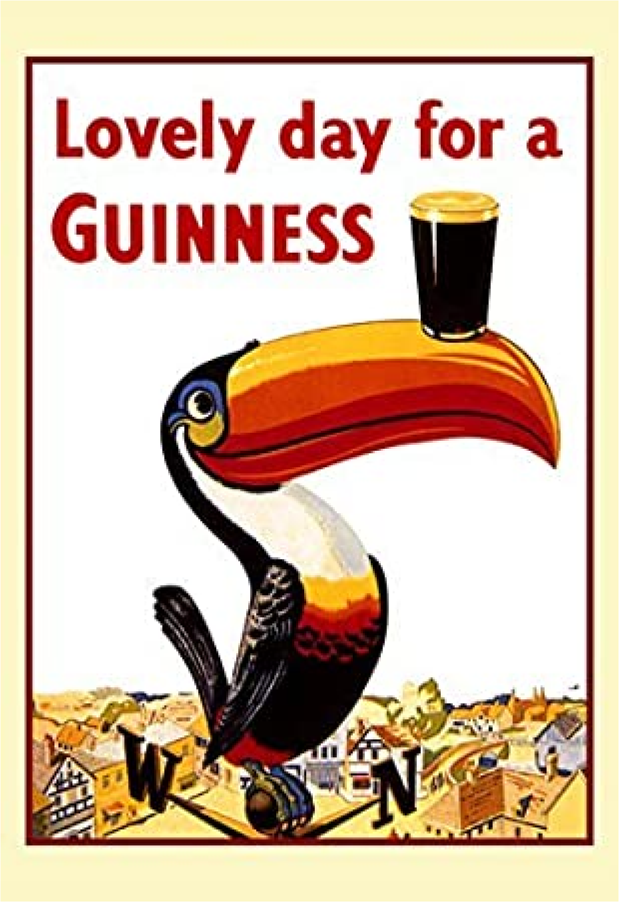
ENGR 298: Engineering Analysis and Decision Making – Statistical Tests

Dr. Jason Forsyth
Department of Engineering
James Madison University

T-Tests: How to know if your samples (or Guinness) is sufficient



Student’s T-Test

- Generally used to perform comparison of means (μ) between one or more *sampled* distribution and *known* distribution.

- Comparison can determine whether means are ==, $\neq$, <, or >. Assumption is usually == and a hypothesis test confirms or proves otherwise.

- When a sample is compared against a fixed/known value is it a 1-sample test. When two samples are compared it is 2-sample. In 2-sample case the samples can be ‘paired’ or not.

T-Test as Hypothesis Testing

- Recall from statistics, when conducting hypothesis testing, will determine a p-value for the relevant test and an alpha (α==0.05).

- Comparison is between a null hypothesis (H_0) and an alternative hypothesis (H_1). H_0 is assumed to be true unless proven otherwise.

- If p-value < α: **reject** the null hypothesis (H_0) and **accept** alternative hypothesis (H_1).

- Else: cannot reject null. Must remain with null hypothesis (H_0). More evidence may be required to overturn H_0.

Lots of good examples online and in textbooks. Follow to confirm process and evidence.

https://online.stat.psu.edu/statprogram/reviews/statistical-concepts/hypothesis-testing/p-value-approach

Brinell Hardness Scores

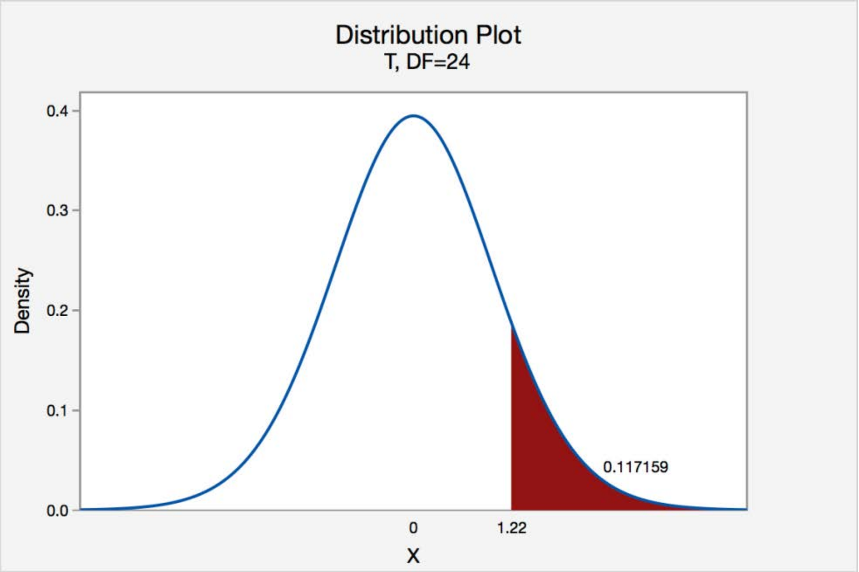
An engineer measured the Brinell hardness of 25 pieces of ductile iron that were subcritically annealed. The resulting data were:

170	167	174	179	179	187	179	183	179
156	163	156	187	156	167	156	174	170
183	179	174	179	170	159	187		

The engineer hypothesized that the mean Brinell hardness of *all* such ductile iron pieces is greater than 170. Therefore, he was interested in testing the hypotheses:

$H_0 : \mu = 170$
$H_A : \mu > 170$

- Test assumes that the mean is 170. Wish to perform test to prove/disprove that assertion
- T-test will be 1-sample (since is comparison against known mean) and will be 'right sided' / left-tailed to the statistic.



https://online.stat.psu.edu/statprogram/reviews/statistical-concepts/hypothesis-testing/examples

Some general statistics… What do you think the hypothesis test result will be?

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Lower Bound
25	172.52	10.31	2.06	168.99

$H_0 : \mu = 170$

$H_A : \mu > 170$

scipy.stats.ttest_1samp

`scipy.stats.ttest_1samp(a, popmean, axis=0, nan_policy='propagate', alternative='two-sided')` [source]

Calculate the T-test for the mean of ONE group of scores.

This is a test for the null hypothesis that the expected value (mean) of a sample of independent observations *a* is equal to the given population mean, *popmean*.

Parameters:

a : *array_like*

Sample observation.

popmean : *float or array_like*

Expected value in null hypothesis. If array_like, then it must have the same shape as *a* excluding the axis dimension.

Returns:

statistic : *float or array*

t-statistic.

pvalue : *float or array*

Two-sided p-value.

Determine test-side for 1-sample test

alternative : *{'two-sided', 'less', 'greater'}, optional*

Defines the alternative hypothesis. The following options are available (default is 'two-sided'):

$H_0 = H_1$

- 'two-sided': the mean of the underlying distribution of the sample is different than the given population mean (*popmean*)

$H_0 > H_1$

- 'less': the mean of the underlying distribution of the sample is less than the given population mean (*popmean*)

$H_0 < H_1$

- 'greater': the mean of the underlying distribution of the sample is greater than the given population mean (*popmean*)

New in version 1.6.0.

```python
# build python example based off: https://online.stat.psu.edu/statprogram/reviews/statistical-concepts/hypothesis-testing/examples
hardness = [170, 167, 174, 179, 179, 187, 179, 183, 179, 156, 163, 156, 187, 156, 167, 156, 174, 170, 183, 179, 174,
            179, 170, 159, 187]

# convert to numpy array
hardness = np.asarray(hardness)

# write down our expected values
expected_mean = 170

# establish a level of significance
alpha = 0.05

# hypothesis H0: mean hardness is 170
# alternative H1: mean hardness is > 170
(stat, p_value) = ttest_1samp(hardness, popmean=expected_mean, alternative='greater')

# based upon the results of our t-test, determine whether means are equal or not....
if p_value < alpha:
    print('Reject H0: sample means and population means are not equal!')

else:
    print('Accept H0: sample means and population means are equivalent')
```

$H_0 : \mu = 170$

$H_A : \mu > 170$

P-value=0.116

Material hardness is not statistically greater than 170.

Height of Sunflowers

A biologist was interested in determining whether sunflower seedlings treated with an extract from *Vinca minor* roots resulted in a lower average height of sunflower seedlings than the standard height of 15.7 cm. The biologist treated a random sample of $n = 33$ seedlings with the extract and subsequently obtained the following heights:

11.5	11.8	15.7	16.1	14.1	10.5	9.3	15.0	11.1
15.2	19.0	12.8	12.4	19.2	13.5	12.2	13.3	
16.5	13.5	14.4	16.7	10.9	13.0	10.3	15.8	
15.1	17.1	13.3	12.4	8.5	14.3	12.9	13.5	

The biologist's hypotheses are:

$$H_0 : \mu = 15.7$$
$$H_A : \mu < 15.7$$

Descriptive Statistics

N	Mean	StDev	SE Mean	95% Upper Bound
33	13.664	2.544	0.443	14.414

```python
# build python example based off: https://online.stat.psu.edu/statprogram/reviews/statistical-concepts/hypothesis-testing/examples
heights = [11.5, 11.8, 15.7, 16.1, 14.1, 10.5, 9.3, 15.0, 11.1, 15.2, 19.0, 12.8, 12.4, 19.2, 13.5, 12.2, 13.3, 16.5,
           13.5, 14.4, 16.7, 10.9, 13.0, 10.3, 15.8, 15.1, 17.1, 13.3, 12.4, 8.5, 14.3, 12.9, 13.5]

# convert to numpy array
heights = np.asarray(heights)


# write down our expected values
expected_mean = 15.7

# establish a level of significance
alpha = 0.05

# hypothesis H0: mean hardness is 15.7
# alternative H1: mean hardness is < 15.7
(stat, p_value) = ttest_1samp(heights, popmean=expected_mean, alternative='less')

# based upon the results of our t-test, determine whether means are equal or not....
if p_value < alpha:
    print('Reject H0: sample means and population means are not equal!')

else:
    print('Accept H0: sample means and population means are equivalent')
```

The biologist's hypotheses are:

$$H_0 : \mu = 15.7$$
$$H_A : \mu < 15.7$$

P-value=3.17E-05

Plant heights are smaller than 15.7cm at a statistically significant level.

Gum Thickness

A manufacturer claims that the thickness of the spearmint gum it produces is 7.5 one-hundredths of an inch. A quality control specialist regularly checks this claim. On one production run, he took a random sample of n = 10 pieces of gum and measured their thickness. He obtained:

7.65	7.60	7.65	7.70	7.55
7.55	7.40	7.40	7.50	7.50

The quality control specialist's hypotheses are:

$$H_0 : \mu = 7.5$$
$$H_A : \mu \neq 7.5$$

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
10	7.550	0.1027	0.0325	(7.4765, 7.6235)

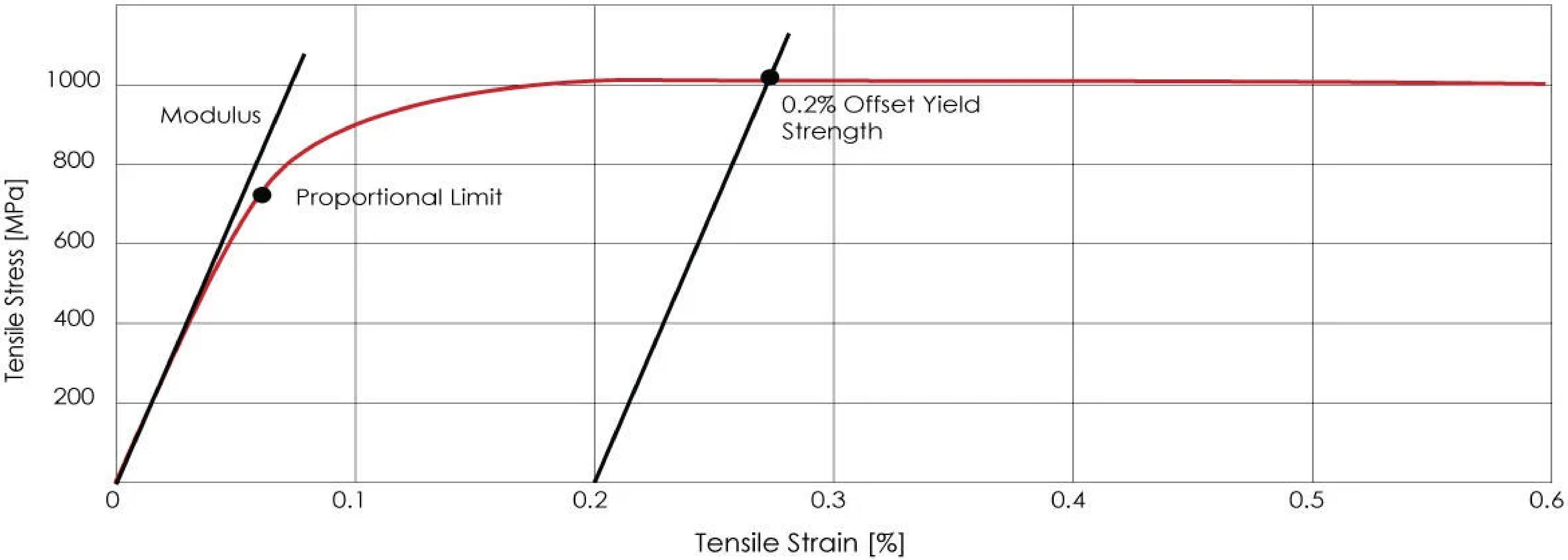
```python
# load sample thickness
thickness = [7.65, 7.60, 7.65, 7.70, 7.55, 7.55, 7.40, 7.40, 7.50, 7.50]

# convert to numpy array
thickness = np.asarray(thickness)


# write down our expected values
expected_mean = 7.5

# establish a level of significance
alpha = 0.05

# hypothesis H0: mean thickness is 7.5
# alternative H1: mean hardness is not 7.5
(stat, p_value) = ttest_1samp(thickness, popmean=expected_mean, alternative='two-sided')

# based upon the results of our t-test, determine whether means are equal or not....
if p_value < alpha:
    print('Reject H0: sample means and population means are not equal!')

else:
    print('Accept H0: sample means and population means are equivalent')
```

$H_0 : \mu = 7.5$

$H_A : \mu \neq 7.5$

P-value=0.158

Applications for Tensile Data



https://www.instron.com/en-us/our-company/library/glossary/o/offset-yield-strength

If a construction job requires minimum yield strength values of 40 MPa, and your XYZ material has an average yield strength of 50 MPa with a standard deviation of +/- 2 MPa, then are you able to assure that 95% of your XYZ materials meet the minimum specified requirement by the owner or agency?

If not, then what minimum yield strength does your XYZ material meet at 95% acceptance rates?

From the 1045CR data, can we show that the yield strength is at least 500 MPa?

- Load tensile data from results CSV; Pull 1045CR data using Pandas; Perform 1-sided T-test against sample population
  - H_0: mean yield strength is 500 MPa
  - H_1: mean yield strength is > 500 MPa

Sample Statistics	Value
Mean	606.03 MPa
Std. Dev.	+/- 17.2 MPa
95% CI on Mean	(571.68, 640.38)

Do you believe the test should pass? What happens if the limit (500) is adjusted? What would happen at 600?

Comparison for various values

Sample Statistics	Value
Mean	606.03 MPa
Std. Dev.	+/- 17.2 MPa
95% CI on Mean	(571.68, 640.38)

Statistical Test	P-value	Inference
H1 > 500	1.88e-06	Average yield strength is > 500
H1 > 550	7.45e-05	Average yield strength is > 550
H1 > 600	0.198	Average yield strength is 600
H1 > 650	0.999	Average yield strength is 600

2-Sample Tests: When Comparing for different sample populations.

1.3.5.3. Two-Sample t-Test for Equal Means

Purpose: Test if two population means are equal

The two-sample t-test (Snedecor and Cochran, 1989) is used to determine if two population means are equal. A common application is to test if a new process or treatment is superior to a current process or treatment.

There are several variations on this test.

1. The data may either be paired or not paired. By paired, we mean that there is a one-to-one correspondence between the values in the two samples. That is, if $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are the two samples, then X_i corresponds to Y_i. For paired samples, the difference $X_i - Y_i$ is usually calculated. For unpaired samples, the sample sizes for the two samples may or may not be equal. The formulas for paired data are somewhat simpler than the formulas for unpaired data.

2. The variances of the two samples may be assumed to be equal or unequal. Equal variances yields somewhat simpler formulas, although with computers this is no longer a significant issue.

3. In some applications, you may want to adopt a new process or treatment only if it exceeds the current treatment by some threshold. In this case, we can state the null hypothesis in the form that the difference between the two populations means is equal to some constant $\mu_1 - \mu_2 = d_0$ where the constant is the desired threshold.

https://www.itl.nist.gov/div898/handbook/eda/section3/eda353.htm

2-Sample Tests

- When comparing two independent populations, test is 2-samples (or independent) t-test.
- Hypothesis testing is the same with p-values
- Scipy functions take two arrays rather than singular array and fixed value.
- Generally used to show statistical differences between population means (e.g. is one group taller…etc than another)

```
scipy.stats.ttest_ind(a, b, axis=0, equal_var=True, nan_policy='propagate',
permutations=None, random_state=None, alternative='two-sided', trim=0)
```

[source]

Calculate the T-test for the means of *two independent* samples of scores.

This is a test for the null hypothesis that 2 independent samples have identical average (expected) values. This test assumes that the populations have identical variances by default.

Parameters:

a, b : *array_like*

The arrays must have the same shape, except in the dimension corresponding to *axis* (the first, by default).

axis : *int or None, optional*

Axis along which to compute test. If None, compute over the whole arrays, *a*, and *b*.

equal_var : *bool, optional*

If True (default), perform a standard independent 2 sample test that assumes equal population variances [1]. If False, perform Welch's t-test, which does not assume equal population variance [2].

New in version 0.11.0.

Returns:

statistic : *float or array*

t-statistic.

pvalue : *float or array*

Two-sided p-value.

Determine test-side for 2-sample test

alternative : *{'two-sided', 'less', 'greater'}, optional*

Defines the alternative hypothesis. The following options are available (default is 'two-sided'):

$H_0 = H_1$

- 'two-sided': the means of the distributions underlying the samples are unequal.

$H_0 < H_1$

- 'less': the mean of the distribution underlying the first sample is less than the mean of the distribution underlying the second sample.

$H_0 > H_1$

- 'greater': the mean of the distribution underlying the first sample is greater than the mean of the distribution underlying the second sample.

New in version 1.6.0.

Rejection of $H_0 = H_1$ implies either $H_0 < H_1$ or $H_0 > H_1$. Can just assert > or < at that point…

Perform a simple test first that should be equal

```
# parameters for distribution 1
dist1_mean = 0
dist1_std = 1

# take samples from distribution
dist1 = np.random.normal(loc=dist1_mean, scale=dist1_std, size=num_samples)

# parameters for distribution 2
dist2_mean = 0
dist2_std = 1

# take samples from distribution
dist2 = np.random.normal(loc=dist2_mean, scale=dist2_std, size=num_samples)

# hypothesis H0: means are equivalent between population samples
# alternative H1: means are NOT equivalent between population samples
(stat, p_value) = ttest_ind(dist1, dist2, alternative='two-sided')
```

Results in p-value=0.781. Clearly cannot reject the null hypothesis.

Perform a simple test first that should be unequal

```
# parameters for distribution 1
dist1_mean = 0
dist1_std = 1

# take samples from distribution
dist1 = np.random.normal(loc=dist1_mean, scale=dist1_std, size=num_samples)

# parameters for distribution 2
dist2_mean = 5
dist2_std = 3

# take samples from distribution
dist2 = np.random.normal(loc=dist2_mean, scale=dist2_std, size=num_samples)

# hypothesis H0: means are equivalent between population samples
# alternative H1: means are NOT equivalent between population samples
(stat, p_value) = ttest_ind(dist1, dist2, alternative='two-sided')
```

Results in p-value=5.019e-41. Must reject the null hypothesis as distributions are not equal.

Summary

- T-test is effective method to compare sample/population *means*. Useful in determining whether a treatment has impacted a population.

- All results vary with number of samples and selected significance level. Easy to change results based upon these ‘custom’ values.

- With great power comes great responsibility; easy to misuse and misapply statistics. Start with simple examples to build up intuition.
  - Personally avoided χ^2 analysis because could not validate results. Reach out to experts to confirm values and to get practice.