**Chapter 7**

**Comparison of Rates**

with Munir Ahmed

# [H1] Learning Objectives

# [INSERT NL]

1. Define distribution of rates
2. Compare statistical significance of a rate
3. Compare two rates to each other
4. Calculate confidence intervals for odds ratios
5. Create a p-chart and related control limits
6. Create risk-adjusted p-charts  
   **[END NL]**

# [H1] Key Concepts

# [INSERT BL]

* Binomial distribution
* Bernoulli distribution
* Normal approximation
* P-chart
* Risk-adjusted p-chart
* Deviance  
  **[END BL]**

# [H1] Chapter at a Glance

This chapter describes how the rate of occurrence of discrete events can be calculated. Discrete events, or discrete variables, take on specific values in a given range. Variables such as race, gender, mortality within six months, or age in different decades are all examples of variables that have countable discrete levels. The rate of occurrence of each of the levels of these variables can be examined. Rates are often reported for binary variables, where the variable assumes two levels. A binary variable can take only two values, 0 and 1, which represent two different groups or levels of a categorical variable of interest. For example, 1 may indicate males and 0 females, if we assume that there are only two genders in our study. Similarly, a binary variable can be used to describe death within six months; dead may be scored as 1 and alive as 0, if we assume that patients are either dead or alive. The group which is represented by 0 is often called the *reference group* or *reference category.*

# [H1] Summarizing Discrete Variables

Discrete variables can be summarized with *frequency distributions*. A frequency distribution counts the number of times each level of a discrete variable occurs in the sample. Exhibit 7.1 is an example of a frequency distribution. Probabilities are calculated by dividing the number of times a particular level of a variable is observed by the total number of observations across all levels of the variable. For example, the probability of admission to an emergency room on Monday is calculated by dividing the count of admissions on Monday by the total number of admissions: 15 ÷ 212 = 0.07. These probabilities can also be multiplied by 100 in order to transform them into percentages, so the probability of 0.07 can be expressed as 7 percent.

**[INSERT EXHIBIT]**

**Exhibit 7.1** Car Accident Victims Seen in an Emergency Room in One Year, by Day of the Admission

|  |  |  |  |
| --- | --- | --- | --- |
| *Days of Week* | *Admissions* | *Relative Frequency* | *Percentage* |
| Mondays | 15 | 0.07 | 7.08 |
| Tuesdays | 12 | 0.06 | 5.66 |
| Wednesdays | 18 | 0.08 | 8.49 |
| Thursdays | 25 | 0.12 | 11.79 |
| Fridays | 49 | 0.23 | 23.11 |
| Saturdays | 53 | 0.25 | 25.00 |
| Sundays | 40 | 0.19 | 18.87 |
| Total | 212 | 1.00 | 100.00 |

**[END EXHIBIT]**

For a discrete variable *X*, the population (or sample) proportion of a value *x* is the number of times that value appears in the population (or sample). For a sample of size *n*, the proportion *p* of a value of interest *x* is thus given by

**[INSERT EQUATION]**

## .

**[END EQUATION]**

## [H1] The Bernoulli Process and the Binomial Distribution

Jacob Bernoulli, a seventeenth-century Swiss mathematician, examined binary variables (e.g., alive vs. dead). An event is assumed to occur with a probability *p*. Repetition of Bernoulli trials produces a binomial distribution. To create a binominal distribution, four conditions need to be satisfied:

**[INSERT NL]**

1. Each repetition consists of the same two mutually exclusive events.

2. The number of repetitions is fixed.

3. Successive repetitions are independent of each other.

4. The probability of the outcome of interest (often classified as a *success*) remains unchanged across repetitions.

**[END NL]**

For example, a healthcare manager might want to know how many children (aged 0–18) are having surgery versus how many adults (aged 18 or older). The manager might count surgeries done in a given month. Because a patient is either a child or adult, the choices are mutually exclusive. The number of surgeries is fixed. Successive surgeries are independent of each other. The probability of a surgery remains unchanged across repetitions. Therefore, calculating the probability that a child would have surgery at PDQ Hospital could be displayed as a binomial distribution.

If we denote the number of successes in a binomial process with *X*, then *X* is a binomial random variable, and its probability distribution is referred to as the *binomial probability distribution*. With *n* repetitions, random variable *X* can thus take any value between 0 and *n* where *X* = 0 represents no success in *n* repetitions and *X* = *n* represents *n* successes in *n* repetitions. Across *n* repetitions, the probability of random variable *X* taking a value *x*, that is, *p(X = x)*,is given by the following expression (also known as the probability mass function):

**[INSERT EQUATION]**

.

**[END EQUATION]**

In this expression, *n* and *p* are called the *binomial parameters* because these values completely determine the binomial probability distribution. The mean and variance of the binomial distribution are equal to  and , respectively. When *p* is close to 0.5 and *n* is sufficiently large, the binomial distribution can be conveniently approximated by the normal distribution. The expression  is the number of possible combinations; when *x* items are selected from a set of *n* items, they can be evaluated as

**[INSERT EQUATION]**

.

**[END EQUATION]**

In this formula, *n*! is read as *n* factorial and is calculated as. By definition, 0! = 1. In Excel, the factorial value can be calculated by using the function “=Fact (a number or cell address).”

## [H2] Example 1: Popularity of Seminars

Based on observed data, a healthcare manager has determined that the probability of an in-house professional development seminar being popular among nurses in a particular hospital ward is only 0.5, or 50 percent. The manager wants to calculate the probability that exactly three out of five independently planned seminars will be popular among nurses. Because the total number of independent seminars, *n,* and the probability of success (seminar being popular), *p,* are known and fixed, we can use the binomial probability distribution to find it. With *n* = 5 and *p* = 0.5,

**[INSERT EQUATION]**

.

**[END EQUATION]**

Thus, the probability that exactly three out of the five planned seminars will be popular among nurses is 0.3125, or 31.25 percent.

## [H2] Example 2: Popularity of at Least Three Seminars

Given the scenario described in example 1, what is the probability that at least three out of five seminars will be popular among nurses? To solve this problem, the manager needs to add probabilities of three or fewer than three seminars being popular. Thus,

**[INSERT EQUATION]**



**[END EQUATION]**

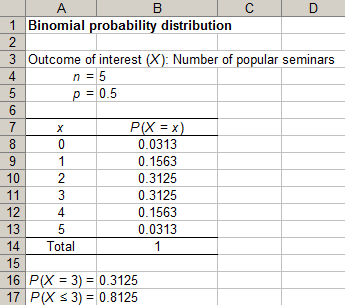
**[H2] Excel Tutorial for Examples 1 and 2**

The problems described in example 1 and example 2 can be conveniently solved with Excel by constructing a binomial probability distribution worksheet. The steps involved in this construction are demonstrated in the following figures. The first step is to specify the values of the binomial parameters, *n* and *p,* in a blank Excel worksheet. Appropriate headings and descriptions may also be added at this stage.

The second step is to construct a table that lists all possible values of binomial random variable *x* and the corresponding probabilities. In exhibit 7.2, cell range A8 to A13 contains all possible values of *x* in an increasing order of magnitude, while cell range B8 to B13 contains the corresponding binomial probabilities. For example, the probability of four out of five seminars being popular among nurses is equal to 0.1563. A check on probability computations is that the sum of the probability column (reported in cell B14) must equal 1. The probabilities required for examples 1 and 2 are reported in cells B16 and B17, respectively, in exhibit 7.2.

**[INSERT EXHIBIT]**

**Exhibit 7.2** Calculating Binomial Probability Distribution in Excel

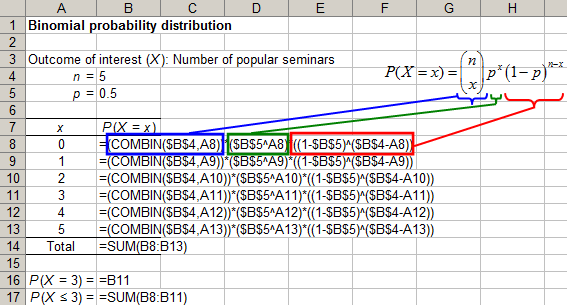


**[END EXHIBIT]**

In exhibit 7.2, each cell in the probability column has a formula embedded in it. These formulas are shown in exhibit 7.3. In order to aid interpretation, the formula in cell B8 has been compared with the binomial probability mass function. Note that the formula in cell B14 simply computes the sum of probabilities in cell range B8 to B13. The use of the $ sign in embedded formulas was necessary in order to type the formula only once in cell B8 and then copy it to cells B9 through B13.

**[INSERT EXHIBIT]**

**Exhibit 7.3** Formulas for Calculating Binomial Probability Distribution in Excel

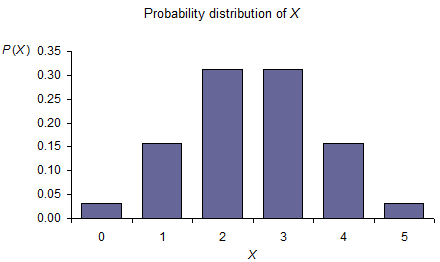


**[END EXHIBIT]**

A bar chart of the binomial probability distribution created with Excel is shown in exhibit 7.4. The distribution is symmetrical, because in this example, the probabilities of success and failure are equal to each other.

**[INSERT EXHIBIT; please render in gray scale; make bars black; remove “probability distribution of X”]**

**Exhibit 7.4** Binomial Probability Distribution



**[END EXHIBIT]**

# [H1] Normal Approximation

In general, the larger the difference between the probability of success and the probability of failure, the more skewed the binomial probability distribution. In exhibit 7.5, we compare two binomial distributions with *p* = 0.3 and *p* = 0.1. For both distributions, *n* = 5. It can be clearly seen that as *p* increases, the binomial probability distribution becomes more and more skewed.

**[INSERT EXHIBIT; please render in gray scale; make bars black; center equations below graphs]**

**Exhibit 7.5** Comparison of Two Binomial Distributions

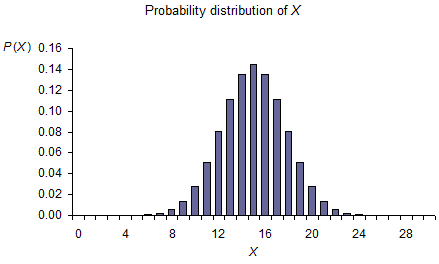
|  |  |
| --- | --- |
| (*a*). *p* = .1, *n* = 5 | (*b*). *p* = .3, *n* = 5 |

**[END EXHIBIT]**

As we noted earlier, when *p* is close to 0.5 and *n* is sufficiently large, the binomial probability distribution can be approximated by the normal distribution. This normal distribution has mean and variance . The approximation may also work when *p* is small, but in that case, a relatively large *n* is required. Many textbooks define *large* as a condition where  and. The binomial probability distribution with *p* = 0.5 and *n* = 30 is shown in exhibit 7.6 to illustrate this idea.

**[INSERT EXHIBIT; please render in gray scale; make bars black; remove “probability distribution of X”]**

**Exhibit 7.6** Binomial Probability Distribution



**[END EXHIBIT]**

When *p* is not close to 0 and *n* is sufficiently large, the binomial distribution can be approximated with the normal distribution. Because the normal distribution is continuous, the probability of an exact value of discrete variable *X* of interest is 0—that is, . In order to avoid this issue, a continuity correction can be introduced. This involves creating an interval of size 1 around the discrete value of interest. For example, if we are interested in calculating the discrete probability, then after applying the continuity, correcting this probability becomes. The next step is to convert values of *X* to the corresponding *z* scores by using the expression

**[INSERT EQUATION]**

.

**[END EQUATION]**

Given that we know that the mean and variance of the binomial distribution are *np* and *np*(1 – *p*), respectively, the expression for *z* becomes

**[INSERT EQUATION]**

.

**[END EQUATION]**

Once the *z* scores, and , have been calculated, the probability can be found as the area under the curve between the two *z* values by reading the standard normal probability table.

## [H2] Example 3: Normal Approximation

In this example, we wish to approximate the binomial distribution described in example 1 with the normal distribution. As before, we estimate the probability as. The first step is to apply the continuity correction to the value of *X*, for which we want to calculate the probability. For *X* = 3, the continuity-corrected corresponding values are 2.5 and 3.5. Thus, we need to calculate the following probability using the normal distribution: . Estimating  and  as  and , respectively, the corresponding *z* scores are 0 and 0.89. Based on the standard normal probability table (see tables for normal distribution on the web), areas under the curve to the left of 0 are 0.50 and to the left of 0.89 are 0.8133, respectively. Thus,

**[INSERT EQUATION]**

.

**[END EQUATION]**

Note that the probability obtained by approximating the binomial distribution with the normal distribution is quite close to the exact probability of 0.3125 that was reported earlier.

## [H2] Excel Tutorial for Example 3

The normal approximation to binomial distribution described in example 3 can be conveniently applied by using Excel. We have shown the required worksheet setup and embedded formulas in exhibit 7.7. The only input required for the setup shown in this image consists of three values: *x*, *n*, and *p*. The probability calculated with Excel is more precise than that calculated from the standard normal probability table as a result of rounding, which is unavoidable with the latter method.

**[INSERT EXHIBIT]**

**Exhibit 7.7** Normal Approximation to Binomial Distribution

|  |  |
| --- | --- |
|  |  |

**[END EXHIBIT]**

# [H1] Inference for a Single Rate

Sample rateis an unbiased estimator of population proportion *p*, and it follows the binomial distribution. When *np* and *n*(1 – *p*) are both greater than or equal to 5, this binomial distribution can be approximated by the normal distribution with mean = *p*, and standard error =. These parameters can be used to test the hypothesis that the population proportion *p* equals the hypothesized value *p*0. The test statistic is

**[INSERT EQUATION]**

.

**[END EQUATION]**

Note that when sampling is done without replacement, the standard error of  needs to be adjusted using the finite population correction factor, like this:

**[INSERT EQUATION]**

 .

**[END EQUATION]**

In this case, the standard error of sample proportion,, is

**[INSERT EQUATION]**

.

**[END EQUATION]**

For a given level of significance,, the confidence interval for *p* has the following form:

**[INSERT EQUATION]**

.

**[END EQUATION]**

## [H2] Example 4: Estimate Population Proportion

In a random sample of 50 hypodermic needles, the number of defective needles is 7. Test the hypothesis that the proportion of defective needles is 10 percent.

**[INSERT UNNUMBERED EXHIBIT]**

|  |  |
| --- | --- |
| Hypotheses |  |
| Level of significance |  |
| Test statistic |  |
| Observed value of test statistic |  |
| Critical value of test statistic | = 1.96 |
| Conclusion: | We fail to reject the null hypothesis and conclude that the proportion of defective needles is not significantly different from 10 percent in the population. |

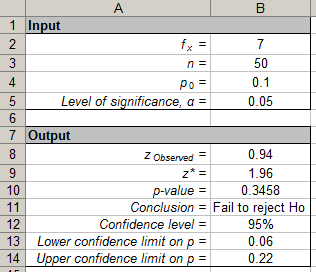
## [END EXHIBIT]

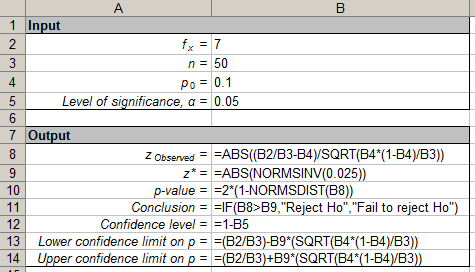
## [H2] Excel Tutorial for Example 4

The test of hypothesis described in example 4 can be conducted in Excel. The required worksheet setup and embedded formulas are shown in exhibit 7.8.

**[INSERT EXHIBIT]**

**Exhibit 7.8** Test of Hypothesis





**[END EXHIBIT]**

# [H1] Comparison of Two Rates

Given sufficiently large independent samplesand, the sampling distribution of difference between their proportionsandthat follow binomial distributions in the corresponding populations is approximately normal with mean, and variance , where

**[INSERT EQUATION]**

.

**[END EQUATION]**

Whenand are unknown and sample sizes are large, the corresponding sample proportions can be used to obtain an estimate of the standard error, where

**[INSERT EQUATION]**

.

**[END EQUATION]**

If it can be assumed that the two population proportions are equal to each other, then and  become estimates of a common population proportion, *pc*, and thus can be combined to produce a weighted mean estimate of *pc*. The expression for  then becomes

**[INSERT EQUATION]**

, where.

**[END EQUATION]**

The test statistic for testing the difference between two population proportions is given by the equation

**[INSERT EQUATION]**

,

**[END EQUATION]**

where andare hypothesized values of  and , respectively. When, the test statistic takes the following form:

**[INSERT EQUATION]**

.

**[END EQUATION]**

When *p* is unknown,

**[INSERT EQUATION]**

.

**[END EQUATION]**

## [H2] Example 5: Test of Differences in Rates

In two independent random samples of 50 and 75 hypodermic needles, the numbers of defective needles are 7 and 15, respectively. Test the hypothesis that this difference is not significantly different from 0 in the population with the following calculation:

**[INSERT EQUATION]**

.

**[END EQUATION]**

**[INSERT UNNUMBERED EXHIBIT]**

|  |  |
| --- | --- |
| Hypotheses |  |
| Level of significance |  |
| Test statistic |  |
| Observed value of test statistic |  |
| Critical value of test statistic | = 1.96 |
| Conclusion | Because, we fail to reject the null hypothesis and conclude that the difference in proportion of defective needles is not significantly different from 0 in the population. |

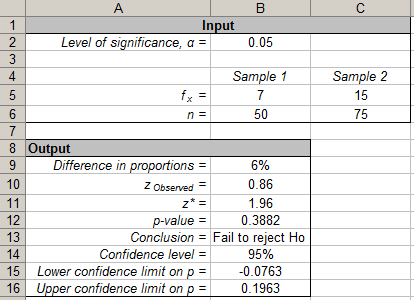
## [END EXHIBIT]

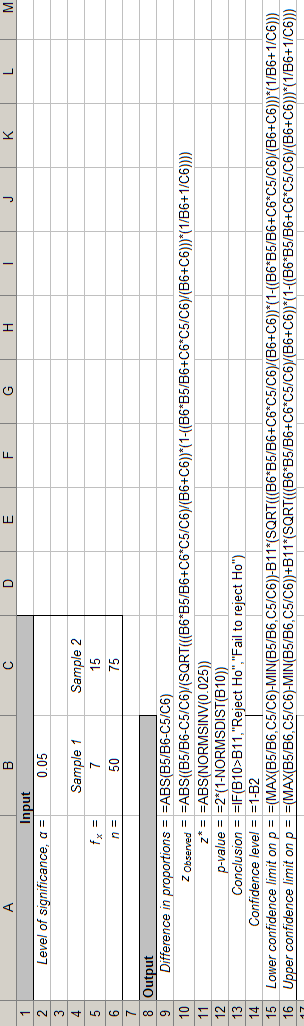
## [H2] Excel Tutorial for Example 5

The test of hypothesis described in Example 5 can be conducted in Excel. The required worksheet setup and embedded formulas are shown in exhibit 7.9.

**[INSERT EXHIBIT]**

**Exhibit 7.9** Test of Hypothesis





**[END EXHIBIT]**

# [H1] Confidence Interval for Odds Ratio

So far we have talked about how to compare two rates and, by extension, the odds of an event such as this:

**[INSERT EQUATION]**

.

**[END EQUATION]**

Odds of an event can also be calculated conditional on occurrences of another event. Suppose we are examining the relationship of two events. We can calculate the odds ratio for the second event as a division of two odds: odds of the second event when the first event occurs and the odds for the second event when the first event does not occur. As a consequence, we have a ratio of two odds, and the change in value of these odds indicates how much the first event contributes to the chances of occurrence of the second event. An odds ratio is a measure of association between one variable and another. In this section, we describe how to calculate a confidence interval for an odds ratio. Suppose we want to understand whether cost overruns at a particular nursing home are associated with severity of the patients’ illness at the time of discharge from our hospital. Exhibit 7.10 shows our data on discharges to various nursing homes.

**[INSERT EXHIBIT]**

**Exhibit 7.10** Consequences of Discharge to Various Nursing Homes

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Nursing Home ABC* | *Other Nursing Homes* | *Total* |
| High score on multimorbidity index | a = 70 | b = 260 | 330 |
| Low score on multimorbidity index | c = 65 | d = 117 | 182 |
| Total | 135 | 377 | 512 |

http://sphweb.bumc.bu.edu/otlt/MPH-Modules/QuantCore/PH717_ComparingFrequencies/ada-reference.gifhttp://sphweb.bumc.bu.edu/otlt/MPH-Modules/QuantCore/PH717_ComparingFrequencies/ada-reference.gif**[END EXHIBIT]**

We start with calculating the odds for Nursing Home ABC as

**[INSERT EQUATION]**

.

**[END EQUATION]**

We also calculate the same odds for other nursing homes to be 2.22. The odds ratio for our referrals is

**[INSERT EQUATION]**

.

**[END EQUATION]**

Odds ratios are not normally distributed and should be transformed by natural log before we calculate the standard error. Once the confidence interval for the log of the odds ratio is calculated, then these values can be transformed to use the same scale as the odds ratio. First, we calculate the natural log of the odds ratio:

**[INSERT EQUATION]**

**[END EQUATION]**

Second, we calculate the standard error of the log of the odds ratio by using the following formula, where *a*, *b*, *c*, and *d* are counts of the cells in exhibit 7.10.

**[INSERT EQUATION]**

**[END EQUATION]**

Third, we calculate the confidence intervals for log of odds ratio as:

**[INSERT EQUATION]**

**[END EQUATION]**

In the last step, we transform the confidence intervals so that they are in the scale of the original odds ratio:

**[INSERT EQUATION]**

.

**[END EQUATION]**

We can conclude from these calculations that our referrals to Nursing Home ABC, compared to other nursing homes, reduced the odds of cost overruns by 0.48. If we repeatedly sample the data, 95 percent of the time, the true odds ratio lies between 0.32 and 0.73.

# [H1] Probability Control Chart

So far, we have focused on rates calculated from one or two samples. In reality, managers want to see if the rates for an event of interest change over time. In this section, we introduce the p‑chart, a control chart used for examining change in rates over time. P-charts are often used to examine the impact of improvement efforts. We assume that you have collected data about a key indicator over several weeks and that you need to analyze the data. Once you create a probability control chart, you can decide if changes have led to real improvements.

# [H2] Assumptions of the P-chart

In the p-chart, we assume that the following five conditions are met:

**[INSERT NL]**

1. The event of interest is dichotomous, mutually exclusive, and exhaustive. *Dichotomous* means that there are only two events. *Mutually exclusive* means that these two events cannot both occur at the same time. *Mutually exhaustive* means than one of these two outcomes must happen. Thus the p-chart may be considered appropriate for analysis of mortality rates if we agree that there are only two outcomes of interest (alive and dead) and that it is not possible to be both alive and dead or to be in a state other than alive or dead. These assumptions are not valid if there is no consensus on what is considered dead or if a stage other than alive or dead can occur.
2. Multiple samples of data are taken over time to track improvements in the process.
3. The observations over time are independent. This means that the probability of adverse outcomes for one patient does not affect the adverse outcome of the other patient. This is not always true. In infectious diseases, one patient affects another. When infection breaks out in a hospital ward, the use of a p-chart to analyze the outcomes of the process is inappropriate, as the observations are not necessarily independent.
4. The patients followed are similar in disease and severity of illness. This is often not true, and adjustments need to be made to reflect the severity of the patients’ illnesses.
5. The analysis assumes that the sample of patients examined represents the population of patients treated during that specific period.

**[END NL]**

These assumptions are important and should be verified before proceeding further with the use of risk-adjusted p-charts. When these assumptions are not met, alternative approaches, such as bootstrapping distributions, should be used.

# [H2] Control Limits for P-chart

We introduce the calculation of p-chart control limits through an example. In this example, we focus on one hospital’s mortality over eight consecutive months. Exhibit 7.11 shows the data we need to analyze.

**[INSERT EXHIBIT]**

**Exhibit 7.11** Monthly Mortality Data for a Hypothetical Hospital

|  |  |  |
| --- | --- | --- |
| *Period* | *Number of Cases* | *Number Dead* |
| 1 | 186 | 49 |
| 2 | 117 | 24 |
| 3 | 112 | 25 |
| 4 | 25 | 3 |
| 5 | 39 | 15 |
| 6 | 21 | 5 |
| 7 | 61 | 16 |
| 8 | 20 | 9 |

**[END EXHIBIT]**

The first step is to create an *x*-*y* plot, for which the *x*-axis is time and the *y*-axis is mortality rates. Calculate mortality rates by dividing the number dead by the number of cases in that month (see exhibit 7.12 for the resulting calculations).

**[INSERT EXHIBIT]**

**Exhibit 7.12** Observed Mortality Rates

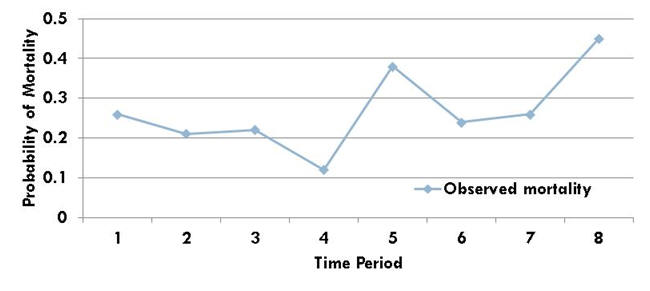
|  |  |  |  |
| --- | --- | --- | --- |
| *Period* | *Number of Cases* | *Number Dead* | *Observed Mortality* |
| 1 | 186 | 49 | 0.26 |
| 2 | 117 | 24 | 0.21 |
| 3 | 112 | 25 | 0.22 |
| 4 | 25 | 3 | 0.12 |
| 5 | 39 | 15 | 0.38 |
| 6 | 21 | 5 | 0.24 |
| 7 | 61 | 16 | 0.26 |
| 8 | 20 | 9 | 0.45 |

**[ENDEXHIBIT]**

Numbers are deceiving, and they can fool you. To understand numbers, you must see them by plotting them. Exhibit 7.13 shows the data plotted against time.

**[INSERT EXHIBIT; don’t bold numbers or labels on either axis. Render in gray scale. Make line black. Remove “Observed Mortality” and sample of line. Simply label line “Observed Mortality” in nonbolded text]**

**Exhibit 7.13** Observed Mortality in Eight Periods



**[END EXHIBIT]**

What does the plot in exhibit 7.13 tell you about these eight months? There are wide variations in the data. It could be that the variation results from random chance or some random measurement error. We cannot be certain whether the apparent improvements result from random chance or real changes in the outcomes. To understand these variations, we can plot two limits on the same chart in such a manner that 95 percent or 99 percent of the points would fall between the lower and upper limits. To set the control chart limits, we need the number of cases and the number of adverse events in each period.

**[INSERT EQUATION]**

cases%20in%20time%20i

adverse%20events%20in%20time%20i

**[END EQUATION]**

In step 1, calculate the average probability of the event across all pre-intervention periods, or if there are no interventions, across all periods. To distinguish this from other calculated probabilities, we call this grand average *p*. This is calculated by dividing the total number of adverse events by the total number of patients. Averaging the rates at different periods will not yield the same results. Calculate the total number of cases and the total number of deaths. The ratio of these two numbers is the grand average probability of mortality across all periods:

**[INSERT EQUATION]**

calculate%20grand%20probability.

Next, calculate the standard deviation of the data. In a binomial distribution, the standard deviation is the square root of grand average *p* multiplied by 1 minus grand average *p* divided by the number of cases in that period. For example, if the grand average *p* is .25 and the number of cases in the period is 186, the standard deviation is the square root of (.25) × (.75) ÷ (186), so

**[INSERT EQUATION]**

standard%20deviation%20of%20probability.

**[END EQUATION]**

Finally, calculate the upper control limits (UCLs) or lower control limits (LCLs) for each period as grand average *p* plus 3 times the standard deviation of the period. This means that you are setting the control limits so that 99 percent of the data should fall within the limits. If you want limits for 90 percent or 95 percent of data, you can use other constants besides 3. The equations would look like

**[INSERT EQUATION]**

UCL%20for%20pchart and

LCL%20for%20pchart.

**[END EQUATION]**

Exhibit 7.14 shows the calculated standard deviations and LCLs and UCLs.

**[INSERT EXHIBIT]**

**Exhibit 7.14** Lower Control Limit and Upper Control Limit

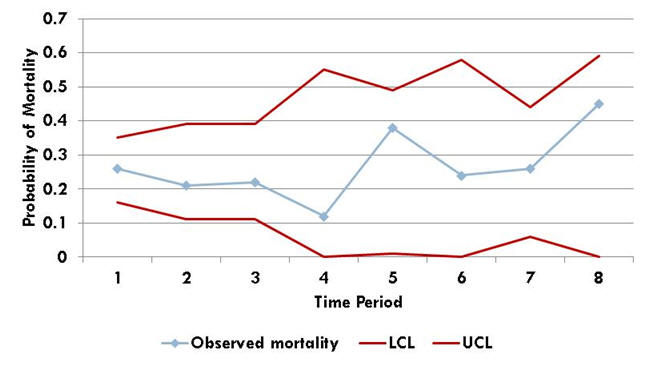
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Period* | *Number of Cases* | *Number Dead* | *Observed Mortality* | *Standard Deviation* | *LCL* | *UCL* |
| 1 | 186 | 49 | 0.26 | 0.03 | 0.16 | 0.35 |
| 2 | 117 | 24 | 0.21 | 0.05 | 0.11 | 0.39 |
| 3 | 112 | 25 | 0.22 | 0.05 | 0.11 | 0.39 |
| 4 | 25 | 3 | 0.12 | 0.10 | 0.00 | 0.55 |
| 5 | 39 | 15 | 0.38 | 0.08 | 0.01 | 0.49 |
| 6 | 21 | 5 | 0.24 | 0.11 | 0.00 | 0.58 |
| 7 | 61 | 16 | 0.26 | 0.06 | 0.06 | 0.44 |
| 8 | 20 | 9 | 0.45 | 0.11 | 0.00 | 0.59 |
| Grand average *p* | | = 0.25 |  |  |  |  |

**[END EXHIBIT]**

Please note that negative control limits in periods 4, 6, and 8 are set to zero because it is not possible to have a negative mortality rate. In addition, the UCLs and LCLs change in each period. This is to reflect the fact that we have different numbers of cases. When we have many observations, we have more precision in our estimates, and the limits become tighter and closer to the average *p*. When we have few observations, the limits go further away from each other. Exhibit 7.15 shows the plot of the observations and the control limits.

**[INSERT EXHIBIT; render in gray scale. remove legend; label upper line “UCL” and make it dashed. label lower line “LCL” and make it dotted. Label middle line “Observed mortality.” None of the numbers or text on the axes should be bold. ]**

**Exhibit 7.15**P-chart for Mortality in Eight Months



**[END EXHIBIT]**

Notice the peculiar construction of the plot, designed to help attract the viewers’ attention to observed rates. The observed rates are shown as single markers connected with a line. Any marker that falls outside the limits can be circled to highlight its unusual nature. The control limits are shown as a line without markers. The UCL and LCL are shown in the same color as their position in the plot, which illustrates which is upper.

The presentation of data is crucial. Make sure that your display of the control chart does not have any of the following five typical errors:

**[INSERT NL]**

1. The chart includes unnamed labels such as “Series 1” and “Series 2.”
2. The markers in the control line were not removed.
3. The *x*-axis is missing a title.
4. The *y*-axis is missing a title.
5. Colors used in the chart and in the cell values do not help in understanding the work—too many or too few colors are used.

**[END NL]**

The control limits help the manager know whether observed changes are real improvements or merely random variations. If we have more mortality than can be expected from chance in a given period, then the process has deteriorated during that period. Any point above the UCL indicates a potential change for the worse in the process. Any point below the LCL indicates that mortality is lower than can be expected from chance. It suggests that the process has improved. In the plot in exhibit 7.15 all data points are within control limits. The process has not changed, and thus the manager can conclude that the rate of mortality is similar to what it was previously.

# [H1] Risk-Adjusted P-chart

P-charts were designed for monitoring the performance of manufacturing firms. These charts assume that the input to the system is the same at each period. In manufacturing, this makes sense. The metal needed for making a car changes little over time. However, the idea makes little sense in healthcare. People are different—in their severity of illness, their ability and will to recover from illness, and their attitudes toward life-saving interventions. These differences affect the outcomes of care. If these differences are not accounted for, we may mistakenly blame the process when poor outcomes were inevitable and praise the process when good outcomes arose from the type of patient.

Some institutions (e.g., tertiary hospitals) receive many severely ill patients. These institutions would be unfairly judged if their outcomes are not adjusted for their case mix before comparing them to other institutions. Similarly, in some months of the year, there are many more severely ill patients. For example, seasonal variations affect the severity of asthma. Holidays affect both the frequency and the severity of trauma cases.

Many process changes lead to changes in the patients attracted to a particular organization. If caregiving staff aggressively try to educate patients to avoid Cesarean section (C-section), their institution may get a reputation for normal vaginal birth delivery and may attract patients with fewer pregnancy complications. In the end, the staff has not really reduced C-sections in its unit; all it has done is to attract healthier patients. Nothing has fundamentally changed in our processes, except for the input.

Risk adjustment of control charts is one method of making sure that the observed improvement does not result from changes in patient type. To help you understand this method of analysis, we will present an example in exhibit 7.16. Suppose we have collected the data in exhibit 7.16 over eight periods. The bottom part of this table shows the patient’s risk of mortality. For example, the risk of mortality in case 1, period 1, was 0.25. A later chapter in this book introduces how regression analysis can be done to calculate the risk of mortality. Chapter 2 introduced the measurement of risk using standard query language (SQL). Whether through regression or SQL, we assume that a measure of risk is available.

**[INSERT EXHIBIT]**

**Exhibit 7.16** Mortality Risks of Individual Patients

|  |  |
| --- | --- |
|  | *Observed Mortality* |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Time 1* | *Time 2* | *Time 3* | *Time 4* | *Time 5* | *Time 6* | *Time 7* | *Time 8* | *Time 9* |
| Number of Deaths | 8 | 6 | 7 | 8 | 5 | 6 | 4 | 5 | 4 |
| Number of Cases | 20 | 20 | 18 | 21 | 20 | 20 | 19 | 20 | 18 |
|  |  |  | *Risk of Mortality* | | | | |  |  |
| *Case* | *Time 1* | *Time 2* | *Time 3* | *Time 4* | *Time 5* | *Time 6* | *Time 7* | *Time 8* | *Time 9* |
| 1 | 0.25 | 0.55 | 0.4 | 0.15 | 0.55 | 0.75 | 0.2 | 0.35 | 0.4 |
| 2 | 0.4 | 0.25 | 0.7 | 0.45 | 0.6 | 0.45 | 0.15 | 0.8 | 0.5 |
| 3 | 0.7 | 0.4 | 0.6 | 0.7 | 0.45 | 0.05 | 0.1 | 0.5 | 0.25 |
| 4 | 0.4 | 0.45 | 0.55 | 0.8 | 0.5 | 0.9 | 0.25 | 0.55 | 0.7 |
| 5 | 0.15 | 0.2 | 0.7 | 0.45 | 0.65 | 0.5 | 0.6 | 0.75 | 0.4 |
| 6 | 0.2 | 0.65 | 0.6 | 0.6 | 0.65 | 0.6 | 0.7 | 0.35 | 0.55 |
| 7 | 0.5 | 0.1 | 0.55 | 0.25 | 0.25 | 0.7 | 0.4 | 0.6 | 0.3 |
| 8 | 0.5 | 0.5 | 0.3 | 0.1 | 0.35 | 0.35 | 0.35 | 0.45 | 0.75 |
| 9 | 0.3 | 0.75 | 0.65 | 0.8 | 0.6 | 0.65 | 0.5 | 0.3 | 0.2 |
| 10 | 0.2 | 0.35 | 0.6 | 0.4 | 0.4 | 0.4 | 0.75 | 0.65 | 0.6 |
| 11 | 0.4 | 0.65 | 0.05 | 0.25 | 0.35 | 0.6 | 0.65 | 0.75 | 0.55 |
| 12 | 0.3 | 0.2 | 0.25 | 0.65 | 0.1 | 0.25 | 0.7 | 0.4 | 0.6 |
| 13 | 0.45 | 0.65 | 0.45 | 0.8 | 0.4 | 0.75 | 0.55 | 0.45 | 0.65 |
| 14 | 0.25 | 0.3 | 0.65 | 0.25 | 0.5 | 0.3 | 0.65 | 0.55 | 0.75 |
| 15 | 0.25 | 0.25 | 0.7 | 0.6 | 0.25 | 0.25 | 0.7 | 0.35 | 0.6 |
| 16 | 0.4 | 0.45 | 0.6 | 0.8 | 0.65 | 0.4 | 0.35 | 0.75 | 0.75 |
| 17 | 0.45 | 0.3 | 0.25 | 0.85 | 0.25 | 0.75 | 0.65 | 0.6 | 0.45 |
| 18 | 0.35 | 0.5 | 0.75 | 0.45 | 0.45 | 0.75 | 0.4 | 0.25 | 0.45 |
| 19 | 0.25 | 0.75 |  | 0.5 | 0.7 | 0.55 | 0.7 | 0.5 |  |
| 20 | 0.1 | 0.6 |  | 0.2 | 0.6 | 0.7 |  | 0.65 |  |
| 21 |  |  |  | 0.45 |  |  |  |  |  |

*Source*: Alemi and Oliver 2001.

**[END EXHIBIT]**

The question we want to answer is whether the observed mortality rate should have been expected from the patient’s severity of illness (the risk of mortality). To answer this question, we need to calculate control limits. Risk-adjusted control limits for probability charts are calculated using the following six steps:

**[INSERT NL]**

1. Calculate the risk of patient *j* in period *i* from an index given in the literature or from a regression of mortality on multiple patient characteristics. Call this .
2. Calculate the expected mortality rate for each period as, where is the number of patients in period *i*.
3. Calculate the expected deviation as .
4. Look up *t* from the Student’s *t*-distribution tables available on the web (e.g., <http://stattrek.com/online-calculator/t-distribution.aspx>).
5. Calculate UCL as
6. Calculate LCL as ; set to zero if negative.

**[END NL]**

The UCLs and LCLs are calculated from the expected risk, *Ei*, the expected deviations, *Di*, and the Student’s *t*-distribution constant. Each of these are further defined and explained in the following sections. The expected mortality rate for each period is calculated as the average of the risks of mortality of all the patients in that period. These calculations are shown in exhibit 7.17.

**[INSERT EXHIBIT]**

**Exhibit 7.17** Observed Mortality Rate

|  |  |  |  |
| --- | --- | --- | --- |
| Period | *Number of Cases* | *Number of Deaths* | *Observed Rate* |
| 1 | 20 | 8 | 0.4 |
| 2 | 20 | 6 | 0.3 |
| 3 | 18 | 7 | 0.39 |
| 4 | 21 | 8 | 0.38 |
| 5 | 20 | 5 | 0.25 |
| 6 | 20 | 6 | 0.3 |
| 7 | 19 | 4 | 0.21 |
| 8 | 20 | 5 | 0.25 |
| 9 | 18 | 4 | 0.22 |
| Total | 176 | 53 |  |
| Grand average *p* | | 0.3 |

**[END EXHIBIT]**

**[INSERT EQUATION]**

**[END EQUATION]**

For example, the observed rate for period 1 can be calculated as

**[INSERT EQUATION]**

.

**[END EQUATION]**

The same calculation should be carried through for each period. These calculations are shown in exhibit 7.18. The same formula can be used to calculate the grand average *p*:

**[INSERT EQUATION]**

**[END EQUATION]**

Before we construct control limits for the expected mortality, we need to measure the variation in these values. The variation is measured by a statistic that we call expected deviation. It is calculated in four steps:

**[INSERT NL]**

1. The risk of each patient is multiplied by 1 minus the risk of the same patient.

2. The multiplied numbers are added for all patients in the same period.

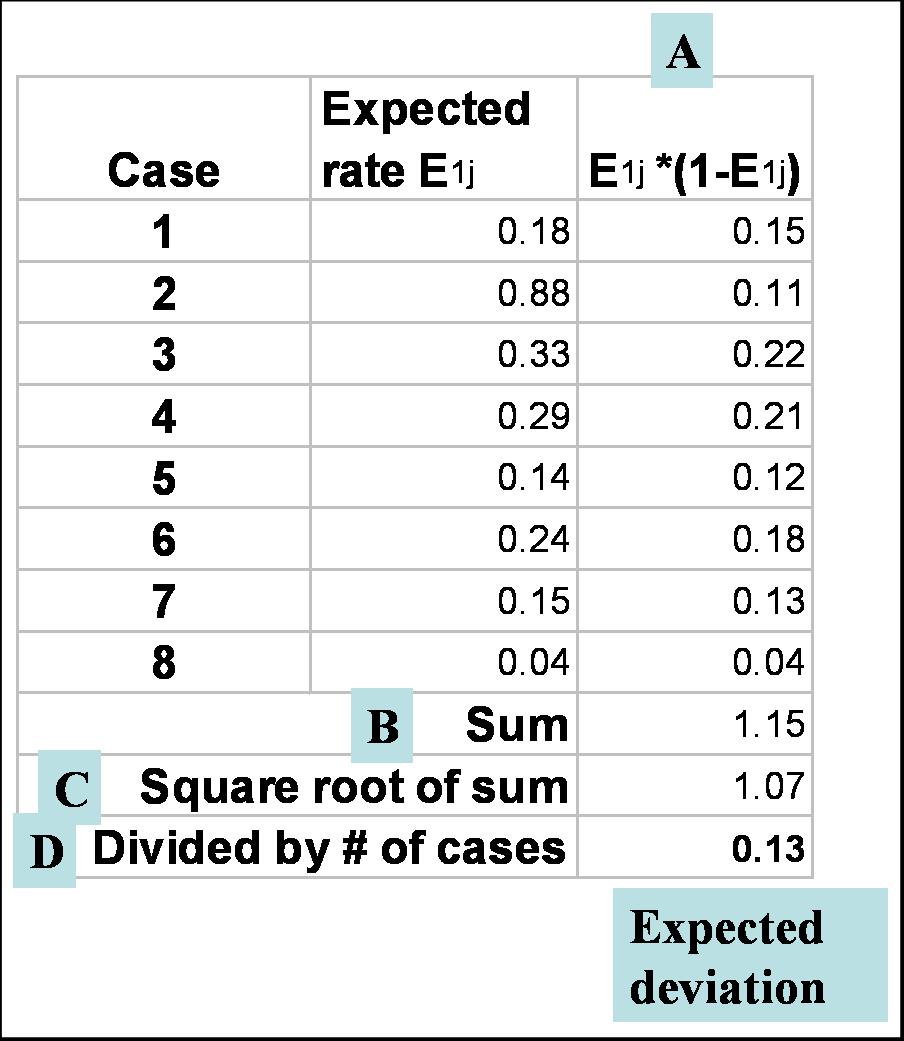
3. The square root of the sum is taken.

4. The expected deviation is the square root of the sum divided by the number of cases.

**[END NL]**

**[INSERT EXHIBIT; render in gray scale. Remove outside border. Make blue medium gray. Remove all bold. Ital column headings. Center righthand and middle column headings. Left-justify content in stub column, including heading.]**

**Exhibit 7.18** Calculation of Expected Deviation



**[END EXHIBIT]**

Exhibit 7.18 shows the calculation of expected deviation for the first period. The same calculation should be carried through for each period, resulting in the data in exhibit 7.19. *Expected rate* is the average of cases in a period. For period 1, the average of cases is 0.34. The same is used to calculate the average of all periods.

**[INSERT EXHIBIT]**

**Exhibit 7.19** Expected Deviation for All Periods

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Time 1* | *Time 2* | *Time 3* | *Time 4* | *Time 5* | *Time 6* | *Time 7* | *Time 8* | *Time 9* |
| Observed rate | 0.40 | 0.30 | 0.39 | 0.38 | 0.25 | 0.30 | 0.21 | 0.25 | 0.22 |
| Expected rate | 0.34 | 0.44 | 0.52 | 0.50 | 0.46 | 0.53 | 0.49 | 0.53 | 0.53 |
| Expected deviation | 0.10 | 0.10 | 0.10 | 0.10 | 0.11 | 0.10 | 0.10 | 0.11 | 0.10 |

**[END EXHIBIT]**

To calculate the control limits, we need to estimate the *t*-statistic that would ensure that 95 percent or 99 percent of data will fall within the control limits. *T*-values depend on the sample size. To see a table of *t*-values for different sample sizes, search the web. Exhibit 7.20 summarizes the estimated *t*-values for all periods.

**[INSERT EXHIBIT]**

**Exhibit 7.20** Estimation of Student’s *t*-Values

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Time 1* | *Time 2* | *Time 3* | *Time 4* | *Time 5* | *Time 6* | *Time 7* | *Time 8* | *Time 9* |
| Observed rate | 0.40 | 0.30 | 0.39 | 0.38 | 0.25 | 0.30 | 0.21 | 0.25 | 0.22 |
| Expected rate | 0.34 | 0.44 | 0.52 | 0.50 | 0.46 | 0.53 | 0.49 | 0.53 | 0.53 |
| Expected deviation | 0.10 | 0.10 | 0.10 | 0.10 | 0.11 | 0.10 | 0.10 | 0.11 | 0.10 |
| *t*-value | 2.09 | 2.09 | 2.11 | 2.09 | 2.09 | 2.09 | 2.1 | 2.09 | 2.11 |
| UCL | 0.55 | 0.66 | 0.73 | 0.71 | 0.68 | 0.74 | 0.70 | 0.75 | 0.74 |
| LCL | 0.13 | 0.23 | 0.31 | 0.30 | 0.24 | 0.32 | 0.28 | 0.31 | 0.31 |

**[END EXHIBIT]**

We are now ready to calculate the control limits and plot the chart (see exhibit 7.21). The UCL and LCL are calculated from the expected mortality and expected deviation, so that 95 percent of the data would fall within these limits (i.e., we use a *t*-value appropriate for 95 percent confidence intervals).

**[INSERT EXHIBIT]**

**Exhibit 7.21** Calculation of UCLs and LCLs

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Time 1* | *Time 2* | *Time 3* | *Time 4* | *Time 5* | *Time 6* | *Time 7* | *Time 8* | *Time 9* |
| 1 | 0.25 | 0.55 | 0.4 | 0.15 | 0.55 | 0.75 | 0.2 | 0.35 | 0.4 |
| 2 | 0.4 | 0.25 | 0.7 | 0.45 | 0.6 | 0.45 | 0.15 | 0.8 | 0.5 |
| 3 | 0.7 | 0.4 | 0.6 | 0.7 | 0.45 | 0.05 | 0.1 | 0.5 | 0.25 |
| 4 | 0.4 | 0.45 | 0.55 | 0.8 | 0.5 | 0.9 | 0.25 | 0.55 | 0.7 |
| 5 | 0.15 | 0.2 | 0.7 | 0.45 | 0.65 | 0.5 | 0.6 | 0.75 | 0.4 |
| 6 | 0.2 | 0.65 | 0.6 | 0.6 | 0.65 | 0.6 | 0.7 | 0.35 | 0.55 |
| 7 | 0.5 | 0.1 | 0.55 | 0.25 | 0.25 | 0.7 | 0.4 | 0.6 | 0.3 |
| 8 | 0.5 | 0.5 | 0.3 | 0.1 | 0.35 | 0.35 | 0.35 | 0.45 | 0.75 |
| 9 | 0.3 | 0.75 | 0.65 | 0.8 | 0.6 | 0.65 | 0.5 | 0.3 | 0.2 |
| 10 | 0.2 | 0.35 | 0.6 | 0.4 | 0.4 | 0.4 | 0.75 | 0.65 | 0.6 |
| 11 | 0.4 | 0.65 | 0.05 | 0.25 | 0.35 | 0.6 | 0.65 | 0.75 | 0.55 |
| 12 | 0.3 | 0.2 | 0.25 | 0.65 | 0.1 | 0.25 | 0.7 | 0.4 | 0.6 |
| 13 | 0.45 | 0.65 | 0.45 | 0.8 | 0.4 | 0.75 | 0.55 | 0.45 | 0.65 |
| 14 | 0.25 | 0.3 | 0.65 | 0.25 | 0.5 | 0.3 | 0.65 | 0.55 | 0.75 |
| 15 | 0.25 | 0.25 | 0.7 | 0.6 | 0.25 | 0.25 | 0.7 | 0.35 | 0.6 |
| 16 | 0.4 | 0.45 | 0.6 | 0.8 | 0.65 | 0.4 | 0.35 | 0.75 | 0.75 |
| 17 | 0.45 | 0.3 | 0.25 | 0.85 | 0.25 | 0.75 | 0.65 | 0.6 | 0.45 |
| 18 | 0.35 | 0.5 | 0.75 | 0.45 | 0.45 | 0.75 | 0.4 | 0.25 | 0.45 |
| 19 | 0.25 | 0.75 |  | 0.5 | 0.7 | 0.55 | 0.7 | 0.5 |  |
| 20 | 0.1 | 0.6 |  | 0.2 | 0.6 | 0.7 |  | 0.65 |  |
| 21 |  |  |  | 0.45 |  |  |  |  |  |
| Observed rate | 0.40 | 0.30 | 0.39 | 0.38 | 0.25 | 0.30 | 0.21 | 0.25 | 0.22 |
| Expected rate | 0.34 | 0.44 | 0.52 | 0.50 | 0.46 | 0.53 | 0.49 | 0.53 | 0.53 |
| Expected deviation | 0.10 | 0.10 | 0.10 | 0.10 | 0.11 | 0.10 | 0.10 | 0.11 | 0.10 |
| *t*-value | 2.09 | 2.09 | 2.11 | 2.09 | 2.09 | 2.09 | 2.1 | 2.09 | 2.11 |
| UCL | 0.55 | 0.66 | 0.73 | 0.71 | 0.68 | 0.74 | 0.70 | 0.75 | 0.74 |
| LCL | 0.13 | 0.23 | 0.31 | 0.30 | 0.24 | 0.32 | 0.28 | 0.31 | 0.31 |

**[END EXHIBIT]**

Because negative probabilities do not make sense, the negative numbers were set to 0. In a risk-adjusted p-chart, we plot the observed rate against the control limits derived from expected values. Exhibit 7.22 shows the resulting chart.

**[INSERT EXHIBIT; render in gray scale. Remove shading in background (make background transparent). Label uppermost line “UCL” and make dashes. Label middle line “Observed rate” and make black. Label bottom line “LCL” and make it dotted]**

**Exhibit 7.22** Risk-Adjusted P-chart for Data in Exhibit 7.21

**[END EXHIBIT]**

More than one of the data points in exhibit 7.22 falls outside the control limits. Points above the UCL limit show periods when outcomes have been higher than expected from the patients’ risks. Points below the control limit show times when outcomes have been lower than expected. In periods 6–9, mortality rates were less than expected.

# [H1] Summary

This chapter has introduced how rates can be compared in two different samples or across time. The first part of the chapter introduced the distribution of rates and the last part introduced p‑charts and risk-adjusted p-charts.

# [H1] Supplemental Resources

A problem set, solutions to problems, multimedia presentations, SQL code, and other related material are on the course website.

## [H1] References

Alemi, F., and D. W. Oliver. 2001. “[Tutorial on Risk-Adjusted P-charts.](https://www.ncbi.nlm.nih.gov/pubmed/11702466)” *Quality Management in Health Care* 10 (1):1–9.