that is available. Despite its apparent simplicity, this model contains the basic elements of more complicated models. The general thrust of interest in the model is to gain some insight into the behaviour of an economic variable y . The model suggests that this variable can be understood in terms of two components: a fixed component μ_y and a random component given by the sequence $\varepsilon_i, i = 1, 2, \dots, n$. Suppose, as is usually assumed, that each ε_i is drawn from a population with zero mean. Then the model claims that each y_i itself has a mean of μ_y . It should be emphasized that the model is not suggesting that y_i is μ_y throughout the whole sample or even through most of the sample. Depending on the behaviour of the ε_i , y_i may not even be particularly close to μ_y . The important point to note is that any variability in the y_i cannot come from μ_y , since this is a constant, and so must come from the variability in the sample of ε_i 's.

It is appropriate now to introduce a third and more fundamental part of the specification of the model. This is that each ε_i is a random selection from the normal law of error, having mean zero and variance σ^2 . Notice carefully that in the model only y_1, y_2, \dots, y_n are observed; μ_y is not known and $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are not observed; moreover, the variance of the ε_i , σ^2 , is not known. Assuming the model to be correct, it is clear that the variance of the y_i , σ_y^2 , must be equal to σ^2 . In effect there are n observations and two unknown parameters μ_y and $\sigma_y^2 = \sigma^2$. The problem is how to estimate μ_y and σ^2 . Estimation is facilitated by writing (1) in n -tuples as:

$$\mathbf{y} = \mathbf{e}\mu_y + \boldsymbol{\varepsilon} \quad \#$$

where \mathbf{e} is the unitary n -tuple, $\boldsymbol{\varepsilon}$ is the n -tuple with elements $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, and \mathbf{y} is the n -tuple of observations y_1, y_2, \dots, y_n .

Let $\hat{\mu}_y$ be an estimate of μ_y and $\hat{\varepsilon}_i$ an estimate of ε_i so that $\hat{\boldsymbol{\varepsilon}}$ is an estimate of $\boldsymbol{\varepsilon}$. One way to estimate $\boldsymbol{\varepsilon}$ is to make use of the assumption that ε_i 's are drawn randomly from a normal distribution with mean zero and variance σ^2 . If the mean of each ε_i is zero then $\hat{\mu}_y$ and $\hat{\boldsymbol{\varepsilon}}$ may be selected by requiring the mean of the $\hat{\varepsilon}_i$'s to be zero. That is,

$$\begin{aligned} (\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'(\mathbf{y} - \mathbf{e}\hat{\mu}_y) &= (\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\hat{\boldsymbol{\varepsilon}} = 0 \\ \Rightarrow (\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{y} &= \hat{\mu}_y. \end{aligned}$$

Of course $(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{y} = \{\sum y_i/n\} = \bar{y}$, the sample mean.

Given that $\hat{\mu}_y = \bar{y}$, $\hat{\boldsymbol{\varepsilon}} = (\mathbf{y} - \mathbf{e}\bar{y}) = \mathbf{y} - \mathbf{e}(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{y} = (\mathbf{I}_n - \mathbf{p})\mathbf{y}$ where $\mathbf{p} = \mathbf{e}(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'$ which is the n -square matrix having every element equal to $\frac{1}{n}$. It follows that

$$\begin{aligned} \frac{1}{n-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} &= \frac{1}{n-1}(\mathbf{y} - \mathbf{e}\hat{\mu}_y)'(\mathbf{y} - \mathbf{e}\hat{\mu}_y) \\ &= \frac{1}{n-1}\mathbf{y}'(\mathbf{I}_n - \mathbf{p})\mathbf{y} \\ &= \frac{1}{n-1}\sum (y_i - \bar{y})^2 = s_y^2. \end{aligned}$$

Thus if $\hat{\sigma}^2 = \hat{\sigma}_y^2$ are the estimates of σ^2 and σ_y^2 , then $\hat{\sigma}^2 = \hat{\sigma}_y^2 = s_y^2$.

Introduction

The sampling distribution theory which will be described here is based on the normal distribution and adequately covers all of the sampling theory that is necessary for Econ 221, and most of what is required for Econ 222. It is assumed that there is available a random sample of n from the population distribution $Y \sim N(\mu_y, \sigma_y^2)$. Since the n observations have been randomly selected from Y , they must be independent of one another. The sample of n is denoted y_1, y_2, \dots, y_n and, when these observations are arranged in a column, they form an n -tuple, and this n -tuple is denoted by \mathbf{y} . The unitary n -tuple is denoted by \mathbf{e} (a column of ones). The sample mean is $\bar{y} = (\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{y}$ and the sample variance is $s_y^2 = \{\mathbf{y}'(\mathbf{I}_n - \mathbf{p})\mathbf{y}/(n-1)\}$; alternatively, $\bar{y} = \{\sum_i y_i/n\}$ and $s_y^2 = \{\sum_i (y_i - \bar{y})^2/(n-1)\} = [\{\sum_i y_i^2 - (\sum_i y_i)^2/n\}/(n-1)]$.

Sampling Distribution

Given the population $Y \sim N(\mu_y, \sigma_y^2)$ it is always possible to draw more than one sample from it. Suppose that k independent random samples of size n are selected from Y . For each sample, \bar{y} and s_y^2 can be calculated. There would then be k sample means $\bar{y}^{(1)}, \bar{y}^{(2)}, \dots, \bar{y}^{(k)}$ which, for sufficiently large k , would form a distribution of their own. The mean and variance of this distribution are straightforward to find. For any $\bar{y} = \frac{1}{n} \sum y_i$,

$$\begin{aligned} E[\bar{y}] &= \frac{1}{n}[E(y_1) + E(y_2) + \dots + E(y_n)] \\ &= \frac{1}{n}(\mu_y + \mu_y + \dots + \mu_y) \quad \# \\ &= \mu_y \end{aligned}$$

Since this is true for any \bar{y} , it must be true for the whole distribution $\bar{y}^{(1)}, \bar{y}^{(2)}, \dots, \bar{y}^{(k)}$ which is denoted \bar{Y} . The variance of any \bar{y} is the variance of $(\frac{1}{n}y_1 + \frac{1}{n}y_2 + \dots + \frac{1}{n}y_n)$. Since the y 's are independently selected, the covariance for any pair is zero and hence

$$\begin{aligned} V(\frac{1}{n}y_1 + \frac{1}{n}y_2 + \dots + \frac{1}{n}y_n) &= \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \dots + \frac{1}{n^2}\sigma^2 \\ &= n\sigma^2/n^2 \\ &= \sigma^2/n. \quad \# \end{aligned}$$

Also \bar{y} , being a linear combination of $\{y_1, y_2, \dots, y_n\}$, must inherit normality. Hence the distribution of \bar{Y} will be normal. Since the mean and variance are given by (1) and (2), it follows that $\bar{Y} \sim N(\mu_y, \frac{\sigma^2}{n})$.

Decomposition

For any random sample of n from $Y \sim N(\mu_y, \sigma_y^2)$, the following decomposition holds:

$$\begin{aligned} \sum_i (y_i - \mu_y)^2 &= \sum_i (y_i - \bar{y} + \bar{y} - \mu_y)^2 \\ &= \sum_i (y_i - \bar{y})^2 + n(\bar{y} - \mu_y)^2 + 2(\bar{y} - \mu_y) \sum_i (y_i - \bar{y}) \end{aligned}$$

Since $\sum_i (y_i - \bar{y}) = 0$, $2(\bar{y} - \mu_y) \sum_i (y_i - \bar{y}) = 0$ and hence

$$\sum_i (y_i - \mu_y)^2 = \sum_i (y_i - \bar{y})^2 + n(\bar{y} - \mu_y)^2. \quad \#$$

Taking expectations, $E[\sum_i (y_i - \mu_y)^2] = \sum_i E (y_i - \mu_y)^2 = n\sigma_y^2$ because for every y_i ,

$E(y_i - \mu_y)^2 = \sigma^2$ and the y_i 's are independent. On the right-hand side of (3), $E[n(\bar{y} - \mu_y)^2] = nE[(\bar{y} - \mu_y)^2] = n\sigma_{\bar{y}}^2$ which from (2) becomes $n(\sigma_y^2/n) = \sigma_y^2$. Hence $E[n(\bar{y} - \mu_y)^2] = \sigma_y^2$, and from (3)

$$n\sigma_y^2 = E \sum_i (y_i - \bar{y})^2 + \sigma_y^2$$

or,

$$(n-1)\sigma_y^2 = E \sum_i (y_i - \bar{y})^2. \quad \#$$

$\sum_i (y_i - \bar{y})^2$ is said to have $(n-1)$ degrees of freedom because the elements in $E \sum_i (y_i - \bar{y})^2$ are subject to one restriction, namely that $\bar{y} = \frac{1}{n} \sum_i y_i$. Moreover, from (4)

$$E\left[\frac{1}{n-1} \sum_i (y_i - \bar{y})^2\right] = \sigma_y^2 \Rightarrow E[s_y^2] = \sigma_y^2. \quad \#$$

Thus s_y^2 is an unbiased estimator of σ_y^2 , since the expectation of the estimator is equal to the true value of the population parameter being estimated.

The Chi-square Distribution

Since $Y \sim N(\mu_y, \sigma_y^2)$, $\{(Y - \mu_y)/\sigma_y\} = Z \sim N(0, 1)$. For any y_i chosen randomly from Y , $\{(y_i - \mu_y)/\sigma_y\} = z_i \sim N(0, 1)$. If z_i is squared, $z_i^2 \sim \chi_{(1)}^2$. If the whole random sample is considered then $\sum_i \{(y_i - \mu_y)/\sigma_y\}^2 = \sum_i z_i^2 \sim \chi_{(n)}^2$: that is, the sum of squares of n independent $N(0, 1)$ variates has the $\chi_{(n)}^2$ distribution. This distribution is called the *central chi-squared distribution with n degrees of freedom* (see pp. 276–282 and pp. 895 of the textbook). The mean of a central $\chi_{(n)}^2$ -distribution is n and its variance is $2n$.

The F - and t -distributions

If $\chi_{(v_1)}^2$ and $\chi_{(v_2)}^2$ are two independent χ^2 -variates with v_1 and v_2 degrees of freedom, then

$$\frac{\chi_{(v_1)}^2/v_1}{\chi_{(v_2)}^2/v_2} \sim F(v_1, v_2). \quad \#$$

This is the *central F -distribution with v_1 and v_2 degrees of freedom* (see pp. 421–427 and pp. 896–899 of the textbook).

In the special case when $v_1 = 1$,

$$\sqrt{\frac{\chi_{(v_1)}^2}{\chi_{(v_2)}^2/v_2}} \sim t_{(v_2)} \quad \#$$

which is the *central t -distribution with v_2 degrees of freedom*. The mean of the central t -distribution is zero; its variance when $n \geq 3$ is $v_2/(v_2 - 2)$.

The Decomposition and the t -test

Returning to the decomposition in (3) above, when both sides are divided by σ_y^2 , the following equation is obtained:

$$\sum \left\{ \frac{y_i - \mu_y}{\sigma_y} \right\}^2 = \sum \left\{ \frac{y_i - \bar{y}}{\sigma_y} \right\}^2 + \left\{ \frac{\bar{y} - \mu_y}{\sigma_y / \sqrt{n}} \right\}^2. \quad \#$$

On the left-hand side, from section 4 above,

$$\sum \left\{ \frac{y_i - \mu_y}{\sigma_y} \right\}^2 = \sum z_i^2 \sim \chi_{(n)}^2.$$

From section 3, it is known that the covariance of $(\bar{y} - \mu_y)$ and $\sum (y_i - \bar{y})$ is zero because $(\bar{y} - \mu_y) \{ \sum (y_i - \bar{y}) \}$ is always zero since $(\bar{y} - \mu_y)$ is constant and $\sum (y_i - \bar{y}) = 0$. Since $(\bar{y} - \mu_y)$ and $\sum (y_i - \bar{y})$ are each normally distributed, a zero covariance (in the case of normal distributions) implies independence and hence the two different sums of square on the right-hand side of (8) are sums of squares of independent normal variates. In particular $\bar{y} \sim N(\mu_y, \sigma_y^2/n)$ and hence

$$\frac{n(\bar{y} - \mu_y)^2}{\sigma_y^2} = \left\{ \frac{\bar{y} - \mu_y}{\sigma_y / \sqrt{n}} \right\}^2 \sim \chi_{(1)}^2. \quad \#$$

Clearly there is only one \bar{y} in the sample and hence the distributions is central chi-square with one degree of freedom. Finally, from section 3, it is known that $\sum (y_i - \bar{y})^2$ has $(n - 1)$ degrees of freedom. Thus the degrees of freedom on the right-hand side of (8), $(n - 1)$ and 1, add up to the degrees of freedom on the left-hand side (n). In these circumstances it is not surprising that, given (9) applies,

$$\sum \left\{ \frac{y_i - \bar{y}}{\sigma_y} \right\}^2 \sim \chi_{(n-1)}^2$$

and the two components on the right-hand side of (8) are independent χ^2 -variates.

Given the independent χ^2 -variates on the right-hand side of (8), from section 5, equation (6):

$$\frac{(\frac{\bar{y} - \mu_y}{\sigma_y / \sqrt{n}}) / 1}{\sum \{ \frac{y_i - \bar{y}}{\sigma_y} \}^2 / (n - 1)} \sim F(1, n - 1). \quad \#$$

Now in (10) the two σ_y^2 's cancel out leaving in the denominator:

$$\frac{1}{n - 1} \sum (y_i - \bar{y})^2 = s_y^2.$$

Hence from (7), by taking the square root of (10),

$$\frac{\bar{y} - \mu_y}{s_y / \sqrt{n}} \sim t_{(n-1)}. \quad \#$$

Equation (11) is the basis of the t -test and rests on the presumption that the original sample comes from $Y \sim N(\mu_y, \sigma_y^2)$.

Summary

In terms of n -tuples, $\bar{y} = (\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{y}$ and $s_y^2 = \{ \mathbf{y}'(\mathbf{I}_n - \mathbf{p})\mathbf{y} / (n - 1) \}$. Notice that,

$$(\mathbf{y} - \mathbf{e}\mu_y) = (\mathbf{I}_n - \mathbf{p})(\mathbf{y} - \mathbf{e}\mu_y) + \mathbf{p}(\mathbf{y} - \mathbf{e}\mu_y), \quad \#$$

$$= (\mathbf{I}_n - \mathbf{p})\mathbf{y} + \mathbf{p}(\mathbf{y} - \mathbf{e}\mu_y), \quad \#$$

because $(\mathbf{I}_n - \mathbf{p})\mathbf{e}\mu_y = \mathbf{e}\mu_y - \mathbf{e}\mu_y = 0$. Also, a property of (13) is

$$\mathbf{y}'(\mathbf{I}_n - \mathbf{p})\mathbf{p}(\mathbf{y} - \mathbf{e}\mu_y) = 0, \quad \#$$

because $(\mathbf{I}_n - \mathbf{p})\mathbf{p} = \mathbf{p} - \mathbf{p}^2 = \mathbf{p} - \mathbf{p} = 0$. Consequently,

$$(\mathbf{y} - \mathbf{e}\mu_y)'(\mathbf{y} - \mathbf{e}\mu_y) = (\mathbf{y} - \mathbf{e}\mu_y)'(\mathbf{I}_n - \mathbf{p})(\mathbf{y} - \mathbf{e}\mu_y) + (\mathbf{y} - \mathbf{e}\mu_y)'\mathbf{p}(\mathbf{y} - \mathbf{e}\mu_y), \quad \#$$

and from equation (12), (13) and (14), equation (15) reduces to

$$(\mathbf{y} - \mathbf{e}\mu_y)'(\mathbf{y} - \mathbf{e}\mu_y) = \mathbf{y}'(\mathbf{I}_n - \mathbf{p})\mathbf{y} + (\mathbf{y} - \mathbf{e}\mu_y)'\mathbf{p}(\mathbf{y} - \mathbf{e}\mu_y), \quad \#$$

Equation (16) is precisely equation (3), that is

$$\sum (y_i - \mu_y)^2 = \sum (y_i - \bar{y})^2 + n(\bar{y} - \mu_y)^2. \quad \#$$

The first component on the right-hand side of (17) is independent of the second, by (14).

There are essentially four results based on (i) the normal distribution and (ii) the decomposition (16) and (17). These results are essential for understanding your work.

Result 1: If $Y \sim N(\mu_y, \sigma_y^2)$ and y_1, y_2, \dots, y_n is a random sample of n from Y , then every $\{(y_i - \mu_y)/\sigma_y\} = z_i \sim N(0, 1)$. The random sample y_1, y_2, \dots, y_n is written as a column and denoted \mathbf{y} .

Result 2: For the random sample \mathbf{y} ,

$$\frac{1}{\sigma_y^2}(\mathbf{y} - \mathbf{e}\mu_y)'(\mathbf{y} - \mathbf{e}\mu_y) = \sum \left\{ \frac{\bar{y} - \mu_y}{\sigma_y} \right\}^2 = \sum z_i^2 \sim \chi_{(n)}^2, \quad \#$$

the central chi-square distribution with n degree of freedom. This distribution has mean n and variance $2n$.

Result 3: If $\chi_{(v_1)}^2$ and $\chi_{(v_2)}^2$ are two independent central χ^2 -variates with v_1 and v_2 degrees of freedom, then

$$\frac{\chi_{(v_1)}^2/v_1}{\chi_{(v_2)}^2/v_2} \sim F(v_1, v_2), \quad \#$$

the central F -distribution with v_1 and v_2 degrees of freedom. In equation (17), $[\{\sum (y_i - \bar{y})^2\}/\sigma_y^2] \sim \chi_{(n-1)}^2$ independently of $\{n(\bar{y} - \mu_y)^2/\sigma_y^2\} \sim \chi_{(1)}^2$. Hence,

$$\frac{n(\bar{y} - \mu_y)^2/\sigma_y^2 \cdot 1}{\sum (y_i - \bar{y})^2/\sigma_y^2(n-1)} \sim F(1, n-1). \quad \#$$

The expression on the left-hand side of (20) reduces to

$$\frac{(\bar{y} - \mu_y)^2}{s_y^2/n} \sim F(1, n-1), \quad \#$$

where $s_y^2 = [\{\sum (y_i - \bar{y})^2\}/(n-1)]$.

Result 4. From Result 3,

$$\frac{\bar{y} - \mu_y}{s_y / \sqrt{n}} \sim \sqrt{F(1, n-1)} \equiv t(n-1), \quad \#$$

the t -distribution with $(n-1)$ degrees of freedom. The mean of $t(n-1)$ is zero and, given $n \geq 3$, its variance is $(n-1)/(n-3)$.

Introduction

On pages 817 – 820 of your textbook there are five different lists of numbers referring to each of 179 cities in Canada which, in 1991, had a city population of at least 10,000 people. From these numbers it is possible to find three cities, Montreal, Toronto and Vancouver, which had a 1991 city population in excess of one million. It is then possible to calculate, for each of these three cities, average personal income per head for 1991 and average household income for 1991, both in thousands of Canadian dollars. These statistics are displayed in Table 1.

Table 1 : Cities in Canada with a city population in excess of 1 million in 1991 and corresponding average personal income per head by city and average income per household by city, in \$000's Canadian

| City | Identification number (i) | Population (x), 000's | Income per head (y) \$000's | Income per household (z) \$000's |
|-----------|-------------------------------|---------------------------|---------------------------------|--------------------------------------|
| Montreal | $i = 1$ | $3,084.1 = x_1$ | $18.73 = y_1$ | $47.3 = z_1$ |
| Toronto | $i = 2$ | $3,822.4 = x_2$ | $22.85 = y_2$ | $65.0 = z_2$ |
| Vancouver | $i = 3$ | $1,586.6 = x_3$ | $19.20 = y_3$ | $49.9 = z_3$ |

Table 1 indicates that Montreal is identified by $i = 1$, Toronto by $i = 2$, and Vancouver by $i = 3$. This numbering permits the values taken on by the variables x, y and z to be identified with a particular city. Thus the population of Montreal is $x_1 = 3,084.1$, for Toronto is $x_2 = 3,822.4$ and for Vancouver is $x_3 = 1,586.6$, all measured in thousands of people. Similarly annual income per head in 1991 was $y_1 = \$18,730$ for Montreal, $y_2 = \$22,850$ for Toronto and $y_3 = \$19,200$ for Vancouver. Corresponding figures for annual household income in 1991 are $z_1 = \$47,300$, $z_2 = \$65,000$, and $z_3 = \$49,900$. Since the three cities constitute all of the cities in Canada which had in 1991 a population in excess of one million people, the sets of numbers $\{x_1, x_2, x_3\}$, $\{y_1, y_2, y_3\}$ and $\{z_1, z_2, z_3\}$ may each be regarded as a statistical population: in terms of x , $\{x_1, x_2, x_3\}$ constitutes the population of Canadian city populations in excess of one million in 1991; for y , the set $\{y_1, y_2, y_3\}$ is the 1991 population of incomes per head in Canadian cities comprising 1 million or more people; and the set $\{z_1, z_2, z_3\}$ is the 1991 population of incomes per household in Canadian cities comprising 1 million or more people.

Had consideration been given to all Canadian cities which in 1991 had a city population of 10,000 people or more, then the identification numbers i would have run from $i = 1$ to $i = 179$, instead of from $i = 1$ to $i = 3$. Thus the set of x -values would have been $\{x_1, x_2, \dots, x_{179}\}$ with corresponding sets of the same size for y and z . Generally, the size of a finite population is denoted by N and hence, for any finite N , $i = 1, 2, \dots, N$. Usually N is quite large, in the hundreds, thousands or even millions, and it should be recognized that a population of $N = 3$ would not normally be encountered in a practical situation. Nevertheless $N = 3$ will suffice for the present purpose which is to illustrate the algebra of calculating population parameters using columns of N observations. This being the case, $N = 3$ has the benefit of being small and hence simple to operate, and its smallness does not detract from the general principles of the argument to be developed.

Operations with N-tuples

In terms of notation it is usual to record N population values on a variable x as $x_i, i = 1, 2, \dots, N$, in a column. The whole column will be represented by the letter of the variable, but in bold type. Using the variables of Table 1,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}; \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}; \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}. \quad \#$$

\mathbf{x}, \mathbf{y} and \mathbf{z} are then said to be N -tuples, since each comprises N real numbers called elements or co-ordinates. N -tuples are worth studying because they are the means by which statistical calculations and theory are facilitated.

An N -tuple of some importance in computing means and variances is the unitary N -tuples \mathbf{e} , of elements e_1, e_2, \dots, e_N , each of which is one:

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad \#$$

The principal reason why \mathbf{e} is important in computing is that it is the means by which the elements of an N -tuple may be summed. To understand this, two fundamental operations with N -tuples need to be explained.

The first operation is called *multiplication by a scalar*. If \mathbf{x} is an N -tuple of elements x_1, x_2, \dots, x_N and λ is a real number, then $\lambda\mathbf{x}$ has elements $\lambda x_1, \lambda x_2, \dots, \lambda x_N$ and the elements of \mathbf{x} are scaled by the factor λ ; λ is therefore referred to as a scalar. For example let $\lambda = 2, N = 3$ and \mathbf{z} be the N -tuple of Table 1:

$$\mathbf{z} = \begin{bmatrix} 47.3 \\ 65.0 \\ 49.9 \end{bmatrix} \Rightarrow 2\mathbf{z} = \begin{bmatrix} 2 \times 47.3 \\ 2 \times 65.0 \\ 2 \times 49.9 \end{bmatrix} = \begin{bmatrix} 94.6 \\ 130.0 \\ 99.8 \end{bmatrix}. \quad \#$$

Thus the elements of \mathbf{z} have been scaled by the factor 2, and 2 is scalar.

The second operation is called an *inner (or scalar, or dot) product*. The idea here is to define an operation which transforms N -tuples of real numbers into one number, a scalar. This is done by defining the inner product of any two N -tuples \mathbf{x} and \mathbf{y} . The inner product is written in various ways: $\mathbf{x}'\mathbf{y}$, (\mathbf{x}, \mathbf{y}) or $\mathbf{x} \cdot \mathbf{y}$. The first of these is called the inner product, the second is the scalar product and the third is the dot product of \mathbf{x} and \mathbf{y} , but all represent exactly the same operation. In economics, the most popular notation is $\mathbf{x}'\mathbf{y}$. Here the prime on \mathbf{x} denotes that the column \mathbf{x} is arranged as the corresponding row \mathbf{x}' as follows,

$$\mathbf{x}'\mathbf{y} = [x_1, x_2, \dots, x_N] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \equiv x_1y_1 + x_2y_2 + \dots + x_Ny_N = \sum_{i=1}^{i=N} x_iy_i. \quad \#$$

The N -tuple \mathbf{x}' is said to be $1 \times N$ (one row of N elements) while \mathbf{y} is $N \times 1$ (a column of N elements). Thus $\mathbf{x}'\mathbf{y}$ is $(1 \times N) \times (N \times 1)$ and the two N 's represent the inner orders that are to be multiplied together as x_iy_i and then summed to form the inner product. The outer orders are 1 and 1 representing a one-tuple or real number, a scalar. Since a scalar is produced, the product is sometimes called a scalar product.

Notice that the inner product is symmetrical, since

$$\mathbf{y}'\mathbf{x} = \sum_{i=1}^{i=N} y_ix_i = \sum_{i=1}^{i=N} x_iy_i = \mathbf{x}'\mathbf{y}. \quad \#$$

Also, if λ is a number, that is a scalar,

$$\lambda(\mathbf{x}'\mathbf{y}) = (\lambda\mathbf{x}')\mathbf{y} = (\lambda\mathbf{y}')\mathbf{x} = \lambda \sum_{i=1}^{i=N} x_iy_i. \quad \#$$

For example, consider \mathbf{x} and \mathbf{y} from Table 1. Since x_i is the population and y_i is the income per head in city i , x_iy_i is the total income in city i ; $\sum_{i=1}^{i=3} x_iy_i$ = aggregate personal income in 1991 in the three cities of Montreal, Toronto, Vancouver:

$$\begin{aligned} \mathbf{x}'\mathbf{y} &= (3,084.1 \times 18.73) + (3,822.4 \times 22.85) + (1,586.6 \times 19.20) \\ &= 57,756.193 + 87,341.84 + 30,462.72 \\ &= 175,569.75 \text{ millions of Canadian dollars.} \end{aligned} \quad \#$$

(The figures for total personal income differ from those in the text because the figures in Table 1 have been rounded.) In (7), it clearly does not matter whether x_i times y_i is calculated or y_i times x_i . This confirms that $\mathbf{y}'\mathbf{x} = \mathbf{x}'\mathbf{y}$. Moreover, if every element of \mathbf{x} , say, were to be multiplied by $\lambda = 2$, this would yield the same answer as $2\mathbf{x}'\mathbf{y}$ or $2\mathbf{y}'\mathbf{x}$ which illustrates (6).

The final property of an inner product is

$$\mathbf{x}'\mathbf{x} = \sum_{i=1}^{i=N} x_i^2 \geq 0. \quad \#$$

Moreover $\mathbf{x}'\mathbf{x} = 0$ only if all $x_i = 0$, that is, only if $\mathbf{x} = 0$.

Calculating the Population Mean

The discussion now turns towards application of the two operations: multiplication by a scalar and inner product. In these applications, the specific role of the unitary N -tuple in computations will become apparent.

$$\mathbf{e}'\mathbf{x} = \sum_{i=1}^{i=N} e_i x_i = \sum_{i=1}^{i=N} x_i = x_1 + x_2 + \dots + x_N. \quad \#$$

$$\mathbf{e}'\mathbf{e} = \sum_{i=1}^{i=N} e_i^2 = N. \quad \#$$

Therefore,

$$(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{x} = \frac{1}{N} \sum_{i=1}^{i=N} x_i = \mu_x, \quad \#$$

is the population mean of x . Of course, μ_x is a scalar.

An important property of a mean is that the sum of deviations from it is equal to zero. This is easily demonstrated: the N deviations from the mean are contained in $(\mathbf{x} - \mathbf{e}\mu_x)$, that is N terms $(x_i - \mu_x)$. The sum of these is

$$\begin{aligned} \mathbf{e}'(\mathbf{x} - \mathbf{e}\mu_x) &= \mathbf{e}'\mathbf{x} - \mathbf{e}'\mathbf{e}[(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{x}] \\ &= \mathbf{e}'\mathbf{x} - \mathbf{e}'\mathbf{x} \\ &= 0. \end{aligned} \quad \#$$

This property holds for any mean. For example, for $\mu_y = (\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{y}$ and $\mu_z = (\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{z}$, $\mathbf{e}'(\mathbf{y} - \mathbf{e}\mu_y) = \mathbf{e}'(\mathbf{z} - \mathbf{e}\mu_z) = 0$.

Deviations from the mean sum to zero because some deviations are positive and some are negative. When these are summed together, the sum of the negative deviations just balances the sum of the positive deviations. When the deviations about the population mean are squared, each is positive or zero and, insofar as they are not all zero, their sum must be positive. The mean of the sum of squared deviations from the mean is the population variance to which attention is now directed.

Calculating the population variance

The N -tuple of deviations from the mean is

$$(\mathbf{x} - \mathbf{e}\mu_x) = \begin{bmatrix} x_1 - \mu_x \\ x_2 - \mu_x \\ \vdots \\ x_N - \mu_x \end{bmatrix}, \quad \#$$

From (8) $(\mathbf{x} - \mathbf{e}\mu_x)'(\mathbf{x} - \mathbf{e}\mu_x) \geq 0$, with equality only if $(\mathbf{x} - \mathbf{e}\mu_x) = 0$, which generally will not be the case since this requires $\mathbf{x} = \mathbf{e}\mu_x$ and no variation in the x_i . Now from (5), (9) and (10), $\mathbf{x}'\mathbf{e} = \mathbf{e}'\mathbf{x} = N\mu_x$ with $N = \mathbf{e}'\mathbf{e}$ and from (13):

$$\begin{aligned}
\sigma_x^2 &= \frac{1}{N}(\mathbf{x} - \mathbf{e}\mu_x)'(\mathbf{x} - \mathbf{e}\mu_x) = \frac{1}{N} \sum (x_i - \mu_x)^2 \\
&= \frac{1}{N} \{ \mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{e}\mu_x - \mu_x\mathbf{e}'\mathbf{x} + \mathbf{e}'\mathbf{e}\mu_x^2 \} \\
&= \frac{1}{N} \{ \mathbf{x}'\mathbf{x} - N\mu_x^2 - N\mu_x^2 + N\mu_x^2 \} \\
&= \frac{1}{N} \{ \mathbf{x}'\mathbf{x} - N\mu_x^2 \}, \tag{#}
\end{aligned}$$

substituting $\mathbf{e}'\mathbf{e}$ for N and $(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{x}(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{x} = \mu_x^2$,

$$\begin{aligned}
\frac{1}{N}(\mathbf{x} - \mathbf{e}\mu_x)'(\mathbf{x} - \mathbf{e}\mu_x) &= \frac{1}{N} \{ \mathbf{x}'\mathbf{x} - (\mathbf{e}'\mathbf{e})(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{x}(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{x} \} \\
&= \frac{1}{N} \{ \mathbf{x}'\mathbf{x} - \mathbf{e}'\mathbf{x}(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{x} \},
\end{aligned}$$

which, since $\mathbf{x}'\mathbf{e} \equiv \mathbf{e}'\mathbf{x}$, becomes

$$\begin{aligned}
\frac{1}{N}(\mathbf{x} - \mathbf{e}\mu_x)'(\mathbf{x} - \mathbf{e}\mu_x) &= \frac{1}{N} \{ \mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{e}(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{x} \} \\
&= \frac{1}{N} \sum_{i=1}^{i=N} x_i^2 - \left\{ \frac{1}{N} \sum_{i=1}^{i=N} x_i \right\}^2. \tag{#}
\end{aligned}$$

The expression $\mathbf{e}'(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}' = \mathbf{p}$ in (15) plays an important role in statistics. In order to understand what \mathbf{p} represents, a new operation, called an outer product, must be considered.

An *outer product* is defined in terms of an inner product. Consider three N -tuples \mathbf{x} , \mathbf{y} and \mathbf{z} . Then the inner product of \mathbf{y} and \mathbf{z} may be written $\mathbf{y}'\mathbf{z}$ and since $\mathbf{y}'\mathbf{z}$ is a scalar, $\mathbf{x}(\mathbf{y}'\mathbf{z})$ is an N -tuple \mathbf{x} multiplied by a scalar $\mathbf{y}'\mathbf{z}$.

Now,

$$\mathbf{x}(\mathbf{y}'\mathbf{z}) = (\mathbf{xy}')\mathbf{z}, \tag{#}$$

and \mathbf{xy}' is called the outer product of \mathbf{x} and \mathbf{y} . From the left-hand side of (16)

$$\mathbf{x}(\mathbf{y}'\mathbf{z}) = \begin{bmatrix} x_1\mathbf{y}'\mathbf{z} \\ x_2\mathbf{y}'\mathbf{z} \\ \vdots \\ x_N\mathbf{y}'\mathbf{z} \end{bmatrix} = \begin{bmatrix} x_1[y_1, y_2, \dots, y_n] \\ x_2[y_1, y_2, \dots, y_n] \\ \vdots \\ x_N[y_1, y_2, \dots, y_n] \end{bmatrix} \mathbf{z}, \tag{#}$$

and the last expression in (17) involves N scalars x_1, x_2, \dots, x_N each times the row N -tuple $\mathbf{y}' = [y_1, y_2, \dots, y_N]$. It then follows that

$$\mathbf{x}(\mathbf{y}'\mathbf{z}) = \begin{bmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_N \\ x_2y_1 & x_2y_2 & \dots & x_2y_N \\ \dots & \dots & \dots & \dots \\ x_Ny_1 & x_Ny_2 & \dots & x_Ny_N \end{bmatrix} \mathbf{z} = (\mathbf{xy}')\mathbf{z}. \tag{#}$$

The square array of N^2 numbers preceding \mathbf{z} in (18) has the number x_iy_j in the i 'th row and j 'th

column, $i, j = 1, 2, \dots, N$. This square array is called a (square) matrix. By the same reasoning, $\mathbf{p} = \mathbf{e}(\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}' = \frac{1}{N}\mathbf{e}\mathbf{e}'$ and $\mathbf{e}\mathbf{e}'$ is the outer product of \mathbf{e} and itself; its i, j 'th element must be $e_i e_j = 1$. Thus

$$\mathbf{p} = \begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix},$$

an $N \times N$ matrix with every element equal to $\frac{1}{N}$; the meaning of $\mathbf{p}\mathbf{x}$ is the same as N rows of the kind

$$\left[\frac{1}{N}, \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}\right]\mathbf{x} = \mu_x.$$

Hence

$$\mathbf{p}\mathbf{x} = \begin{bmatrix} \mu_x \\ \mu_x \\ \vdots \\ \mu_x \end{bmatrix} = \mathbf{e}\mu_x.$$

Another square matrix of importance is

$$\mathbf{I}_N = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

a diagonal matrix with one in each diagonal position, zeros elsewhere. The product $\mathbf{I}_N\mathbf{x}$ is interpreted like $\mathbf{p}\mathbf{x}$ as N inner products of each row of \mathbf{I}_N and \mathbf{x} :

$$\mathbf{I}_N\mathbf{x} = \mathbf{x}.$$

Returning to (15) using \mathbf{p} and \mathbf{I}_N ,

$$\begin{aligned} \frac{1}{N}(\mathbf{x} - \mathbf{e}\mu_x)'(\mathbf{x} - \mathbf{e}\mu_x) &= \frac{1}{N}\{\mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{p}\mathbf{x}\} \\ &= \frac{1}{N}\mathbf{x}'(\mathbf{I}_N - \mathbf{p})\mathbf{x}. \end{aligned} \quad \#$$

Thus the population variance, σ_x^2 , is

$$\frac{1}{N}(\mathbf{x} - \mathbf{e}\mu_x)'(\mathbf{x} - \mathbf{e}\mu_x) = \frac{1}{N}\mathbf{x}'(\mathbf{I}_N - \mathbf{p})\mathbf{x} = \frac{1}{N}\sum_{i=1}^{i=N} x_i^2 - \left\{\frac{1}{N}\sum_{i=1}^{i=N} x_i\right\}^2. \quad \#$$

Notice two important properties of \mathbf{p} and $(\mathbf{I}_N - \mathbf{p})$: both are symmetric matrices in the sense that, if the rows were to be written as columns, the matrices would remain unaltered; also $\mathbf{p}^2 = \mathbf{p}$ and

$(\mathbf{I}_N - \mathbf{p})^2 = (\mathbf{I}_N - \mathbf{p})$. Matrices of this kind are said to be symmetric and idempotent.

A Digression

Suppose you were asked to add up all the integers identifying the observations in a population: $i = 1, 2, \dots, N$. This could be done by defining the N -tuple:

$$\mathbf{r} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ N \end{bmatrix},$$

and solving the inner product $\mathbf{e}'\mathbf{r}$. Alternatively, if the sum of the elements is S_N then

$$S_N = 1 + 2 + 3 + \dots + N. \quad \#$$

Or, in reverse order,

$$S_N = N + (N - 1) + (N - 2) + \dots + 3 + 2 + 1. \quad \#$$

Adding (21) to (22) yields $2S_N$ and the sum

$$\begin{aligned} 2S_N &= (N + 1) + (N + 1) + \dots + (N + 1) \\ &= N(N + 1). \end{aligned} \quad \#$$

$$\Rightarrow S_N = \frac{N}{2}(N + 1). \quad \#$$

Thus if $N = 179$ (as in the *Introduction*),

$$S_{179} = \frac{179}{2}(179 + 1) = \frac{179}{2}(180) = 179 \times 90 = 16,110. \quad \#$$

This is a useful way to check that $\mathbf{e}'\mathbf{r}$ has been calculated correctly. In this case, $\mathbf{e}'\mathbf{r}$ is a ponderous route to reaching a conclusion which may also be reached by mental arithmetic applied to (24). There are many cases in Econ 221 when a little thought may be the means to achieve improved accuracy with less work compared with the straightforward computation. Thus the most accurate route to computing a population variance is the last formula of (15), not the direct route as given in (14).

Model Formulation and Estimation

The principal model of Econ 221 takes the form

$$y_i = \mu_y + \varepsilon_i, \quad i = 1, 2, \dots, N. \quad \#$$

In terms of N -tuples, (26) is equivalent to

$$\mathbf{y} = \mathbf{e}\mu_y + \boldsymbol{\varepsilon}, \quad \#$$

$\boldsymbol{\varepsilon}$ being the N -tuple of errors which is clearly equivalent to $(\mathbf{y} - \mathbf{e}\mu_y)$. By (12) above it is also clear that

$$\frac{1}{N}\mathbf{e}'(\mathbf{y} - \mathbf{e}\mu_y) = \frac{1}{N}\mathbf{e}'\boldsymbol{\varepsilon} = 0. \quad \#$$

and so the population mean of $\boldsymbol{\varepsilon}$ is zero.

In practice (26) and (27) are taken to be correct but μ_y is unknown; and μ_y is unknown because the whole population is not observed. What is observed is a (random) sample (or subgroup) of the population $\{y_1, y_2, \dots, y_n\}$, $n < N$, and use has to be made of this sample to estimate μ_y and σ_y^2 . How is this done?

A natural approach is to make use of (27) and (28), presuming that these must hold for the sample. This permits the estimation μ_y . If $\hat{\mu}_y$ is an estimate of μ_y , then for the n sample observations, that is for (28) formulated in n -tuples as opposed to N -tuples,

$$\begin{aligned} \mathbf{e}'(\mathbf{y} - \mathbf{e}\hat{\mu}_y) &= 0 \\ \Rightarrow \hat{\mu}_y &= (\mathbf{e}'\mathbf{e})^{-1}\mathbf{e}'\mathbf{y} = \bar{y} \end{aligned} \quad \#$$

the sample mean. It turns out that \bar{y} is also the estimate of μ_y that minimizes the sum of squared errors in the sample $\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$, where $\hat{\boldsymbol{\varepsilon}}$ is any linear estimate of $\boldsymbol{\varepsilon}$.

To estimate the variance σ_y^2 it might seem appropriate to use

$$\frac{1}{n} \sum_{i=1}^{i=N} (y_i - \bar{y})^2 = \frac{1}{n} \mathbf{y}'(\mathbf{I}_n - \mathbf{p})\mathbf{y}, \quad \#$$

where \mathbf{I}_n and \mathbf{p} are defined for n observations, not N as in (20). (30) is not the best estimate of σ_y^2 , because \bar{y} has to be used, rather than the true value of the mean μ_y . A better estimate is

$$\begin{aligned} s_y^2 &= \frac{1}{n-1} \mathbf{y}'(\mathbf{I}_n - \mathbf{p})\mathbf{y} \\ &= \left\{ \frac{1}{n-1} \sum y_i^2 \right\} - \left\{ \frac{(\sum y_i)^2}{n(n-1)} \right\}. \end{aligned} \quad \#$$

s_y^2 is called the sample variance. Of course in general $\bar{y} \neq \mu_y$ and $s_y^2 \neq \sigma_y^2$.