

SEARCH 

RESOURCES

CONCEPTS

- ✓ 1. Video: Introduction
- ✓ 2. Video: From Sampling Distributi...
- ✓ 3. ScreenCast: Sampling Distributi...
- ✓ 4. Notebook + Quiz: Building Confi...
- ✓ 5. ScreenCast: Difference In Means
- ✓ 6. Notebook + Quiz: Difference in ...
- ✓ 7. Video: Confidence Interval Appli...
- ✓ 8. Video: Statistical vs. Practical Sig...
- ✓ 9. Quiz: Statistical vs. Practical Sig...
- ✓ 10. Video: Traditional Confidence I...
- ✓ 11. ScreenCast: Traditional Confi...



Mentor Help

Ask a mentor on our Q&amp;A platform


Peer Chat ☐

Chat with peers and alumni



**jupyter** Traditional Confidence Intervals Last Checkpoint: an hour ago (autosaved)

File Edit View Insert Cell Kernel Help



```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
np.random.seed(42)

coffee_full = pd.read_csv('coffee-dataset.csv')
coffee_red = coffee_full.sample(200)

In [2]: diff = []

for _ in range(10000):
    bootsample = coffee_red.sample(200, replace=True)
    mean_coff = bootsample[bootsample['drinks_coffee'] == True]['height'].mean()
    mean_nocoff = bootsample[bootsample['drinks_coffee'] == False]['height'].mean()
    diff.append(mean_coff - mean_nocoff)

np.percentile(diff, 2.5), np.percentile(diff, 97.5)

Out[2]: (0.39656867909086274, 2.2432588681124224)

In [3]: import statsmodels.stats.api as sms

X1 = coffee_red[coffee_red['drinks_coffee'] == True]['height']
X2 = coffee_red[coffee_red['drinks_coffee'] == False]['height']

cm = sms.CompareMeans(sms.DescrStatsW(X1), sms.DescrStatsW(X2))
cm.tconfint_diff(usevar='unequal')

/Users/clamps/anaconda/envs/RoboND/lib/python3.5/site-packages/statsmodels/compat/pandas.
ndas.core.datetools module is deprecated and will be removed in a future version. Please
e instead.
from pandas.core import datetools

Out[3]: (0.39600106159185644, 2.2734131570228908)

In [ ]:
```

In this video you saw a comparison of the traditional method for calculating a confidence interval using a python built in to the bootstrapping method you have been using through the course.

With large sample sizes, these end up looking very similar. With smaller sample sizes, the traditional methods likely has assumptions that are not true of your interval. So while bootstrapping is not ideal for bootstrapping methods though either, as they can lead to misleadingly not accurately representing your entire population well.