

# 電車の遅延データを収集、Excelでデータベース化し分析する

東京大学教育学部附属中等教育学校 5年 近藤 琉海

## 1. 動機

通学で毎日一定の時間に電車を利用するようになりより多くの電車の遅延に遭遇するようになった。そして利用路線のひとつである東武東上線は遅延が多い(特に人身事故が多い)と言われることが多いのに気づいた。これは事実なのか？事実ならそれはなぜなのか？疑問に思い調べてみたが遅延に関する詳細な統計情報は開示されていないようである。そこで電車の遅延について総合的に分析し、鉄道の安全運行について考察を行うことを目的として東上線を含む私鉄3線の毎日の遅延情報を収集、Excelでデータベース化することとした。

## 2. 研究概要

収集対象：東武東上線、西武線、東急線の各路線の遅延データ（東上線と近郊、乗り入れ、接続している路線を選定）  
データの収集方法：各鉄道会社が利用者向けにメール、アプリ等で発信しているリアルタイムの遅延のデータを一つのメールアドレスに集約して収集する  
研究方法：メール、アプリで発信された内容から各項目のデータをExcelのテーブルに手動で書き込むことによりデータベースを作成する  
収集期間：2022年7月～2023年6月以降

## 3. データベースの作成

### 3.1 データベース設計の方針

収集するデータ項目の選定にあたってはデータベース作成後にデータを抽出・分析しやすいよう心掛けた。特にデータの分析時にExcelの関数で複数の条件を組み合わせたことが想定されたため番号、コードをつけ判定しやすくした。またデータの収集期間が1年以上でデータの総数が多くなることも予想されたことから当初の設計からデータベースの構成を変えないように項目を考えた。

### 3.2 データ項目とデータベース

以上より次のデータ項目を設定し(表1)データベースを作成した(表2)。

表1 データ項目

項目	例	概要	項目	例	概要
A 番号	00318	5桁の通し番号	L 影響	●	他路線が原因の場合(●または空欄)
B 月別番号	2301003	年月別番号 YYMM001	M 人的		人的事象が原因の場合(●または空欄)
C 年月コード	THJAN	年月英字コード TW 22年 TH 23年	N 他	●	その他の事象が原因の場合(●または空欄)
D 年月日	2023/1/2	年月日	O 発生理由	人身事故	発生理由
E 曜日	月	曜日	P 発生場所	下赤塚駅	発生駅・区間
F 時刻	13:15	発生時刻	Q 影響区間	池袋～和光市駅	影響区間
G 天気	晴れ	天気	R 影響理由	運転見合わせ	影響理由
H 鉄道会社	東武鉄道	鉄道会社 東武鉄道、西武鉄道、東急電鉄	S 振替	●	振替輸送の有無(●または空欄)
I 路線	東武東上線	路線	T 運転再開	14:33	運転再開時刻
J 路線コード	TJ	路線コード	U 平常通り	20:59	平常運行時刻
K 自線		自路線が原因の場合(●または空欄)	V 時間	13	遅延発生時刻(時)

I: 路線 (J: 路線コード) 東武東上線 (TJ)、西武新宿線 (SS)、西武池袋線・秩父線 (SI)、東急東横線 (TY)、東急田園都市線 (DT)、東急目黒線 (MG)、東急大井町線 (OM)、東急池上線 (IK)、東急多摩川線 (TM)、東急新横浜線 (SH)

表2 データベース

番号	月別番号	年月コード	年月日	曜日	時刻	天気	鉄道会社	路線	路線コード	自線	影響	人的	他	発生理由	発生場所	影響区間	影響理由	振替	運転再開	平常通り	時間	
00001	2207001	TWJUL	2022/7/1	金	7:03	晴れ	東急電鉄	東急東横線	TY	●	●	●	●	異音確認	自由が丘駅～田園調布駅		遅れが発生	●		12:00	7	
00002	2207002	TWJUL	2022/7/1	金	7:03	晴れ	西武鉄道	西武池袋線・秩父線	SI	●	●	●	●	東急東横線内で線路内点検を行った影響			一部列車に遅れや運休				7	
00003	2207003	TWJUL	2022/7/1	金	7:24	晴れ	西武鉄道	西武新宿線	SS	●	●	●	●	旅客対応	東村山駅		一部列車に遅れ				7	
00004	2207004	TWJUL	2022/7/1	金	7:03	晴れ	東急電鉄	東急目黒線	MG	●	●	●	●	東急東横線内で線路内点検を行った影響			遅れ				7	
00005	2207005	TWJUL	2022/7/1	金	7:03	晴れ	東武鉄道	東武東上線	TJ	●	●	●	●	東急東横線内で線路内点検を行った影響		和光市～小川町駅	一部列車に遅れ			13:41	7	
00006	2207006	TWJUL	2022/7/1	金	15:25	晴れ	西武鉄道	西武新宿線	SS	●	●	●	●	沿線火災	下井草～井荻駅	西武新宿～上石神井駅	運転見合わせ	●			15	
00007	2207007	TWJUL	2022/7/2	土	10:10	晴れ	西武鉄道	西武池袋線・秩父線	SI	●	●	●	●	人身事故	大泉学園駅		運転見合わせ	●			10	
集計	7	7	7	7	7	7	7	7	7	4	3	2	5		7	4	2	7	3	0	2	7

データ総数：849件 2022年7月～2023年9月(15カ月間) \*23年11月10日現在もデータ収集継続中

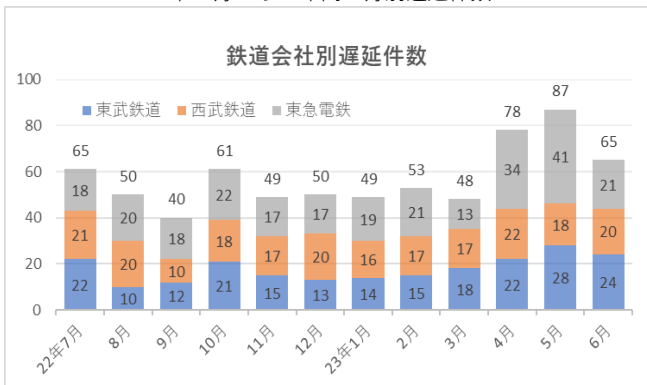
## 4. データベースの利用例

データベースを基にして集計、分析作業を行う(以下は集計表、グラフの例)。

総数(22年7月～23年9月)

西武鉄道	259
西武新宿線	95
西武池袋線・秩父線	164
東急電鉄	323
東急新横浜線	20
東急多摩川線	7
東急大井町線	36
東急池上線	29
東急田園都市線	54
東急東横線	108
東急目黒線	69
東武鉄道	267
東武東上線	267
総計	849

22年7月から1年間の月別遅延件数



## 5. 今後の課題

今後は得られたデータを抽出して集計、分析を行う予定である。主に関数を利用し日付、曜日、時間ごとの遅延件数や鉄道会社、発生場所、発生理由等による遅延の特徴、因果関係の発見などを行ってきたい。そして電車の遅延を減らすにはどのような対策が必要か総合的な考察を行うことが研究の最終目標である。これまでのデータ集計により遅延理由などの文字データは抽出が難しいことがわかった。この処理をうまくできる方法を考えていきたい。今回は手動でデータ収集を行ったが今後もデータベース作成を継続していくには自動的なデータ収集も考えたい。