PGE 338 Data Analytics and Geostatistics

Lecture 7b: Confidence Intervals and Hypothesis Testing

Lecture outline . . .

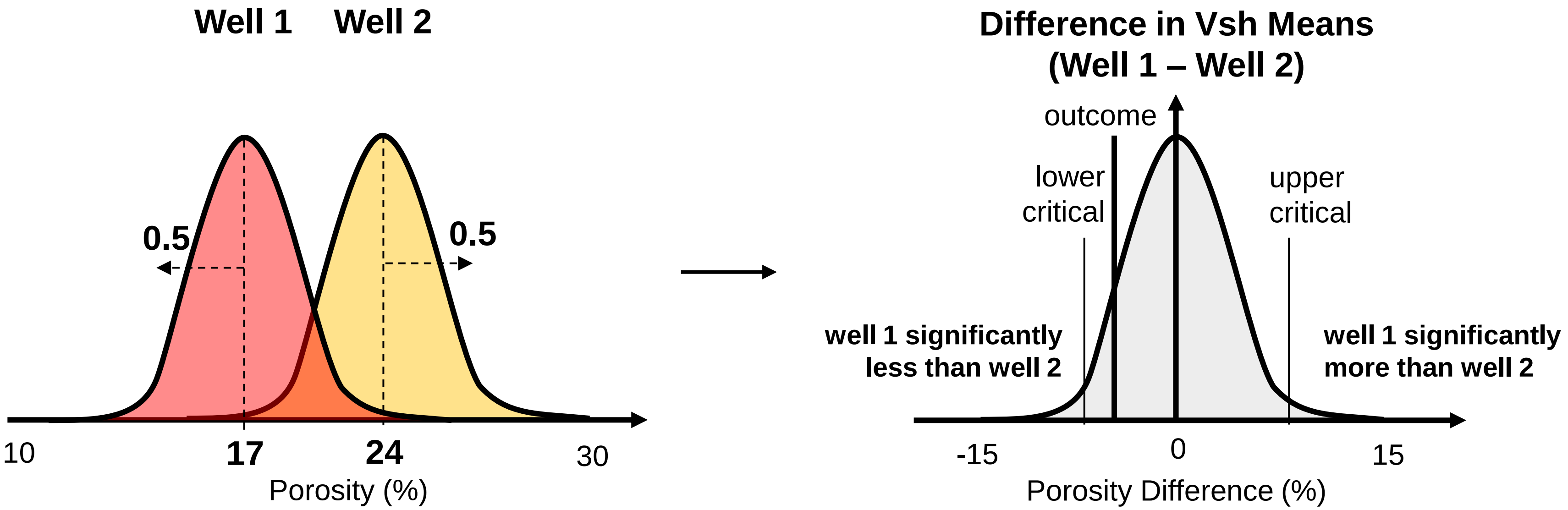
- Concepts
- Analytical Hypothesis Testing
- Bootstrap Hypothesis Testing
- Examples

Michael Pyrcz, The University of Texas at Austin

Motivation

We need to report uncertainty and significance!

- Otherwise, we do not know if any of our results are meaningful.



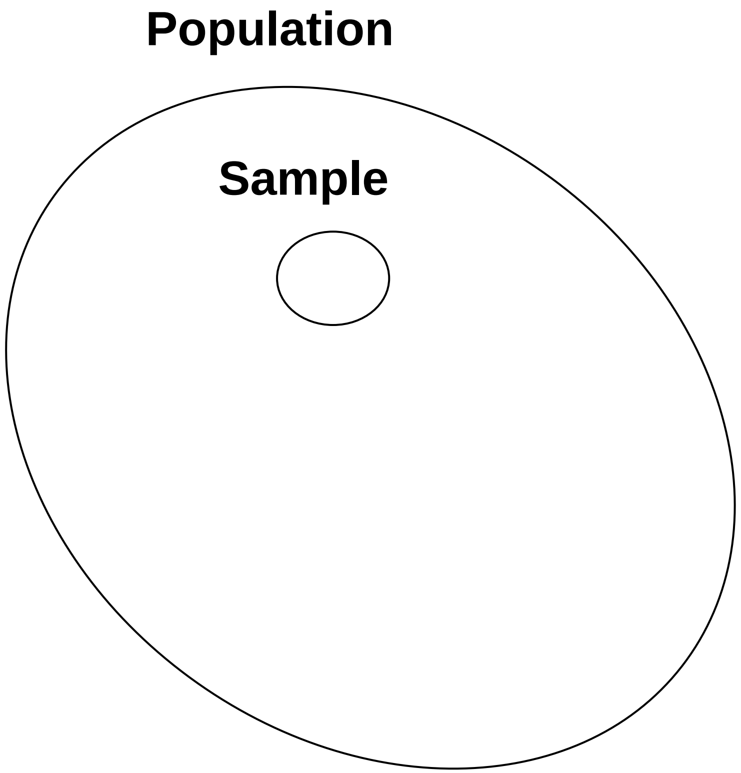
Average fraction of shale uncertainty models that appear to be different (left), but is that different significant (right).

- For example, are these wells from the same reservoir?

Review of Nomenclature

Recall our Nomenclature for sample statistics and population parameters.

	Sample	Population
proportion	$\hat{p}$	p
mean	$\bar{x}$	μ
standard deviation	s	σ
variance	s^2	σ^2



PGE 338 Data Analytics and Geostatistics

Lecture 7b: Confidence Intervals and Hypothesis Testing

Lecture outline . . .

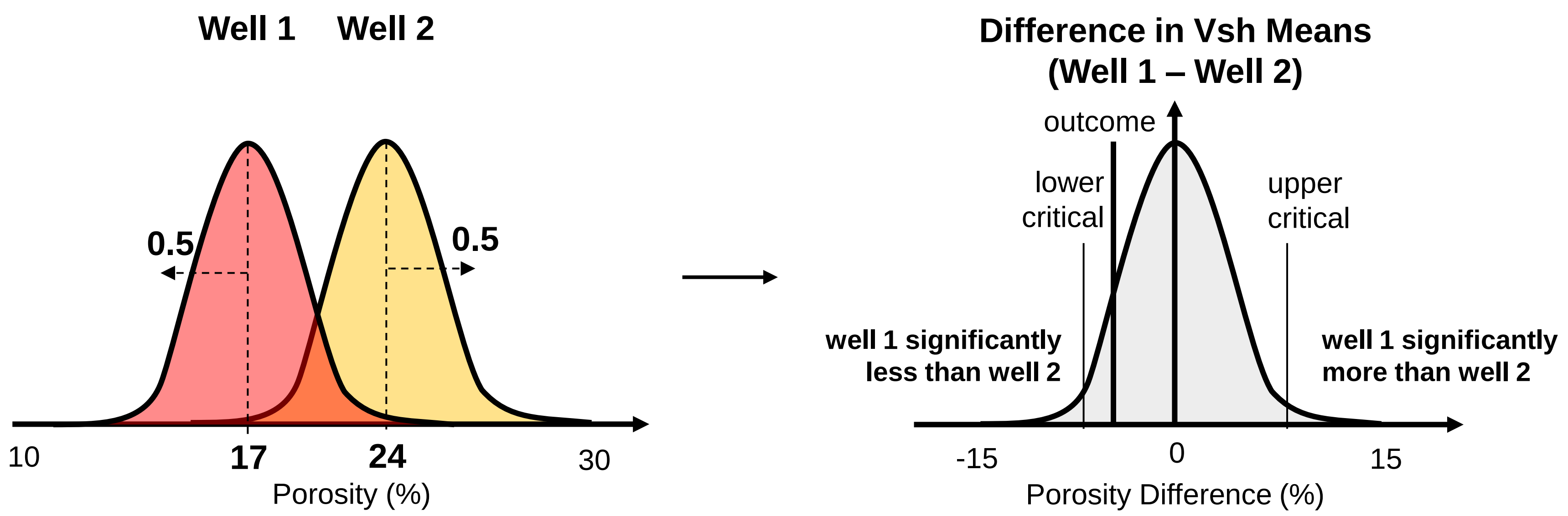
- Concepts

Michael Pyrcz, The University of Texas at Austin

From Confidence Intervals to Hypothesis Testing

Confidence Intervals are Useful – Error in your Measure!

- What about this situation? Two measures with error, did they come from the same population?



Average porosity uncertainty models that appear to be different (left), but is that difference significant (right).

- **Impact:** Is the new offset well drilling through the same type of rock as the previous wells? If so, lets use our previously optimized completion method. If not, we have to determine new completion parameters. This may delay and increase cost of additional drilling.
- **This is hypothesis testing.** e.g. Our metric and the theoretical distribution is the difference in means of samples.

Hypothesis Testing Definitions and Design

- **Hypothesis:** A statement about a population parameter

- **Method:** start by accepting the null hypothesis, H_0, and then test it.

- **Null Hypothesis (**H_0**) and Alternative Hypothesis (**H_1) are two complementary hypotheses in a hypothesis testing problem
  - $\mathbf{H_0}$ – no effect, e.g. means are the same, one is not larger, difference due to limited samples and random effect
  - $\mathbf{H_1}$ – a significant difference, the effect we are checking for

- **Sampling Distributions** are the theoretical distributions given the null hypothesis is true.
  - **Student's t-test, Independent 2-sample**: Compares the means of two sample sets
  - **F-test of equality of 2 variances**: Compares the variances of two sample sets
  - **Chi-Square test**: Compares entire histograms of two sample sets

- **Result:**
  - **Reject the null hypothesis**, evidence to support the alternative hypothesis H_1
  - **Fail to reject the null hypothesis**, retain the null hypothesis, H_0

hypothesis lives to fight another day

- **Error Types:**
  - **Type 1 error**: incorrectly rejecting the null hypothesis (false positive)
  - **Type 2 error**: incorrectly retaining the null hypothesis (false negative)

Hypothesis Testing Example Designs

Example Hypotheses

Collected data from a new well. Does this new data come from the same population as the previous wells or is there a geological discontinuity?

- Test if the means are the same between the new data and the old data.

H_0: $\mu_n = \mu_p$, the well data is from the same population as the previous wells.
H_1: $\mu_n \neq \mu_p$, the well is from a different population than the previous wells.

Applied a new lab measurement for permeability from core data. Does the new method result in consistent amount of permeability variance?

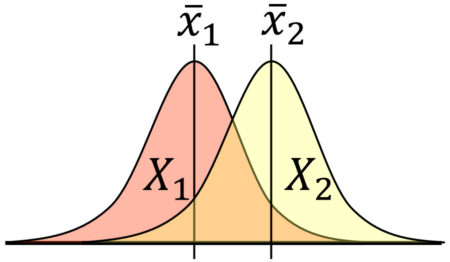
- Test if the variance is the same between the two methods.

H_0: $\sigma_{m1}^2 = \sigma_{m2}^2$, the permeability variance of the two methods is the same.
H_1: $\sigma_{m1}^2 \neq \sigma_{m2}^2$, the permeability variance of the two methods is different.

Slide modified from Dr. Zoya Heidari's PGE 337.

Hypothesis Testing Difference in Means

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$



Given 2 samples are the means significantly different?

Confidence Level

Degrees of freedom

	90%	95%	99%
	Tail area probability, α		
df	0.05	0.025	0.005
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
18	1.734	2.101	2.878
20	1.725	2.086	2.845
22	1.717	2.074	2.819
24	1.711	2.064	2.797
26	1.706	2.056	2.779
28	1.701	2.048	2.763
30	1.697	2.042	2.750
40	1.684	2.021	2.704
60	1.671	2.000	2.660
∞	1.645	1.960	2.576

$t_{critical}$ table from Jensen et al., (2000).

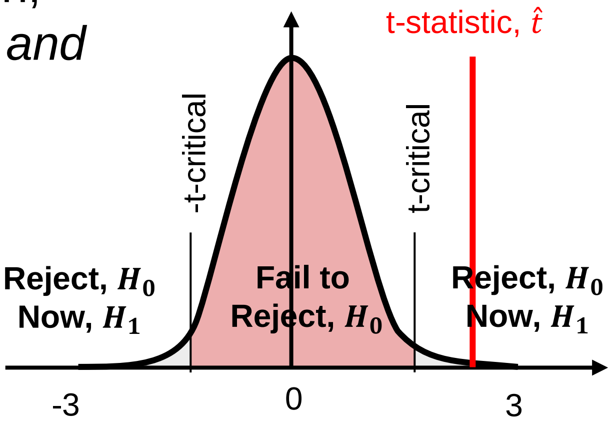
Method:

1. Calculate test statistic, e.g., t-statistic, $\hat{t}$, (measure, $\bar{x}_1 - \bar{x}_2$, divided by the standard error), in *standard deviation, from the null hypothesis*.

2. Look up the critical value, e.g., $t_{critical}$, from a table or a function, *confidence level interval in standard deviations based on alpha and degrees of freedom*. t-dist mean is 0.0.

3. Compare the test statistic to the critical value, e.g.:

Fail to reject H_0 if $-t_{critical} < \hat{t} < t_{critical}$
Reject H_0 if $\hat{t} < -t_{critical}$ or $\hat{t} > t_{critical}$



Sampling distribution, the Student's t distribution (mean = 0, standard deviation = 1.0)

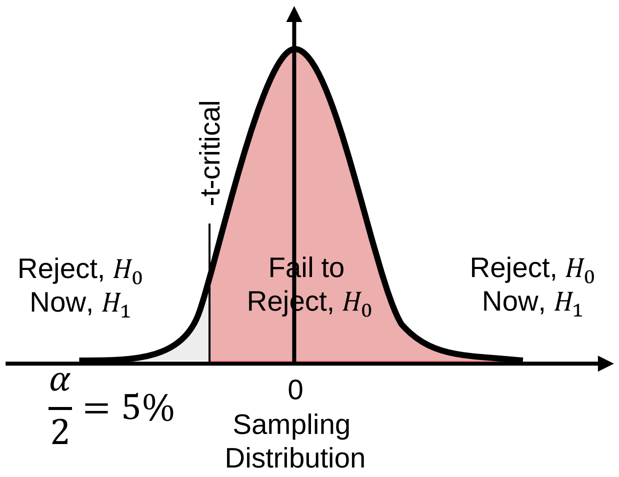
Hypothesis Testing Difference in Means

Test: is well 1 in better rock than well 2 (by average porosity)?

One Tail Test:

H_0: $\mu_1 \leq \mu_2$, the well 1 average porosity is less than or equal to the well 2 average porosity.

H_1: $\mu_1 > \mu_2$, the well 1 average porosity is greater than well 2 average porosity.



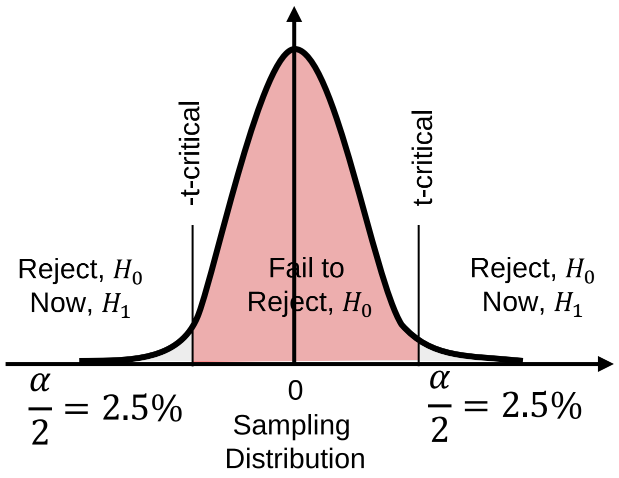
One tail test at alpha = 0.05.

Test: are wells 1 and 2 in different rock (according to average porosity)?

Two Tail Test:

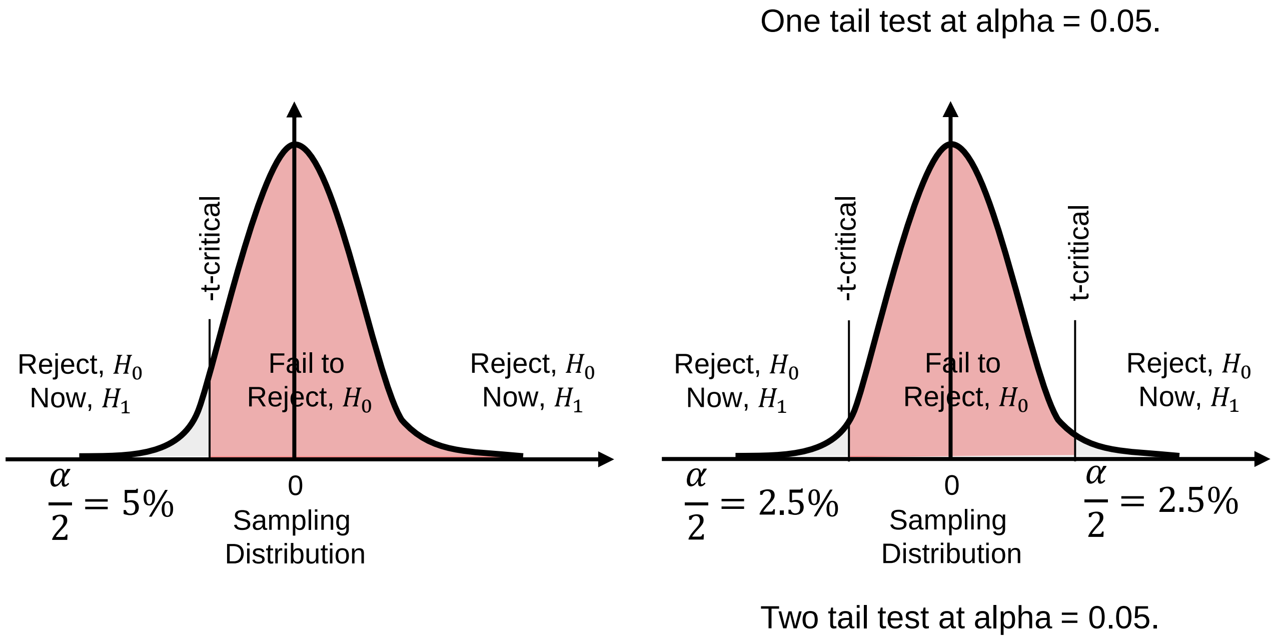
H_0: $\mu_1 = \mu_2$, the well 1 & 2 average porosity are the same.

H_1: $\mu_1 \neq \mu_2$, the well 1 & 2 average porosity are the different.



Two tail test at alpha = 0.05.

Hypothesis Testing Difference in Means



Limitations of Hypothesis Testing

Limitations of Hypothesis Testing:

- Publication Bias – only publish results that reject the null
  - Recall, α probability of false positive
  - Data mining for any effect will find insignificant differences that look significantly different (by random)
  - Many studies have selectively reported significant results
- Very small sample sizes, poorly understood phenomenon
  - Can't demonstrate conformity with the test assumptions
- Other Data Issues, Poor Sampling Practice
  - Contamination of sample 'clever Hans effect'
  - Placebo effect, impact of subject motivation

Clever Hans, the horse that could do math in the early 1900s.

Due to these there is some mistakes and skepticism concerning the use of hypothesis testing in publications.

Image from https://www.damninteresting.com/clever-hans-the-math-horse/

PGE 338 Data Analytics and Geostatistics

Lecture 7b: Confidence Intervals and Hypothesis Testing

Lecture outline . . .

- Analytical Hypothesis Testing

Michael Pyrcz, The University of Texas at Austin

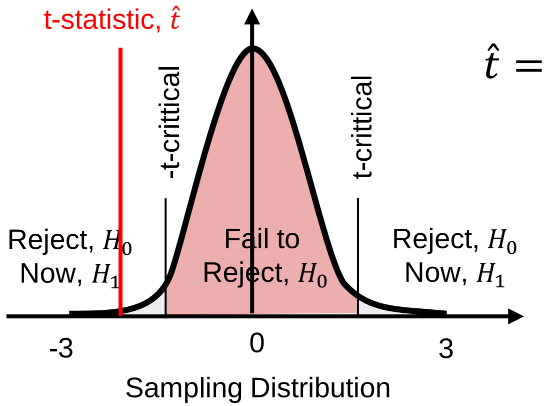
Introduction
General Concepts
Univariate
PDF / CDF
Statistics
Distributions
Heterogeneity
Hypothesis
Bivariate
Time Series Analysis
Spatial Analysis
Machine Learning
Uncertainty Analysis

Hypothesis Testing
Difference in Means

The Student's t-test, equal variances, pooled variance method (William Gosset, 1876-1937)

The t-statistic is:

The Pooled Variance Test in Python: SciPy.stats.ttest_ind(X1,X2)



$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)}} = \frac{measure}{standard\ error} = t\ statistic, \hat{t}$$

$$t_{critical} = \left| t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right) \right|$$

Cumulative probability

Degrees of freedom

α = 1.0 – confidence level

Alpha level

Excel t.inv($\frac{\alpha}{2}$,$n_1 + n_2 - 2$)

Python SciPy.stats.t.ppf($\frac{\alpha}{2}$,$n_1 + n_2 - 2$)

If $-t_{critical} < \hat{t} < t_{critical}$ then we fail to reject the null hypothesis. Our current hypothesis remains that there is no difference between the means.

We can get the $t_{critical}$ value from our $\hat{t}$ and d.f. using a table or a calculator (http://www.statisticshowto.com/t-score-formula/), see Excel t.inv or Python SciPy.stats.t.ppf functions.

This method assumes that the **variables are Gaussian distributed, and the standard deviations are not significantly different**.

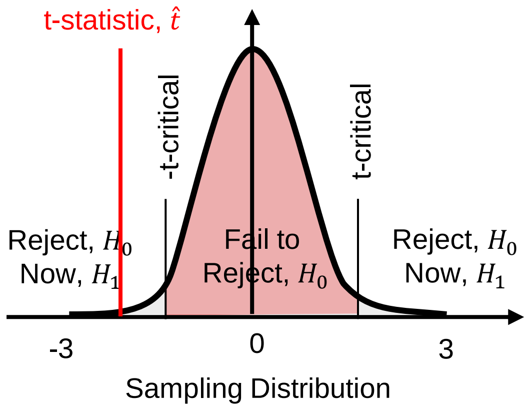
Hypothesis Testing
Difference in Means

The Student's t-test for unequal variances, Welch's t-test (Bernard Welch, 1911-1989)

The Welch's t-test in Python: SciPy.stats.ttest_ind(X1,X2,equal_var=False)

The t statistic is:

α = 1.0 – confidence level



$$\hat{t} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \frac{measure}{standard\ error} = t\ statistic, \hat{t}$$

$$u = \frac{s_2^2}{s_1^2}$$

$$t_{critical} = \left| t\left(\frac{\alpha}{2}, \nu\right) \right|$$

Alpha level

Degrees of freedom

$$\nu = \frac{\left(\frac{1}{n_1} + \frac{u}{n_2}\right)^2}{\frac{1}{n_1^2(n_1 - 1)} + \frac{u^2}{n_2^2(n_2 - 1)}}$$

If $-t_{critical} < \hat{t} < t_{critical}$ then we fail to reject the null hypothesis. Our current hypothesis remains that there is no difference between the means.

We can get the $t_{critical}$ value from our $\hat{t}$ and d.f. using a table or a calculator (http://www.statisticshowto.com/t-score-formula/), see Excel t.inv or Python SciPy.stats.t.ppf functions.

This method assumes that the variables are Gaussian distributed, and variances are unequal. Welch's t-test.

Hypothesis Testing
Difference in Means

Example #1 Hypothesis Test for Difference in Means, Equal Variance Method

The Problem:

- Well 1 (20 samples, mean porosity = 13%, st. dev. = 2%)
- Well 2 (25 samples, mean porosity = 15%, st. dev. = 3%)
- At a 95% confidence level (alpha = 0.05) test $H_0\!: \mu_1 = \mu_2$

Recall:

$$Alpha = 1 - Confidence\ Level$$

Measure

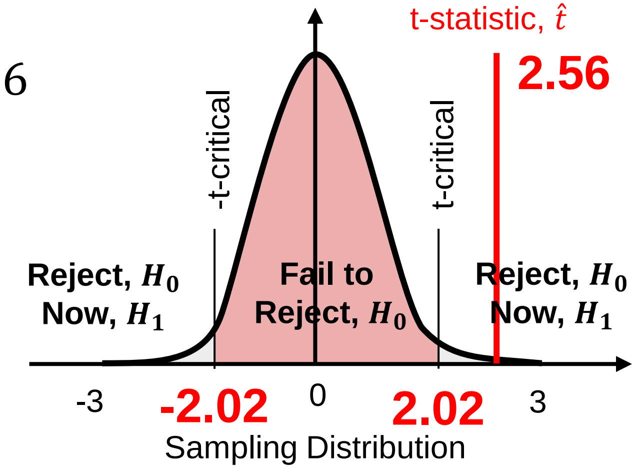
t-statistic

$$\hat{t} = \frac{|13\% - 15\%|}{\sqrt{\left(\frac{1}{20}+\frac{1}{25}\right)\left(\frac{(20-1)4+(25-1)9}{20+25-2}\right)}} = \frac{2}{0.78} = 2.56$$

SE

t-critical

$$t_{critical} = \left|t\left(\frac{0.05}{2}, 20+25-2\right)\right| = \pm 2.02$$



If $\hat{t} = 2.56$ is outside interval $-2.02 < \hat{t} < 2.02$; therefore, we **reject the null hypothesis**. We adopt the alternative hypothesis that $\boldsymbol{H_1\!: \mu_1 \neq \mu_2}$

If the means are significantly different, then the distributions are significantly different, and we are drilling potentially new rock.

Hypothesis Testing
Difference in Means

Example #2 Hypothesis Test for Difference in Means, Equal Variance Method

The Problem:

- Determine if the two data sets came from the same reservoir unit. Compare the two sets of sample for porosity uploaded on Canvas website (data set PorositySample2Units.xlsx).

- At a 95% confidence level (alpha = 0.05) test $H_0\!: \mu_1 = \mu_2$

- Use pooled variance method.

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)}}$$

$$t_{critical} = \left|t\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right)\right|$$

$t_{critical}$ →T.INV.2T($\alpha, n_1 + n_2 - 2$) or T.INV($\alpha/2, n_1 + n_2 - 2$) in Excel.

Hypothesis Testing
Difference in Means

Example #2 Hypothesis Test for Difference in Means, Equal Variance Method

The Problem:

- Determine if the two data sets came from the same reservoir unit. Compare the two sets of sample for porosity uploaded on Canvas website (data set PorositySample2Units.xlsx).

- At a 95% confidence level (alpha = 0.05) test $H_0: \mu_1 = \mu_2$

- Use pooled variance method.

$$\hat{t} = \frac{|0.16 - 0.20|}{\sqrt{\left(\frac{1}{20} + \frac{1}{20}\right)\left(\frac{(20-1)0.0008 + (20-1)0.0021}{20+20-2}\right)}} = 2.981$$

$$t_{critical} = \left| t\left(\frac{0.05}{2}, 20+20-2\right) \right| = \pm 2.02$$

Result:

Reject $H_0: \mu_1 = \mu_2$

Adopt $H_1: \mu_1 \neq \mu_2$

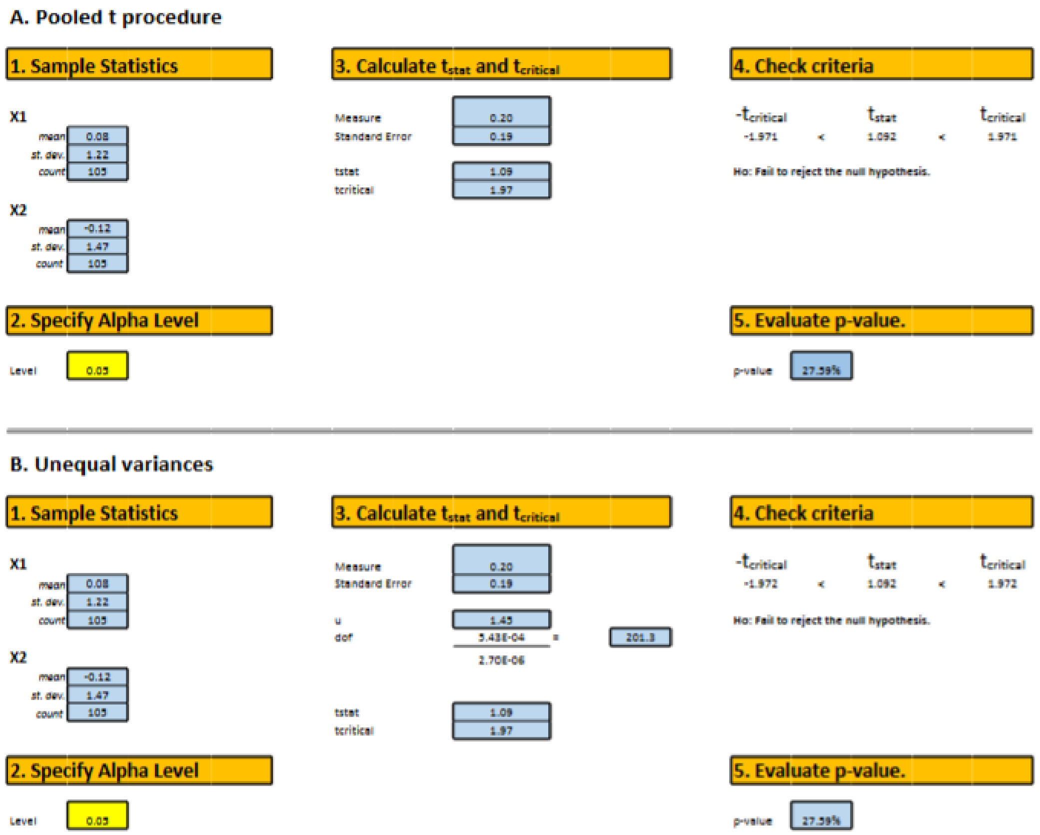
$\therefore$ **2 different reservoir units.**

Hypothesis Testing
Difference in Means in Excel

Example #3 Hypothesis Test for Difference in Means, Equal Variance & Welch's Methods

The Problem:

- set up in an Excel sheet for any dataset. It calculates the sample statistics and completes the test.



Difference in means hypothesis test demonstration in Excel, file is Difference_in_means.xlsx.

Hypothesis Testing
Difference in Means in Python

Hypothesis Test for Difference in Means, Equal Variance & Welch's Methods

The Problem:

- Difference in means hypothesis tests in Python

Data Analytics

Confidence Intervals and Hypothesis Testing in Python in Python

Michael Pyrcz, Associate Professor, The University of Texas at Austin

Twitter | *GitHub* | *Website* | *GoogleScholar* | *Book* | *YouTube* | *LinkedIn*

This is a tutorial / demonstration of **Confidence Intervals and Hypothesis Testing in Python**. In Python, the SciPy package, specifically the Stats functions (https://docs.scipy.org/doc/scipy/reference/stats.html) provide excellent tools for efficient use of statistics.

I have previously provided these examples worked out by-hand in Excel (https://github.com/GeostatsGuy/LectureExercises/blob/master/Lecture7_CI_Hypoth_eg_R.xlsx) and also in R (https://github.com/GeostatsGuy/LectureExercises/blob/master/Lecture7_CI_Hypoth_eg.R). In all cases, I use the same dataset available as a comma delimited file (https://git.io/fxLAt).

This tutorial includes basic, typical confidence interval and hypothesis testing methods that would commonly be required for Engineers and Geoscientists including:

1. Student-t confidence interval for the mean and proportion
2. Student-t hypothesis test for difference in means (pooled variance)
3. Student-t hypothesis test for difference in means (difference variances), Welch's t Test
4. F-distribution hypothesis test for difference in variances

Caveats

I have not included all the details, specifically the test assumptions in this document. These are included in the accompanying course notes, Lec08_hypothesis.pdf.

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as stats
```

```python
df = pd.read_csv('https://raw.githubusercontent.com/GeostatsGuy/GeoDataSets/master/PorositySample2Units.csv')
X1 = df['X1'].values; X2 = df['X2'].values
```

```python
alpha = 0.05
t_critical = stats.t.ppf(alpha/2,len(X1)+len(X2)-2)

t_pooled, p_pooled = stats.ttest_ind(X1,X2) # assuminng equal variance
if(t_pooled < t_critical or t_pooled > -1*t_critical):
    print('t-critical and t-statistic are ' + str(np.round(t_critical,2)) + ' ≤ ' + str(np.round(t_pooled,2)) +
      ' ≤ ' + str(np.round(-1*t_critical,2)) + '; therefore, reject the null hypothesis')
else:
    print('t-critical and t-statistic are ' + str(np.round(t_critical,2)) + ' ≤ ' + str(np.round(t_pooled,2)) +
      ' ≤ ' + str(np.round(-1*t_critical,2)) + '; therefore, fail to reject the null hypothesis')
```

```
t-critical and t-statistic are -2.02 ≤ -2.98 ≤ 2.02; therefore, reject the null hypothesis
```

```python
alpha = 0.05
t_critical = stats.t.ppf(alpha/2,len(X1)+len(X2)-2)

t_welch, p_welch = stats.ttest_ind(X1,X2,equal_var = False) # assuminng unequal variance, Welch's t-test
if(t_welch < t_critical or welch > -1*t_critical):
    print('t-critical and t-statistic are ' + str(np.round(t_critical,2)) + ' ≤ ' + str(np.round(t_welch,2)) +
      ' ≤ ' + str(np.round(-1*t_critical,2)) + '; therefore, reject the null hypothesis')
else:
    print('t-critical and t-statistic are ' + str(np.round(t_critical,2)) + ' ≤ ' + str(np.round(t_welch,2)) +
      ' ≤ ' + str(np.round(-1*t_critical,2)) + '; therefore, fail to reject the null hypothesis')
```

```
t-critical and t-statistic are -2.02 ≤ -2.98 ≤ 2.02; therefore, reject the null hypothesis
```

Short Python demonstration with PorositySample2Units dataset.

Difference in means hypothesis test demonstration in Python, file is PythonDataBasics_ConfidenceInterval_HypothesisTesting.xlsx.

Hypothesis Testing
Difference in Variances

The F-test for Difference in Variance (Snedecor and Cochran, 1989)

- Compares variances of two distributions
  - Example: comparison of the heterogeneity of two samples sets (e.g., from 2 wells) to determine if the wells have different heterogeneity

Requirements:

- The sample and population distributions are of both samples are Gaussian, but at 5% alpha with similar number of samples it is robust if non-normal (not Gaussian distributed).

The test:

- **Null hypothesis,** $H_0 \colon {}^{\sigma_2^2}\!/_{\sigma_1^2} = 1.0, where\ \ \sigma_2^2 > \sigma_1^2.$ The data comes from independent random samples from normal distributions with equal variances

- **Alternative hypothesis,** $H_1 \colon {}^{\sigma_2^2}\!/_{\sigma_1^2} > 1.0$: The data comes from populations with unequal variances

Slide modified from Dr. Zoya Heidari's PGE 337.

Hypothesis Testing
Difference in Variances

The F-test (Snedecor and Cochran, 1989)

- The F-test statistic

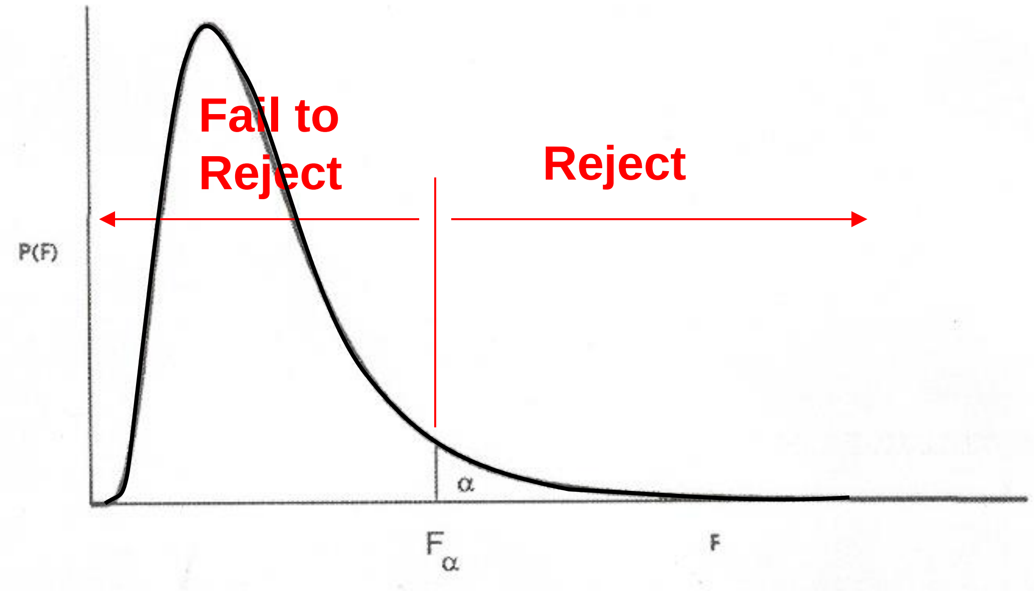
$$\hat{F} = \frac{s_a^2}{s_b^2} \quad \text{where} \quad s_a^2 > s_b^2$$

if $\hat{F} < F_{critical}$ then we fail to reject H_0 and state that the two variances are not significantly different.

$$F_{critical} = f(n_a - 1, n_b - 1, \alpha)$$

Degrees of Freedom

Alpha Level



F-distribution with reject and fail to reject regions.

Image from http://uregina.ca/~gingrich/f.pdf

Hypothesis Testing
Difference in Variances

How to calculate the $F_{critical}$ values?

- A table of $F_{critical}$ at a specific α level.
- Excel - F.INV(1-α, $n_a - 1$, $n_b - 1$)
- Python - scipy.stats.f.ppf(1-α, $n_a - 1$, $n_b - 1$),

F	Degrees of Freedom in the Numerator																			
α = 0.05	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	50	100	200	1000
Degress of Freedom in the Denominator 1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.3	250.1	251.1	251.8	253.0	253.7	254.2
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.49	19.49	19.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.55	8.54	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.70	5.66	5.65	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.41	4.39	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.75	3.71	3.69	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.32	3.27	3.25	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.02	2.97	2.95	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.80	2.76	2.73	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.64	2.59	2.56	2.54
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.40	2.35	2.32	2.30
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.18	2.12	2.10	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.97	1.91	1.88	1.85
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.84	1.78	1.75	1.72
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.76	1.70	1.66	1.63
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.66	1.59	1.55	1.52
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.87	1.78	1.73	1.69	1.63	1.60	1.52	1.48	1.45
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.77	1.68	1.62	1.57	1.52	1.48	1.39	1.34	1.30
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.72	1.62	1.56	1.52	1.46	1.41	1.32	1.26	1.21
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.68	1.58	1.52	1.47	1.41	1.36	1.26	1.19	1.11

Table to look up $F_{critical}$ values at α = 0.05.

Hypothesis Testing
Difference in Variances

Example #1 Hypothesis Test for Difference in Variance
The Problem:

- Are the variances different for well 1 and well 2 at α = 0.05?

- Variance: well 1 variance = 4%^{2} with 20 samples, well 2 variance = 9%^{2} with 25 samples

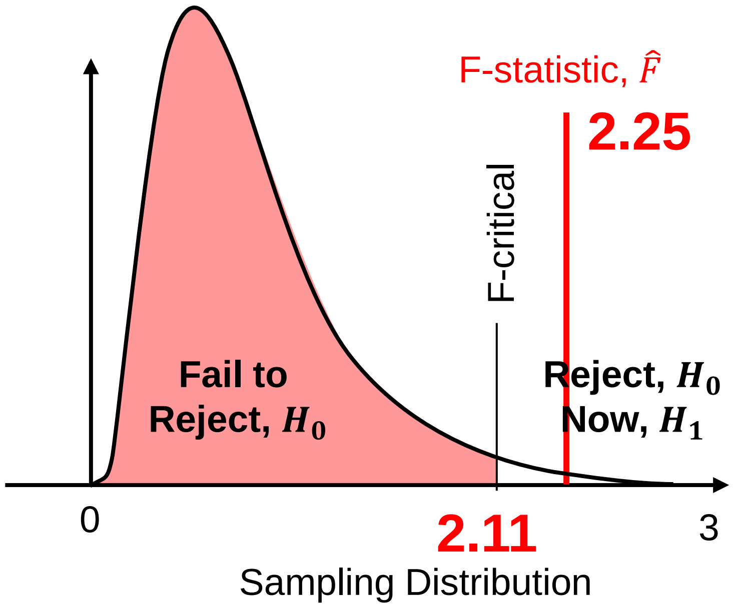
$$\hat{F} = \frac{9}{4} = 2.25$$

if $\hat{F} \leq F_{critical}$ then we fail to reject H_0 and state that the two variances are not significantly different.

$$F_{critical} = f(25-1, 20-1, 0.05) = 2.11$$

Degrees of Freedom

Significant Level



$\hat{F} > F_{critical}, Reject\ H_0$, variances are significantly different.

Hypothesis Testing
Difference in Variances

Example #2 Hypothesis Test for Difference in Variance

The Problem:

- Determine if the two data sets came from the same reservoir unit. Compare the two sets of sample for porosity uploaded on Canvas website (dataset PorositySample2Units.xlsx).

- Are the variances significantly different at a 95% confidence level?

$$\hat{F} = \frac{s_a^2}{s_b^2} \quad \text{where} \quad s_a^2 > s_b^2$$

$$F_{critical} = f(n_a - 1, n_b - 1, \alpha)$$

if $\hat{F} < F_{critical}$ then we fail to reject H_0 and state that the two variances are not significantly different.

Hypothesis Testing
Difference in Variances in Python

Hypothesis Test for Difference in Variance

The Problem:

- Determine if the two data sets came from the same reservoir unit. Compare the two sets of sample for porosity uploaded on Canvas website (dataset PorositySample2Units.xlsx).

- Are the variances significantly different at a 95% confidence level?

Data Analytics

Confidence Intervals and Hypothesis Testing in Python in Python

Michael Pyrcz, Associate Professor, The University of Texas at Austin

Twitter | *GitHub* | *Website* | *GoogleScholar* | *Book* | *YouTube* | *LinkedIn*

This is a tutorial / demonstration of **Confidence Intervals and Hypothesis Testing in Python**. In Python, the SciPy package, specifically the Stats functions (https://docs.scipy.org/doc/scipy/reference/stats.html) provide excellent tools for efficient use of statistics.

I have previously provided these examples worked out by-hand in Excel (https://github.com/GeostatsGuy/LectureExercises/blob/master/Lecture7_CI_Hypoth_eg_R.xlsx) and also in R (https://github.com/GeostatsGuy/LectureExercises/blob/master/Lecture7_CI_Hypoth_eg.R). In all cases, I use the same dataset available as a comma delimited file (https://git.io/fxLAt).

This tutorial includes basic, typical confidence interval and hypothesis testing methods that would commonly be required for Engineers and Geoscientists including:

1. Student-t confidence interval for the mean and proportion
2. Student-t hypothesis test for difference in means (pooled variance)
3. Student-t hypothesis test for difference in means (difference variances), Welch's t Test
4. F-distribution hypothesis test for difference in variances

Caveats

I have not included all the details, specifically the test assumptions in this document. These are included in the accompanying course notes, Lec08_hypothesis.pdf.

Difference in means hypothesis test demonstration in Python, file is PythonDataBasics_ConfidenceInterval_HypothesisTesting.xlsx.

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as stats
```

```python
df = pd.read_csv('https://raw.githubusercontent.com/GeostatsGuy/GeoDataSets/master/PorositySample2Units.csv')
X1 = df['X1'].values; X2 = df['X2'].values
```

```python
var_X1 = np.var(X1); var_X2 = np.var(X2)
f_stat = np.max([var_X1,var_X2])/np.min([var_X1,var_X2])
if var_X1 > var_X2:
    f_critical = stats.f.ppf(1-alpha,len(X1)-1,len(X2)-1)
else:
    f_critical = stats.f.ppf(1-alpha,len(X2)-1,len(X2)-1)
if f_stat > f_critical:
    print('f-statistic and f-critical are ' + str(np.round(f_stat,2)) + ' > ' + str(np.round(f_critical,2))
          + '; therefore, reject the null hypothesis')
else:
    print('f-statistic and f-critical are ' + str(np.round(f_stat,2)) + ' ≤ ' + str(np.round(f_critical,2))
          + '; therefore, fail to reject the null hypothesis')
```

```
f-statistic and f-critical are 2.67 > 2.17; therefore, reject the null hypothesis
```

Short Python demonstration with PorositySample2Units dataset.

Hypothesis Testing
Difference in Variances in Excel

Hypothesis Test for Difference in Variance

The Problem:

- Determine if the two data sets came from the same reservoir unit. Compare the two sets of samples for porosity variance in dataset Porosity2Units.xlsx.
- Spreadsheet is ‘Difference_in_variances_demo.xls’

f-test for difference in variances

1. Sample Statistics

X1

mean	0.16
var	0.00077
count	20

X2

mean	0.20
var	0.0021
count	20

2. Specify Alpha Level

Level 0.05

3. Calculate f_{stat} and $f_{critical}$

fstat	2.67	
fcritical	2.17	F_{α}, N_1-1, N_2-1
P(F<=f) one tail	1.92%	

4. Check criteria

f_{stat}		$f_{critical}$	*Test*
2.668	<	2.168	$F > F_{\alpha}, N_1-1, N_2-1$

H1: Reject the null hypothesis.

5. Evaluate p-value.

p-value 1.92% < 5.0%

Reject if probability is less than confidence level.

6. Check with EXCEL built-in F-test

p-value 1.92%

Difference in variance hypothesis test demonstration in Excel, file is Difference_in_variance.xlsx.

Hypothesis Testing
Comparing Two Distributions

The Chi-Square Test (Karl Pearson, 1900)

- Compares two distributions

- Test the entire distribution, not just the mean or variance!

- Apply to any distribution

- **Null hypothesis**: The two distributions are the same
- **Alternative hypothesis**: The two distributions are different

- Commonly applied to see if a sample distribution matches an ‘expected’ theoretical distribution, e.g., is the data histogram Gaussian?

Hypothesis Testing
Comparing Two Distributions

The Chi-Square Test for Difference in Histograms (Karl Pearson, 1900)

- *n* data are grouped in K classes (*n* should be above 30)
- The appropriate test statistic is:

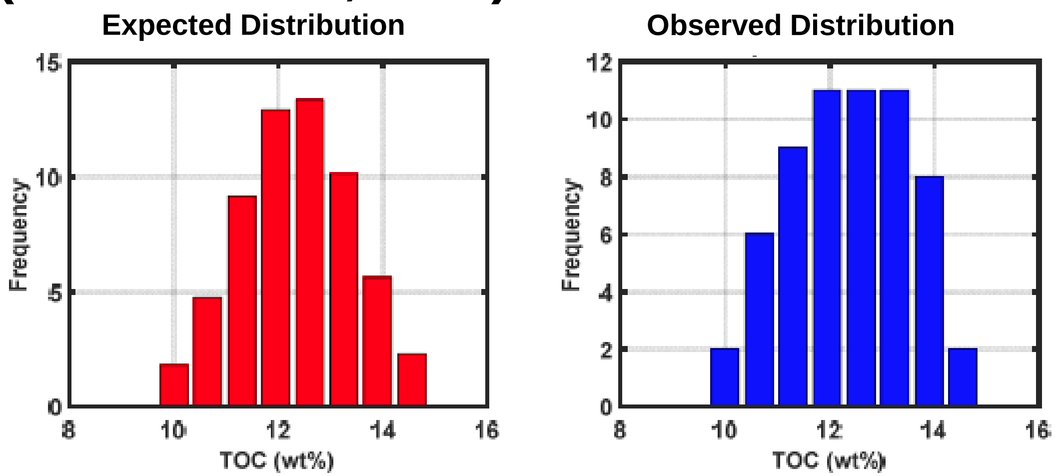
Observed Frequency

$$\hat{\chi}^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

Expected Frequency

$$\hat{\chi}^2{}_{critical} \rightarrow \hat{\chi}^{2^{-1}}(1-\alpha, K-3)$$

if $\hat{\chi}^2 < \hat{\chi}^2{}_{critical}$, fail to reject the null hypothesis. The two distributions are not significantly different.



Expected and observed histograms, with the same bins for comparison of frequencies.

Degrees of Freedom

$$\Phi = K - Z$$

# of parameters + 1, e.g., for Gaussian Z = 3.

number of classes (number of histogram bins)

if $\hat{\chi}^2 > \hat{\chi}^2{}_{critical}$, reject the null hypothesis two histograms are significantly different.

$\hat{\chi}^2{}_{critical}$ →CHISQ.INV$(1-\alpha, K-3)$ in Excel

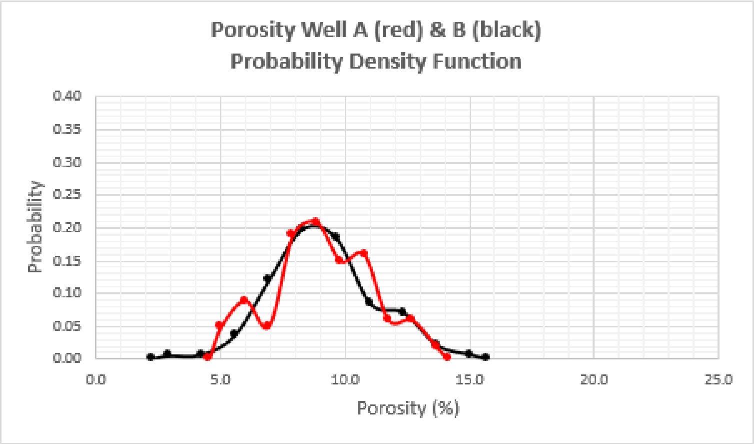
Slide modified from Dr. Zoya Heidari's PGE 337.

Hypothesis Testing
Comparing Two Distributions

Example #1 Chi-Square Test for Difference in Histograms

The Problem:

- compare 2 histograms in the Excel spreadsheet, Q-Qplot_Chi_Sq_Demo.xlsx, that calculates:
    1. *stochastic realizations of two porosity distributions*
    2. *Q-Q plot*
    3. *Chi Square test*



Two normalized histograms. Note, should be a bar chart.

Chi-Square Test for Difference In Distributions

min	2.2
max	15.7

Bins	2.24	3.59	4.94	6.29	7.63	8.98	10.33	11.68	13.03	14.38
	3.59	4.94	6.29	7.63	8.98	10.33	11.68	13.03	14.38	15.72
Freq A	0	1	6	16	28	24	18	9	3	0
Freq B	1	0	6	17	28	26	12	10	3	2
(A-B)^2/B	1.0		0.0	0.1	0.0	0.2	3.0	0.1	0.0	2.0

Degrees of Freedom	*nbins*	9	Φ	6

Chi-Square Test Result	H_0:	χ^2	6.3	<	$\chi^2_{critical}$	12.6	Fail to Reject

Calculation steps for Chi-Square test for difference in histograms.

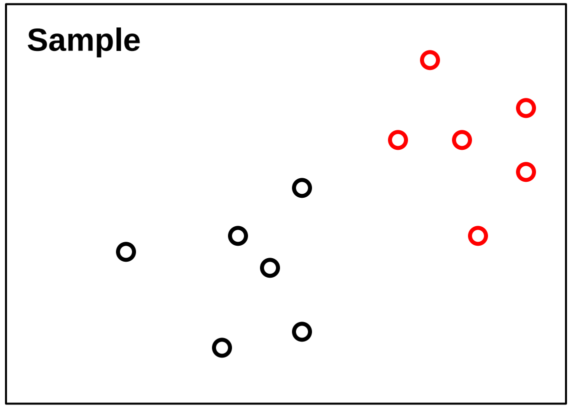
$\hat{\chi}^2_{critical}$ →CHISQ.INV(1 − α,K-3) or CHISQ.INV.RT(α,K-3) in Excel

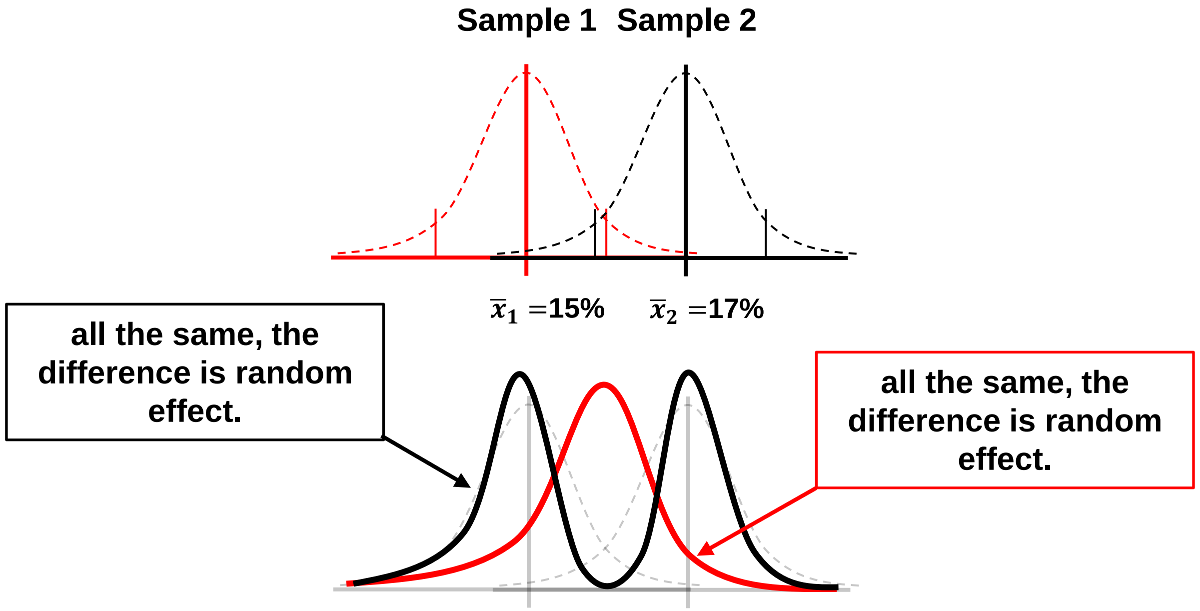
Hypothesis Testing
Take 2

The Problem:

- You have 2 datasets (1 and 2), did they come from the same population?

- If you had 2 datasets from the same population. They could look different!





- There is structure in random!

Hypothesis Testing
Take 2

The Problem:

- Belief in the law of small numbers

Psychological Bulletin
1971, Vol. 76, No. 2, 105–110

BELIEF IN THE LAW OF SMALL NUMBERS

AMOS TVERSKY AND DANIEL KAHNEMAN [1]

Hebrew University of Jerusalem

People have erroneous intuitions about the laws of chance. In particular, they regard a sample randomly drawn from a population as highly representative, that is, similar to the population in all essential characteristics. The prevalence of the belief and its unfortunate consequences for psychological research are illustrated by the responses of professional psychologists to a questionnaire concerning research decisions.

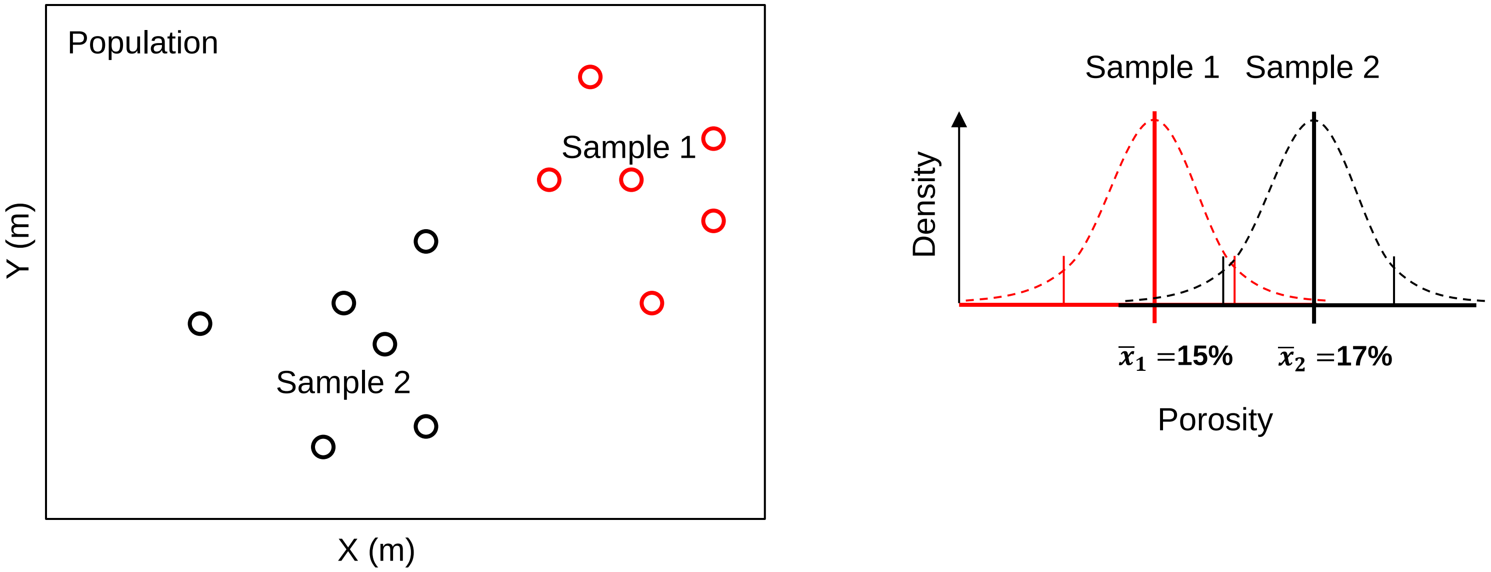
- That samples randomly drawn from a population as highly representative.

- Any statistic, e.g., mean, variance, P13 etc. will be the same!

Hypothesis Testing
Take 2

The Solution:
- What are you going to compare?
  - Mean, variance, binned frequencies?

- To start, set up the hypothesis test: $\mu_1 = \mu_2$ they come from populations with the same mean; therefore, they could be the same population. **If the means are different, can't be the same population.**



Sample (red and black) locations (left) and schematic of sample distributions (rights).

Hypothesis Testing Take 2

The Solution:

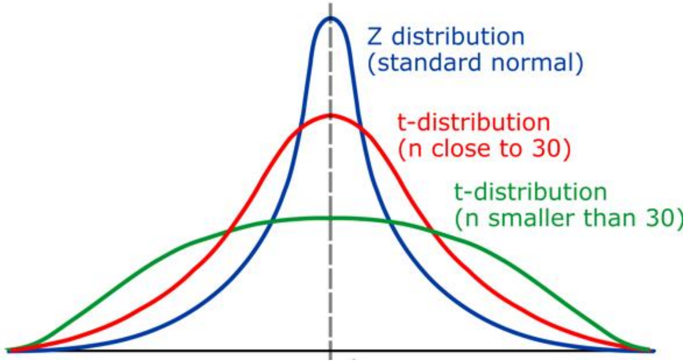
- Hypothesis test:

$$H_0\colon \mu_1 = \mu_2$$
$$H_1\colon \mu_1 \neq \mu_2,$$

Recall hypothesis is with regard to population parameters not sample statistics.

- Decide on the metric? $\bar{x}_1 - \bar{x}_2$

- What is the sampling distribution we would expect
  - Difference of 2 Gaussian random variables with small sample size and unknown σ



student's t distribution

Schematic of possible sampling distributions.

Image from https://andyjconnelly.wordpress.com/2017/05/16/uncertainty-and-repeats/

Hypothesis Testing
Take 2

The Solution:

- Hypothesis test:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2,$$

Recall hypothesis is with regard to population parameters not sample statistics.

- Decide on the metric? $\bar{x}_1 - \bar{x}_2$

- What is the sampling distribution we would expect – Student's t distribution.

- But how much difference is significant?
  - Standard error tells us how much spread due to random effect, small sample and variability in samples.

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)}$$

standard error for Student's t difference in means with equal variance

The Solution:

- Hypothesis test:

$$H_0\colon \mu_1 = \mu_2$$
$$H_1\colon \mu_1 \neq \mu_2,$$

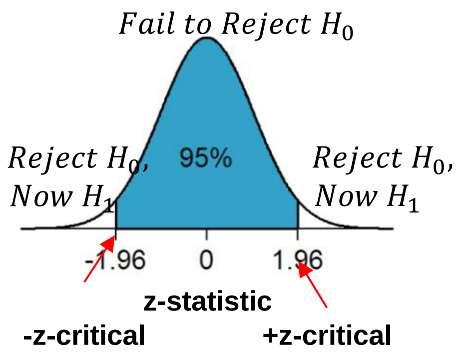
Recall hypothesis is with regard to population parameters not sample statistics.

- Decide on the metric? $\bar{x}_1 - \bar{x}_2$

- What is the sampling distribution we would expect – Student's t distribution.

- But how much difference is significant? $\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left(\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\right)}$

- What is the threshold for random effect at an alpha level?
  - this is how our metric should be distributed if both samples were sampled from distributions with the same mean.



Hypothesis Testing
Take 2

What is the p-value?

- **Common output, communication tool for a hypothesis test.**

- Indicates how close you were to rejecting / failing to reject.

- Could conduct the test by comparing p-value to α.

- For a 2 tailed test p-value is summed over both tails and we reject if p-value < α.

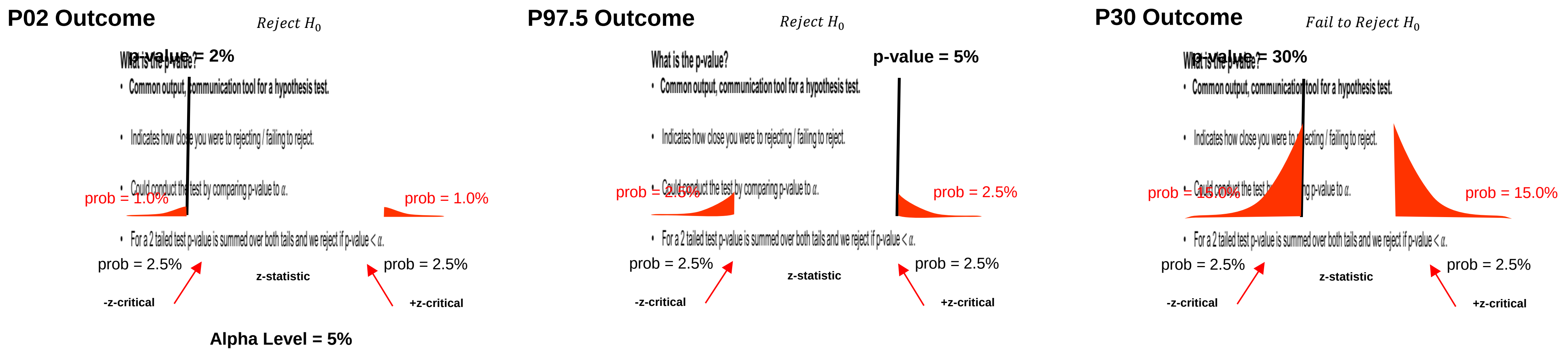


Illustration of t statistics that result in p-values of 2%, 5% and 30% respectively.

PGE 338 Data Analytics and Geostatistics

Lecture 7b: Confidence Intervals and Hypothesis Testing

Lecture outline . . .

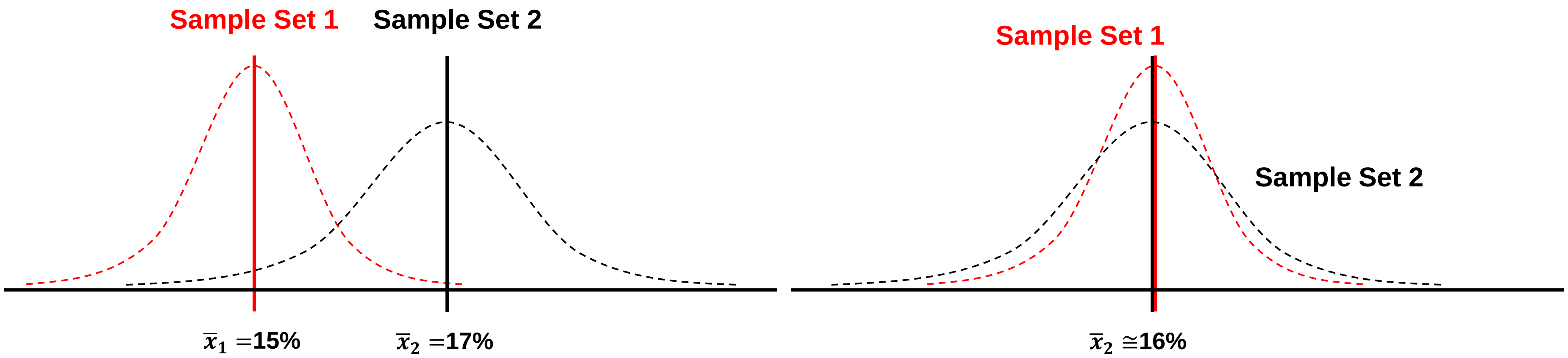
- Bootstrap Hypothesis Testing

Michael Pyrcz, The University of Texas at Austin

Hypothesis Testing
Take 3!

We can apply bootstrap to calculate the sampling distribution for hypothesis testing. E.g., difference in means hypothesis test:

1. Shift the 2 sample sets to have the mean of the combined sample set (1 and 2 together), ensure $H_0\colon \mu_1 = \mu_2$ state is true.



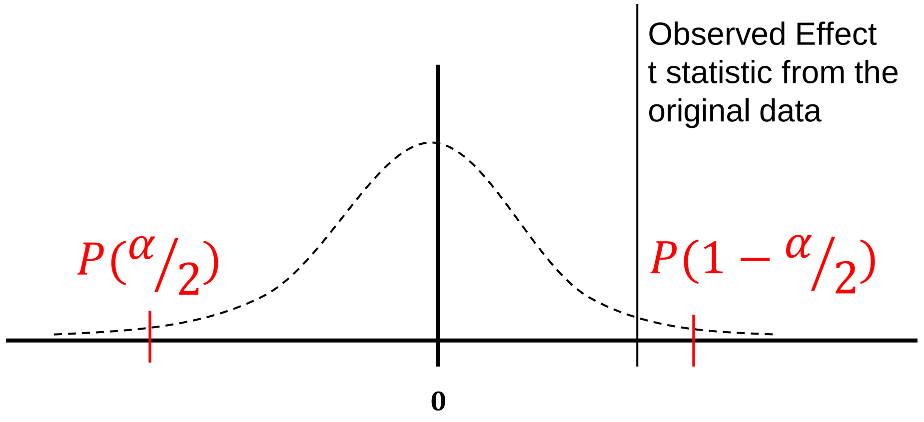
2. Bootstrap, n_1 and n_2 samples with replacement from each distribution.
3. Calculate a realization of the t statistic.

$$\hat{t}^{\ell} = \frac{\bar{x}_1{}^{\ell} - \bar{x}_2{}^{\ell}}{\sqrt{\left(\frac{s_1^{2\,\ell}}{n_1} + \frac{s_2^{2\,\ell}}{n_2}\right)}}$$

Hypothesis Testing
Take 3!

We can apply bootstrap to calculate the sampling distribution for hypothesis testing. E.g., difference in means hypothesis test:

4. Repeat 'L" (number of realizations) times to calculate the t statistic sampling distribution (empirically), given the H_0: $\mu_1 = \mu_2$ size of the sample sets, n_1 and n_2, and associated dispersion, s_1^2 and s_2^2



t statistic sampling distribution given H_0: $\mu_1 = \mu_2$ is true.

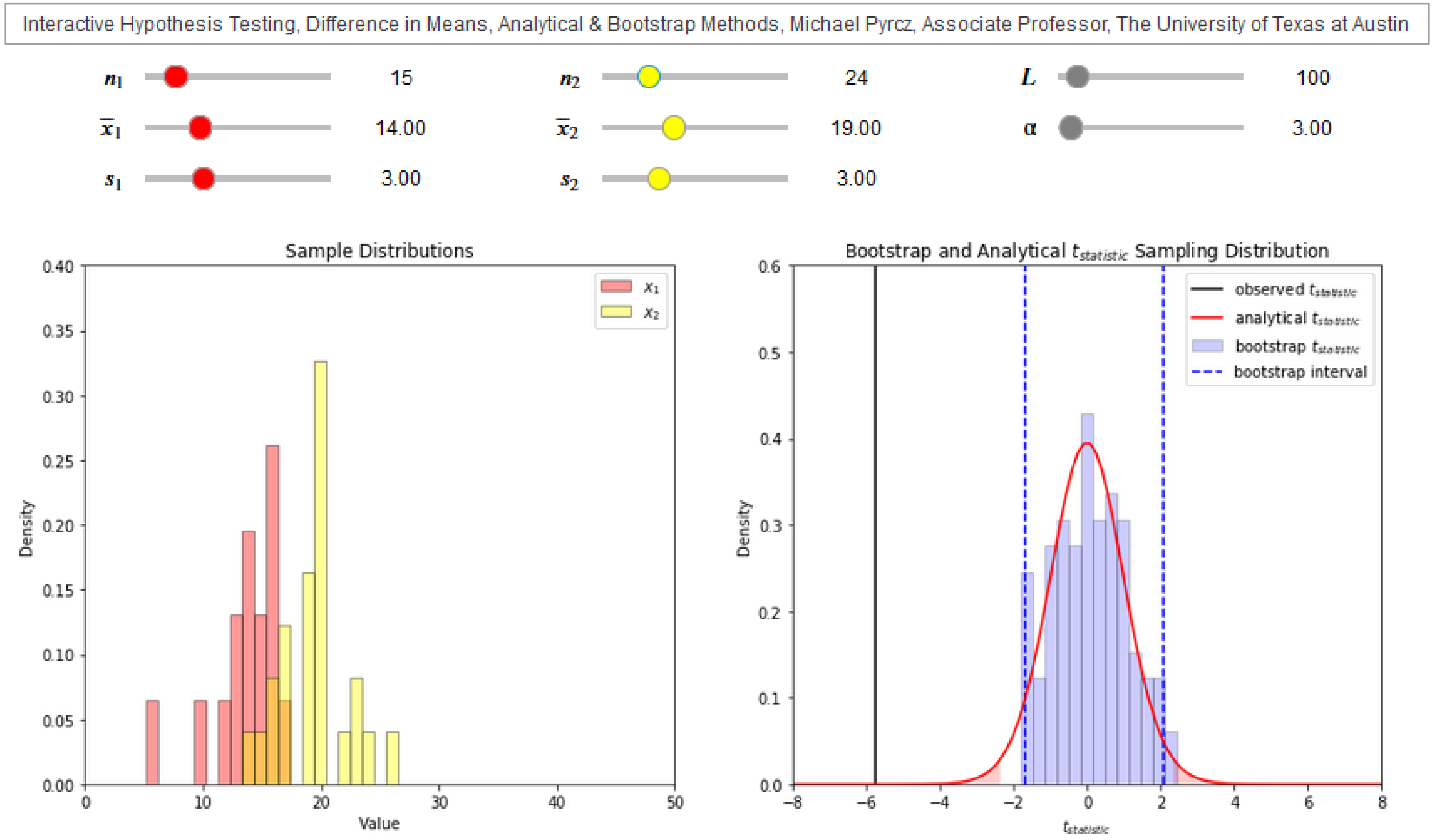
4. Calculate empirical bounds, percentiles, $P(\alpha/2)$ and $P(1-\alpha/2)$

5. Calculate the observed effect, t statistic from the original data with original means and compare with the empirical bounds.

Hypothesis Testing
Take 3!

Example #1 Bootstrap Hypothesis Test for Difference in Means
The Problem:

- Interactive demonstration
- Specify the sample distributions
- Monte Carlo simulate the samples
- Perform analytical and bootstrap hypothesis tests



2 sample sets. Bootstrap and analytics sampling distributions and observed effect.

Jupyter notebook Python interactive demonstration 'Interactive_Hypothesis_Testing.ipynb'.

PGE 338 Data Analytics and Geostatistics

Lecture 7b: Confidence Intervals and Hypothesis Testing

Lecture outline . . .

- Examples

Michael Pyrcz, The University of Texas at Austin

Confidence Intervals / Hypothesis Testing in Python

Walk Through:

Confidence Intervals and Hypothesis Testing in Python demo.

Data Analytics

Confidence Intervals and Hypothesis Testing in Python in Python

Michael Pyrcz, Associate Professor, The University of Texas at Austin

Twitter | ***GitHub*** | ***Website*** | ***GoogleScholar*** | ***Book*** | ***YouTube*** | ***LinkedIn***

This is a tutorial / demonstration of **Confidence Intervals and Hypothesis Testing in Python**. In Python, the SciPy package, specifically the Stats functions (https://docs.scipy.org/doc/scipy/reference/stats.html) provide excellent tools for efficient use of statistics.

I have previously provided these examples worked out by-hand in Excel (https://github.com/GeostatsGuy/LectureExercises/blob/master/Lecture7_CI_Hypoth_eg_R.xlsx) and also in R (https://github.com/GeostatsGuy/LectureExercises/blob/master/Lecture7_CI_Hypoth_eg.R). In all cases, I use the same dataset available as a comma delimited file (https://git.io/fxLAt).

This tutorial includes basic, typical confidence interval and hypothesis testing methods that would commonly be required for Engineers and Geoscientists including:

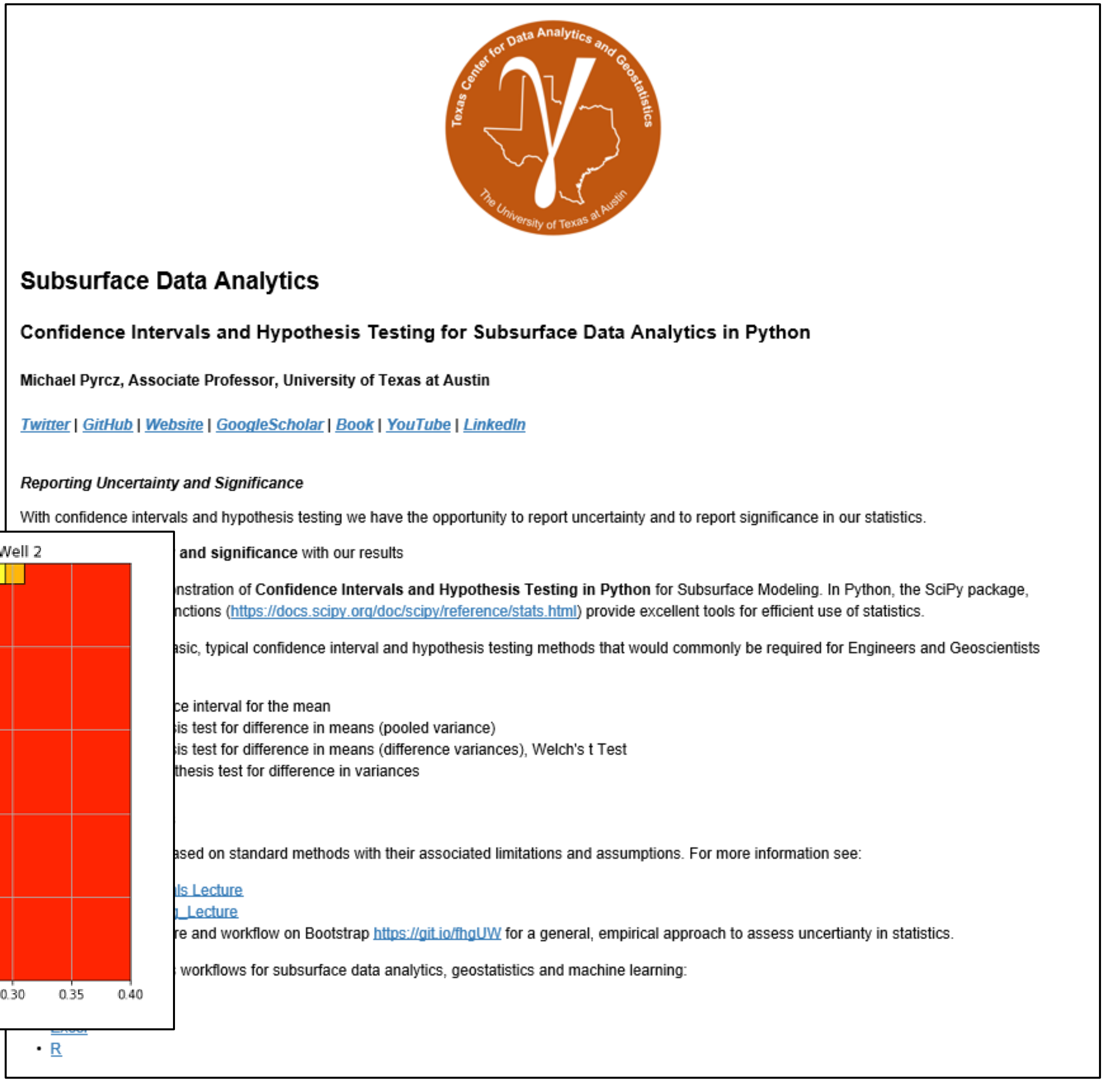
1. Student-t confidence interval for the mean
2. Student-t hypothesis test for difference in means (pooled variance)
3. Student-t hypothesis test for difference in means (difference variances), Welch's t Test
4. F-distribution hypothesis test for difference in variances

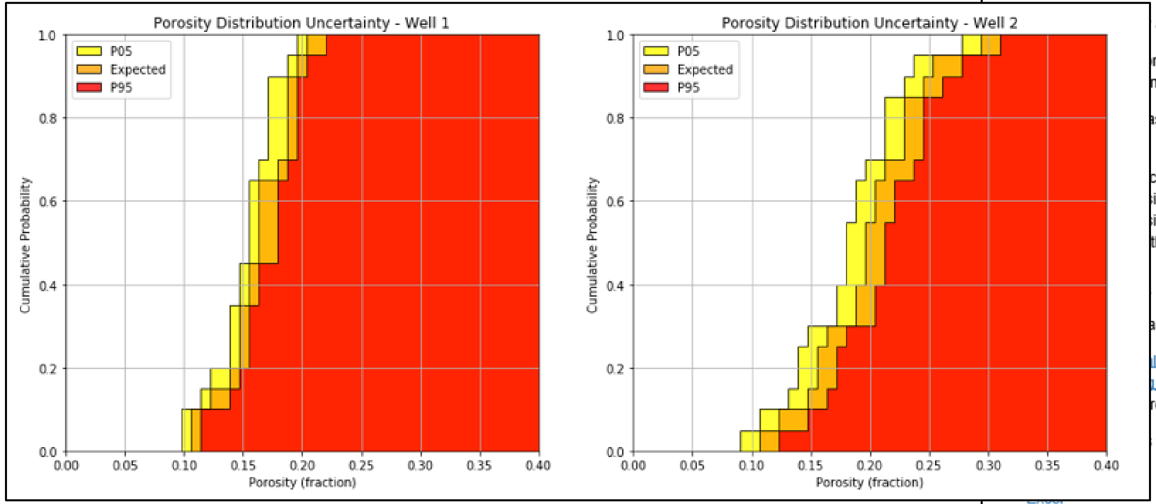
Confidence intervals and hypothesis testing in Python with file PythonDataBasics_ConfidenceInterval_HypothesisTesting.ipynb.

Confidence Intervals / Hypothesis Testing in Python

Walk Through:

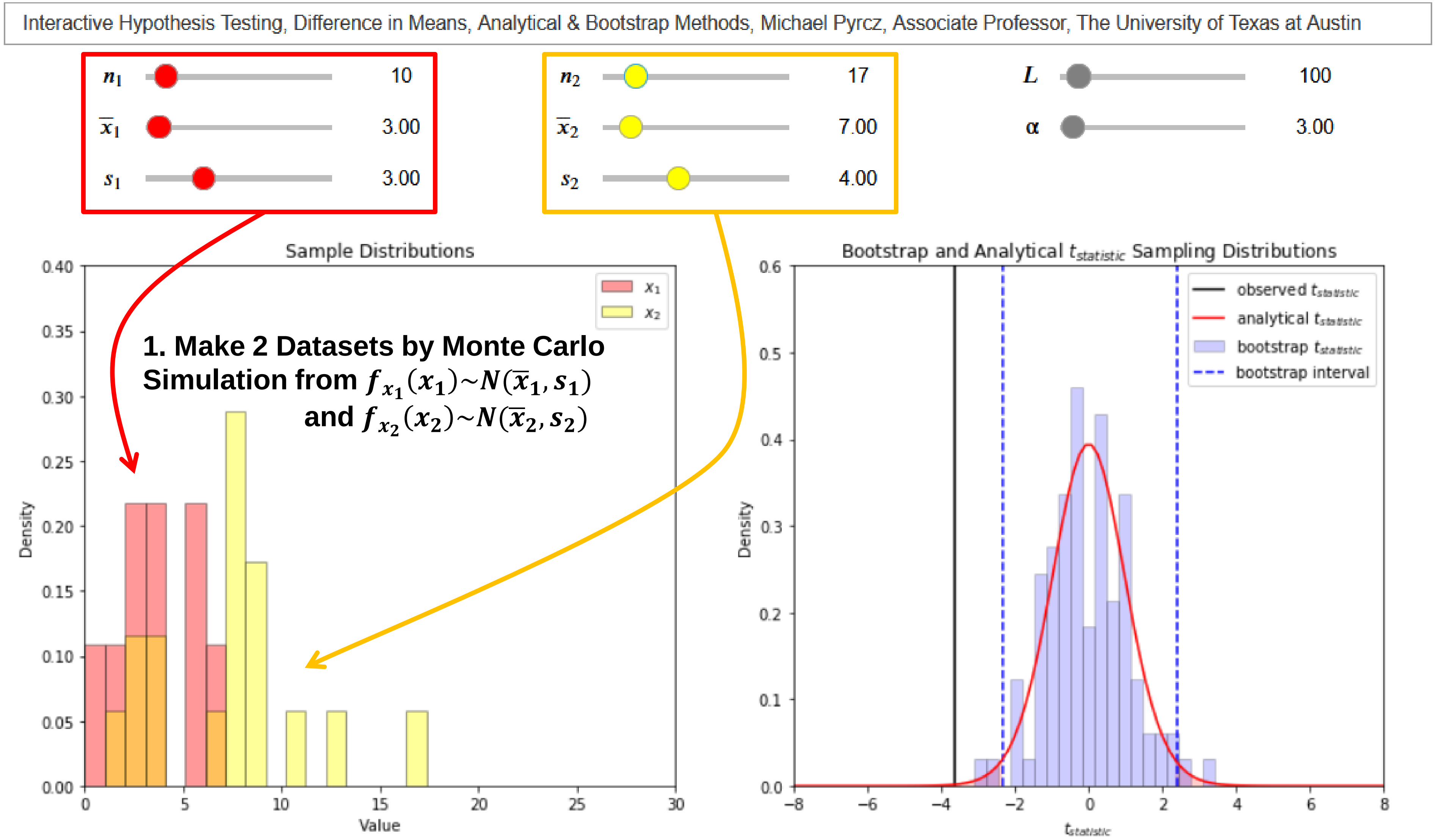
Confidence Intervals and Hypothesis Testing in Python demo in file:





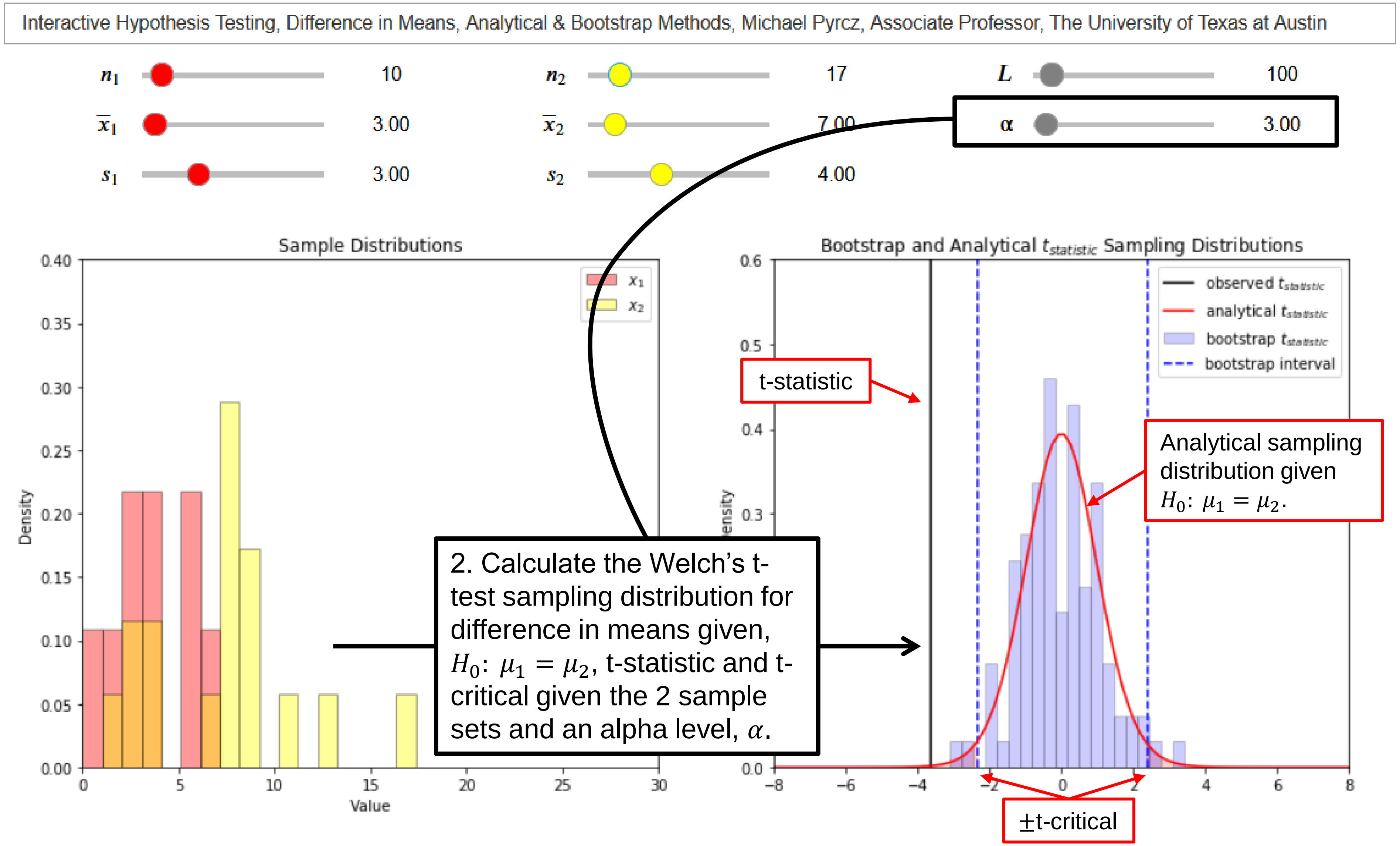
Confidence intervals and hypothesis testing in Python with file Subsurface_MachineLearning_Confidence_Hypothesis.ipynb

Confidence Intervals / Hypothesis Testing in Python



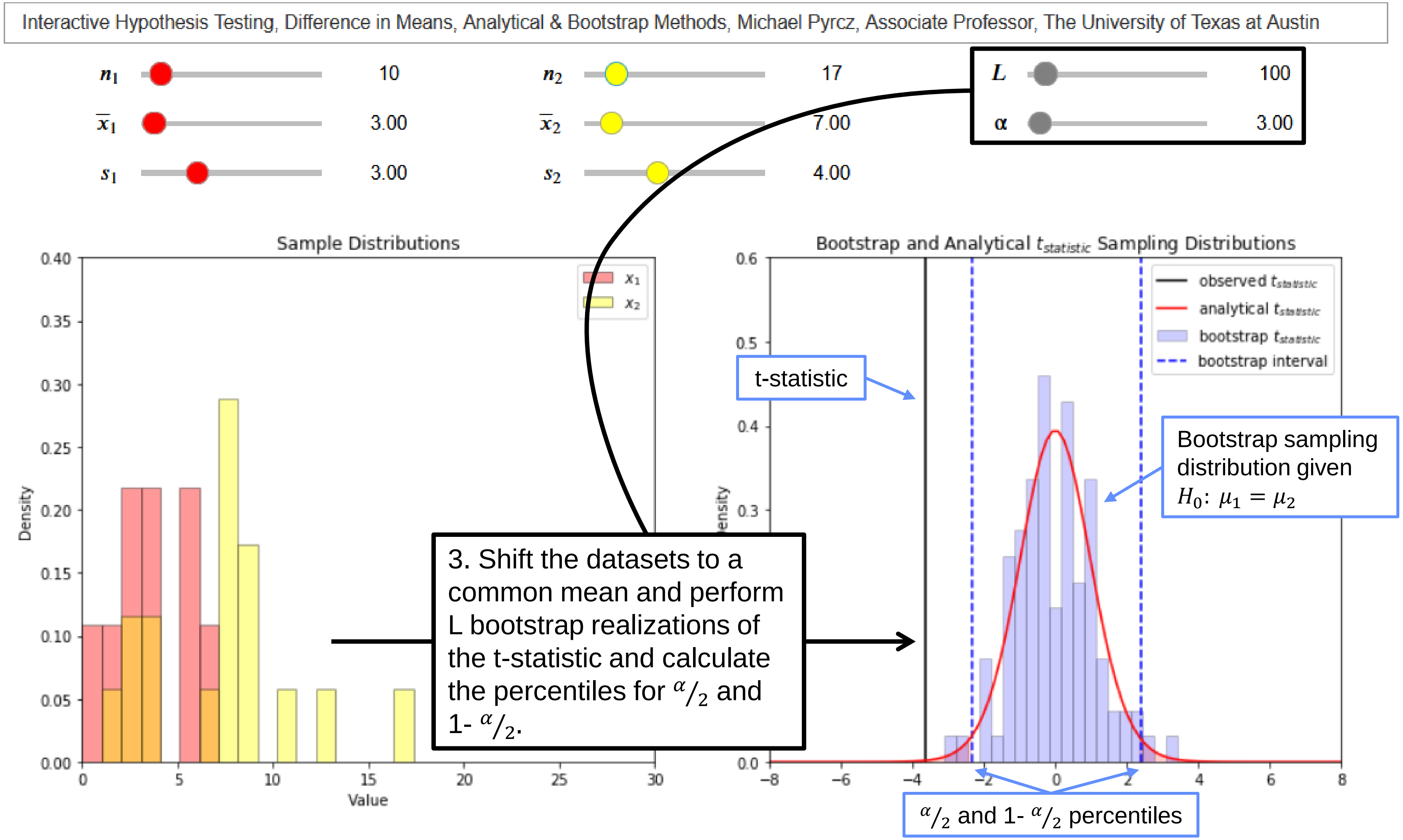
Interactive_Hypothesis_Testing.ipynb @ https://git.io/JU2as on GitHub/GeostatsGuy

Confidence Intervals / Hypothesis Testing in Python



Interactive_Hypothesis_Testing.ipynb @ https://git.io/JU2as on GitHub/GeostatsGuy

Confidence Intervals / Hypothesis Testing in Python

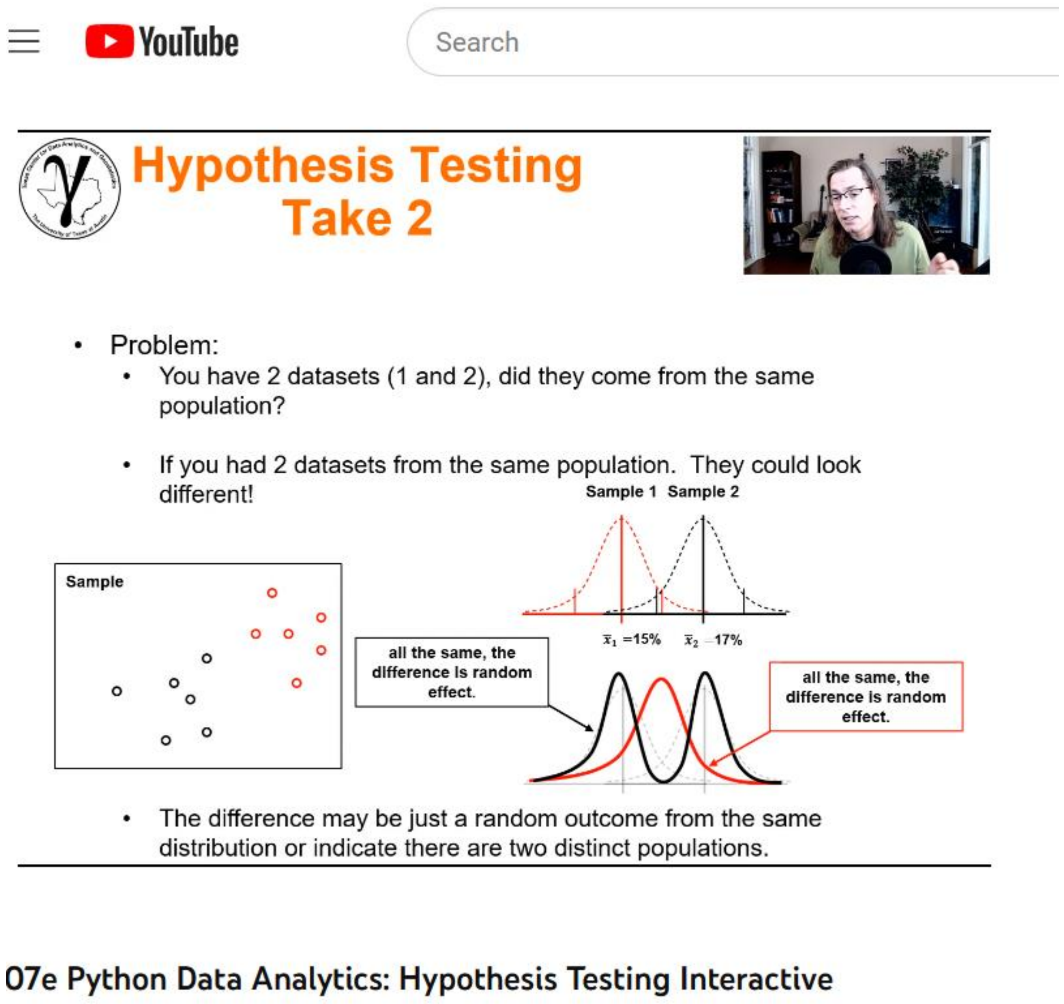


Interactive_Hypothesis_Testing.ipynb @ https://git.io/JU2as on GitHub/GeostatsGuy

Confidence Intervals / Hypothesis Testing in Python

Experiential Learning, Evaluate the Impact of Each of These:

- Sample Sizes for dataset 1 n_1, and 2, n_2.
- Sample Means for dataset 1, $\bar{x}_1$, and 2, $\bar{x}_2$.
- Sample Standard Deviations for dataset 1 s_1, and 2, s_2.
- Alpha level, α
- Number of bootstrap realizations, L



I recorded a walkthrough of this interactive demonstration, https://youtu.be/bcb3m3LBtRk.

PGE 338 Data Analytics and Geostatistics

Lecture 7b: Confidence Intervals and Hypothesis Testing

Lecture outline . . .

- Concepts
- Analytical Hypothesis Testing
- Bootstrap Hypothesis Testing
- Examples

Michael Pyrcz, The University of Texas at Austin