Australasian Transport Research Forum 2021 Proceedings
8-10 December, Brisbane, Australia
Publication website: http://www.atrf.info

Case Study: Standard deviation as a reliability measure for public transport managed by headways

Mathew Hounsell
Institute for Sustainable Futures, University of Technology Sydney,
PO Box 123, Ultimo, NSW, Australia, 2007
Email for correspondence: Mathew.F.Hounsell@student.uts.edu.au

Abstract

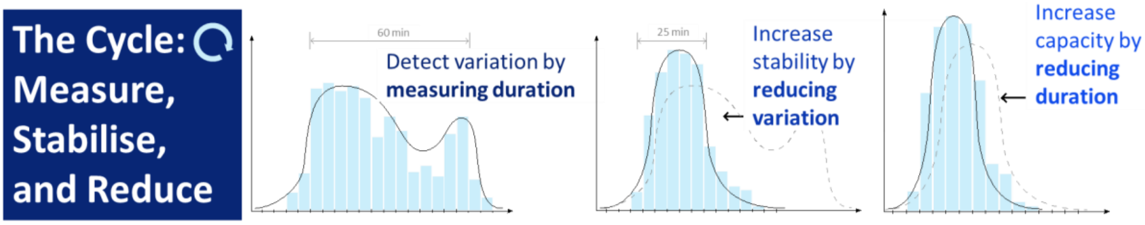
This paper examines empirical data from public transport operations in Sydney, to assess if the mean end-to-end runtime, and its standard deviation, are useful measurements to monitor the reliability and efficiency of the public transport services delivered. This paper shows that the mean plus two standard deviations for end-to-end runtimes was an effective estimator of the 97th percentile for the both the headway and timetable-based services studied.

1. Introduction

This paper and the paper Hounsell (2021) form a pair. Hounsell (2021) describes why the transport operator must deliver an efficient and reliable public transport service, in order to maintain their legitimacy. Hounsell (2021) then explains how *statistical process control* (SPC) could theoretically be used to assist in achieving reliable and efficient operations.

This paper examines observations of the running times for three transport services in Sydney to assess whether *observed* running times on the transport services can be treated as either normally distributed or a valid estimator for use in SPC. UTS has already demonstrated that efficiency can be improved through the application of the Measure, Stabilise, Reduce (MSR) framework outlined by Dr M.E. Zeibots (Hounsell 2018a). SPC can be considered an application of the MSR framework focused on measuring and reducing variation.

Figure 1: Runtime variability impacts what services are targeted and what services can be delivered



In a natural system, the measurements of key parameters will often form a normal distribution. As such measuring the system and controlling variance is a key technique for reducing waste and achieving efficiency. The wider business community has expanded upon the principles of SPC since the early twentieth century, such as the work of W.E. Deming (Deming 2000, 2018).

The six-sigma approach to achieving SPC monitors the mean ($\bar{\boldsymbol{e}}$) and the standard deviation ($\boldsymbol{\sigma_e}$) of the product or service because in a normal distribution 99.8% of the values will be less than three standard deviations from the mean. Therefore, if the operator controls the variance of the delivery to ($\bar{e} - 3\sigma_e \geq e_{min}$ and $\bar{e} + 3\sigma_e \leq e_{max}$) then the output will be between less than the upper quality tolerances of the process 99.9% of the time (Brussee 2006).

This paper examines Sydney's Inner West Light Rail (IWLR) and the City to South-East Light Rail (CSELR). The IWLR is about 12km long, and both branches of the CSELR are about 9km long. These two light rails provide major high frequency services in Sydney and are operated to a Headway Adherence based KPI. These services do not operate to conventional timetables, and do not have substantial built-in 'fat' where vehicles wait at intermediate stops for several minutes to maintain timetable adherence. As such, these services operate as natural processes.

This paper also examines the autonomous Sydney Metro during testing. Although the data shows the metro operated to a timetable, the timetable 'fat' built into the testing was minimal. Thus, it was examined to see if a well-controlled timetabled service could support SPC. Remember that on a frequent route timetable 'fat' is just wasting your passengers time to meet a product centric timetable adherence KPI.

This paper shows that the method could be applied to any frequent service in a dedicated right of way, as well as to any frequent tram or bus service in any city. However, further research is required into the limits of this approach for on-street routes longer than twelve kilometres

Most of the literature reviewed was examining schedule adherence, i.e. failure rates (Abkowitz & Engelstein 1984; Currie, Douglas & Kearns 2012; El - Geneidy, Horning & Krizek 2011; Mazloumi, Currie & Rose 2008). A case study examining the below estimation hypothesis is needed before further examining the application of SPC, as SPC measures the behaviour of a system during successful delivery of a service, with the aim of ensuring the mean output and its variation remains within the acceptable tolerances. SPC is not monitoring failure rates (like lateness) as transport operators often do.

2. Method

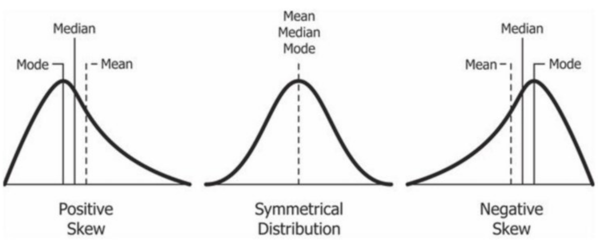
The objective of this research was to examine if the Measure Stabilise Reduce framework and Statistical Process Control, could be empirically demonstrated as being applicable to real world public transport operations, such as those in Sydney.

A realistic hypothesis for real world public transport operations is:

> the mean end-to-end runtime ($\bar{e}$) plus two standard deviations ($2\sigma_e$) **provides a reasonable estimate** for the 97th percentile of the observed end-to-end runtimes ($\mathbf{e}$).

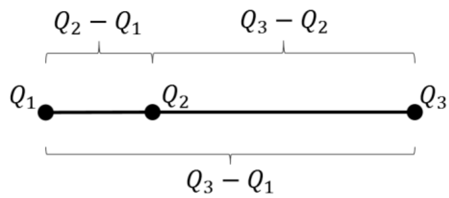
In response to Hounsell (2018a) and Hounsell (2018b), noted transport economist Neil Douglas, suggested that public transport runtimes would be positively skewed with a long tail. However, as the results below demonstrate, public transport runtimes are not *naturally skewed*, they are *artificially skewed* by timetables. This observation is clearly demonstrated in the results below for the IWLR and CSELR versus the Sydney Metro. During timetable creation, waiting time at stops/stations are added for buses/trams/trains to timetabled services to compensate for the natural variability of public transport operations. See also (NIST 2018).

Figure 2: Skewness of a statistical distribution - (Jain 2018)



This paper provides the Quantile/Galton/Bowley skewness for the measurements because it is simple to compute and provides a comparable value due to its bounded range of [-1, 1].

> ‘The quantile definition of skewness uses Q1 (the lower quartile value), Q2 (the median value), and Q3 (the upper quartile value). You can measure skewness as the difference between the lengths of the upper quartile (Q3-Q2) and the lower quartile (Q2-Q1), normalized by the length of the interquartile range (Q3-Q1). For a symmetric distribution, the quantile skewness is 0 because the length Q3-Q2 [equals] Q2-Q1. If the right length (Q3-Q2) is larger than the left length (Q2-Q1), then the quantile skewness is positive. If the left length is larger, then the quantile skewness is negative. … whereas the Pearson skewness can be any real value, the quantile skewness is bounded in the interval [-1, 1].’ (Wicklin 2017)



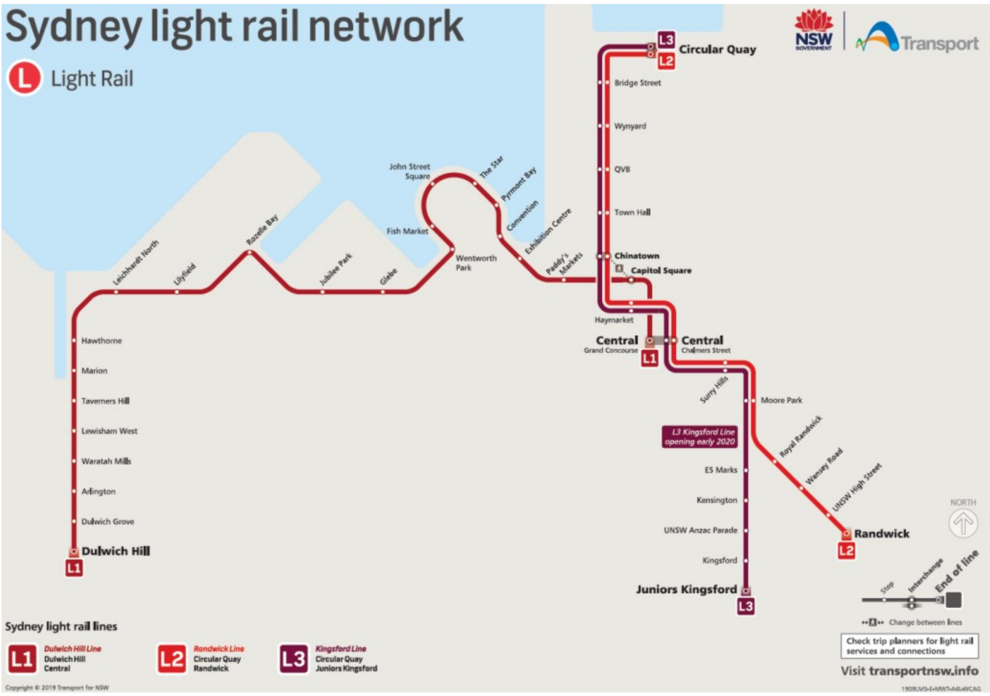
The light rail runtime data was provided by Transdev Sydney. The IWLR measurements are for the up-running time between Dulwich Hill and Central from September to December 2017. The CSELR dataset was derived from their measurements using their Automated Vehicle Location System for inbound services running between Central and Circular Quay. That section was chosen to maximise the number of data points. The data used was from May 2020 after the pandemic restrictions had loosened, and TfNSW had improved the traffic light sequencing, as examined in McRoberts-Smith (2020). Sydney Metro runtimes were from system testing of the automated control systems right before opening and was provided in response to an FOI.

For the CSELR study, only the L2 branch was operational. Central Station was a timing point. There were additional services between Central Station and Circular Quay. During the study that section had the most patronage. Since the purpose of this case study is to see if the delivered times can be estimated, this paper examines running time on just that primary section.

Figure 3: Sydney Metro within the Sydney Rail Network Map – (TfNSW 2016) [R2]

Due to its size, this map was appended as the last page of this paper.

Figure 4: Sydney Light Rail Network Map - (TfNSW 2016) [R1]



3. Results

The box plots for the light rail runtimes show a number of outliers and a slight positive skew for both datasets. It is important to note they do not have the same scale — the IWLR usually ranged between (31..41) minutes and the CSELR ranged between (15..29) minutes.

Figure 5: Distribution of all end-to-end upline runtimes for the IWLR Q4 2017 and CSELR May 2020

Table 1: Key metrics for end-to-end runtimes for the IWLR Q4 2017 and CSELR May 2020

From All Data Points	IWLR	CSELR	Units
Sample Count	8762	2766	samples
Min	29.3	12.8	minutes
Lower Quartile	34.7	19.1	minutes
Mean (95% CI) – $\bar{e}$	35.9 (±0.038)	21.3 (±0.127)	minutes (minutes)
Median	35.7	21.0	minutes
Upper Quartile	36.9	23.0	minutes
Max	65.8	61.8	minutes
Range	36.5	49.0	minutes
Standard Deviation – σ_e	1.78 (107 sec)	3.35 (201 sec)	minutes
Standard Error – $\hat{\sigma}_e$	0.019	0.064	minutes
Skew	1.539	2.507	
Galton/Bowley skew	0.078	0.013	
Four Sigma – $\bar{e}+2\sigma_e$	39.49	27.96	minutes
97.72 Percentile	39.99 (+0.5)	27.98 (+0.02)	minutes
Outliers	< 31.5 and > 40.1	< 13.2 and > 28.8	minutes
Outlier Count	203 (2.3%)	55 (2.0%)	samples

The metro performance tests in Figure 6 below, appear to have more spread until you see they normally range from (36.8 to 37.8) minutes. That was achieved through the use of a dedicated carriageway, lightweight construction, excellent acceleration/braking, and being computer controlled. Those operating benefits will in future be weighed against the construction costs.

In contrast to the light rails, the skew of these metro distributions is strongly positively. However, these plots and tables are not enough to prove/disprove the above hypothesis.

Figure 6: Distribution of all end-to-end runtimes for Sydney Metro final performance testing Sep 2019

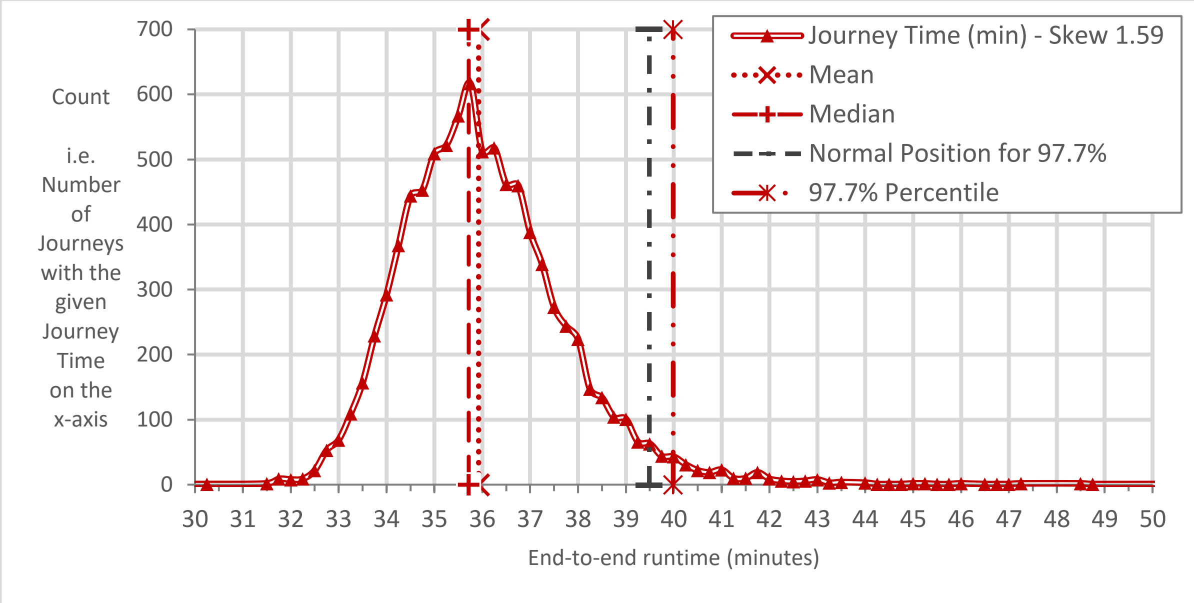
Table 2: Key metrics for end-to-end runtimes for the Sydney Metro final performance testing Sep 2019

From All Data Points	Up First	Down First	Up Second	Down Second	Units
Sample Count	1601	1601	1601	1601	samples
Min	36.8	36.8	35.9	36.0	minutes
Lower Quartile	37.0	37.1	36.9	37.0	minutes
Mean (95% CI) – $\bar{e}$	37.1 ±0.021	37.3 ±0.036	37.0 ±0.030	37.2 ±0.037	minutes
Median	37.1	37.2	36.9	37.1	minutes
Upper Quartile	37.2	37.3	37.0	37.2	minutes
Max	45.0	54.0	49.6	53.7	minutes
Range	8.3	17.2	13.7	17.8	minutes
Std Deviation – σ_e	0.43	0.71	0.60	0.75	minutes
Std Error - $\hat{\sigma}_e$	0.011	0.012	0.015	0.018	minutes
Skew	10.06	12.58	11.40	11.07	
Galton/Bowley skew	0.111	0.077	-0.200	0.077	
Four Sigma – $\bar{e}+2\sigma_e$	37.98	38.73	38.22	38.69	minutes
97.72 Percentile	38.00 (+.02)	38.58 (-.15)	38.23 (+.01)	38.67 (-.02)	minutes

3.1. Distribution of all measured running times

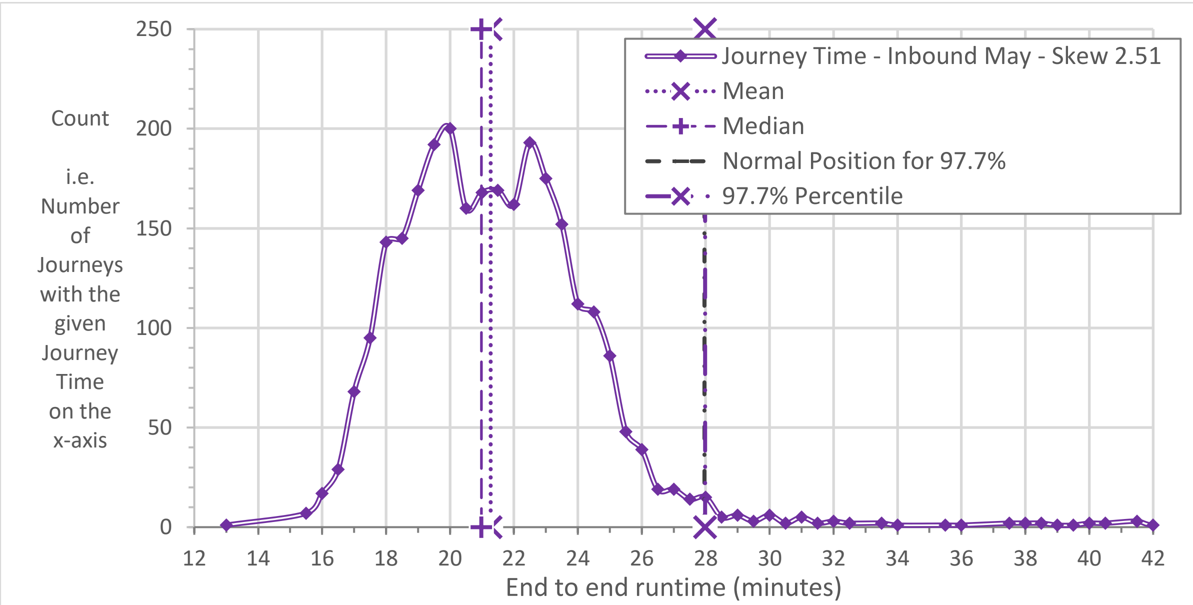
The IWLR is authorised to provide a public transport service with a vehicle-headway that TfNSW describes as a 'turn-up-and-go'. Figure 7 below shows the distribution for the IWLR end-to-end runtimes, the median is below the mean and the skew is positive. The measured 97th percentile of 39.9 minutes is 0.5 minutes greater than predicted by the normal distribution ($\bar{e} + 2\sigma_e$). Given the large sample size (>8,000), it is reasonable to conclude that the normal model does provide a reasonable estimate for the 97th percentile of end-to-end runtimes.

Figure 7: Count of all end-to-end upline runtimes for the IWLR Q4 2017



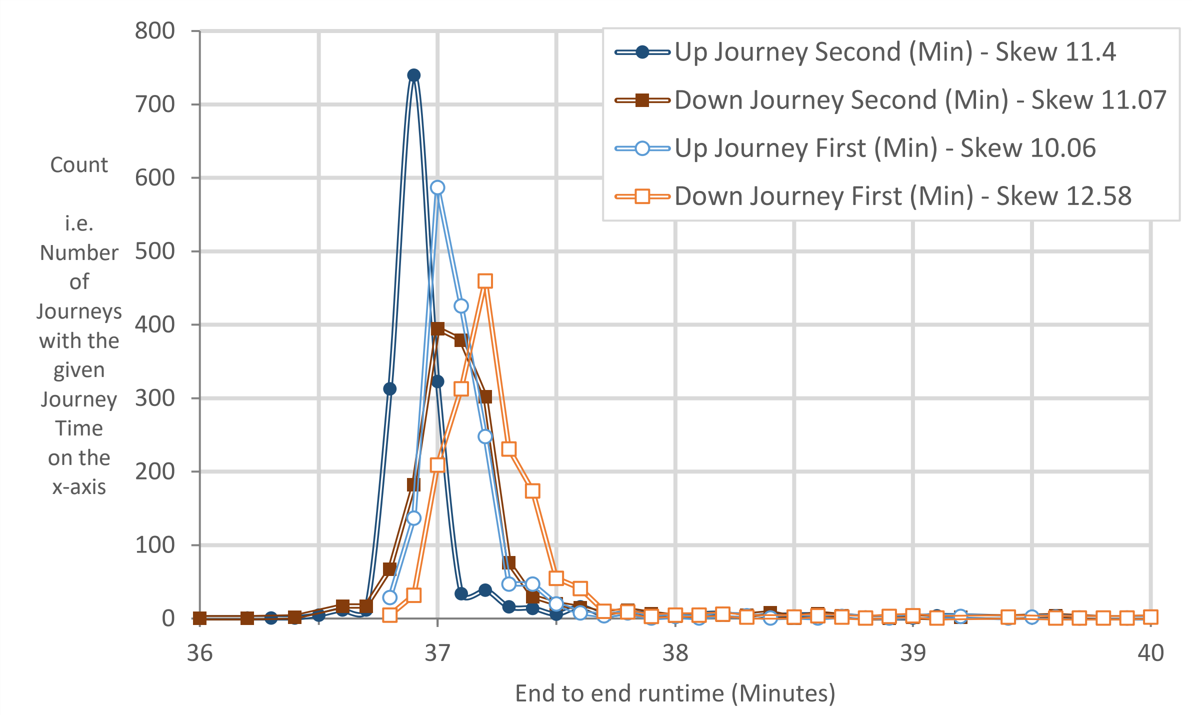
The CSELR is also authorised with Key Performance Indicators (KPI) focused on *headway-adherence*. In contrast to the IWLR, the CSELR from Central to Circular Quay is primarily on-street running, through many traffic lights and long stretches of pedestrian malls. As such the distribution of CSELR runtimes has a much larger range, and due to variability in signalling it displays multiple peaks. However, the difference between the normal model and measure 97th percentile is only 0.02 minutes which shows the CSELR does not invalidate the hypothesis.

Figure 8: Count of all end-to-end inbound runtimes – CS to CQ – for the CSELR May 2020



The Sydney Metro demonstrated a strongly positive skew suggesting timetable adherence. However, the absolute difference between the normal model estimates for the 97th percentile and the observed measurements was less 9 seconds, which does not invalidate the hypothesis.

Figure 9: Count of all end-to-end runtimes for Sydney Metro final performance testing Sep 2019



As such, these initial results suggest that the hypothesis is valid for frequent service which are operated to either a timetable or a specific vehicle-headway. As such, it looks like the mean end-to-end runtime ($\bar{e}$) plus two standard deviations ($2\sigma_e$) **provides a reasonable estimate** for the 97 percentile of the observed end-to-end runtimes (**e**). Additional analysis is presented below to expand upon these results.

3.1. Normal probability plots

A normal probability plot uses cumulative probability, z-scores, and a linear regression to assess whether a dataset has a nearly normal distribution. Figure 10 and Figure 11 show the IWLR and CSELR runtimes are nearly normally distributed with a larger tail (R^2>98%). Figure 12 indicates that as expected the Sydney Metro runtimes is not normally distributed.

Figure 10: Normal probability plot of end-to-end upline runtimes for the IWLR Q4 2017

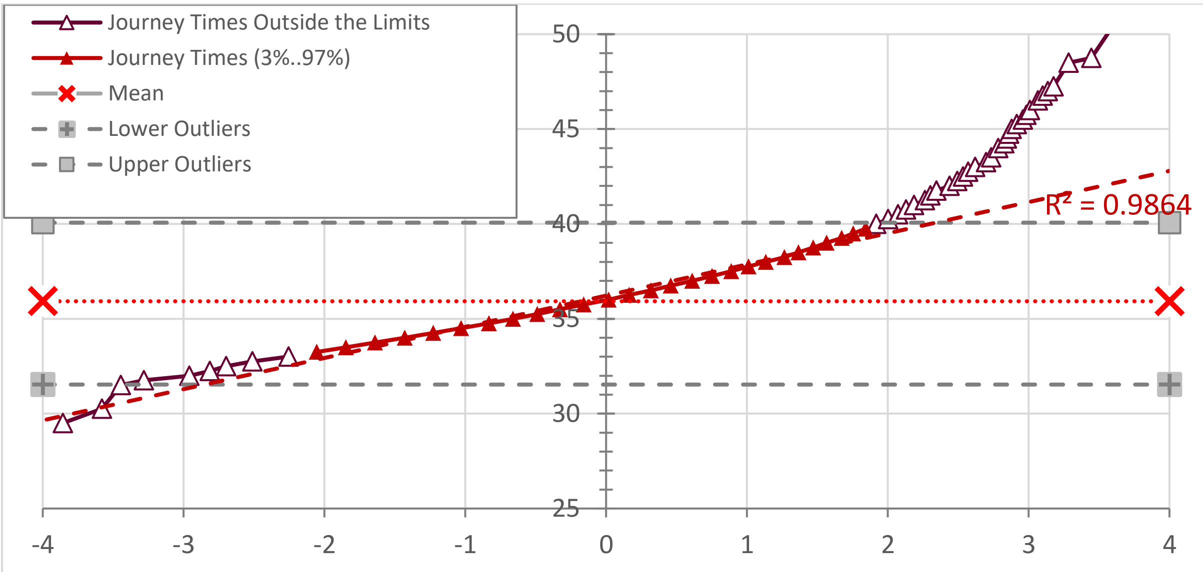


Figure 11: Normal probability plot of end-to-end inbound runtimes – CS to CQ – for the CSELR

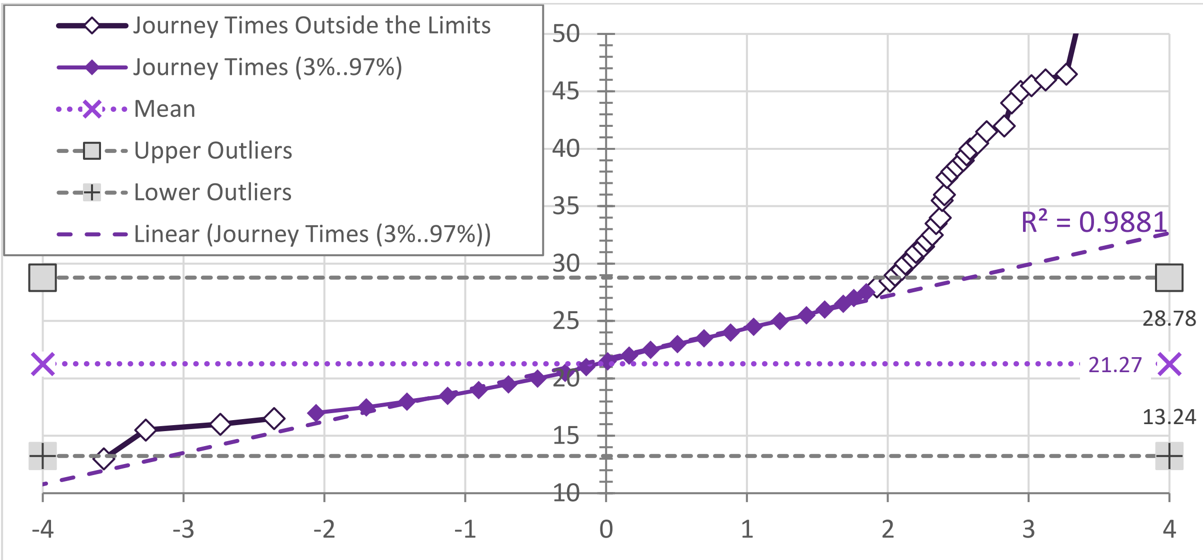
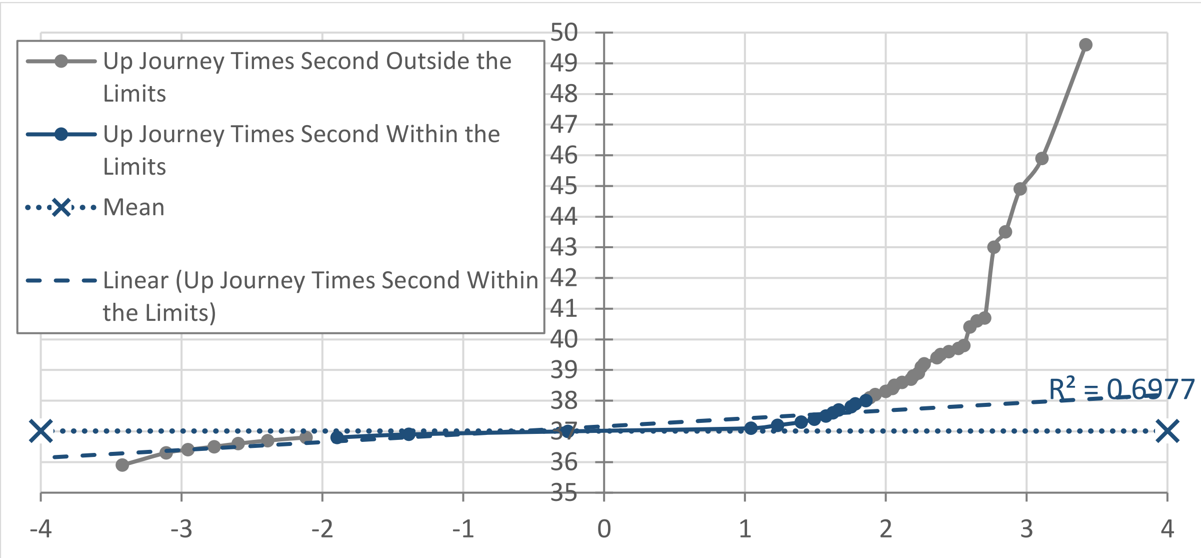


Figure 12: Normal probability plots of end-to-end runtimes for Metro final performance testing Sep 2019



3.1. Distribution and cumulative probability

Two other ways to examine the similarity between the normal distribution and a dataset are the cumulative distribution and percent point function plot. Figure 13 and Figure 14 below show the observed distribution for the IWLR in red and the normal distribution model generated from that dataset in black. Both charts mark the mean and the 97th percentiles. Both charts, but especially Figure 13, show that the observed data is nearly normal. Both charts show that the mean end-to-end runtime ($\bar{e}$) provides a reasonable estimate for the middle observed end-to-end runtimes ($\mathbf{e}$), and also that the mean plus two standard deviations ($\bar{e} + 2\sigma_e$) does provide a reasonable estimate for the 97th percentile, which does not invalidate the hypothesis.

Figure 13: Cumulative probability of end-to-end upline runtimes for the IWLR Q4 2017

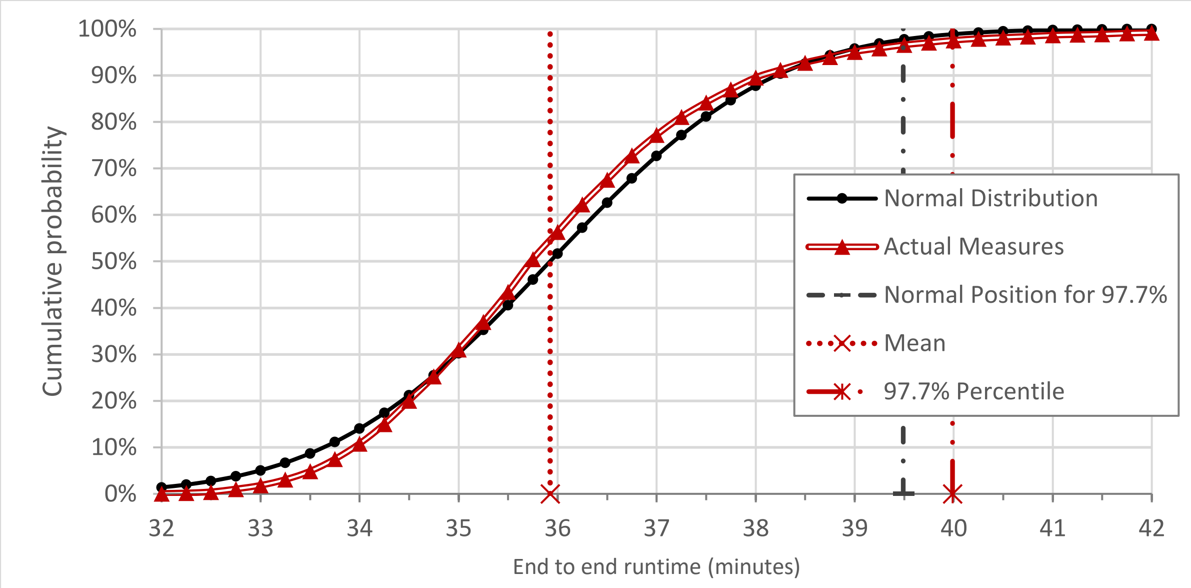
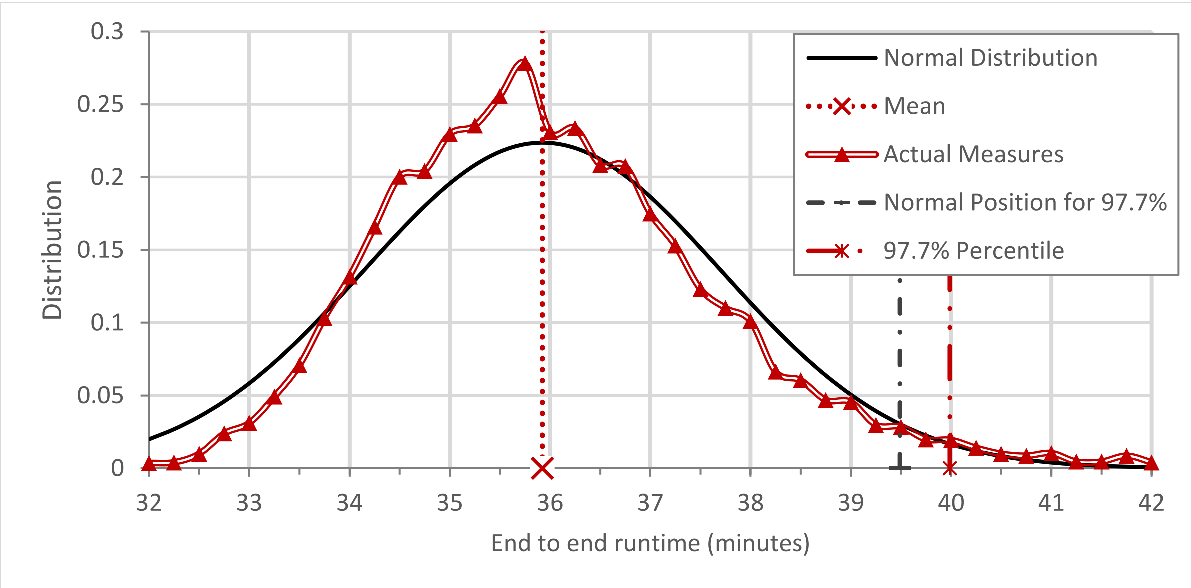


Figure 14: Distribution of end-to-end upline runtimes for the IWLR Q4 2017



Similarly, Figure 15 and Figure 16 below show that the CSELR runtimes could also be estimated using a normal distribution model as the middle and the 97th percentile can be accurately estimated using a normal model. That is despite the observed data having two peaks due to the design and signalling on the George St section of the CSELR.

Figure 15: Cumulative probability trimmed end-to-end runtimes – CS to CQ – for the CSELR May 2020

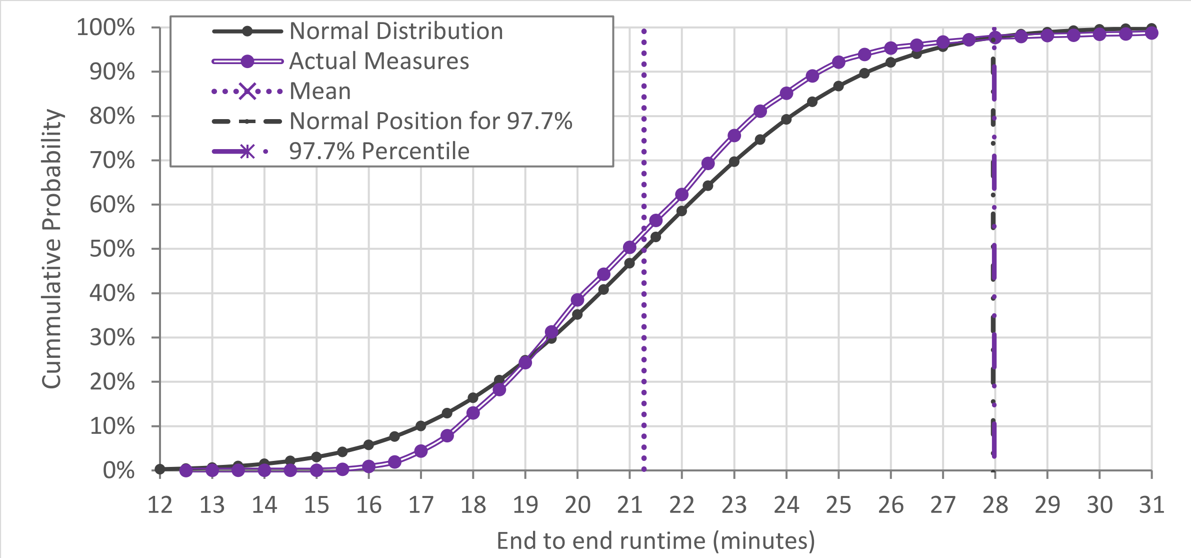
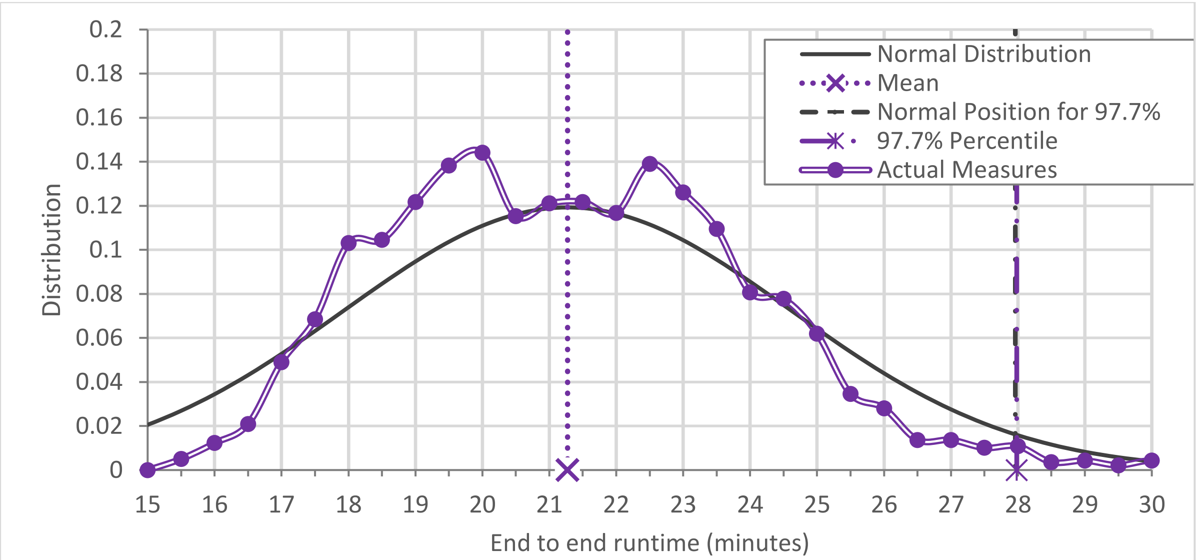


Figure 16: Distribution of end-to-end inbound runtimes – CS to CQ – for the CSELR May 2020



3.1. Distribution and cumulative probability for a right skewed distribution

Figure 17 below is the plot of the cumulative probability of measurements for the Sydney Metro. In contrast to the nearly normal natural runtimes of the IWLR and CSELR, the highly skewed artificial runtimes of the Sydney Metro are clear when plotted against the cumulative probability of their normal distribution model.

The left half of Figure 18 shows the full set of observations plotted against their normal distribution model. These plots also show these end-to-end runtimes are artificial and tightly controlled, but with a substantial and impactful tail that affects the normal model and making it too shallow to model the full runtimes. However, the right half of these figures plot the trimmed set of observations where outliers have been removed. Further research should examine if a normal or another model could better represent the skewed dataset.

However, the Sydney Metro plots below indicate that $\bar{e} + 2\sigma_e$ for the normal models may provide an accurate estimate of the 97th percentile for artificial and tightly controlled end-to-end runtimes at the human-scale resolution of minutes.

Figure 17: Cumulative probability of end-to-end runtimes for Metro final performance testing Sep 2019

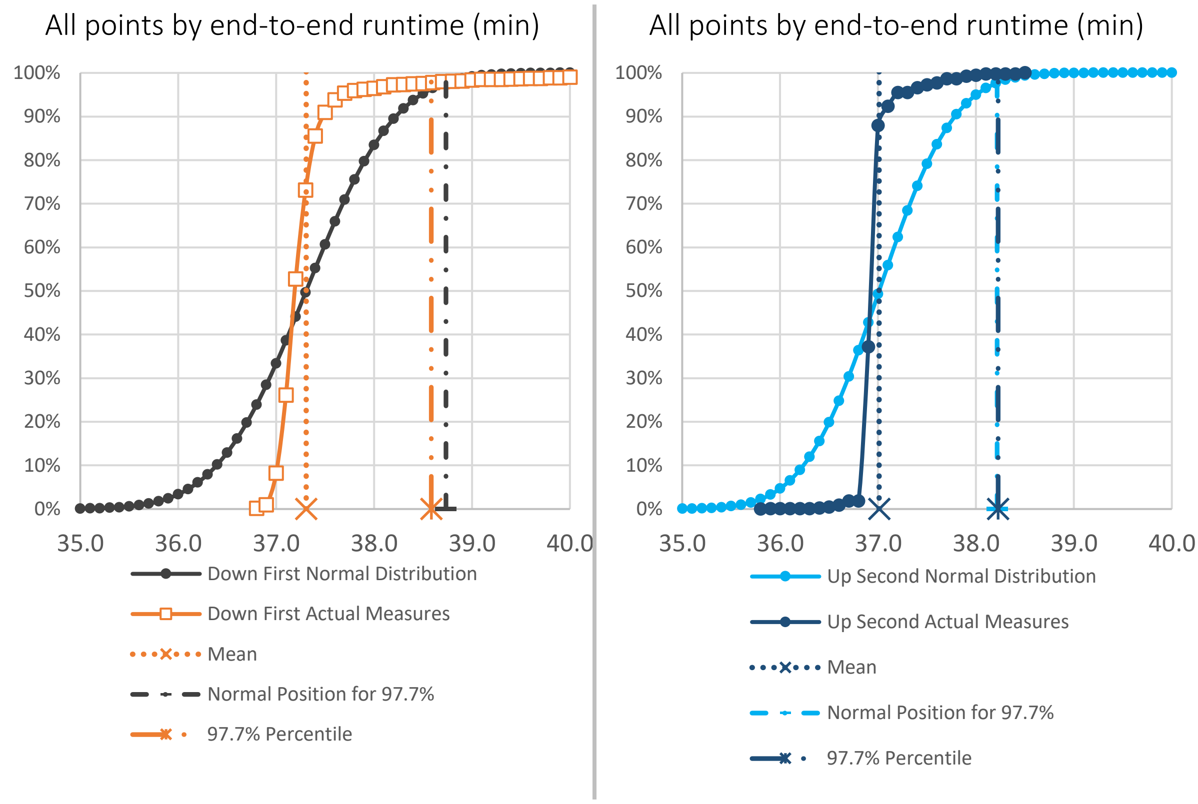
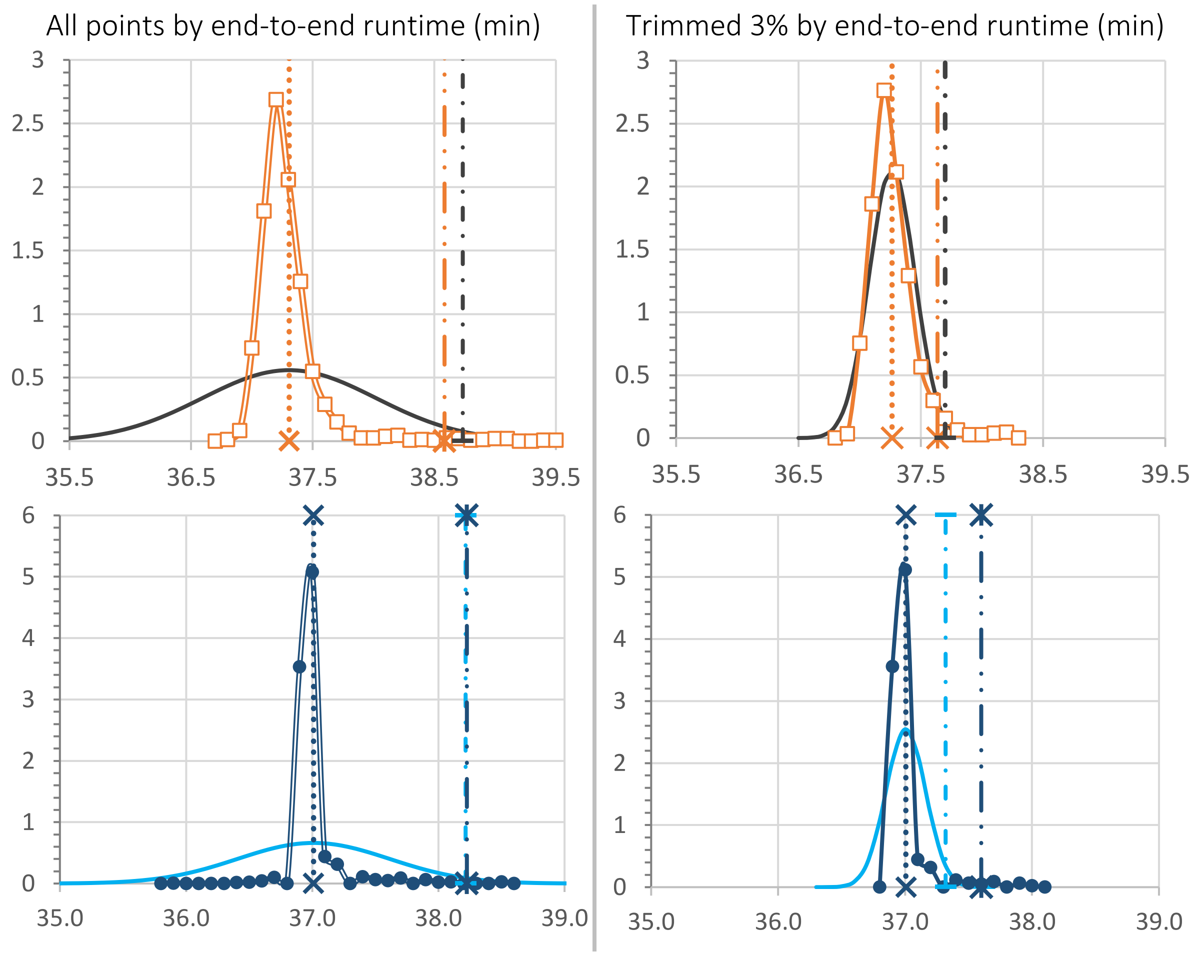


Figure 18: Distribution of end-to-end runtimes for Metro final performance testing Sep 2019



4. Discussion

This paper examined the hypothesis that for real world natural public transport operations, the normal-distribution model, the mean end-to-end runtime ($\bar{e}$), and the mean plus two standard deviations ($\bar{e} + 2\sigma_e$) **provide reasonable estimators** for the distribution, middle, and 97th percentile of the observed end-to-end runtimes respectively. As shown above, when the system is operating with natural runtimes (especially when focused on delivering a given vehicle-headway) then the normal model, mean, and standard deviation are useful estimators for observed runtimes, thus ***the hypothesis has not been invalidated***.

With tightly controlled and artificial runtimes on an automated grade-separated railway, such as the Sydney Metro, then the normal model is not applicable for these highly skewed runtimes. However, even in such an artificial situation the mean and the standard deviation still **provide reasonable estimators** for the middle and the 97th percentile of service delivery.

Therefore, then Statistical Process Control may be applicable to deliver an efficient system, especially if a system is operating with natural runtimes. This paper has shown that the mean, standard deviation, and other statistical measures used above could provide inputs to continuous optimisation frameworks, such as Measure, Stabilise, and Reduce, see Figure 1.

In conclusion, considering all the results above, the mean end-to-end runtime ($\bar{e}$), and the standard deviation (σ_e) are clearly useful measurements and provide reasonable estimators to measure the reliability and efficiency of the public transport services delivered.

As a postscript, the efficiency of the service will be determined by the reliability of the running times, since the number of vehicles is affected by running time variability. In Equation 1, the number of vehicles and crew required to deliver a service (**V**) for a targeted vehicle-headway (**h**) and minimum turn-around times (**f**, **l**), is proportional to the key variable of delivered end-to-end runtime (**e**) for inbound services (e_i) and for outbound services (e_o). Vehicles will wait at the first and last terminus to maintain a headway, so the actual terminal time (t_t) at each terminus will be longer than minimums turnaround times given (**f**, **l**). If an operator minimises variability on a route, they can minimise wasted waiting time.

Equation 1: End-to-end runtime including variation for a given number of vehicles at a given headway

97.72% Confidence

$$e_{normal} = \bar{e} + 2\sigma_e \qquad \therefore V = \left\lceil \frac{\bar{e_i} + 2\sigma_{e_i} + \bar{e_o} + 2\sigma_{e_o} + f + 2\sigma_f + l + 2\sigma_l}{h} \right\rceil$$

For a PSO to maintain legitimacy and public support they must transparently demonstrate they are reliably achieving efficient operations and minimising resource appropriations. As such, for every public transport run, the end-to-end runtimes and any waiting times (for timetable adherence) should be reported to the service authorisers to allow independent analysis of the PSO's efforts at reducing variability while providing their public transport services.

5. References

R1	TfNSW, (2019-12-04), 'Sydney Light Rail Network Map', State of NSW, UUID ae2340c7-88b6-4ff4-ba66-c08fa5eb17d5, https://opendata.transport.nsw.gov.au/dataset/network-maps/resource/571b96ec-bcd0-48fc-8fb2-2de09dcf3902
R2	TfNSW (2021-05-14), 'Sydney Rail Network 14 May 2021', State of NSW UUID: 543108c1-f8c9-4330-a69d-30e7836e490e, https://opendata.transport.nsw.gov.au/dataset/network-maps/resource/395b5ff5-aff2-43b2-a539-162bec0a2f80

Abkowitz, M. & Engelstein, I. 1984, 'Methods for maintaining transit service regularity', *Transportation Research Record*, vol. 961, pp. 1-8.

Brussee, W. 2006, *All about six sigma : the easy way to get started*, McGraw-Hill, New York.

Currie, G., Douglas, N. & Kearns, I. 2012, 'An assessment of alternative bus reliability indicators', paper presented to the *Australasian Transport Research Forum*.

Deming, W.E. 2000, *Out of the crisis*, MIT Press, Cambridge, Mass.

Deming, W.E. 2018, *The new economics : for industry, government, education*, Third edition. edn, The MIT Press, Cambridge, Massachusetts.

El-Geneidy, A.M., Horning, J. & Krizek, K.J. 2011, 'Analyzing transit service reliability using detailed data from automatic vehicular locator systems', *Journal of Advanced Transportation*, vol. 45, no. 1, pp. 66-79.

Hounsell, M. 2018a, 'Better service through runtime savings – Inner West Light Rail case study', paper presented to the *Australasian Transport Research Forum*, Darwin, <https://atrf.info/papers/2018/files/ATRF2018_paper_61.pdf>.

Hounsell, M. 2018b, 'Implications of running time variability for passengers and operators–a case study using Bondi's 333 buses', paper presented to the *Australasian Transport Research Forum (ATRF)*, Darwin, <https://atrf.info/papers/2018/files/ATRF2018_paper_60.pdf>.

Hounsell, M. 2021, ' Arbitrated Service Quality — a synthesis of the Strategic Triangle and the Service Quality Loop — in the context of public transport customers' desire for reliability and the public's desire for value-for-money', paper presented to the *Australasian Transport Research Forum*, Brisbane.

Jain, D. 2018, *Relationship between mean and median under different skewness*, Bitmap, Wikimedia Commons, <https://en.wikipedia.org/wiki/File:Relationship_between_mean_and_median_under_different_skewness.png \ CC-BY-SA-4.0>.

Mazloumi, E., Currie, G. & Rose, G. 2008, 'Causes of travel time unreliability – a Melbourne case study', paper presented to the *Australasian Transport Research Forum*, Gold Coast, Queensland, Australia, <https://www.researchgate.net/profile/Ehsan_Mazloumi/publication/267680401_Causes_of_travel_time_unreliability_-_A_Melbourne_case_study/links/547457fb0cf29afed60f7968/Causes-of-travel-time-unreliability-A-Melbourne-case-study.pdf>.

McRoberts-Smith, D. 2020, 'The comparison of signal changes to the operation capability of the Sydney CBD and South Eastern Light Rail, its impacts on delays and runtimes and methods to measure, stabilise, and reduce variability', Capstone thesis, University of Technology Sydney, Ultimo, New South Wales, Australia.

National Institute of Standards and Technology 2018, *Engineering Statistics Handbook*, USA, United States of America,, viewed 2021-09-25, <https://www.itl.nist.gov/div898/handbook/index.htm>.

TfNSW 2016, *Network Maps*, Transport for NSW, State of New South Wales, Sydney, New South Wales, Australia, <https://opendata.transport.nsw.gov.au/dataset/network-maps>.

Wicklin, R. 2017, 'A quantile definition for skewness', *The DO Loop*, weblog, SAS, viewed 2021-09-25, <https://blogs.sas.com/content/iml/2017/07/19/quantile-skewness.html>.