Call center data modeling

In [1]:
```
# import libraries
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt

# load call center data set
waiting_times_day = np.loadtxt('call_center.csv')
```

In [3]:
```
# data pre-processing
current_time = 0
waiting_times_per_hour = [[] for _ in range(24)]  # Make 24 empty lists,
 one per hour
for t in waiting_times_day:
    current_hour = int(current_time // 60)
    current_time += t
    waiting_times_per_hour[current_hour].append(t)

for hour, waiting_times_hour in enumerate(waiting_times_per_hour):
    print('%02i:00-%02i:00 - %i calls' % (hour, hour + 1, len(waiting_ti
mes_hour)))
```

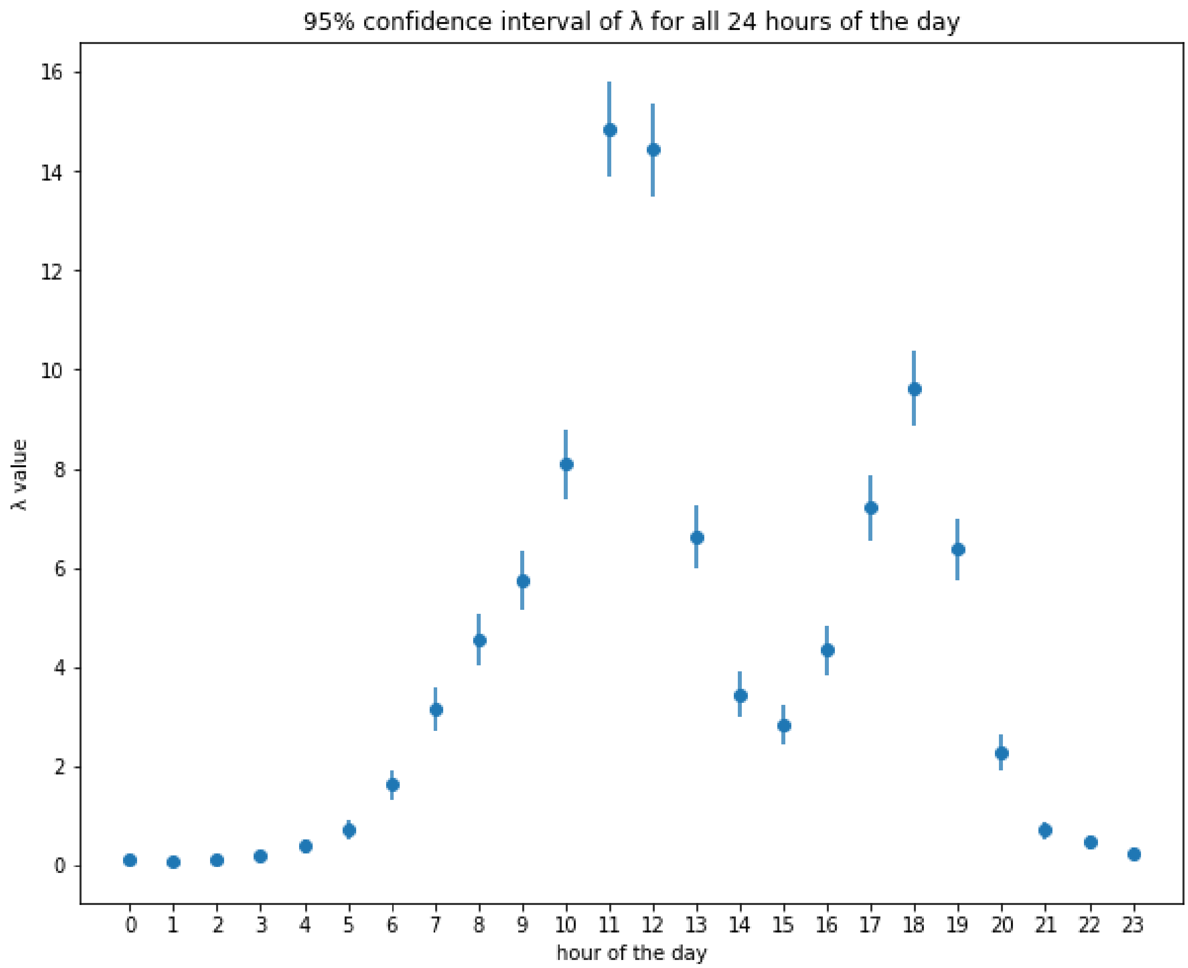
```
00:00-01:00 - 7 calls
01:00-02:00 - 5 calls
02:00-03:00 - 8 calls
03:00-04:00 - 7 calls
04:00-05:00 - 21 calls
05:00-06:00 - 42 calls
06:00-07:00 - 96 calls
07:00-08:00 - 189 calls
08:00-09:00 - 274 calls
09:00-10:00 - 344 calls
10:00-11:00 - 487 calls
11:00-12:00 - 892 calls
12:00-13:00 - 869 calls
13:00-14:00 - 401 calls
14:00-15:00 - 206 calls
15:00-16:00 - 169 calls
16:00-17:00 - 261 calls
17:00-18:00 - 430 calls
18:00-19:00 - 579 calls
19:00-20:00 - 383 calls
20:00-21:00 - 136 calls
21:00-22:00 - 45 calls
22:00-23:00 - 28 calls
23:00-24:00 - 12 calls
```

In [18]:
```
alpha_per_hour = []
beta_per_hour = []
posterior_per_hour = []

for hour, waiting_times_hour in enumerate(waiting_times_per_hour):

    # set prior hyperparameters
    alpha_0 = 1
    beta_0 = 0.25

    # calculate posterior hyperparameters
    alpha = alpha_0 + len(waiting_times_hour)
    beta = beta_0 + sum(waiting_times_hour)
    print("Posterior for hour {:d} is: α = {:.2f}, β = {:.2f}".format(ho
ur, alpha, beta))

    # create posterior distribution
    posterior = stats.gamma(a=alpha, scale=1/beta)
    print ("95% confidence interval over λ: [{:.2f}, {:.2f}]".format(post
erior.ppf(0.025), posterior.ppf(0.975)))
    print ("mean of λ: {:.2f}".format(posterior.mean()))
    # add data to set
    alpha_per_hour.append(alpha)
    beta_per_hour.append(beta)
    posterior_per_hour.append(posterior)
```

```
Posterior for hour 0 is: α = 8.00, β = 60.51
95% confidence interval over λ: [0.06, 0.24]
mean of λ: 0.13
Posterior for hour 1 is: α = 6.00, β = 67.87
95% confidence interval over λ: [0.03, 0.17]
mean of λ: 0.09
Posterior for hour 2 is: α = 9.00, β = 69.97
95% confidence interval over λ: [0.06, 0.23]
mean of λ: 0.13
Posterior for hour 3 is: α = 8.00, β = 44.65
95% confidence interval over λ: [0.08, 0.32]
mean of λ: 0.18
Posterior for hour 4 is: α = 22.00, β = 58.83
95% confidence interval over λ: [0.23, 0.55]
mean of λ: 0.37
Posterior for hour 5 is: α = 43.00, β = 60.18
95% confidence interval over λ: [0.52, 0.94]
mean of λ: 0.71
Posterior for hour 6 is: α = 97.00, β = 59.89
95% confidence interval over λ: [1.31, 1.96]
mean of λ: 1.62
Posterior for hour 7 is: α = 190.00, β = 60.18
95% confidence interval over λ: [2.72, 3.62]
mean of λ: 3.16
Posterior for hour 8 is: α = 275.00, β = 60.50
95% confidence interval over λ: [4.02, 5.10]
mean of λ: 4.55
Posterior for hour 9 is: α = 345.00, β = 60.06
95% confidence interval over λ: [5.15, 6.37]
mean of λ: 5.74
Posterior for hour 10 is: α = 488.00, β = 60.28
95% confidence interval over λ: [7.39, 8.83]
mean of λ: 8.10
Posterior for hour 11 is: α = 893.00, β = 60.09
95% confidence interval over λ: [13.90, 15.85]
mean of λ: 14.86
Posterior for hour 12 is: α = 870.00, β = 60.26
95% confidence interval over λ: [13.49, 15.41]
mean of λ: 14.44
Posterior for hour 13 is: α = 402.00, β = 60.72
95% confidence interval over λ: [5.99, 7.28]
mean of λ: 6.62
Posterior for hour 14 is: α = 207.00, β = 60.24
95% confidence interval over λ: [2.98, 3.92]
mean of λ: 3.44
Posterior for hour 15 is: α = 170.00, β = 59.93
95% confidence interval over λ: [2.43, 3.28]
mean of λ: 2.84
Posterior for hour 16 is: α = 262.00, β = 60.54
95% confidence interval over λ: [3.82, 4.87]
mean of λ: 4.33
Posterior for hour 17 is: α = 431.00, β = 59.80
95% confidence interval over λ: [6.54, 7.90]
mean of λ: 7.21
Posterior for hour 18 is: α = 580.00, β = 60.25
95% confidence interval over λ: [8.86, 10.43]
mean of λ: 9.63
```

```
Posterior for hour 19 is: α = 384.00, β = 60.30
95% confidence interval over λ: [5.75, 7.02]
mean of λ: 6.37
Posterior for hour 20 is: α = 137.00, β = 60.24
95% confidence interval over λ: [1.91, 2.67]
mean of λ: 2.27
Posterior for hour 21 is: α = 46.00, β = 66.21
95% confidence interval over λ: [0.51, 0.91]
mean of λ: 0.69
Posterior for hour 22 is: α = 29.00, β = 63.24
95% confidence interval over λ: [0.31, 0.64]
mean of λ: 0.46
Posterior for hour 23 is: α = 13.00, β = 53.39
95% confidence interval over λ: [0.13, 0.39]
mean of λ: 0.24
```

In [42]:
```
mean_per_hour = [post.mean() for post in posterior_per_hour]
confidence_95_per_hour = [post.mean() - post.ppf(0.025) for post in post
erior_per_hour]

plt.rcParams['figure.figsize'] = [10, 8]
plt.errorbar(x=list(range(0, 24)), y=mean_per_hour, yerr=confidence_95_p
er_hour, fmt='o')
plt.title("95% confidence interval of λ for all 24 hours of the day")
plt.xlabel("hour of the day")
plt.xticks(list(range(0, 24)))
plt.ylabel("λ value")
plt.show()
```



This graph visually represents the estimated incoming call rate per minute in the 24 hours a day. The dot represents where the actual average number of phone calls per minute (population mean) is most likely to occur, and the line above and below the dot shows the range where we are 95% confidence that the accrual population mean is. In addition, by looking through the pattern by the hour, the number of call increase as noon, peaking at 10-11th hour, and there is another increase of call between 17-19th hour. By using Bayesian approach, the 95% confidence interval is obtained, and by supplying the handling capacity represented by the upper bound, it means at current demand there is 97.5% chance that there will be no wait time on the call service.