Statistics: The Science of Decisions Project

Gorkem Berk Sahan

9/12/2017

Questions For Investigation

1. What is our independent variable? What is our dependent variable?

Our independent variables are word's meaning and word's color, dependent variable is time.

2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.

H0: There is no difference in population means of response time under incongruent and congruent scenarios

$\mu_c = \mu_i$, μ_c is mean of congruent population and μ_i is mean of incongruent population

H1 : The population mean of the response time under incongruent scenarios is different from congruent scenarios

$\mu_c \neq \mu_i$

We have two particular samples,
we need to compare two sample means,
we don't know population mean and standart deviation.

So I use two-tailed t-test for paired sample for this.

3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

When we look at the summary of datas we can see that mean and median is close each other, so we can say that both distrubition is normal distributed.

Central tendencys (mean, meadian) and measure of variability (IQR, range) like below ;

```
Congruent :

 Minimum value   :8.63
 Maximum value   :22.33
 Median          :14.36
 Mean            :14.05
 std.dev         :3.56
 1st Qu.         :11.90
 3rd Qu.         :16.20
 IQR             :4.3
 lower - upper (Q1 - 1.5*IQR , Q3+1.5*IQR) : (5.45, 22.25)

Incongruent :

 Minimum value   :15.69
 Maximum value   :35.26
 Median          :21.02
 Mean            :22.02
 std.dev         :4.80
 1st Qu.         :18.72
 3rd Qu.         :24.05
 IQR             :5.33
 lower - upper  (Q1 - 1.5*IQR , Q3+1.5*IQR) : (13.39, 29.38)
```
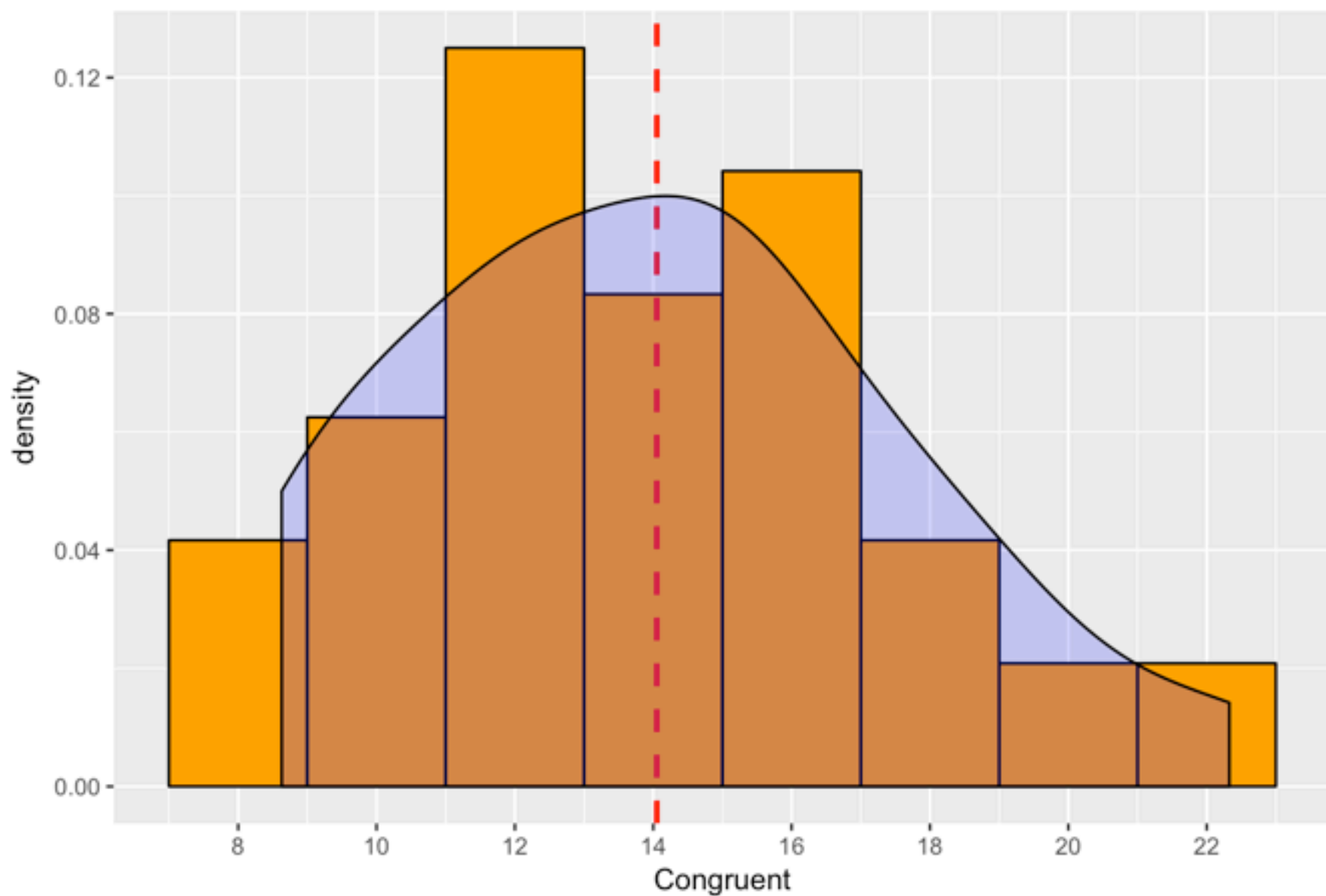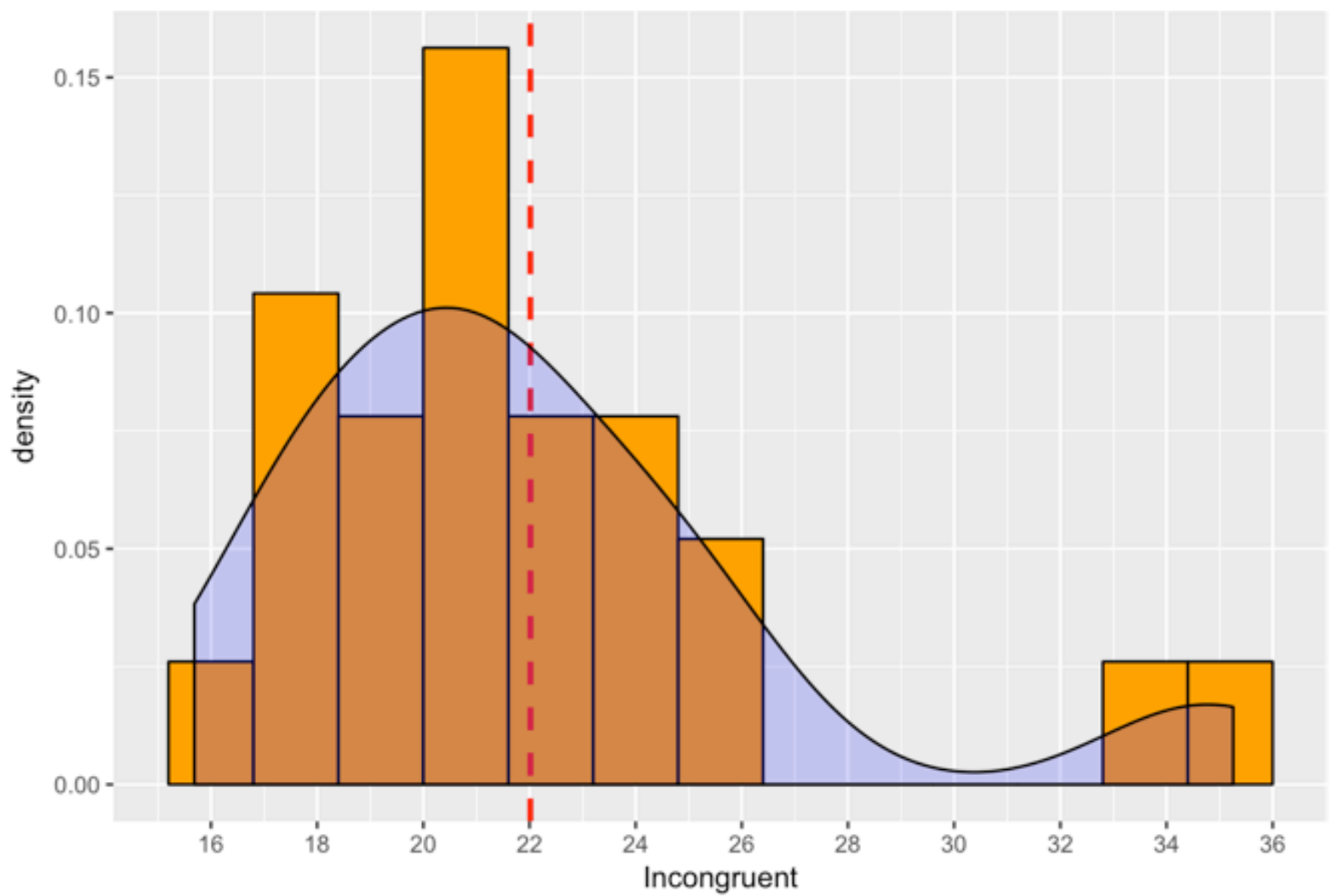
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

Congruent Histogram

Congurent distribution is normal distrubuted, means of this distribution is 14.05
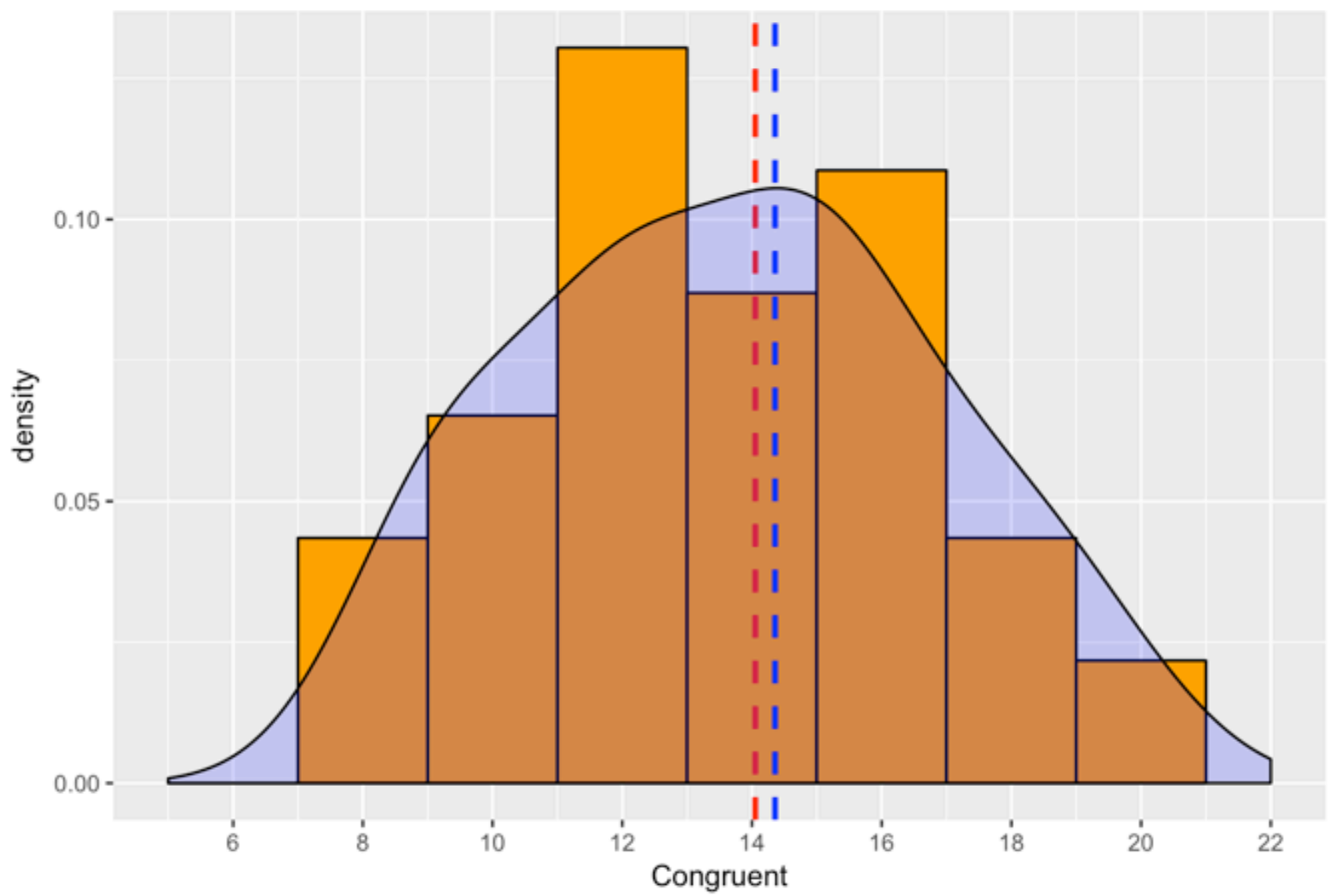
Incongurent distribution is bimodal distrubuted, means of this distribution is 22.02
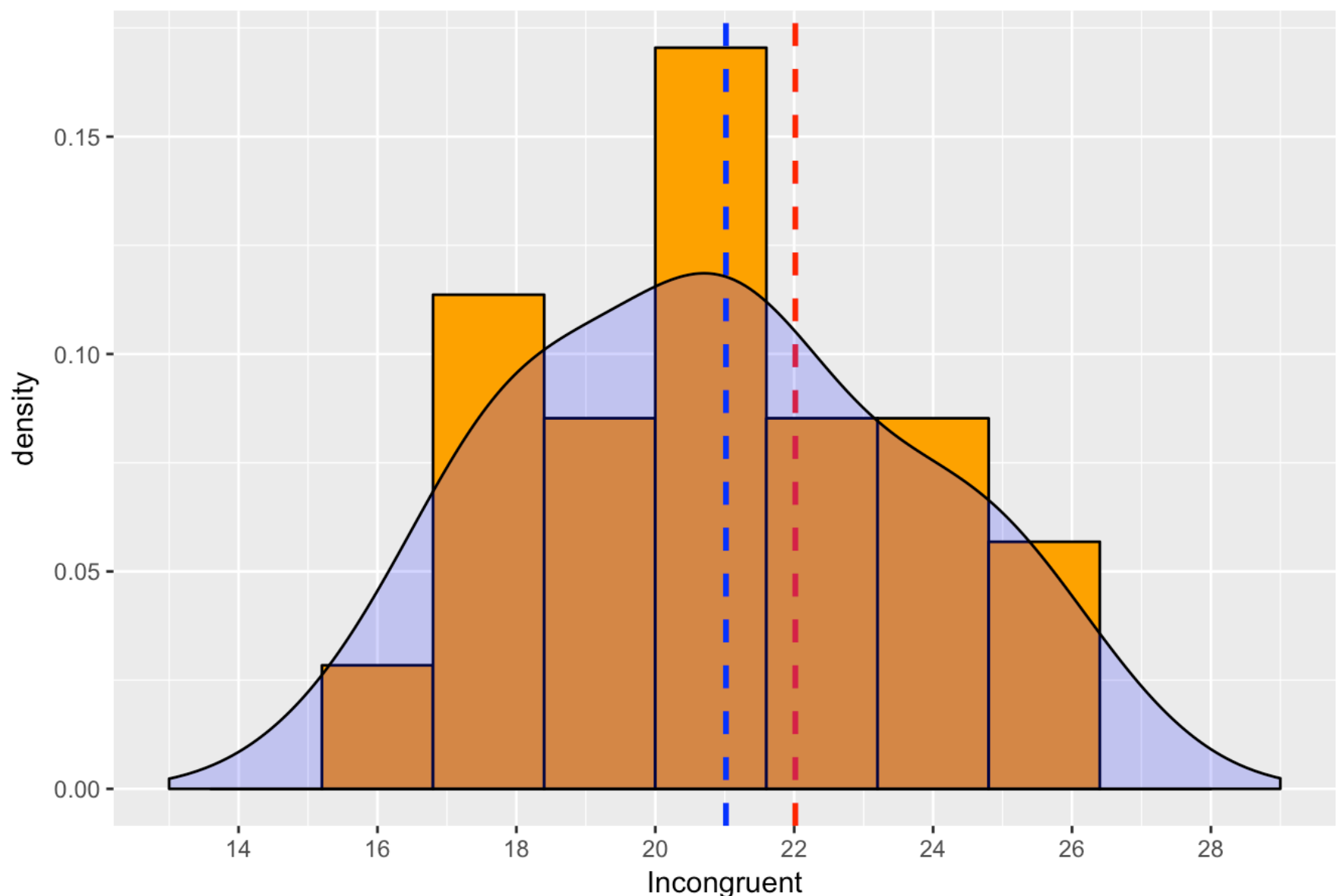
Also if we remove the outliers from these distributions ;

Congruent Histogram without OUTLIERS

it seems normal distrubuted, blue line is median and red line is mean of distribution.

Incongruent Histogram without OUTLIERS

As we can see this non-outliered distribution is normal distribution too.And blue line is median and red line is mean of this distribution.

According to this sample data , we can easily say that Incongurent test takes more time that congruent test.

5. Now, perform the statistical test and report your results. What is your confidence level and your critical statistic value? Do you reject the null hypothesis or fail to reject it? Come to a conclusion in terms of the experiment task. Did the results match up with your expectations?

a = 0.05
it is 95 percent confidence level
sample size = 24 so df = 23

two tailed 0.05 and 23 ;

$P = +-2.069$ and
t* = -2.069 and t* = 2.069
Mean of Congruent sample , Mc = 14.051125
Mean of Incongruent sample , Mi = 22.015916

and step by step ;

```
1 - Calculate the Differences (Congruent - Incongruent) of each test
2 - Calculate the squared deviations difference - ((Differences of Congruent and I
ncongruent)-Mean difference)^2
3 - Sum the squared deviations difference :
4 - Divide it by the number of samples - 1 which is equals to the Variance Differe
nce :23.66654088
5 - Square root the Variance Difference : 4.864826911
```

Congruent	Incongruent	d	d - mean diff.		mean of Congruent	mean of Incongruent	mean difference
12.079	19.278	-7.199	0.58629649		14.051125	22.01591667	-7.964791667
16.791	18.741	-1.95	36.17661609				
9.564	21.214	-11.65	13.58143609				
8.63	15.687	-7.057	0.82391929				
14.669	22.803	-8.134	0.02866249		23.66654088	<--- Sum(d:d)/23	
12.238	20.878	-8.64	0.45603009		4.864826911	<--- S (sqrt(23.66))	
14.692	24.572	-9.88	3.66837409				
8.987	17.394	-8.407	0.19562929				
9.401	20.762	-11.361	11.53485369				
14.48	26.282	-11.802	14.72487129				
22.328	24.524	-2.196	33.27789969				
15.298	18.644	-3.346	21.33238969				
15.073	17.51	-2.437	30.55546729				
16.929	20.33	-3.401	20.82735769				
18.2	35.255	-17.055	82.63355409				
12.13	22.158	-10.028	4.25720689				
18.495	25.139	-6.644	1.74424849				
10.639	20.429	-9.79	3.33172009				
11.344	17.425	-6.081	3.54832569				
12.369	34.288	-21.919	194.7224885				
12.944	23.894	-10.95	8.91201609				
14.233	17.96	-3.727	17.95810129				
19.71	22.058	-2.348	31.54731889				
16.004	21.157	-5.153	7.90565689				

according to the dependent t-test for paired sample ;

$$S_d = \sqrt{\left(\frac{\sum(d - (\mu_c - \mu_i))}{n - 1} \right)}$$

S : 4.864826911

n : 24

$$t = \left(\frac{\mu_c - \mu_i}{\frac{S}{\sqrt{n}}} \right)$$

t = (14.051125 - 22.015916) / (4.8648 / 4.8989)
t = -8.02

Result : reject to null , t value is significently low

Conclusion

First of all we found dependent variable, time interval is dependent variable and it depends on words meaning and its color. According to the data I've appropriated a hypothesis and to test this hyphothesis, Because of we have dependent variable and two sets, I've choosed dependent t-test.

Choosing critical zone is 5% (a = 0.05), I've calculate mean of samples, according to alpha and deep of freedom values, found P values. and I've calculated t value.

Calculated t value is in critical zone in distribution so null hypothesis has been rejected.

Referances

- http://www.stat.cmu.edu/~cshalizi/rmarkdown/ (http://www.stat.cmu.edu/~cshalizi/rmarkdown/)
- https://statistics.laerd.com/statistical-guides/dependent-t-test-statistical-guide.php (https://statistics.laerd.com/statistical-guides/dependent-t-test-statistical-guide.php)
- http://www.stat.cmu.edu/~cshalizi/rmarkdown/ (http://www.stat.cmu.edu/~cshalizi/rmarkdown/)
- http://onlinestatbook.com/2/summarizing_distributions/variability.html (http://onlinestatbook.com/2/summarizing_distributions/variability.html)