Social Studies 201

February 4, 2005

Interquartile Range (IQR)

See text, section 5.8.2, pp. 218-224.

The interquartile range (IQR) is the range of the values of a variable over the middle part of a distribution. Specifically it is the range from the 25th to the 75th percentile of a variable. The IQR is a range, once the upper one-quarter and lower one-quarter of cases of a distribution are eliminated from consideration.

> **Definition**. The interquartile range (IQR) is the seventy-fifth percentile minus the twenty-fifth percentile. In symbols,
>
> $$\text{IQR} = P_{75} - P_{25}$$
>
> Alternatively, the interquartile range is the third quartile minus the first quartile, since the third quartile is defined as the seventy-fifth percentile and the first quartile is the twenty-fifth percentile.

The unit of measure for the IQR is the same as the unit of measure for the variable X. For example, if X represents income of individuals in dollars, each percentile is also in dollars and the IQR is in dollars.

Compared with the range, one advantage of the interquartile range is that it indicates the spread or concentration for the middle one-half of the distribution, ignoring the extremes of the distribution. This is worthwhile for statistical analysis when the extremes are of less interest and more consideration of the middle part of the distribution is required. That is, the difference between the seventy-fifth and twenty-fifth percentiles provides a good indication of the range of the values for the middle, or more typical cases, of the distribution.

The IQR eliminates the effect of extreme values. For example, a distribution such as an income distribution may be primarily composed of people with low and middle income. Adding a few individuals with very high incomes (say millionaires) to this distribution could dramatically increase the range

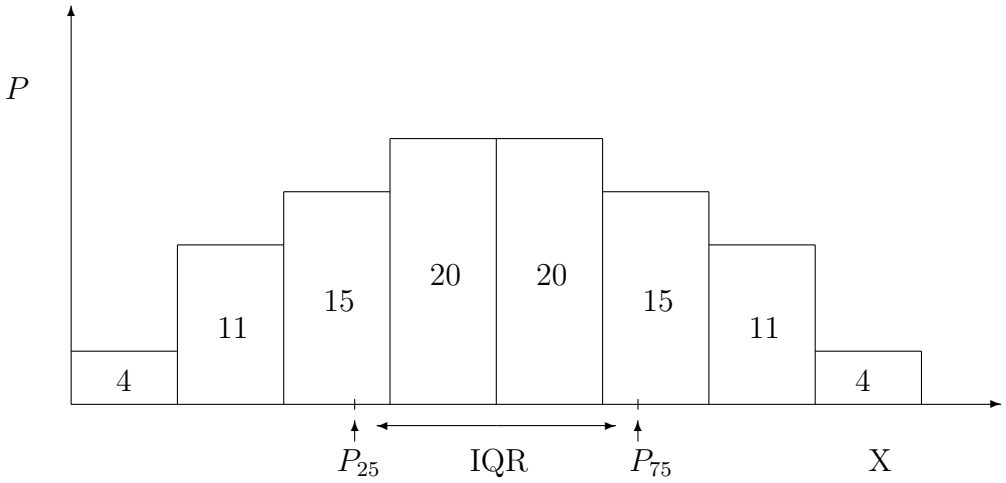
of the distribution. In contrast, the interquartile range would change little, since it indicates the range of the middle part of the income distribution.

One defect of the IQR as a measure of variation is that it is based on only two specific percentiles, and does not take other values of the variable into account. This occurs because the IQR is a positional measure, indicating only the difference between two other positional measures, P_{75} and P_{25}. In general though, this defect is outweighed by the advantages noted earlier.

Diagrammatic illustration of the interquartile range – Figure 1

The distribution in Figure 1 is a histogram, with the approximate percentage of cases in each bar listed in the middle of the bar; the sum of these percentages is one hundred per cent.

Figure 1: Diagrammatic illustration of interquartile range



The approximate position of the twenty-fifth percentile is indicated by P_{25} – about two-thirds of the way across the bar labelled 15. That is, there are 4% of cases in the first interval and another 11% in the second interval, for a total of 15%. The location of P_{25} is the value of X such that 25% of cases are less than or equal to P_{25}. Since there are 15% of cases in the first two categories, and another 15% in the third category, the twenty-fifth per cent point is reached at about two-thirds of the way across the third category.

By similar reasoning, the seventy-fifth percentile occur about one-third of the way across the sixth category – there are seventy per cent of cases in the first five categories and about one-third of the way across the sixth category is where P_{75} is located. This is indicated by P_{75} in Figure 1, the value of X such that 75% of cases are less than or equal to P_{75}.

The IQR is the distance between P_{25} and P_{75}. There are 25% of cases less than P_{25}, another 25% of cases greater than P_{75}, and the middle 50% of cases occur within the interval represented by the IQR.

Example 4.3 – Variation in attitudes to social issues. A sample of just under seven hundred undergraduate students was asked to state their view about two statements, "Tax laws and job benefits should recognize gay and lesbian couples as married" and "More provincial tax dollars should be devoted to universal health care." Respondents stated their degree of agreement or disagreement about this statement on a five-point scale, from 1, indicating strongly disagree with the statement, to 5, indicating strongly agree with the statement. Responses are reported in Table 1.

Table 1: Frequency distributions of views of undergraduate students to statements concerning gay and lesbian couples and expenditures for health care

Response	Frequency	
X	Gay/lesbian couples	Health care
1 (strongly disagree)	141	35
2 (somewhat disagree)	86	71
3 (neutral)	183	230
4 (somewhat agree)	161	222
5 (strongly agree)	124	126
Total	695	686

Question. Obtain the interquartile range for responses to the two statements. In words, briefly compare the variability of the two distributions.

Answer. For each variable, the scale is ordinal, since these are attitudes or opinions concerning the respective issue, with responses measured on a 1-5 scale. Responses are ordered from strongly disagree to strongly agree, an ordinal scale, so percentiles and the IQR can be meaningfully obtained.

Since the IQR is the difference between the seventy-fifth and twenty-fifth percentiles, it makes most sense to reconstruct the frequency distributions into percentage and cumulative percentage distributions, so percentiles can be easily determined. This is done in Table 2, with the two distributions separated with a vertical line.

Table 2: Percentage (P) and cumulative percentage (Cum. P) distributions of views about gay and lesbian couples and expenditures for health care

Response	Gay/lesbian couples		Health care	
X	P	Cum. P	P	Cum. P
1 (strongly disagree)	20.3	20.3	5.1	5.1
2 (somewhat disagree)	12.4	32.7	10.3	15.4
3 (neutral)	26.3	59.0	33.5	48.9
4 (somewhat agree)	23.2	82.2	32.4	81.3
5 (strongly agree)	17.8	100.0	18.7	100.0
Total	100.0		100.0	

From Table 2, percentiles can be readily determined. Attitude responses are measured on a discrete scale (1, 2, 3, 4, or 5). Interpolation is not necessary; the percentiles are the values of the response X where the appropriate percentage of cases is first obtained.

Gay and lesbian couples. For the statement concerning gay and lesbian couples, the twenty-fifth percentile is at $X = 2$. There are 20.3% of cases at $X = 1$ and another 12.4% of cases at $X = 2$, totalling 32.7% of cases with values of 2 or less. Thus $P_{25} = 2$.

The seventy-fifth percentile is at $X = 4$, since there are fewer

than 75% of cases at $X = 3$ or less but more than 75% at 4 or less. Thus $P_{75} = 4$.

The IQR is thus two.

$$IQR = P_{75} - P_{25} = 4 - 2 = 2$$

The interquartile range is 2 for responses to the statement about treating gay and lesbian couples as married.

Health care expenditures. For the health care issue, the twenty-fifth percentile is not obtained until $X = 3$, since there are only 15.4% of cases at values 1 and 2 of responses, and the 25 per cent point is first reached at a response of 3. Also $P_{75} = 4$, since seventy-five per cent of cases is first reached when $X = 4$. Thus the IQR is 1.

$$IQR = P_{75} - P_{25} = 4 - 3 = 1$$

The interquartile range is 1 for responses to the statement about health care expenditures.

Comparison of variability. Comparing these two distributions, views are more varied on the gay and lesbian couples issue ($IQR = 2$) and less varied on the health care expenditure issue ($IQR = 1$). The range for the two distributions is identical, from an attitude response of 1 to and attitude response of 5. But across these possible values for the variable, responses are more concentrated on the health care issue, with few respondents disagreeing and most respondents having a response of 3 or more. This produces a relatively small IQR of only 1 point on the five-point attitude scale. In contrast, responses are more varied on the gay and lesbian couples issue, with considerable percentages of responses at each of the five possible responses.

In summary, there is great variation of student views concerning treating gay and lesbian couples as married, with many respondents in support of it but with many also opposed to it. In contrast, views concerning tax dollars for health care are more uniform or concentrated, with respondents generally in agreement that this would be a good policy.

Example 4.4 – Variation in earnings in Saskatchewan and Alberta The distributions of earnings for respondents in Saskatchewan and Alberta were examined in Exercise 3.xxx. These distributions are provided again in Table 3.

Table 3: Distribution of earnings, Saskatchewan and Alberta respondents, 2002

Income in thousands of $	Percentage of respondents Saskatchewan	Alberta
Less than 5	16.2	11.7
5-10	12.9	11.4
10-15	8.8	8.8
15-20	8.2	7.9
20-30	15.6	14.0
30-40	12.3	13.3
40-50	9.4	9.9
50-60	6.6	6.8
60 plus	10.0	16.2
Total	100.0	100.0

Adapted from Table 202-0101, Statistics Canada, *Income Trends in Canada 1980-2002*, catalogue number 13F0022XCB, 2004.

Question. Obtain the interquartile range for the distribution of earnings among respondents in each of Saskatchewan and Alberta. Briefly comment on differences in the variation of earnings in the two provinces.

Answer. In Exercise 3.xxx, the twenty-fifth and seventy-fifth percentiles (P_{25} and P_{75}, respectively) were calculated. The value of these percentiles, in dollars, is given in Table 4. If you do not recall how to obtain these percentiles, review Exercise 3.xxxx.

The interquartile range for Saskatchewan is

$$IQR = P_{75} - P_{25} = 41,100 - 8,400 = 32,700$$

Table 4: Summary of statistics for Saskatchewan and Alberta earnings

	Value of statistic ($)	
Measure	Saskatchewan	Alberta
P_{75}	41,100	48,000
P_{25}	8,400	11,100
IQR	32,700	36,900

and for Alberta is

$$IQR = P_{75} - P_{25} = 48,000 - 11,1400 = 36,900$$

The interquartile range is $32,700 for Saskatchewan earners and $36,900 for Alberta earners.

From these two measures, and examining the table, the variation in earnings is greater in Alberta than in Saskatchewan. There is a difference of just over $4,000 in the interquartile ranges ($36,900 minus $32,700) for the two provinces, with Alberta having greater variation than Saskatchewan. The range of earnings is the same in the two provinces, from less than $5,000 to $60,000 plus. But in Saskatchewan there is a greater concentration of respondents at lower and middle earnings levels. For Alberta, there is a greater percentage of respondents at the high end of earnings, resulting in a greater variation for Alberta. The respective interquartile ranges are consistent with the larger variation for Alberta than Saskatchewan, with the range over the middle one-half of earnings being $36,900 for Alberta and $32,700 for Saskatchewan.

Recap of Interquartile range (IQR)

The interquartile range is a useful measure of variation in that it describes the extent of variation over the middle part of a distribution. The advantage of the IQR over the range is that the influence of the extremes of a distribution are eliminated – only the difference between the seventy-fifth and twenty-fifth percentiles is considered. For variables such as income or earnings, the IQR is especially useful. These variables may have a few extreme values, with either very large or very small incomes or earnings. By eliminating the lower one-quarter and the upper one-quarter of cases, the IQR provides a good summary of how varied the more typical incomes are.

vspace0.1cm

The IQR is also an easy measure to calculate, at least once the percentiles are obtained. It is merely the difference between the seventy-fifth and twenty-fifth percentiles.

The disadvantage of the interquartile range is that it is a positional measure, based on only the twenty-fifth and seventy-fifth percentiles. The next measures of variation to be examined in these notes, the standard deviation and variance, remedy this defect. For these next measures, the value associated with each case is taken into account.

In summary, the IQR is used when the researcher wishes to eliminate the influence influence of extreme values and consider the variation for the more typical cases in a distribution.