HW2 Part II

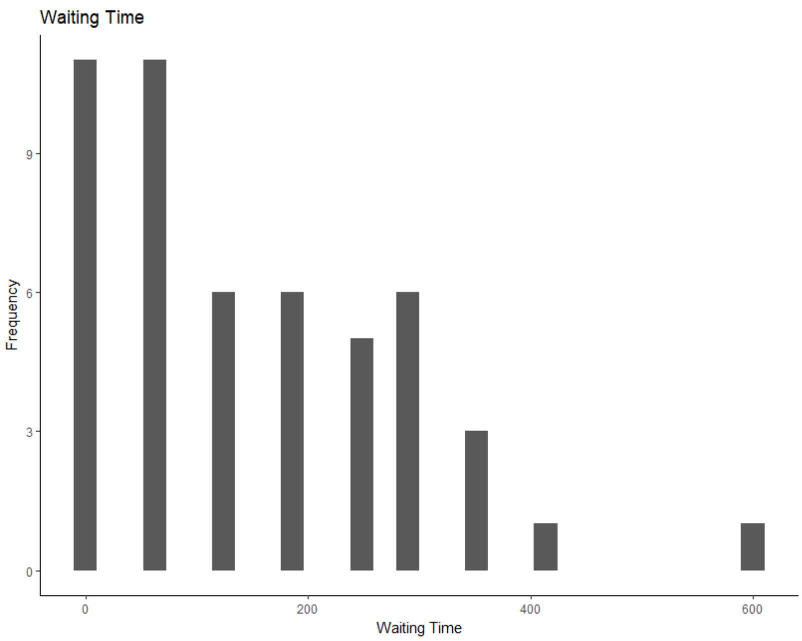
Monday, January 21, 2019 9:33 PM

Collected arriving time of 50 buses at the bus stop in front of Communications Building from 14:16 to 16:22 Jun 21st.

30	372E	15:31:00	60
31	67	15:31:00	0
32	277	15:36:00	300
33	32	15:36:00	0
34	540E	15:42:00	360
35	372E	15:47:00	300
36	31	15:48:00	60
37	67	15:48:00	0
38	78	15:51:00	180
39	855	15:57:00	360
40	32	15:57:00	0
41	67	15:59:00	120
42	67	16:02:00	180
43	540E	16:06:00	240
44	372E	16:06:00	0
45	31	16:07:00	60
46	67	16:12:00	300
47	277	16:15:00	180
48	372E	16:16:00	60
49	32	16:17:00	60
50	67	16:22:00	300

(a) The population here is the waiting time for any buses among 8 lines of buses ("277" "31" "32" "372E" "540E" "67" "78" "855") that way to the bus stop in front of CMU.

(b) The waiting time T should follow an exponential distribution.
The plotted distribution is based on my data.



(c) $\bar{T}_{20}$calculated from sample of waiting time for first 20 buses can be used as an unbiased estimator for mean of population distribution.

```
> mean(waiting_time[1:20])
[1] 159
> mean(waiting_time)
[1] 151.2
```

Based on my data
$\bar{T}_{20}$ = 159 seconds, 2.65 in minutes.

(d) I choose sample mean value of waiting time $\bar{T}_{50}$ as unbiased estimator.
Based on my data
$\bar{T}_{50}$ = 151.2 seconds, 2.52 in minutes.

(e) The estimator $\bar{T}_{50}$ is better in estimation of waiting time.
Since the size of $\bar{T}_{20}$ is 20 and $\bar{T}_{50}$ is 50, the standard error $\left(SE = \frac{\sigma}{\sqrt{n}}\right)$ of these two sample statistics are different. Standard error indicateds how variable a sample statistic is if the experiment is repeated multiple times. Small sample error indicates the sample statistic only varies by a small amount. The standard error of $\bar{T}_{50}$ is much smaller than $\bar{T}_{20}$ for its bigger denominator. True population parameter lies in the interval of sample statistics, so less the sample error is, the higher precision we we have to pinpoint the population parameter. Also, outliers will have less effect on $\bar{T}_{50}$.

(f) (Since I only did the observation one time, I'm unable to calculate $Var(\bar{T})$ from real data.)
Assuming population parameter a fixed number λ, so the variance of each trial, $Var(T_i) = \frac{1}{\lambda^2}$.
Both estimators are unbiased, that is $Bias(\bar{T}_{20}) = Bias(\bar{T}_{50}) = 0$.
Therefore, in this case, $MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2(\hat{\theta}) = Var(\hat{\theta}) + 0 = Var(\hat{\theta})$

$$MSE(\bar{T}_{20}) = Var(\bar{T}_{20}) = Var\left(\frac{\sum_{i=1}^{20} T_i}{20}\right) = \frac{\sum_{i=1}^{20} Var(T_i)}{20^2} = \frac{1}{20\lambda^2}$$

$$MSE(\bar{T}_{50}) = Var(\bar{T}_{50}) = Var\left(\frac{\sum_{i=1}^{50} T_i}{50}\right) = \frac{\sum_{i=1}^{50} Var(T_i)}{50^2} = \frac{1}{50\lambda^2}$$

Since we get $MSE(\bar{T}_{50}) < MSE(\bar{T}_{20})$, $\bar{T}_{50}$ is a better estimator.

(g) There are several ideas I perceived from this part of questions.

- Even if we choose the same type of estimator, in this case, I chose sample mean based on different size of sample as estimators, the operation of sample size has an influence on efficiency of the estimator. Sample mean of size of 50 has smaller mean square error than sample mean of size of 20.
- Bias is not the only one factor that we should consider when choosing an estimator. In this case, both estimators are unbiased since $E(\bar{X}) = \mu$, however, $\bar{T}_{50}$ has better efficiency than $\bar{T}_{20}$.
- Through my own observation, the waiting time of a bus does follow an exponential distribution.