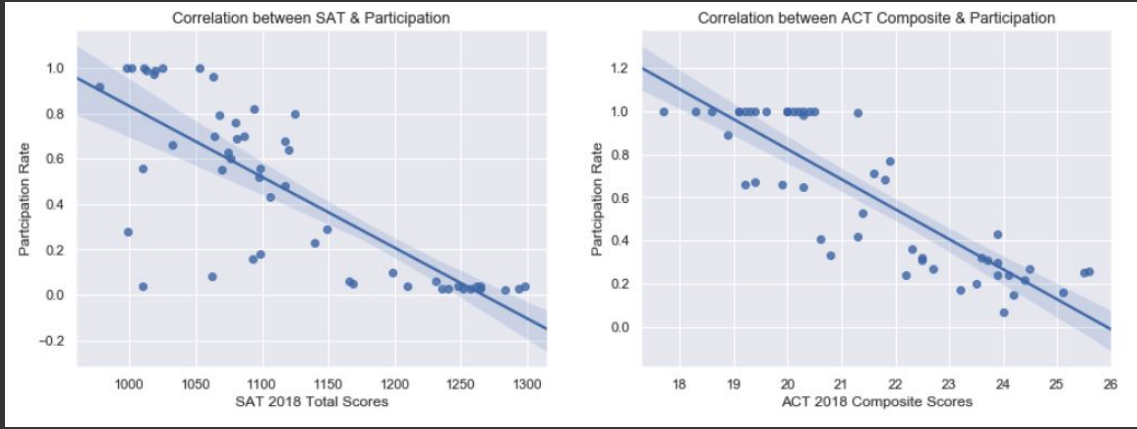
ACT & SAT Analysis - United States

DSI Project 01

Jerome Chua

Table of contents

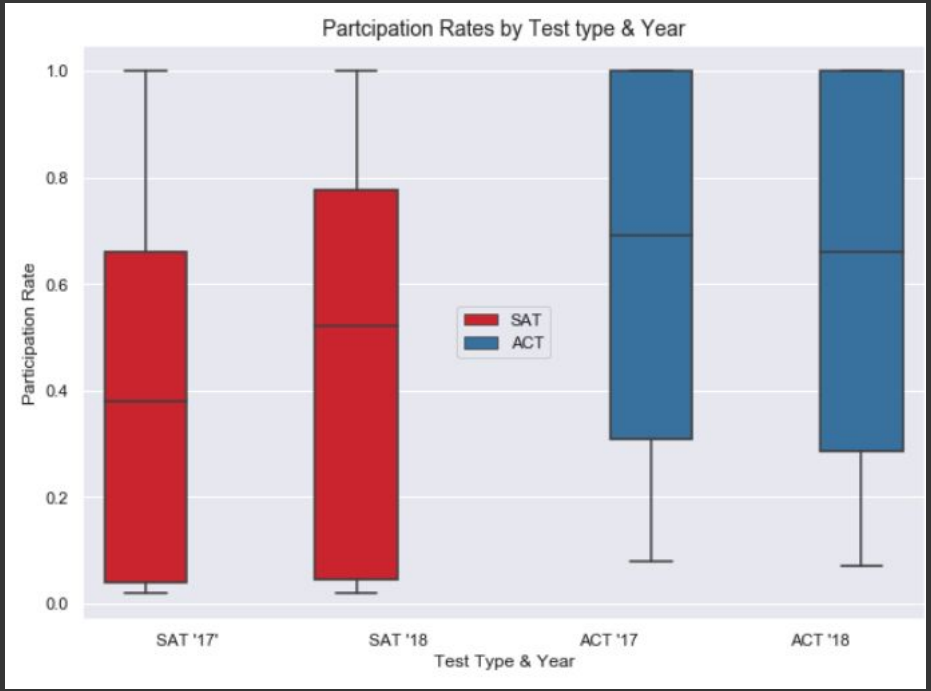
What's the relationship between participation rates & total/composite scores?



There is a negative relationship between participation rates & total/composite scores likely due to more "students from disadvantaged backgrounds taking the test".

Reference: *https://time.com/4021795/sat-test-scores-dip/*

What is the trend in SAT & ACT participation rates?



The boxplots show an increase in participation rates in most states for the SAT test, while the ACT test remains fairly popular ever since 2013.

It's not that the SAT is losing customers. On the contrary, the number of test takers has grown. It's that the ACT is growing much faster. One of the main reasons that though there is an increase in SAT test takers, the "real shift in the behavior of top high school students, with many more choosing to work toward impressive scores on **both tests.**"

Reference: *https://www.nytimes.com/2013/08/04/education/edlife/more-students-are-taking-both-the-act-and-sat.html*

Is the data we are sampling from a normally distributed population?

Step 1 - Construct Null & Alternative Hypothesis
Null Hypothesis: The data follows a normal distribution
Alternative Hypothesis: The data does not follow a normal distribution

If p-value <= α, we reject the null
If p-value > α, we fail to reject the null

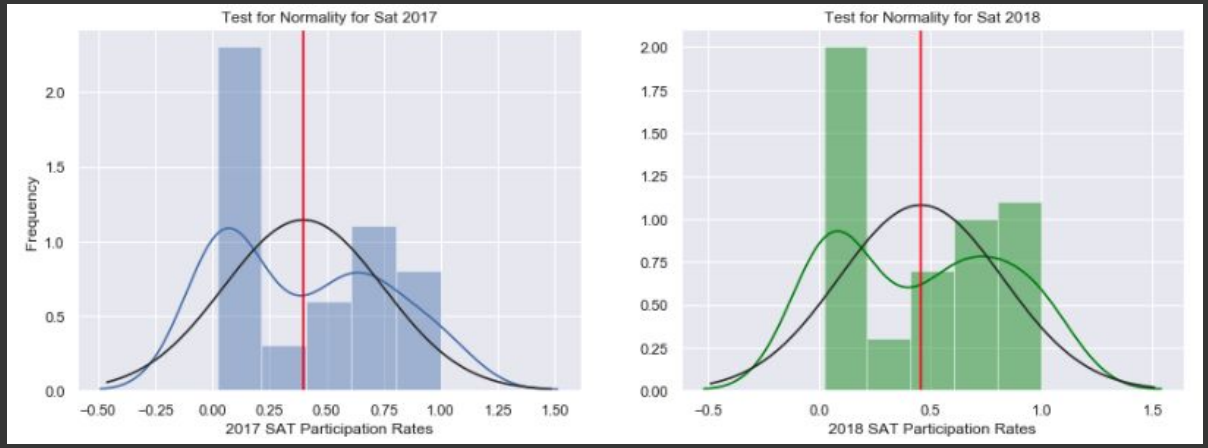
Step 2 - Specify level of significance
Significance level, α = 0.05

Step 3 - Calculate test statistic & Step 4 - Calculate p-value

```
# test for normality
d, p_val = stats.kstest(final['sat_17_participation'], 'norm')
print('Test statistic:', d)
print('p-value: ', p_val)
```

```
Test statistic: 0.5079783137169019
p-value:  1.0082794525495568e-12
```

Is the data we are sampling from a normally distributed population?



After using the SAT tests' participation rates to visualise normality, we can clearly see that the data does not follow a normal distribution.

Data analysis process code:

Data cleaning

```
# ensure respective columns are floats
sat_17_clean['Participation'].dtype, act_17_clean.Composite.dtype, act_17_clean.Participation.dtype
```

```
(dtype('float64'), dtype('float64'), dtype('float64'))
```

```
# preview data to check
act_17_clean.tail()
```

	State	Participation	English	Math	Reading	Science	Composite
47	Virginia	0.29	23.5	23.3	24.6	23.5	23.8
48	Washington	0.29	20.9	21.9	22.1	22.0	21.9
49	West Virginia	0.69	20.0	19.4	21.2	20.5	20.4
50	Wisconsin	1.00	19.7	20.4	20.6	20.9	20.5
51	Wyoming	1.00	19.4	19.8	20.8	20.6	20.2

```
# remove '%' from str and change dtype to float
sat_18_clean['Participation'] = sat_18_clean.Participation.str.replace('%','').astype(float)
# convert Participation column to proportion
sat_18_clean['Participation']  = sat_18_clean.Participation / 100
```

```
# check
sat_18_clean.head()
```

	State	Participation	Evidence-Based Reading and Writing	Math	Total
0	Alabama	0.06	595	571	1166
1	Alaska	0.43	562	544	1106
2	Arizona	0.29	577	572	1149
3	Arkansas	0.05	592	576	1169
4	California	0.60	540	536	1076

Merging

10. Merge Dataframes ¶

Join the 2017 ACT and SAT dataframes using the state in each dataframe as the key. Assign this to a new variable.

```
# merge sat_17_clean and act_17_clean DataFrames on their State columns
combined_17 = pd.merge(sat_17_clean, act_17_clean, left_on='sat_17_state', right_on='act_17_state', how='outer')
```

```
# rename 'sat_17_state' to 'state'
combined_17.rename(columns={'sat_17_state' : 'state'}, inplace=True)
# drop duplicated state name from ACT dataset
combined_17.drop(columns=['act_17_state'],inplace=True)
combined_17.head()
```

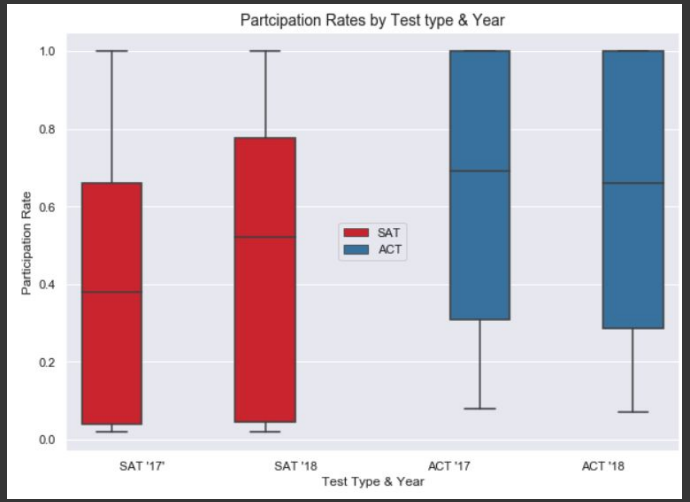
	state	sat_17_participation	sat_17_erw	sat_17_math	sat_17_total	act_17_participation	act_17_english	act_17_math	act_17_reading	act_17_science	act_
0	Alabama	0.05	593	572	1165	1.00	18.9	18.4	19.7	19.4	
1	Alaska	0.38	547	533	1080	0.65	18.7	19.8	20.4	19.9	
2	Arizona	0.30	563	553	1116	0.62	18.6	19.8	20.1	19.8	
3	Arkansas	0.03	614	594	1208	1.00	18.9	19.0	19.7	19.5	
4	California	0.53	531	524	1055	0.31	22.5	22.7	23.1	22.2	

Plotting & testing code:

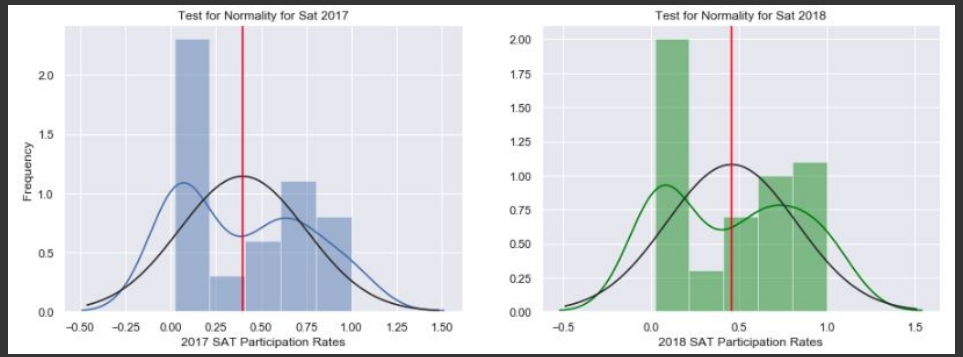
```
# make a copy of original dataframe
df = final.copy()
df.reset_index(inplace=True)
```

```
# filter copied dataframe
participation_df = df[['state','sat_17_participation', 'sat_18_participation', 'act_17_participation', 'act_18_partici

# rename columns
participation_df.columns = ['state', "SAT '17'", "SAT '18", "ACT '17", "ACT '18"]

# unpivot participation_df for plotting in later stage
participation_df = pd.melt(participation_df, id_vars='state')

# create column to be able to differentiate by test type
participation_df['test_type'] = 'SAT'
participation_df.iloc[102:, -1] = 'ACT'
```

```
plt.figure(figsize=(10,7))

sns.boxplot(data=participation_df, x='variable', y='value', hue="test_type", palette="Set1")
plt.title('Partcipation Rates by Test type & Year', fontsize=14)
plt.ylabel('Participation Rate')
plt.xlabel('Test Type & Year')
plt.legend(loc='center');
```




```
# plot the Figure object
fig = plt.figure(figsize=(15,5))

# add the Axes 1 object to the Figure
ax1 = fig.add_subplot(1,2,1)
sns.distplot(final['sat_17_participation'],bins=5, fit=norm)
ax1.axvline(final['sat_17_participation'].mean(), c='red')

# plot titles and labels
plt.xlabel('2017 SAT Participation Rates')
plt.ylabel('Frequency')
plt.title('Test for Normality for Sat 2017');

# add the Axes 2 object to the Figure
ax2 = fig.add_subplot(1,2,2)
sns.distplot(final['sat_18_participation'],bins=5, fit=norm, color='green')
ax2.axvline(final['sat_18_participation'].mean(), c='red')

# plot titles and labels
plt.xlabel('2018 SAT Participation Rates')
plt.title('Test for Normality for Sat 2018');
```



Summary

- Based on the findings found in the 2017 & 2018 data, and after some analysis, it is amazing to see how the news confirms the data findings, i.e. that there is indeed a negative relationship between performance & participation.

- It would have been useful to find other data sources to do more complex analysis, such as demographics data and how that can affect over all State test scores.

- I initially tried to compile percentile data so to uniformalise the scores between SAT & ACT, however, there was not enough time to be able to conduct this investigation.

Thank you.