

# TOPIC 8

## PROPORTIONS

*The one formula to which every political problem is reduced:  
the necessity of statistics, and the ruling principle for their compilation.*

*Rosmini*

### Introduction

In this topic some of the main types of statistical indicators used in statistical reports are defined and discussed. The objectives are to become aware of the way statistical indicators are used in data analysis; and to become familiar with the different types of indicators and the terms and concepts used in constructing indicators.

### Proportions

To provide meaningful and useful statistics it is usually necessary to compare the counts of a set of observations using some kind of proportion. Proportions may be expressed as a percentage, a ratio or a rate.

Using these statistical measures we can measure the change occurring in a country or region from year to year. As the rates are calculated the same way in each country or region, we can compare the rate in one country or region with that in another. There is, however, a difficulty in using proportions.

#### **Compositional variables**

When comparing populations care should be taken that compositional variables are taken into consideration. That is, different populations may have different structures that affect the comparability of the measures.

For example, crude birth and death rates are affected by the proportions of persons in the different ages and can give misleading comparisons. It is therefore necessary to adjust these rates for differences in age composition – i. e. age-specific rates.

Gender is also a compositional variable that needs to be taken into account during data analysis. It is good business practice to disaggregate proportions by gender and age group, especially where there are significant differences in the relative distributions.

The Asia Pacific Economic Cooperation (APEC) Ministerial Meeting on Women, held in Manila in 1998, recognized the need to reinforce the important role of women in economic development. That meeting initiated the development of the Framework for the Integration of Women in APEC, which ... included the following as principal recommendations for areas of improvement: Gender Analysis, Collection and Use of Sex-Disaggregated Data, and the Involvement of Women in APEC. Sex-disaggregated data can be used to illustrate the differing realities of women and men in society and provide the foundation for policies and programs to address the needs of women.

**Data Disaggregation**

It is important to note that presenting data at an aggregate level may be misleading if compositional variables are not taken into account. This is called the Simpson’s Paradox. It is usually desirable to report statistics at lower levels of aggregation and provide rates, ratios and percentages separately for males and females, and for different age groups.

**Simpson's paradox** is a statistical paradox described by E. H. Simpson in 1951: an association in sub-populations may be reversed in the population. It appears that two sets of data separately support a certain hypothesis, but, when considered together, they support the opposite hypothesis.

**Standardisation**

One way of adjusting the population for compositional effects is to standardise the population. Standardisation involves choosing a standard or reference population and substituting that population’s composition for the population of interest. This technique makes the composition of the populations identical and calculates the rates so that the difference in the rates is not due to the compositional variables.



The aggregation of data may hide the effects of compositional variables and can sometimes provide misleading results. The following section looks at the implications of the Simpson’s Paradox for data analysis.

**Simpson’s Paradox – an example**

A good way to see how Simpson’s Paradox occurs is to look at an example. Suppose a researcher at the University of the South Pacific released the following data.

Table 1 Percentage of Students Passing First Year University Courses

Male	60%
Female	44%

**males are doing better than females ?**

The researcher suggests that this is good evidence that males are doing better at university than females. You think this is a bit strange, so you decide to investigate further and ask for the results by subject. For the purpose of this exercise, suppose students can only do one of two subjects, Science or English.

Table 2 Percentage of Students Passing First Year USP by Type of Course

Course	Male	Female
Science	70%	80%
English	20%	40%

**females are doing better than males?**

By breaking down the results, you learn that females are in fact doing better in both the courses offered. So what is happening here? Could it be that the compositional structure of the population (that is, the number of students doing Science and the number of students doing English) are affecting the results? Lets look at the numbers that make up the percentages in Table 2.

**Table 3 Number of Students Passing First Year USP by Type of Course**

Course	Male		Female	
	Number of Students who passed	Number of Students	Number of Students who passed	Number of Students
Science	560	800	80	100
English	40	200	360	900
<b>Total</b>	600	1,000	440	1,000

Table 3 shows that the vast majority of male students take science relative to females. So if we combine the numbers of students who passed both science and English then the aggregate pass rate is in favour of males (i.e. 60% compared to 44%).

The paradox arises because we assume that combining the data will average the proportions. However, by combining the data we weight the individual proportions so that the aggregate proportion is pulled in the direction of the largest weight. That is, the proportion in the combined data is the average of the proportions in each group contributing, weighted by the size of the contribution to the combined data.

**TIP**



It is good practice to break down aggregated data into its compositional parts, especially data that is composed of gender, age and geographical information.

## Rates

A rate is the number of cases of a particular outcome of interest that occur over a given period of time divided by the size of the population in that time period. That is, a rate is a measurement of one quantity in relation to another for a specific time period. It is similar to a proportion except a multiplier or base is used e.g. per 1,000, 10,000, 100,000.

You can use rates to express the number of cases or occurrences of an event so that the proportions can be easily measured and compared. Note that it is common practice for rates to be calculated separately for different sub-populations, such as: for each gender, by geographical location and separate age groups.

### Examples of rates used in official statistics are:

Literacy rate

Infant mortality rate

Unemployment rate

### Example: Adult Literacy Rate

#### definition

The **adult literacy rate** is defined as the percentage of population aged 15 years and over who can both read and write with understanding a short simple statement on his/her everyday life.

The purpose of the adult literacy rate is to show the accumulated achievement of primary education and literacy programmes in imparting basic literacy skills to the population, thereby enabling them to apply such skills in daily life and to continue learning and communicating using the written word.

To calculate the adult literacy rate divide the number of literates by the corresponding age group population and multiply the result by 100.



**Formula**

$$\frac{\text{Number of literates}}{\text{Population aged 15 years or over}} \times 100$$

**example**

For example, from the Tonga 1996 census the adult literacy rate is calculated as:

$$\frac{57416}{58056} \times 100 = 98.9\%$$

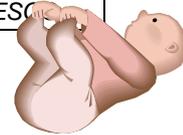
**meaning**

A high literacy rate indicates a wide coverage of the primary education system and/or literacy programmes in that a large proportion of the population has acquired the ability of using the written word in daily life and to continue learning.

Note that it is usual to present and analyse literacy rates together with the absolute number of adult illiterates as improvements in literacy rates may sometimes be accompanied by increases in the illiterate population due to the changing demographic structure.

The youth literacy rate for persons aged 15 to 24 years old reflects the accomplishment of a country's basic education system. It measures the population's ability to read, write, and communicate and thus reflects to some extent their ability to continue learning using the written word. It is also often conceptualized as a proxy measure of social progress and economic achievement. Literacy are being used as core indicators for monitoring and assessing progress towards the UN Millennium Development Goals as well as the Education For All Goals.

*UNESCO Institute for Statistics. Education Indicators. Technical Guidelines. UNESCO*



**Example: Infant Mortality Rate**

**definition**

The infant mortality rate is the number of deaths of infants under one year of age for each 1,000 live births that occur in a given time period, usually a given calendar year.

**formula**

$$\frac{\text{Number of deaths of infants}}{\text{Total live births}} \times 1,000$$

**example**

For example, if in a given year there were 38 infant deaths and 3,734 live births, the infant mortality rate would be calculated:

$$\frac{38}{3,734} \times 1,000 = 10.2$$

This means there were 10 (rounded) deaths of infants under 1 year of age for every 1,000 births.

**Example: Unemployment Rate**

**definition**

The number of persons who are unemployed is also expressed as a rate. The unemployment rate is the number of unemployed persons in relation to the labour force, not the total population.



**formula**

$$\frac{\text{Number of persons unemployed}}{\text{Number of persons in the labour force}} \times 100$$

**example**

For example, in the 1991 population census of Samoa the number of unemployed enumerated was 1,175 and 55,967 were employed. Therefore the total unemployment rate was:

$$\frac{1,175}{57,142} \times 100 = 2.1\%$$

**meaning**

You will note that the unemployment rate is also expressed as a percentage. The reason the unemployment rate is also expressed as a percentage is because the rate is calculated as a proportion of 100. For instance, if the unemployment rate was 10% that would mean that there were 10 unemployed persons for every 100 persons in the labour force.

The youth unemployment rate (youth unemployment as a percentage of the youth labour force) is a general measure of utilization of the labour force of young persons. Youth unemployment rate alone cannot fully gauge the ability of youth to meet their full-productive capacity; however, it can serve as a useful proxy of the labour market situation facing this group. Other indicators of interest in assessing issues facing youth employment issues should include, therefore, measures of underemployment, the informal sector, educational access and labour force participation, amongst others.

*Castles, I. Surviving Statistics. A user guide to the basics, Australian Bureau of Statistics, 1992*

**Ratios**

The ratio is the relation which one quantity bears to another of the same kind. The definition of a ratio is the number of observations in a given group with a certain characteristic divided by the number of observations without the given characteristic. Ratios are used to compare the frequency of one category with that of another category.

The ratio of the frequency of one category to another is written as “frequency of one category: frequency of the other category”, but is commonly expressed as a proportion or percentage. The ratio is measured by the fraction “frequency of one category/ frequency of the other category”. Note that some of the ratios used in the construction of statistical indicators do not conform to this definition.

As with rates, ratios can be disaggregated to provide comparisons between different sub-populations. For example, ratios can be calculated separately for each gender, geographical region or age groups.

**Examples of ratios used in official statistics:**

Sex ratio

Enrolment ratio

Maternal mortality ratio

**Example: Sex Ratio****definition**

The sex ratio is the ratio of males to females in a given population, expressed as the number of males for every 100 females.

**formula**

$$\frac{\text{Number of males}}{\text{Number of females}} \times 100$$



**example**

For example, in a population of 25,643 males and 23,721 females, the sex ratio is calculated as:

$$\frac{25,643}{23,721} \times 100 = 107.3$$

**meaning**

This means there are 107 (rounded) males per 100 females.

Gender equality in terms of access to all levels of education has been identified as a priority at both national and international levels, and is one of the goals of the Education For All (EFA) initiative. Within the Millennium Development framework, the indicators used to measure progress towards gender equality in education are the ratio of female to male enrolment by level of education and the ratio of literate females to males aged 15-to-24 years.

*UNESCO Institute for Statistics. Education Indicators. Technical Guidelines. UNESCO*



**Example: Enrolment Ratio**

If we want to show the extent of participation of children and youths in a given level of education we use the enrolment ratio. There are two types of enrolment ratio – the net enrolment ratio and the gross enrolment ratio.

The net enrolment ratio is defined as the enrolment of the **official age-group** for a given level of education as a percentage of the corresponding population. The gross enrolment ratio is defined as the **total enrolment** for a given level of education as a percentage of the official school-aged population.

To calculate the net enrolment ratio we divide the number of pupils enrolled by the population for the same age-group and multiply the result by 100.

**formula**

$$\frac{\text{Enrolment of the population of school age}}{\text{Population in school age group}} \times 100$$

**example**

For example, in Samoa there were 35,330 primary enrolments in 1999 and a projected population aged 5-13 of 38,206. The net enrolment ratio is calculated:

$$\frac{35330}{38206} \times 100 = 92.5\%$$

**meaning**

A high net enrolment ratio denotes a high degree of participation of the official school-age population. Increasing trends can be considered as reflecting improved participation at the specified level of education.

Note that the ratio should be based on the total enrolment in all types of schools and education institutions, including public, private and all other institutions that provide organized programmes.

Achievement of universal primary education for all has been identified as a priority at both national and international levels. In the Millennium Declaration, as in the Education For All (EFA) initiative, the overriding goal is for all countries to ensure that by 2015, children everywhere will be able, at the minimum, to complete a full course of primary schooling of high quality. One of the three indicators used to measure progress towards this goal is the net enrolment ratio for primary education.

UNESCO Institute for Statistics. Education Indicators. Technical Guidelines. UNESCO.

### Example: Maternal Mortality Ratio

#### definition

The maternal mortality ratio is the number of women who die as a result of child-bearing in a given year per 100,000 births in that year. Maternal deaths are those caused by complications of pregnancy and childbirth.

#### formula

$$\frac{\text{Number of maternal deaths} \times 100,000}{\text{Total number of live births}}$$



#### example

For example, if in a given year there were 6 maternal deaths and a total of 3,867 live births, the maternal mortality ratio is calculated as:

$$\frac{6}{3,867} \times 100,000 = 155.15$$

#### meaning

This means there were 155 (rounded) deaths per 100,000 live births. Note that this calculation gives rise to problems when applied to Pacific Island countries as the number of births is relatively small, and therefore the ratio expressed per 100,000 births may be misleading.

In countries without comprehensive registration of deaths and good attribution of causes of death, ...survey methods have to be used to estimate levels [of maternal mortality] and these are subject to wide margins of uncertainty and cannot be used to monitor trends. Therefore, there is consensus that for assessing trends over time, other indicators should be used.

*Report of the Inter-agency Expert Group on MDG Indicators, New York, April 2002*

## Percentages

A proportion is the number of observations or responses with a given characteristic divided by the total number of observations. In other words, a proportion is a part divided by a whole. A percentage is a proportion multiplied by 100.

#### Percentage of Total

We often use percentages to compare and assess numbers in frequency tables. It is easier to compare numbers when they have a common base, especially when the numbers are large. We tend not to use percentages when the numbers are small (i.e. less than 50).

To calculate a percentage, divide the amount by the total and multiply by 100.

$$\frac{\text{Amount} \times 100}{\text{Total}}$$

#### Examples: Proportion of 1 year old children immunized against measles

#### definition

If we wanted to calculate the number of 1 year old children who were immunized against measles as a percentage of the total population of 1 year olds, we would calculate

**formula**

$$\frac{\text{Number of immunized children aged 1 year}}{\text{Total number of children aged 1 year}} \times 100$$

**definition**

For example, if there were 3210 children who were immunized against measles out of a total of 3500 children aged one year in the population then the proportion would be calculated:

$$\frac{3210}{3500} \times 100 = 91.7\%$$



Immunization is an essential component in reducing mortality rates for children under five years of age. Among the childhood vaccine-preventable diseases, measles is the leading cause of child mortality... Measles immunization coverage can be assessed annually and hence is a useful indicator of progress on one of the key causes of under-five deaths.

*United Nations. Millennium Development Goals Data and Trends, 2002*

**Comparison of percentages**

When we compare two different percentages we refer to the difference in percentage points. For example, the difference between 92% and 95% is 3 percentage points. If the percentages refer to the same characteristic over different time periods we can say that there was an increase of 3 percentage points.

**Percentage Change**

Often we need to find the amount of percentage change between two numbers over time. That is, we need to calculate the percentage one number is of another between different time periods.

**Example: Population Percentage Increase**

**definition**

If we wanted to calculate the percentage increase in a population between two years, we would subtract the two populations and divide by the population of the earlier year.



**formula**

$$\frac{P1 - P0}{P0} \times 100$$

Where: P1 = population at time of most recent census  
 P0 = population at time of census before

**example**

For example, the population of Matereva village was 2,376 in 1984 and in 1994 it was 2,769. What is the percentage increase between 1984 and 1994.

$$\frac{2769-2376}{2376} \times 100 = 16.5\%$$

**meaning**

We call the denominator the base, since we are comparing the number of increase with the base number. The percentage is calculated by dividing the amount of increase by the base and multiplying by 100.

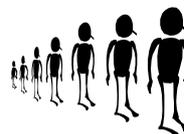
**Example: Population Percentage Growth**

**definition**

We sometimes need to find the annual percentage change between two numbers. That is, we need to calculate the annual rate of percentage change between two observations made at different time periods.

In cases where there is no or little compounding effect the calculation can be made by dividing the percentage change by the exact number of years (i.e. the arithmetic growth rate). However for population-based data the calculation requires the use of an exponential growth formula to take account of the compounding effect of the growth rate.

The formula for exponential population growth rate is:

**formula**

$$r = \frac{\ln(P_1/P_0)}{t} \times 100$$

Where  $P_1$  = population at time of most recent census  
 $P_0$  = population at time of census before  
 $r$  = annual growth rate between two censuses  
 $t$  = exact time between these two censuses  
 $\ln$  = natural logarithm

To calculate the annual rate of growth, divide the most recent population by the earlier population and take the natural logarithm of the resultant, and then divide by the time difference between the two populations. The annual percentage growth rate is the result multiplied by 100.

**example**

For example, we can calculate the annual population growth for the Marshall Islands between 1980 (pop. 30,873) and 1988 (pop. 43,380). The exact difference between the censuses is 8.17 years.

$$\frac{\ln(43,380/30,873)}{8.17} \times 100 = 4.16\%$$

In excel we can use the LN natural log function to calculate the exponential rate of increase for a population. The function syntax is:  $r = \text{LN}(P_1/P_0)/t*100$

**References:**

Castles, I. Surviving Statistics. A user guide to the basics, Australian Bureau of Statistics, 1992

Hunter, D. Getting started in statistics: A handbook for Pacific Women Leaders, Secretariat of the Pacific Community, 1998

Westbrooke, I. Simpson's Paradox: An example in a New Zealand Survey of Jury Composition. Statistics New Zealand, 1997

## ... Exercises ...

**1. Indicate which of the following proportions are rates, ratios or percentages:**

- (a) The proportion of children immunized in a country \_\_\_\_\_
- (b) The proportion of males to females \_\_\_\_\_
- (c) The proportion of students completing school \_\_\_\_\_
- (d) The proportion of 15-24 year olds in a population \_\_\_\_\_
- (e) The proportion of under 5 year olds to over 60 year olds \_\_\_\_\_
- (f) The proportion of households with access to safe water \_\_\_\_\_

**2. From the information in the tables on the next page calculate:**

- (a) the sex ratio of total males and females \_\_\_\_\_
- (b) the percentage of total population in the 5 - 9 age group \_\_\_\_\_
- (c) the percentage of males over 65 years of age \_\_\_\_\_
- (d) the maternal mortality ratio \_\_\_\_\_
- (e) the infant mortality rate \_\_\_\_\_
- (f) the crude birth rate \_\_\_\_\_
- (g) the teenage fertility rate (for women aged 15-19 years) \_\_\_\_\_

**3. Why should we disaggregate statistical data by sex and age groups?**

---



---

4. If the unemployment rate went from 5% in December 2002 to 8% in December 2003, how would you describe the increase?
- 
- 

**Population by Age and Sex,  
1990**

Age Group	Male	Female	Total
0 - 4	3,589	3,354	6,943
5 - 9	2,924	2,716	5,640
10 -14	2,812	2,417	5,229
15 -19	2,396	2,322	4,718
20 -24	2,301	2,363	4,664
25 - 29	1,986	2,175	4,161
30 - 34	1,807	1,706	3,513
35 -39	1,359	1,362	2,721
40 - 44	1,178	1,068	2,246
45 -49	974	805	1,779
50 - 54	835	687	1,522
55 - 59	598	495	1,093
60 - 64	467	456	923
65 - 69	324	343	667
70 -74	223	201	424
75 +	241	280	521
<b>Total</b>	<b>24,014</b>	<b>22,750</b>	<b>46,764</b>

**Vital Statistics, 1990**

Total live births	1,953
Infant deaths	15
Maternal deaths	4
Live births to women aged 15-19	114

Source: *Getting Started in Statistics: A Handbook for Pacific Women Leaders*  
 Secretariat of the Pacific Community  
 1998

