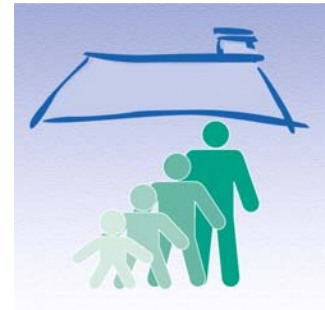


Study on the reliability of the weighing scales used in the 2008 Canadian Community Health Survey (CCHS) for the measured height and weight (MHW) module

José Pedro, methodologist, CCHS, June 2009



1. Introduction

This document presents the results of the study on the reliability of the weighing scales used to collect measured height and weight (MHW) module data. Physical height and weight measurements were taken for a subsample of about 5,000 respondents of the 2008 Canadian Community Health Survey (CCHS). The data collection for this survey is managed in 4 regional offices: Edmonton, Montréal, Halifax and Toronto.

We will first briefly describe the goal of this study. Then we will discuss the methodology used. Finally, we will present the conclusions drawn from tables showing all the statistics used to measure pre-survey and post-survey scale accuracy.

2. Reliability study goal

To ensure the accuracy of physical measurements data, we conducted a study on the reliability of the scales used in the survey. Data for this reliability study was collected in two stages. Measurements were first collected before the survey began (December 2007) to determine that the scales were accurate and operating properly before being sent to the regional offices. Measurements were again collected after the survey (January 2008) to ensure the accuracy of the scales used in the field.

3. Methodology

To test scale accuracy, a sample of scales was selected for each regional office (the document Validation Instructions.doc describes sampling details). Initially, 20 scales were selected per regional office. In cases where 2 scales did not successfully pass the validation process, an additional sample of 10 scales was selected. For a scale to fail the validation process, it had to meet at least one of the following criteria: batteries do not work, the scale does not begin at 0, at least one of the 4 weight measurements (40 Kg, 80 Kg, 120 Kg and 140 Kg) is outside the range of acceptable values of 2%, the scale is broken or the scale does not work at all.

Here is a brief description of the measurement procedure followed. At each regional office and for each scale selected, a few descriptive characteristics were recorded on an Excel spreadsheet: the number of the box from which the scale was taken, the location of the box in the warehouse and the scale serial number. For the pre-survey study, battery

operation was also tested. However, for the post-survey study, new batteries were used for each scale selected. Then measurements were collected for each weight (40 Kg, 80 Kg, 120 Kg and 140 Kg) and the value indicated by the scale was recorded. Finally, the values obtained were checked to ensure that they were not outside the 2% range of acceptable values.

It should be noted that no additional sample had to be selected, since the original samples complied with the validation procedures.

Once measurements had been taken, scale accuracy was checked by looking at the relationship between the measured weight value and the true weight value based on a Student's hypothesis test (T-TEST). This method is used to determine whether or not there is a significant difference between the average of the measurements taken with the sampled scales and the reference value. The t-test was applied for each of the weights tested at 40, 80, 120 and 140 Kg. The P value¹ and the upper and lower t-test confidence interval boundaries were calculated for all regions and for each region. The significance level used was 5%. That means that if we find a P value greater than 5%, there is no significant difference between the observed average and the reference weight value. Otherwise, we can say that there is a significant difference between the average and the reference weight.

SAS 9.0 and Excel 2002 were used to perform our analyses and produce our tables.

4. Analysis – Descriptive statistics and t-test

4.1 Pre-survey analysis

Before the survey, the accuracy of each of the 80 scales (20 for each of the 4 regional offices) was tested. The tests were performed on 40 Kg, 80 Kg, 120Kg and 140 Kg reference weights. When we look at Table 4.1 for all regional offices, we see that the minimum and maximum at each reference weight is within the range of acceptable values of 2% established when measurements were taken with calibrated weights. For the 40 Kg weight, the minimum and maximum are reached at 39.80 and 40.70 respectively and are within the range of acceptable values of (39.2; 40.8). For the 80 Kg weight, the minimum and maximum are reached at 79.65 and 80.70 and are within the range of acceptable values of (78.4; 81.6). For the 120 Kg weight, the minimum and maximum are reached at 119.85 and 120.90 and are within the range of acceptable values of (117.6; 122.4). Finally, for the 140 Kg weight, the minimum and maximum are reached at 139.85 and 141.50 and are also within the range of acceptable values of (137.2; 142.8).

When we look at the t-test P value at each of the reference weights (40, 80, 120 and 140 Kg), we note a significant difference between the sample average and the reference

¹ The P value of a statistical hypothesis test is the probability of erroneously rejecting the basic hypothesis even though it is true. Thus, the smaller the P value, the smaller the probability of erroneously rejecting the basic hypothesis knowing that it is true. The basic hypothesis will then be accepted as false. A P limit value of 5% is commonly used in hypothesis testing.

weight ($p < 5\%$). However, looking at the upper and lower confidence interval boundaries, we see that even though 40, 80, 120 and 140 are not within their respective confidence levels, the upper and lower boundary values are clearly always very close to the reference values of (40.1;40.06), (80.03;80.1), (120.1;120.17) and (140.12;140.23) respectively. It can thus be said that the scales are sufficiently accurate.

Then, similar analyses were performed for each of the regional offices with a sample of 20 scales per region. The results should be used with care since they apply to a sample of less than 30 units. In the case of Edmonton and Toronto, the minimums and maximums are all also within the 2% range of acceptable values. A P value greater than 5% is noted for the measurements tested at 40 Kg. This means that there is no significant difference between the average and the 40 Kg reference value. For the other measurements, i.e. 80, 120 and 140 Kg, the P value is less than 5% and thus significantly different from the reference weights. Again, since the upper and lower boundaries are always very close to the reference values, it can be said that the scales are sufficiently accurate.

In the case of Halifax and Montréal, the minimums and maximums are within the range of acceptable values. A P value greater than 5% is noted for the measurements tested at 40 and 80 Kg. That means that there is no significant difference between the average at 40 Kg and the reference value or between the average at 80 Kg and the reference value for each of the 2 regions. For the 120 and 140 Kg tests, the P values are under 5% but the confidence intervals are such that the conclusions are the same as in the preceding analyses.

Table 4.1: Pre-survey scale statistics

Regional office	Statistics	Test at 40 Kg	Test at 80 Kg	Test at 120 Kg	Test at 140 Kg
All regional offices	Maximum	40.7	80.7	120.9	141.5
	Mean	40.03	80.07	120.14	140.17
	Minimum	39.8	79.65	119.85	139.85
	Count	80	80	80	80
	Standard error	0.11	0.16	0.16	0.25
	Lower bound of the confidence interval	40.01	80.03	120.1	140.12
	Upper bound of the confidence interval	40.06	80.1	120.17	140.23
	Standard error	0.01	0.02	0.02	0.03
	P-value	0.0127	0.0003	<.0001	<.0001

Regional office	Statistics	Test at 40 Kg	Test at 80 Kg	Test at 120 Kg	Test at 140 Kg
EDMONTON	Maximum	40.1	80.15	120.25	140.35
	Mean	40.02	80.04	120.07	140.11
	Minimum	39.9	79.9	119.85	139.95
	Count	20	20	20	20
	Standard Error	0.07	0.07	0.09	0.12
	Lower bound of the confidence interval	39.99	80.01	120.02	140.06
	Upper bound of the confidence interval	40.05	80.07	120.11	140.17
	Standard error	0.01	0.02	0.02	0.03
	P-value	0.3157	0.011	0.0044	0.0004
	HALIFAX	Maximum	40.1	80.2	120.35
Mean		40.02	80.03	120.11	140.12
Minimum		39.95	79.85	119.95	139.9
Count		20	20	20	20
Standard error		0.06	0.1	0.11	0.11
Lower bound of the confidence interval		39.99	79.98	120.05	140.07
Upper bound of the confidence interval		40.05	80.08	120.16	140.17
Standard error		0.01	0.02	0.03	0.02
P-value		0.1485	0.1864	0.0006	<.0001
MONTREAL		Maximum	40.25	80.7	120.9
	Mean	40.02	80.1	120.2	140.19
	Minimum	39.8	79.65	119.9	139.85
	Count	20	20	20	20
	Standard error	0.09	0.25	0.2	0.28
	Lower bound of the confidence interval	39.98	79.98	120.1	140.06

Regional office	Statistics	Test at 40 Kg	Test at 80 Kg	Test at 120 Kg	Test at 140 Kg
MONTRÉAL	Upper bound of the confidence interval	40.06	80.21	120.29	140.33
	Standard error	0.02	0.06	0.05	0.06
	P-value	0.3915	0.1031	0.0004	0.0069
TORONTO	Maximum	40.7	80.45	120.65	141.5
	Mean	40.08	80.1	120.19	140.27
	Minimum	39.9	79.85	120.05	139.95
	Count	20	20	20	20
	Standard error	0.19	0.16	0.17	0.37
	Lower bound of the confidence interval	39.99	80.02	120.11	140.1
	Upper bound of the confidence interval	40.17	80.18	120.27	140.44
	Standard error	0.04	0.04	0.04	0.08
	P-value	0.0811	0.0129	<.0001	0.0041

4.2 Post-survey analysis

After the survey, the accuracy of each of the 91 scales (20 in Edmonton and Halifax, 21 in Montréal and 30 in Toronto) was tested. The test measurements were used for 87 of the 91 scales, since 4 scales were found to be defective. One of those scales was in Montréal and the three others in Toronto. The broken scales represent 4.4% of all the scales tested in the study. During the analysis, we noted that 2 of those defective scales were used during collection for 14 measurements. Those 14 cases constitute only 0.28% of the 5,000 MHW module cases. Those cases were excluded from the study.

Again, the tests were performed on reference weights of 40 Kg, 80 Kg, 120Kg and 140 Kg for the operating 87 scales and the results were similar to the pre-survey results. Here are the details.

We note in Table 4.2 that the minimum and maximum at each reference weight are within the range of acceptable values – i.e. 2%. For the 40 Kg weight, the minimum and the maximum are reached at 39.90 and 40.50 and are within the range of acceptable values of (39.2; 40.8). For the 80 Kg weight, the minimum and maximum are reached at 79.70 and 80.50 and are within the range of acceptable values (78.4; 81.6). For the 120 Kg weight, the minimum and the maximum are reached at 119.95 and 120.65 and are within the range of acceptable values (117.6; 122.4). Finally, for the 140 Kg weight, the

minimum and maximum are reached at 139.95 and 140.40 and are within the range of acceptable values (137.2; 142.8).

When we look at the t-test P value at each reference weight (40, 80, 120 and 140 Kg), we see a significant difference between the sample average and the reference weight. We come to the same conclusions as for the pre-survey analysis – i.e. that the upper and lower boundaries are very close to the reference values, which are (40.04;40.07), (80.07;80.11), (120.11;120.16) and (140.13;140.17) respectively. The scales can thus be considered sufficiently accurate.

Then, similar analyses were done by regional office. Again, the results should be used with care, since they apply to a sample of less than 30 units. For each regional office and each weight value, we note a P value under 5%, which means there are significant differences between the average and the reference value. Again, since the upper and lower boundaries are always very close to the reference values, it can be concluded that the scales are sufficiently accurate.

Table 4.2: Post-survey scale statistics

Regional office	Statistics	Test at 40 Kg	Test at 80 Kg	Test at 120 Kg	Test at 140 Kg
All regional offices	Maximum	40.5	80.5	120.65	140.4
	Mean	40.06	80.09	120.14	140.15
	Minimum	39.9	79.7	119.95	139.95
	Count	87	87	87	87
	Standard error	0.06	0.09	0.11	0.09
	Lower bound of the confidence interval	40.04	80.07	120.11	140.13
	Upper bound of the confidence interval	40.07	80.11	120.16	140.17
	Standard error	0.01	0.01	0.01	0.01
	P-value	<.0001	<.0001	<.0001	<.0001
EDMONTON	Maximum	40.1	80.2	120.25	140.25
	Mean	40.06	80.09	120.13	140.13
	Minimum	40	79.95	120	140
	Count	20	20	20	20
	Standard error	0.03	0.06	0.06	0.06
	Lower bound of the confidence interval	40.04	80.06	120.1	140.1
	Upper bound of the confidence interval	40.08	80.12	120.16	140.16
	Standard error	0.01	0.01	0.01	0.01
	P-value	<.0001	<.0001	<.0001	<.0001

Regional office	Statistics	Test at 40 Kg	Test at 80 Kg	Test at 120 Kg	Test at 140 Kg
HALIFAX	Maximum	40.1	80.5	120.5	140.4
	Mean	40.05	80.13	120.18	140.2
	Minimum	40	80	120	140.05
	Count	20	20	20	20
	Standard error	0.03	0.11	0.12	0.09
	Lower bound of the confidence interval	40.04	80.08	120.12	140.16
	Upper bound of the confidence interval	40.07	80.18	120.23	140.24
	Standard error	0.01	0.03	0.03	0.02
	P-value	<.0001	<.0001	<.0001	<.0001
MONTRÉAL	Maximum	40.1	80.2	120.65	140.3
	Mean	40.04	80.08	120.14	140.12
	Minimum	39.95	80	120	140
	Count	20	20	20	20
	Standard error	0.04	0.06	0.15	0.07
	Lower bound of the confidence interval	40.03	80.06	120.07	140.09
	Upper bound of the confidence interval	40.06	80.11	120.21	140.16
	Standard error	0.01	0.01	0.03	0.02
	P-value	<.0001	<.0001	0.0005	<.0001
TORONTO	Maximum	40.5	80.2	120.25	140.3
	Mean	40.06	80.06	120.11	140.14
	Minimum	39.9	79.7	119.95	139.95
	Count	27	27	27	27
	Standard error	0.11	0.1	0.08	0.1
	Lower bound of the confidence interval	40.02	80.02	120.08	140.1
	Upper bound of the confidence interval	40.11	80.1	120.14	140.18
	Standard error	0.02	0.02	0.02	0.02
	P-value	0.0037	0.0036	<.0001	<.0001

5. Conclusion

The scale reliability study led to the conclusion that the scales are sufficiently accurate. Although we noted significant differences between the averages and the reference values, the upper and lower confidence interval boundaries are very close to the reference weight value. This conclusion is valid for both the pre-survey and post-survey data.