An Introduction to t-Tests

Contents

1 Definition

- **Motivation**: T-test is a statistical test that is used to compare the means of two groups.
- **Application**: often used in hypothesis testing to test whether the two means are the same. Mathematically, it establishes the problem by assuming that the means of the two distributions are equal (H_0: $\mu_1=\mu_2$). If the t-test rejects the null hypothesis (H_0: $\mu_1=\mu_2$), it indicates that the groups are highly probably different. **t-value**: larger score indicates that the groups are different while a small t-score indicates that the groups are similar.
- **Meanings of t-value and p-value**: The less the distributions of values in the two data sets overlap, the larger the *t* value will tend to be. We can then estimate the probability that the observed difference occurred simply by chance, rather than due to a true difference — this is the *p-value*.
- **p-values**: Computed based on the t-value and the degree freedoms over t-Distribution table. reflect the probability of having sufficient proof to negate the indifference between the mean of the two samples. since the p-value is just a value, we need to compare it with the critical value (α):
  p_value > α (Critical value): Fail to reject the null hypothesis of the statistical test.
  p_value ≤ α (Critical value): Reject the null hypothesis of the statistical test.
  In general, the critical value is $\alpha = 0.05$.
- **Note**: T-test only works for two groups. For more than 2 groups, use an ANOVA test or a post-hoc test. Furthermore, it works when the number of samples is around 30.

2 T-test example

- **Example**: null hypothesis H_0 and alternate hypothesis H_a

  ***t* test example**

  You want to know whether the mean petal length of iris flowers differs according to their species. You find two different species of irises growing in a garden and measure 25 petals of each species. You can test the difference between these two groups using a *t* test and null and alterative hypotheses.

  - The null hypothesis (H_0) is that the true difference between these group means is zero.
  - The alternate hypothesis (H_a) is that the true difference is different from zero.

- **Assumptions**: t-test is a parametric test (regression, comparison, or correlation). The data is assumed to satisfy the following properties.

  1. are independent
  2. are (approximately) normally distributed
  3. have a similar amount of variance within each group being compared (a.k.a. homogeneity of variance)

3 How to choose t-test and their formulation

- One-sample, two-sample, or paired t test?

- **One-sample**: compare one group with a true value. The formulation is given below. The degree freedom is $n-1$.

one-sample t-tests

$$t = \frac{\overline{X} - u}{s/\sqrt{n}}$$

1. Find the mean
2. Find the variance/standard deviation (subtract the mean from each score, square that number, sum them and divide by n-1)
3. Find the square root of n
4. *u* is given
5. Plug them into the formula
6. df = n - 1

- **Two-sample t test (independent T-test)**: two groups from two independent groups. The formulation is shown as (s_1^2 and s_2^2 are obtained by dividing $(n-1)$.) The degree freedom is (n_1+n_2-2)

t-test independent sample

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{diff}}$$

1. Find the means for each sample.
2. Find the variance/standard deviation for each sample
3. Find the pooled standard deviation:

$$s_p = \sqrt{\frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{(n_1 - 1) + (n_2 - 1)}}$$

4. Find the standard error of the difference:

$$s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = s_{diff}$$

5. Plug into the formula
6. df = $n_1 + n_2 - 2$

- **paired t test**: compare two groups from the same distribution.

t-test for dependent samples (pre-test and post-test)

$$t = \frac{\overline{X}_{diff}}{s_D/\sqrt{n}}$$

1. Find the mean for the difference ($\overline{X}_{diff}$) between first score and second score (subtract second score from first score and then find the mean in the usual way)
2. Find the variance/standard deviation for the differences (subtract the mean difference from each difference, square that number, sum them, and divide by n (# of pairs) - 1
3. Plug these into the formula
4. df = n (# of pairs) - 1

- **Equal Variance T-Test (also called pooled T-test)**: the test is used when the sample size in each group or population is the same or the variance of the two data sets is similar.
- **Unequal Variance T-Test (also called Welch's test)**: The testing is used when the variance and the number of samples in each group are different.

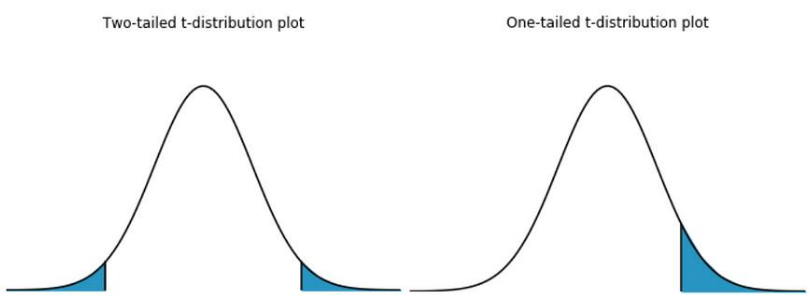
One-sample, two-sample, or paired *t* test?

- If the groups come from a single population (e.g., measuring before and after an experimental treatment), perform a **paired *t* test**. This is a within-subjects design.
- If the groups come from two different populations (e.g., two different species, or people from two separate cities), perform a **two-sample *t* test** (a.k.a. **independent *t* test**). This is a between-subjects design.
- If there is one group being compared against a standard value (e.g., comparing the acidity of a liquid to a neutral pH of 7), perform a **one-sample *t* test**.

- One-tailed or two-tailed t test?

One-tailed or two-tailed *t* test?

- If you only care whether the two populations are different from one another, perform a **two-tailed *t* test**.
- If you want to know whether one population mean is greater than or less than the other, perform a **one-tailed *t* test.**

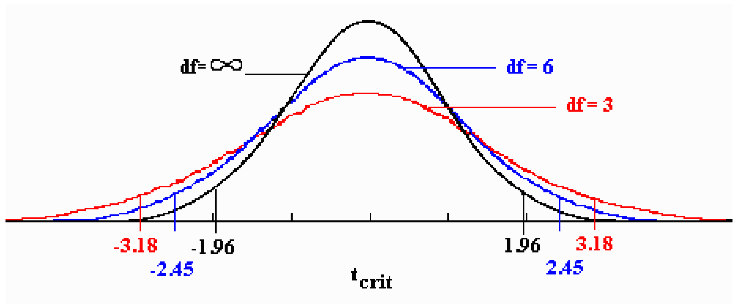


2-tailed *p* values: less to reject the Null hypothesis. The p-value will be divided by 2 such that reflects a probability of 2.5% that the result occurred by chance in one direction.

- For the example, choose the right t-test

> *t* test example
>
> In your test of whether petal length differs by species:
>
> - Your observations come from two separate populations (separate species), so you perform a two-sample *t* test.
> - You don't care about the direction of the difference, only whether there is a difference, so you choose to use a two-tailed *t* test.

- **T-statistic** (pdf): affected by the degree freedom. By increasing the number of samples, the t-static pdf gets closer to the normal distribution.



4 Perform a t-test

- Acquire data sets
- Decide the suitable t-test (1-sample, 2-sample, or paired)
- Compute the t-value
- Find the p-value of the t-value (based on the t-Distribution table with n-1 degrees of freedom.)
- Compare the p-value with the critical **α (Critical value). If $p \leq \alpha$, reject null hypothesis; otherwise, not reject.**
- the p-value means that there is a p% chance of your results occurring if the null hypothesis is true.

5 two types of errors[5]

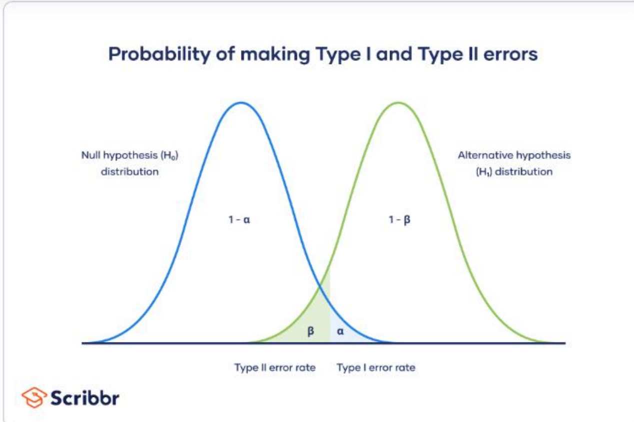
- **Type I error (false positive)**: A Type I error means rejecting the null hypothesis when it's actually true, called false positive. If you value your null hypothesis more, use a smaller critical value α. The Type I error rate is α.
- **Type II error (false negative)**: not rejecting the null hypothesis when it's actually false, which is affected by the statistical power to detect an effect of a certain size. The higher the statistical power, the lower the probability of making a Type II error. The statistical power may be affected by many parameters. However, increasing sample size and significance level will reduce Type II error.

Statistical power is determined by:

- **Size of the effect**: Larger effects are more easily detected.
- **Measurement error**: Systematic and random errors in recorded data reduce power.
- **Sample size**: Larger samples reduce sampling error and increase power.
- **Significance level**: Increasing the significance level increases power.

To (indirectly) reduce the risk of a Type II error, you can increase the sample size or the significance level.

- **Tradeoff**: The Type I and Type II error rates influence each other. That's because the significance level α (the Type I error rate) affects statistical power, which is inversely related to the Type II error rate.



It's important to strike a balance between the risks of making Type I and Type II errors. Reducing the alpha always comes at the cost of increasing beta, and vice versa.

- For statisticians, a Type I error is usually worse.

6 t-test by python and Pandas

```
'''
Import data from pandas
'''

data=pd.read_csv('flower.data.csv')
#print(data.head())
#(data.tail())

group1=data[data['Species']=='setosa']
group2=data[data['Species']=='virginica']

results=ttest_ind(group1['Petal.Length'], group2['Petal.Length'])    # By assuming the same variance

print(results)


'''
Example 2: Welch's t-Test in Pandas
 two populations that the samples came from have different variance.
'''
results1=ttest_ind(group1['Petal.Length'], group2['Petal.Length'], equal_var=False)

print(results1)

'''
Example 3: Paired Samples t-Test in Pandas
determine if two population means are equal in which each observation in one sample
 can be paired with an observation in the other sample.

By defualt: return 2-tailed p values.
 1-tailed p value should be used if we have a specific prediction of a "direction" of the difference.
'''

from scipy.stats import ttest_rel    # Paired t-test function
results2=ttest_ind(group1['Petal.Length'], group2['Petal.Length'])    # By assuming the same variance

print(results2)
```

```
In [9]: runfile('D:/Study_in_UVA/PhD_jobs/Preparing/Risk Modeling/Yuanyuan/t-test_model.py', wdir='D:/Study_in_UVA/PhD_jobs/Preparing/Risk Modeling/Yuanyuan')
Ttest_indResult(statistic=-33.719454860157065, pvalue=4.4484174283799283e-35)
Ttest_indResult(statistic=-33.719454860157065, pvalue=1.5855478441588353e-25)

In [10]: runfile('D:/Study_in_UVA/PhD_jobs/Preparing/Risk Modeling/Yuanyuan/t-test_model.py', wdir='D:/Study_in_UVA/PhD_jobs/Preparing/Risk Modeling/Yuanyuan')
Ttest_indResult(statistic=-33.719454860157065, pvalue=4.4484174283799283e-35)
Ttest_indResult(statistic=-33.719454860157065, pvalue=1.5855478441588353e-25)
Ttest_indResult(statistic=-33.719454860157065, pvalue=4.4484174283799283e-35)
```

[1]. An Introduction to t Tests | Definitions, Formula and Examples, https://www.scribbr.com/statistics/t-test/#:~:text=A%20t%20test%20is%20a,are%20different%20from%20one%20another.
[2]. Basic Statistics in Python: t tests with SciPy, https://neuraldatascience.io/5-eda/ttests.html
[3]. One important application, https://www.wallstreetmojo.com/t-test/
[4]. https://speech.pfw.edu/582/t-tests.pdf
[5]. Two types of errors, https://www.scribbr.com/statistics/type-i-and-type-ii-errors/#type-ii-error

An introduction of ANOVA

1 Definition and application

- **Definition of ANOVA (Analysis of Variables)**: used to analyze the difference between the means of more than two groups. The null hypothesis (H0) of ANOVA is that there is no difference among group means. The alternative hypothesis (Ha) is that at least one group differs significantly from the overall mean of the dependent variable.
- **A one-way ANOVA:** uses one independent variable, which is used when the collected data includes one categorical independent variable and one quantitative dependent variable. The independent variable should have at least three levels (i.e. at least three different groups or categories).
- **Example**: Your independent variable is brand of soda, and you collect data on Coke, Pepsi, Sprite, and Fanta to find out if there is a difference in the price per 100ml.
- **How does it work**: ANOVA uses the F test for statistical significance. This allows for comparison of multiple means at once. The F test compares the variance in each group mean from the overall group variance.
- **Shortcomings:** ANOVA will tell you if there are differences among the levels of the independent variable, but not which differences are significant. The post-hoc test can find how the treatment levels differ from one another.
- **The post-hoc test (also called TukeyHSD (Tukey's Honestly-Significant Difference))**: runs **pairwise comparisons** among each of the groups, and uses a conservative error estimate to find the groups which are statistically different from one another.

```
 Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = yield ~ fertilizer, data =
crop.data)

$fertilizer
         diff         lwr       upr     p adj
2-1 0.1761687 -0.19371896 0.5460564 0.4954705
3-1 0.5991256  0.22923789 0.9690133 0.0006125
3-2 0.4229569  0.05306916 0.7928445 0.0208735
```

- **Assumptions of ANOVA**: three conditions

Assumptions of ANOVA

The assumptions of the ANOVA test are the same as the general assumptions for any parametric test:

1. **Independence of observations**: the data were collected using statistically valid sampling methods, and there are no hidden relationships among observations. If your data fail to meet this assumption because you have a confounding variable that you need to control for statistically, use an ANOVA with blocking variables.
2. **Normally-distributed response variable**: The values of the dependent variable follow a normal distribution.
3. **Homogeneity of variance**: The variation within each group being compared is similar for every group. If the variances are different among the groups, then ANOVA probably isn't the right fit for the data.

2 F-test formulation

- **Definition**: The F test compares the variance in each group mean from the overall group variance.

$$\text{SST (stands for } \textit{sum of squares total}) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2.$$

This variability has two sources:

1. Variability between group means (specifically, variation around the overall mean $\bar{x}$)

$$\text{SSG} := \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2, \quad \text{and}$$

2. Variability within groups means (specifically, variation of observations about their group mean $\bar{x}_i$)

$$\text{SSE} := \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2.$$

It is the case that

$$\text{SST} = \text{SSG} + \text{SSE}.$$

- **F-test**: measure the ratio of variation between groups to variation within groups, with group degree freedom k-1 and error freedom n-k within group, where n is the number of samples and k is the number of groups.

Source	SS	df	MS	F
Model/Group	SSG	$k-1$	$\text{MSG} = \frac{SSG}{k-1}$	$\frac{MSG}{MSE}$
Residual/Error	SSE	$n-k$	$\text{MSE} = \frac{SSE}{n-k}$	
Total	SST	$n-1$		

What are these things?

- The *source* (of variability) column tells us SS=Sum of Squares (sum of squared deviations):

  SST measures variation of the data around the overall mean $\bar{x}$
  SSG measures variation of the group means around the overall mean
  SSE measures the variation of each observation around its group mean $\bar{x}_i$

- Degrees of freedom

  $k-1$ for SSG, since it measures the variation of the k group means about the overall mean
  $n-k$ for SSE, since it measures the variation of the n observations about k group means
  $n-1$ for SST, since it measures the variation of all n observations about the overall mean

3 The process of ANOVA

- **Decision**: based on the two degree freedoms $(k-1,\ n-k)$ and the critical level α, find the F value in the F -test table. Then compare the computed f with the value in the table. If $f \geq F$, reject the null hypothesis.
- **Apart from that**: they may also compute the p-value and the R^2.

Null Hypothesis, H_0**:** $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$

Alternative Hypothesis, H_1**:** The means are not equal

Decision Rule: If test statistic > critical value then reject the null hypothesis and conclude that the means of at least two groups are statistically significant.

The steps to perform the one way ANOVA test are given below:

- **Step 1:** Calculate the mean for each group.
- **Step 2:** Calculate the total mean. This is done by adding all the means and dividing it by the total number of means.
- **Step 3:** Calculate the SSB.
- **Step 4:** Calculate the between groups degrees of freedom.
- **Step 5:** Calculate the SSE.
- **Step 6:** Calculate the degrees of freedom of errors.
- **Step 7:** Determine the MSB and the MSE.
- **Step 8:** Find the f test statistic.
- **Step 9:** Using the f table for the specified level of significance, α, find the critical value. This is given by $F(\alpha, df_1. df_2)$.
- **Step 10:** If f > F then reject the null hypothesis.

4 Experience of ANOVA

- **Key**: Apply the f-oneway() function from python scipy.stats to compute the F-val and P-val.
- **Data**: import four sets of values into pandas. dataframe

	A	B	C	D
0	25	45	30	54
1	30	55	29	60
2	28	29	33	51
3	36	56	37	62
4	29	40	27	73

- **Apply the f-oneway() function**: with 4 sets of inputs

```
from scipy.stats import f_oneway
fval, pval=f_oneway(data_F['A'],data_F['B'],data_F['C'],data_F['D'] )
print('f=',fval, ', P-val=',pval)
```

- **Output**: F-val and P-val

```
f= 17.492810457516338 , P-val= 2.639241146210922e-05
```

- **Decision**: Based on the p-val, reject the null hypothesis. Note that F value is inversely related to p value and higher F value (greater than F critical value) indicates a significant p-value.

5 Post-hoc

- **Motivation**: ANOVA shows that the treatment differences are statistically significant, but ANOVA does not tell which treatments are significantly different from each other.
- **Solution**: perform multiple pairwise comparison (post hoc comparison) analysis by Tukey's HSD to analyze the difference for specific categories. Tukey's HSD test accounts for multiple comparisons and corrects for family-wise error rate (FWER) (inflated type I error)
- **HSD formulation**: either use the following HSD

Tukey's HSD (When equal sample size in each group),

$$HSD = q_{A,\alpha,dof}\sqrt{\frac{MS_E}{n}}$$

Tukey-Kramer method (When unequal sample size in each group),

$$HSD = q_{A,\alpha,dof}\sqrt{\frac{MS_E}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

Where,
$q_{A,\alpha,dof}$ = studentized range statistic with A number of groups, α significance level (0.05 or 0.01), and dof degrees of freedom
MS_E = mean square error from ANOVA
n = sample size in each group (when the sample size is equal in two comparing groups)
n_i, n_j = sample size in group i and j (when the sample size is unequal in two comparing groups)

Alternatively, Scheffe's method is completely coherent with ANOVA and considered as more appropriate post hoc test for significant ANOVA for all unplanned comparisons. However, it is highly conservative than other post hoc tests.

- **Example**: the previous case with 4 treatments A, B, C, and D.
Computed by post-hoc from the scikit-posthocs package.

```
import statsmodels.api as sa
import statsmodels.formula.api as sfa
import scikit_posthocs as sp

#re=sp.posthoc_ttest(df_melt, val_col='value', group_col='treatments', p_adjust='holm')  # With two input arrays
re=sp.posthoc_tukey(df_melt, val_col='value', group_col='treatments')
print('tukey_hsd=',re)
```

```
-----------------------Post-hoc Start---------------------------
tukey_hsd=          A         B         C         D
A  1.00000  0.025070  0.900000  0.001000
B  0.02507  1.000000  0.048178  0.029578
C  0.90000  0.048178  1.000000  0.001000
D  0.00100  0.029578  0.001000  1.000000
```

Computed by tukey_hsd() from bioinfokit.

```
# output
  group1 group2  Diff      Lower      Upper   q-value   p-value
0      A      B  15.4   1.692871  29.107129  4.546156  0.025070
1      A      C   1.6 -12.107129  15.307129  0.472328  0.900000
2      A      D  30.4  16.692871  44.107129  8.974231  0.001000
3      B      C  13.8   0.092871  27.507129  4.073828  0.048178
4      B      D  15.0   1.292871  28.707129  4.428074  0.029578
5      C      D  28.8  15.092871  42.507129  8.501903  0.001000
```

Based on the observations, different hsc functions may generate different post-hoc values. However, their pairwise relationship is very similar.

[1]. One-way ANOVA | When and How to Use It (With Examples), https://www.scribbr.com/statistics/one-way-anova/

[2]. https://sites.calvin.edu/scofield/courses/m143/materials/handouts/anova1And2.pdf

[3]. ANOVA with python and cooperate with post-hoc, https://www.reneshbedre.com/blog/anova.html

[4]. Post-hoc by scikit-posthocs, https://scikit-posthocs.readthedocs.io/en/latest/tutorial.html#parametric-anova-with-post-hoc-tests

T-test over linear regression

1 Introduction

- **Motivation**: Linear regression is an important model for prediction and inference between the input and the output. Furthermore, linear regression model is an important model to analyze the contributions of some features, presented by the corresponding coefficients. For some feature, if the corresponding coefficient of some feature is zero, the feature is considered to be no influence on the output. To estimate each coefficient, we consider the t-test over the linear regression.

2 Linear regression model

- **Data**: Given a dataset D with n samples, i.e., $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, where each input $x_i \in R^p$ consists of p features.
- **Model**: the linear regression model is

$$Y = X\theta + e$$

Where $E[e] = 0$ and $var(e) = \sigma^2 I$. (Note that σ^2 is not known in real application. We will estimate σ^2 later.)
Then the output Y is $Y \sim N(X\theta, \sigma^2 I)$. By using the least square method, we have

$$L(\theta) = e^2 = \| y - X\theta \|^2$$

$$\overset{LS}{\Longrightarrow} \quad \frac{\partial L(\theta)}{\partial \theta} = -2X^T(y - X\theta) = 0$$

$$\Rightarrow \quad X^T y = X^T X \hat{\theta}$$

$$\Rightarrow \quad \hat{\theta} = (X^T X)^{-1} X^T y$$

- Then the distribution of parameters θ satisfies $\hat{\theta} \sim N(\theta,\ (X^T X)^{-1}\sigma^2)$, where we have

$$E[\hat{\theta}\hat{\theta}^T] = E\left[(X^TX)^{-1} X^T y \cdot y^T X (X^TX)^{-T}\right]$$
$$= (X^TX)^{-1} X^T E[yy^T] X (X^TX)^{-T}$$
$$= (X^TX)^{-1} X^T (\sigma^2 I + X\theta\theta^T X) X (X^TX)^{-1}$$
$$= \sigma^2 (X^TX)^{-1} + \theta\theta^T$$

- If σ^2 is known, then for the ith parameter, e.g., $\hat{\theta}_1$, the corresponding t-test is

$$t = \frac{\hat{\theta}_1 - \theta_1}{\sigma S_1},$$

Where $S_1 = (X^TX)^{-1}_{11}$.

- ***T-test***: estimate whether the first figure contributes to the output. Null hypothesis H_0: $\theta_1 = 0$, alternative hypothesis H_a: $\theta_1 \neq 0$.
- **Question**: however, at present, σ is unknown. Our next step is to estimate σ.
- **The distribution of error with covariance:**

$$\hat{\theta} = (X^TX)^{-1} X^T y.$$

$$e = y - X\theta = y(I - X(X^TX)^{-1}X^T)$$
$$E[ee^T] = y^T(I - X(X^TX)^{-1}X^T)(I - X(X^TX)^{-1}X^T)y$$
$$= y^T(I - X(X^TX)^{-1}X)y \sim \chi^2_{n-p}(0)$$

- The relationship between $E[e^2]$ and the σ is

$$E\left[\frac{e^2}{\sigma^2}\right] = n - p,$$

Where n is the number of samples, and p is the dimension of features. Then the estimation of $\hat{\sigma}^2$ is

$$\hat{\sigma}^2 = \frac{e^T e}{n - p}.$$

- **Final formulation of the t-test**: Null hypothesis H_0: $\theta_1 = 0$, alternative hypothesis H_a: $\theta_1 \neq 0$. The t-distribution with degree freedom n-p is

$$t_{n-p} = \frac{\hat{\theta}_1 - \theta_1}{\hat{\sigma} S_1} = \frac{\hat{\theta}_1 - 0}{\hat{\sigma} S_1},$$

Where $S_1 = (X^T X)^{-1}_{11}$ and $\hat{\sigma}^2 = \frac{e^T e}{n-p}$.

- Finally, based on the p-val, accept or reject the null hypothesis.

[1]. https://stats.stackexchange.com/questions/344006/understanding-t-test-for-linear-regression

[2]. https://www.geo.fu-berlin.de/en/v/soga/Basics-of-statistics/Hypothesis-Tests/Inferential-Methods-in-Regression-and-Correlation/Inferences-About-the-Slope/index.html

An introduction to ADF test

1 Introduction

- **Stationary**: a time series is considered stationary if it has constant mean μ_x, constant variance $R_{xx}(0) = \sigma^2$, and constant covariance $R_{xx}(\tau)$

A time series has to satisfy the following conditions to be considered stationary:

- **Constant mean** — average value doesn't change over time.
- **Constant variance** — variance doesn't change over time.
- **Constant covariance** — covariance between periods of identical length doesn't change over time.

- **Motivation**: A nonstationaty time series is called integrated if it can be transformed by first differencing once or a very few times into a stationary process. The order of integration is the minimum number of times the series needs to be first differenced to yield a stationary series. An **integrated** of order 1 time series is denoted by I(1) or AR(1). A stationary time series is said to be integrated of order zero, I(0).
- **Augmented Dickey Fuller Test (ADF)**: a *unit root* test for stationarity
- **Formulation**: a unit root test is represented by $Y_t = \rho Y_{t-1} + u_t$. If $\rho = 1$, the process can be considered as a random walk with a drift. Then it is not stationary.
  For ADF test: the hypothesis is test $H_0 : \phi = 0$ versus $H_1 : \phi < 0$.

$$\Delta Y_t = \phi Y_{t-1} + \alpha + \beta t + u_t$$

- Based on the P-val, we decide whether reject null hypothesis (nonstationary).

2 Hypothesis

- **Null hypothesis**: there is a unit root, i.e., $\phi = 0$, the model is not stationary.
- **Alternative hypothesis**: the time series is stationary (or trend-stationary).

3 Model selection

- Before running a ADF test, inspect your data to figure out an appropriate regression model.
- There are three basic regression model: $H_0{:}\,\gamma = 0,\ H_1{:}\,\gamma < 0$.

  - No constant, no trend: $\Delta y_t = \gamma y_{t-1} + v_t$
  - Constant, no trend: $\Delta y_t = \alpha + \gamma y_{t-1} + v_t$
  - Constant and trend: $\Delta y_t = \alpha + \gamma y_{t-1} + \lambda_t + v_t$

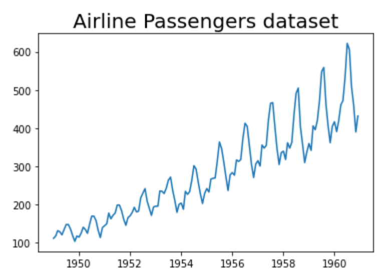
- **ADF test**: we also will choose **a lag length**. Then the corresponding models with m lags are shown below:

- No constant, no trend: $\Delta y_t = \gamma y_{t-1} + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t$
- Constant, no trend: $\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t$
- Constant and trend: $\Delta y_t = \alpha + \gamma y_{t-1} + \lambda_t + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t$

- **How to choose lag length**: The lag length should be chosen so that the residuals aren't serially correlated.
- **Three options for choosing lags**:
  a) Minimize Akaike's information criterion (AIC)
  b) Bayesian information criterion (BIC),
  c) drop lags until the last lag is statistically significant.

4 Example[4][5]

- **Dataset**: choose the airline passengers



- **Interpretation of data (below)**: The first output is the ADF test, the second output is the **p-value**. P-value is just over 0.99, showing that the dataset isn't stationary.

```
import matplotlib.pyplot as plt
from statsmodels.tsa.stattools import adfuller


df_adf=pd.read_csv("ADF_data.txt", index_col='Month', parse_dates=True)
#df_adf=pd.read_csv("ADF_data.txt")

plt.figure(3)
plt.title('Airline Passengers dataset', size=20)
plt.plot(df_adf)
'''
ADF verify stationary
'''
R5=adfuller(df_adf['Passengers'])
print(R5)
```

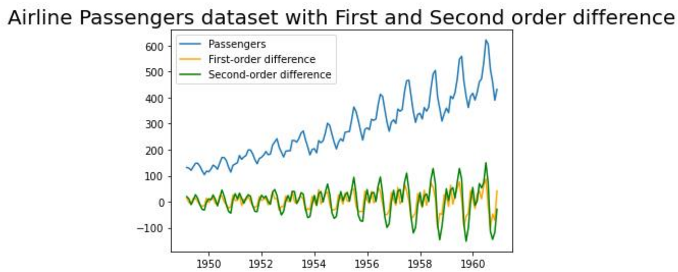
```
-----------------------ADF test Start---------------------------
(0.8153688792060502, 0.991880243437641, 13, 130, {'1%': -3.4816817173418295,
'5%': -2.8840418343195267, '10%': -2.578770059171598}, 996.692930839019)
```

- **First order and second order difference**: generate the 1st and 2nd order difference values. The two time series look promising to be stationary.

```python
'''
# First and second order difference
'''
df_adf['Passengers_Diff1'] = df_adf['Passengers'].diff()
df_adf['Passengers_Diff2'] = df_adf['Passengers'].diff(2)

# Don't forget to drop missing values
df_adf = df_adf.dropna()


# Plot the data
plt.figure(6)
plt.title('Airline Passengers dataset with First and Second order difference', size=20)
plt.plot(df_adf['Passengers'], label='Passengers')
plt.plot(df_adf['Passengers_Diff1'], label='First-order difference', color='orange')
plt.plot(df_adf['Passengers_Diff2'], label='Second-order difference', color='green')
plt.legend();
```



- **Run ADF for two above time series**: Given the critical value $\alpha = 0.05$, the 2nd order difference is stationary.

```python
# Perform ADF test
adf_diff_1 = adfuller(df_adf['Passengers_Diff1'])
adf_diff_2 = adfuller(df_adf['Passengers_Diff2'])

# Extract P-values
p_1 = adf_diff_1[1]
p_2 = adf_diff_2[1]

# Print
print(f'P-value for 1st order difference: {np.round(p_1, 5)}')
print(f'P-value for 2nd order difference: {np.round(p_2, 5)}')
```

```
----------------------ADF test Start--------------------------
(0.8153688792060502, 0.991880243437641, 13, 130, {'1%': -3.4816817173418295,
'5%': -2.8840418343195267, '10%': -2.578770059171598}, 996.692930839019)
P-value for 1st order difference: 0.05366
P-value for 2nd order difference: 0.03863
```

- **Best order** difference: By iteratively search from 0-max order, we can find some order difference such that it is stationary.

[1]. Augmented Dickey-Fuller Test with Python, https://www.exfinsis.com/tutorials/python-programming-language/augmented-dickey-fuller-test-with-python/

[2]. Stationarity and Unit Roots Tests, https://nscpolteksby.ac.id/ebook/files/Ebook/Accounting/Financial%20Econometrics%20with%20EViews%20(2010)/5.%20Chapter%204%20-%20Stationarity%20and%20Unit%20Roots%20Tests.pdf
[3]. https://www.exfinsis.com/tutorials/python-programming-language/augmented-dickey-fuller-test-with-python/
[4]. https://www.statology.org/dickey-fuller-test-python/
[5]. https://towardsdatascience.com/time-series-from-scratch-stationarity-tests-and-automation-14b02fa5ca4d
[6].
[7].