BUSINESS INTELLIGENCE AND DATA ANALYTICS

University of Petra

Institute of Analytics
Recognized Programme

307102
Descriptive Statistics for Business

Introduction to Hypothesis Testing

2024-2

Content

- Introduction to Hypothesis Testing
    - Business Examples
    - Significance Level – Alpha
    - Computing T-Statistic
    - Critical Value Approach vs P-Value Approach
    - Type I and Type II Errors
- T-Test: Used for normally distributed data when the variance is unknown.
    - One-sample t-test: Tests the mean of a single group against a known mean.
    - Two-sample t-test: Compares the means of two dependent groups (paired test).
    - Two-sample t-test: Compares the means of two independent groups (variance assumed to be equal).
    - Two-sample t-test: Compares the means of two independent groups (variance assumed to be not equal).
- Proportion Test: Compare sample proportion to a benchmark (e.g., open rate, conversion rate).
- ANOVA (Analysis of Variance): Used to compare the means of three or more samples.
    1. One Way ANOVA
    2. Two Way ANOVA
- Chi-square test: Tests for relationships between categorical variables.

Hypothesis Testing

- **Hypothesis Testing** is a statistical method that uses sample data to evaluate a hypothesis about a population parameter.
- A **hypothesis** is a claim about a population parameter that can be examined.
- The purpose of hypothesis testing is to determine whether there is enough evidence in a sample of data to support a particular belief about a population.
- Examples of claims that can be checked:
    1. Testing if a new marketing campaign increases customer engagement.
    2. Checking whether a redesigned website improves conversion rates.
    3. Evaluating if a new product packaging affects sales volume.
    4. Determining if weekend promotions boost in-store foot traffic.
    5. Assessing whether remote work impacts employee productivity.
    6. Comparing customer satisfaction between two service locations.
    7. Testing if a loyalty program increases repeat purchase rates.
    8. Seeing if a price change affects total revenue.
    9. Checking whether a new supplier leads to fewer product defects.
    10. Measuring if training improves sales team performance.

Business Related Examples in Hypothesis Testing

Category	Scenario	Type of Test	Detailed Experiment Design
Marketing	Testing if a new marketing campaign increases customer engagement.	Proportion	• Compare engagement rates (clicks, opens) between users exposed to the new vs. old campaign. • Randomly assign at least 500 customers to two groups: one receiving the new campaign, one the old. • Track engagement (clicks/opens) over a week.
	Seeing if a promotional email improves open or click-through rates.	Proportion	• Compare open/click-through rates between two versions of an email. • Send Version A to 1,000 customers and Version B to another 1,000. • Ensure equal demographic spread. • Compare open and click-through rates.
	Measuring whether influencer partnerships drive more traffic to the site.	Mean	• Compare average daily website visits with and without influencer campaigns. • Track daily website traffic for 30 days before and after the influencer campaign. • Use rolling averages to minimize noise.
	Evaluating if social media ads result in more conversions.	Proportion	• Compare conversion rates for users who saw the ads vs. those who did not. • Randomly expose 5,000 users to social media ads while keeping a matched control group unexposed. • Track conversion behavior over 7 days.
Sales & Revenue	Determining if a price change affects total revenue.	Mean	• Compare average revenue per user before and after price change. • Collect revenue data for 4 weeks before and after price change. • Use at least 1,000 transactions in each period to ensure reliability.
	Testing whether bundling products increases average order value.	Mean	• Compare average order value for bundled vs. unbundled products. • Assign 500 customers to view bundled product pricing and 500 to unbundled. • Track order values across one sales cycle.
	Assessing if offering free shipping improves purchase completion rates.	Proportion	• Compare checkout completion rates with and without free shipping. • Split 2,000 checkout users evenly into groups with and without free shipping. • Track completion rate within 24 hours.

Category	Scenario	Type of Test	Detailed Experiment Design
Operations	Checking whether a new supplier leads to fewer product defects.	Proportion	• Compare defect rates between products from the old vs. new supplier. • Collect defect data from 1,000 units produced using the old supplier and 1,000 from the new. • Monitor under similar production conditions.
	Evaluating if a new production method reduces waste or error rates.	Mean	• Compare mean waste/error amounts using old vs. new methods. • Run the production process using each method for a week, generating at least 500 measurements of waste or errors per method.
	Determining if delivery time decreases with a new logistics provider.	Mean	• Compare average delivery times between the two providers. • Record 1,000 delivery times for each provider across the same set of shipping regions and similar order types.
Customer Behavior	Testing if a loyalty program increases repeat purchase rates.	Proportion	• Compare proportion of returning customers with and without loyalty program. • Track purchase frequency over 3 months for 1,000 loyalty program users and 1,000 non-users, matched by purchase history.
	Seeing if a new mobile app design boosts daily active users.	Mean	• Compare average daily users before and after redesign. • Monitor app usage data for 30 days before and after the redesign, using data from at least 5,000 active users.
	Measuring if customer satisfaction differs after a policy change.	Mean	• Compare satisfaction ratings before and after the policy change. • Survey at least 500 customers before and after the policy change or use NPS ratings from two equivalent time periods.
Human Resources	Assessing whether remote work impacts employee productivity.	Mean	• Compare average productivity scores of remote vs. in-office workers. • Randomly select 200 remote and 200 in-office employees with similar roles. Track weekly productivity for 1 month.
	Measuring if training improves sales team performance.	Mean	• Compare average sales figures before and after training. • Analyze sales data from 100 employees for the 3 months before and after training. Normalize for seasonal trends.
	Testing if flexible hours reduce employee turnover.	Proportion	• Compare turnover rates between employees with and without flexible hours. Track turnover across 1,000 employees with flexible hours and 1,000 without, matched on role and tenure, over a year.

Key Concepts

- **Null Hypothesis (H_0):** A statement of no effect or no difference (usually we want to test this null state against a difference made by an intervention)
- **Alternative Hypothesis (H_1 or Ha):** What you want to prove – a statement of effect or difference.
- **Significance Level (α):** Probability of rejecting H_0 when it is actually true. Common choices: 0.05, 0.01.
- **p-value:** Probability of observing a result at least as extreme as the one obtained, assuming H_0 is true.
- **Test Statistic:** A value calculated from the sample to decide whether to reject H_0.

The Null and The Alternative Hypothesis

- When formulating a hypothesis test, we define two hypotheses: the null hypothesis (H_0) and the alternative hypothesis (H_1 or H_a).
- These two hypotheses complement each other—only one of them can be valid.
- For example, suppose a company claims that the average delivery time of its orders is 3 days.
- To test this claim, the company defines the hypotheses as following:
  - **The Null hypothesis H_0:** The average delivery time **IS EQUAL TO 3 days.**
  - **The Alternative hypothesis H_a:** The average delivery time **IS NOT 3 days.**
- Hypotheses are often expressed with symbols like this:

  H_0: μ = 3 days

  H_a: μ ≠ 3 days
- Notice that the null hypothesis sets a **baseline** assuming no difference or effect—in this context, that the mean delivery time is not different from 3 days.
- The burden of proof lies with the alternative hypothesis, which must show sufficient evidence that the average delivery time is different from 3 days.

Choosing the Right Test

- **One-sample z-test or t-test**: Compare sample mean to a known value.
- **Proportion test**: Compare sample proportion to a benchmark (e.g., open rate, conversion rate).
- **Two-sample t-test**: Compare means between two groups (e.g., control vs test).
- **Chi-square test**: Compare categorical variables (e.g., customer satisfaction vs region).
- **ANOVA Test:** Compare means between three or more groups.

The Process

1. **State** H_0 and Ha.
2. **Choose** α (e.g., 0.05).
3. **Collect data** and compute a test statistic.
4. **Compute** the p-value.
5. **Compare** p-value to α.
6. **Decision:** Reject or fail to reject H_0.

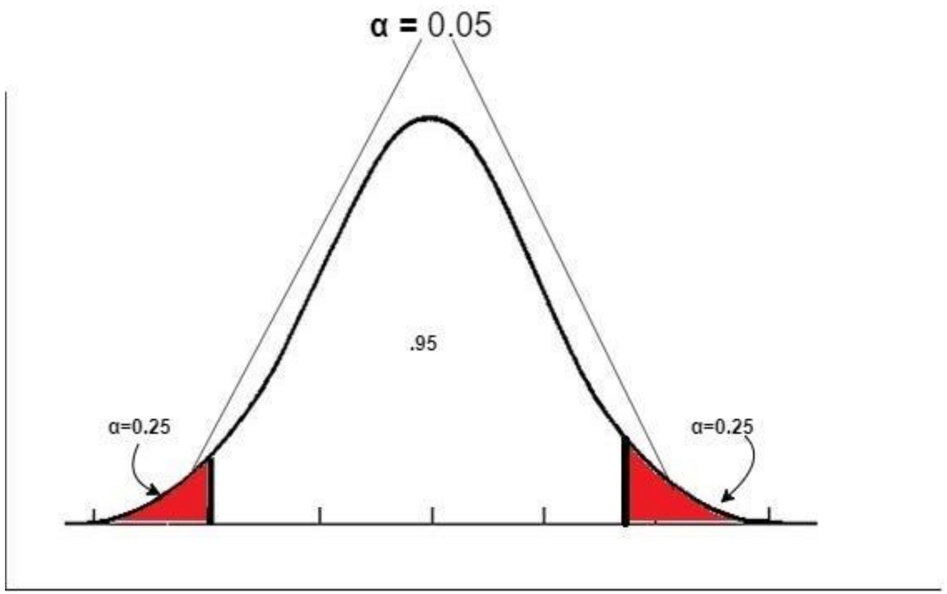
- Note: We cannot we accept the null hypothesis, we can only reject it or fail to reject it.
- The basis for this testing process comes from logic.
- For example, if our null hypothesis = All students live in Amman, it will be challenging to check if all students live in Amman.
- On the other hand, it will be much easier to find just one student who does not live in Amman and therefore reject the null hypothesis.
- Rejecting is much easier that proving or accepting the hypothesis.

More about Hypothesis Testing Core Concepts

Alpha value, The Test Statistic and Critical Values, P-Value, Type I and Type II Errors

The Significance Level α

- Before conducting the test, we need to set the significance level α.
- The significance level is a measure of the strength of the evidence that must be present in our sample to reject the null hypothesis and conclude that the effect is statistically significant.
- Typical significance levels are: α=0.1 (10%), α=0.05 (5%), α=0.01 (1%)
- A lower significance level means that the evidence in the data needs to be stronger to reject the null hypothesis.



There is no "correct" significance level - it only states the amount of uncertainty of the conclusion.
A 5% significance level means that we expect to reject a true null hypothesis 5 out of 100 times (Type I Error).

The Test Statistic

- The test statistic is the value we compute based on the sample data that we collect.
- The test statistic is used to decide the outcome of the hypothesis test.
- We standardize test statistic using:
  - Z Standard Normal Distribution for large samples or when the variance is known.
  - The Student's T-Distribution (T) for small samples and the population variance is unknown.

Computing the Test Statistic Dependence on our Knowledge about Variance

Populations	Variance	Samples	Statistic	Variance	Statistic
One	Known	-	Z	σ^2	$Z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$
One	Unknown	-	t	s^2	$t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$
Two	Known	Independent	Z	σ_x^2, σ_y^2	$Z = \frac{(\bar{x}-\bar{y})-\mu_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}$
Two	Unknown, Assumed Equal	Independent	t	$s_p^2 = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2}$	$t = \frac{(\bar{x}-\bar{y})-\mu_0}{\sqrt{s_p^2\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}}$
Two	Unknown, Not Assumed Equal	Independent	t	s_x^2, s_y^2	$t' = \frac{(\bar{x}-\bar{y})-\mu_0}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}}$
Two	Unknown	Dependent/Paired	t	s_d^2	$t = \frac{\bar{d}-\mu_0}{\frac{s_d}{\sqrt{n}}}$

The Critical Value and the p-Value Approach to Hypothesis Testing

- To decide whether to reject the null hypothesis, a test statistic is calculated.
- The decision is made based on the numerical value of that test statistic.
- There are two approaches how to arrive at that decision: The critical value approach and the p-value approach.

The critical value approach

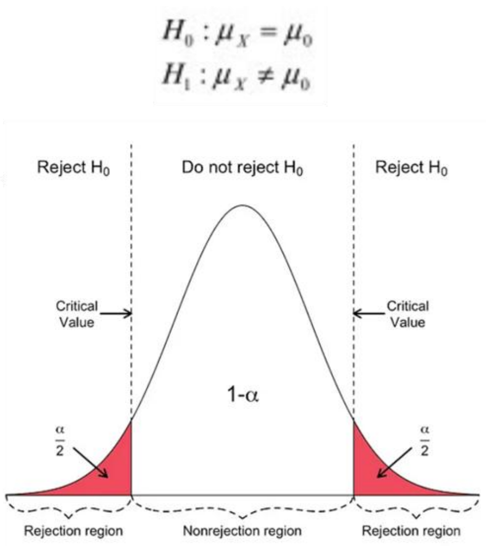
- By applying the critical value approach, it is determined, whether the observed test statistic is more extreme than a defined critical value.
- Therefore, the observed test statistic (calculated based on sample data) is compared to the critical value (a kind of cutoff value).
- If the test statistic is more extreme than the critical value, the null hypothesis is rejected.
- If the test statistic is not as extreme as the critical value, the null hypothesis is not rejected.
- The critical value is computed based on the given significance level α and the type of probability distribution of the idealized model.
- The critical value divides the area under the probability distribution curve in rejection region(s) and non-rejection region.

References

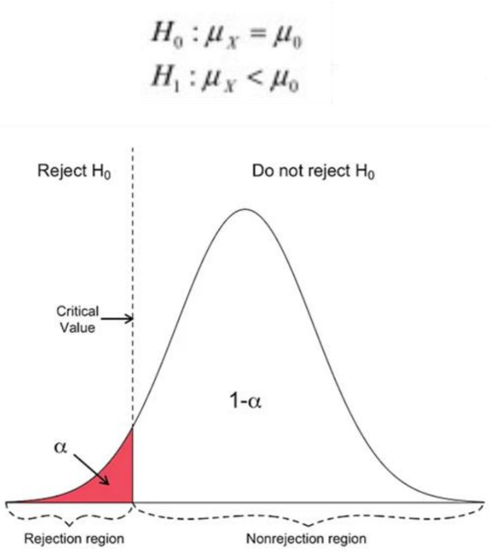
The Critical Value Approach to Hypothesis Testing

- The following three figures show a two tailed test, a left tailed test and a right sided test.
- In a two-sided test the null hypothesis is rejected if the test statistic is either too small or too large. Thus, the rejection region for such a test consists of two parts: one on the left and one on the right.
- For a left-tailed test the null hypothesis is rejected, if the test statistic is too small. Thus, the rejection region for such a test consists of one part, which is left from the center.
- For a right-tailed test the null hypothesis is rejected, if the test statistic is too large. Thus, the rejection region for such a test consists of one part, which is right from the center.
- Note: if p≤α , reject H_0 ; otherwise, if p>α , do not reject H_0

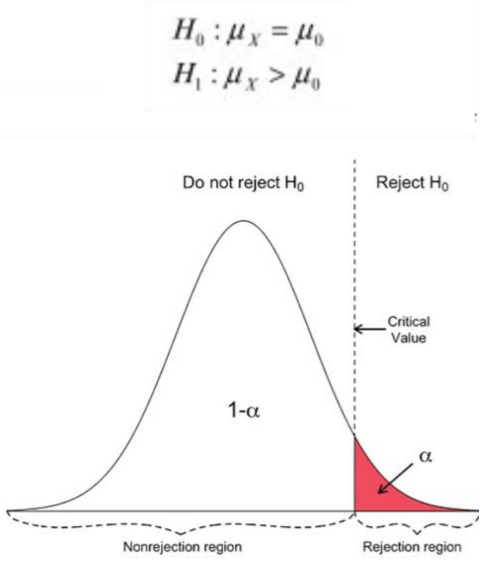
Two Tails Test



One Tail Test - Left Tail

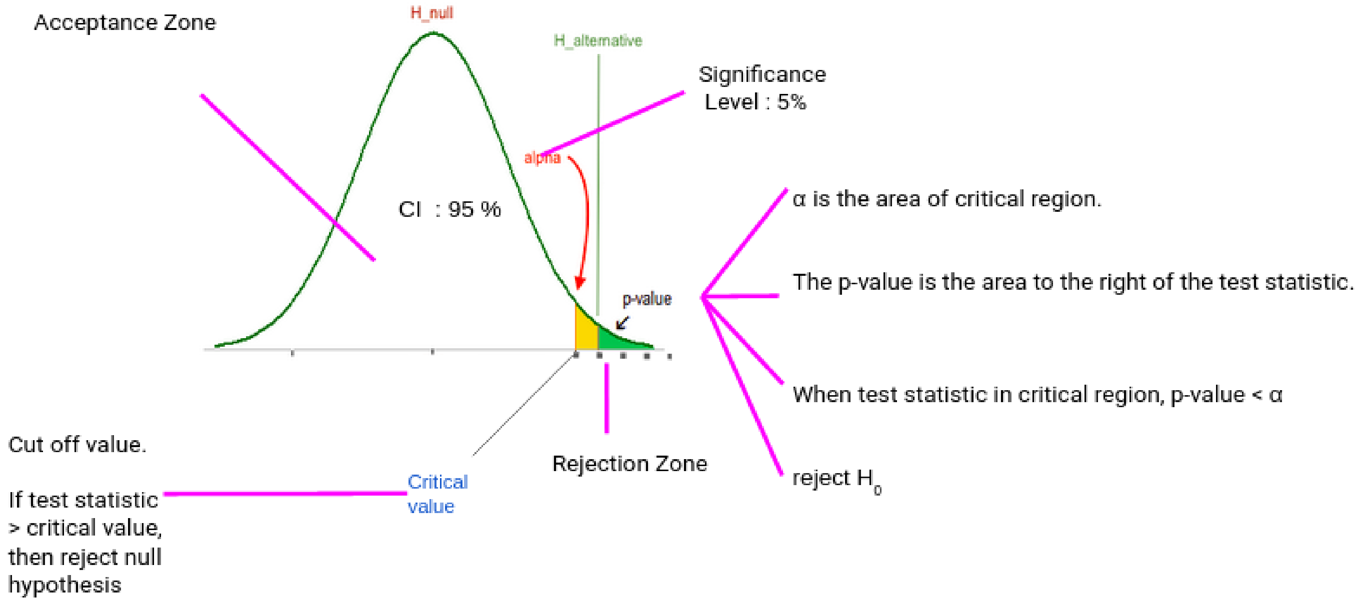


One Tail Test - Right Tail



The p-Value Approach to Hypothesis Testing

- The p-value represents:
  - The probability of observing the sample data or more extreme values, given that the null hypothesis is true.
  - The amount of evidence against the null hypothesis, the smaller (closer to 0) the p-value, the stronger is the evidence against the null hypothesis.
- If the p-value (**Green Area**) is less than or equal to the specified significance level α (**Yellow Area**) , the null hypothesis is rejected. Otherwise, we fail to reject the null hypothesis.

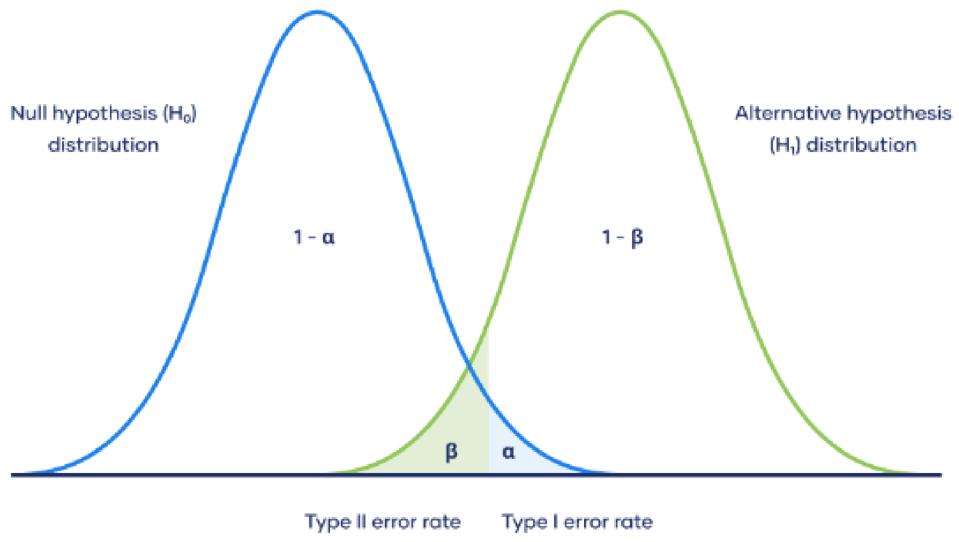


References

Type I and Type II Errors

Type I and Type II errors are two potential forms of incorrect conclusions in hypothesis testing.

- Type I Error (False Positive): Rejecting H_0 when it is actually true.
- Type II Error (False Negative): Failing to reject H_0 when it is false.
- The acceptable level of Type I error is designated by alpha (α), while the acceptable level of Type II error is designated beta (β).



- A Type I error occurs when the null hypothesis is incorrectly rejected when it is actually true.
- Conversely, a Type II error happens when the null hypothesis is not rejected when it is in fact false, akin to a false negative, where we fail to detect a true effect or difference.
- Consider a pharmaceutical company deciding whether a new drug is effective. A Type I error would occur if the company concludes the drug works when it does not, potentially leading to a costly recall and reputation damage when the truth emerges.
- On the other hand, a Type II error would be failing to recognize the drug's efficacy, missing out on the opportunity to bring a beneficial product to market, which can result in lost revenue and competitive disadvantage. The company must balance these risks, often prioritizing the avoidance of one type of error over the other based on the potential costs and impacts associated with each.

Hypothesis Testing Examples

One Sample Test – Testing for Two Tails

- **Scenario:** A logistics company believe that its average delivery time is equal to 3 days. The company wants to test this belief.
- The company collected a sample data of 40 delivery times to test if there enough evidence (at $\alpha = .05$) to conclude that the population mean delivery time for the company is equal to 3 days?

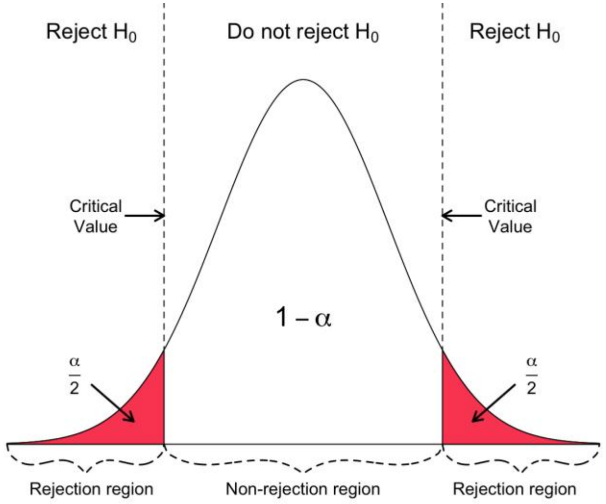
  $H_0: \mu = 3$

  $H_a: \mu \neq 3$

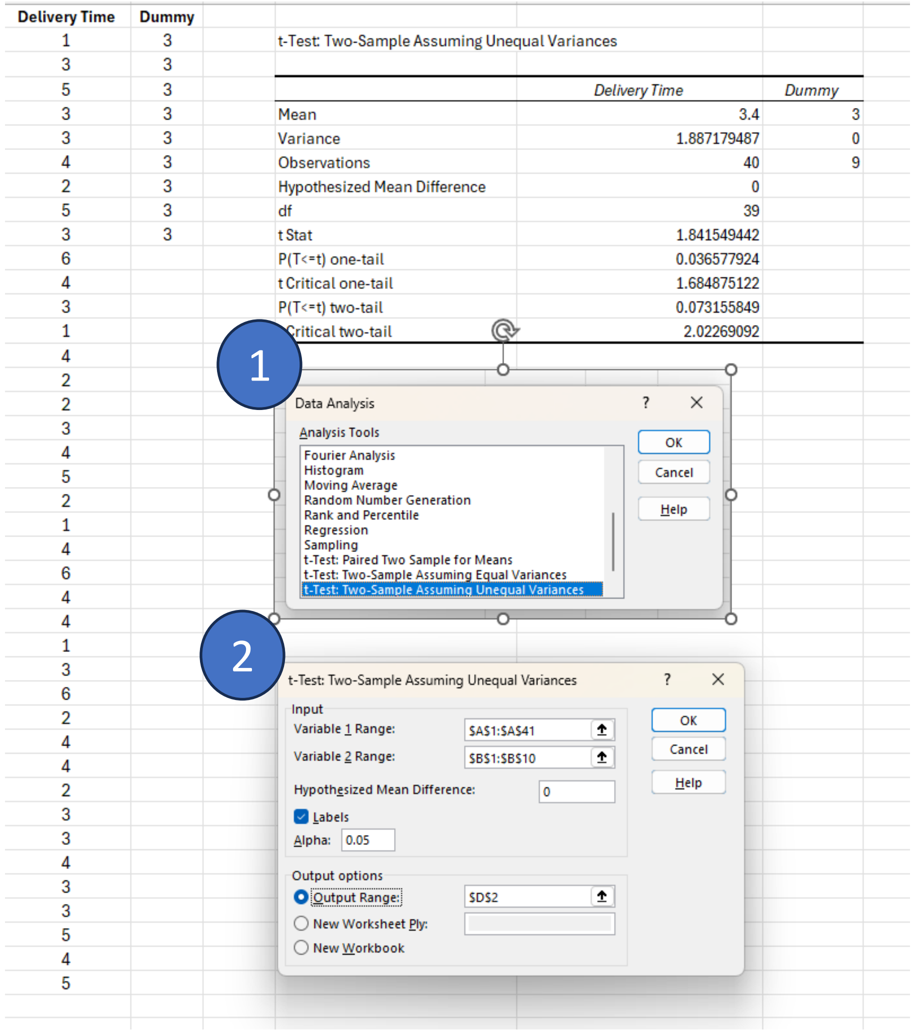
- Using the collected data, the t-statistic will be computed:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

- If the test statistic is within the rejection region, we reject the null hypothesis, otherwise we fail to reject the null hypothesis noting that the sample data did not show enough evidence against the null hypothesis.



Implement the Test Using Excel Data Analysis Add In



t-Test: Two-Sample Assuming Unequal Variances

	Delivery Time	Dummy
Mean	3.4	3
Variance	1.887179487	0
Observations	40	9
Hypothesized Mean Difference	0	
df	39	
t Stat	1.841549442	
P(T<=t) one-tail	0.036577924	
t Critical one-tail	1.684875122	
P(T<=t) two-tail	0.073155849	
t Critical two-tail	2.02269092	

Delivery Times for 40 Days Sample – In Days			
1	4	1	4
3	3	4	2
5	1	6	3
3	4	4	3
3	2	4	4
4	2	1	3
2	3	3	3
5	4	6	5
3	5	2	4
6	2	4	5

Test Results	
	Delivery Times
Mean	3.4
Variance	1.887179487
Observations	40
Hypothesized Mean Difference	0
df	39
t Stat	1.841549442
P(T<=t) one-tail	0.036577924
t Critical one-tail	1.684875122
P(T<=t) two-tail	0.073155849
t Critical two-tail	2.02269092

- According to the test results, we fail to reject the null hypothesis which states that the mean delivery time is equal to 3 days.
- The collected sample data did not provide us with sufficient evidence to conclude that the mean delivery time is different from 3.
- We reached this conclusion because the t statistic falls inside the non-rejection region.
- We also reached that conclusion because the p-value (0.07315) which is greater than alpha 0.05.
- Computing t-statistic manually, $t = (3.4 - 3) / (1.37 / \sqrt{40}) \approx 1.84$

One Sample – One Tail Test – Upper Tail

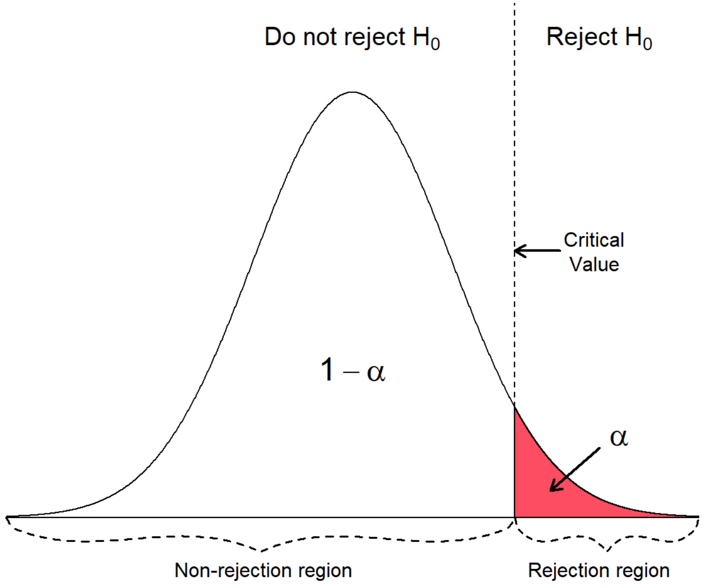
- **Business Context:**
  A manufacturing company believes that its daily production output is larger than 500 units.
- The company collect sample outputs for 30 days.

Hypothesis Formulation:

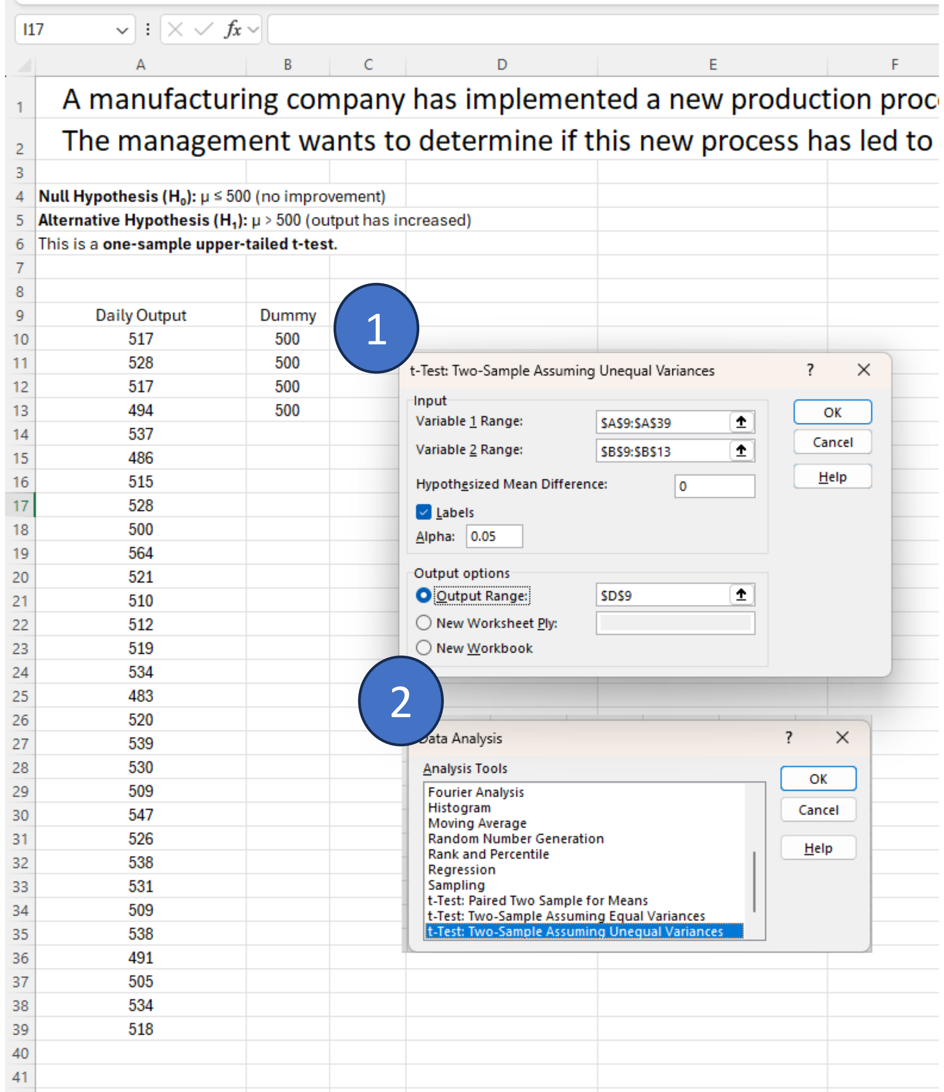
Null Hypothesis (H_0): $\mu \leq 500$ (Baseline)

Alternative Hypothesis (H_1): $\mu > 500$ (Output is larger)

- This setup calls for a **one-tailed (upper-tailed) t-test**, as the interest lies specifically in detecting an increase in the mean output.



Implement the Test Using Excel Data Analysis Add In



Collect Daily Output Data	
517	534
528	483
517	520
494	539
537	530
486	509
515	547
528	526
500	538
564	531
521	509
510	538
512	491
519	505
518	534

t-Test: Two-Sample Assuming Unequal Variances		
	Daily Output	Dummy
Mean	520	500
Variance	340.0689655	0
Observations	30	4
Hypothesized Mean Difference	0	
df	29	
t Stat	5.940282826	
P(T<=t) one-tail	9.38286E-07	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	1.87657E-06	
t Critical two-tail	2.045229642	

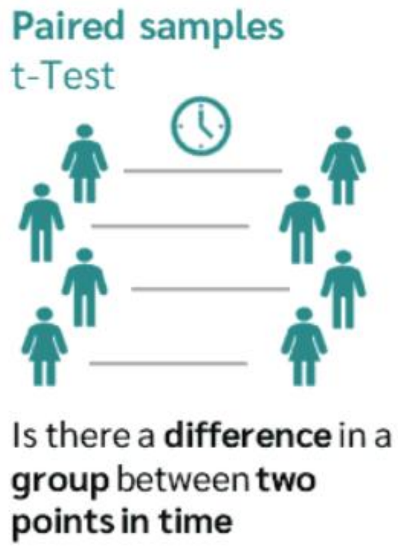
T statistic is above the critical value and P-Value is very small < alpha, therefore we reject the null hypothesis

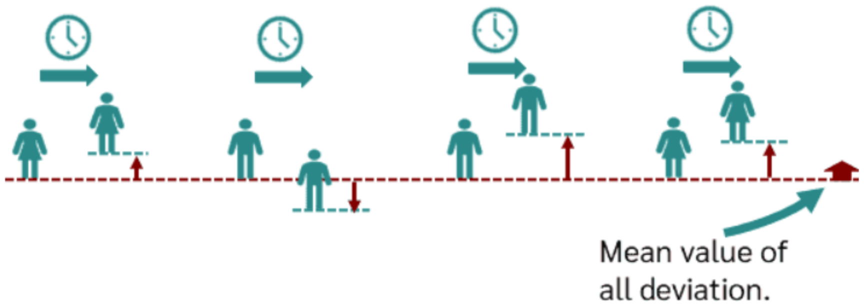
Two Sample T-Tests

1. Two-sample t-test: Compares the means of two dependent groups (Paired t test).
2. Two-sample t-test: Compares the means of two independent groups (variance assumed to be equal).
3. Two-sample t-test: Compares the means of two independent groups (variance assumed to be not equal).

Two Samples T-Test – Testing Dependent Groups (Paired T Test)

- The dependent samples T-test (or paired samples t-test) is a statistical test that determines whether there is a difference between two dependent groups or samples.
- We need the paired t-test whenever we survey the same group or sample at two points in time.
- This can be the case, for example, in longitudinal studies with several measurement points (time series analyses) or in intervention studies with experimental designs (before-after measurement).
- An example of dependent sampling is when the weight of a group of people is measured at two points in time (before and after going into a diet).
- The weight of a person can be measured individually before and after going into the diet, and the difference in the measured values can be calculated in each case.





References

Two Samples T-Test – Testing Dependent Groups (Paired T Test)

- We are interested in the question whether an online statistics learning tutorial helps students improve their grades.
- There are three variables of interest in the student's data set (score1, score2 and online.tutorial).
- The variables score1 and score2 show the grades (0-100) for two exams on mathematics after attending online training. The higher the value the better the student performed.
- The variable online.tutorial is a binary variable, which is 1 if the student completed the online statistics learning tutorial, or 0 otherwise.
- The first exam takes place before the students attended the online statistics learning tutorial.
- The participation in the online statistics learning tutorial is not mandatory, however the two exams are obligatory for all students.
- We want to examine if the group of students, which attended the online statistics learning tutorial, performs better on the second exam compared to the first exam.

$$H_0: \mu_{score2} \le \mu_{score1}$$

Null Hypothesis states that there is no effect for the intervention

$$H_1: \mu_{score2} > \mu_{score1}$$

Alternative Hypothesis claims that the intervention caused an effect

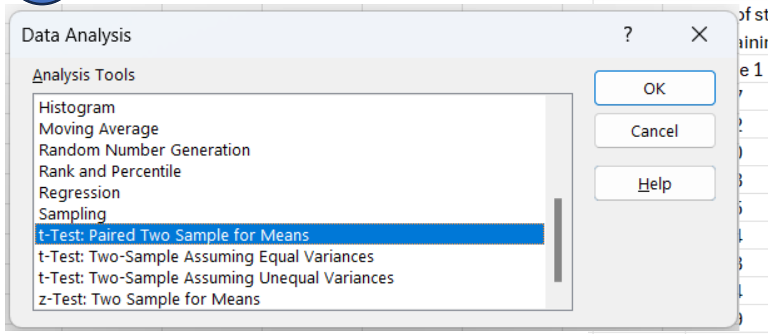
Random sample of students who took the online training

`	Score 1	Score 2
1	67	71
2	62	64
3	60	57
4	53	52
5	75	77
6	64	79
7	43	48
8	74	70
9	59	53
10	54	58
11	66	66
12	81	90
13	55	57
14	64	68
15	72	71
16	76	78
17	75	73
18	90	92
19	83	90
20	64	73
21	49	51
22	41	51
23	75	87
24	66	66
25	75	74
26	49	57
27	91	89
28	68	69
29	72	72
30	87	93

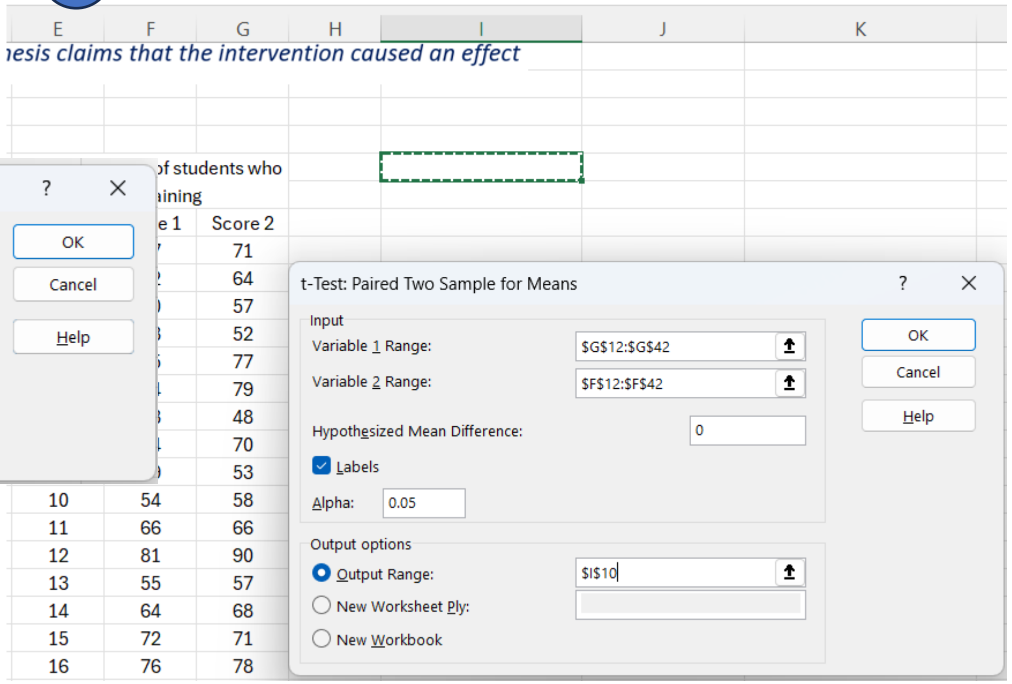
References

Two Samples T-Test – Testing Dependent Groups (Paired T Test)

1



2



3

t-Test: Paired Two Sample for Means		
	Score 2	*Score 1*
Mean	69.86666667	67
Variance	180.6712644	171.3793103
Observations	30	30
Pearson Correlation	0.931421115	
Hypothesized Mean Diff	0	
df	29	
t Stat	**3.187976131**	
P(T<=t) one-tail	**0.00171094**	
t Critical one-tail	**1.699127027**	
P(T<=t) two-tail	0.003421879	
t Critical two-tail	2.045229642	

- Based on the sample data and the test statistic, we found that there is a statistically significant difference between the before and after score, the after score (score2) is higher than score1.
- This finding is supported by the test statistic and the p-value, therefore we REJECT the null hypothesis, and we conclude that the online tutorial helped students in enhancing their exam performance.

References

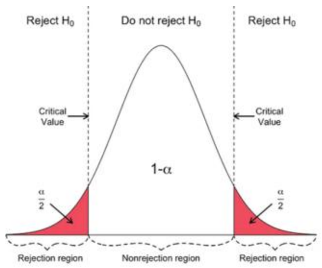
Two Samples T-Test – Testing Independent Groups (Unpaired T Test) (variance assumed to be equal)

- In this section we perform a hypothesis test for the means of two populations.
- We examine the mean annual salary (in Euro) of graduates.
- The first population consist of male students and the second population of female students.
- The standard deviations of the two populations are unknown but assumed to be equal.
- The question is, whether there is a difference (at α 5%) in the mean annual salary of graduates related to gender (two tailed test)

	Male Random Sample	Female Random Sample
1	$39,647.68	$29,543.60
2	$55,726.10	$30,972.76
3	$34,717.77	$33,216.12
4	$29,999.35	$24,143.27
5	$50,763.66	$40,466.69
6	$52,433.73	$29,543.60
7	$44,385.84	$47,760.48
8	$36,599.27	$45,054.95
9	$47,842.44	$42,742.79
10	$30,854.42	$42,202.39
11	$58,614.72	$40,606.41
12	$62,887.03	$29,783.76
13	$48,288.53	$46,945.17
14	$43,191.15	$29,555.94
15	$43,756.58	$45,042.30
16	$48,453.00	$29,555.94
17	$23,090.67	$45,625.09
18	$55,660.20	$28,017.87
19	$57,527.17	$35,840.90
20	$65,776.23	$33,648.38
21	$50,011.69	$46,858.67
22	$46,213.64	$42,463.48
23	$42,211.41	$33,200.27
24	$61,384.55	$28,338.17
25	$49,780.18	$26,265.35
26	$50,474.57	$34,373.99
27	$43,281.92	$37,312.85
28	$55,484.64	$37,709.73
29	$24,409.73	$35,106.01
30	$40,489.49	$37,807.11

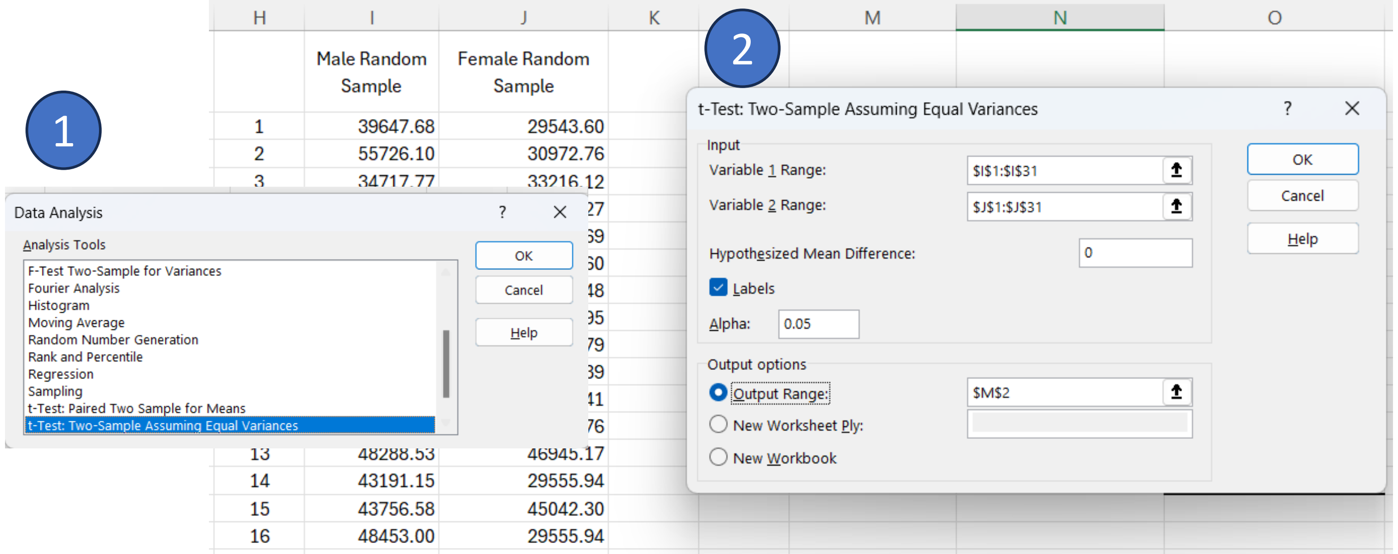
$H_0: \mu_{male} = \mu_{female}$ *Null Hypothesis states that there is no difference*

$H_1: \mu_{male} \neq \mu_{female}$ *Alternative Hypothesis claims that there is a difference*



References

Two Samples T-Test – Testing Independent Groups (Unpaired T Test) (variance assumed to be equal)



3

t-Test: Two-Sample Assuming Equal Variances

	Male Random Sample	*Female Random Sample*
Mean	46465.24594	36323.46873
Variance	117433271.4	48853686.8
Observations	30	30
Pooled Variance	83143479.09	
Hypothesized Mean	0	
df	58	
t Stat	**4.30770147**	
P(T<=t) one-tail	3.22648E-05	
t Critical one-tail	1.671552762	
P(T<=t) two-tail	**6.45296E-05**	
t Critical two-tail	**2.001717484**	

- Based on the sample data and the test statistic, we found that there is a statistically significant difference between male salaries and female salaries.
- This finding is supported by the test statistic and the p-value, therefore we REJECT the null hypothesis, and we conclude that the salaries of male graduates and female graduates are not equal.

References

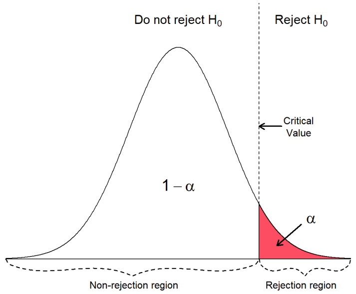
Two Samples T-Test – Testing Independent Groups (Unpaired T Test) (variance assumed to be NOT equal)

- In this section we perform a hypothesis test for the means of two populations.
- To showcase the non-pooled t-test we test a claim that states that the annual salaries for political science female students is higher than the annual salaries of the social science female students.
- The standard deviations of the two populations are unknown and assumed to be unequal.
- We want to test at α 1%, whether the mean political science salary is higher than the mean salary of the social science (one tail test, upper tail).

	Political Science Random Sample	Social Science Random Sample
1	$30,796.16	$33,641.73
2	$32,038.41	$29,635.67
3	$32,038.41	$38,738.01
4	$32,996.39	$21,453.99
5	$33,648.38	$31,484.80
6	$41,430.16	$34,287.96
7	$40,032.78	$35,583.43
8	$39,710.80	$28,672.85
9	$36,911.26	$30,062.71
10	$45,900.13	$34,420.48
11	$35,270.19	$33,641.73
12	$43,318.41	$26,291.32
13	$41,282.13	$31,151.42
14	$41,430.16	$38,738.01
15	$21,485.62	$33,565.74
16	$29,195.50	$31,484.80
17	$45,141.81	$32,194.59
18	$34,492.08	$35,746.22
19	$27,893.26	$28,281.25
20	$38,591.54	$31,405.48
21	$33,726.58	$32,194.59
22	$29,832.22	$35,987.41
23	$32,154.94	$35,583.43
24	$34,686.49	$34,227.49
25	$25,607.82	$25,758.88
26	$33,726.58	$34,373.99
27	$37,782.86	$25,658.83
28	$35,507.83	$34,201.58
29	$33,955.21	$31,151.42
30	$25,883.03	$33,641.73

$H_0: \mu_{political\ science} \leq \mu_{social\ science}$ *Null Hypothesis states that the salaries are equal*

$H_1: \mu_{political\ science} > \mu_{social\ science}$ *Alternative Hypothesis claims that political science is higher*

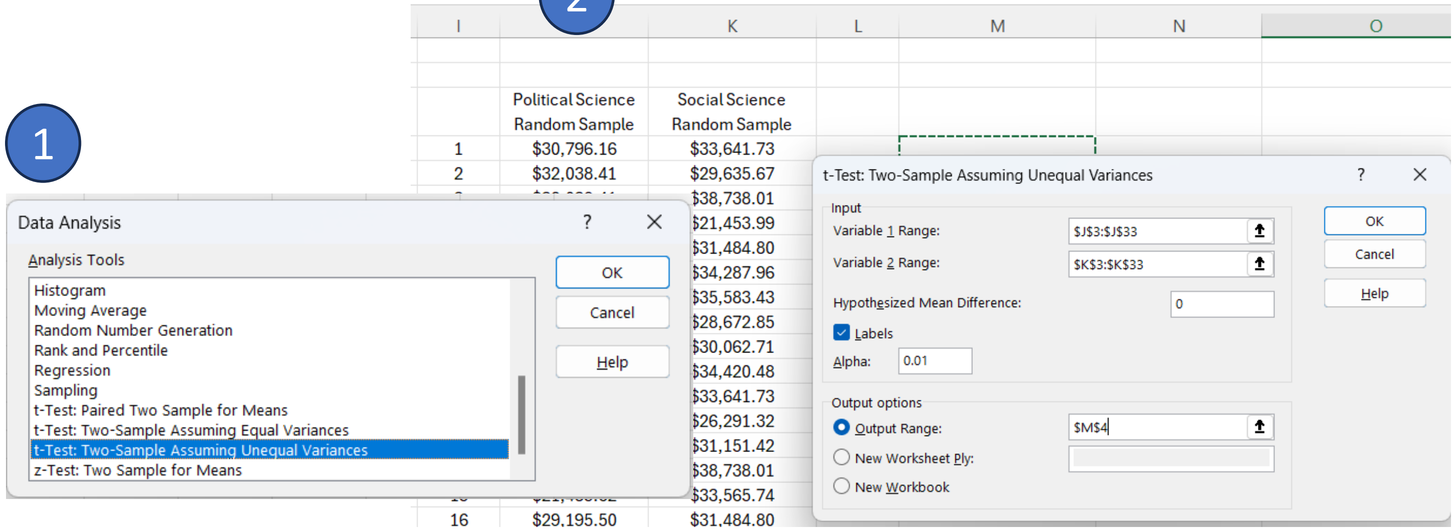


References

Two Samples T-Test – Testing Independent Groups (Unpaired T Test) (variance assumed to be NOT equal)

1

2



3

t-Test: Two-Sample Assuming Unequal Variances

	Political Science Random Sa	Social Science Random Sample
Mean	34882.23824	32108.71836
Variance	34602476.07	15267236.06
Observations	30	30
Hypothesized Mean Differe	0	
df	50	
t Stat	**2.151163793**	
P(T<=t) one-tail	**0.018159982**	
t Critical one-tail	**2.403271917**	
P(T<=t) two-tail	0.036319964	
t Critical two-tail	2.677793271	

- Based on the sample data and the test statistic, we found that there is no statistically significant difference (at α 1%) between the salaries of political science and social science.
- This finding is supported by the test statistic and the p-value, therefore we FAIL TO REJECT the null hypothesis which states that the salaries of political science and social science are equal for female graduates.

References

Testing Proportions

Introduction to Proportion Tests

- **A Proportion** is a fraction or percentage of a population exhibiting a specific trait (e.g., success/failure).
- Proportion tests allow us to **infer population parameters** from sample data.

Key Differences from t-tests:

- Applied to **categorical data** (e.g., success/failure), not continuous variables.
- Uses the **normal approximation** to the binomial distribution. The **central limit theorem** says: if the sample size is large enough, the binomial distribution starts to look like a **normal distribution**.
- Therefore, to use the normal distribution as an approximation for the binomial, both of these must be true:
  - $n.p \geq 10$
  - $n.(1-p) \geq 10$
  - Where:
    - n = sample size
    - p = hypothesized population proportion (under the null hypothesis)

One-Sample Proportion Test

Purpose:

To test whether a population proportion equals a specific hypothesized value.

Test Statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Where:

- $\hat{p}$: sample proportion
- p_0: hypothesized population proportion
- n: sample size

Assumptions:

- Random sample
- Independent observations
- Large sample: $np_0 \geq 10$ and $n(1 - p_0) \geq 10$

Example

A company claims 75% of customers are satisfied. In a sample of 200 customers, 140 are satisfied. Is the true satisfaction rate different?

Step-by-Step:

- **H₀:** $p = 0.75$
- **Hₐ:** $p \neq 0.75$

$$\hat{p} = \frac{140}{200} = 0.70, \quad z = \frac{0.70 - 0.75}{\sqrt{0.75 \times 0.25/200}} = \frac{-0.05}{0.0306} \approx -1.63$$

- **p-value (two-tailed):**

$$p = 2 \times P(Z < -1.63) = 2 \times 0.0516 = 0.1032$$

- **Decision (α = 0.05):**
  Fail to reject H_0 (p-value > 0.05)
- **Conclusion:**
  Insufficient evidence to say the satisfaction rate differs from 75%.

Two-Samples Proportion Test

Purpose:

To compare proportions from two **independent** groups.

Test Statistic (pooled):

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where:

- $\hat{p}_1, \hat{p}_2$: sample proportions
- n_1, n_2: sample sizes
- $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$: pooled proportion

Assumptions:

- Independent random samples
- Each group satisfies: $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

Example: Does the New Treatment Work?

A pharmaceutical company claims that its new treatment is more effective than the standard treatment. In a clinical trial:

- 65 out of 100 patients improved with the new treatment.
- 50 out of 100 patients improved with the standard treatment.

To evaluate this claim, we formulate the hypotheses as follows:

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2$$

This is a one-tailed test for comparing two proportions

$$\hat{p}_1 = 0.65, \quad \hat{p}_2 = 0.50, \quad \hat{p} = \frac{115}{200} = 0.575$$

$$z = \frac{0.65 - 0.50}{\sqrt{0.575 \times 0.425 \times \left(\frac{1}{100} + \frac{1}{100}\right)}} = \frac{0.15}{0.0699} \approx 2.15$$

- **p-value:**
  $p = P(Z > 2.15) = 0.0158$
- **Decision ($\alpha = 0.05$):**
  Reject H_0 (p-value < 0.05)
- **Conclusion:**
  There is **evidence** the new treatment is more effective.

Hypothesis Testing for Categorical Data

Chi-Square Tests

Introduction to Chi-Square Tests

Definition: Chi-square tests are statistical methods used to determine whether there's a significant association between categorical variables or whether observed data differs significantly from expected values.

- Types of Chi-Square Tests:
  - Test of Independence: Determines if two or more categories are independent.
  - Goodness of Fit Test: Determines if a sample matches an expected distribution.

Examples

Test of Independence

- **Marketing**: Is customer gender associated with product preference?
- **HR Analytics**: Is employee department linked to job satisfaction level?
- **Retail**: Does store location affect the type of payment method used?
- **Healthcare Management**: Is patient age group related to preferred appointment times?
- **Customer Service**: Is complaint category associated with region?

Chi-Square Steps – Test of Independence

Step-by-Step Process:

1. Set a contingency table to show the expected frequencies and the observed frequencies.
2. Calculate the expected frequencies for each cell using the following formula.

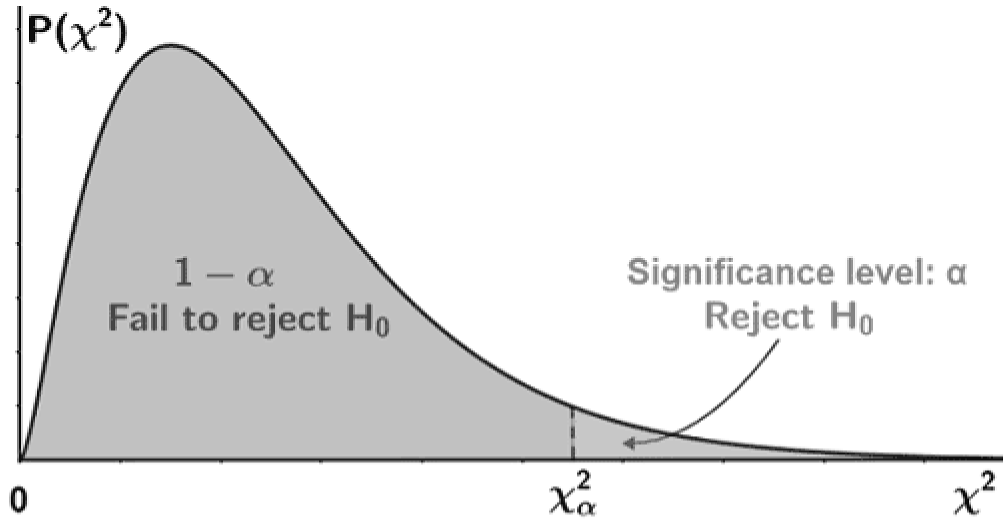
$$E_{ij} = \frac{(Row\ Total) \times (Column\ Total)}{Grand\ Total}$$

3. Compute Chi-Square statistic using the formula

- **Formula:** $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$
  - O_i: Observed frequency
  - E_i: Expected frequency
- **Degrees of Freedom:** $(r-1) \times (c-1)$
  - r: Number of rows in the table
  - c: Number of columns in the table

Chi-Square Steps – Test of Independence

4. Use the Chi-Square distribution to find the Chi-Square statistic and compare it to the critical values at common significance level e.g. (0.05)



- **Decision Rule:**
  - If χ^2 calculated > χ^2 critical, reject the null hypothesis (significant association).
  - If χ^2 calculated ≤ χ^2 critical, fail to reject the null hypothesis (no significant association).

Example: Product Preference Across Age Groups

- A company wants to determine if there is a significant association between the age group of its customers and their preference for one of two new beverage products.
- The marketing team has conducted a survey and **collected data** from different age groups.
- The table below shows the **observed frequency** table based on the survey:

	Product A	Product B	Total
Under 30	30	40	70
30 to 50	50	30	80
Over 50	20	30	50
Total	100	100	200

Task: Perform a Chi-Square Test of Independence to determine if there's a significant relationship between age group and product preference.

Steps to Solve in Excel

1. **Prepare the Data in Excel:**
   - Create a new Excel spreadsheet.
   - Enter the data into a worksheet as shown in the table above.
2. **Calculate Expected Frequencies:**
   - Each cell's expected frequency (E) can be calculated using the formula: $E_{ij} = \frac{(Row\ Total) \times (Column\ Total)}{Grand\ Total}$.
   - For instance, the expected frequency for Product A and Under 30 would be $E_{11} = \frac{70 \times 100}{200} = 35$.
3. **Setup Chi-Square Calculation:**
   - Create another table next to the first one to fill in the expected frequencies.
   - Below this table, calculate the Chi-Square statistic using the formula: $\chi^2 = \sum \frac{(O-E)^2}{E}$, where O is the observed frequency and E is the expected frequency.
4. **Use Excel Formula:**
   - For cell A1 (Under 30, Product A): `` `=((B2-B8)^2)/B8` `` where B2 is observed and B8 is expected frequency.
   - Drag and fill this formula for all cells.

Calculate Chi-Square Value

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Category (Age Group, Product)	Observed (O)	Expected (E)	O−E	(O−E)^2	E(O−E)^2
Under 30, Product A	30	35	-5	25	25/35≈0.714
Under 30, Product B	40	35	5	25	25/35≈0.714
30 to 50, Product A	50	40	10	100	100/40=2.500
30 to 50, Product B	30	40	-10	100	100/40=2.500
Over 50, Product A	20	25	-5	25	25/25=1.000
Over 50, Product B	30	25	5	25	25/25=1.000
Total	200	200			χ2≈8.428

Steps to Solve in Excel

5. **Sum the Results:**
   - Add up all the individual values calculated in the above step to get the total Chi-Square statistic.
6. **Determine the Degrees of Freedom:**
   - Calculate degrees of freedom as $(rows - 1) \times (columns - 1)$. Here, it would be $(3 - 1) \times (2 - 1) = 2$.
7. **Compare to Critical Value:**
   - Use a Chi-Square distribution table or Excel's CHISQ.DIST.RT function to find the critical value or p-value. The formula in Excel would be: `` `=CHISQ.DIST.RT(chi_square_statistic, degrees_freedom)` ``
   - Generally, if p-value < 0.05, the result is significant.

Interpretation

After calculating the Chi-Square statistic and comparing it with the critical value from a Chi-Square distribution table (or using p-value), you can conclude whether there is a statistically significant association between age group and product preference.

Check the Solution in The Excel File

Observed Values - Data Collected by the Syrvey

	Product A	Product B	Total
Under 30	30	40	70
30 to 50	50	30	80
Over 50	20	30	50
Total	100	100	200

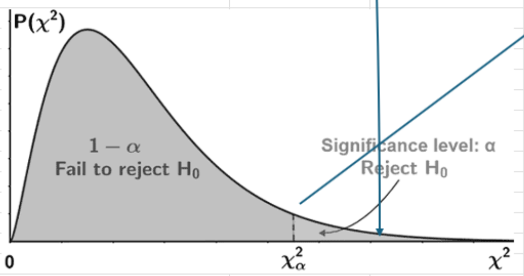
Expected Values if they were Independent

	Product A	Product B
Under 30	35	35
30 to 50	40	40
Over 50	25	25

(O-E)^2/E

	Product A	Product B
Under 30	0.714285714	0.714285714
30 to 50	2.5	2.5
Over 50	1	1

	X2	**8.428571429**
Compute the P-Value Manually	=CHISQ.DIST.RT(I22,2)	**0.014782877**
Compute the P-Value Using Excel Function	=CHISQ.TEST(B6:C8,I6:J8)	**0.014782877**



Significance level (α)

Degrees of freedom (df)	.99	.975	.95	.9	.1	.05	.025	.01
1	-------	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379
70	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329
100	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116
1000	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807

Hypothesis Testing for More than Two Populations Means

One Way ANOVA

Introduction to One-Way ANOVA

- **Definition:** One-Way ANOVA is a statistical test that compares the means of three or more independent groups to determine if at least one group mean is statistically different from the others.
- **Purpose:** To test for significant differences between the means of multiple groups when you have one categorical independent variable and one continuous dependent variable.
- Appropriate Use Cases:
  - Comparing customer satisfaction levels across different store locations.
  - Examining student performance across different teaching methods.
- Key Advantage:
  - Allows simultaneous comparison of more than two groups, providing a broader analysis than t-tests.

One Way ANOVA

Hypothesis in One-Way ANOVA:

- Null Hypothesis (H0): All group means are equal.
- Alternative Hypothesis (H1): At least one group mean is different from the others.

How to Calculate?

1. Calculate Group Means and Overall Mean
2. Compute Sum of Squares:
   - Total Sum of Squares (SST): Variability of all data points around the overall mean.
   - Within-Group Sum of Squares (SSW): Variability within each group.
   - Between-Group Sum of Squares (SSB): Variability due to the interaction between the groups.
3. Calculate Mean Squares:
   - Mean Square Between (MSB): SSB divided by the degrees of freedom between groups (k-1).
   - Mean Square Within (MSW): SSW divided by the degrees of freedom within groups (N-k).
4. F-Statistic:
   - F = MSB / MSW

ANOVA Example

- Scenario: A school tests three different teaching methods to determine which one results in the best performance on a standardized test.
- The students are randomly assigned to one of three groups, each receiving instruction through a different method.
- Their scores on a standardized test are recorded.
- Variables:
  - Group: The teaching method group (A, B, C)
  - Score: The score each student received on the standardized test.

Student ID	Group	Score
1	A	82
2	A	88
3	A	84
4	A	90
5	A	85
6	B	78
7	B	75
8	B	80
9	B	82
10	B	77
11	C	90
12	C	92
13	C	95
14	C	91
15	C	94

Interpreting Results

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Group A	5	429	85.8	10.2		
Group B	5	392	78.4	7.3		
Group C	5	462	92.4	4.3		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	490.5333	2	245.2667	33.75229	1.18232E-05	3.885293835
Within Groups	87.2	12	7.266667			
Total	577.7333	14				

- F-Statistic and P-Value:
  - An F-Statistic higher than the critical value or a P-value less than the alpha level (e.g., 0.05) indicates significant differences between group means.
  - In our example, both the F-Statistic and the P-value indicate significant difference between groups means, therefore we reject the null hypothesis.
- Post Hoc Tests: If the ANOVA is significant, perform post hoc tests (e.g., Tukey’s HSD) to determine which specific groups differ from each other.

2-Way ANOVA

- **Case Study: Effect of Teaching Method and Gender on Student Performance**

Objective:

- A school administrator wants to investigate whether **student performance** is affected by:
    1. **Teaching method** (Traditional vs. Online)
    2. **Student gender** (Male vs. Female)
- This leads to a **2x2 factorial design**:
- Factor A: Teaching Method (Traditional, Online)
- Factor B: Gender (Male, Female)
- Response Variable: Final Exam Score (numerical)

2-Way ANOVA

Gender	Teaching Method	Exam Score
Male	Traditional	75
Male	Traditional	78
Female	Traditional	85
Female	Traditional	82
Male	Online	70
Male	Online	68
Female	Online	80
Female	Online	83

	Traditional		Online	
	Male	Female	Male	Female
Score1	75	85	70	80
Score2	78	82	68	83

How to Select Test Type