

D5_inference (Score: 1.8 / 2.0)

1. Test cell (Score: 0.1 / 0.1)
2. Test cell (Score: 0.5 / 0.5)
3. Test cell (Score: 0.25 / 0.25)
4. Coding free-response (Score: 0.0 / 0.2)
5. Test cell (Score: 0.2 / 0.2)
6. Test cell (Score: 0.1 / 0.1)
7. Test cell (Score: 0.2 / 0.2)
8. Test cell (Score: 0.15 / 0.15)
9. Test cell (Score: 0.2 / 0.2)
10. Test cell (Score: 0.1 / 0.1)

D5: Inference¶

In this notebook, we'll use statistical approaches to attempt to determine what countries the best soccer (football, for the rest of the world) players come from. Some argue that Brazil produces the best footballers, while others argue it's got to be France. (And other have different opinions still.) We'll investigate this question using FIFA data from 2020.

Run the following cell. These are all you need for the assignment. Do not import additional packages.

In [1]:

```
# Imports
%matplotlib inline

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import math
import seaborn as sns
sns.set()
sns.set_context('talk')

import warnings
warnings.filterwarnings('ignore')
pd.set_option("display.max_columns", 104)
import patsy
import statsmodels.api as sm
import scipy.stats as stats
from scipy.stats import ttest_ind, chisquare, normaltest
# Note: the statsmodels import may print out a 'FutureWarning'. Thats fine.
```

Part 1: Data & Wrangling¶

Fixing messy data makes up a large amount of the work of being a data scientist.

The real world produces messy measurements and it is your job to find ways to standardize your data such that you can make useful analyses out of it.