A Statistical Method for Calculating Trail Usage from Counter Data

1/17/17

Stephen Martin, Ph.D.

Note: This method has been superseded by a simpler method [documented here]. However, unlike the methods developed here, the new method calculates total trail usage only, not the distribution of trail usage.

1. Introduction

Automatic counters are used on trails to measure how many people are using the trail. A fundamental question is how to use the counter data to estimate the total trail usage. The answer to this question is not at all obvious or simple.

Generally there are only a few counters on a trail. However, users may be able to enter the trail at numerous access points and may go different distances after entering the trail. This means that some trail users may not be counted at all, while others may pass several counters and be counted multiple times.

This report describes a method for solving this problem. Three approaches (models) for implementing the method were developed and are documented here. The simplest and most usable model is the Simple Usage model. The other two models are more complex and require the use of computer programs.

- The Simple Usage (SU) model is particularly elegant because its solution is a very simple formula. This makes this model easy to use. A remarkable feature of this model is that only the average distance traveled by trail users is required. It is not necessary to know the exact form of the distribution of distances traveled. Using this model, it is shown that the ratio of pedestrians to cyclists is not the same as the ratio obtained by simply counting trail users at a fixed location.

- The Counts to Density (C2D) model is more accurate than the SU model but requires a computer program to evaluate the solution. It assumes that trail users can access the trail at any point along the trail.

- The Counts to Access Points (C2A) model is also more accurate than the SU model but also requires a computer program to calculate the solution. This model assumes that trail users can only enter the trail at specific locations.

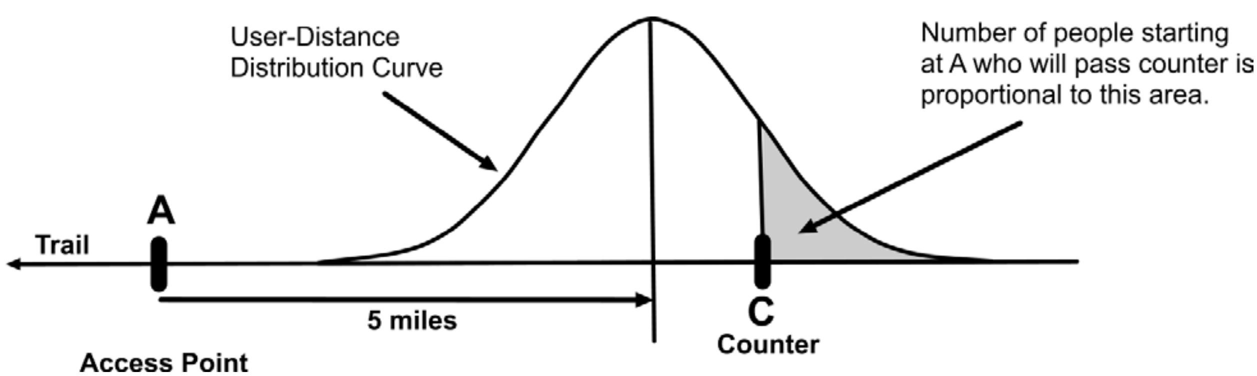
2. The Basic Modeling Concepts

Although the ultimate objective is to calculate usage from the counts on the counters, the method developed here starts by solving the problem the other way around; calculate the count from the usage. We can then use various methods to turn this solution around so that we can solve for the number of users from the counter data.

The User-Distance Distribution

Trail users travel different distances. In statistics this variation is represented using a distribution curve such as the one shown in Figure 1.

Figure 1



Suppose a group of people start from point A on the trail and there is a counter a few miles away at point C. We would like to know how many of those people will pass that counter. As shown, in this example the average user goes about 5 miles. However, some people will go farther. The fraction of people who pass the counter is proportional to the shaded area under the curve to the right of the counter. If the counter is moved farther away from the access point in the above figure, the area will be smaller and fewer people will pass the counter.

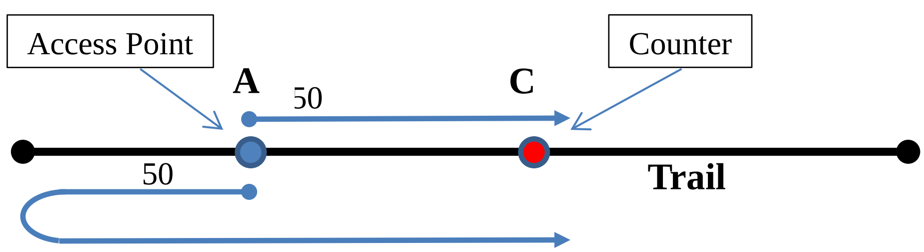
The approximate shape of the user-distance distribution is determined by surveying trail users and asking how far they typically travel on the trail. Once the distance-traveled distribution curve is known, the fraction of users who pass the counter will depend only upon the distance to the counter. This fraction can be calculated using a computer program.

Alternatively, it will be proven that with the SU (Simple Usage) model, we do not need to know the shape of the distribution function. Only the average distance travelled by trail users is required.

Paths from the Access Point to the Counter

Figure 2 shows the trail as a horizontal line. There is an access point (A) and a counter (C). Assume that a group of people, say 100, start at A. It is assumed that half of these people will go toward the counter and the other half go in the opposite direction (50 each way).

Figure 2



Because people go different distances, of the 50 people who started toward the counter, not all of them will make it there. As described above, there is a statistical distribution of how far people go. The fraction of these 50 people who will make it to the counter can be calculated from their travel distance distribution and the distance to the counter.

The other 50 people who went in the opposite direction from the counter eventually reach the end of the trail, turn around, and head back toward the counter. They have to go a longer distance to reach the counter, so fewer of them will reach it. Again, the number of these people who pass the counter can be calculated from the distance that they have to go.

We can add the number reaching the counter from both groups to get the total number of people who will reach the counter from access point A. The total count on the counter at C is obtained by simply adding the contributions from every possible access point.

Discrete vs. Continuous Trail Access

If there are very few access points to the trail then the above method of keeping track of access at specific points is probably appropriate. This will be called the discrete access point approach. On the other hand, if there a many access points it is simpler to just assume that the trail can be accessed at any point along its length. With this continuous access approach it is natural to think and work in terms of the number of accesses (usages) per unit length of trail. This quantity will be referred to here as the usage density. It is a measure of the concentration of people entering the trail at different points along the trail. This is similar to the concept of population density, or people per square mile. For usage density we work with usages per mile of trail.

For example, if 100 people enter the trail along a 1 mile segment of the trail, the usage density would be 100 users/mile. Similarly, if 100 people entered in a ½-mile trail segment the usage density would be 200 users/mile.

This technique of approximating something as continuous is often used in engineering and physics. For example, to simplify computations, materials are modeled as continuous media, even though they actually consist of separate atoms.

In the following sections, two of the models use the continuous access approach (the SU and C2D models) and one of them models access as only occurring at specific access points (the C2A model).

3. The Models

Three statistical models were developed as part of this study. Each model looks at the problem in a slightly different way. The following summarizes the features of each model:

- **Simple Usage (SU) Model:**
  - Access is available at any point on the trail.
  - The trail is very long compared to the distance trail users travel.
  - Access density is the same everywhere.

  - The solution is a very simple algebraic formula.
  - Only the average distance travelled by trail users is required; we do not need to know the shape of the user-distance distribution.

- **Counts to Density (C2D) Model:**
  - Continuous trail access.
  - Unlike the Simple Usage model, this model allows usage to vary over the length of the trail and the trail can be any length.
  - Requires a computer program to solve.

- **Counts to Access Points (C2A) Model:**
  - The purpose of this model is to test whether allowing access at only specific points on the trail makes a difference in the results.
  - The trail can be of any length.
  - Requires a computer program to solve.

The goal of these models is to calculate the quantity and distribution of trail usage from the counter data. All models can only model a linear trail, and therefore only The Legacy Trail and the eastern side of the Venetian Waterway Park are actually modeled. The total trail usage is estimated by assuming that the western Venetian Waterway Park has the same average usage density as the combined Legacy Trail and eastern Venetian Waterway Park.

4. The Simple Usage (SU) Model

For simplicity this model assumes that the trail is infinitely long. It is also assumed that the usage density is the same everywhere. These assumptions imply that every counter will record the same usage count.

Although these simplifying assumptions seem rather restrictive, this model has the advantage that the solution is a very simple equation. It can be shown that the usage density u is given by:

$$u = \frac{C}{(1+r)\mu} \tag{1}$$

where

- C is the count on the counters,
- r is the fraction of users who make a round-trip on the trail, and
- μ is the average one-way distance traveled per usage. For one-way travelers this would be their total distance on the trail. For people who make a round trip on the trail the one-way distance would be half of their total distance.

The quantities r and μ are characteristics of the trail user population. The values of these quantities are obtained from surveys of trail users.

One of the remarkable aspects of this result is that it is not necessary to know the details of the user-distance statistical distribution; only the average distance traveled is required.

The derivation of Equation (1) is somewhat complex. The complete details of the derivation are given in the Appendix.

Example Calculation:

From Reference 1, the yearly totals for counters on The Legacy Trail and eastern part of the Venetian Waterway Park for 2013 are shown in Table 1.

Table 1

Central Sarasota Parkway	81,984
Hatchett Creek	97,827
Oscar Scherer State Park	99,507
Circus Bridge	91,183
Average	92,625

The fact that these counts are relatively close to each other would seem to support the model's assumption that the usage is relatively uniform along the trail.

To apply Equation (1) we need a single representative count number, and therefore it is reasonable to use the average of these four numbers, or

$$C = 92{,}625.$$

Next, from the survey results reported in Reference 2 it was found that the fraction of people making a round trip on the trail was 0.93 and the average distance traveled for all trail users was 8.6 miles. Therefore:

r = 0.93 and μ = 8.6 miles

Plugging these numbers into Equation (1) gives

$$u = \frac{92{,}625}{(1 + 0.93)(8.6)}$$

$$= 5580 \text{ users/mile}$$

The length of the combined Legacy Trail and Venetian Waterway Park is 17.2 miles. Multiplying the above users per mile by this length gives approximately 96,000 users per year.

An Interesting Consequence of the Model: Estimating the Ratio of Pedestrians to Cyclists

One practical use of the SU model is the estimation of the number of cyclists vs. pedestrians on a trail. If a person doing a survey stands at a particular point on the trail and counts four times as many cyclists going by as pedestrians, this does not mean that there are four times as many cyclists on the trail as pedestrians. In fact, there may be more pedestrians than cyclists.

Example

For cyclists, assume

Number counted = 100
Average distance traveled = 9 miles
All make a round trip, r = 1.0

Equation (1) gives: u = 5.6 cyclists per mile

For pedestrians, assume

Number counted = 25
Average distance traveled = 2 miles
All make a round trip, r = 1.0

Equation (1) gives: u = 6.25 pedestrians per mile

Therefore, there are more pedestrians than cyclist on the trail, even though the surveyor counts four times as many cyclists.

In the next section it will be shown that the Simple Usage model under-predicts the number of users for The Legacy Trail. The reason for this is that a large fraction of our users are cyclists, and the trail is relatively short compared to the distance that the average cyclist rides. Therefore the assumption that the trail has infinite length is not quite met.

One of the primary virtues of this model is its simplicity of calculation. It will be shown in Section 7 that the accuracy of the model can be improved by applying a correction factor for trail length. For the combined Legacy Trail and Venetian Waterway Park the correction factor is 1.25 and the corrected formula for calculating usage density becomes

$$u = \frac{1.25C}{(1+r)\mu} \tag{2}$$

5. The Counts to Density (C2D) Computer Model

The Simple Usage model assumed that trail usage did not vary significantly over the length of the trail and the trail was infinite in length. The C2D model was developed to determine how much these modeling simplifications affect the results. With the C2D model the trail usage can vary over the length of the trail and the trail can be of any length. This case cannot be solved with hand calculations. A computer program is required.[1]

[1] Unlike the Simple Usage model, the computer models require a full knowledge of the shape of the distance-travelled statistical distribution. The log-normal distribution was used in all analyses because the distance traveled cannot be a negative number. The mean and standard deviation of the distribution was taken from user survey data.

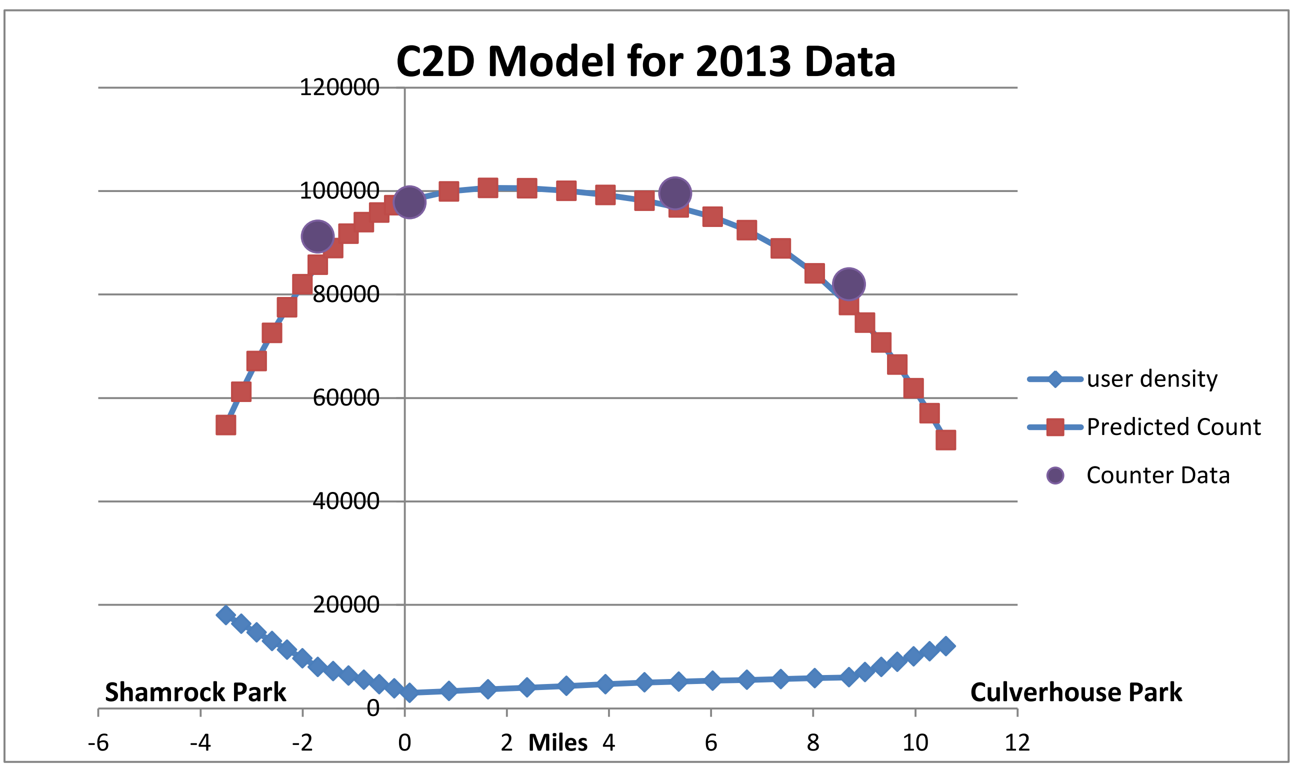
The program works by subdividing the trail into a large number of small segments. The program user starts by making an initial estimate of the shape of the usage density along the trail. The program then calculates the value that counters would record at each point on the trail. This solution is compared to the known counter data at specific points on the trail. If the predicted counts do not agree with the data, the program user adjusts the usage estimate and runs the program again. This process continues until a good fit is obtained between the prediction and the data.

Analysis of 2013 Year Data

Figure 3 shows the prediction of the C2D model for the 2013 counter data given in Table 1[2]. The upper curve (squares) was obtained by manually adjusting the shape of the lower curve (diamonds) until a good fit to the data was obtained.

For this analysis the total number of trail usages for the combined Legacy Trail and Venetian Waterway Park was calculated to be approximately 113,000 usages. This result agrees with the 113,000 users/year originally estimated in Reference 1.

Figure 3

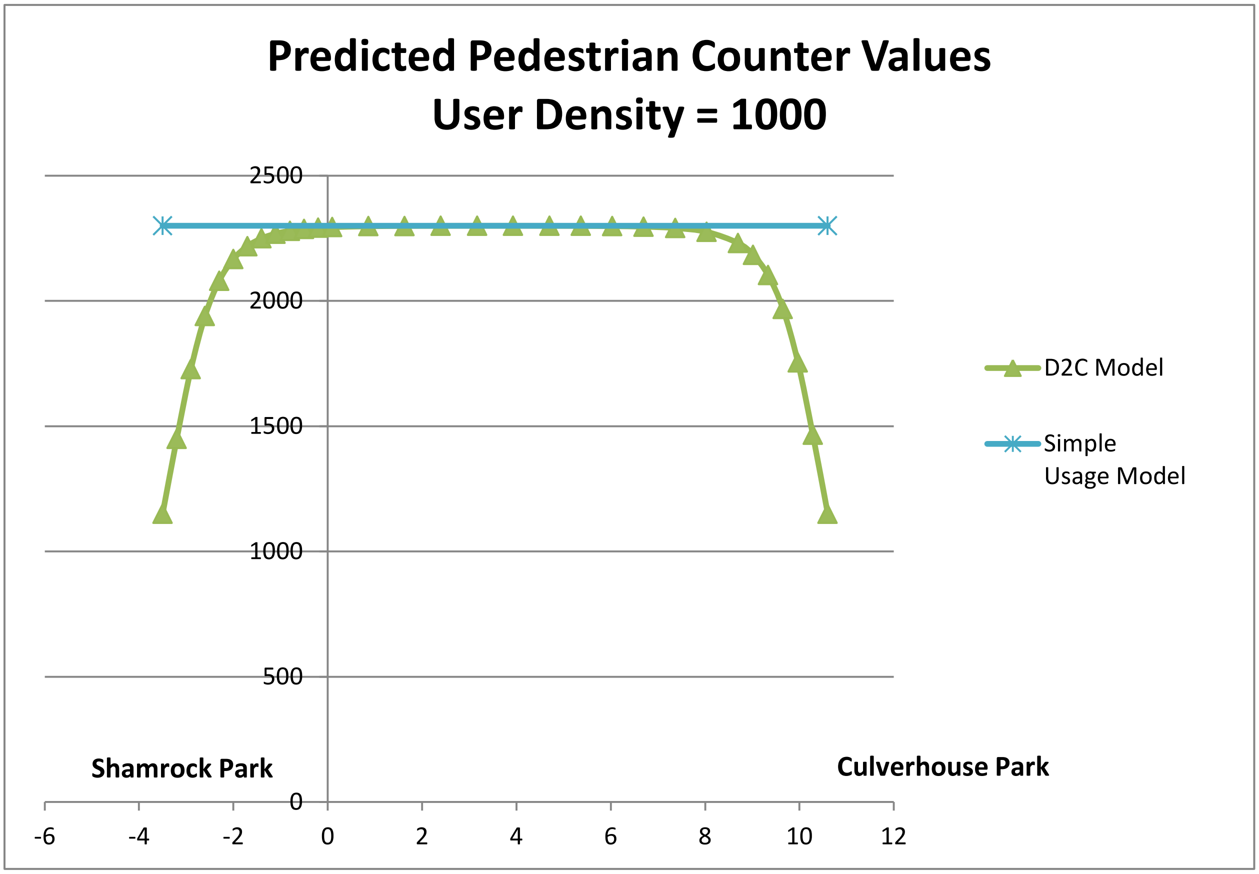


[2] As shown in Reference 2, parameter values for this population are r=0.93, μ =8.6 miles, and standard deviation = 4.7 miles.

Figure 3 shows that counter predictions drop at the ends of the trail. This reason for this is that counters near the ends of the trail only receive traffic from people starting on one side of the counter. Counters near the middle of the trail receive traffic from people starting on both sides of the counter. Thus, if usage along the trail is approximately uniform, counters near the end of the trail would be expected to record only about half the counts of a counter located near the center of the trail. This has the effect of creating a zone near the ends of the trail where the count decreases. The length of this end-affected zone is approximately equal to the average one-way distance traveled by trail users.

To demonstrate this further a special version of the C2D model, called D2C (density to counts), was created which takes user density as input and calculates the resulting counter values. Figure 4 shows the predicted shape of the counter curve if all users are pedestrians and the usage density is 1000 users per mile.

Figure 4

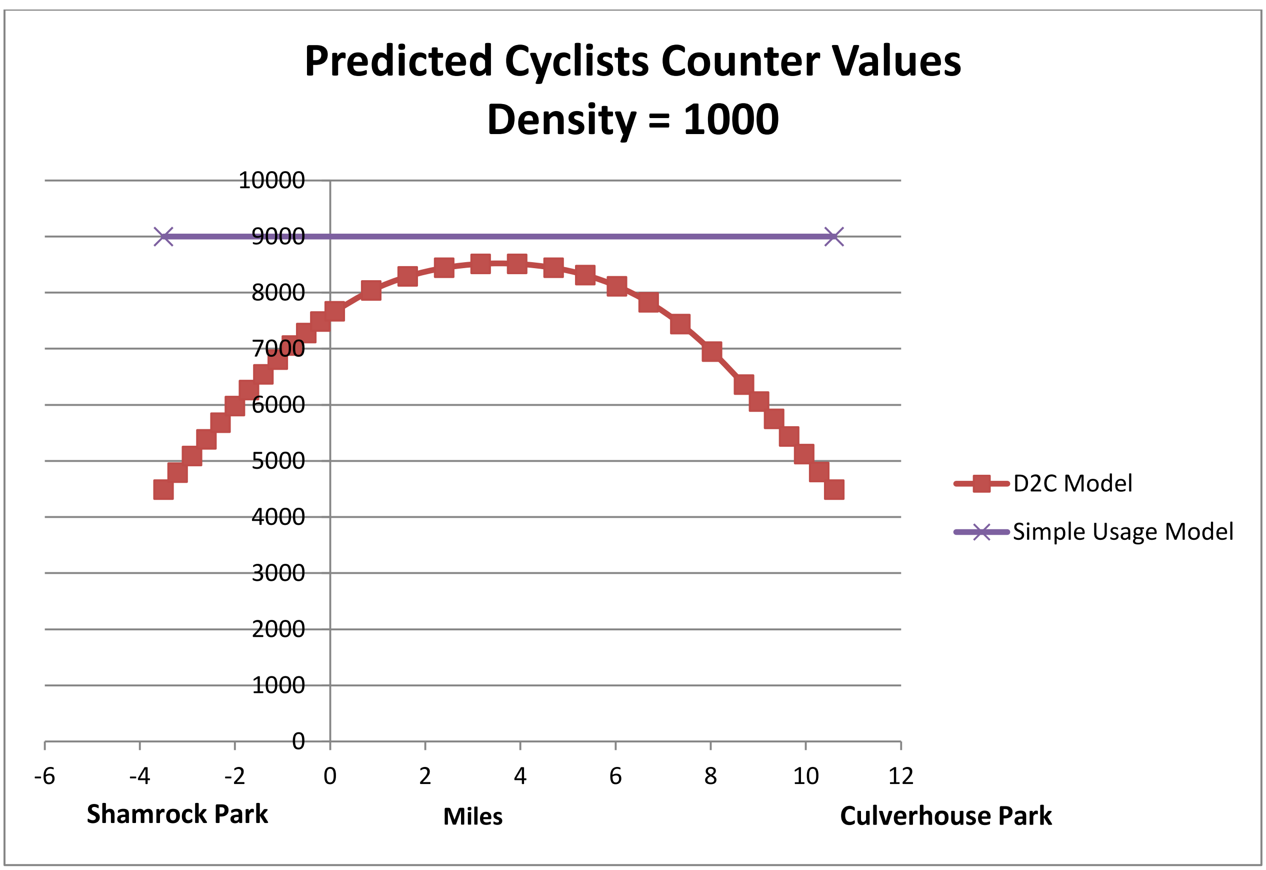


The graph shows a comparison of the count predictions for the D2C model and the Simplified Usage model. The two models agree well in the center region of the trail. However the D2C values decrease as the end of the trail is approached. The effect of this reduction is that for a given set of counter data the Simple Usage model will under-predict total trail usage. The degree of under-counting depends on the size of the end-affected zone.

Figure 5 shows the corresponding graph for cyclists. As shown, the end-affected zones are considerably larger in this case. In fact, counts along the entire length of the trail are reduced by the end effect.

The counts predicted by the SU model are shown as the upper horizontal line in the chart. The counts predicted by the D2C model never quite reach and flatten out along the prediction of the SU model as they did in Figure 4. This means that the SU model will under-predict trail usage more for cyclist than for pedestrians.

Figure 5



6. The Access Point to Counts (C2A) Model

Both the Simplified Usage model and the C2D model assume that the trail can be accessed at any point. The C2A model was developed to test whether allowing access at only specific locations would make a significant difference in the results.

For this model the program user specifies the number and locations of the access points. The user must also specify the fraction of the total number of trail users that enter at each access point.

The program estimates the total number of trail users and compares the predicted counts at the actual counter locations to the measured counter date. The program automatically continues to correct its guess until a good fit to the data is obtained.

Figure 6 shows the 15 access points and their associated usage fractions used in this analysis. This information was obtained from an online survey reported in Reference 2.

Figure 6

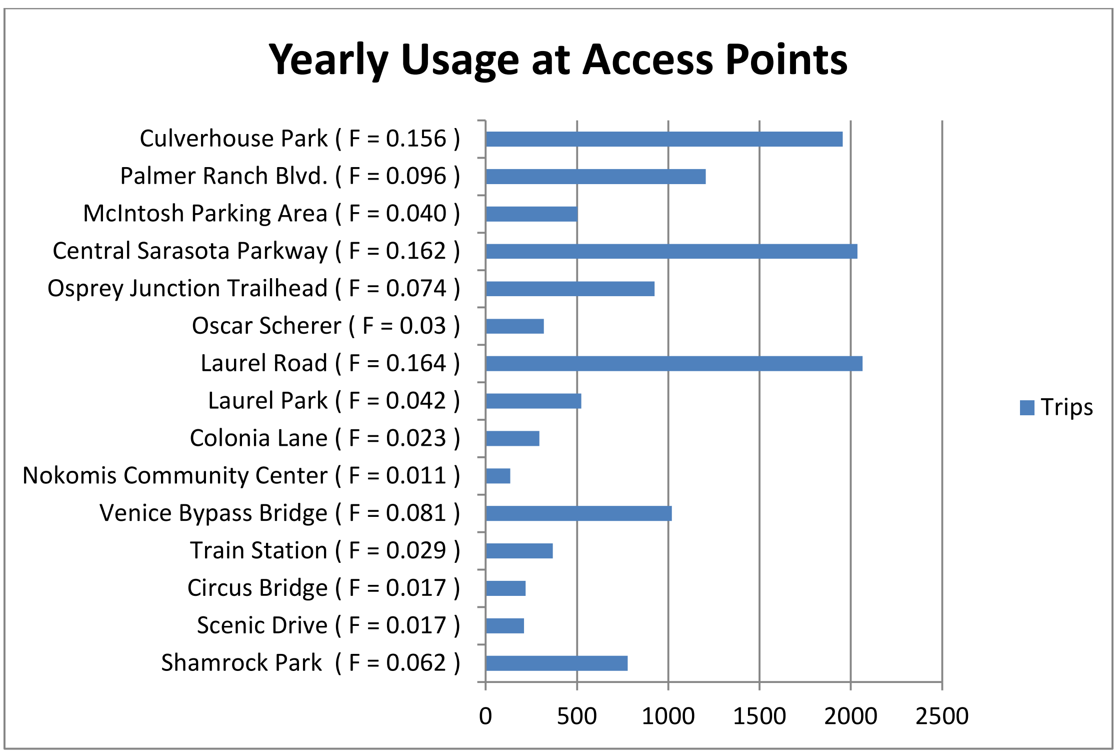
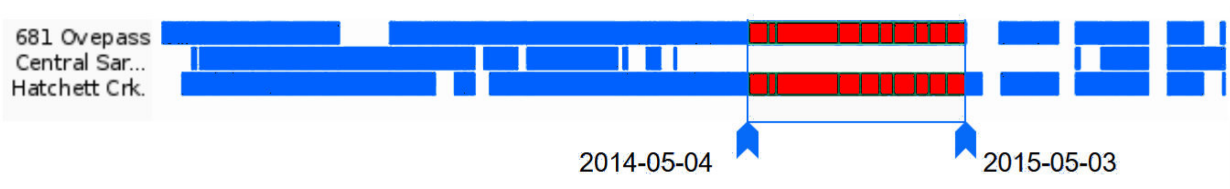


Figure 7 shows the times for which counter data is currently available for counters at the Hatchett Creek, 681 Bridge, and Central Sarasota Parkway stations.

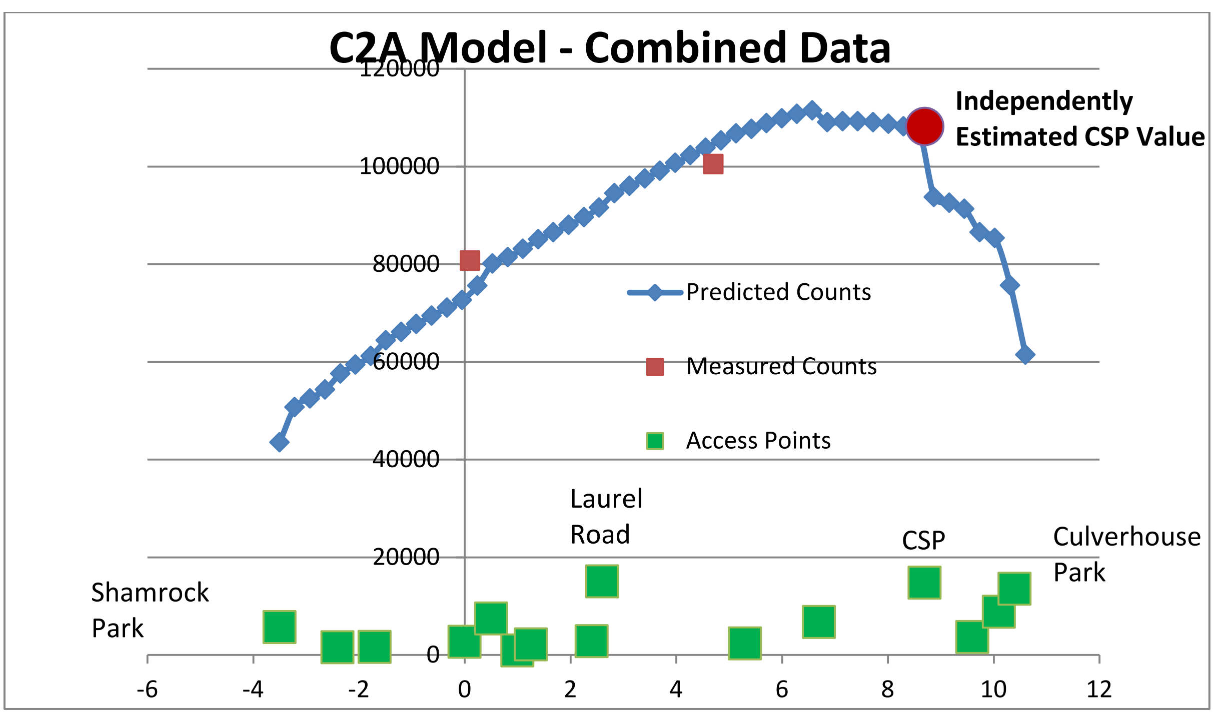
Figure 7 – Available Counter Data



The period from 5/4/2014 to 5/3/2015 was chosen for this analysis because it is the most recent one-year period of continuous data from two of the counters. Only the total year data from Hatchett Creek and the 681 Bridge were used.

Figure 8 shows the results for the combined data; in other words data that includes cyclists, walkers, runners and skaters. The squares at the bottom of the graph show the access points and indicate the number of people starting at each location. For this one-year time period, this analysis predicts a total of 110,000 usages.

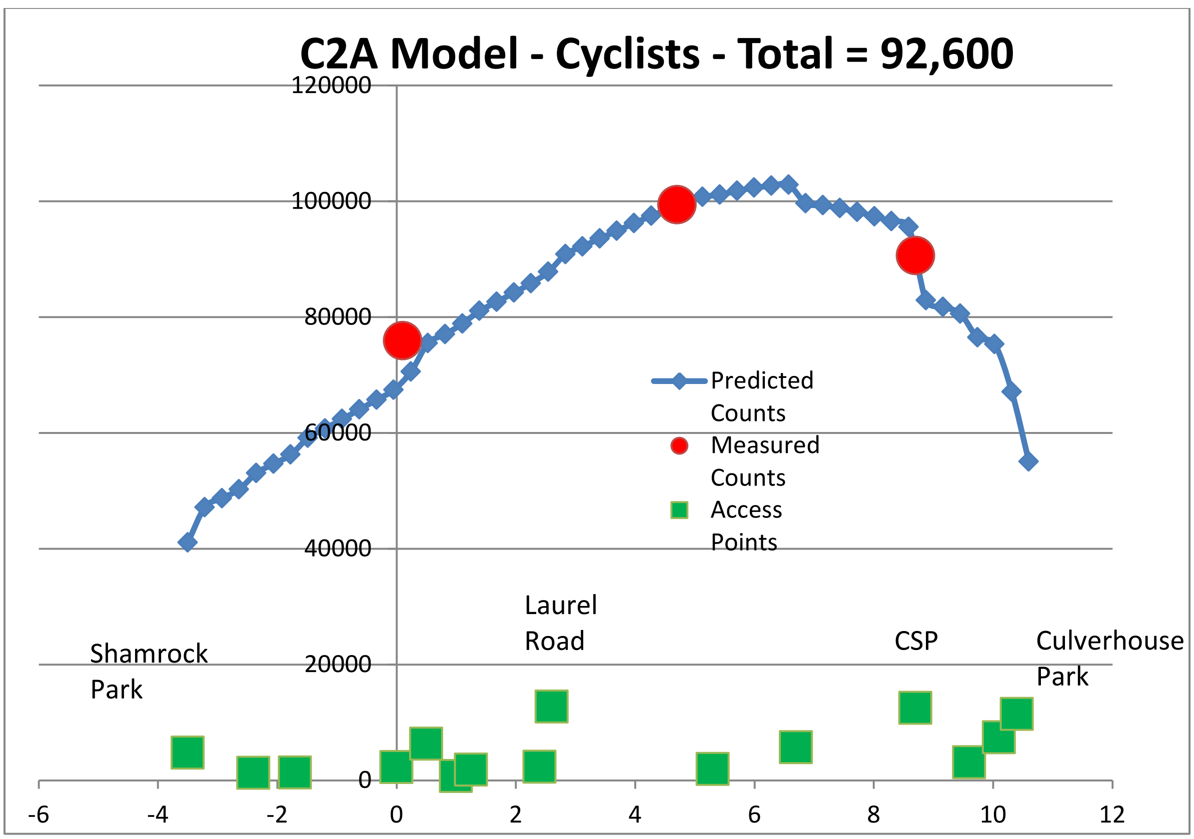
Figure 8



Although no data exists during this time period for the Central Sarasota Parkway station, it is possible to estimate what its value might have been. Data from the survey reported in Reference (2) show that over a 4-hour period the number of trail users at Central Sarasota Parkway was 50.0 per hour, compared with 37.3 per hour at Hatchett Creek. Based on these rates, and the yearly user count at Hatchett Creek, the yearly number of usages at CSP is estimated to be 108,180. This point is plotted as the circle in the upper right of Figure 8. The good agreement between this independently estimated value and the model prediction provides some assurance that the model results are reasonable.

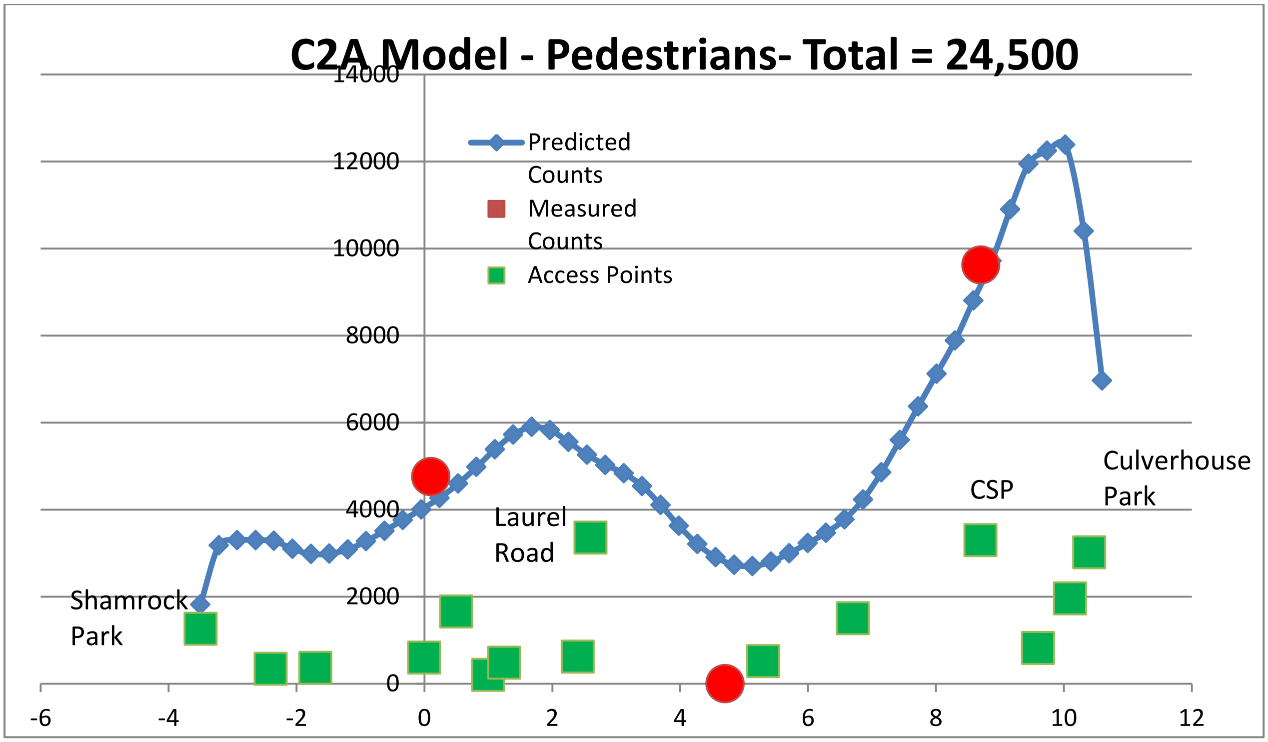
Next, the C2A model was applied to the cyclist and pedestrian populations independently[3]. The results are shown in Figures 9 and 10.

Figure 9



[3] From Reference 2: For pedestrians, average distance = 1.99 miles and standard deviation = 1.05 miles. For cyclists, average distance = 9.42 miles and standard deviation = 4.27 miles.

Figure 10



The cyclist graph is very similar to the combined data graph. The number of cyclists is predicted to be 92,600. The curve for pedestrians is considerably different, having a minimum near the center of the trail. This is caused by the much shorter distances traveled by pedestrians and the fact that usage is low near the middle of the trail. In the trail survey, no walkers were observed at the 681 Bridge during the 2-hour survey period, so this translates into a zero count for pedestrians at this station. The model, however, does not predict this zero value. Instead the predicted count for pedestrians is approximately 2500. To put this error in perspective, it can be shown that if 5 pedestrians had been counted during the 2-hour survey period, there would have been very close agreement between the model and the data. The number of pedestrians counted was small, so the sampling error is much larger than the sampling error for cyclists.

The number of pedestrians was predicted to be 24,500. Therefore, the combined number of pedestrians and cyclists is predicted to be 117,000 for this model.

7. A Simple Approximate Formula for The Legacy Trail and Venetian Waterway Park

We can compare the above results with the results obtained using the Simple Usage model. Using an average counter number of 90,000 (average of 80,000 and 100,000), a round-trip fraction of 0.93, and a mean distance traveled of 8.56 miles. Plugging these numbers into Equation (1) gives:

$$\mathrm{u}=\frac{90{,}000}{(1+0.93)(8.56)}=5448\,users/mile$$

For the combined length of the Legacy Trail and VWP of 17.2 miles, this gives a total user count of 93,700. Comparing this to the 117,000 users predicted by of the C2A model shows that the C2A result is larger than the Simple Usage result by a factor of 1.25. Multiplying the results of the Simple Usage model by this factor will bring them into close agreement with the more accurate results of the C2A model.

Applying this correction factor to Equation (1), using, as above, r = 0.93, u = 8.56 miles, and a trail length of 17.2 miles, we get

$$\mathrm{U}=\left[\frac{(1.25)(17.2)}{(1+0.93)(8.56)}\right]C$$

or

$$\mathrm{U}=1.3C$$

where U is the total number of usages and C is the average of the counts on the trail counters in a given time period. For this approximation to be valid it is assumed that that there is a reasonable distribution of the counters along the trail. It should be kept in mind that this formula is derived assuming the current length of the The Legacy Trail plus eastern Venetian Waterway Park. It is also based on the statistical user behavior reported in Reference 2. If either of these changes, the formula should be modified.

References

1. “Legacy Trail and Venetian Waterway Park User Count Report,” by Steve Christian, January 11, 2014, available at bikevenicefloridа.com.

2. “Results and Analysis of a Survey of Users of The Legacy Trail and Venetian Waterway Park Performed During June and July of 2016,” The Friends of The Legacy Trail Survey Team, 1/17/17, available under Documents at www.friendsofthelegacytrail.org.

Appendix – Derivation of the Simple Usage Model

Counters are used on trails and roads to measure the number of people or vehicles passing a particular location. The problem is to determine how the counts relate to the number of users. This relationship is complicated by the fact that some users may not travel far enough to pass a counter, and others may pass several counters.

The model and analysis in this document show that the trail usage density u (users per mile, for example) is approximately given by

$$u = \frac{C}{(1 + r)\mu} \tag{1}$$

where C is the counter reading, r is the fraction of users making a round trip on the trail, and μ is the mean distance that people travel each way on the trail. A surprising result of the analysis is that only the mean of the distance traveled is required; it is not necessary to know the shape of the distribution function.

The Model

Consider the case of a very long trail with many access points, and assume there are many counters. This situation is shown in Figure 1, where the x axis represents length along the trail and the y axis plots counter readings (upper curve) and usages per mile (lower curve).

For convenience, one can visualize trail users as entering the trail at any point along the side of the trail and then going either left or right along the trail. The counters are represented by vertical lines in the upper graph.

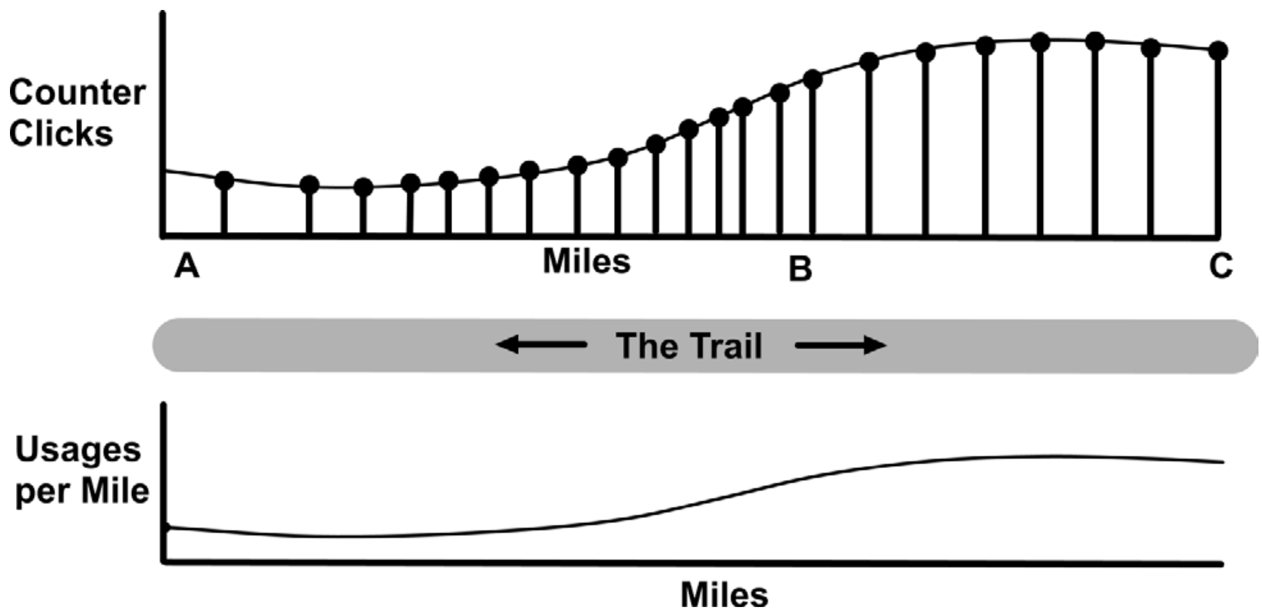


Figure 1

There are several observations that can be made from Figure 1:

- As the number of counters goes to infinity, the counts form a continuous curve.
- In a given region, higher usage per mile translates into higher counter readings. The section A-B might be a rural area, whereas section B-C could be an urban area.
- The longer the trail, the more total users there will be. Therefore, it makes sense to think in terms of usage density (users per mile) rather than total trail usage. Total trail usage can be calculated later from the usage density and the length of the trail.
- The curve of usage density (bottom) determines the counter curve (upper), and conversely, the counter curve determines the usage density curve.
- Over short distances, the counter curve and the usage density are approximately constant.
- The number of counter clicks on a counter will depend upon how far users normally travel. We expect users to hit a nearby counter, but if users travel large distances on average, then users from far away will also hit that counter. Thus, the counter will record users from a larger region and will show a higher count.

Given these observations, a model has been developed. To simplify the model and the mathematics, the following assumptions were made:

- The trail is a single continuous line that is infinitely long.
- All counters record the same number of clicks C in a given time period. In other words, if there were an infinite number of counters, the curve of counter readings would be flat. Therefore, we can talk about the count as a single number C.
- Users can enter the trail at any point along the trail. The number of users entering the trail per mile is the usage density u. The usage density is assumed to be constant along the trail. The objective is to calculate the usage density from the count.
- The distance that people travel on the trail is a random variable with a known statistical distribution. This information can be obtained from user surveys.

Analysis of the Simple Usage Model

In this model we assume that the one-way distance that people travel is a random variable with a probability density function $p(y)$. Since distance traveled is never a negative number, the only restriction on $p(y)$ is that $p(y) = 0$ for $y < 0$.

Figure 2 shows two parallel horizontal axes. The lower one is the x axis and represents locations on the trail. It is assumed that there is a counter located at the origin, x=0, as shown. The upper axis is the y axis and is used to locate points on the probability distribution curve $p(y)$. The x axis increases from left to right, and the y axis increases from right to left. The initial objective is to calculate the count C from the usage density u.

We begin by calculating the portion of the count from users who start from the right of the counter. Referring to Figure 2, consider the users who enter a small segment of the trail of length dx at location x. The number of users in this group is udx.

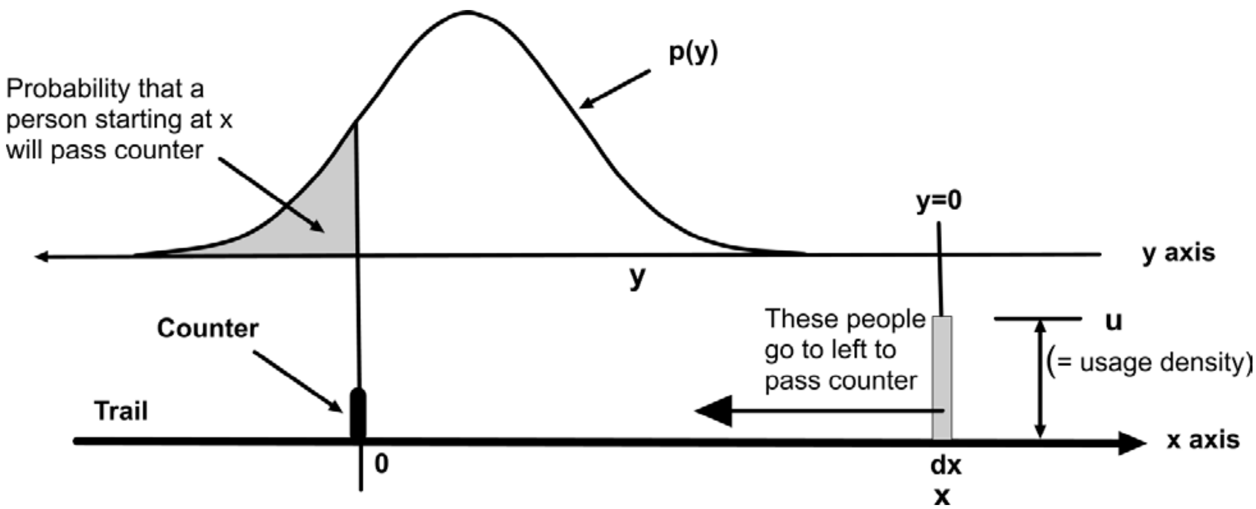


Figure 2

For simplicity, we assume that half the people entering the trail at x turn to the left and head toward the counter, and the other half go right and do not get counted[4].

The number of users from the dx segment that are headed for the counter is then given by

$$n = \left(\frac{1}{2}\right) u dx \qquad (2)$$

Not all of these people will make it to the counter because not everyone travels that far. We assume that the distance that people travel is governed by a distribution function $p(y)$ as shown in the figure. The origin of the y axis is taken to coincide with the location x of the interval dx.

The shaded area under the distribution curve represents the fraction of the people who started toward the counter who will pass it. This fraction F is given by

$$F = \int_x^\infty p(y) dy \qquad (3)$$

---

[4] It can be shown that if a fraction of the users go left, and the remainder go right, the result is the same as the 50% assumption. The change in the number of users coming to the counter from the right is compensated by a corresponding change in users coming from the left.

Next, let dC denote the number of people passing the counter who started in the trail segment dx. Then, using Equation (2) and (3), dC is given by

$$dC = \left(\frac{1}{2}\right) u \left[\int_{x}^{\infty} p(y)dy\right] dx \tag{4}$$

To get the contribution to the count from everyone starting from the right of the counter we integrate Equation (4) with respect to x from $x = 0$ to ∞. This gives

$$\hat{C} = \left(\frac{1}{2}\right) u \int_{0}^{\infty}\int_{x}^{\infty} p(y)dy\, dx \tag{5}$$

Next we change the order of integration and make the corresponding change in the limits of integration to get

$$\hat{C} = \left(\frac{1}{2}\right) u \int_{0}^{\infty}\int_{0}^{y} p(y)dx\, dy \tag{6}$$

It's now possible to evaluate the inner integral to obtain

$$\hat{C} = \left(\frac{1}{2}\right) u \int_{0}^{\infty} yp(y)\, dy \tag{7}$$

Next, for any probability distribution $p(y)$, by definition, the mean μ is given by

$$\mu = \int_{-\infty}^{\infty} yp(y)\, dy \tag{8}$$

For the distribution considered here, $p(y) = 0$ for $y < 0$, and therefore in this case it is only necessary to integrate from 0 to ∞, giving

$$\mu = \int_{0}^{\infty} yp(y)\, dy \tag{9}$$

Replacing the integral in Equation (7) with μ from Equation (9) gives

$$\hat{C} = \left(\frac{1}{2}\right) u\mu \tag{10}$$

Equation (10) gives the total number of people passing the counter who started to the right of the counter. However, it is only counting these people on the outbound leg of their trip. Some of these people will be counted again on their return trip. Let the fraction of people who make a return trip be r. The fraction who go only one way will be $(1 - r)$. The total number of counts from people starting to the right of the counter will be given by

$$C_{right} = (2)r\left(\frac{1}{2}\right)u\mu + (1)(1 - r)\left(\frac{1}{2}\right)u\mu \tag{11}$$

which simplifies to

$$C_{right} = \left(\frac{1}{2}\right)(1 + r)u\mu \tag{12}$$

By symmetry, the people starting on the left side of the counter contribute the same number of counts as those starting on the right, so the total number of counts on the counter is given by

$$C = (1 + r)u\mu \tag{13}$$

Solving for u gives

$$u = \frac{C}{(1 + r)\mu} \tag{14}$$

which is the same as Equation (1).