Output Analysis

Barry L. Nelson & Seong-Hee Kim &
Chuljin Park

Output Analysis

- “Output Analysis” covers…
  - The design of the simulation experiment
    (# of replications, which systems to simulate)
  - Analysis of the data generated by the simulation
- A key goal of output analysis is to produce, or control, a measure of error on the simulation performance estimates.

Types of Simulation

- There are two fundamental types of systems simulations:
  - *Nonterminating (steady-state) simulation*: performance measures are over a conceptually infinite planning horizon. **LATER…**
  - *Terminating simulation*: performance measures are with respect to well defined initial and final conditions, very much like spreadsheet simulations.

Terminating Simulations

- A store that opens (empty) at 7 am, locks its doors at 11 pm, and closes when the last customer departs.
- A project that begins on a designated day and completes when all activities are finished.
- A portfolio’s performance is tracked over 6 months.

Initial Conditions

- Key to a terminating simulation is that the initial conditions at time 0 are well defined *and matter!*
  - Store opens empty
  - The project begins with no activities underway
  - The portfolio starts with the securities already purchased and at their current values.

Stopping Time

- Also important in terminating simulation is that the stopping time T_E is defined by the system *and matters.*
  - Store closes when the last customer leaves
  - Project ends when nothing left to do
  - Portfolio evaluation ends after 6 months

Output Analysis for Terminating Simulation

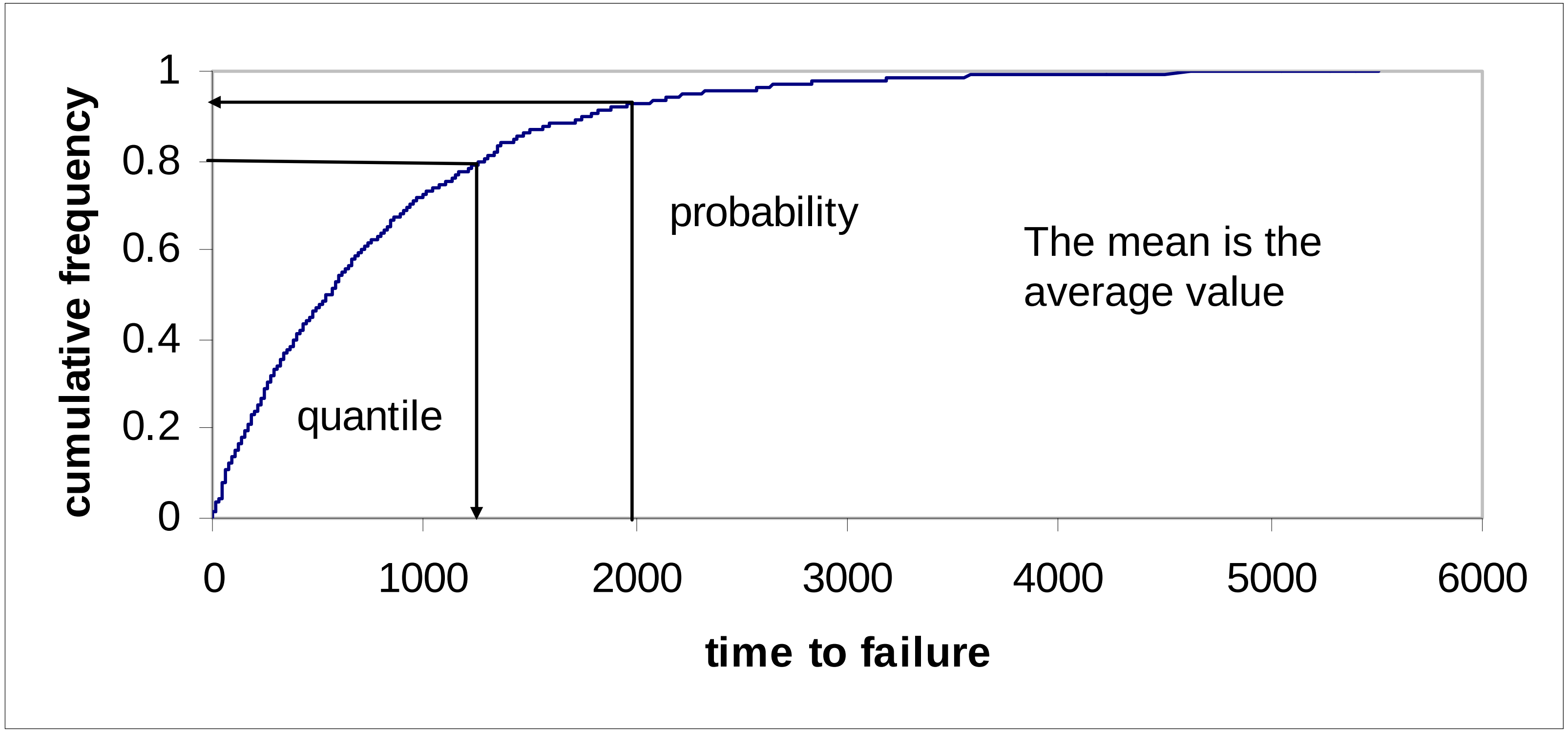
- We will start with design & analysis for terminating simulations.
- In terminating simulation the natural experiment design is to make **multiple replications** each from time 0 to T_E.
- The key issues are
  - What performance measure do we want?
  - How do we evaluate the error in the estimate?
  - How many replications should we make?

Performance Measures

- Suppose we are simulating a highly reliable system, and Y is the time until it fails.
- To evaluate different scenarios we might look at…
  - **Mean time to failure**: $\mu = E(Y)$
  - **Probability of early failure:** $p = \Pr\{Y \le c\}$
  - **Quantile of time to failure**: γ such that
    $$\Pr\{Y \le \gamma\} = p$$
    (a level of performance that can be delivered with a prespecified probability, also called **percentile**)

Performance Measures

Results from 1000 replications



Measure of Risk/Error

- Measure of Error (MOE): Tells whether estimated values are accurate enough to use in a decision making
  - standard error
  - confidence intervals
- Measure of Risk (MOR): Anything related to the distribution of outputs
  - standard deviation (S)
  - quantiles and probabilities
- Ignoring MOR can easily make a good simulation go bad.

Output Data

- When we make *R* replications, we generate output data $Y_1, Y_2, \ldots, Y_R$
- Since different random numbers are used on each replication, the outputs are *independent.*
- Since the same model is used on each rep, they are also *identically distributed.*

Caution!

- Output data *within* a replication may be neither independent nor identically distributed.
  - Waiting times of successive customers in a queue are dependent (correlated)
  - Waiting times when the queue opens may tend to be shorter than later in the day; thus not identically distributed

When are Output Data i.i.d.?

- Statistical tools for i.i.d. data apply to *summary measures* across replications.
  - If on each replication we extract a **single observation** of each performance measure, like time to failure, then we are ok.
  - If the output from each replication is a **summary measure**, like the **average** of **all** customer waiting times, then we are ok.

Review: Mean

$$\mu = E(Y)$$

$$\bar{Y} \pm t_{\alpha/2,R-1}\frac{S}{\sqrt{R}}$$

This is often called the "standard error;" it is the average deviation of the sample mean from the true value μ

$$\bar{Y} = \frac{1}{R}\sum_{r=1}^{R} Y_r$$

$$S^2 = \frac{1}{R-1}\sum_{r=1}^{R}(Y_r - \bar{Y})^2$$

Simio automatically computes 95% confidence intervals (α = 0.05) for the mean when we make multiple replications.

Review: Probabilities

$$p = \Pr\{Y \le c\}$$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{R-1}}$$

This CI is valid when data are i.i.d. and *R* is large enough that the *t* quantile is the same as the standard normal *z* quantile.

$$\hat{p} = \frac{1}{R}\#\{Y_r \le c\}$$

To get such a count from Simio we have to add additional statistics collection. This works for both dependent and independent data.

Review: Quantiles/Percentiles

- A quantile is the inverse of a probability; it tells us what level of performance can be delivered with a prespecified probability.

This is given.

$$\Pr\{Y \leq \gamma\} = p$$

- The best way is to sort the outputs and use the *Rp*th smallest value. This works for both dependent and independent data.
- In our example, the .75(100) = 75th smallest value.

CI For Quantiles

- Since a quantile is the inverse of a probability, we can also invert (sort of) the CI for a probability to get a CI for a quantile, provided that data are i.i.d.
- The approach is valid when the number of reps *R* is large; and the more extreme the quantile, the larger *R* needs to be.

Specifics

We define "confidence limits," but make use of the fact that *p* is known:

$$p_\ell = p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{R-1}}$$

$$p_u = p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{R-1}}$$

The end points of our c.i. for γ are the Rp_l and Rp_u smallest values (after sorting).

Quantile Confidence Interval

$$p_\ell = 0.75 - 1.96\sqrt{\frac{.75(1-.75)}{100-1}} = 0.66$$

$$p_u = 0.75 + 1.96\sqrt{\frac{.75(1-.75)}{100-1}} = 0.83$$

The c.i. is the 66th and 83rd smallest values

Computing a confidence interval for a quantile

$$p_\ell = 0.75 - 1.96\sqrt{\frac{.75(1-.75)}{100-1}} = 0.66$$

$$Rp_\ell = (0.66)(100) = 66$$

$$Rp = (0.75)(100) = 75$$

$$p_u = 0.75 + 1.96\sqrt{\frac{.75(1-.75)}{100-1}} = 0.83$$

$$Rp_u = (0.83)(100) = 83$$

J	K
62	0.63985
63	0.66355
64	0.6643
65	0.68903
66	0.70913
67	0.73676
68	0.76407
69	0.76775
70	0.79182
71	0.80682
72	0.82576
73	0.83048
74	0.83606
75	0.84137
76	0.84877
77	0.85618
78	0.86222
79	0.8663
80	0.8683
81	0.87523
82	0.88909
83	0.92094
84	0.99571
85	1.01067
86	1.067

What sorted values would we use if R=500 reps?

$$p_\ell = 0.75 - 1.96\sqrt{\frac{.75(1-.75)}{500-1}} = 0.71$$

$$Rp_\ell = (0.71)(500) = 355$$

$$Rp = (0.75)(500) = 375$$

$$p_u = 0.75 + 1.96\sqrt{\frac{.75(1-.75)}{500-1}} = 0.79$$

$$Rp_u = (0.79)(500) = 395$$

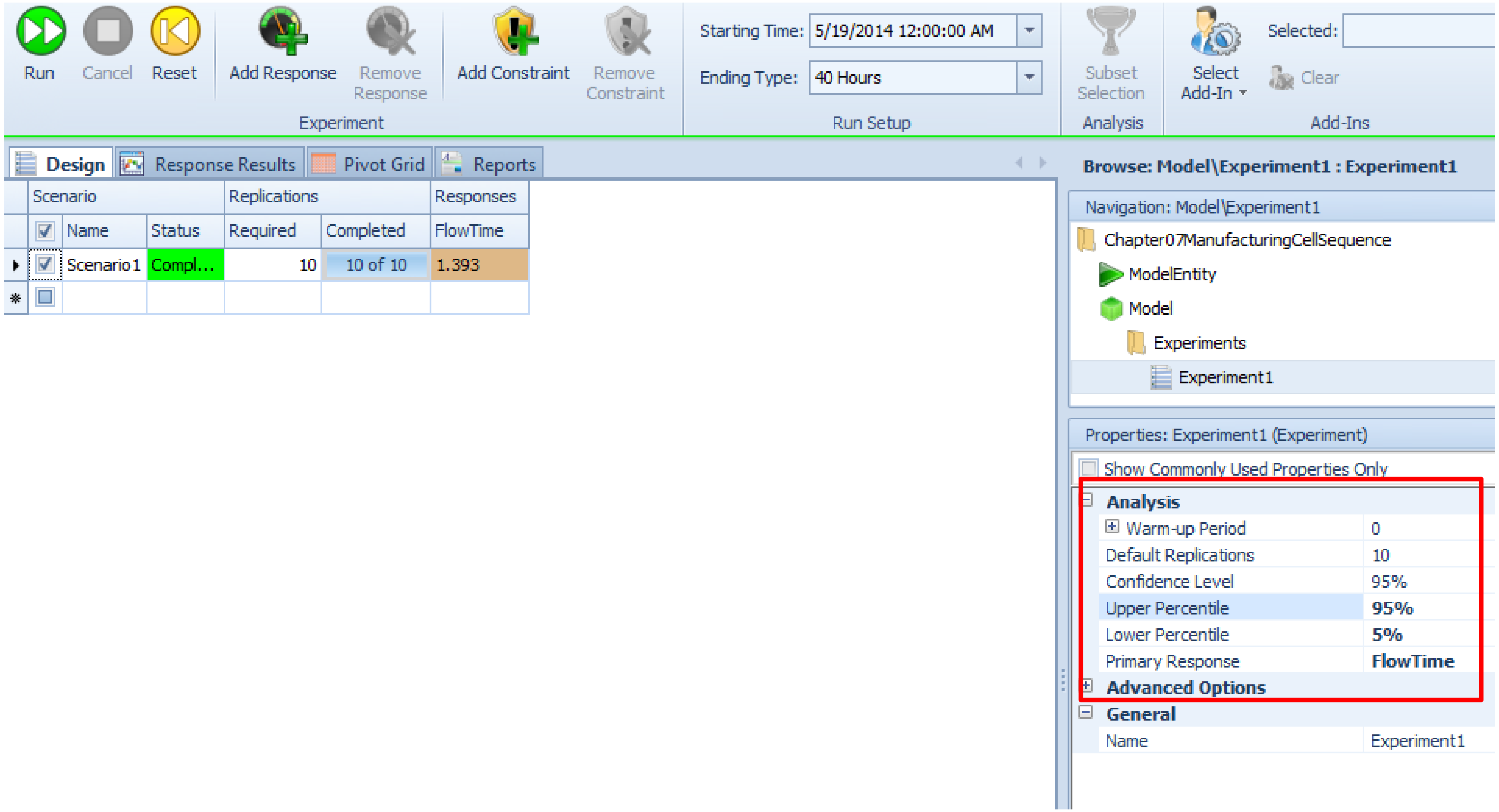
Confidence Interval vs. Prediction Interval

- A 95% confidence interval and a 95% prediction interval (sort of quantiles) are different:
  - A CI is a *measure of the error of a point estimate*; its length will shrink to 0 as we get more reps.
  - A PI captures what will *actually happen* in the future; its length will stabilize as we get more reps.
- Simio has automatic PIs.

What Simio Does

- From Simio Experiment, we can get
  - Histogram of across-replication data
  - Across-replication averages
  - Quantiles of across-replication data.

Simio Experiment



Simio MORE (SMORE) plot

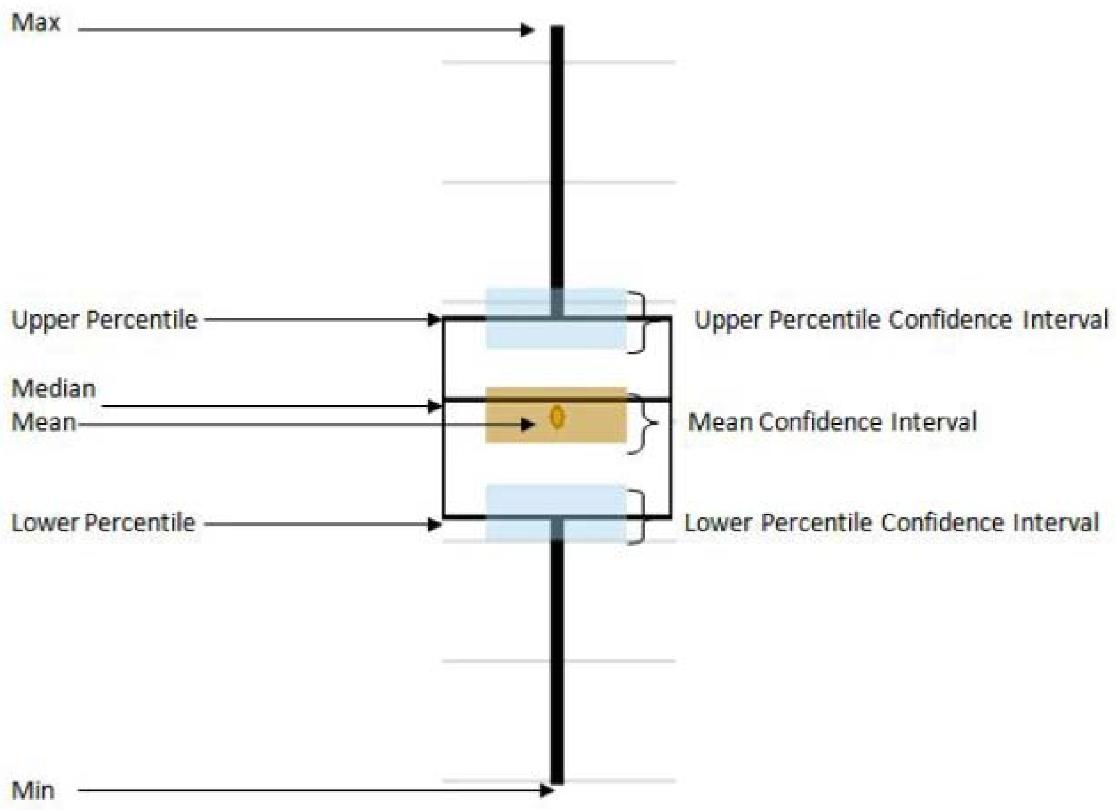
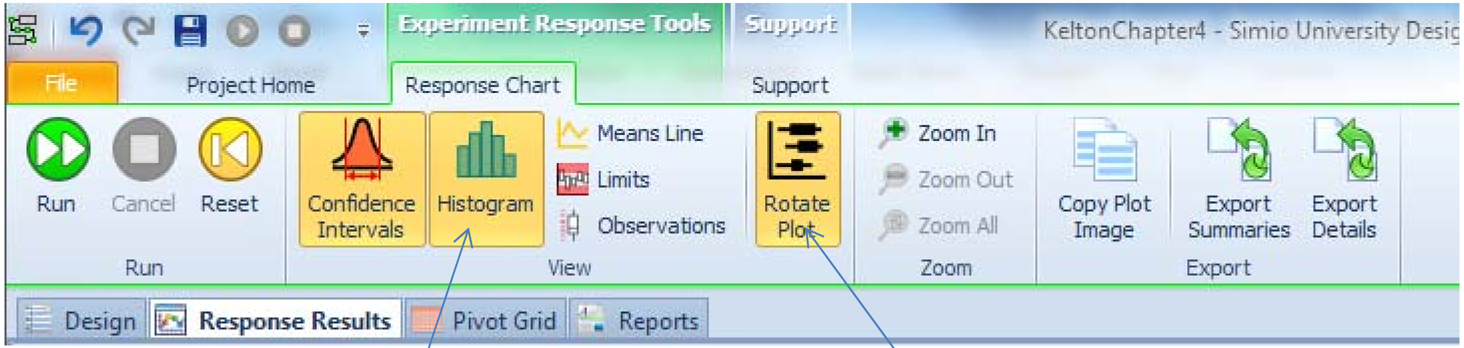


Figure 4.21: SMORE plot components (from the *Simio Reference Guide*).

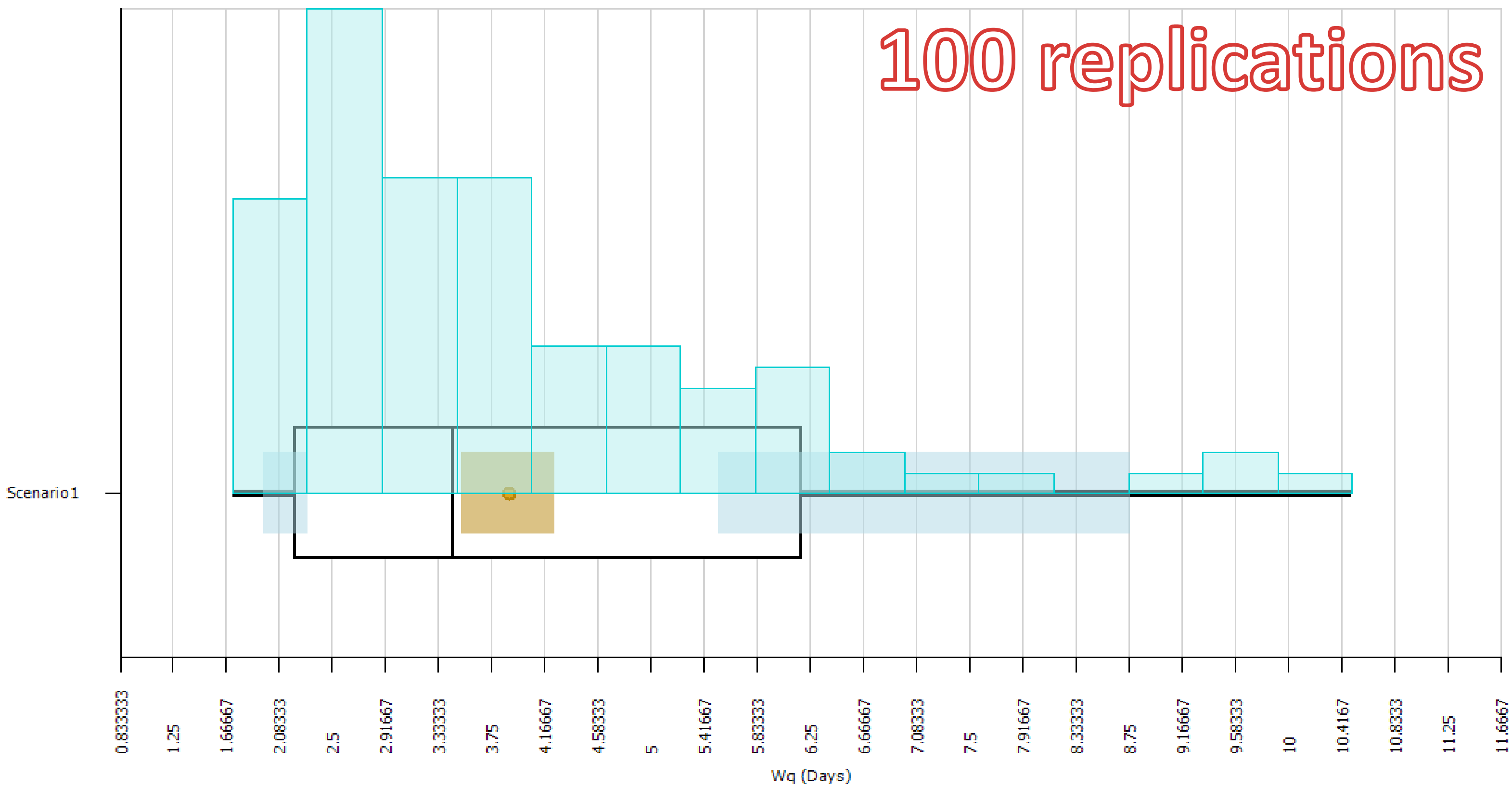


Click this button for
A histogram (an estimated pmf/pdf) of output data corresponding to a performance measure

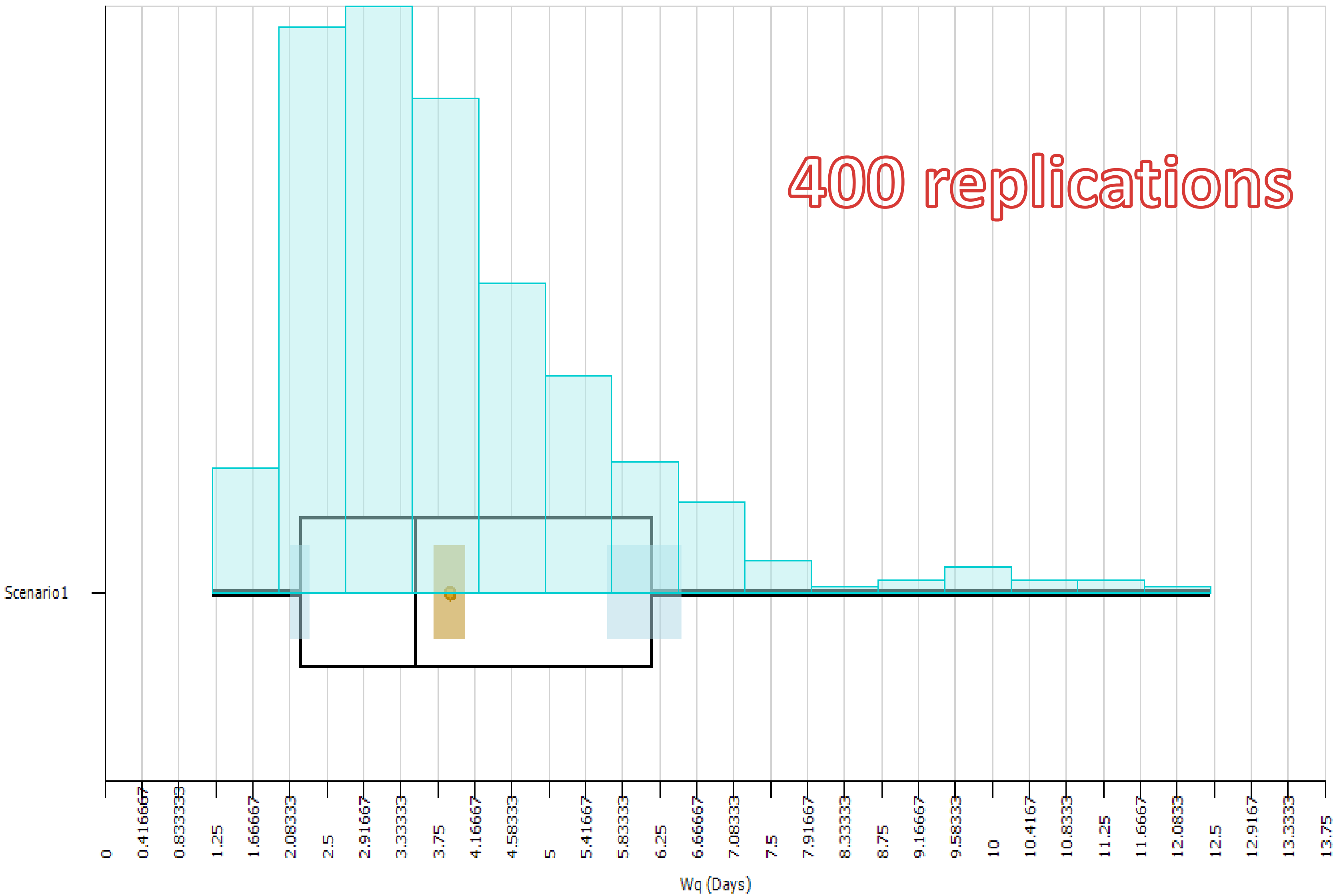
You may want to rotate SMORE plots for a horizontal view.

SMORE Plot

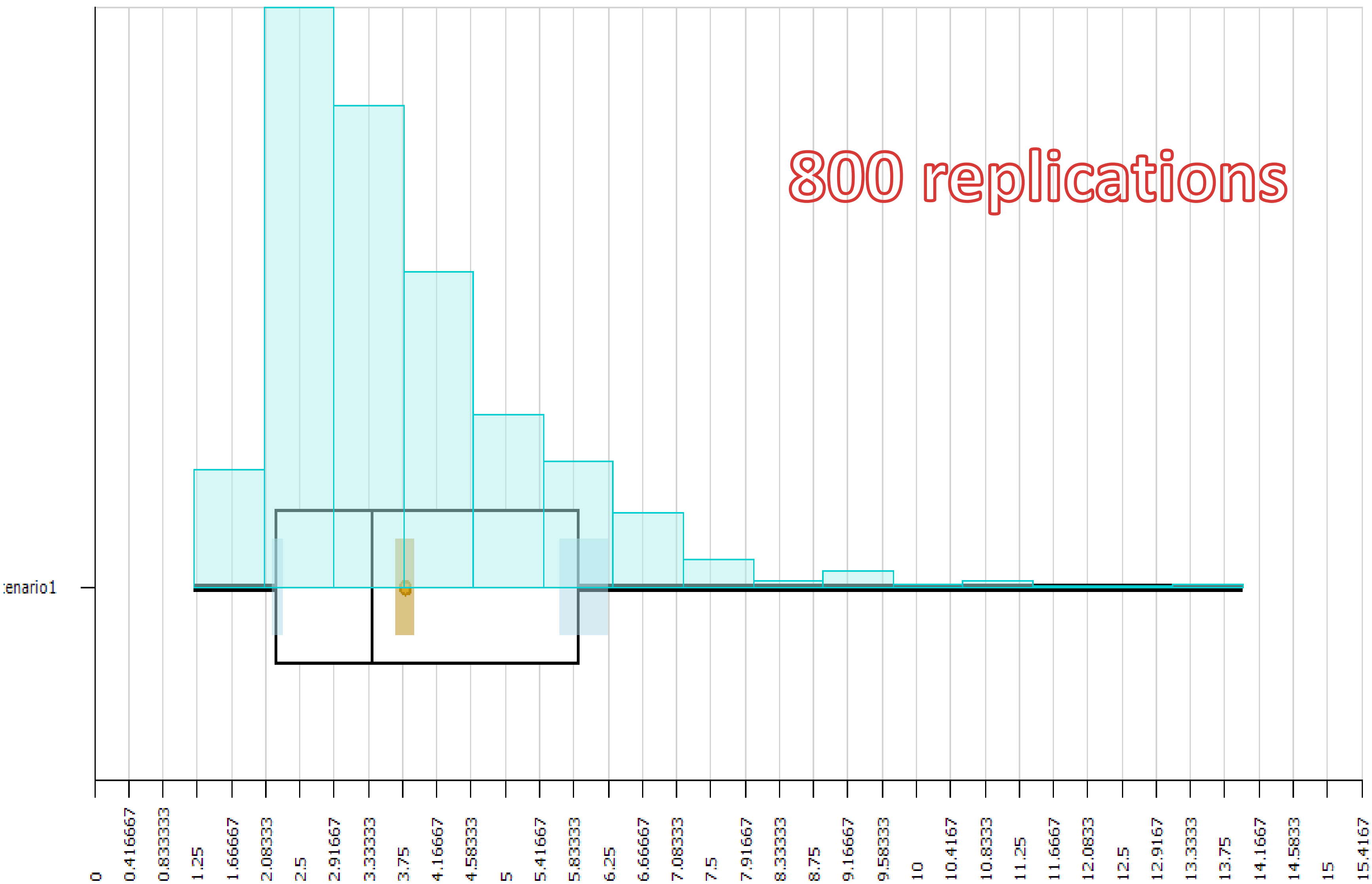
New package delivery system: 100 package delivery times
Is mean delivery time smaller than 4 days?
Does it ensure that at least 90% packages are delivered within 7 days?



- Average delivery time is 3.88 days.
  - But the true mean wait time in queue can be anywhere between 3.5 days and 4.2 days (with 95% confidence).
- A likely delivery time is between 1.96 days (10^{th} percentile) and 6.19 days (90^{th} percentile).
  - However, this is not an accurate interval because the 90^{th} percentile can be as large as 8.77 days.
  - Estimation error is large so we are not sure if 90% or more will be delivered within 7 days.


400 replications
Scenario1
0
0.416667
0.833333
1.25
1.66667
2.08333
2.5
2.91667
3.33333
3.75
4.16667
4.58333
5
5.41667
5.83333
6.25
6.66667
7.08333
7.5
7.91667
8.33333
8.75
9.16667
9.58333
10
10.4167
10.8333
11.25
11.6667
12.0833
12.5
12.9167
13.3333
13.75
Wq (Days)

- Average delivery time is 3.88 days.
  - The true mean wait time in queue can be anywhere between 3.7 days and 4.1 days (with 95% confidence).
- A likely delivery time is between 2.22 days (10^{th} percentile) and 6.17 days (90^{th} percentile).
  - The 90^{th} percentile can be as large as 6.51 days.
- The new system seems to ensure that at least 90% packages will be delivered within 7 days but we need more replications to check if mean delivery time is less than 4 days.


800 replications
cenario1
0
0.416667
0.833333
1.25
1.66667
2.08333
2.5
2.91667
3.33333
3.75
4.16667
4.58333
5
5.41667
5.83333
6.25
6.66667
7.08333
7.5
7.91667
8.33333
8.75
9.16667
9.58333
10
10.4167
10.8333
11.25
11.6667
12.0833
12.5
12.9167
13.3333
13.75
14.1667
14.5833
15
15.4167

Setting the Number of Reps

- Reminder: We don't pick *R* arbitrarily, we pick it large enough to control the error in our estimates.

$$R \geq \left(\frac{t_{\alpha/2, n_0-1} S}{\varepsilon} \right)^2$$

where ε is the acceptable error in the estimate of the mean.

If we have only the half width…

- Since the Category Overview Report only gives us the half width *H*, we can approximate *R* by

$$R \geq R_0 \left(\frac{H}{\varepsilon} \right)^2$$

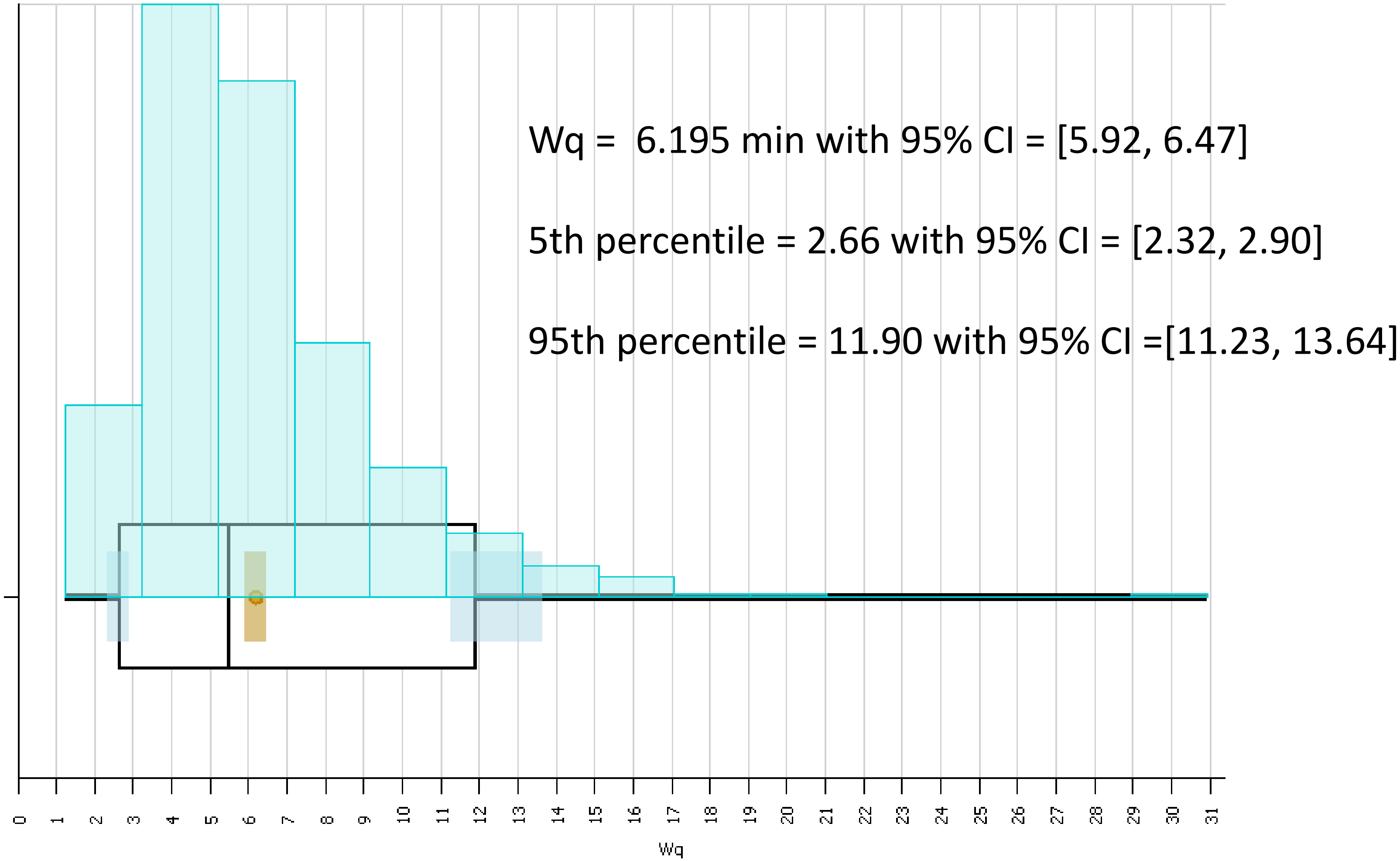
where *H* is the half length based on R_0 replications.

A Final Caution about Across-Rep Data

- Data are i.i.d., but the default stuff (Y_i) may be **averages** from within each rep (i.e., within-replication averages).
  - Note the difference between the Max and Min **Values** and the Min and Max **Averages**.
  - If we use these (Y_i) to derive probability and quantile estimates then they relate to the average, not the individual observations (e.g., individual wait times).

Fast-food Store

- A fastfood store runs for 8 hours per day.
- In Experiment, a response was SrvName.InputBuffer.Contents.AverageTimeWaiting.



Wq = 6.195 min with 95% CI = [5.92, 6.47]
5th percentile = 2.66 with 95% CI = [2.32, 2.90]
95th percentile = 11.90 with 95% CI =[11.23, 13.64]
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Wq

Interpretation

5th percentile = 2.66 mins with 95% CI = [2.32, 2.90]

1. Point estimate?
   a. A daily average wait time for the store is less than 2.66 mins 5% of time (i.e., approximaly 5 days out of 100 days).
   b. 5% customers experience <= 2.66 mins waiting time.
2. 2. Interval estimate: The true 5th percentile is between 2.32 and 2.90 mins with 95% confidence.

95th percentile = 11.90 mins with 95% CI =[11.23, 13.64]

1. Point estimate?
   a. A daily average wait time for the store is less than 11.90 mins approximately 95 days out of 100 days.
   b. 95% customers experience <= 11.23 mins waiting time.
2. Interval estimate: The true 95 percentile of daily averages is between 11.23 mins and 13.64 mins with 95% confidence.