Statistics Canada Statistique Canada



Statistics Canada
Analytical Studies Branch

BAN

# Research Paper Series

Canada

# UNEMPLOYMENT AND TRAINING*

by

Garnett Picot

No. 2

Social and Economic Studies Division
Statistics Canada
1987

* This paper is an excerpt from a larger paper entitled "The Participation in Training by Women, the Unemployed, and the Educationally Disadvantaged, Research Paper No. 24, Social and Economic Studies Division, 1986.

## ABSTRACT

Training if often discussed as a principle means of improving the labour adjustment process for the unemployed. But if training is to be effective for particular target groups of unemployed, it is necessary to know to what degree training is actually utilized by the group. That is the question addressed in this paper. Using logistic regression and data from two surveys, the probability of taking training is determined for the unemployed with various characteristics. It is found that being unemployed increases significantly the likelihood of training. It is also found that often groups of the unemployed who face the most difficult adjustment experiences and the most difficult labour markets are those who are least likely to turn to training.

## ACKNOWLEDGEMENTS

The author wishes to thank Ted Wannell and Marie-Claire Couture for their assistance throughout this project.

## Introduction

Expectations for training and education in Canada have been high and multi-faceted. Individuals and families have looked to the education system to foster social mobility (usually meaning a better job with higher pay). Economists and politicians have looked to the system to promote productivity and international competitiveness with a resulting increase in wealth, profits, and standard of living. During the 1960s and early 1970s, expectations focussed particularly on the young, on what training and education could do for them, and through them, for the economy. The extent to which these expectations have been met has been debated at length.

By the 1980s, the number of young people started to decline, concerns about technological change and decreasing international competitiveness were rising, and the worst economic slowdown since the 1930s sent unemployment to record levels. In this environment, adult training came to the forefront.

Confronted with turmoil and uncertainty in the labour market, people began to realize that the most important component of education and training may not be specific technical knowledge acquired in any given program, but rather the process of learning how to learn. Many argue that since the work of the future is so uncertain, specificity is folly. Adaptation and adjustment to changing demands, achieved largely through retraining, will be of paramount importance. Confronted with a rapidly changing workplace and technology, the ability to learn may emerge as a premium skill. It may not be what one knows, but what one can learn - and how fast - that will determine how good a worker is. At this particular time the focus is on retraining, continued training and upgrading of adults.

Government analysts, special interest groups, politicians, economists, and educators are particularly interested in retraining the unemployed. This applies to those with obsolete skills, as well as the more general need to retrain adults so as to produce a flexible, highly skilled labour force able to cope with technological change. Furthermore,the training of particular target groups is perceived as a means of reducing income disparities, particularly those that afflict women, the unemployed and the educationally disadvantaged.

Assessing success of training programs usually entails an examination of wage gains after training, the employability of graduates in a related field, changes in productivity levels, and related economic measures. But training is often directed at a particular target group, and in such cases, one very fundamental and almost elementary piece of information is frequently neglected - the probability that people in the target group actually take training, that is, the proportion who enroll. Knowing the number who train is not enough; it is also necessary to know the number who do not. It is, thus, essential to ascertain the probability that persons in any particular target group will train before employment objectives of training can be evaluated. It is unrealistic to assume that adult training will play a major role in improving skills, employability, productivity and wages if only a very small percentage of a target group participates. Programs may help those who do enroll, but without adequate representation they are unlikely to impact on the group as a whole.

It is also necessary to know something about kinds of training received as well as who offers and sponsors it (eg., colleges, employers, public training programs). For example, officials who design and implement government training programs should be aware of the extent to which adults in a particular target group take training in colleges, universities, industry, unions, professional organizations or other organizations. Without such knowledge, a program implemented and funded with tax dollars may duplicate existing programs. Referring to short "upgrading" courses offered by employers under the National Training Act, the Economic Council, (1982) noted: "The paucity of data on training in industry is a glaring example of basic deficiencies with respect to current detailed occupational information ... risks are involved in instituting government programs without greater knowledge of the training effort in industry."

The aforementioned considerations may seem elementary, but they have been largely ignored in the past, probably because of the lack of data on who takes training and, who does not. This study attempts to provide just such information for the long-term unemployed. In their case, training or retraining is often viewed by policy-makers as an important facet of the labour market adjustment process. This accords with the premise that people permanently laid-off from industries and occupations in which there is little hope of securing stable employment require training to locate new jobs (Economic Council:1983; Saunders:1984; Pearson and Salembier:1983).

## Focus on Long-Term Unemployment

Government at all levels in Canada is struggling with the problem of improving employment prospects of the unemployed, and in particular, the long-term unemployed. When the 1981-82 recession dramatically increased the frequency and duration of spells of unemployment, it was an increase in duration that contributed most to rising unemployment rates. About 80% of the change in the national unemployment rate over the 1979-84 period was due to an increase in the duration of unemployment spells (Barrett:1985). In other words, rising unemployment has resulted more in a problem of increasing long-term unemployment than a problem of increasing numbers of workers becoming unemployed. Investments in training to redirect the unemployed toward new jobs are more likely to be justified for those who have been unsuccessfully seeking work for some time. Retraining may be particularly necessary when structural economic problems cause the unemployment. Furthermore, adults who have been unemployed for a considerable period may be more apt to regard training as a potential solution. Thus, the emphasis here is on training among those unemployed more than six months during 1983.

An additional reason for concentrating on this group is their large numbers. It is generally known that between 1.3 and 1.5 million people are unemployed at any one time (according to the monthly counts), but it is less well-known that approximately one million Canadians were unemployed for more than six months during 1983. They represented 30% of all persons who were unemployed at any time during 1983, and fully 7.7% of the labour force. In other words, one in every thirteen labour force members experienced more than six months of unemployment during the year. And this is an underestimate, because the data refer to the 1983 calendar year; some longer periods of unemployment at the beginning and end of the year are truncated and appear in the data as short spells.

Training is Canada's most important active labour market policy (versus direct job creation, work sharing), to tackle chronic unemployment. For some time, training the unemployed has been a large part of Canadian labour market policy. Employment and Immigration's Consultation Paper on Training (E&IC: 1984c) notes: "Existing government-sponsored training is aimed at people who are unemployed because they do not have needed skills." Of the 63,000 trainees who started skill-development courses funded through the National Training Act in 1982-83, fully 76% were unemployed. But training the unemployed has not been a totally expedient solution, as there has often been a mismatch between the areas in which the unemployed are trained and the availability of jobs.

Criticism of training the unemployed has focussed on the (i) type of training and (ii) the provinces where it is offered. A frequent observation is that training is often concentrated on occupations or in geographical areas where there are few jobs. In its 1982 review of training, the Economic Council observed "Linking the Adult Occupatinal Training Act (AOTA) to equity considerations through its function of 'soaking up unemployment' does impede the efficiency of the program.[1] Not only has the distribution of training been inversely related, to some extent, to regional patterns of economic activity ..., but the strong orientation of AOTA towards alleviating unemployment does not facilitate the development of those skills which are in greatest demand. Rather than developing people with productive and desirable skills, much of the activity under AOTA has seemingly served merely as a temporary palliative for unemployment" (Economic Council:1982, p. 86).

Employment and Immigration acknowledged these criticisms in their Task Report (E&IC:1981b) and developed the National Training Act (NTA) as a move toward confronting them. NTA continues to embrace the premise that training has a strong role to play in combatting unemployment. It is in the context of NTA programs, initiated between 1982-85, that results of this study are relevant.

## Objectives

The major objective of this chapter is to estimate probabilities of taking training among long-term unemployed Canadians who differ in terms of age, sex, marital status, education, occupation, etc. Why is this important?

Segments of the long-term unemployed population are confronted with poor employment prospects because of declining job opportunities in some industries (eg. textiles, clothing, furniture, construction, primary resource industries, etc.) and in some occupations (eg. processing occupations, fabricating and assembling occupations, construction trades, some less skilled service occupations, etc.). Poor employment prospects are related to changes in the industrial and occupational structure of the economy (an on-going process) which may have accelerated since the 81-82 recession (Picot:1986). Personal characteristics are also associated with difficulty in labour adjustment, notably lower levels of education (in part because severe employment problems are in less skilled occupations) and older age.

However, there are various reasons why many of the long-term unemployed who face the toughest labour market conditions are unlikely to turn to training to help locate new employment in new occupations or regions. The less educated, for example, are a case in point. Those with below average education in their youth may be unlikely to seek out training as adults. This may be related to their family background, the degree of difficulty they have in education or training programs, the norms of the social group in which they were raised, their beliefs about benefits of education (or lack of), or various other reasons. Furthermore, many of these people work in occupations and industries where there is no history of training. Thus, the habit of training does not develop among these workers and industries. Whatever the reason (having cut short their education in their youth or being unfamiliar with education and training systems), such works may be unlikely candidates for training. This is particularly true of training offered in an institutional context. Older persons experiencing long-term unemployment are also unlikely to turn to training to assist their adjustment process, because many have been away from such activities for many years. Also, the personal benefits of training decrease with age, simply because the pay-back is shortened.

## Findings

Findings of this paper largely substantiate the viewpoint above. Some of these are:

(1) In spite of the increased likelihood of training associated with unemployment in general, the probability of taking training is observed to decline with age (approaching a very low level over 40). It is also very low among the less educated (elementary or some secondary education) as compared with the more highly educated, and is relatively low among unemployed married women.

(2) When the long-term unemployed are classified in target groups according to age, education and marital status, it is observed that one-half the long-term unemployed in 1983 belonged to groups with very low training rates. Again, these tend to be less educated, older workers, and married women.

(3) The long-term unemployed in occupations with low demand for labour (i.e., the "toughest" labour markets) were less likely to take short-term training than their counterparts with similar characteristics (age, education, sex, etc.) in higher-demand occupations. Both groups, however, were equally likely to enter a full-time training program.

In sum, though unemployment in general definitely increases the probability of training, approximately one-half of the unemployed - the long-term unemployed who face the toughest labour market conditions among the unemployed - are in groups which tend not to turn to training as a means of adapting to changing labour market conditions.

## Methodology

Participation in two types of training programs is examined. The first is short term training offered by employers, colleges and universities, and organizations such as unions, professional associations and school boards. These programs offer courses ranging in duration from 10 to 400 hours. On average, each course consists of approximately 84 hours or the equivalent of three weeks of full-time attendance. Approximately 1.8 million people, or 13% of Canada's labour force aged 18 to 64 participated in some form of short-term training during 1983.

The second type of training is exclusively aimed at adults (18 years and over) enrolled in a full-time program at a school, college or university.[2] The data show that approximately 250,000 adults entered a full-time program in a school, college or university during 1983, 70% of them in September.[3]

We estimate the probability that people with particular combinations of socio-economic characteristics will take short-term training or enter long-term training during the year (here, 1983).[4] Note that entry rather than participation rates or probabilities are used for the full-time training. Multivariate techniques are used because they control for a number of variables while allowing the analyst to determine the influence of a particular variable (say, education) on the participation rate. This is necessary because what appears to be a strong correlation between two variables may, in part, be attributable to the influence of another variable. For example, participation rates are much higher among people with university degrees than among those with an elementary education. But the university educated tend also to be much younger, and participation rates are higher among the young than the old. Thus, part of the large variation in participation rates between different educational groups would be due to the effect of age. By examining the relationship between the probability of training and several independent variables (characteristics) simultaneously, multivariate techniques can more accurately assess the effect of particular characteristics.

Specifically we employ a technique called logit regression, details on which are provided in Methodological Appendix B. Two comments are in order about the "probability" of training as estimated by logit regression. First, it is the probability of training during one year that is examined here, not the probability of ever training, or of training over, say a five-year period. Clearly, the probabilities of training over longer durations would be much higher. Second, for any particular group (e.g., 35-39-year-old men working full-time in managerial jobs), the estimated probability is to be construed as an "average". That is, within each group, probabilities for particular individuals will vary considerably.

## Participation in Training

According to Table 1, those unemployed for more than six months participated only half as much in short-term training as did those who were employed full-time (ie., 9% versus 16.7%). Those unemployed less than six months were also more likely to be enrolled in training than the long-term unemployed.

The impressions above are crude, of course, because they do not control for other important influences such as age, sex, education and province of residence. Much of the difference in participation rates may be due not to unemployment but to other characteristics which reduce their participation in training such as low educational backgrounds. To capture, or rather to control, the effects of such influences more precisely, estimates of the impact of unemployment on probabilities of taking training are presented in Table 2. These probabilities have been derived specific to worker groups, each of which differ in terms of selected characteristics. These groups were selected for no other reason than to indicate what the regression model estimates when groups with the same levels of the control variables such as age, sex, are compared.

In Table 2, Group 1 pertains to workers in the sample who - as indicated - are males, aged 25-34, with a university degree, in a managerial occupation and residents of BC/Alberta. They have been disaggregated in Table 2 into three categories according to labour force status. Probabilities reported for each category derive from logit regressions which convey that the probabi-

**TABLE 1:** Participation Rates in Short-Term Training, by Labour Force Status, 1983

| Labour Force Status | Participation Rates | Number Training | Percent Distribution of Persons Training |
|---|---|---|---|
| | % | (thousands) | % |
| Employed Full-Time/Full-Year | 16.7 | 1,105.2 | 61.2 |
| Unemployed Less Than Six Months | 11.3 | 275.2 | 15.2 |
| Unemployed More Than Six Months | 9.0 | 88.2 | 4.9 |
| Employed Part-Time or Part-Year (no unemployment) | 10.5 | 287.4 | 15.9 |
| Sub-Total: Labour Force | 13.7 | 1,756.0 | 97.3 |
| Not in Labour Force | 1.8 | 48.6 | 2.7 |
| Total Population | 11.6 | 1,804.1 | 100.0 |

**TABLE 2:** Estimated Probability of Taking Short-Term Training, by Labour Force Status, Selected Groups

- Output from Logistic Regression -

| | Control Variables | Labour Force Status | | |
|---|---|---|---|---|
| | | Unemployed < 6 Months | Unemployed > 6 Months | Employed Full-Time |
| | | (1) | (2) | (3) |
| | Group 1 | | | |
| Sex: | Male | | | |
| Age: | 25-34 | | | |
| Education: | University | 0.352 | 0.366 | 0.426 |
| Occupation: | Managerial | | | |
| Region: | Alta/B.C. | | | |
| | Group 2 | | | |
| Sex: | Female | | | |
| Age: | 35-44 | | | |
| Education: | Some Postsec. | 0.193 | 0.203 | 0.246 |
| Occupation: | Prof/Tech | | | |
| Region: | Quebec | | | |
| | Group 3 | | | |
| Sex: | Female | | | |
| Age: | 35-44 | | | |
| Education: | Completed Sec. | 0.109 | 0.116 | 0.144 |
| Occupation: | Clerical/Sales | | | |
| Region: | Ontario | | | |
| | Group 4 | | | |
| Sex: | Male | | | |
| Age: | 45-54 | | | |
| Education: | Some Sec. | 0.037 | 0.039 | 0.049 |
| Occupation: | Blue Collar | | | |
| Region: | Atlantic | | | |

lity of taking training is .366 if the individual has been unemployed more than six months versus .426 if employed full-time. That is, the probability of taking training if unemployed long-term is approximately 86% that of full-time employees - not one-half as implied by Table 1. The same impression emerges when we compare probabilities of taking training among the remaining group (column 1). On balance, however, those who would seem to need training most are not as well represented in short-term courses as are those employed full-time. (We will examine this question in fuller detail later.)

Turning to full time training, cross-tabulated data in Table 3, column 1 reveals little difference in entry rates of unemployed and employed 18-20 year-olds.[5] For those aged 21-44 years, however, the incidence of entry into training among unemployed individuals is more than double that of employed workers[6].

The impressions above are not altered when we control for additional influences in logit regression.[7] Estimates of the probabilities for four groups, Profiles A, B, C and D, are shown in Table 4. These groups have been selected to be representative of those with high, medium and low probabilities. In general, the probability of an unemployed person (21 years or older) entering a full-time program is approximately 3 times that of an employed person. Since full-time training is likely to be seen as a more effective means of assisting the labour market process of unemployed workers, these results are encouraging. For the 18-20 age group, however, there is little difference in the probability of entering a full-time program among different labour force categories.

## Characteristics of Full-Time Trainees

Perhaps the most important issue concerns not differences between unemployed and employed workers in their participation in training, but differences among the unemployed themselves who take training. Do all unemployed, regardless of personal characteristics, exhibit similar participation rates in training? To evaluate this question, we employ multivariate analysis to evaluate characteristics of the long-term unemployed population only.

TABLE 3: Entry Rates in Full-Time Training of Persons Aged 18-20 and 21-44, by Labour Force Status Prior to Entry, 1982 and 1983*

| Labour Force Status Prior to Entry | Age 18-20 | | | Age 21-44 | | | |
|---|---|---|---|---|---|---|---|
| | Entry Rates 1982 & 1983 Combined* | Number Training | Percent Distribution | Entry Rates 1982 | Entry Rates 1983 | Number Training 1983 | Percent Distribution 1983 |
| | (1) % | (2) (thousands) | (3) % | (4) % | (5) % | (6) (thousands) | % |
| Employed | 6.7 | 60.9 | 50.4 | 1.4 | 1.3 | 84.7 | 51.8 |
| Not in the Labour Force | 7.2 | 23.9 | 19.7 | 1.8 | 1.8 | 29.9 | 18.3 |
| Unemployed | 6.4 | 16.5 | 13.6 | 3.6 | 4.2 | 26.6 | 16.2 |
| No Predominant Status | 7.8 | 19.5 | 16.1 | 2.9 | 3.8 | 22.4 | 13.7 |
| All Levels | 6.9 | 120.9 | 100 | 1.7 | 1.8 | 163.5 | 100 |

Note : * Entry rates represent the proportion of the population (excluding full-time students) who started a full-time program during the year specified. It is necessary to pool 1982 and 1983 data to obtain reliable estimates in some cells.

**TABLE 4:** Logistic Regressions for Long-Term Training

Probabilities of Entering Full-time Training

| Control Variables | Labour Force Status: Not in Labour Force | Labour Force Status: Unemployed | Labour Force Status: Employed |
|---|---|---|---|
| | (1) | (2) | (3) |
| Profile 'A'<br>Sex: Male<br>Age: 18-20<br>Education: Some Post-sec.<br>Region: Alta/B.C. | .116 | .163 | .126 |
| Profile 'B'<br>Sex: Female<br>Age: 21-24<br>Education: Some Post-sec.<br>Region: Alta/B.C. | .066 | .095 | .031 |
| Profile 'C'<br>Sex: Female<br>Age: 25-29<br>Education: Completed Sec.<br>Region: Alta/B.C. | .033 | .048 | .015 |
| Profile 'D'<br>Sex: Male<br>Age: 35-39<br>Education: Some Sec.<br>Region: Quebec | .006 | .009 | .003 |

"Explanatory" variables include age, sex, education, marital status, occupation and region of residence. Pooled data from both 1982 and 1983 were used in this regression, in order to render a sample sufficiently large to reliably estimate the model coefficients. Major findings are presented in Figures 1 and 2 and regression results are presented in the Appendix Table 3. They convey the following:

Age: The probability of the unemployed entering a full-time program falls with age. Once over 40, the probability is quite low - under 3.0%. This applies to most unemployed workers, regardless of their education, occupation, sex, marital status or province of residence.[8]

Education: The unemployed who had completed postsecondary education were much more likely to start training than their counterparts with less education. They were four times as likely to start training as the elementary educated who had similar characteristics, and approximately twice as likely as those with some secondary education. These differences are shown in Figure 3, where the probability of starting training is shown by educational attainment for selected groups.

Marital Status: Besides age and education, marital status significantly influenced the probability of unemployed women starting full-time training. Among adult men, marital status was not significant; unemployed married and unmarried men were equally likely to train. But among women, the difference was surprisingly large. Unemployed women who were unmarried were almost three times as likely to embark on full-time training as their married counterparts of similar age, education, and occupation. And unemployed, unmarried women had a slightly higher probability of entering full-time training than men (married or unmarried) with similar characteristics. For example, the estimates from the regression model indicate that the "average" probabilities of starting training for 25-39 year-olds with a postsecondary education were as follows: for married women 3.2%, unmarried women 8.8%, and men (married or unmarried) 7.2% (see Figure 1).

FIGURE 1: ESTIMATED PROBABILITY OF ADULT LONG-TERM UNEMPLOYED ENTERING FULL-TIME TRAINING, BY AGE, SEX, MARITAL STATUS AND EDUCATIONAL ATTAINMENT, 1983

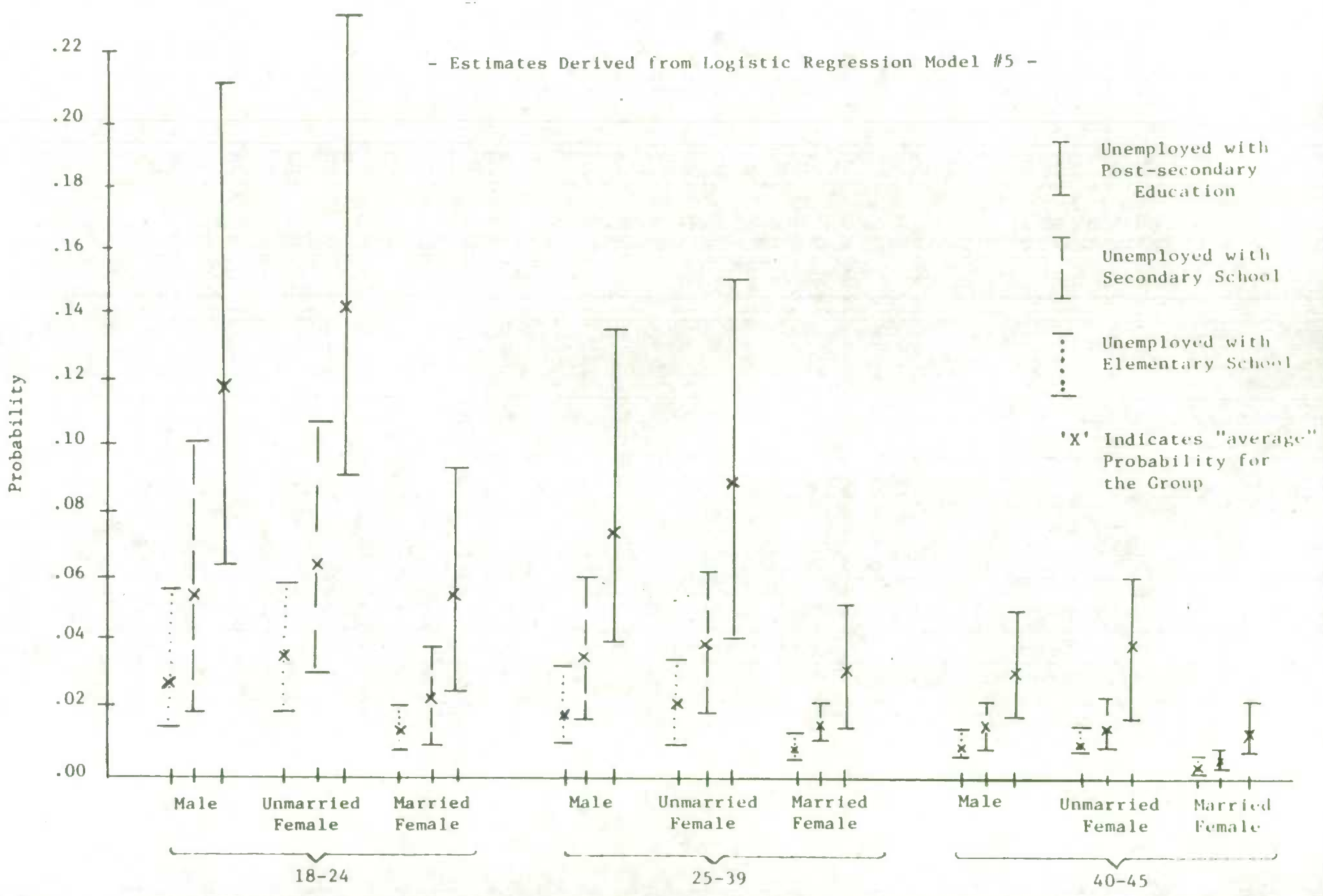
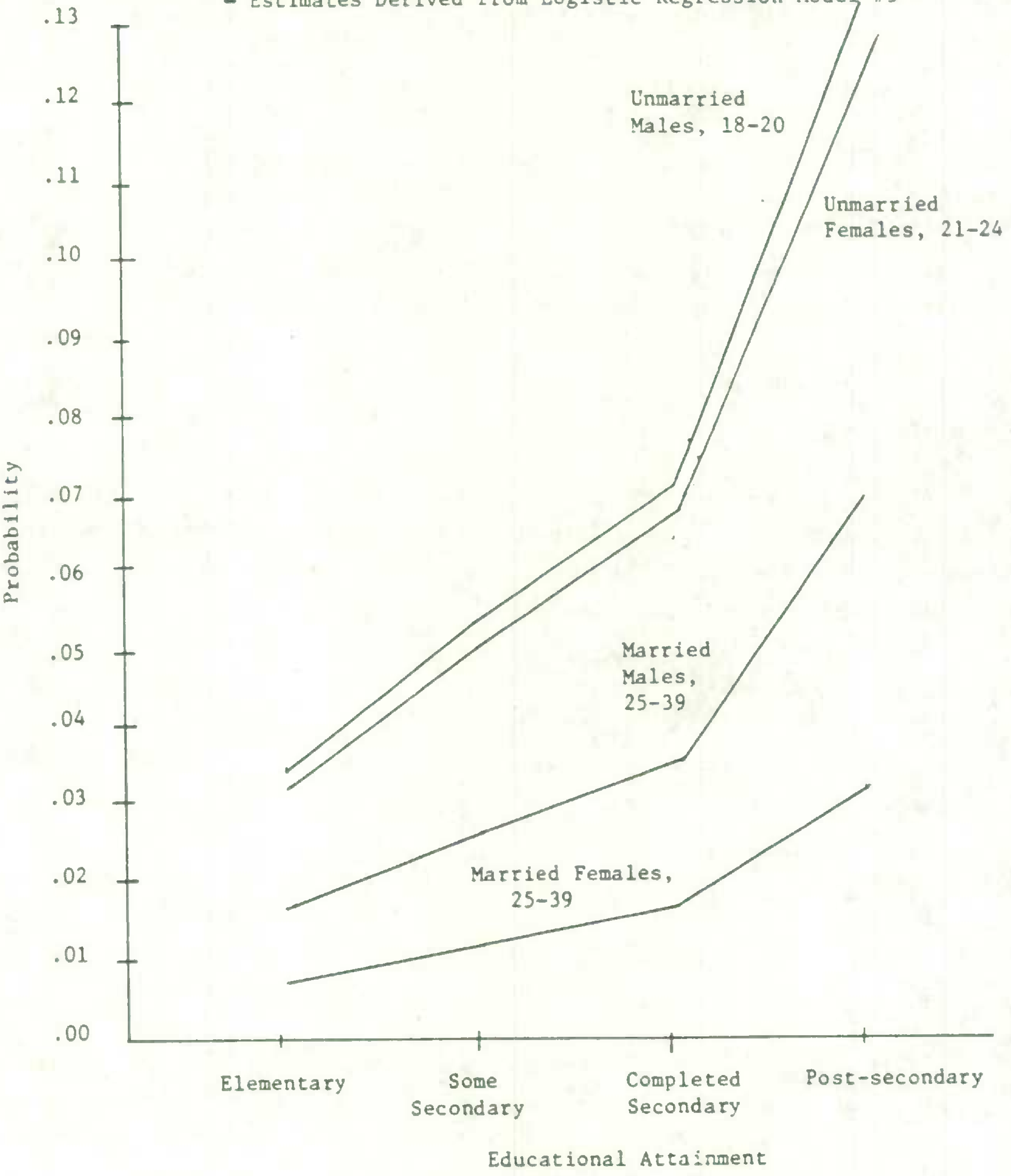


FIGURE 2: ESTIMATED PROBABILITY OF ENTERING FULL-TIME TRAINING AMONG THE LONG-TERM UNEMPLOYED, BY EDUCATIONAL ATTAINMENT, SELECTED GROUPS, 1983

- Estimates Derived from Logistic Regression Model #5 -



The higher probability of training among unmarried, unemployed women may be related to the fact that the financial burden of unemployment depends on family status. It was concluded in Shaw (1984)[9] that:

> "While unemployment contributes to "financial hardship", the relationship is not direct. Family status is a powerful mediating variable. All else held constant, the impact of unemployment on financial hardship is greatest among single-parent family heads, unattached individuals and husbands in husband/wife families. It is far less among wives, adult children, and other relatives in husband/wife families."

Hence, for unmarried women who are divorced, separated, widowed, or single, unemployment is much more likely to lead to financial hardship than for married women. Individuals in the latter groups are usually in families with other income earners. This difference may lead to a greater motivation and need for unmarried women to seek retraining in an effort to locate new employment.

It is also possible that family responsibilities decrease the opportunity for married women to take full-time training. The fact that a great many married women work part-time might also be part of the explanation. The motivation to devote considerable time, energy and money to full-time training may not be great when the rewards are diminished because one is seeking only part-time employment.

Remaining variables had minor effects on the probability of the unemployed entering full-time training. They include:

Region: Generally, the probability of the unemployed starting training in the western region (Manitoba/Saskatchewan/Alberta/British Columbia), was approximately 1.5 times that of the unemployed in Quebec, which had the lowest probabilities when groups with similar characteristics are compared.

Occupation: The unemployed whose last occupation was in the clerical or services area displayed the highest probability of starting a full-time program. When groups with similar characteristics were compared, the probabilities for these two groups were approximately 1.7 times that of the unemployed in sales occupations, where the unemployed had the lowest likelihood of training full-time. The unemployed in management/professional/technical and blue collar occupations (mining, processing, construction, etc.) displayed probabilities which were in a sense intermediate. As with the observed regional differences, variation among the five occupational groups i probabilities of starting full-time training were not statistically significant.

Why is the probability of training highest among the clerical and services occupations. It is possible that young people employed in clerical and service occupations on a temporary basis returned to a full-time program to develop an alternative career path following some time in the labour force. Also, a substantial share of full-time training is government-sponsored, implying financial training assistance of various kinds.[10]

## Long-Term Unemployed Not Likely to Train Full-Time

Thinking in terms of target groups, it is useful to identify whether particular groups, when confronted with unemployment, do or do not turn to full-time training. Moreover, it is important to know how large each group is. How many of the long-term unemployed belong to groups with low training probabilities? If the numbers are very small, then it really matters much less that their training probabilities are low. However, if a significant number belong to groups with low training probabilities, then the usefulness of training as a potential solution to employment problems could be questioned.

Who are the long-term unemployed? It has been established by Shaw (1984) that they (i) tend to have lower than average levels of education, (ii) live east of the Ontario-Quebec border, (iii) work in the less skilled occupations, and (iv) are younger than the labour force as a whole. The first three factors - less education, residing in the east, and low-skilled occupations - are all associated with a low probability of training. Therefore, many of the characteristics that are apt to increase the likelihood of being unemployed also tend to decrease the probability of training. In addition, one-quarter of the long-term unemployed are over 40 years of age, when significant re-training tends to be rare, at least given current participation patterns.

How many of the long-term unemployed are in groups with characteristics that suggest they would be unlikely to turn to training when confronted with employment problems? To assess this question, Tables 5 and 6 classify the long-term unemployed by age, education, sex and marital status, (ie., those variables which displayed the most influence on the probabilities in the regression model).[11] Calculated for each group are: (1) the probability of starting full-time training when unemployed long-term (Table 5) and (2) the proportion of the long-term unemployed in each group (Table 6). Again, we calculate probabilities from logit regression towards determining the proportion of the long-term unemployed who belong to groups with a low probability of taking full-time training.[12]

The groups we are interested in have the following characteristics:

- all unemployed over 40, regardless of other characteristics;
- males aged 25-39 with less than high school graduation;
- married women aged 25-39 with high school graduation or less, and unmarried women age 25-39 with elementary education;
- young married women (18-24) who do not have a postsecondary education.

Almost without exception, probabilities for workers with these characteristics are low in Table 5. Of course, this also means that one-half of the unemployed belong to groups which tend to have relatively high training rates, as shown in Table 6. Those groups are:

- most 18-24-year-olds (except married women without postsecondary education);
- 25-39-year-old males with at least high school graduation
- most 25-39-year-old unmarried women (except those with elementary education)
- 25-39-year-old married women with a postsecondary education.

**TABLE 5:** Estimated Probability of the Long-Term Unemployed Entering Full-Time Training, by Age, Sex, Marital Status and Education

Output from regression model

| | 18-24 | 25-39 | 40 and over | |
|---|---|---|---|---|
| MALES | | | | |
| Elementary | $\boxed{.040}$ | $\boxed{.017}$ | $\boxed{..}$ | (.008) |
| Some Secondary | .046 | $\boxed{.027}$ | $\boxed{..}$ | (.012) |
| Completed Secondary | .061 | .037 | $\boxed{..}$ | (.016) |
| Postsecondary | .117 | .072 | $\boxed{..}$ | (.032) |
| UNMARRIED FEMALES | | | | |
| Elementary | .036 | $\boxed{.021}$ | $\boxed{..}$ | (.009) |
| Some Secondary | .057 | .034 | $\boxed{..}$ | (.015) |
| Completed Secondary | .075 | .045 | $\boxed{..}$ | (.019) |
| Postsecondary | .141 | .088 | $\boxed{..}$ | (.039) |
| MARRIED FEMALES | | | | |
| Elementary | $\boxed{.012}$ | $\boxed{.007}$ | $\boxed{..}$ | (.003) |
| Some Secondary | $\boxed{.020}$ | $\boxed{.012}$ | $\boxed{..}$ | (.005) |
| Completed Secondary | $\boxed{.027}$ | $\boxed{.016}$ | $\boxed{..}$ | (.007) |
| Postsecondary | .053 | .032 | $\boxed{..}$ | (.014) |

.. No estimate for this group as a whole, but certainly all fall below .03.
() Indicates probability for 40-45 age group.

**TABLE 6:** Percent of the Long-Term Unemployed in Each Group, by Age, Sex, Marital Status and Education

| | 18-24 | 25-39 | 40 and over |
|---|---|---|---|
| | % | % | % |
| MALES | | | |
| Elementary | 2.1 | $\boxed{3.5}$ | $\boxed{6.3}$ |
| Some Secondary | 7.7 | $\boxed{7.6}$ | $\boxed{4.1}$ |
| Completed Secondary | 7.8 | 7.3 | $\boxed{3.8}$ |
| Postsecondary | 4.0 | 6.8 | $\boxed{2.9}$ |
| UNMARRIED FEMALES | | | |
| Elementary | 0.5 | $\boxed{0.3}$ | $\boxed{1.1}$ |
| Some Secondary | 2.5 | 0.9 | $\boxed{0.8}$ |
| Completed Secondary | 3.7 | 1.2 | $\boxed{0.5}$ |
| Postsecondary | 1.4 | 1.6 | $\boxed{0.7}$ |
| MARRIED FEMALES | | | |
| Elementary | $\boxed{0.4}$ | $\boxed{1.0}$ | $\boxed{2.4}$ |
| Some Secondary | $\boxed{1.0}$ | $\boxed{2.5}$ | $\boxed{1.7}$ |
| Completed Secondary | $\boxed{1.8}$ | $\boxed{3.8}$ | $\boxed{1.9}$ |
| Postsecondary | 1.0 | 2.8 | $\boxed{1.0}$ |
| | | TOTAL (100%) is 985.1 thousand | |

**Note:** 49.1% of the long-term unemployed are in groups (as defined by age, sex, marital status, education) where the probability of entering a full-time program is less than 3% (those indicated by the boxes in Table 5).

For the long-term unemployed in the groups with low training rates -- fully half the long-term unemployed -- full-time training initiatives can hardly be considered a major avenue to more stable employment, simply because so few of them actually enroll. Full-time training offered in schools, colleges or universities may help specific individuals, but for the target group as a whole, it cannot make a major contribution toward solving their employment problems, unless patterns of participation were to change. This says nothing, of course, about the likelihood that various groups would succeed in training even if they entered a program. Such ssues are beyond the scope of this study.

## Long-Term Unemployed in Low-Demand Occupations

Another important aspect of training as a potential solution to long-term unemployment is the sector in which the individuals are seeking jobs. Some industries and occupations experienced very slow employment growth, if not declines, between 1975-85. In addition to recession, this may be caused by changing patterns of trade (imports and exports) and domestic demand or technological and productivity changes. Consequently, some industries and occupations have suffered chronic unemployment as the supply of labour has exceeded demand. It is often argued that the unemployed (particularly the long-term unemployed) in these occupations and industries should acquire new skills and train for employment in economically healthier occupations and/or industries (e.g., Saunders, 1984). A large labour surplus (as indicated by high unemployment) suggests a difficult milieu in which to locate work. Training, it is argued, can help workers move to a more hospitable environment.

But do the long-term unemployed in low-demand occupations and industries seek training? To determine this, let us first identify these occupations and industries in terms of aggregate unemployment rates. Low-demand occupations and industries are defined as those with above-average unemployment rates.[13] Occupations in this category include processing; product fabricating, assembling and repairing; construction; and some service occupations. Affected industries were in manufacturing, consumer services (food and accommodation, amusement and recreation, etc.), and primary sectors (Table 7).

The data indicate that the long-term unemployed from occupations with the most serious employment problems were less likely to take short-term training (a 6% participation rate) than other long-term unemployed (12%) (Table 8), and much less likely than the full-time employed (17%).

Educational attainment is responsible for much of this difference in participation. As noted previously, the long-term unemployed generally have lower levels of education than the labour force in general and the long-term unemployed in low-demand occupations tend to have even less (58% did not have high school graduation). And as demonstrated earlier, people who did not graduate from high school had low probabilities of taking short-term training, no matter what their other characteristics.

In contrast, the probability of the long-term unemployed in low-demand occupations embarking on a full-time program was not significantly different from that of their counterparts in other occupations. This applies despite variations in educational attainment. It may be attributable, in part, to full-time training offered under the National Training Act, much of which is oriented towards the less educated who have been unemployed in these industries and occupations.

**TABLE 7:** Industries and Occupations with Above Average Unemployment Rates

| Occupations with Above Average Labour Surpluses (higher than average unemployment rates) | | Industries with Above Average Labour Surpluses (higher than average unemployment rates) | |
|---|---|---|---|
| CCDO | Name | SIC | Name |
| 415 | Material recording & scheduling | 031-049 | Forestry, fishing and trapping |
| 612 | Food & beverage preparation | 051-099 | Mining* (except mineral, fuels) |
| 613 | Occns. in lodging & accommodation | 101-109 | Food & beverage Ind. |
| 619 | 'Other' service occns. | 172-179 | Leather Ind. |
| 718/719 | 'Other' farming occns. | 181-189 | Textile Ind. |
| 731 to 771 | Fishing, trapping and related occns. | 231-239 | Knitting Mills |
| 821/822 | Food & beverage processing | 243-249 | Clothing Ind. |
| 823 | Wood processing occns. | 251-259 | Wood Ind. |
| 825 | Pulp & paper making occns. | 262-268 | Furniture & Fixtures |
| 826/827 | Textile processing occns. | 301-309 | Metal Fabricating Ind. |
| 829 | 'Other' processing occns. | 391-399 | Misc. Manufacturing |
| 831 to 839 | Machinery & related occns. | 404-421 | Contractors (construction) |
| 851 to 859 | Product Fabricating, Assembling & Repairing (except electrical assembling & repair (853) and mechanics and related (858)) | 841-849 | Amusement & Recreational Services |
| | | 881-886 | Accommodation & Food Services |
| | | 891-899 | Misc. Services |
| 871 to 879 | Construction Trades (except electrical trades - 873) | | |
| 971 | Motor Transportation operating occns. | | |
| 931 | 'Other' Material handling occns. | | |

* Higher than average unemployment rates in 1982 and 1983 only. All other industries had higher than average rates in every year since 1975 (when data first available).

**TABLE 8:** Participation Rates in Short-Term Training of People Unemployed More than Six Months, by Type of Industry and Occupation, 1983

| Industry/ Occupation | Participation Rates | Number Training |
|---|---|---|
| | % | (thousands) |
| Industry | | |
| High Unemployment Rate | 7.7 | 29.5 |
| Low Unemployment Rate | 10.0 | 55.2 |
| Occupation | | |
| High Unemployment Rate | 6.3 | 32.7 |
| Low Unemployment Rate | 12.4 | 52.0 |
| Total Unemployed More than Six Months | 9.0 | 84.7 |

## Summary

Generally speaking, unemployed adults are more likely to enter a full-time education or training program, than are employed persons of similar characteristics. But this does not imply that all types of unemployed individuals have a high probability of training. In fact, about one-half of the *long-term unemployed* belong to groups (defined by age, sex, education, marital status, occupation) which have very *low* probabilities of entering full-time training. These groups are heavily predominated by the less-educated, older persons, and married women. Furthermore, the unemployed in labour surplus occupations (i.e., high unemployment rate) are much less likely to have taken training than are persons in occupations with lower unemployment rates. Persons in both occupational groups are equally likely, however, to enter a full-time training program.

We conclude that training programs do not appear to be serving as a major source of labour market adjustment for all groups of the unemployed. For some, training is presently a feasible approach to assisting the labour market adjustment process. But for others, the simple fact that participation rates are so low renders training relatively ineffective for the group as a whole. Training may help individual members of target groups to locate new employment, but for these groups *as a whole*, it is likely to be relatively ineffective, simply because participation rates are so low.

## APPENDIX 1: Notes on Sample Design

The data used in this analysis were collected through the Labour Force Survey. The respondents from this survey do not form a simple random sample from the Canadian population. Rather the survey has a complex design, including stratification and clustering (multiple stages of selection), with unequal probabilities of selection for the respondents. These unequal selection probabilities mean that some areas (those where the selection probabilities are high) are overrepresented in the sample, relative to their representation in the population, while others are underrepresented.

Using data from such complex surveys presents problems for the analyst, both in ensuring that the over- and underrepresented is adjusted for in estimation procedures and in incorporating the effect of survey design in "variance calculations". These problems are especially difficult for the analyst because most standard statistical procedures are not suitable in the complex survey framework.

### Design Effects

The variance of a statistic estimated from a complex survey will be different from the variance of the same statistic estimated from a simple random sample of the same size. In general, stratification decreases the variance, while clustering often increases the variance. It has been observed that for many variables measured by the Labour Force Survey, the variance is larger than would have been obtained had a simple random sample of the same size been taken.

Most standard statistical tests use variances that are calculated assuming that the data are from a simple random sample. These variances tend to be smaller than those observed in complex surveys when accounting for the survey design effect. Hence the 't' or 'F' statistics reported by the standard statistical procedures are overestimates of the true values.

The variances on which the tests in this report are based were calculated using customized software developed in the Institutional and Agricultural Survey Methods Division of Statistics Canada which account for the survey design.

## The Use of Sample Weights

The Labour Force Survey employs a highly stratified design, with significant differences in sampling fractions (proportion of the population surveyed) among strata. This is necessary to ensure that reliable statistics will be available for provincial and certain sub-provincial levels of aggregation. For example, the sampling fraction is much higher in Prince Edward Island than in Ontario, since for a given desired precision, the proportion of the population that must be surveyed increases as the population decreases.

Since some regions are overrepresented and some underrepresented in the Labour Force Survey, the unweighted sample is not representative of the Canadian population. This disproportionate representation can be accounted for through the use of the sample weights, but any statistics or estimates based on the unweighted data will underrepresent those areas with small sampling fractions (the large provinces for instance) and overrepresent those with large sampling fractions (the small provinces). Thus, if the unweighted data are used in an analysis the results cannot be interpreted as applying to the Canadian population, but rather will apply to some conceptual population. Under some circumstances, with large samples and certain models, the difference between a weighted and an unweighted analysis may be insignificant, but in general a weighted analysis is to be preferred.

But there may be problems in using weighted data as well. In most standard statistical packages the meaning or definition of the weight differs from that used in a sample survey. The result is that while estimates are often correct, the variances calculated are almost meaningless and not useful. Hence the tests of significance from these packages are meaningless when weighted data are used. The calculation of design based variances using custom software for this report avoided the problems of using the standard packages.

For theoretical details of the approach used, see Binder (1983). Further discussion of the problems encountered in applying multivariate analysis to complex survey data can be found in Kish and Frankel (1974).

## APPENDIX 2: Logit Regression Tables on Training

**APPENDIX TABLE 1:** Logistic Regression Results for <u>Short-Term Training</u>

$$\text{Model: } \ln \frac{P(Y=1)}{1-P(Y=1)} = BX$$

where Y = 1 if individual took short-term training
0 if individual did not take training

| Reference Group | Variable | Beta | 't' |
|---|---|---|---|
| - | Intercept | -.444 | -7.26 |
| Male | SEX: Female | -.036 | -0.92 |
| | AGE: 18-20 | -.439 | -5.18 |
| 25-34 | 21-24 | -.071 | -1.18 |
| | 35-44 | -.094 | -2.25 |
| | 45-54 | -.482 | -7.75 |
| | 55-64 | -.876 | -10.71 |
| | EDUCATION: | | |
| | Elementary | -1.852 | -18.62 |
| Univ. Degree | Some Sec. | -1.090 | -13.39 |
| | Compl. Sec. | -.692 | -11.97 |
| | Some Postsec. | -.187 | -2.85 |
| | College Dips./ Cert. | -.100 | -1.77 |
| | EMPLOYMENT STATUS | | |
| | Unempl. Six Months | -.249 | -5.18 |
| Employed Full-Time/ Full-Year | Unempl. Six Months | -.313 | -3.19 |
| | Empl. Part-Time or Part-Year | -.469 | -8.66 |
| | REGION | | |
| | Atlantic | -.403 | -7.55 |
| Alta./B.C. | Quebec | -.356 | -6.41 |
| | Ontario | -.156 | -3.04 |
| | Man./Sask. | -.036 | -.64 |
| | OCCUPATION | | |
| | Managerial | .146 | 2.56 |
| Prof./Tech. | Clerical/Sales | -.362 | -7.09 |
| | Blue Collar | -.548 | -7.79 |
| | Not Worked in 5 Years | -.537 | -10.16 |
| | | -.478 | -4.07 |

**APPENDIX TABLE 2:** Logistic Regression Results for Full-Time Training

Model: $\ln \dfrac{P(Y=1)}{1-P(Y=1)} = BX$

where Y = 1 if individual entered a full-time program
0 if individual did not

| Reference Group | Variable | Beta | 't' value |
|---|---|---|---|
| - | Intercept | -1.417 | -6.36 |
| Male | SEX: Female | -0.275 | -2.72 |
| 18-20 | AGE: 21-24 | -.343 | -2.66 |
| | 25-29 | -1.028 | -7.45 |
| | 30-34 | -1.307 | -6.49 |
| | 35-39 | -1.774 | -8.38 |
| | 40-45 | -2.171 | -9.00 |
| | EDUCATION: | | |
| | Elementary | -1.420 | -5.28 |
| Univ. Degree | Some Sec. | -0.942 | -4.43 |
| | Compl. Sec. | -0.264 | -1.61 |
| | College or some postsec. | -0.218 | -0.87 |
| | LABOUR FORCE STATUS | | |
| | Employed | -1.180 | -7.00 |
| Employed Prior to Entry | Not in Labour Force | -0.391 | -2.28 |
| | No Predominant Status | -0.211 | -1.216 |
| | REGION | | |
| Alta./B.C. | Atlantic | -0.294 | -2.34 |
| | Quebec | -0.538 | -4.39 |
| | Ontario | -0.314 | -2.59 |
| | Man./Sask. | -0.088 | -0.65 |
| | Interaction Variable Employed and Age 18-20 | 0.879 | 5.68 |
| Occupations With Above Average Unemployment Rates | Occupations With Below* Average Unemployment Rate | -.035 | -0.35 |
| Population | | Total population (excluding students) 18 to 45 | |
| Sample Size | | 42,700 | |
| $\chi^2$ | | 430 | |
| $\rho^2$ | | .10 | |

* This variable used only with those unemployed prior to entry. It has a value 1 if individual was unemployed prior to entry and was in a low-unemployment occupation; it was zero otherwise.

**APPENDIX TABLE 3:** Logistic Regression Results for Full-Time Education and Training, Taking by the Long-Term Unemployed

Model: $\ln \dfrac{P(Y=1)}{1-P(Y=1)} = BX$

where Y = 1 if long-term unemployed person entered a full-time program during 1982 or 1983
0 otherwise

| Reference Group | Variable | Model 4 Beta | Model 4 't' value | Model 5 Beta | Model 5 't' value |
|---|---|---|---|---|---|
| | Intercept | -1.977 | -6.88 | -1.872 | -9.23 |
| Male | Female | +.102 | 0.42 | +1.66 | 0.83 |
| | AGE: 21-24 | -.214 | -1.01 | | |
| | 25-29 | -.574 | -2.15 | | |
| | 30-34 | -.698 | -2.28 | | |
| | 35-39 | -.547 | -1.51 | | |
| | 40-45 | -1.452 | -3.01 | | |
| | AGE: 21-24 | | | -.220 | -1.05 |
| | 25-39 | | | -.631 | -2.55 |
| | 40-45 | | | -1.480 | -3.07 |
| | EDUCATION: | | | | |
| | Elementary | -1.536 | -4.71 | -1.486 | -4.77 |
| Some or Completed postsecondary | Some Secondary | -1.072 | -5.22 | -1.011 | -5.11 |
| | Completed Secondary | -0.807 | -4.16 | -0.713 | -3.84 |
| | REGION: | | | | |
| | Atlantic | -0.116 | -0.61 | | |
| Alta/BC | Quebec | -0.385 | -1.82 | | |
| | Ontario | -0.047 | -0.22 | | |
| | Man/Sask | +.144 | +0.63 | | |
| | OCCUPATION: | | | | |
| | Clerical | +0.531 | +1.77 | | |
| Managerial/ Professional/ Technical | Sales | -0.119 | -0.33 | | |
| | Service | +0.492 | +1.62 | | |
| | Blue Collar | +0.291 | +1.14 | | |
| | INTERACTION VARS: | | | | |
| | Married and Female | -1.085 | -3.50 | -1.073 | -3.47 |
| | Married and Male | -0.109 | -0.48 | -0.096 | -0.43 |
| | Population | Persons unemployed for four months or longer during 1982 or 1983 | | Persons unemployed for four months or longer during 1982 or 1983 | |
| | Sample Size | 7200 | | 7200 | |
| | $\chi^2$ | 80 | | 70 | |
| | $\rho^2$ | .06 | | .06 | |
| | d.f. | 671 | | 671 | |

## APPENDIX 3: A Brief Description of Logistic Regression Analysis

Only a very brief discussion of this multivariate technique is provided here. It is intended to allow readers not familiar with the technique to acquire enough information to be able to interpret the results. None of the details of the estimation procedure and other technical aspects of the analysis are discussed. Those wishing a better understanding of the strengths and weaknesses of this technique are referred to Amemiya (1981), Jackson (1983), Hanushek and Jackson (1977) and Stopher and Meyburg (1979).

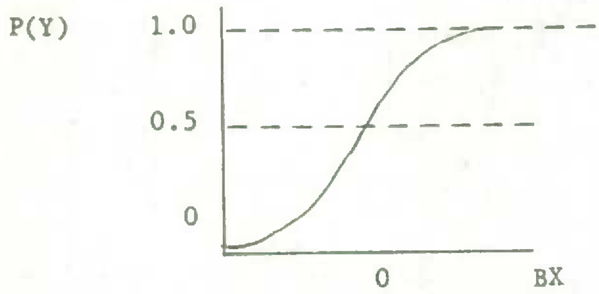
Logistic regression is appropriate where the dependent variable takes on only the values of 1 and 0, indicating the presence (or absence) of some event or condition. For example, in this case the dependent variable Y has the value 1 if the individual took training in 1983, and 0 if he/she did not. Using such data, one wants to estimate the likelihood or probability of taken training for particular values of the independent variables (i.e., Pr (Y=1)).

If ordinary least-squares - as opposed to logistic -regression- is used to estimate the probabilities (referred to as linear probability models), a number of problems are encountered. Likely the most serious is that the predicted values of the dependent variable are not necessarily bounded between 0 and 1; the model may predict probabilities above 1 or below 0. As well, the assumptions of homoscedasticity (constant variance of the error terms) and normality are violated. To overcome these problems, one can turn to logistic regression.

In the logistic model, the relationship between $P = Pr\{Y=1\}$ and the vector X of independent variables is assumed to be:

$$P = \frac{1}{1 + e^{-BX}}$$

where B is the vector of coefficients associated with the respective independent variables. In the model, the value of P can range from 0 to 1 as BX goes from $-\infty$ to $+\infty$. The curve is S shaped, and approaches the limits of 0 and 1 asymptotically, as shown below.




This results in one important characteristic of logistic models that must be kept in mind when interpreting the results. The change in the the probability of training associated with a fixed change in the value of an independent variable is not the same over all values of P. The relationship between BX and P is not linear, as in ordinary regression (ie., the slope of the curve is not constant over all possible ranges). Rather, the marginal change in P associated with a fixed change in an independent variable is larger in the midrange of the curve than it is closer to the limits of 0 or 1. This makes some sense intuitively since it suggests that it is difficult to increase a probability already near one, or decrease one already close to zero.

Hence, when interpreting the results, one must remember that the marginal change in P associated with a fixed change in any independent variable depends on the value of P itself. Thus, no single value can reflect the marginal effects of an independent variable on P. For this reason, in the text when the influence of an independent variable (say, age) on P is reported (with all other variables held constant), it is reported for a range of values of P. Profile groups with particular characteristics are selected such that P has high, medium and low values, and the relationship between P and the independent variable is illustrated over different ranges of P.

To convert the model to a form suitable for estimation of the coefficients using a regression technique, the following transformation is made.

Note that if $P = \dfrac{1}{1 + e^{-XB}}$ , then

$$1-P = e^{-XB}/[1 + e^{-XB}]$$

$$= \frac{1}{1 + e^{XB}}$$

Then let

$$L = \ln \left(\frac{P}{1-P}\right) = \ln(P) - \ln(1 - P)$$

$$= \ln(1 + e^{-xB}) - \ln(e^{-xB}) - \ln(1 + e^{-xB})$$

$$= \ln(e^{-XB})$$

$$= XB$$

<u>L is called the logit or the log of the odds ratio</u>, and this becomes the dependent variable in the regression model. As P goes from 0 to 1, L goes from - to + ; thus, while the probabilities are bounded, the logits are unbounded with respect to values of the independent variables. And while P is not a linear function of the independent variables, the logit is. Thus, the regression equation involves the logit and the independent variables. But it is the relationship between P and and the independent variables that is of interest, not that between $\ln \left(\frac{P}{1-P}\right)$ and the independent variables. When the the independent variables are <u>continuous</u>, the relationship between the change in P and a change in <u>one</u> of the independent variables can be determined by taking the partial derivative of P with respect to that variable (ie., determine the slope).

In doing this, it turns out that

$$\frac{\partial P}{\partial X_k} = \frac{\partial}{\partial X_k} [1/(1+e^{-XB})] = B_k P(1-P)$$

where the kth independent variable $X_k$, whose coefficient is $B_k$ has been selected. Thus, for continuous variables a change in P with respect to $X_k$ is simply the coefficient in the logit equation, $B_k$, times P(1-P). Thus, it can be seen that the rate of change of P w.r.t. $X_k$ depends upon the value of P itself, as indicated earlier.

However, when the independent variables are qualitative (categorical) as all are in this analysis, taking the first derivative has little meaning, since it represents the instantaneous change of P w.r.t. $X_k$ immediately at some given value of P. What is of interest with categorical independent variables is the extent to which P changes when one moves from one level to another of a qualitative independent variable (all other variables being held constant).

This is done in a very straight forward manner, and is described after the regression model itself is outlined.

## Estimation of the Regression Model

The model expresses the logit as a linear combination of the independent variables:

In our case, the independent variables are qualitative. For the sake of demonstration, assume there are two independent variables, sex, and educational attainment, the first having two levels, the second four. It is necessary to select a "reference" category for each variable, and this is omitted in the equation to prevent singularity. Suppose the variables are as follows: Sex: MALE, FEMALE, Education: ELEM, SEC, COLL, UNIV, with the reference categories selected as MALE and ELEM.

Then the model becomes:

$$\hat{L}_{ij} = \hat{B}_0 X_0 + \hat{B}_1 \text{ FEM} + \hat{B}_2 \text{ SEC} + \hat{B}_3 \text{ COLL} + \hat{B}_4 \text{ UNIV}$$

where i = the level of the sex variable (1=male, 2=female);

j = the level of the education variable (1=elem, 2=sec, 3=coll, 4=univ);

$Xo$ = 1 for all cases;

FEM = 1 if the case if female, = 0 if the case is male;

UNIV = 1 if the case has university education, = 0 otherwise;

COLL = 1 if the case has college as highest education, = 0 otherwise;

SEC = 1 if the case has secondary as highest, = 0 otherwise.

This is an additive model, since the effect on the logit value, $L_{ij}$, is simply the sum of the B's for any particular group. For example, the logit value for Females with secondary education is:

$$\hat{L}_{22} = \hat{B}_0 + \hat{B}_1 + \hat{B}_2$$

Here, the effect of any one variable does not depend on the values for the other explanatory variables. For example, the effect on the logit of moving from level 1 (elementary) to level 2 (secondary) in education is the same whether a person is male or female. This assumption is sometimes not supportable by the data, in which case interaction variables may be added to the equation, as has been done in some cases in the analysis in this report. For example, if the above assumption was not supportable, an interaction variable FEM·SEC could be added. The variable would have the 1 if the case were both female and had a secondary education, and 0, otherwise. The model becomes

$$\hat{L}_{ij} = \hat{B}_0 X_0 + \hat{B}_1 \text{ FEM} + \hat{B}_2 \text{ SEC} + \hat{B}_3 \text{ COLL} + \hat{B}_4 \text{ UNIV} + \hat{B}_5 \text{ FEMSEC}$$

Now,

$$\hat{L}_{22} = \hat{B}_o + \hat{B}_1 + \hat{B}_2 + \hat{B}_5 \qquad \text{(i.e. female, secondary educ.)}$$

$$\hat{L}_{12} = \hat{B}_0 + \hat{B}_2 \qquad \text{(i.e. male, secondary educ.)}$$

Now the difference in the logit estimate between the elementary and secondary level is not independent of sex, as it was before. This difference in the logit is $\hat{B}_2+\hat{B}_5$ for Females, and $\hat{B}_2$ for males.

Note that for the "reference" category (ie., male and elementary), the logit value is simply that of the intercept parameter, $\hat{B}_0$.

The regression yields predicted values, $\hat{L}_{ij}$, of the logit. These can be converted to predicted value of the probability, $P_{ij}$, for any category ij, as follows:

$$\text{Since } \hat{L}_{ij} = \ln \frac{\hat{P}_{ij}}{1-\hat{P}_{ij}}$$

rearranging, we get

$$\hat{P}_{ij} = \frac{e^{\hat{L}_{ij}}}{1 + e^{\hat{L}_{ij}}}$$

The t values reported in the output refer to a test of the null hypothesis that a coefficient is zero. This is testing the null hypothesis that the effect on the logit of the variable associated with the coefficient is statistically insignificant. In our example, the t value associated with $B_3$ (college education) can be interpreted as testing the null hypothesis that there is not a statistically significant difference in the value of the logit (and hence in the probability) between persons with elementary education (the reference group) and those with college education. It is also possible to test hypothesis involving a <u>set of coefficients</u> (e.g., all those related to education).

To estimate the effect on P of moving from one level of a qualitative variable to another (with all other variables held constant), it is necessary to start with some given value of P, since as mentioned earlier, the effect of a variable on P depends upon the value of P itself. One can often start with the value of P for some particular group of interest. Suppose in the examples, one wanted to evaluate the effect on P of moving from the secondary ($B_2$) to the college ($B_3$) categories for females. One would simply then calculate P for the two cases, and determine the difference. It may be desirable, however, to see how the effect of moving between these two categories would vary over low and high values of P. In this case, we would start with some given value $P_0$, calculate

$$\hat{L}_0 = \ln \frac{\hat{P}_0}{1-\hat{P}_0}$$

then add the effect on the logit of moving between the two categories, that is $B_3 - B_2$, and calculate the new P, $P_1$, where

$$\hat{L}_1 = \hat{L}_0 + \hat{B}_3 - \hat{B}_2$$

and
$$P_1 = \frac{e^{\hat{L}_1}}{1 + e^{\hat{L}_1}}$$

The effect on P of moving between two levels of a qualitative variable, given a starting point of $P_0$, is simply $P_1 - P_0$.

This is similar to calculating the partial derivative of the logit with respect to a particular variable, $X_k$, ie. $\left[\frac{\partial P}{\partial X_k} = B_k\ P(1-P)\right]$ and evaluating it at different values of P to determine the relationship between $X_k$ and P, where $X_k$ is a continuous variable. In essence, with a continuous variable one calculates the slope, and with qualitative variables the method outlined above calculates the chord between the two levels of the qualitative variable.

To estimate the regression coefficients, a maximum likelihood procedure was used with microdata (individual records) for a large number of persons in the labour force. The very large samples (from 41,000 to 49,000) were necessary because the proportion of <u>some</u> subgroups taking training was quite small. With small samples, there would have been very few cases of persons taking training in some subgroups.

One major difference between logistic regression using micro-data, (or any regression with discrete dependent variables) and that with continuous dependent variables is the use of the $R^2$ statistic. In logistic regression, $R^2$ as it is commonly used is of very little value in determining the goodness of fit of the overall model.

This is because the observed values for a particular observation are either 0 or 1, but the predicted values take the form of a probability, lying between 0 and 1. If the predicted probabilities range between 0.05 and 0.5, as is the case in this work, then the residuals will be very large, and the $R^2$ very low. The $R^2$, when computed, is invariably extremely low when using microdata where the dependent variable is binary. They are not very useful, and hence not reported here.

What is reported is the overall model $\chi^2$, which tests the null hypothesis that the probabilities predicted by the full fitted model (with the independent variables) are not significantly different from those predicted by the intercept term only. But this $\chi^2$ test is sensitive to sample size, increasing as sample size increases, although the critical value of $\chi^2$ is not a function of sample size. Thus, models with very large samples, such as is used here, will rarely fail the $\chi^2$ test.

Another statistic which was used by Stopher and Meyburg (1979) is the likelihood ratio index $p^2$.

$$p^2 = 1 - \frac{L_1'}{L_0'}$$

where $L_1'$ = log likelihood for fitted model

$L_0'$ = log likelihood for model with intercept term only

If the fitted model adds nothing to the null hypothesis, then $L_1'$ and $L_0'$ will be approximately equal, and $p^2$ will approach zero. As the fitted model diverges further from the null hypothesis (ie., the fit with intercept term only) $p^2$ will approach 1. The difficulty with this statistic, however, is that large values (approaching 1) are quite unlikely, and thus one does not know in what range $p^2$ should be to constitute a "good" fit. The author of the statistic suggests that a $p^2$ of .230 (achieved in his example) indicated "an excellent" fit, but the level of $p^2$ below which a fit is considered "poor" remains unclear.

In any case, the concentration in this work is on the t values and the predicted values of P itself, as the major hypotheses and question of interest concern the degree of the difference in the probability of training among different populations of particular interest.

## FOOTNOTES

1. AOTA is the forerunner of the National Training Act.

2. A person was an adult trainee if they were between ages 18 and 45 and, had not been a full-time student for the previous six months (i.e., were not "continuing" students).

3. The 250,000 figure is likely an underestimate. Approximately 80,000 to 100,000 people in full-time programs of one, two or more months duration were classified as taking short-term training because of the way they responded to the question. Furthermore, data from other sources suggest that the 250,000 likely consist of 35,000 to 50,000 entrants to post-secondary college programs (excluding trades), an equal number to university programs, and 150,000 to 180,000 to other programs, many of which were funded through the National Training Program of Employment and Immigration Canada. Thus, the majority would have enrolled for employment-related reasons, since even among those entering university, this was true of over 50%.

4. The distinction between short-term and full-time training is not entirely clear-cut. Some people who took "short-term" training attended day-long classes for some period of time (two to four months). In essence, distinguishing betwen the two categories of training was left up to the respondents. If they believed they attended a school, college or university as a full-time student, then their training was recorded as such; otherwise, it was classified as short-term.

5. This may have been because many young people held part-time or temporary jobs and would leave them to become full-time students more readily than older persons with permanent jobs.

6. Even though the unemployed were more likely to enroll in the 21-44 age group, almost half the adults entering full-time training had been predominantly employed during the previous six months; only 16% had been predominantly unemployed. This apparent anomaly simply reflects the fact that the employed population is far larger, and even if a smaller percentage of them enroll, they still outnumber the unemployed.Another 18% had not been in the labour force, and the remaining 14% had no predominant labour force status.

7. In the regression, an age-employment interaction variable was introduced. The coefficient for the unemployed groups was significantly different from the reference group, (the employed) indicating a statistically significant difference in their probabilities of entering.

8. It is interesting to note, however, that there was not a significant change in the probability of entering training over the 25 to 39 age span. There appeared to be three groupings with significantly different probabilities: 18-24, 25-39 and over 40. For example, men with a secondary school education in the 18-24 age group had a probability of entering a program in the 2.7% to 9.9% range (depending upon their occupation, province, etc.), while for the 25-39 age group the range fell to 1.7% to 5.9% and in the 40-45 group to 0.7% to 2.3% range (Figure 5.1).

9. Shaw, Paul, Sizing Up the Burden of Unemployment in Canada, Part II: A Perspective on Financial Hardship, Social and Economic Studies Division, Research Paper .., Statistics Canada, 1984, page 34.

10. Training in the clerical and services occupations forms a substantial component (25%) of all institutional training offered under the federal National Training Act. The other large component is in the blue collar (machining, assembling, repairing, construction, transportation equipment operating) occupations, which accounted for 45% of total training sponsored by Employment and Immigration Canada (E&IC) in 1982-83. And the majority (76%) of people taking this training were unemployed prior to starting the program.

11. These variables were most strongly associated with the probability of the unemployed entering full-time training. Categories have been defined so that variations in probabilities are minimized <u>within</u> the groups, and maximized <u>among</u> the groups.

12. What constitutes a "low" probability is somewhat arbitrary. If the "low" probability cut-off was set at .03 (i.e., 3% or less of the long-term unemployed in the group started full-time training in 1983), then almost one-half (49%) of the unemployed belonged to groups with low training rates. If the "low" cut-off was set at .04, then 60% of the unemployed fell into low groups. If set at .02, 39% were in such groups.

13. For occupations, the above-average rates must apply between 1982 and 1984, for industries, between 1975 and 1984. Unemployment data on occupations at the 3-digit CCDO level - used here - did not exist before 1982. It is recognized that, in some cases, high annual unemployment rates reflect a very large supply of labour, seasonal or voluntary unemployment or other patterns of lay-offs. But in most cases, high unemployment was associated with slower-than-average growth in demand for labour.

## REFERENCES

BARRETT, Alan J. (1985), Sources of Recent Changes in the Unemployment Rate in Canada, Labour Market Bulletin, Employment and Immigration Canada, July.

BINDER, D.A. (1983), "On the Variance of Asymptotically Normal Estimates from Complex Surveys", International Statistical Review, 51-279-292.

E&IC (1981b), Labour Market Developments in the 1980s, Supply and Services Canada.

E&IC (1984c), Consultation Paper: Training, Supply and Services Canada.

ECONOMIC COUNCIL (1983), On the Mend, (Ottawa: Ministry of Supply and Services).

ECONOMIC COUNCIL OF CANADA (1982), In Short Supply: Jobs and Skill in the 1980s, Canadian Government Publishing Centre, Supply and Services Canada.

KISH, L. and FRANKEL, M.R. (1974), Inference from Complex Samples, J.R. Statistical Society, 36, 1-37.

PEARSON, C., and SALEMBIER, G. (1983), Trade, Employment and Adjustment, The Institute for Research on Public Policy, Montreal.

SAUNDERS, R.S. (1984), Aid to Workers in Declining Industries, Ontario Economic Council.

SHAW, P. (1984), Sizing Up the Burden of Unemployment, Part I: A General Labour Market Perspective, Research Paper No. 4, Social and Economic Studies Division, Statistics Canada.

SHAW, P. (1984), Sizing Up the Burden of Unemployment, Part II: A Perspective on Financial Hardship, Research Paper No. 5, Social and Economic Studies Division, Statistics Canada.