

Technion - Israel Institute of Technology
The William Davidson Faculty
of Industrial Engineering and Management

Center for Service Enterprise Engineering (SEE)
<http://ie.technion.ac.il/Labs/Serveng/>

SEESat Tutorial



June 2010

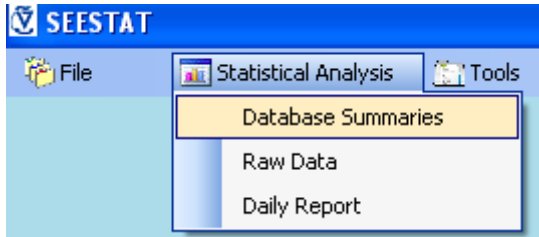
SEESat Tutorial

Part 1	2
Example 1.1: Distributions	3
Example 1.2: Intraday time series.....	10
Example 1.3: Time series (Daily totals)	15
Part 2	19
Example 2.1: Distribution fitting.....	19
Example 2.2: Distribution mixture fitting.....	21
Example 2.3: Survival analysis with smoothing of hazard rates.....	23
Example 2.4: Smoothing of intraday time series.....	25
Part 3	27
Example 3.1: Queue regulated by a protocol (Quick&Raily)	27
Example 3.2: Queue length (Business and Platinum).....	29
Example 3.3: Average wait time (all) and Unhandled proportion.....	31
Example 3.4: Agents on line.....	33
Example 3.5: Daily flow of Totals calls (Tuesday, April 2, 2002).....	35

Part 1

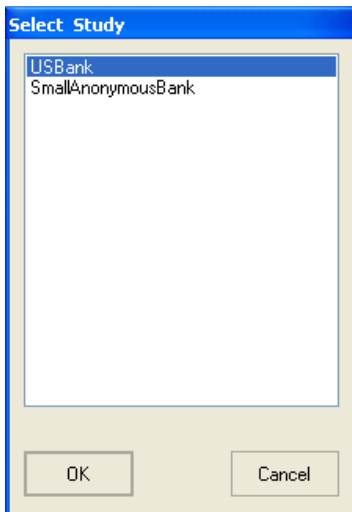
After connecting to the server, click the SEEStat icon to open the program.

On the top of the screen you see the main menu. Click "**Statistical Analysis**". We shall work with "**Database Summaries**". Click it.

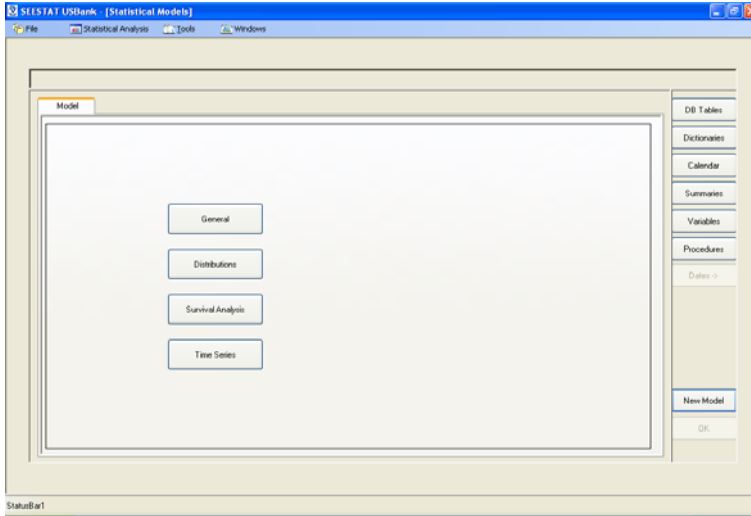


A list-box with names of SEEStat *studies* appears (two databases in our case). Select **USBank** (the database we shall be working with), click "**OK**" and wait for a few seconds.

Background: The source of this example database is a large call center of a US Bank. This call center has sites in 4 states. It has about 900-1200 agent positions on weekdays, and 200-500 agent positions on weekends. Agents process up to 300,000 calls per day. Calls are queued up, when appropriate, in a central queue; they are then served by agents across sites, by fitting service types to agent skills.



Now you see the “**Model**” panel.



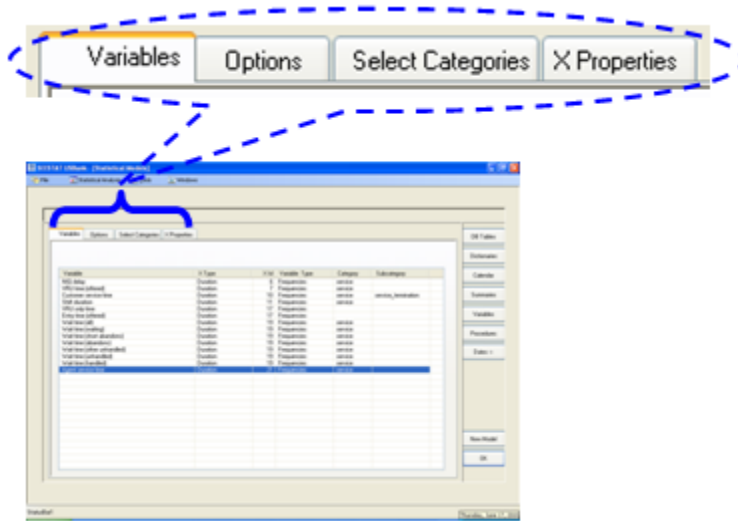
Example 1.1: Distributions

We shall now create a histogram of the distribution of *service time* (duration), at 1-second resolution.

Click the "**Distributions**" button. Three available distribution- models appear. Select "**Estimates**".



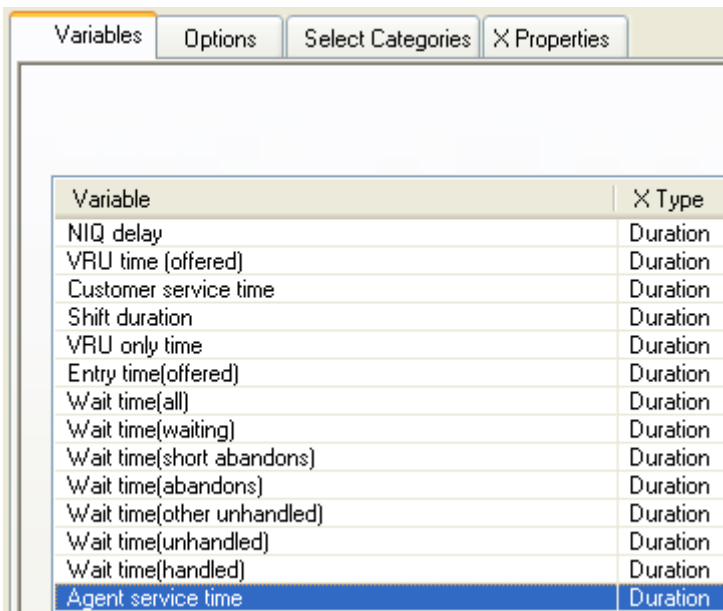
You see the tab control, that has 4 tabs: "Variables", "Options", "Select Categories" and "X Properties".



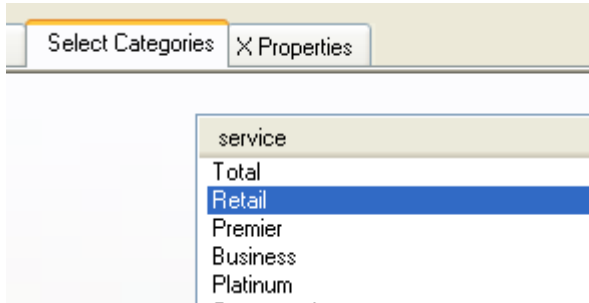
The first one "**Variables**" is active. This tab is *mandatory*, which means you must select variable(s) before moving forward. The three other tabs are *optional*, which means that they already have default values.

NOTE: You can select (click) several variables simultaneously by pressing the **Ctrl** button and, in parallel, clicking on the variables one by one.

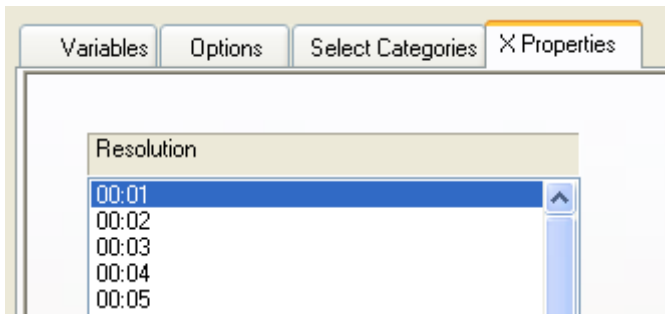
Select "**Agent service time**"(the last entry in the list).



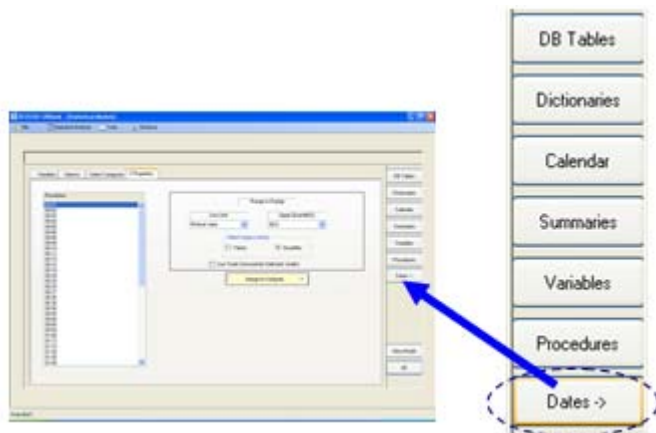
Now click tab **"Select Categories"**. You see a list box with all the service types that are offered by USBank. Select **"Retail"**, which is the Bank's main service.



Open the tab **"X Properties"**. It is used to set properties of charts and tables. At the left side there is a list box **"Resolution"**. The default resolution (bin-size of the histogram) of 5 seconds is marked. Select the minimal resolution **00:01** = 1 second, in order to not miss any details of the histogram.



Now you must select the dates we focus on. Click button **"Dates ->"** at the right side.



You see the list of months for which the requested data are available.

Select **"April 2001"**.

Below the list of months, you see two options for date-selection (Date type): "Aggregated days" and "Individual days". **"Aggregated days"** is a chosen-default, which we now follow.

The screenshot shows a window with two tabs: "Months" and "Days". The "Months" tab is active. It contains a table with two columns: "Month" and "Year". The "Month" column lists months from March to March of the following year. The "Year" column lists years from 2001 to 2002. The row for "April 2001" is highlighted in blue. Below the table, there is a section labeled "Days type" with two radio buttons: "Individual days" (unselected) and "Aggregated days" (selected).

Month	Year
March	2001
April	2001
May	2001
June	2001
July	2001
August	2001
September	2001
October	2001
November	2001
December	2001
January	2002
February	2002
March	2002

Days type

Individual days Aggregated days

Click **"Days"** to make the selection of days, and select **"Week days"** – an aggregation of all 5 working days of the week. (*Holidays and some special days, such as system failure, are excluded*).

The screenshot shows the same window as before, but now the "Days" tab is active. It contains a table titled "Aggregated days" with three columns: "Name", "n days", and "n normal days". The "Name" column lists days of the week and "Week days" and "All days". The "n days" column shows the total number of days, and the "n normal days" column shows the number of normal days. The "Week days" row is highlighted in blue.

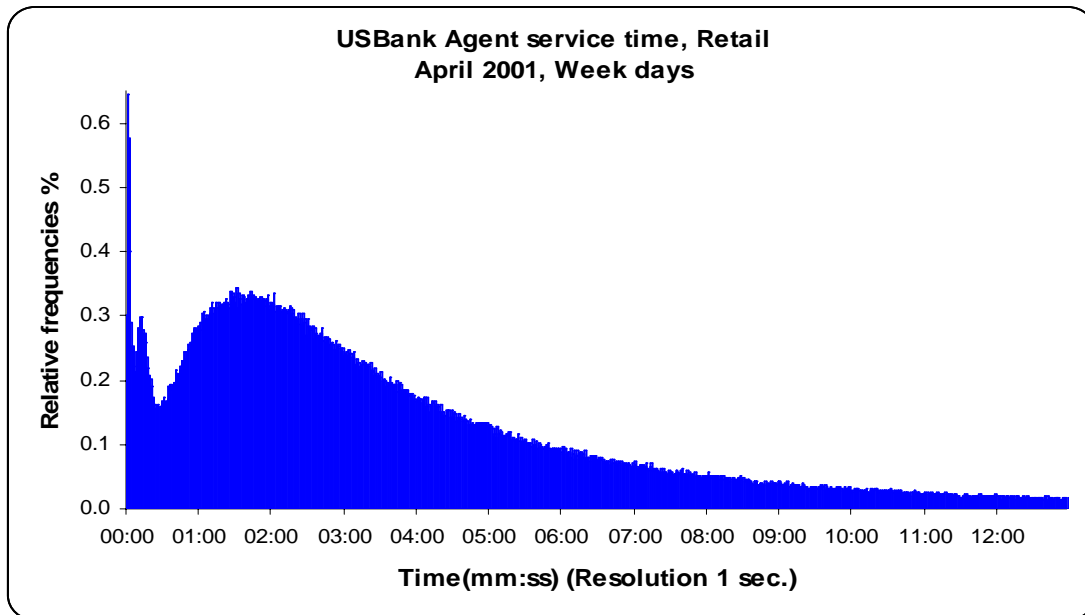
Aggregated days		
Name	n days	n normal days
Mondays	5	5
Tuesdays	4	4
Wednesdays	4	3
Thursdays	4	4
Fridays	4	4
Saturdays	4	4
Sundays	5	5
Week days	21	20
All days	30	29

All selections are have now been completed: click "OK" at the bottom right.



Wait for a few seconds – SEEStat is processing your request: you now see the chart/histogram, produced as "Sheet 2" within an Excel spreadsheet.
NOTE: All the examples in this tutorial will, from now on, be accumulated in this Excel file. DO NOT close this Excel file.

Looking at the chart, you see irregularities at the left (near the origin). We will look at these more carefully later.

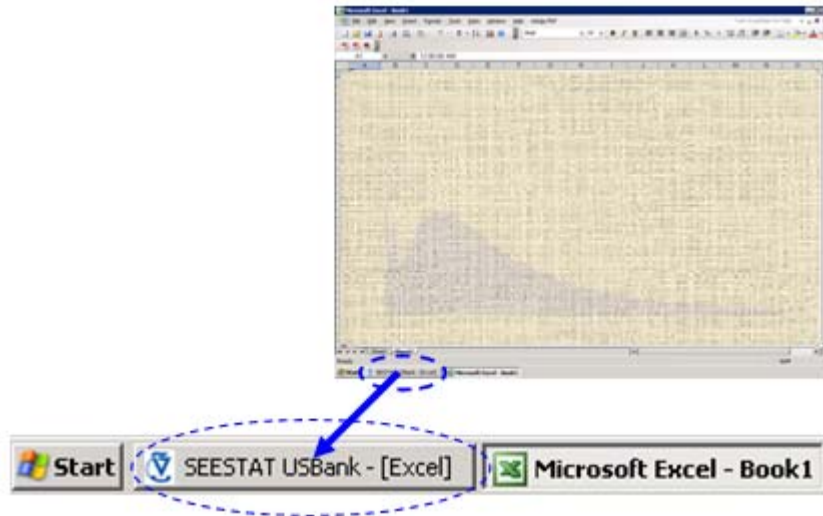


In fact, two sheets have been created: The chart is in “Sheet 2”. "Sheet 1" includes Table(s) that are associated with the chart, with the default one being "Statistics" table. Click [Sheet 1](#) to see the contents of this table (N=619,082 is the number of observations), then return to [Sheet 2](#).

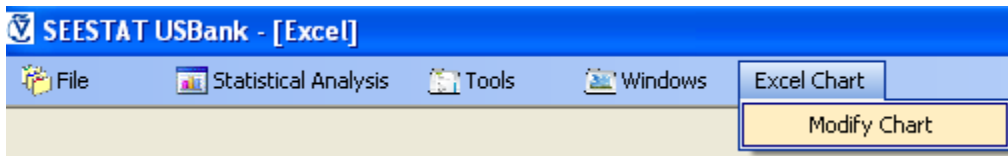
Statistics	
	Agent service time
N	619082
N (average per day)	30954.1
Mean	259.64
Standard Deviation	275.66
Variance	75989
Median	177
Minimum	0
Maximum	3595
Skewness	3.062
Kurtosis	15.38
Standard Error Mean	0.35
Interquartile Range	233
Gini's Mean Difference	NA
Mean Absolute Deviation	182.95
Median Absolute Deviation(MAD)	102
Coefficient of Variation(CV) (%)	106.17

You can easily make modifications to charts and tables, as long as they do not require the loading of new data from the database. You will now go through several such modifications.

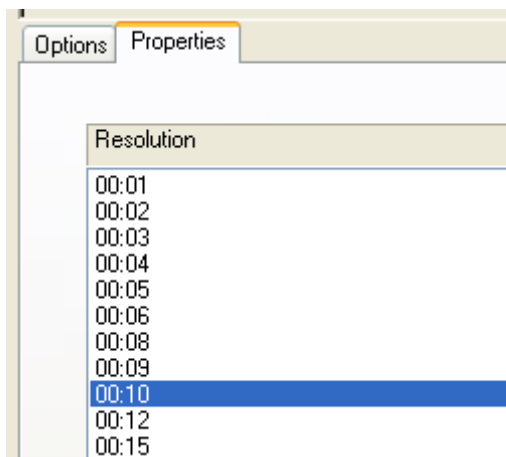
First, return to the SEESTAT main menu by clicking the **SEESTAT USBank** button, on the task bar at the lower-left side of the screen – this is a click that you will be exercising each time that you wish to transfer from Excel to SEESTat.



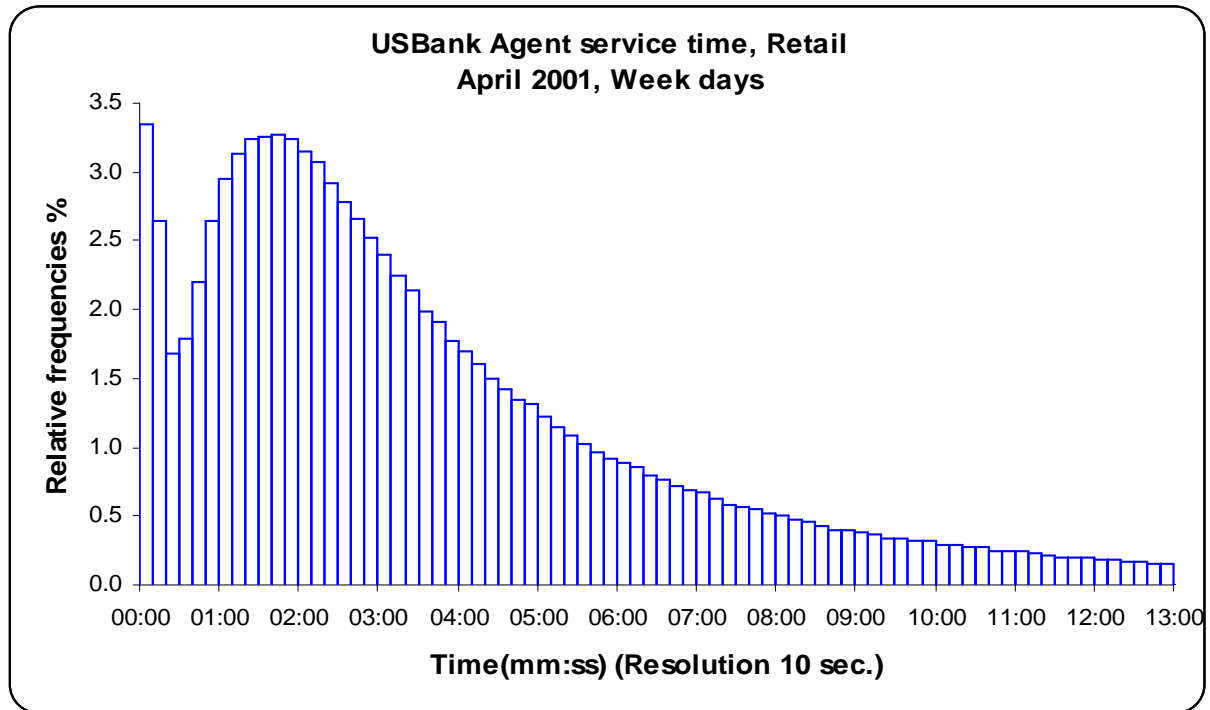
Click **Excel Chart** on the top main menu at the right; after this click **Modify Chart**.



Two tabs are available: "Options" and "Properties". Open **Properties** and change the resolution to **00:10** = 10 seconds.



Click **OK**.

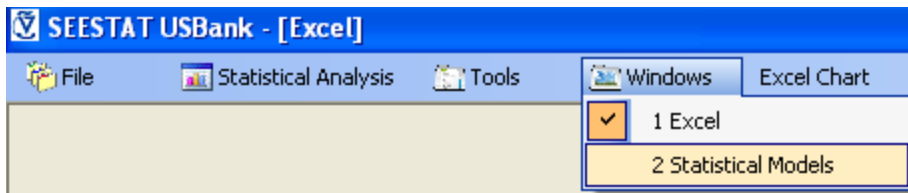


The chart is becoming smoother, but at the cost of losing some details at the left, near the origin.

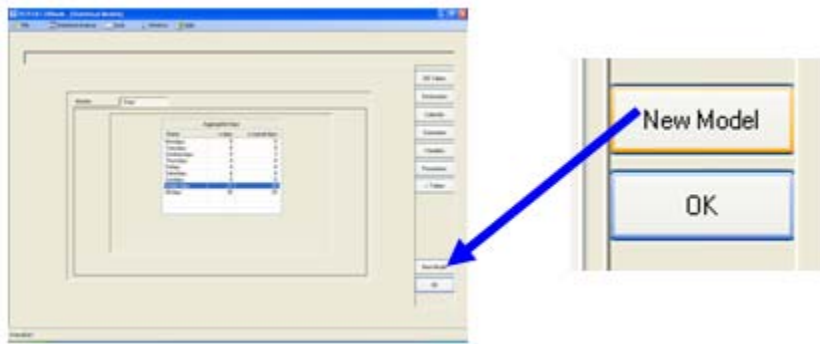
Example 1.2: Intraday time series

We now create a chart for arrival-counts to the call center(s) of USBank, during several days in a September.

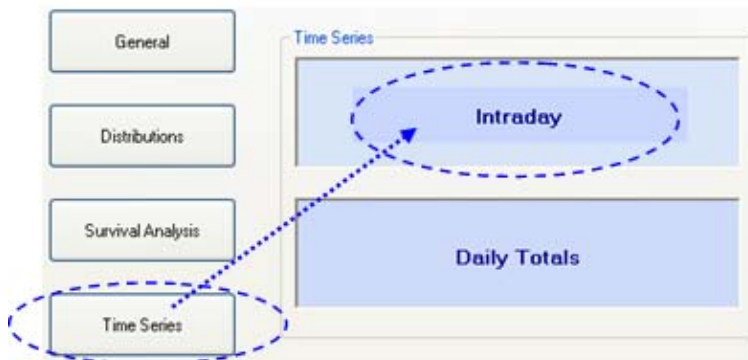
First you must return to the **"Statistical Models"** window. Click the **SEESTAT** button on the task bar (left-bottom), next click **"Windows"** on the main menu (at the top) and select **"Statistical Models"**



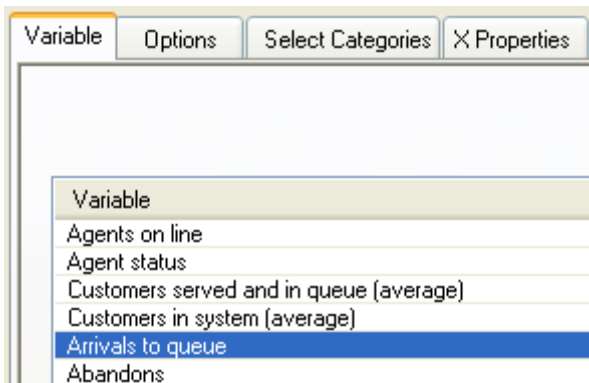
We are now changing models. To this end, select the **"New Model"** button (bottom-right).



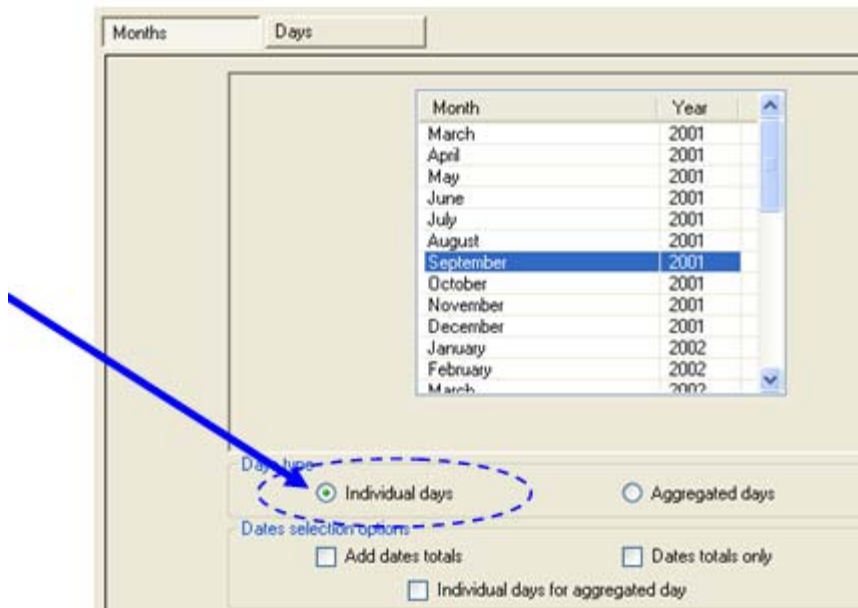
Select now **"Time Series"** and then select **"Intraday"**.



As in [Example 1.1](#), four tabs appear. In tab **"Variable"**, select **"Arrivals to queue"**.



Now select dates: Click "**Dates ->**"; Select from "**Months**" the month of **September 2001**; Mark "**Individual days**", and click the "**Days**" tab.



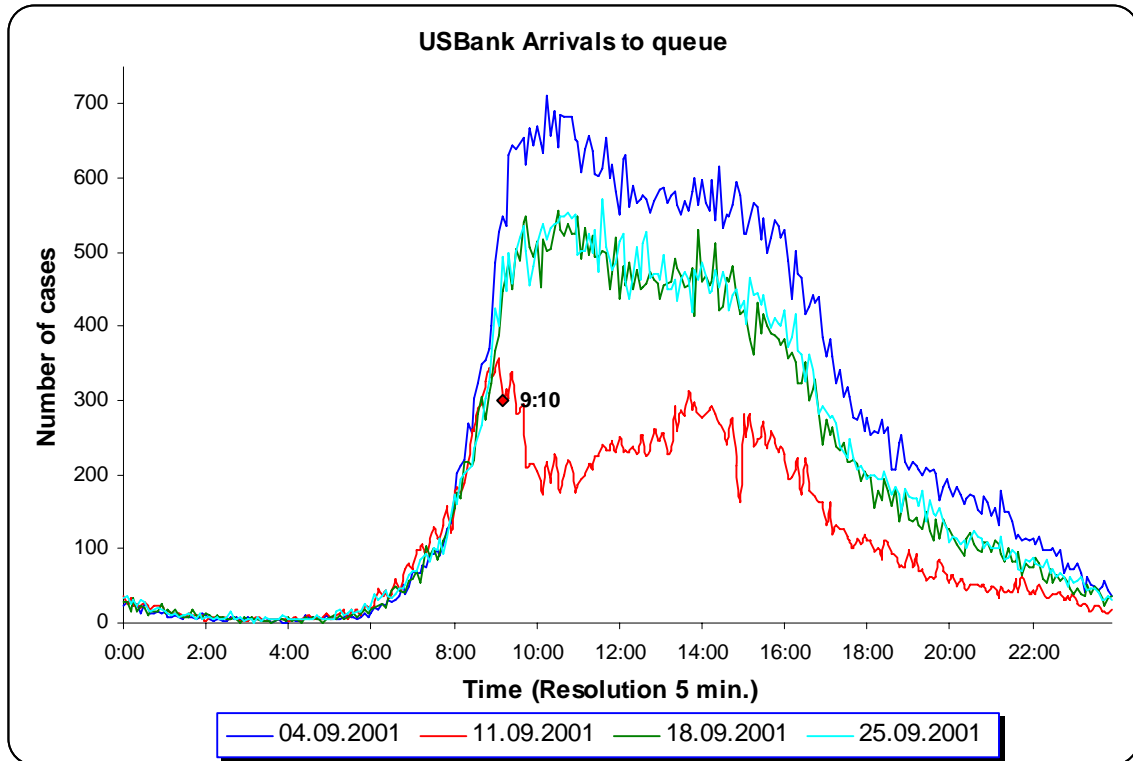
The list of days contains the date, the name of week-day and comments if any. For example, Monday, September 3rd, was Labor Day.

It is expected that the Tuesday following a holiday will be a busy day. We thus compare all Tuesdays of the month: September 4, September 11, September 18 and September 25.

Hold down the "**Ctrl**" key on the keyboard, and in parallel click, one after one, on the **four Tuesdays** of September 2001.

Day	Month	Year	Week Day	Comments
1	September	2001	Saturday	
2	September	2001	Sunday	
3	September	2001	Monday	Labor Day
4	September	2001	Tuesday	
5	September	2001	Wednesday	
6	September	2001	Thursday	
7	September	2001	Friday	
8	September	2001	Saturday	
9	September	2001	Sunday	ShutDown from 6:30
10	September	2001	Monday	
11	September	2001	Tuesday	
12	September	2001	Wednesday	
13	September	2001	Thursday	

Then click "**OK**" (bottom right).



Note: The graphs appear in “Sheet 4” of Excel. As before, “Sheet 3” contains the corresponding numerical data.

*You see a sharp drop in the number of calls around 09:00 AM on **September 11, 2001** – which of course is not surprising!*

You also see that the Tuesday after Labor Day is indeed a heavily-loaded Tuesday, as anticipated.

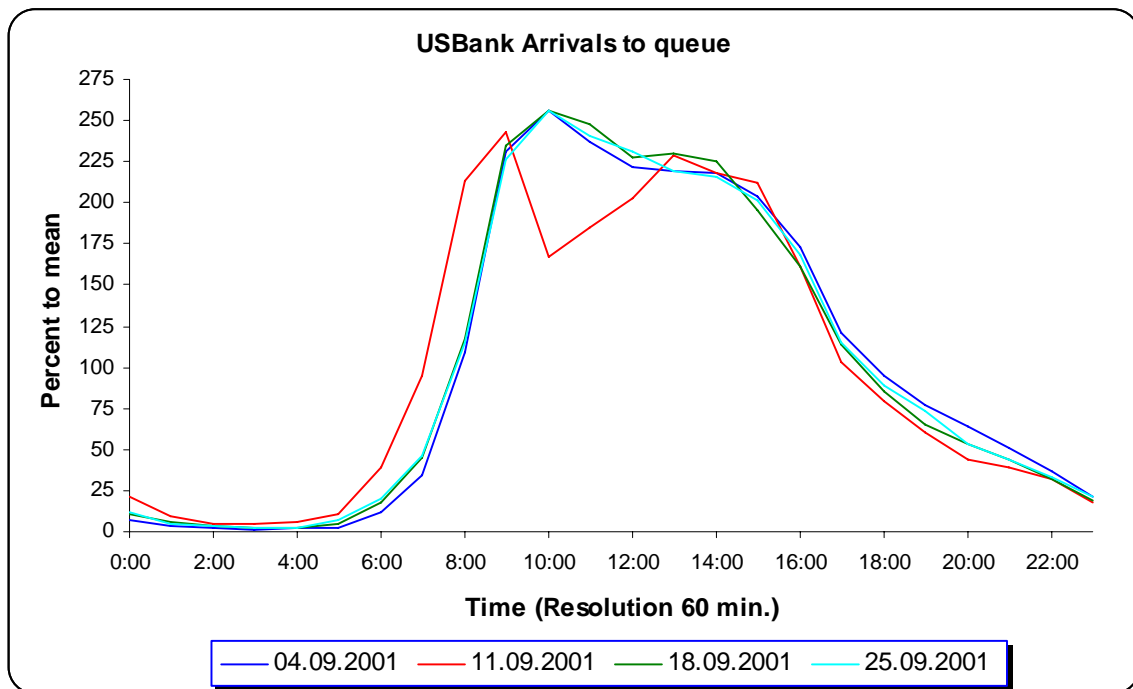
The chart is noisy, due to its 5 minute. resolution. We shall momentarily increase the resolution to 1 hour (60 minutes). We also note the following:

On the two Tuesdays after September 11, the number of calls is low, relative to the Tuesday after Labor Day. A natural question now arises: Is there a “shape of a Tuesday”? To seek a common pattern for (the shape of) a Tuesday, if there is any, we change the graphs from absolute counts to “percent to mean” (mean = average number of calls per resolution period).

Go back to the main menu via the **SEESTAT** tab (bottom-left). In the main menu select **"Excel Chart"** then **"Modify Chart"**. In the **"Options"** tab, in the **"Convert to"** table on the left, select **"Percent to mean"**, and on the **"Properties"** tab set resolution to **60:00** = 1 hour,



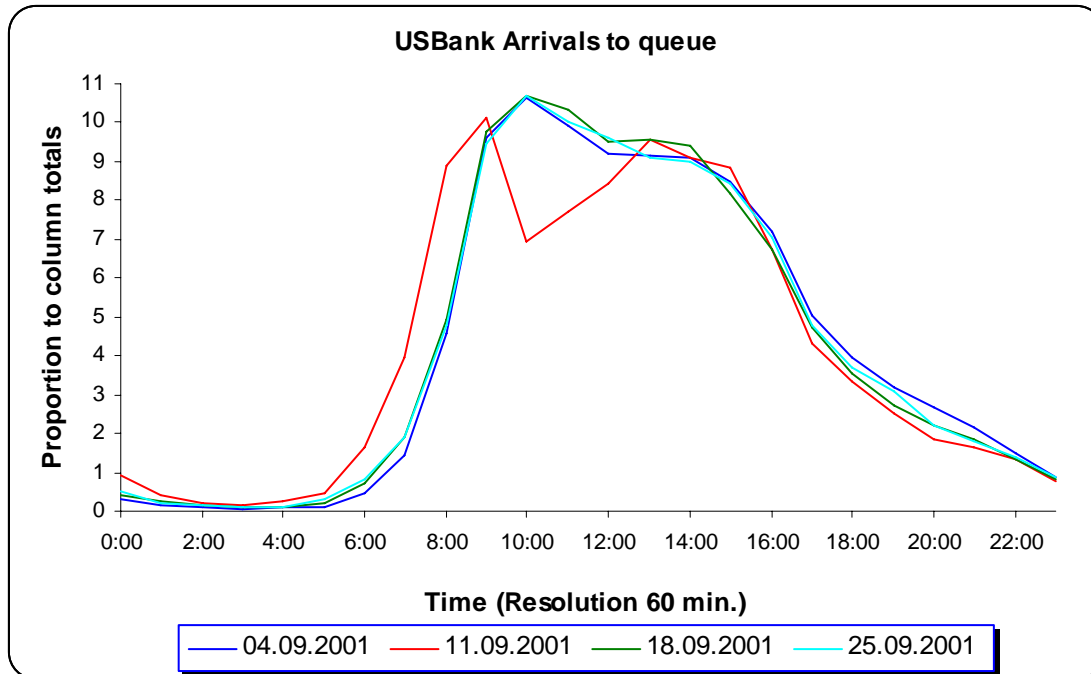
Click **"OK"**.



The "Shape of a Tuesday" is clearly manifested: the distribution of calls over the day is almost the same for the three Tuesdays, both normal and heavily-loaded. (Surprisingly, September 11 also catches up from around 13:00 or so.) For example, the arrival rate during the peak hour – from 10:00-11:00 – is about 2.5 times that of an average hour.

*Instead of "Percent to mean", you can plot according to **"Proportion to column totals"** which, in simple words, means the "hourly fraction of load".*

Going via the "SEESTAT tab", **"Excel Chart"**, **"Modify Chart"** **"Proportions to column totals"**, then **"OK"**.

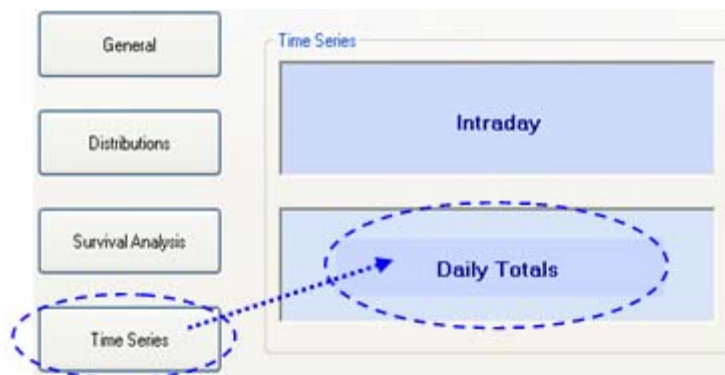


You see that the arrivals during the peak hour 10:00-11:00 constitute 10% of the daily total. (Such observations make load-predictions much easier: indeed, only the daily total must be predicted. Once the daily total is determined, the number of arrivals per hour is allocated according to the shape of the day; e.g. 10% allocated to 10:00-11:00.)

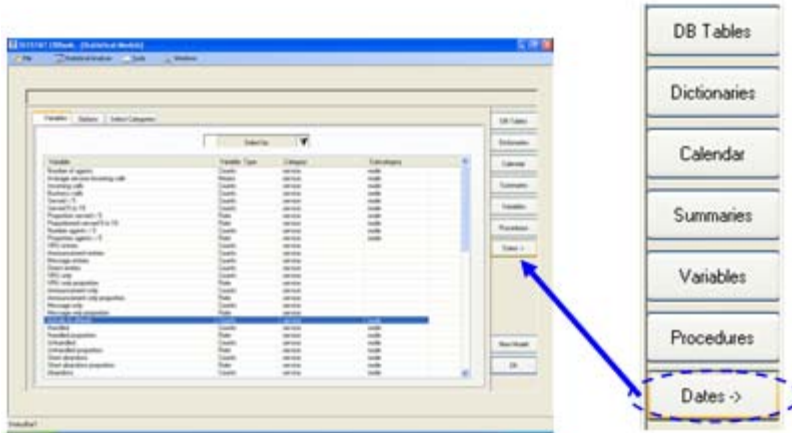
Example 1.3: Time series (Daily totals)

There are two types of daily-totals time-series: individual days during a specific month and aggregated days by months. We now demonstrate these concepts.

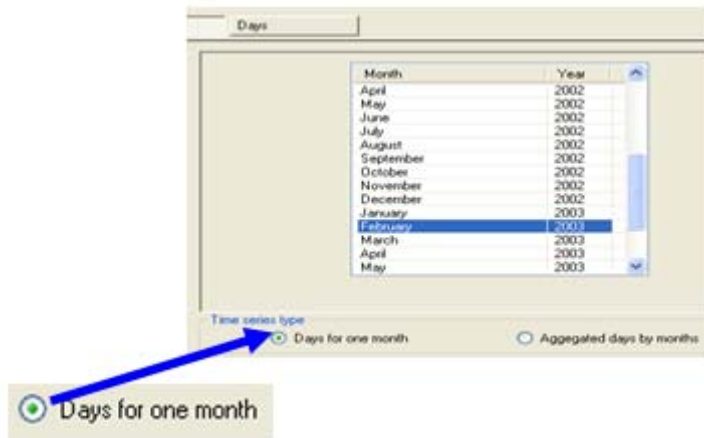
Return to the **"Statistical Models"** window, via **SEESTAT** and using **"Windows"** on the main menu. Press the **"New Model"** button. Select **"Time Series"**, then **"Daily totals"**.



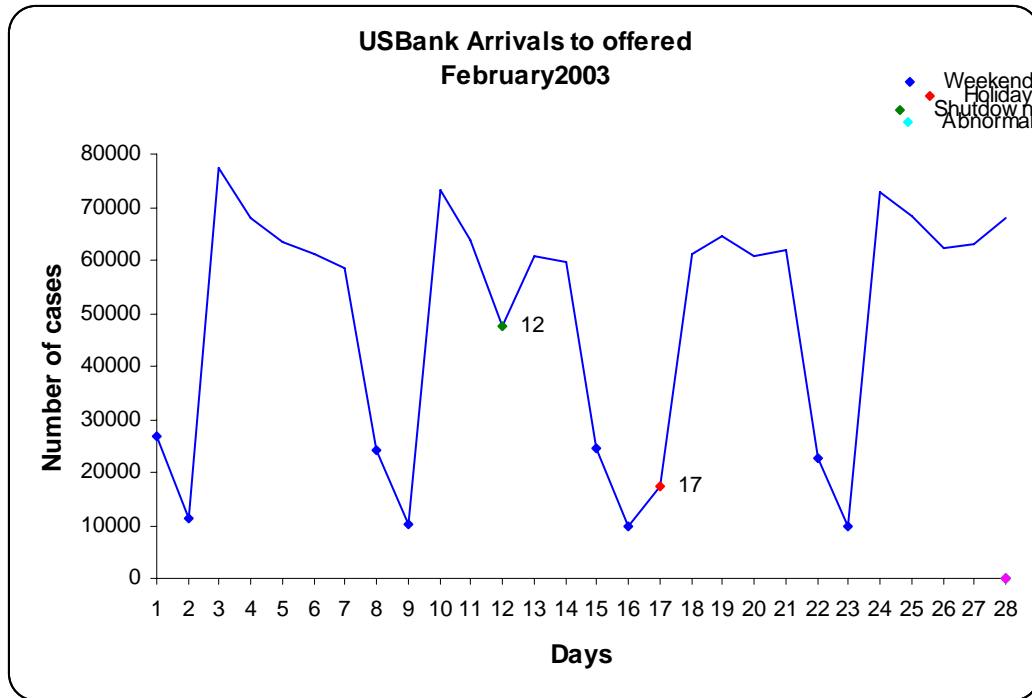
From the variables list select "**Arrivals to offered**" (around the middle of the list – it counts arrivals to the phone queue).
 Press the "**Dates->**" button.



Mark "**Days for one month**" and select (after scrolling down) **February 2003**.

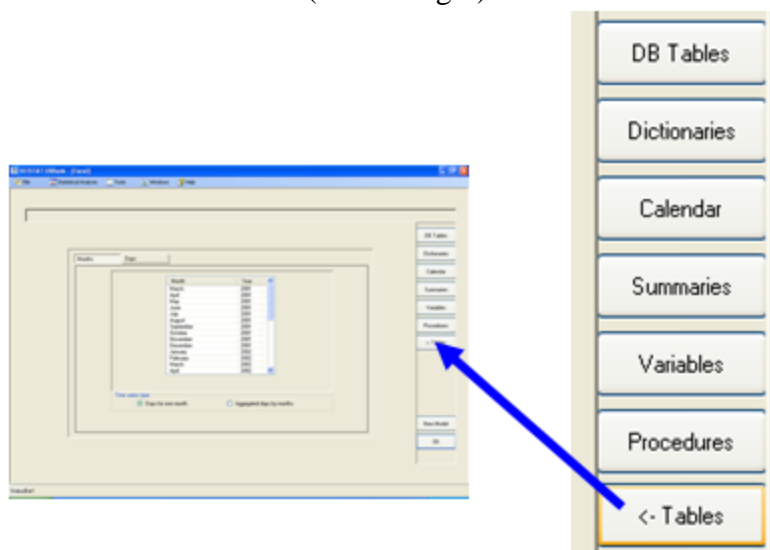


Open tab "**Days**" (there is no need for you to select anything, but note the Comments).
 Click "**OK**".



Two comments are worth making: On February 12, the system stopped working at 4:00 PM, and February 17 was a holiday - Washington's birthday. This is manifested on the chart, where these special days are marked as Abnormal (Shutdown) (green) and Holiday (red). Note that weekends are also marked (blue).

Return again to the **"Statistical Models"** window via the **SEESTAT** tab. Press button **"<-Tables"**(middle-right)

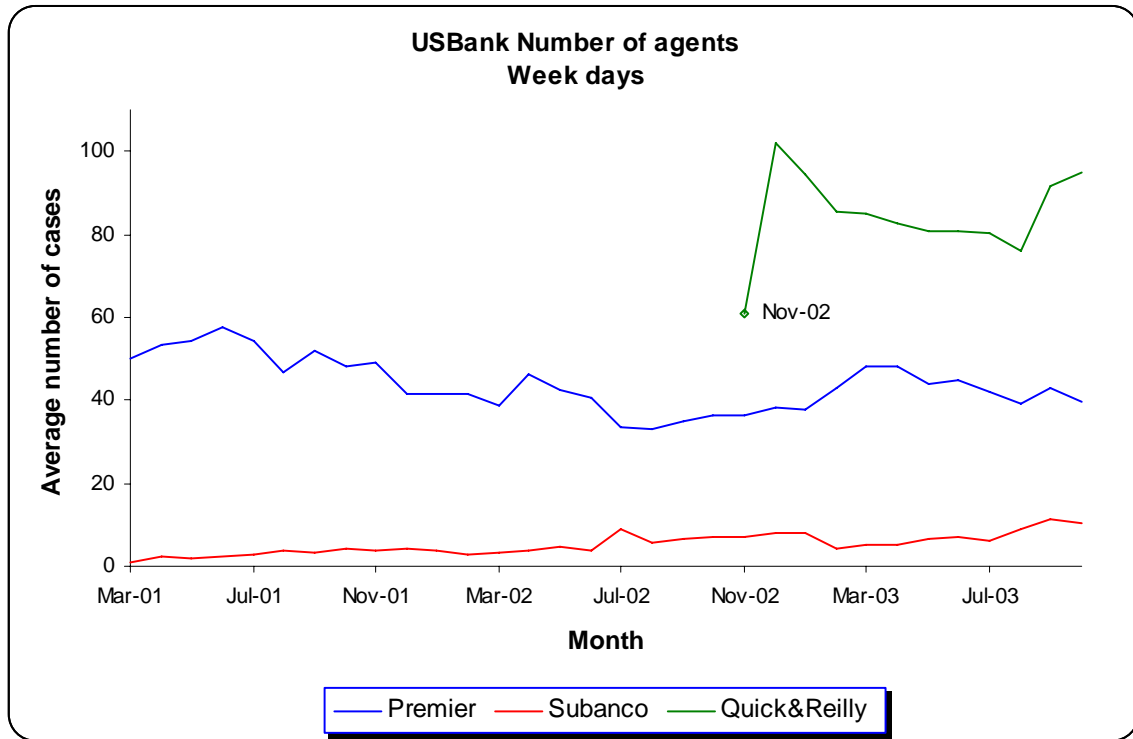


From the variables list select **"Number of agents"**. Open the **"Select Categories"** tab. Select the following **three** services: select **"Premier"** (priority Retail service), press **"Ctrl"**, and select **"Subanco"** (Spanish language), **"Quick&Reilly"** (brokerage).

Now press the "Dates->" button. Mark "Aggregated days by months" and click "Select all".

Open the "Days" tab and select "Week days".

Click "OK".



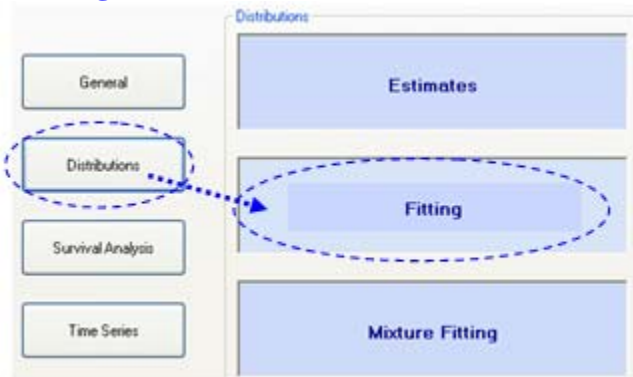
You see that one of the selected services (Quick&Reilly) was integrated into the Call Center of USBank only in November 2002.

Part 2

Example 2.1: Distribution fitting.

We now fit a parametric service-time distribution to the service-time data from [Example 1.1](#)

Open window "**Statistical Models**". Click "**New Model**" and select "**Distributions**" and "**Fitting**".

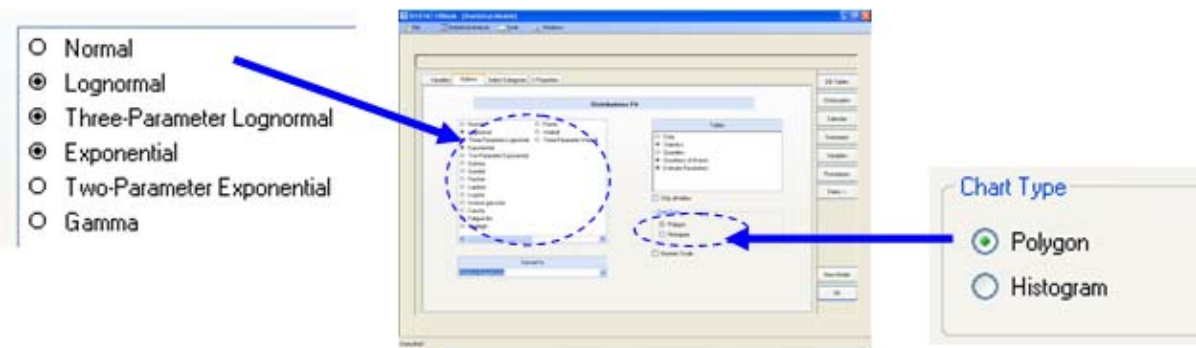


From the variables list select "**Agent service time**".

Open tab "**Options**". You see the list of distributions available for fitting.

Mark simultaneously **3** of them: **Lognormal**, **Three-Parameter Lognormal** and **Exponential**.

Set chart type to "**Polygon**".

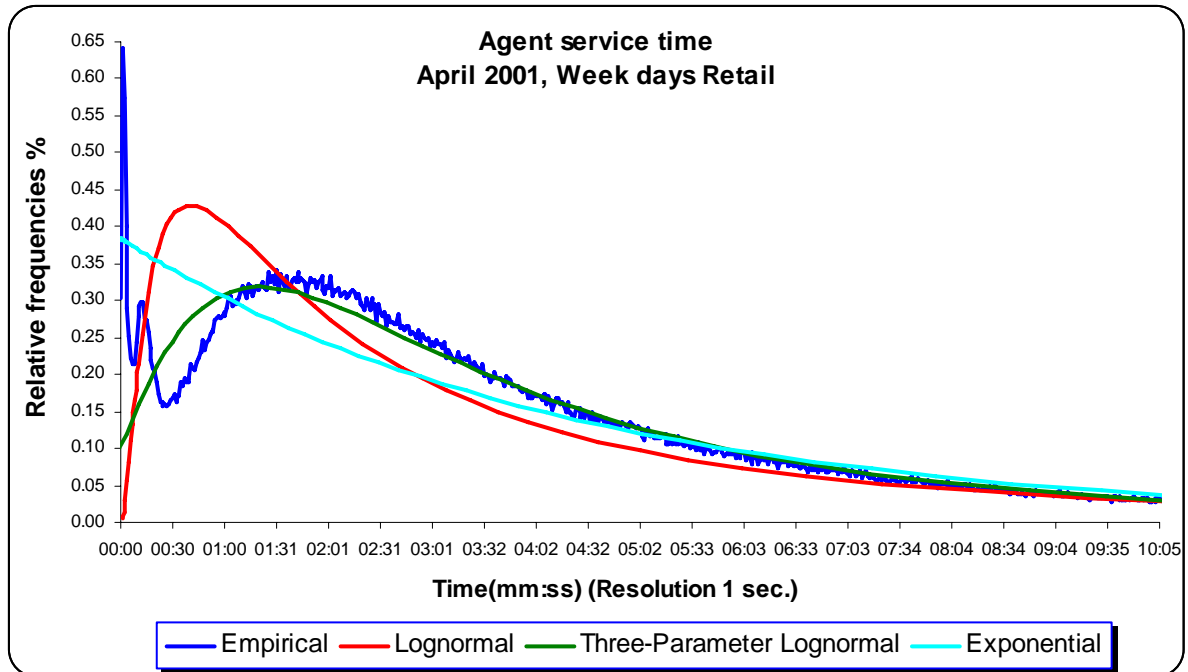


On the tab "**Select Categories**" select **Retail**.

Open the "**X Properties**" tab and set resolution to **00:01 = 1 second**.

Click the "**Dates->**" button. Select **April 2001** and "**Aggregated days**", open tab "**Days**" and select "**Week days**".

Click "**OK**".



Observe again the irregularities near the origin. It looks as though there are at least three distributions involved: very short calls, abnormally short calls and, after around 30 seconds, the pattern looks rather regular. The best fit is produced by the Lognormal distribution with 3 parameters, which amounts to a shifted Lognormal curve. But clearly, close the origin from the right, the fit is inadequate.

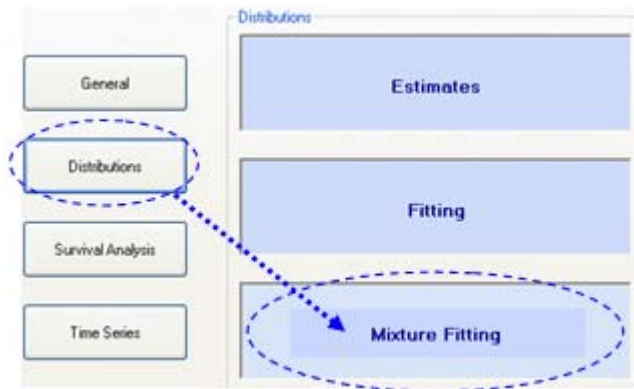
You could use the Tables on the previous *Sheet* (the one accompanying the graph-sheet) to statistically validate the fit: **scroll down** until reaching the tables "**Parameter-Estimates**" and "**Goodness-of-Fit tests**".

Distribution	Goodness-of-Fit Tests				
	Residuals Std	Kolmogorov-Smirnov		Cramer-von Mises	
		Statistic	p Value	Statistic	p Value
Lognormal	0.0471167	0.0823192	<.0001	1374.35	<.0001
Three-Parameter Lognormal	0.0068412	0.0173089	<.0001	28.97	<.0001
Exponential	0.0338746	0.0659468	<.0001	710.39	<.0001

Example 2.2: Distribution mixture fitting

We now try to accommodate the behavior to the right of the origin by a *mixture* of distributions for “Agent service time”.

Via **SEESTAT** return to the "**Statistical Models**" window, click "**New Model**", select "**Distributions**" and "**Mixture fitting**".



Open the "**Options**" tab. You can select a homogeneous or heterogeneous (mixture of various distributions) option. The former is the default. Select "**Lognormal**". Set the number of mixture components to **5**, select chart type **Polygon**.

Mixture Type

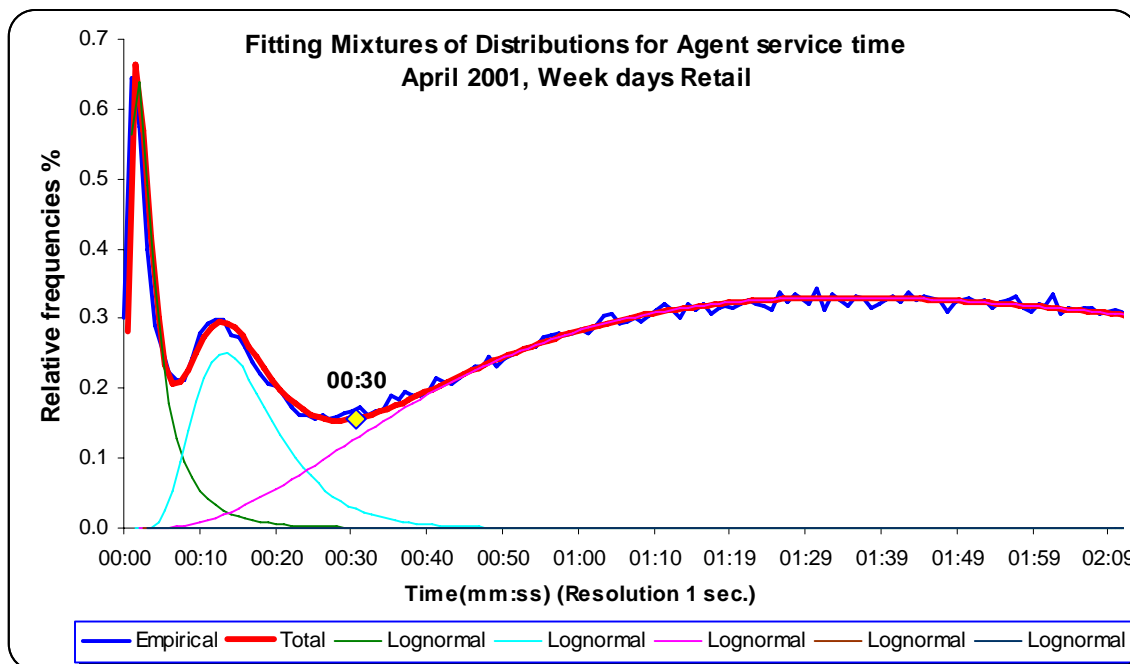
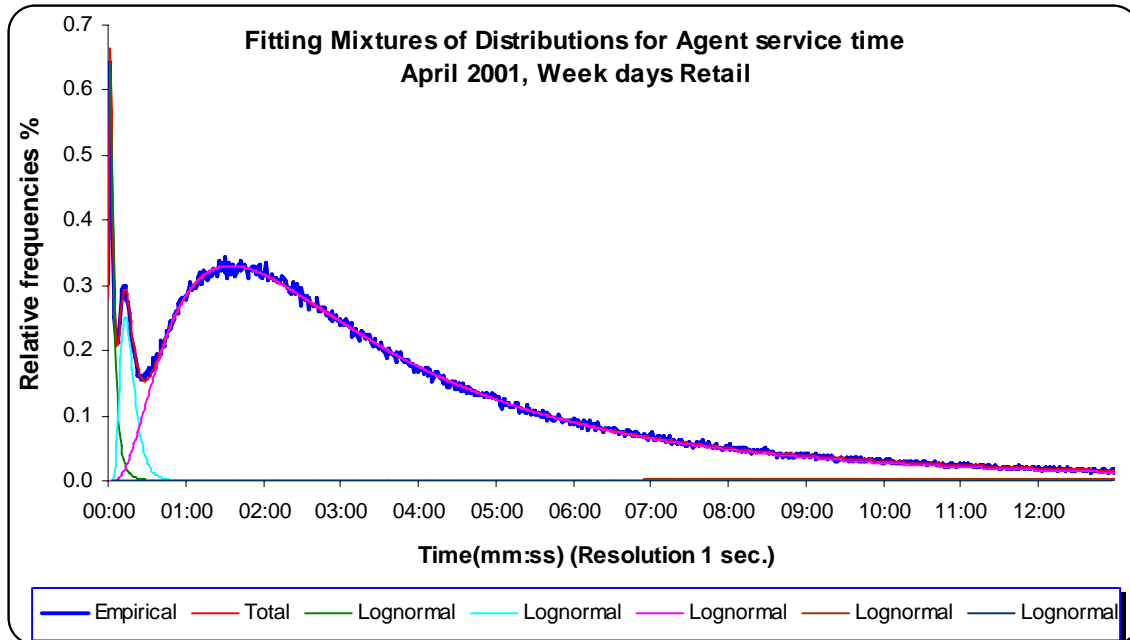
Homogeneous Heterogeneous

Normal Inverse gaussian
 Lognormal
 Exponential
 Gamma
 Weibull

Number of Components : 5

Algorithm: Proportions grid	2
	3
	4
Select Algorithm	5
	6
	7

Click "**OK**".



You observe an almost excellent fit (Red line). In particular, on the left side, there are two components, accommodating the very short and short calls.

Going to the previous Excel Sheet, to view the corresponding Tables (by scrolling it down), one notes that the main component has weight 91% in the mixture – its role in the chart is to fit the part beyond 30 seconds, which it does very well.

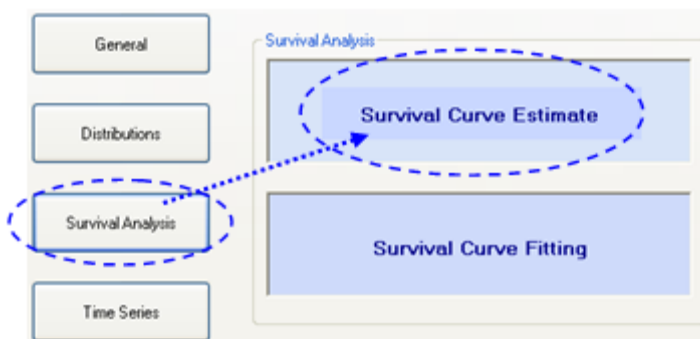
Parameter Estimates	
Components	Mixing Proportions (%)
1. Lognormal	3.08
2. Lognormal	3.58
3. Lognormal	91.14
4. Lognormal	1.86
5. Lognormal	0.34

Example 2.3: Survival analysis with smoothing of hazard rates.

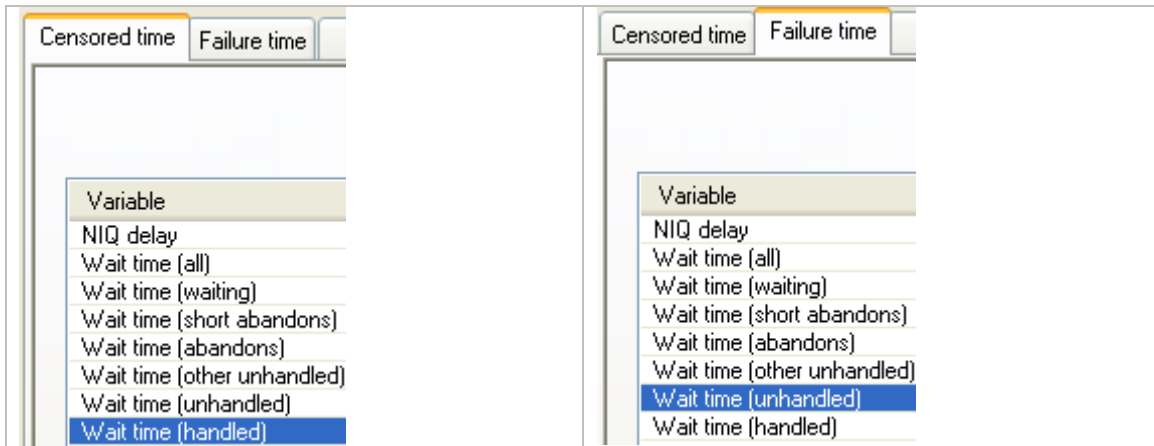
SEESTat supports several survival models. These are required, for example, in order to get insight into customers' **(im)patience**, namely the time they are willing to wait prior to hanging up. Indeed, for those customers who got served, their waiting time provides only a lower bound on how long they are willing to wait - their (im)patience constitutes **censored** observations. One must thus "uncensor" the data to produce adequate estimates. To this end, we now use a simple survival curve estimate. It will produce hazard-rate functions, which provide natural statistical summaries of (im)patience.

Return again, via **SEESTAT**, to the "Statistical Models" window, click "New Model".

Select "**Survival analysis**" and "**Survival Curve Estimate**".



There are two variable tabs. The first tab "**Censored time**" is open. Select "**Wait time (handled)**": this corresponds to the waiting time of the customers who received service. Open the "**Failure time**" and select "**Wait time (unhandled)**": this corresponds to the waiting of customers who joined the queue but did not receive service (mainly due to abandonment, though there are sometimes other reasons such as system malfunction).



Open the "**Options**" tab. SEESStat supports several methods of smoothing, which are applicable to hazard rates and beyond.

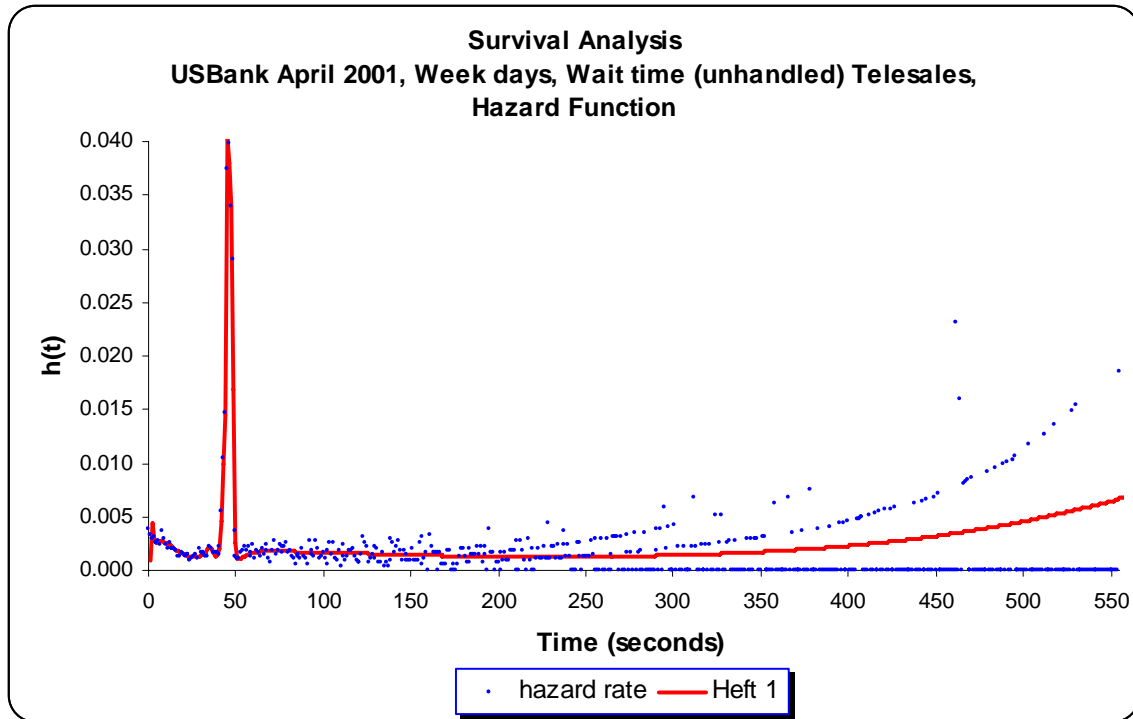
We shall use the default algorithm (HEFT).



From the tab "**Select categories**" select "**Telesales**".

Click "**Dates**". Select "**April 2001**" and on the tab "**Days**" select "**Week days**".

Click "**OK**".

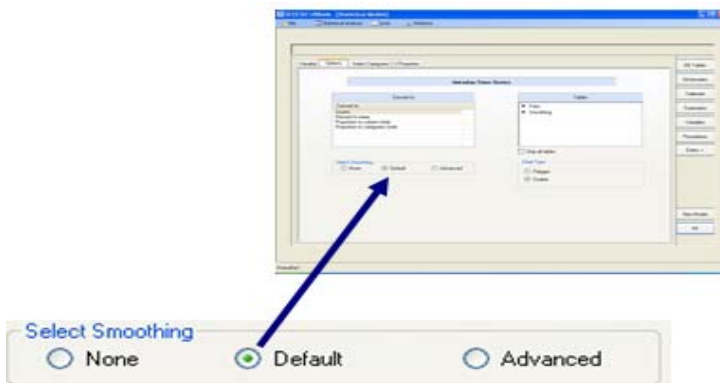


A noticeable peak in the hazard rate indicates that there is a trigger for customers to abandon after about 50 seconds of waiting (which, based on our experience, could be the result of a voice-announcement at that time).

Example 2.4: Smoothing of intraday time series.

Smoothing algorithms are available for several statistical models. We now demonstrate the application of smoothing to the data used in [Example 1.2](#).

Return as usual to "Statistical Models", click "New Model", select "Time Series" and "Intraday". Select "Arrivals to queue". In "Options" tab select "Default" smoothing (this time, the default is the method of Cubic Splines).



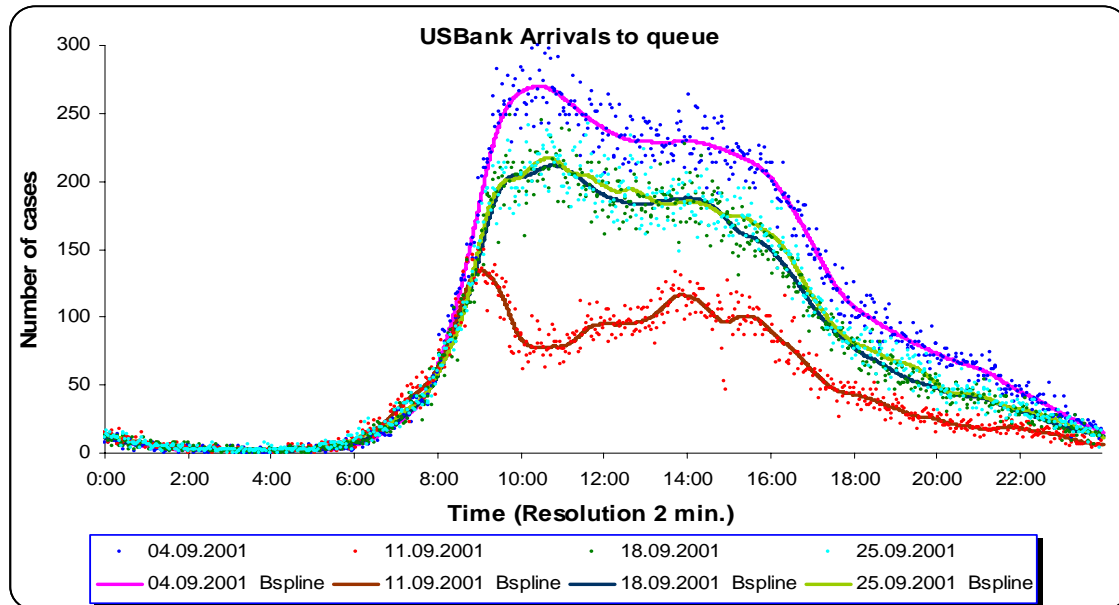
Select **"Scatter"** as chart type

In tab **"X Properties"**, set resolution to **02:00** = 2 minutes.

Click **"Dates"**, mark **"Individual days"** and select **"September 2001"**.

On the **"Days"** tab select (with **"Ctrl"** and click) all four **Tuesdays** of September.

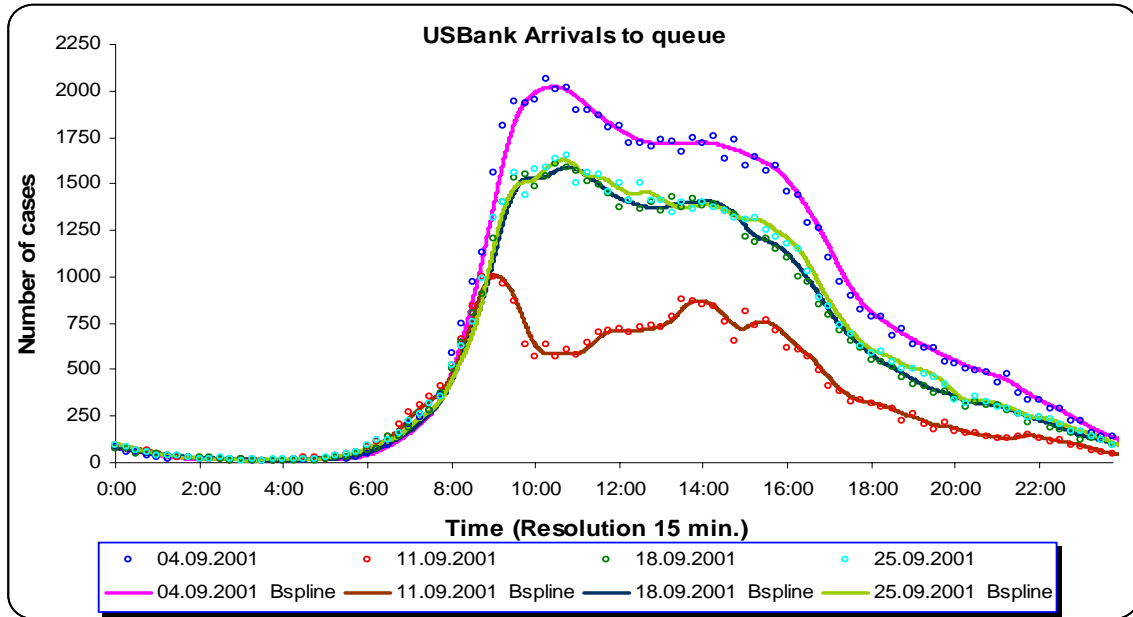
Press **"OK"**.



For this small resolution of 2 minutes, there is plenty of noise, but the smoothed data clearly identifies the regular pattern that was discovered before. (Note that the smoothed curves are computed with the minimal resolution for this variable, which is 30 seconds).

Click **"Excel Chart"** on the main menu, then click **"Modify Chart"**.

Open the **"Properties"** tab, set resolution to **15 min.** and click **"OK"**.



The Averaged Data (over 15 minutes) are now much closer to the smoothed curves, as expected.

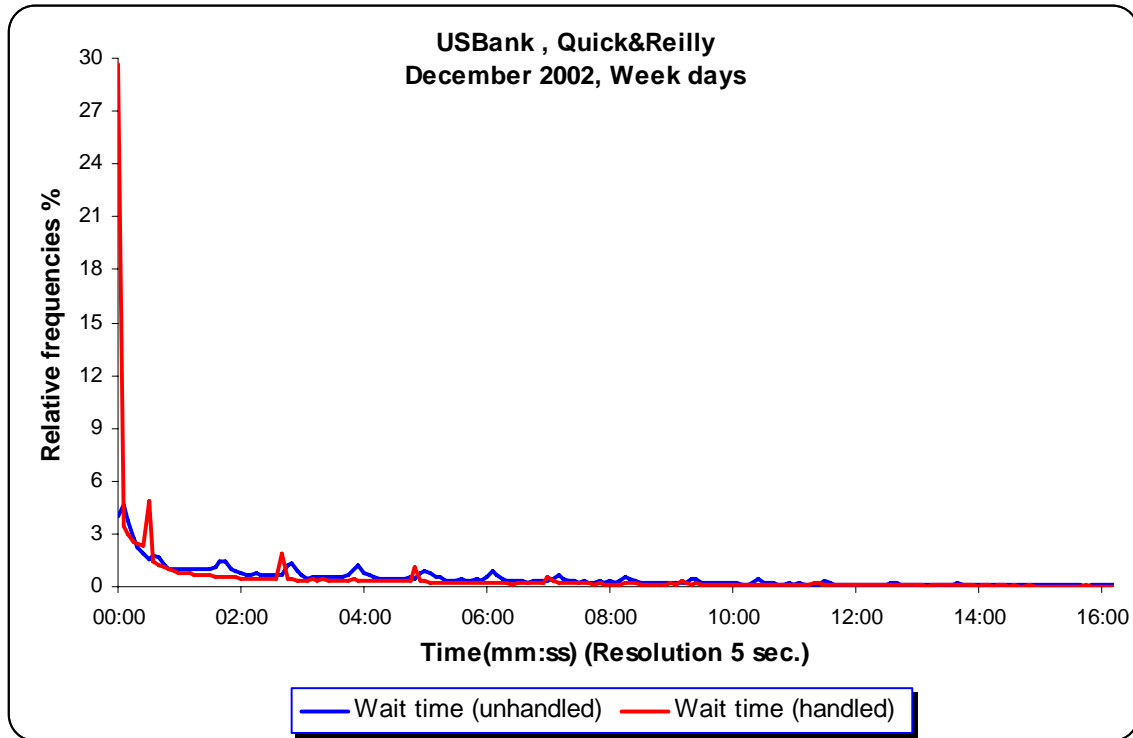
Part 3

Some additional interesting examples.

Example 3.1: Queue regulated by a protocol (Quick&Raily)

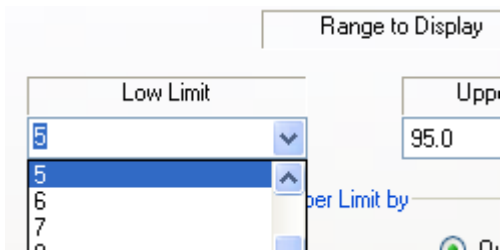
Via **SEESTAT** return to the **"Statistical Models"** window, click **"New Model"**, then click the **"Distributions"** button. Three available distribution models appear. Select **"Estimates"**. In tab **"Variables"** select (using **Ctrl**) both **"Wait time (unhandled)"** and **"Wait time (handled)"**.

In tab **"Options"** select chart type **Polygon**. Click **"Dates->"**, select **December 2002**, make sure the **"Aggregated days"** option is selected, and in **"Days"** select **Week days**. Click **"<-Tables"**. In **"Select Categories"** select **"Quick&Reilly"**. Press **"OK"**.

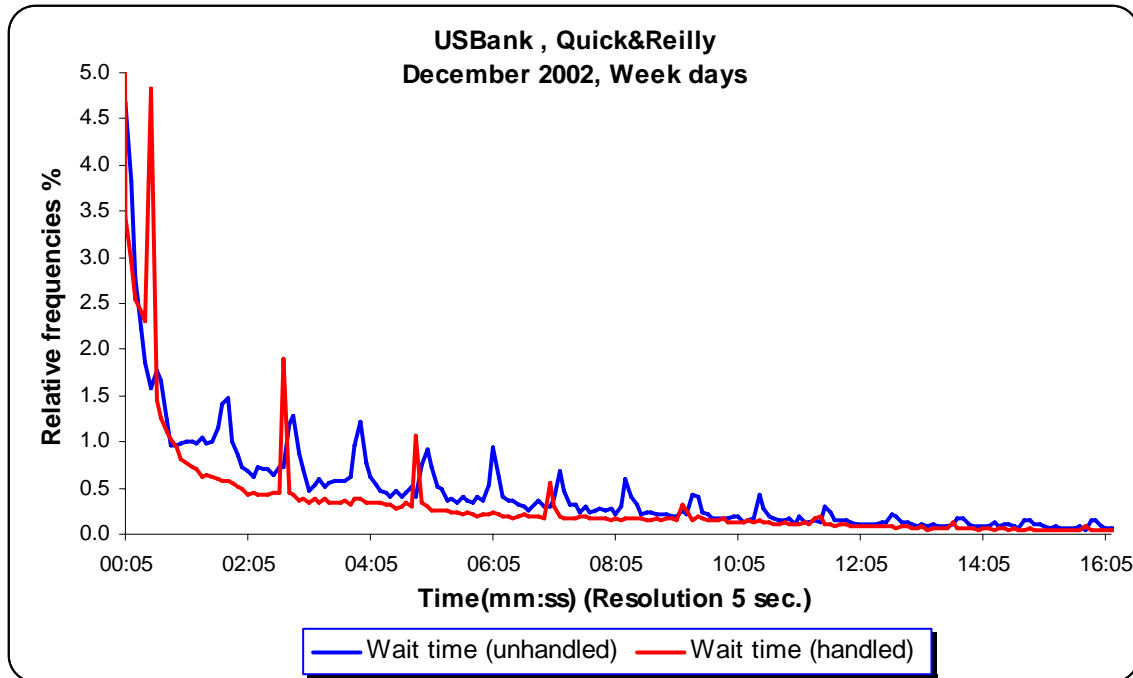


Both lines are periodical. To get a better focus, you will cut the chart at the left side.

Click **"Excel Chart"** on the main menu and then **"Modify Chart"**.
Open **"Properties"**, set the low limit **5 seconds**.



Click **"OK"**.



As you see, the Wait time (unhandled), in blue, peaks every 65 sec. The Wait time (handled), in red, peaks every 130 seconds. These interesting observations are yet to find their explanations.

Example 3.2: Queue length (Business and Platinum).

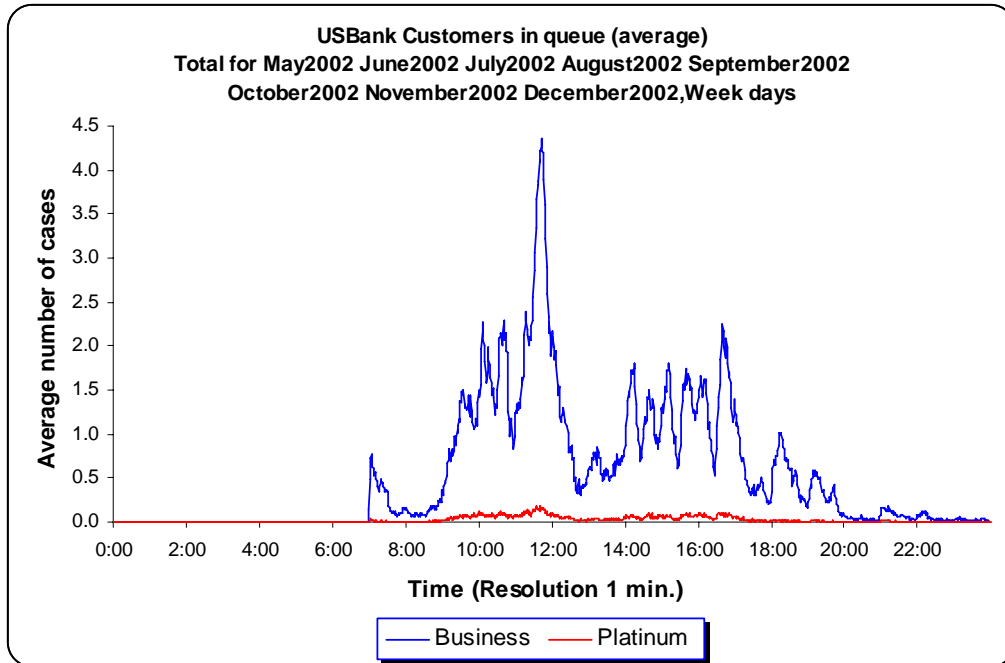
Via **SEESTAT** return to the **"Statistical Models"** window, click **"New Model"**. Click the **"Time Series"** button.

Two available models for time series appear. Select **"Intraday"**. On tab **"Variables"** select **"Customers in queue (average)"**.

On tab **"Options"** select smoothing **"None"** and chart type **Polygon**.

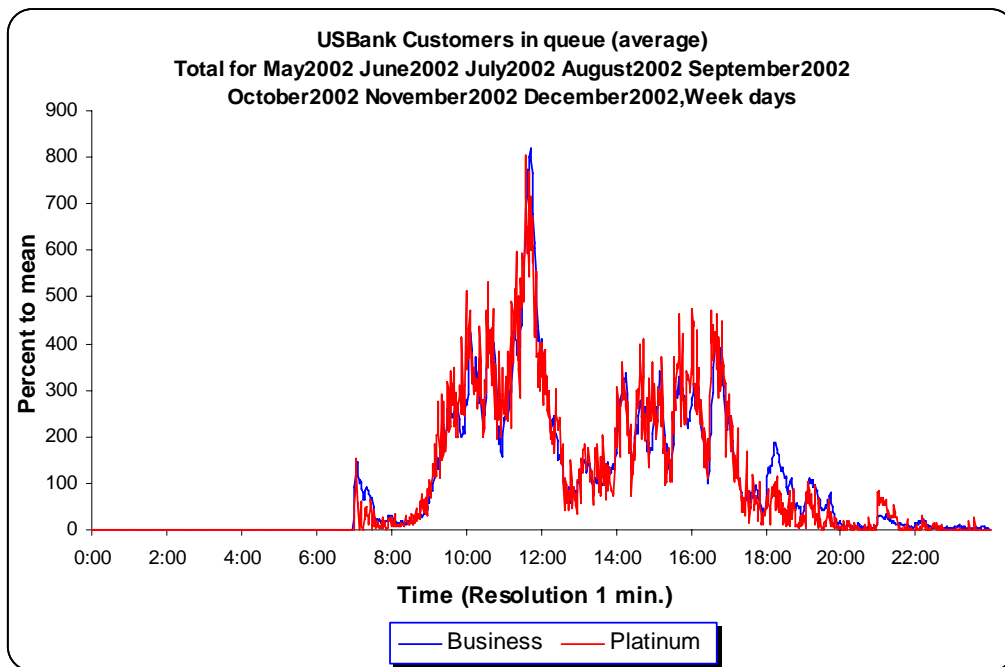
In **"Select Categories"** tab select (with **Ctrl** and click) **Business** and **Platinum**.

Click **"Dates->"**, select **"Dates totals only"**, select the 8 months from **May 2002** to **December 2002** and select **Week days** on the **"Days"** tab .Click **"OK"**.



Platinum is a small-scale service. You will now normalize the chart in order to identify patterns.

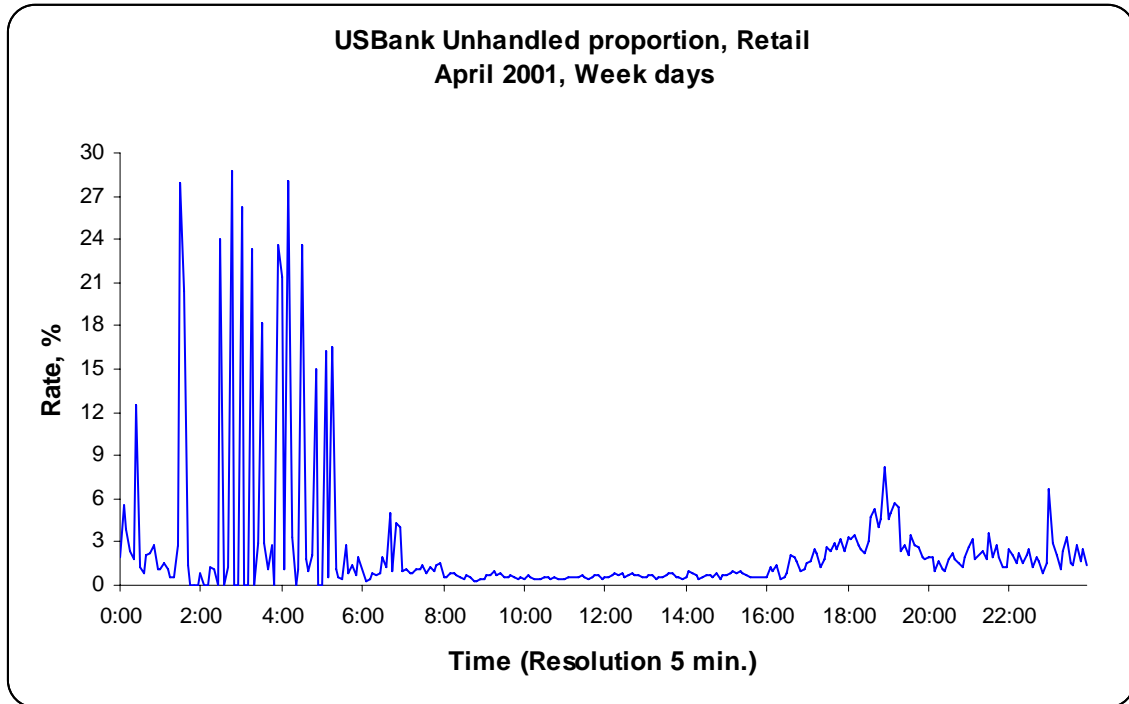
Click **"Excel Chart"** on the main menu and then **"Modify Chart"**.
 Open the **"Options"** tab and select **Percent to mean**. Click **"OK"**.



Note the overlapping patterns of the queue lengths of the two customer types. (This phenomenon is called State-Space-Collapse, in Queueing Theory.)

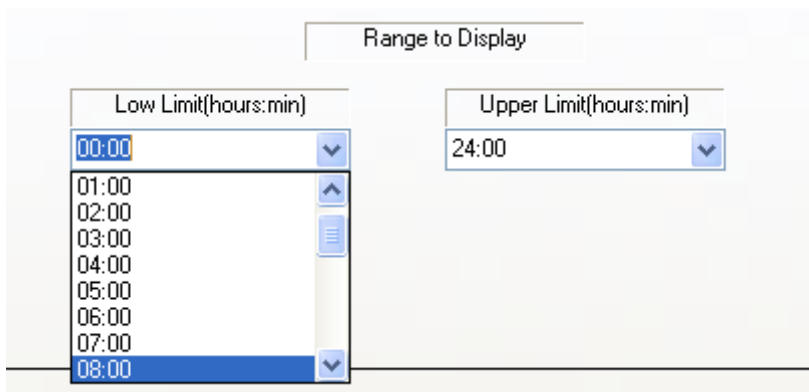
Example 3.3: Average wait time (all) and Unhandled proportion

Via **SEESTAT** return to the **"Statistical Models"** window, click **"<-Tables"**. On tab **"Variables"** select **"Unhandled proportion"**. In **"Select Categories"** tab select **"Retail"**. Click **"Dates->"**, select **"April 2001"**, and select **Week days**. Click **"OK"**.

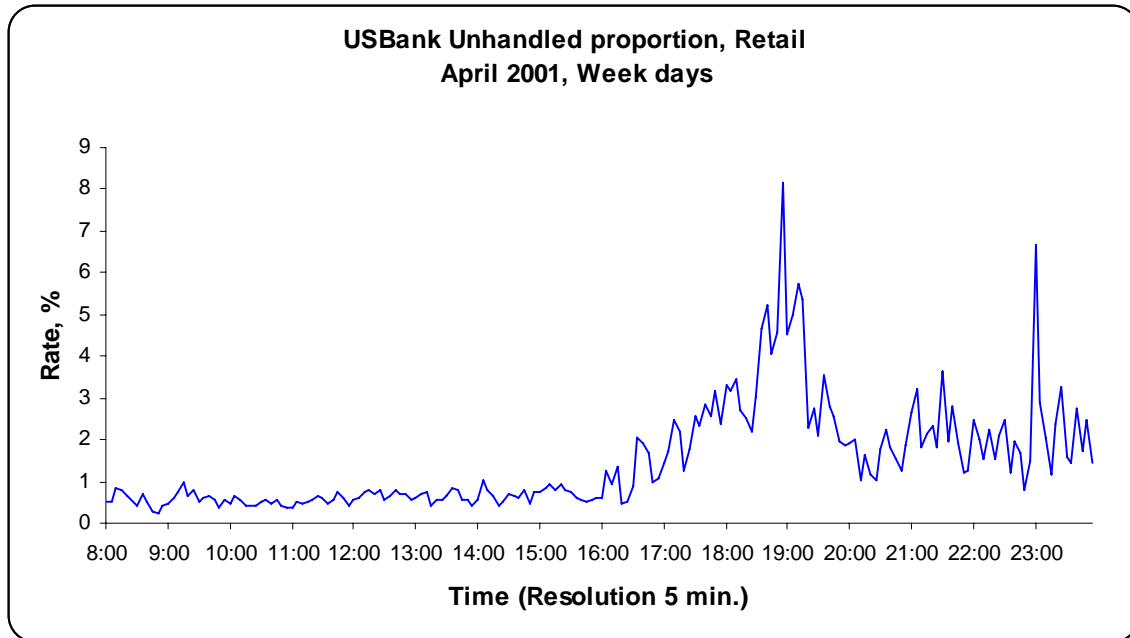


You observe a lot of noise before 8:00 AM. There are only few agents working then, and few customers are calling. We now cut this irrelevant part of the chart, until 8:00 AM.

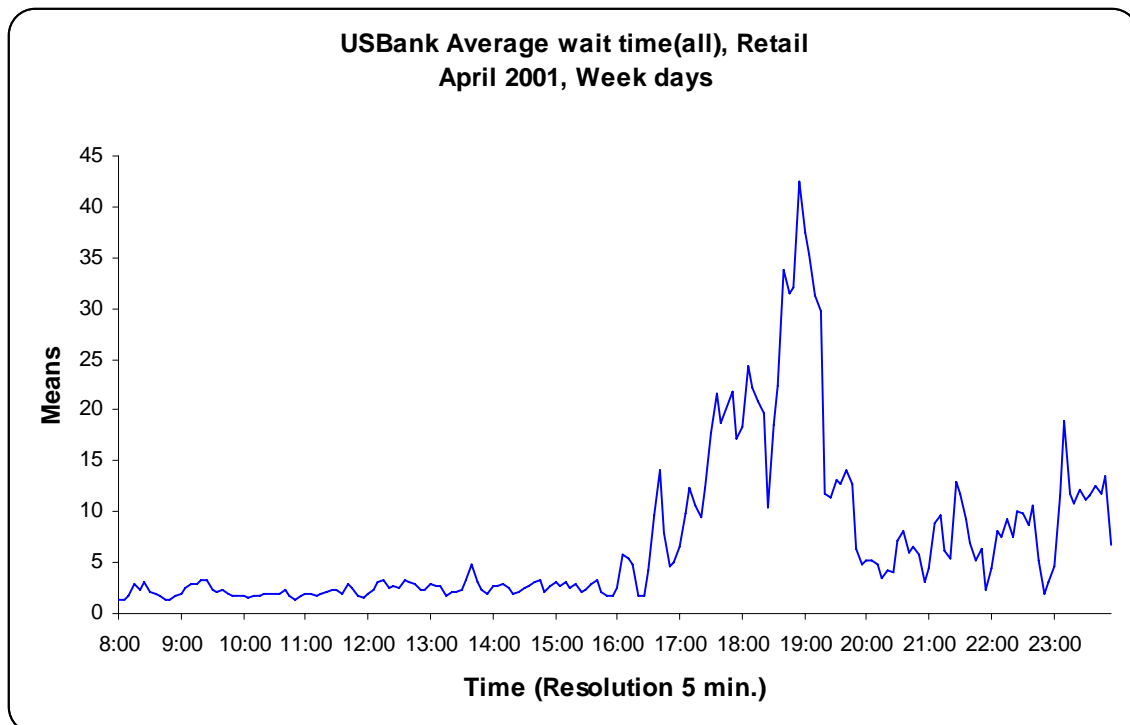
Click **"Excel Chart"** on the main menu and then **"Modify Chart"**. Open the **"Properties"** tab and change low limit to **08:00**.



Click **"OK"**.

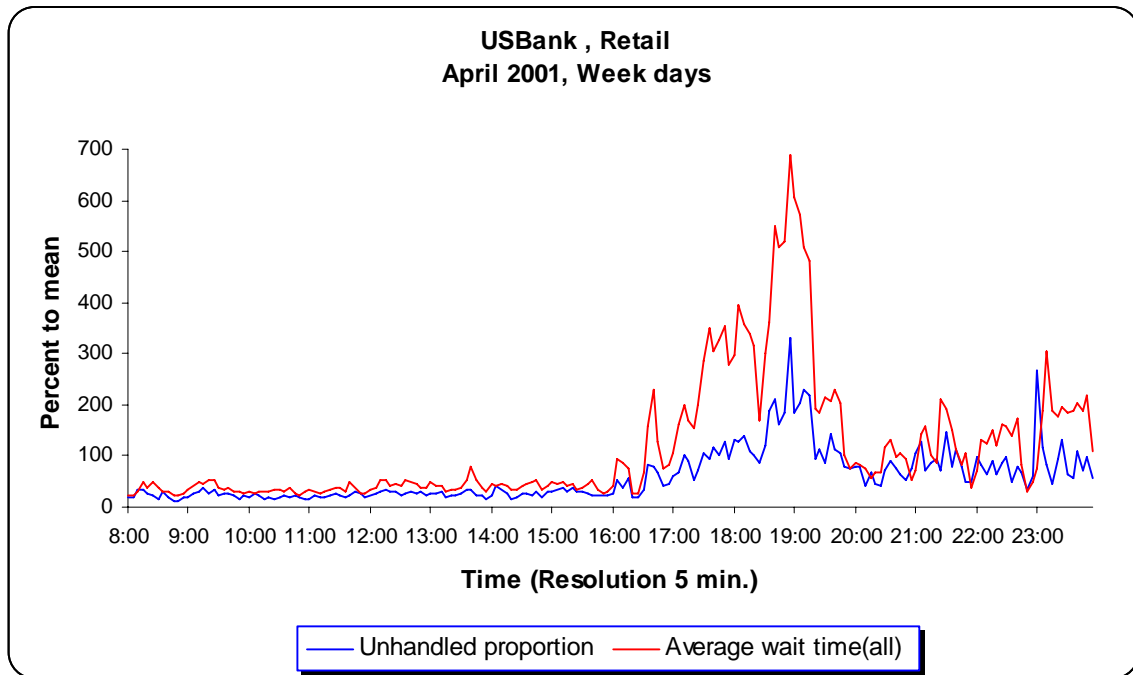


Via **SEESTAT** return to the **"Statistical Models"** window.
 Click **"<-Tables"**. On tab **"Variables"** select **"Average wait time (all)"**. Click **"OK"**



We see that the patterns for two variables ("Unhandled proportion" and Wait Time (all)) are similar. We now compare them more closely.

Via **SEESTAT** return to the "**Statistical Models**" window. On tab "**Variables**" select "**Unhandled proportion**" and "**Average wait time (all)**". Click **OK**.

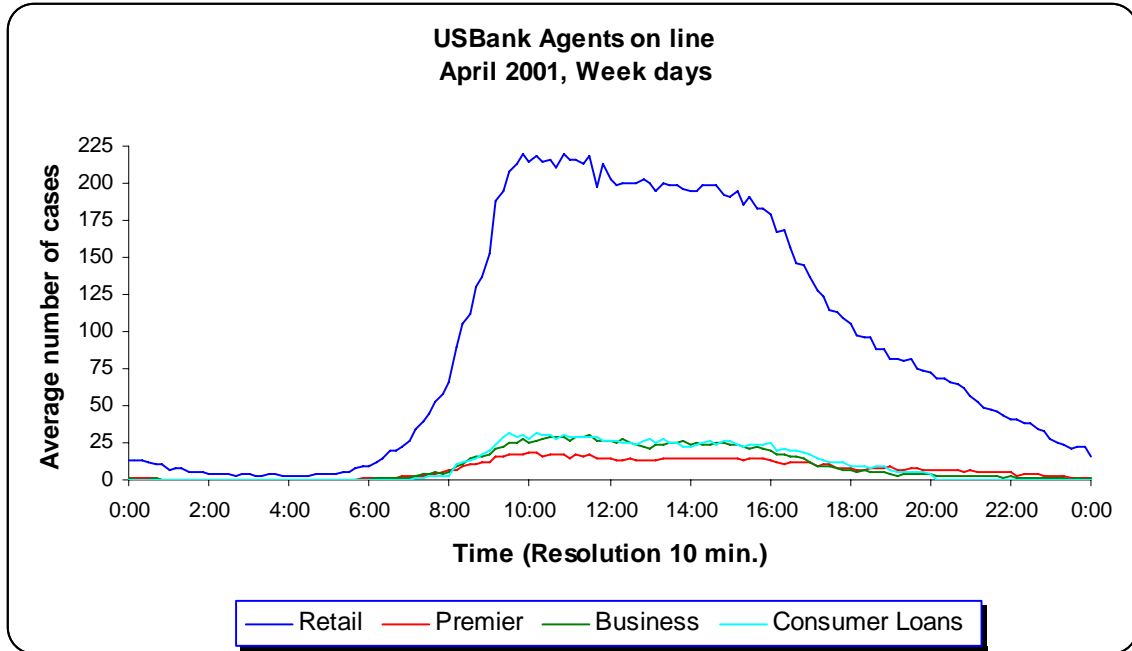


You observe an increase in "unhandled proportion" and "average wait time" from 17:00 to 20:00. During this period, a lot of agents are leaving their shifts. The number of arrivals is also going down, but the schedule of agent exits is not synchronized with arrivals – agents are leaving prematurely. This was a problem for USBank for some period of time.

Example 3.4: Agents on line

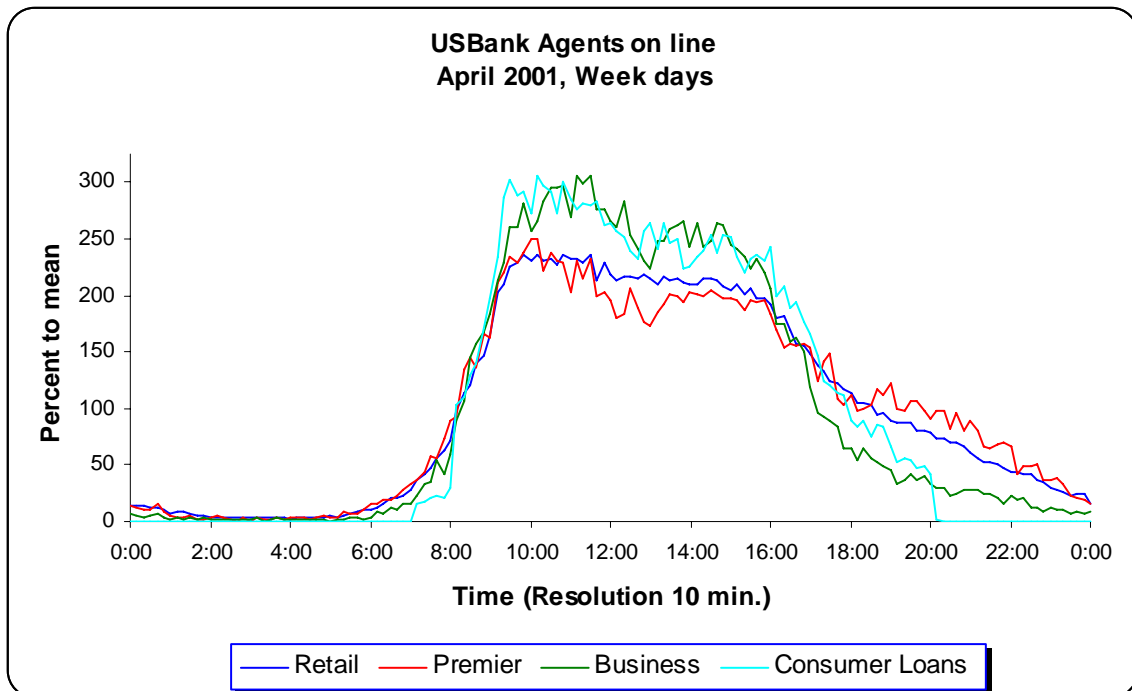
Via **SEESTAT** return to the "**Statistical Models**" window. On tab "**Variable**" select "**Agents on line**". In "**Select Categories**" tab select **Retail, Premier, Business** and **Consumer Loans**.

In tab "**X Properties**" select resolution **10:00** minutes. Click "**Dates->**", select **April 2001, Week days**. Click **OK**.



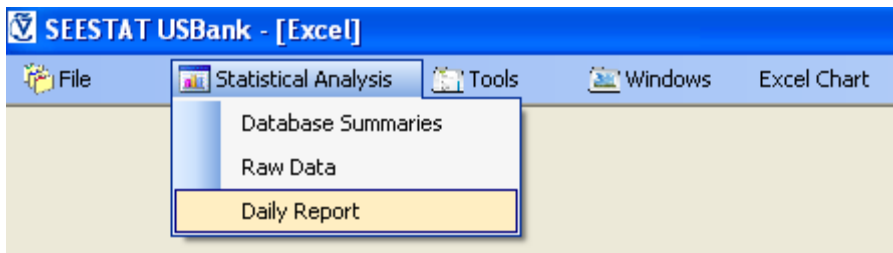
Due to the differing volume of services, you will normalize the chart in order to explore the existence of a common *pattern*.

Click **"Excel Chart"** on the main menu and then **"Modify Chart"**. In tab **"Options"** select **Percent to mean**. Click **OK**.

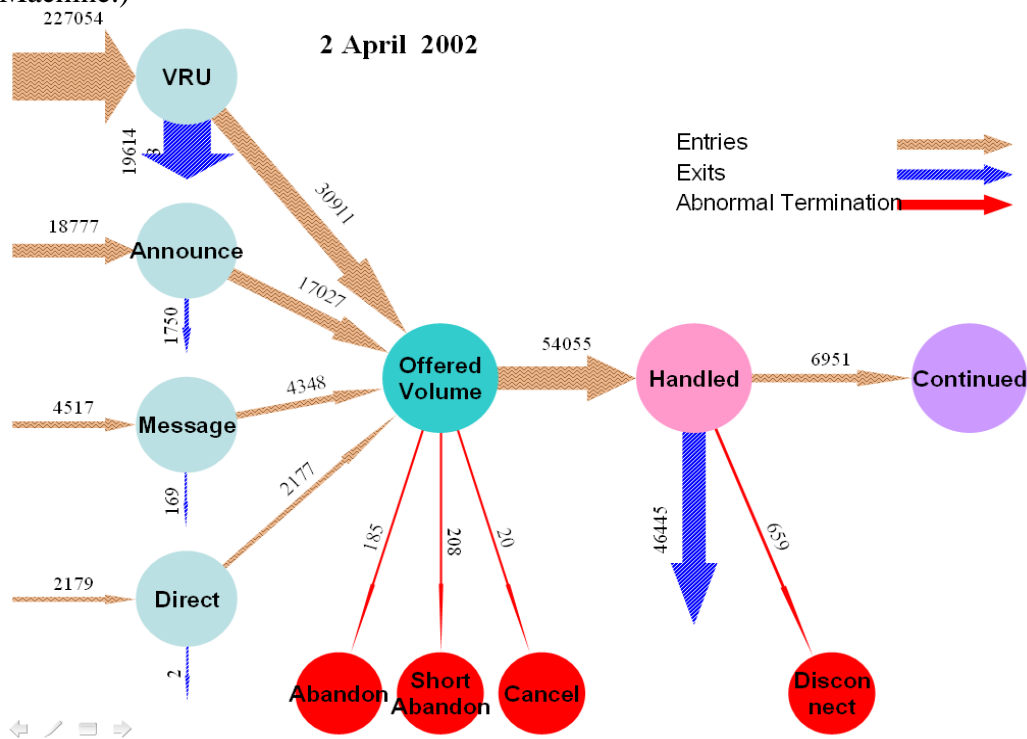


Example 3.5: Daily flow of Totals calls (Tuesday, April 2, 2002)

Via SEESTAT return to the "Statistical Analysis" menu. Select "Daily Report".



Click "Dates->" select April 2002, "Individual days", click "Days" and select "2 April 2002 Tuesday". Click OK. (Note; VRU = Voice Response Unit, or simply an Answering Machine.)



We have chosen a typical day – Tuesday, April 2, 2002 – since this day has virtually no problematic calls. There is a total of 261,143 calls on that day. The PowerPoint slide describes the process-flow of calls. There are 4 significant entry points to the system: through the VRU ~227054 calls (87%), Announcement ~18777 calls, Message ~4517 calls and Direct group (callers that directly connect to an agent) 2179 calls. 196143 calls (about 79% of all calls) exit from the system through the VRU, Announcement, Message and Others groups; while another 21% of callers entering the system seek service by an agent (Offered Volume).

The served callers include those who will request other services 6951 calls (about 13% of the handled calls), while 46445 calls (86% of callers) exit the system after receiving service by a single agent.