

# 词义的力量

## —— 斯特鲁普效应实验数据分析

### Words matter?

## —— Data Analysis of Stroop Effect Experiment

### 1. 前言 Introduction

#### 1.1 斯特鲁普效应简介 Introduction of Stroop Effect

当文字的颜色与词义不一致时，词义会对反应时间产生干扰。1935年 John Ridley Stroop 发表了文章以后，该现象就被称为斯特鲁普效应（Stroop effect）[1]。具体的实验是：说出文字的颜色，而不是文字本身。比如“红色”这两个字以蓝色印刷（"RED" is printed in a blue colour），这时参与者读出“蓝色”（BLUE）[2]。当文字的字义与印刷的颜色一致时，称为 Congruent；不一致时，称为 Incongruent。实验发现 Incongruent 的时候，参与者阅读的时间更长，因此实验的结论是词义（semantic meaning）对颜色的识别产生了干扰，称为词义干扰（semantic interference）。

#### 1.2 本项目总览 Project overview

现有一份斯特鲁普效应的实验数据（随文件所带）。在此基础上拟进行的工作包括：

- 描述性统计
  - 用描述统计的方法分析实验数据
- 推论统计
  - 建立假设
  - 用样本数据进行假设检验
  - 得出检验结果
- 结论部分

### 2. 描述性统计 Descriptive statistics

两种实验数据，一种是 Congruent（词义与颜色一致），另一种是 Incongruent（词义与颜色不一致）。

为了进行后面的工作，先引入必要的 Python 包。

In [2]:

```
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
```

用 pandas 读入数据文件：

In [4]:

```
stroop = pd.read_csv('stroopdata.csv')
```

查看数据头部，确认数据读入正常：

In [5]:

```
stroop.head()
```

Out[5]:

	Congruent	Incongruent
0	12.079	19.278
1	16.791	18.741
2	9.564	21.214
3	8.630	15.687
4	14.669	22.803

用 numpy 分别对 Congruent 与 Incongruent 数据进行处理：

In [6]:

```
congruent = np.array(stroop['Congruent'])  
incongruent = np.array(stroop['Incongruent'])
```

## 2.1 Congruent 部分数据描述性统计

统计性描述结果：

In [7]:

```
print("-"*5,"Congruent 部分","-"*5)
print("平均阅读时间 (mean) : ", congruent.mean())
print("方差 (variance) : ", congruent.var(ddof=1)) # ddof=1, 表示除以 n-1
print("标准差 (standard deviation) : ", congruent.std(ddof=1))
print("最小值 (min) : ", congruent.min())
print("最大值 (max) : ", congruent.max())
print("第一四分位数 (25th percentile) : ", np.percentile(congruent, 25))
print("中位数 (median) : ", np.median(congruent))
print("第三四分位数 (75th percentile) : ", np.percentile(congruent, 75))
print("样本数量 (sample size) : ", congruent.size)
```

```
----- Congruent 部分 -----
平均阅读时间 (mean) :  14.051125
方差 (variance) :  12.6690290707
标准差 (standard deviation) :  3.55935795765
最小值 (min) :  8.63
最大值 (max) :  22.328
第一四分位数 (25th percentile) :  11.89525
中位数 (median) :  14.3565
第三四分位数 (75th percentile) :  16.20075
样本数量 (sample size) :  24
```

如果把数值都 **round** 到个位 (**round**表示取值最靠近的整数) :

In [8]:

```
print("-"*5,"Congruent 部分","-"*5)
print("平均阅读时间", np.round(congruent.mean()))
print("方差", np.round(congruent.var(ddof=1)))
print("标准差", np.round(congruent.std(ddof=1)))
print("最小值", np.round(congruent.min()))
print("最大值", np.round(congruent.max()))
print("第一四分位数", np.round(np.percentile(congruent, 25)))
print("中位数", np.round(np.median(congruent)))
print("第三四分位数", np.round(np.percentile(congruent, 75)))
print("样本数量", congruent.size)
```

```
----- Congruent 部分 -----
平均阅读时间 14.0
方差 13.0
标准差 4.0
最小值 9.0
最大值 22.0
第一四分位数 12.0
中位数 14.0
第三四分位数 16.0
样本数量 24
```

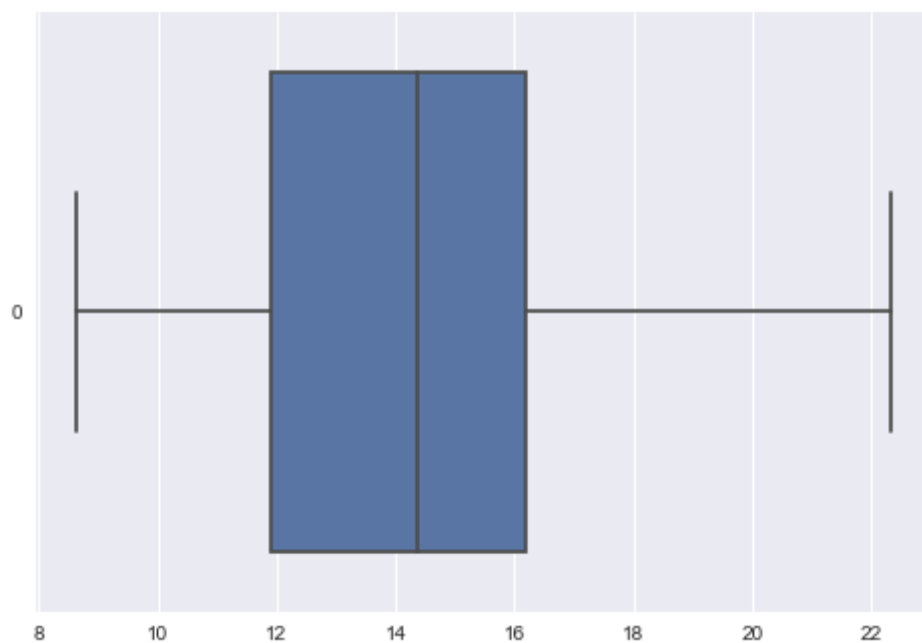
数据的分布用盒形图 (**box plot**) 可视化, 没有出现异常值 (**outlier**) 。数据呈现正偏斜。

In [9]:

```
sns.boxplot(data=stroop['Congruent'], orient='h')
```

Out[9]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10c53de10>



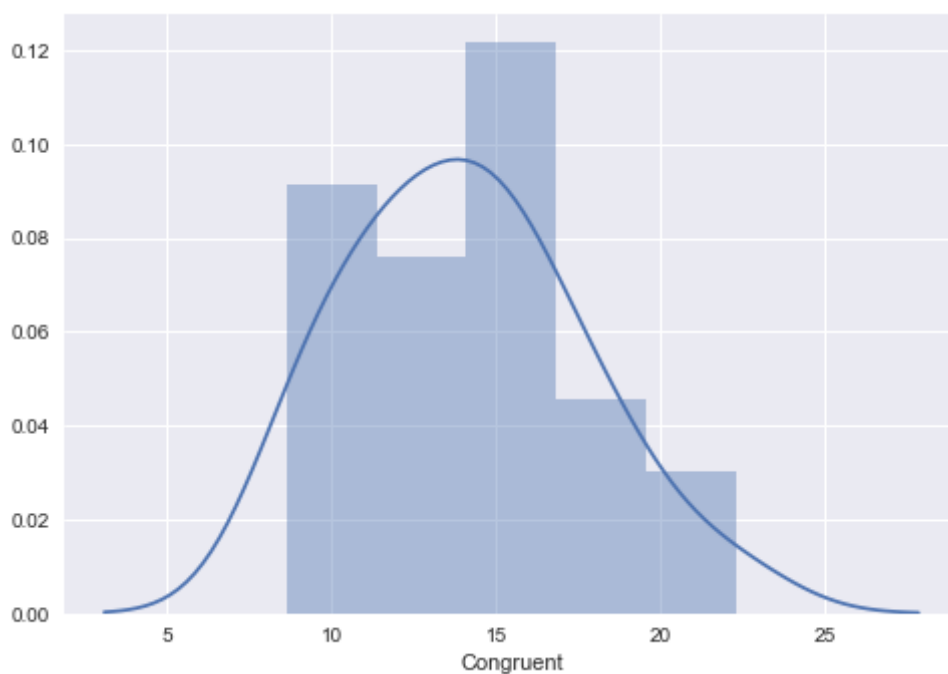
直方图发现 **Congruent** 的样本数据接近正态分布，有一定程度的正偏斜 (**positive skewed**)

In [10]:

```
sns.distplot(stroop['Congruent'])
```

Out[10]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10c4fed30>



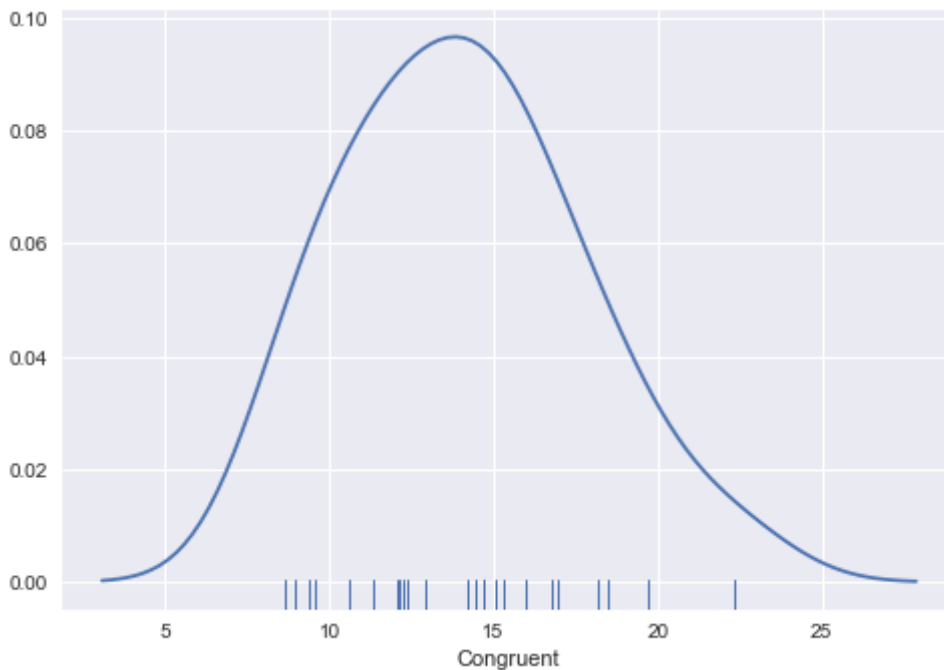
用 rug 图也可以看见数据在 10-15 区间的密度较大。

In [11]:

```
sns.distplot(stroop['Congruent'], hist=False, rug=True)
```

Out[11]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10c664198>



## 2.2 Incongruent 部分数据描述性统计

统计性描述结果:

In [12]:

```
print("-"*5,"Incongruent 部分","-"*5)
print("平均阅读时间: ", incongruent.mean())
print("方差: ", incongruent.var(ddof=1))
print("标准差: ", incongruent.std(ddof=1))
print("最小值: ", incongruent.min())
print("最大值: ", incongruent.max())
print("第一四分位数", np.percentile(incongruent, 25))
print("中位数: ", np.median(incongruent))
print("第三四分位数", np.percentile(incongruent, 75))
print("样本数据: ", incongruent.size)
```

----- Incongruent 部分 -----

平均阅读时间: 22.0159166667

方差: 23.0117570362

标准差: 4.79705712247

最小值: 15.687

最大值: 35.255

第一四分位数 18.71675

中位数: 21.0175

第三四分位数 24.0515

样本数据: 24

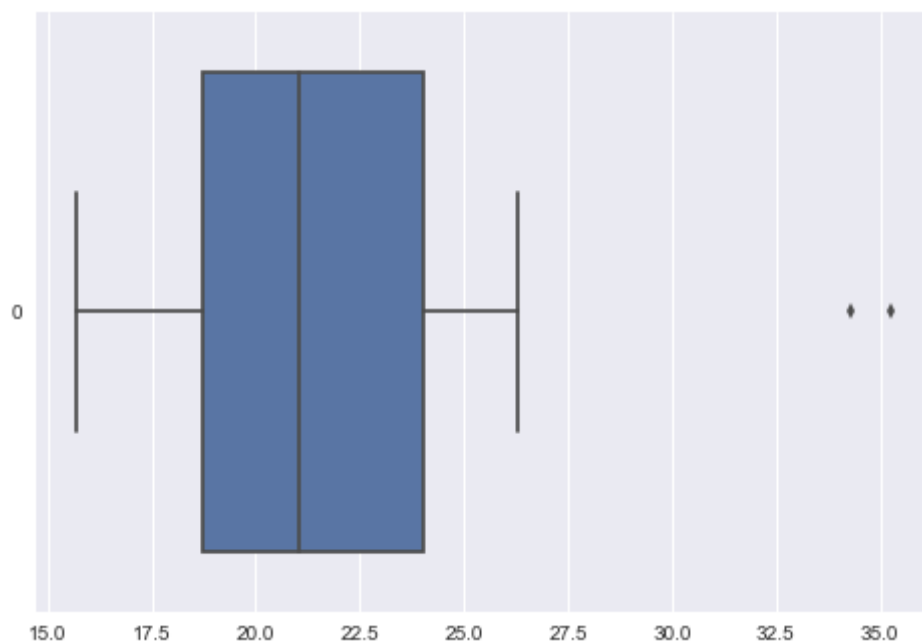
盒形图中发现 Incongruent 数据中最大的两个值是异常值

In [13]:

```
sns.boxplot(data=stroop['Incongruent'], orient='h')
```

Out[13]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10c63c160>



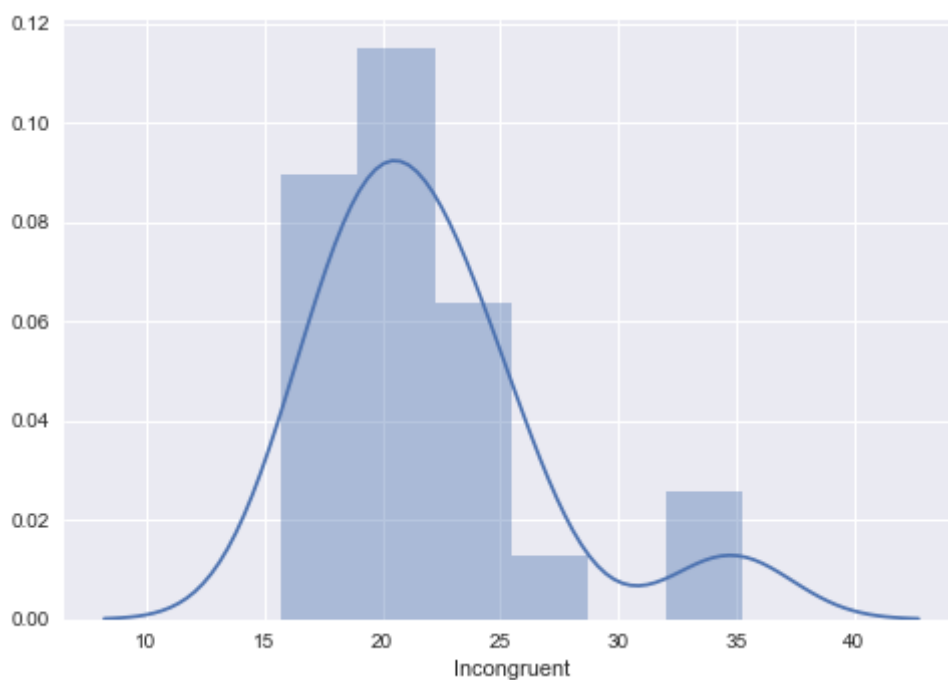
用直方图发现出现了两个峰。

In [14]:

```
sns.distplot(stroop['Incongruent'])
```

Out[14]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10c912da0>



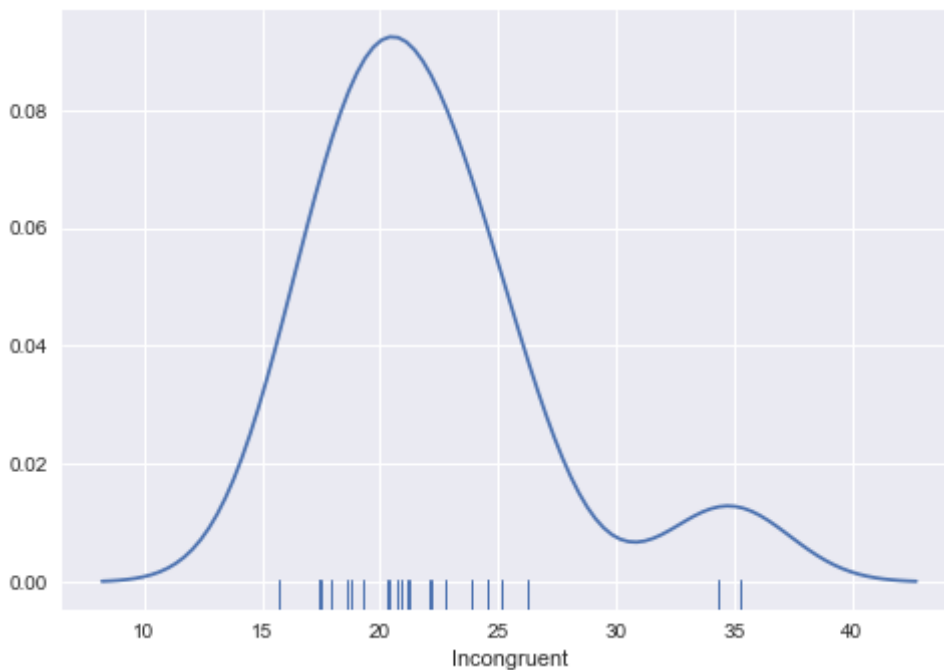
较小的峰是由两个异常值产生的。

In [15]:

```
sns.distplot(stroop['Incongruent'], hist=False, rug=True)
```

Out[15]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10cb00630>



设置函数删除异常值，输入参数分别是：

- 需要删除异常值的 numpy array
- 异常值常数 (outlierConstant, 一般是1.5)
- IQR 表示 Interquartile ranges

In [16]:

```
def removeOutliers(nparray, outlierConstant):
    """
    to remove the outliers for a numpy array,
    return a new numpy array without outliers
    """
    upper_quartile = np.percentile(nparray, 75)
    lower_quartile = np.percentile(nparray, 25)
    IQR = (upper_quartile - lower_quartile) * outlierConstant
    quartileSet = (lower_quartile - IQR, upper_quartile + IQR)
    # print(IQR)
    # print(quartileSet)
    resultList = []
    for y in nparray.tolist():
        if y > quartileSet[0] and y < quartileSet[1]:
            resultList.append(y)
    resultList_nparray = np.array(resultList)
    return resultList_nparray
```

Incongruent 数据，在删除异常值之前先备份，并且确定样本数量是 24。

In [17]:

```
incongruent_old = incongruent
print(incongruent_old)
print("Sample size: ", incongruent_old.size)
```

```
[ 19.278  18.741  21.214  15.687  22.803  20.878  24.572  17.394  2
 0.762
 26.282  24.524  18.644  17.51  20.33  35.255  22.158  25.139  2
 0.429
 17.425  34.288  23.894  17.96  22.058  21.157]
Sample size: 24
```

用 removeOutliers 函数处理 incongruent 数据之后，可以发现数组中的两个异常值已经被删除。

In [18]:

```
incongruent = removeOutliers(incongruent, 1.5)
print(incongruent)
print("Sample size: ", incongruent.size)
```

```
[ 19.278  18.741  21.214  15.687  22.803  20.878  24.572  17.394  2
 0.762
 26.282  24.524  18.644  17.51  20.33  22.158  25.139  20.429  1
 7.425
 23.894  17.96  22.058  21.157]
Sample size: 22
```

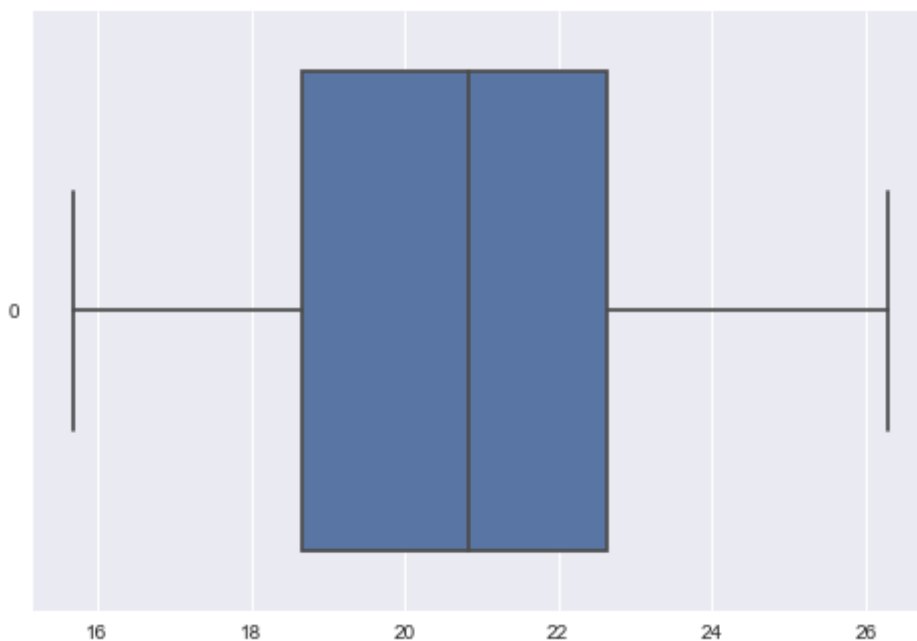
再次用盒形图确认没有异常值了。

In [19]:

```
incongruent_dataframe = pd.DataFrame(data=incongruent)
sns.boxplot(data=incongruent_dataframe, orient='h')
```

Out[19]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x10cc91c88>
```



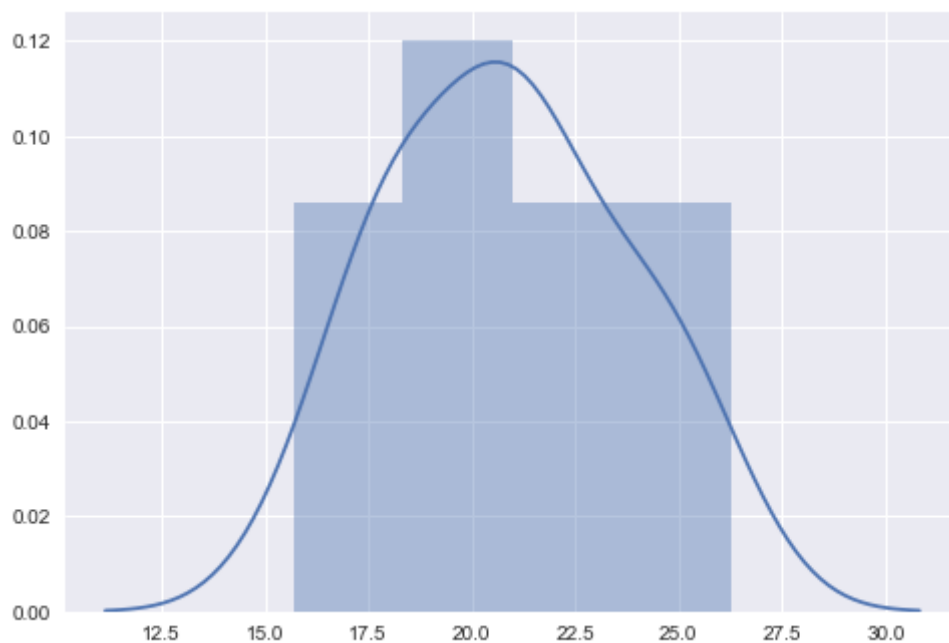


In [20]:

```
sns.distplot(incongruent_dataframe)
```

Out[20]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10cd7a6d8>



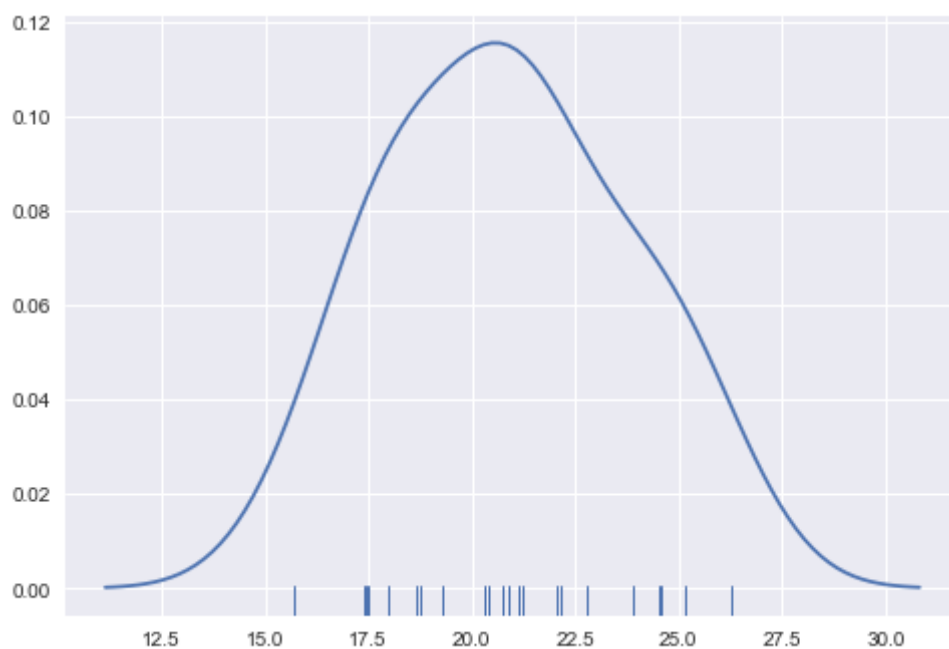
取样分布也只剩一个峰，接近于正态分布。

In [21]:

```
sns.distplot(incongruent_dataframe, hist=False, rug=True)
```

Out[21]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x10d015fd0>



再次计算 Incongruent 部分描述性统计数据

In [22]:

```
print("-"*5,"Incongruent 部分","-"*5)
print("平均阅读时间: ", incongruent.mean())
print("方差: ", incongruent.var(ddof=1))
print("标准差: ", incongruent.std(ddof=1))
print("最小值: ", incongruent.min())
print("最大值: ", incongruent.max())
print("第一四分位数", np.percentile(incongruent, 25))
print("中位数: ", np.median(incongruent))
print("第三四分位数", np.percentile(incongruent, 75))
print("样本数据: ", incongruent.size)
```

```
----- Incongruent 部分 -----
平均阅读时间:  20.8563181818
方差:  8.27668308442
标准差:  2.87692250233
最小值:  15.687
最大值:  26.282
第一四分位数 18.66825
中位数:  20.82
第三四分位数 22.64175
样本数据:  22
```

## 2.3 描述性统计部分总结 Conclusion of descriptive statistics part

对斯特鲁普效应实验的数据进行了描述性统计分析。用盒形图发现 Incongruent 在最大值方向有两个异常值。用 removeOutliers 函数删除了异常值，并形成新的 numpy array。

最终得到两个数组如下所示。后面的推论统计部分将在这两个数据的基础上进行。

In [23]:

congruent

Out[23]:

```
array([ 12.079,  16.791,   9.564,   8.63 ,  14.669,  12.238,  14.69
 2,
        8.987,   9.401,  14.48 ,  22.328,  15.298,  15.073,  16.92
 9,
        18.2  ,  12.13 ,  18.495,  10.639,  11.344,  12.369,  12.94
 4,
        14.233,  19.71 ,  16.004])
```

In [24]:

incongruent

Out[24]:

```
array([ 19.278,  18.741,  21.214,  15.687,  22.803,  20.878,  24.57
 2,
        17.394,  20.762,  26.282,  24.524,  18.644,  17.51 ,  20.33
 ,
        22.158,  25.139,  20.429,  17.425,  23.894,  17.96 ,  22.05
 8,
        21.157])
```

### 3. 推论统计 Inference statistics

#### 3.1 结论预览 Preview of conclusion

双样本的零假设为两种实验结果一样。T检验结果表明：实验结果的 t 值为-7.16，实验结果一致的概率  $P = 7.30e-9$ ，远小于 0.1%。即使按照 99.9% 的要求，零假设也被拒绝。结论意味着两种实验的结果并不一致，Congruent 实验的平均阅读时间确实小于 Incongruent 的平均阅读时间。

In [27]:

```
stats.ttest_ind(a= congruent,  
                b= incongruent,  
                equal_var=False)
```

Out[27]:

```
Ttest_indResult(statistic=-7.1570631082927409, pvalue=7.298508143705  
8256e-09)
```

即使 Incongruent 数据的异常值没有排除，也是一样的结论。

In [28]:

```
stats.ttest_ind(a= congruent,  
                b= incongruent_old,  
                equal_var=False)
```

Out[28]:

```
Ttest_indResult(statistic=-6.5322505539032285, pvalue=6.510167390442  
6379e-08)
```

下面按照推论统计的步骤详细推导。

## 3.2 推论统计具体步骤（不含异常值） Detailed inference statistics (without outliers)

先考虑排除异常值的情况。

详细的推论统计步骤包括：

- (A) 假设检验
  - (A-1) 识别因变量与处理条件
    - 因变量 (dependent variable)
    - 处理条件 (treatment)
  - (A-2) 建立假设
    - 零假设 (null hypothesis)
    - 备择假设 (alternative hypothesis)
    - 假设类型 (type of hypothesis)
  - (A-3) 找到 t-critical
    - 设定alpha值 (alpha leve)
    - 计算自由度 (degree of freedom)
    - 获得 t-critical
  - (A-4) 计算 t-statistic
    - 计算两种样本的平均值 (mean)
    - 找到两种样本大小 (sample size)
    - 计算两种样本的方差之和 (sum of variance)
    - 获得合并方差 (pooled variance)
    - 计算标准误差 (standard error)
    - 得到样本的 t 值 (t-statistic)
  - (A-5) 对比 t 值，得出检验结论
- (B) 置信区间 (confidence interval)
  - (B-1) 误差范围 (margin of error)
  - (B-2) 置信区间 (CI)
- (C) 测量效应强度 (effect size measures)
  - (C-1) 相关性分析 (correlation)

### (A) 假设检验

#### (A-1) 识别因变量与处理条件

##### **因变量 (dependent variable)**

参与者识别出颜色的总时间。

##### **处理条件 (treatment)**

颜色与词义是否一致。颜色与词义一致时，称为 Congruent；不一致时则是 Incongruent。因此也可以把处理的类型作为自变量(independent variable)。

#### (A-2) 建立假设

**零假设 (null hypothesis)**

词义与颜色是否一致，都不会影响阅读者识别颜色的时间。其中 1 表示 Congruent，2 表示 Incongruent。

$$H_0 : \mu_1 - \mu_2 = 0$$

**备择假设 (alternative hypothesis)**

词义与颜色不一致，就会影响阅读者识别颜色的时间。

$$H_A : \mu_1 - \mu_2 \neq 0$$

**假设类型 (hypothesis type)**

双尾检测 (two-tailed test)。当 p 落入概率密度曲线两侧的 critical region 时，判定检测统计上显著。

**(A-3) 找到 t-critical****设定alpha值 (alpha leve)**

$$\alpha = 0.05$$

**计算自由度 (degree of freedom)**

$$df = n_1 + n_2 - 2 = 24 + 22 - 2 = 44$$

**获得 t-critical**

根据以上两个条件，使用 [GraphPad \(http://www.graphpad.com/quickcalcs/\)](http://www.graphpad.com/quickcalcs/) 的计算：

$$t_{critical} = \pm 2.0154$$

**(A-4) 计算 t-statistic****计算两种样本的平均值 (mean)**

根据描述性统计的结果，把两个样本的算术平均值四舍五入到小数点后两位得到：

In [44]:

```
x1 = congruent.mean() # mean of sample Congruent
x2 = incongruent.mean() # mean of sample Incongruent
print ("Mean of sample Congruent:", x1)
print ("Mean of sample Incongruent:", x2)
```

```
Mean of sample Congruent: 14.051125
Mean of sample Incongruent: 20.8563181818
```

$$\bar{X}_1 = 14.05$$

$$\bar{X}_2 = 20.86$$

样本的均值差 (mean difference) 为:

In [47]:

```
x_dif = x1 - x2
print ("Difference of two sample means:", x_dif)
```

Difference of two sample means: -6.80519318182

$$\bar{X}_1 - \bar{X}_2 = 14.05 - 20.86 = -6.81$$

### 找到两种样本大小 (sample size)

In [37]:

```
n1 = congruent.size
n2 = incongruent.size
print("Congruent sample size, n1:", n1)
print("Incongruent sample size, n2:", n2)
```

Congruent sample size, n1: 24

Incongruent sample size, n2: 22

$$n_1 = 24$$

$$n_2 = 22$$

### 计算两种样本的方差之和 (sum of variance)

In [38]:

```
ss1 = congruent.var()*congruent.size
print("Sum of sample Congruent variances, ss1:",ss1)
ss2 = incongruent.var()*incongruent.size
print("Sum of sample Incongruent variances, ss2:",ss2)
```

Sum of sample Congruent variances, ss1: 291.387668625

Sum of sample Incongruent variances, ss2: 173.810344773

$$SS_1 = 291.39$$

$$SS_2 = 173.81$$

### 获得合并方差 (pooled variance)

获得两个样本的自由度 (df)，然后计算合并方差。

In [40]:

```
df1 = n1 - 1
df2 = n2 - 1
sp2 = (ss1 + ss2)/(df1 + df2)
print("Pooled variance, sp2:", sp2)
```

Pooled variance, sp2: 10.5726821227

$$S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = 10.57$$

### 计算标准误差 (standard error)

In [43]:

```
SE = np.sqrt( sp2/n1 + sp2/n2 )
print("Standard error:", SE)
```

Standard error: 0.959742091345

$$S_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} = 0.96$$

### 得到样本的 $t$ 值 ( $t$ -statistic)

In [45]:

```
t_statistic = (x1 - x2) / SE
print("t-statistic:", t_statistic)
```

t-statistic: -7.09064783465

$$t_{statistic} = -7.09$$

该值与 `ttest_ind` 方法得出 -7.16 较为接近，两者可以相互印证计算数值的正确性。

### (A-5) 对比 $t$ 值，得出检验结论

对比  $t_{statistic} = -7.09$  与  $t_{critical} = \pm 2.0154$ ，发现  $t_{statistic}$  值已经落入了临界区域 (critical region)。

同样通过  $t_{statistic}$  与自由度  $df = 44$  也可以得到双尾检测的P值小于 0.001。因此我们可以认为两个样本的t-检验 (two-sample t-test) 在统计上显著，结论是拒绝零假设。

*Result : Reject  $H_0$*

### (B) 置信区间 (confidence interval)

#### (B-1) 误差范围 (margin of error)

In [46]:

```
t_critical = 2.0154
margin_of_error = t_critical * SE
print("Margin of error", margin_of_error)
```

Margin of error 1.9342642109

$$\text{margin of error} = (t_{\text{critical}} \times S_{\bar{X}-\bar{Y}}) = 2.0154 \times 0.96 = 1.93$$

## (B-2) 置信区间 (CI)

In [49]:

```
upper_bound = x_dif + margin_of_error
lower_bound = x_dif - margin_of_error
print ("Upper bound is", upper_bound)
print ("Lower bound is", lower_bound)
```

Upper bound is -4.87092897092

Lower bound is -8.73945739272

置信区间 (CI) 是 (-8.74, -4.87)

## (C) 测量效应强度 (effect size measures)

### (C-1) 相关性分析 (correlation)

In [53]:

```
correlation = (t_statistic ** 2) / ((t_statistic ** 2) + (df1 + df2))
print("The correlation is", correlation)
```

The correlation is 0.533291617386

$$r^2 = \frac{t_{\text{statistic}}^2}{t_{\text{statistic}}^2 + df_{\text{total}}} = 0.53 = 53\%$$

意味着两种实验结果不一致，其中53%的原因可以用实验条件的差异来解释。



### 推论统计总结（不含异常值）

在推论统计部分，我们认为斯特鲁普效应实验中，词义与颜色的一致或不一致可能会产生不同的结果。为了进行检验，进行了以下工作：

- 自变量是实验的种类（词义与颜色一致为 Congruent，词义与颜色不一致为 Incongruent），因变量是阅读者按照要求读完材料的时间。
- 建立零检验：两种实验的阅读时间相同；备择检验是时间不同。因此检验类型是双尾t检验。
- 设定临界区域，临界值为0.05，得到临界t值为0.2154。
- 计算双样本的统计t值是-7.09。
- 样本统计t值已经落入临界区域，因此在统计上显著。
- 结果是拒绝零检验。

通过双样本t检验，在统计上不能认为斯特鲁普效应实验中的两种实验类型的结果一致。并且这种实验类型的不一致，可以解释53%的阅读时间不同。

### 3.3 推论统计具体步骤（包含异常值） Detailed inference statistics (with outliers)

In [52]:

```
x1 = congruent.mean() # mean of sample Congruent
x3 = incongruent_old.mean() # mean of sample Incongruent
print ("Mean of sample Congruent:", x1)
print ("Mean of sample Incongruent (with outliers):", x3)

x_dif3 = x1 - x3 # mean difference
print ("Difference of two sample means:", x_dif3)

n1 = congruent.size # sample size of Congruent
n3 = incongruent_old.size # sample size of Incongruent with outliers
print("Congruent sample size, n1:", n1)
print("Incongruent (with outliers) sample size, n2:", n3)

ss1 = congruent.var()*congruent.size # sum of Congruent
print("Sum of sample Congruent variances, ss1:",ss1)
ss3 = incongruent_old.var()*incongruent.size # sum of Congruent with outliers
print("Sum of sample Incongruent variances, ss2:",ss3)

df1 = n1 - 1 # degree of freedom of Congruent
df3 = n3 - 1 # degree of freedom of Incongruent
sp2_outlier = (ss1 + ss3)/(df1 + df3) # pooled variance
print("Pooled variance, sp2:", sp2_outlier)

SE_outlier = np.sqrt( sp2_outlier/n1 + sp2_outlier/n2 )
print("Standard error:", SE_outlier)

t_statistic_outlier = (x1 - x3) / SE_outlier
print("t-statistic:", t_statistic_outlier)

t_critical_outlier = 2.0154
margin_of_error_outlier = t_critical_outlier * SE_outlier
print("Margin of error", margin_of_error_outlier)

upper_bound_outlier = x_dif3 + margin_of_error_outlier
lower_bound_outlier = x_dif3 - margin_of_error_outlier
print ("Upper bound is", upper_bound_outlier)
print ("Lower bound is", lower_bound_outlier)

correlation_outlier = (t_critical_outlier ** 2)/((t_critical_outlier ** 2) + (df
1 + df3))
print("The correlation is", correlation_outlier)
```

```
Mean of sample Congruent: 14.051125
Mean of sample Incongruent (with outliers): 22.0159166667
Difference of two sample means: -7.96479166667
Congruent sample size, n1: 24
Incongruent (with outliers) sample size, n2: 24
Sum of sample Congruent variances, ss1: 291.387668625
Sum of sample Incongruent variances, ss2: 485.164544181
Pooled variance, sp2: 16.8815698436
Standard error: 1.21274186342
t-statistic: -6.56759027369
Margin of error 2.44415995153
Upper bound is -5.52063171514
Lower bound is -10.4089516182
The correlation is 0.08113639831111624
```

$t_{statistic} = -6.57$  与 `ttestind` 方法计算的  $-6.532$  基本一致。 $t_{statistic}=-6.57$  也是在临界区域中，因此即使包含了异常值，样本数据同样可以拒绝零假设。

结论是：异常值不会影响t检验结论。包含或者不包含异常值的t检验结果，都表明不能认为两种类型的斯特鲁普效应实验结果一致。

## 4. 结论 Conclusion

斯特鲁普效应结论是词义会产生干扰。为了验证斯特鲁普效应，分别进行了两种实验：词义与文字颜色一致，以及词义与文字颜色一致。分别对两种实验的样本进行了描述性统计与推论性统计。在推论统计部分，以两种实验中阅读时间相同作为零假设，设定临界值alpha为0.05的水准，在经过双样本t检验以后，统计上显著地拒绝了零假设，即不能认为两种类型的实验阅读时间会一致。

In [ ]: