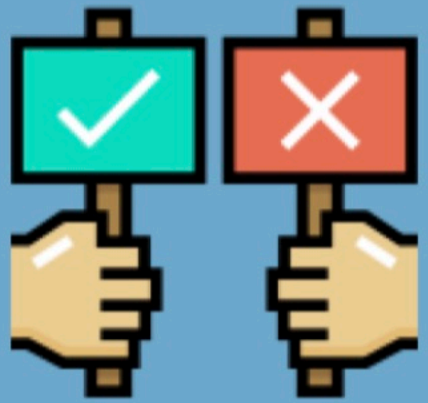
Statistics Session-6

CLARUSWAY©
WAY TO REINVENT YOURSELF

Did you finish Statistics (Hypothesis Tests) pre-class activity?



Students choose an option

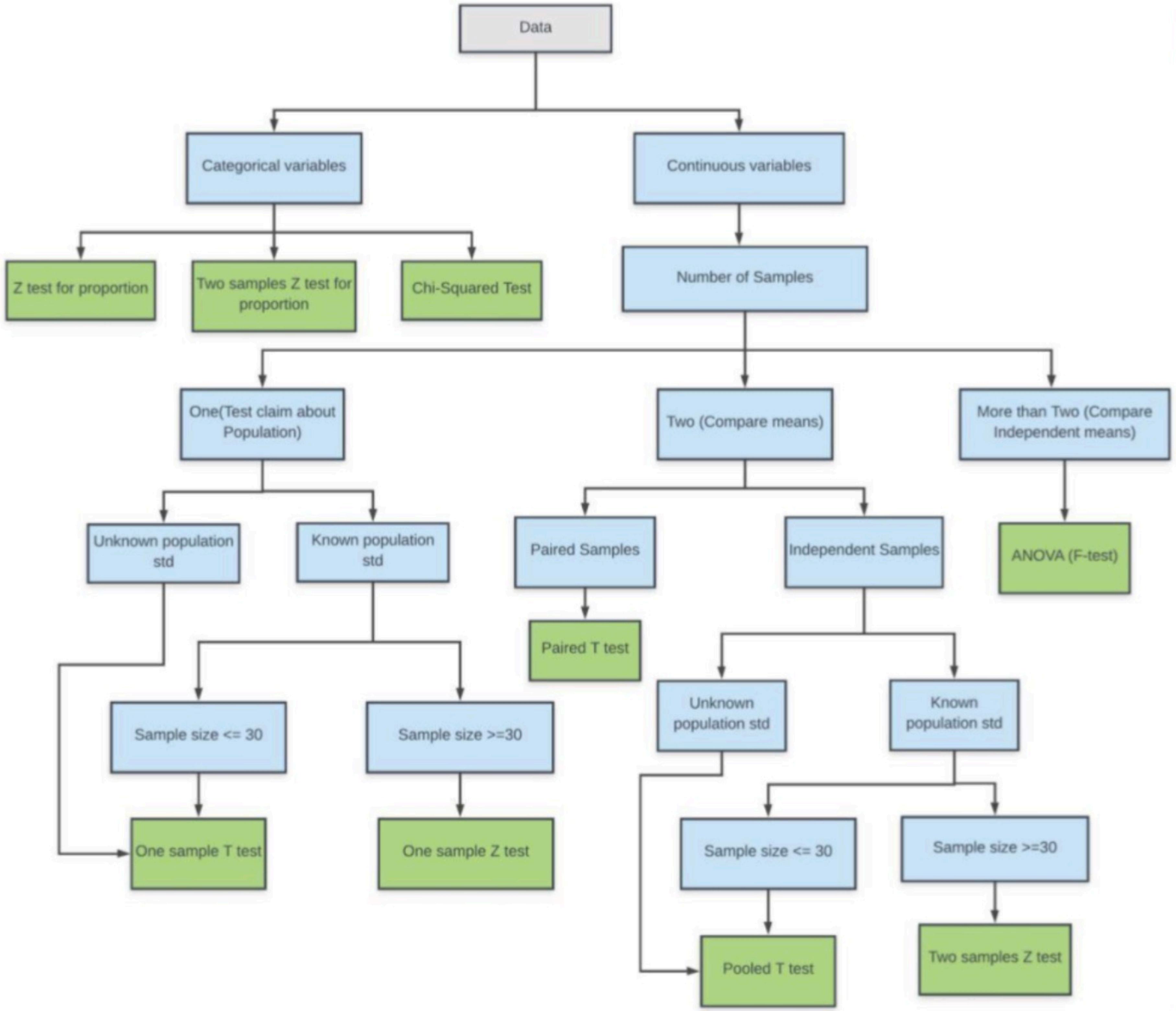
Pear Deck Interactive Slide
Do not remove this bar

No Multiple Choice Response

You didn't answer this question

Table of Contents

Test Types



CLARUSWAY©
WAY TO REINVENT YOURSELF

1 Basic Concepts Review

Independent Samples T Test

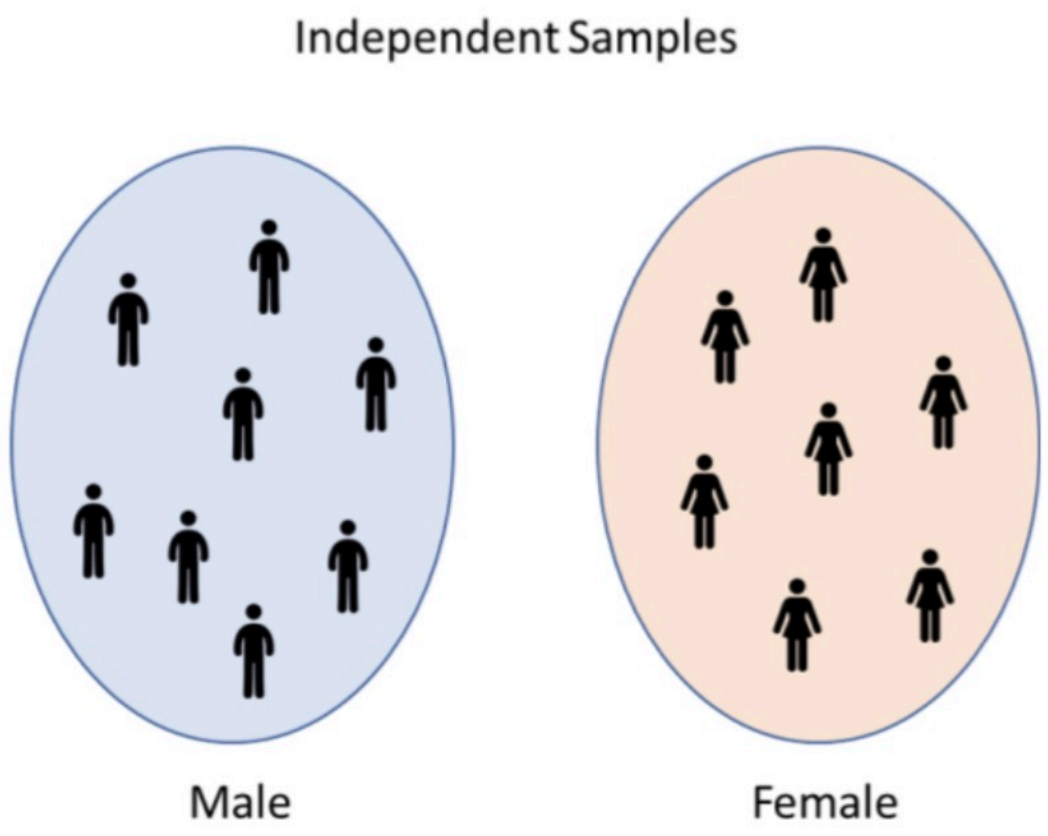
The independent-samples t-test (or independent t-test) compares the means between two independent groups on the same continuous, dependent variable.

- Does first year graduate salaries differ based on gender?

Dependent Variable

Two Independent Groups

Salary	Gender
$35,000	Male
$45,000	Male
$38,000	Female
$50,000	Female
$28,000	Male



The Steps of a Significance Test

A significance test has five steps.

- **Step 1** : Assumptions
- **Step 2** : Hypotheses
- **Step 3** : Test Statistic
- **Step 4** : P-Value
- **Step 5** : Conclusion

2

Hypothesis Test: Independent Samples Test

Independent Samples T Test

Assumptions

- A quantitative response variable for two groups
- Independent random samples
- Normal population distribution for each group.

Hypotheses

The null hypothesis	$H_0: \mu_1 = \mu_2$
The alternative hypothesis	$H_a: \mu_1 \neq \mu_2$

Test Statistic

Equal Variances NOT Assumed

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

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

Equal Variances Assumed

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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

$df = n_1 + n_2 - 2$

Independent Samples T Test - *Example*

- Under normal conditions, is the average body temperature the same for men and women?
- Medical researchers interested in this question collected data from a large number of men and women, and random samples from that data are presented in the accompanying table.
- Is there sufficient evidence to indicate that mean body temperatures differ for men and women?

Body Temperatures (°F)	
Men	Women
96.9	97.8
97.4	98.0
97.5	98.2
97.8	98.2
97.8	98.2
97.9	98.6
98.0	98.8
98.6	99.2
98.8	99.4

Step 1: Assumptions

★ The three basic assumptions of a test about difference between means are as follows:

Independent Samples T-Test Assumptions

1	A quantitative response variable for two groups
2	Independent random samples
3	Approximately normal population distribution for each group.

Step 2 : Hypotheses

★ Each significance test has two hypotheses about a population parameter: the null hypothesis and an alternative hypothesis.

The null hypothesis $H_0: \mu_1 = \mu_2$

The alternative hypothesis $H_a: \mu_1 \neq \mu_2$

The pooled estimate of σ

- Note: both s_1 and s_2 are estimators of σ.

- These can be combined to form a single estimator of σ, $\mathbf{s_{Pooled}}$.

- Sample Sizes
  - $n_1 = 9$
  - $n_2 = 9$
- Sample Std.Dev.
  - $s_1 = 0.5833$
  - $s_2 = 0.5487$

$$s_{Pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$s_p = \sqrt{\frac{(9-1)0.5833^2 + (9-1)0.5487^2}{9+9-2}} = 0.5663$$

Step 3 : Test Statistic

Calculation of test statistic

t

- $\bar{x}_1 = 97.856$, $\bar{x}_2 = 98.489$, $s_p = 0.5663$
- Test statistic: $t_0 = \dfrac{\bar{x}_1 - \bar{x}_2 - 0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$
- We find that: $t = \dfrac{\bar{x} - \bar{y} - \Delta_0}{s_{Pooled}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \dfrac{97.856 - 98.489 - 0}{0.5663 \times \sqrt{\dfrac{1}{9} + \dfrac{1}{9}}} = -2.371$
- df = $n_1 + n_2 - 2 = 9 + 9 - 2 = 16$

Step 4 : P-Value

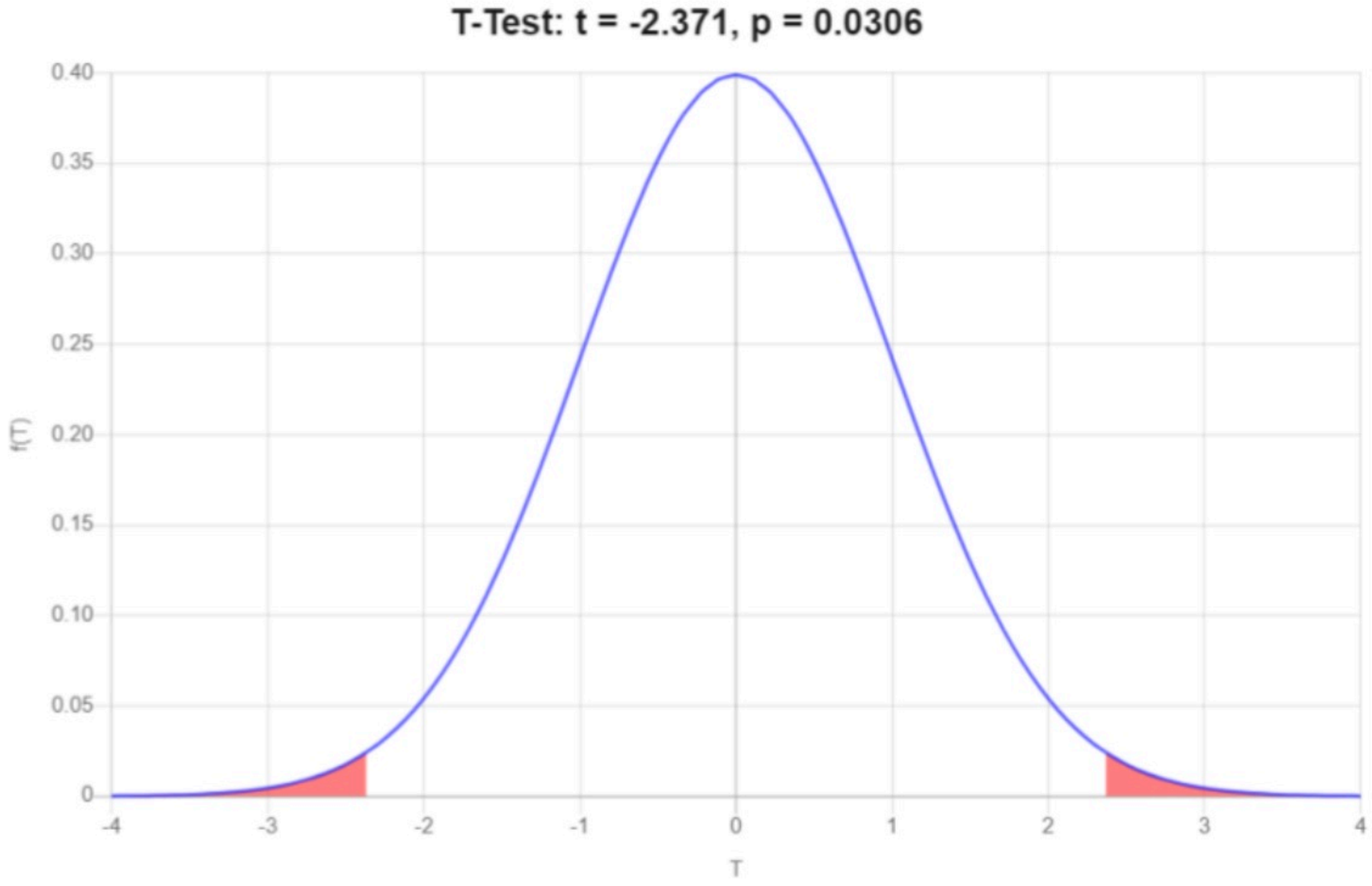
```
In [1]: import scipy.stats as stats

In [2]: 2*stats.t.cdf(-2.371, 16)

Out[2]: 0.030634485990952903
```

```
> 2*pt(-2.371,16)
[1] 0.0306345
```

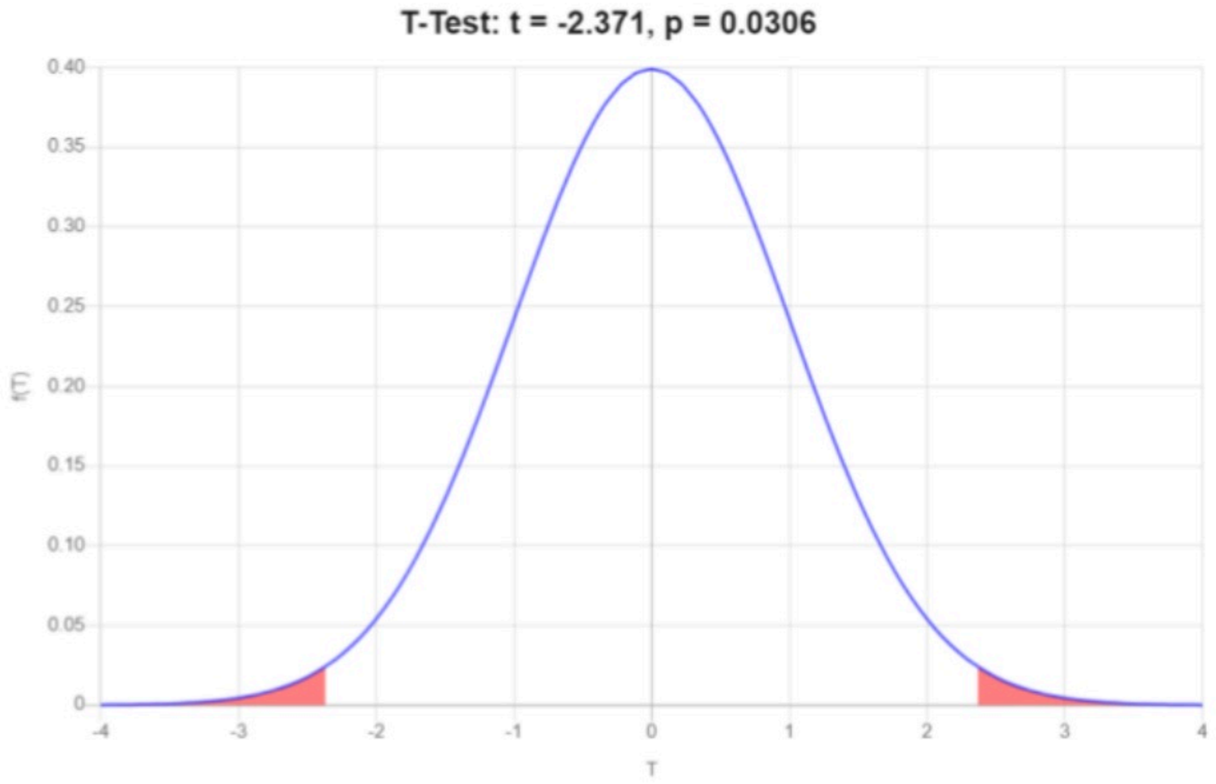
- P-value =
- $2^{*}t_{-2.371,\,16} = .0306$



Step 4 : P-Value

One line code

```
men=[96.9, 97.4, 97.5, 97.8, 97.8, 97.9, 98, 98.6, 98.8]
women = [97.8, 98, 98.2, 98.2, 98.2, 98.6, 98.8, 99.2, 99.4]
```




```
[11] # Calculate test statistics using stats.ttest_ind()
     indTest = stats.ttest_ind(men, women, equal_var = True)
     indTest
```

```
Ttest_indResult(statistic=-2.3724271468993643, pvalue=0.03054788637798765)
```

CLARUSWAY©
WAY TO REINVENT YOURSELF

Step 5 : Conclusion

- P-value < α
- We reject the null hypothesis.

- P-value = .0306
- Significance Level (α) = 0.05
- P-value < α   Reject the Null

- We have sufficient evidence to indicate that mean body temperatures differ for men and women. (at the α=0.05)


3
Hypothesis Test:
Dependent T Test
CLARUSWAY©
WAY TO REINVENT YOURSELF

Dependent T Test (Paired)

The dependent t-test (or the paired-samples t-test) compares the means between two related groups on the same continuous, dependent variable.

- Is there statistical difference in smokers' daily cigarette consumption before and after a 6 week hypnotherapy programme?

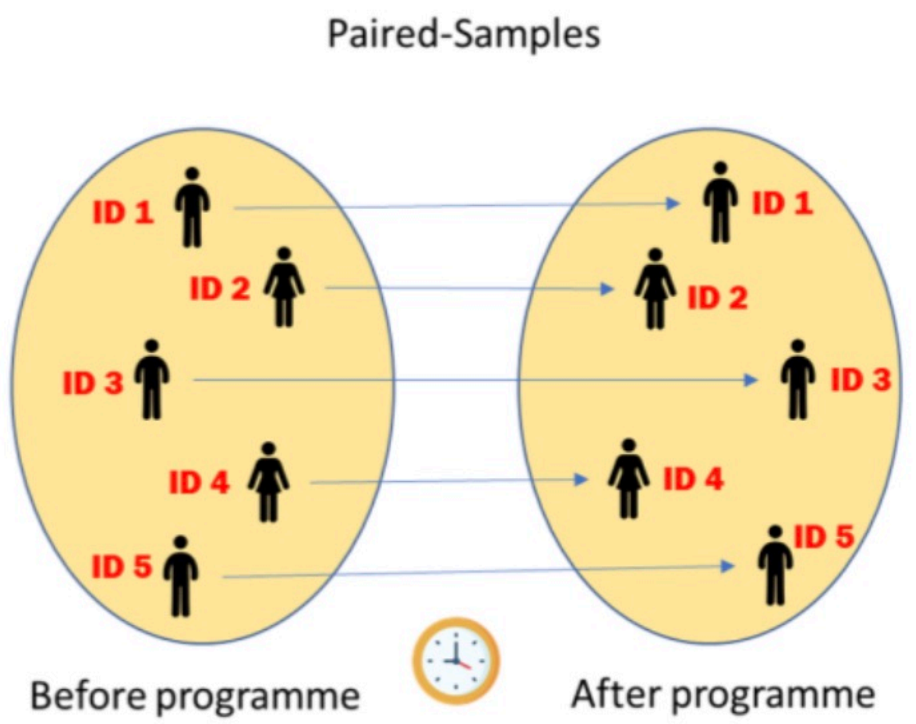
Participant ID

daily cigarette consumption before programme

daily cigarette consumption after programme

Before-After

ID	Before	After	Difference
1	12	10	2
2	18	7	11
3	23	22	1
4	10	12	-2
5	8	4	4



Dependent T Test

Assumptions

- The dependent variable is continuous, and it is measured twice on the same sample of subjects.
- The independent variable should consist of two categorical, "related groups" or "matched pairs".
- The difference between the scores of the variables are normally distributed.

Hypotheses

The null hypothesis	$H_0: \mu_D = 0$	$(\mu_D = \mu_1 - \mu_2)$
The alternative hypothesis	$H_a: \mu_D \neq 0$	

Test Statistic

$$t = \frac{\bar{x}_{diff} - 0}{s_{\bar{x}}}$$

$$s_{\bar{x}} = \frac{s_{diff}}{\sqrt{n}}$$

$\bar{x}_{\text{diff}}$ = Sample mean of the differences

n = Sample size (i.e., number of observations)

s_{diff} = Sample standard deviation of the differences

$s_{\bar{x}}$ = Estimated standard error of the mean (*s*/sqrt(*n*))

Dependent Samples T Test - *Example*

- An article* reports a comparison of several methods for predicting the shear strength for steel plate girders.
- Data for two of these methods, the Karlsruhe and Lehigh procedures, when applied to nine specific girders, are shown in the table below.
- Determine whether there is any difference (on the average) for the two methods. ($\alpha = 0.05$)

*Journal of Strain Analysis [1983, Vol. 18(2)]

Girder	Karlsruhe Method	Lehigh Method	Difference d_j
S1/1	1.186	1.061	0.125
S2/1	1.151	0.992	0.159
S3/1	1.322	1.063	0.259
S4/1	1.339	1.062	0.277
S5/1	1.2	1.065	0.135
S2/1	1.402	1.178	0.224
S2/2	1.365	1.037	0.328
S2/3	1.537	1.086	0.451
S2/4	1.559	1.052	0.507

Step 1: Assumptions

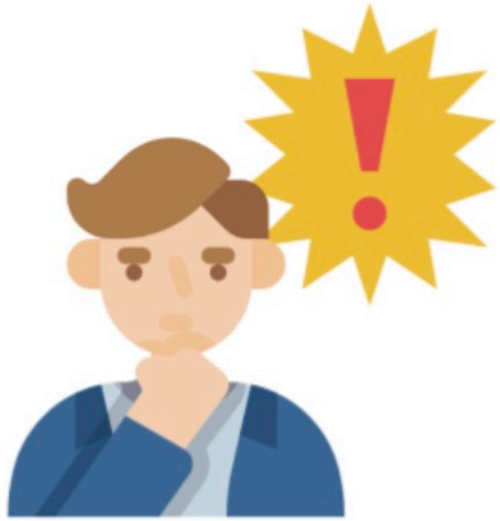
★ The three basic assumptions of a test about difference between means are as follows:

Paired Samples T-Test Assumptions

1	The dependent variable is continuous.
2	Observations are independent of one another.
3	Approximately normal population distribution for each group.

Step 2 : Hypotheses

★ Each significance test has two hypotheses about a population parameter: the null hypothesis and an alternative hypothesis.



The null hypothesis $H_0: \mu_D = 0$

The alternative hypothesis $H_a: \mu_D \neq 0$

Step 3 : Test Statistic

Calculation of test statistic

- $\bar{d}$= 0.2769, s_d = 0.1350, n=9
- Test statistic: $t_0 = \dfrac{\bar{d}}{s_d / \sqrt{n}}$
- We find that: $t_0 = \dfrac{\bar{d}}{s_d/\sqrt{n}} = \dfrac{0.2769}{0.1350/\sqrt{9}} = 6.15$

Step 4 : P-Value

```
In [1]: import scipy.stats as stats
```

```
In [2]: 2*(1-stats.t.cdf(6.15, 8))
Out[2]: 0.00027399606897193785
```

- P-value = 0.0003

Step 4 : P-Value

One line code

```
karlsruhe=[1.186, 1.151, 1.322, 1.339, 1.2, 1.402, 1.365, 1.537, 1.559]
lehigh=[1.061, 0.992, 1.063, 1.062, 1.065, 1.178, 1.037, 1.086, 1.052]
```

```
# Calculate test statistics using stats.ttest_rel()
pairedtest = stats.ttest_rel(karlsruhe, lehigh)
pairedtest
```

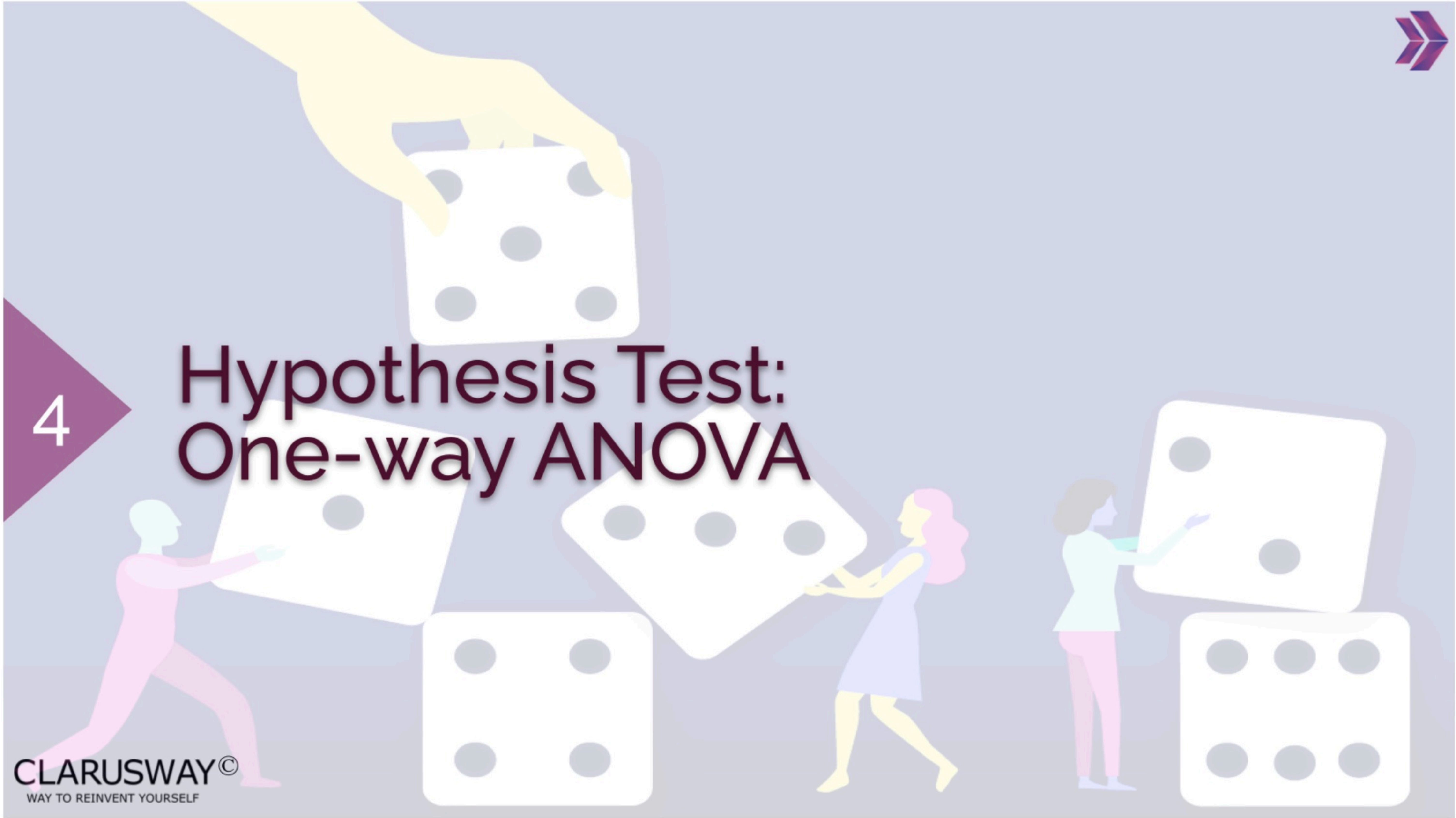
```
Ttest_relResult(statistic=6.0819394375848255, pvalue=0.00029529546278604066)
```

Step 5 : Conclusion

- P-value < α
- We reject the null hypothesis.

- P-value = .0003
- Significance Level (α) = 0.05
- P-value < α   Reject the Null

- The data indicate that the Karlsruhe method produces, on the average, higher strength predictions than does the Lehigh method.

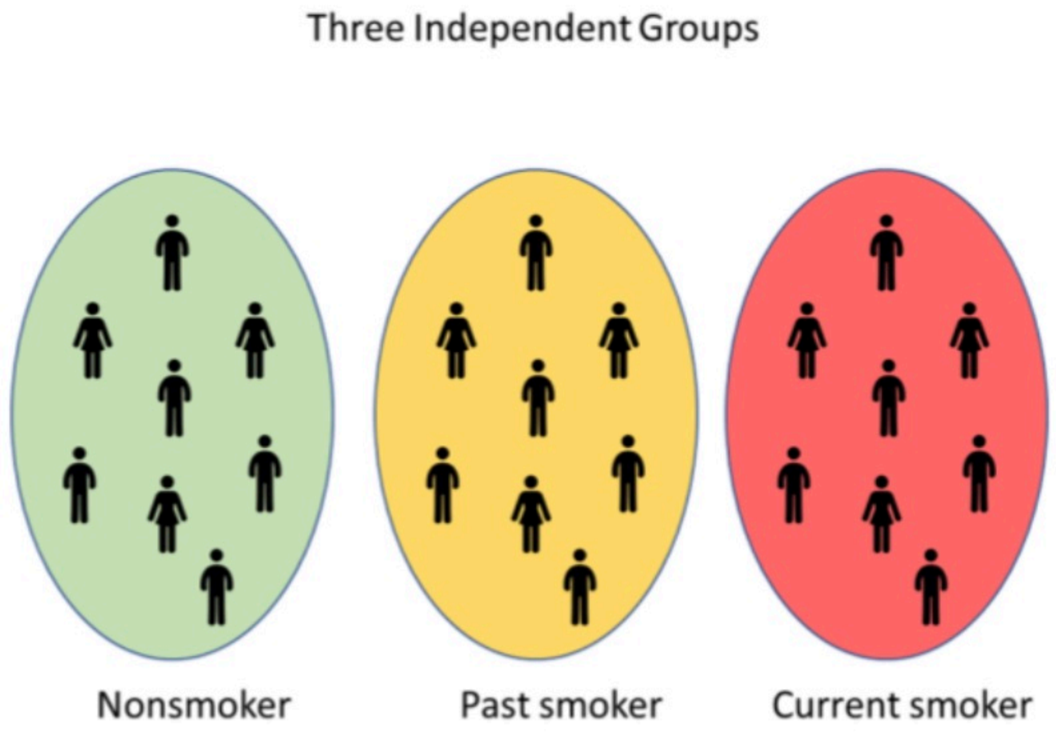

4
Hypothesis Test:
One-way ANOVA
CLARUSWAY©
WAY TO REINVENT YOURSELF

One-way ANOVA

Both the One-Way ANOVA and the Independent Samples t Test can compare the means for two groups. However, only the One-Way ANOVA can compare the means across three or more groups.

- Is there a statistically significant difference in sprint time with respect to smoking status? (0 = Nonsmoker, 1 = Past smoker, 2 = Current smoker)

Dependent Variable	Independent Groups
Sprint	**Smoking**
5.1	0
7.8	2
7.1	1
8.6	2
4.9	0
7.7	1



One-way ANOVA

Assumptions

- Dependent variable that is continuous
- Independent variable that is categorical (two or more groups)
- Independent samples/groups
- Normal distribution of the DV for each group
- Homogeneity of variances

Hypotheses

The null hypothesis	H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ (all k population means are equal)
The alternative hypothesis	H_a: At least one μ_i different

Test Statistic *ANOVA Table*

$SSR = \Sigma(\hat{y}_i - \bar{y})^2$

regression sum of squares

model degrees of freedom
k: group number

regression mean square

F statistic

	Sum of Squares	df	Mean Square	F
Group (Between)	SSR	k-1	MSR = SSR/(k-1)	MSR/MSE
Error (Within)	SSE	n-k	MSE = SSE/(n-k)	
Total	SST = SSR+SSE	n-1		

error sum of squares

$SSE = \Sigma(\hat{y}_i - y_i)^2$

total sum of squares

error degrees of freedom

total degrees of freedom

mean square error

One-way ANOVA - *Scenarios*

★ Variation in Time to Relief of Symptoms Between and Within 3 Treatment

★ Music compressed by four MP3 compressors are with the same quality

★ Three new drugs are all as effective as a placebo

★ Lectures, group studying, and computer assisted instruction are equally effective

One-way ANOVA - *Example*

- As an example, we analyze the Cushings data set, which is available from the MASS package.
- The Type variable in the data set shows the underlying type of syndrome, which can be one of four categories:
  - adenoma (a),
  - bilateral hyperplasia (b),
  - carcinoma (c),
  - unknown (u).

- Our objective is to find whether the four groups are different with respect to urinary excretion rate of Tetrahydrocortisone.

One-way ANOVA - *Example*

- The total number of observations is $n=27$,
- The number of observations in each group is
  - $n_1 = 6$, $n_2 = 10$, $n_3 = 5$, and $n_4 = 6$.
- We also find the group specific means, which are
  - 3.0, 8.2, 19.7, and 14.0, respectively.
- The degrees of freedom parameters are
  - $df_1 = 4-1=3$ (k-1) and $df_2 = 27-4=23$ (n-k).
  - (k: number of groups)
- $SS_B = 893.5$ and $SS_W = 2123.6$.

Step 1: Assumptions

★ The basic assumptions of one-way ANOVA test are as follows:

One-way ANOVA Assumptions

1. The dependent variable is continuous.
2. Independent variable that is categorical.
3. Approximately normal population distribution for each group.
4. Homogeneity of variances

Step 2: Hypotheses

★ The null and alternative hypotheses of one-way ANOVA can be expressed as:

The null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

The alternative hypothesis H_a: At least one μ_i different

Step 3: Test Statistic

F

ANOVA Table

	Sum of Squares	df	Mean Square	F
Group (Between)	SSR = 893.5	k-1 = 3	MSR = 893.5 / 3 = 297.8	MSR/MSE = 297.8 / 92.3 = 3.226
Error (Within)	SSE = 2123.6	n-k = 23	MSE = 2123.6 / 23 = 92.3	
Total	SST = 3017.1	n-1 = 27		

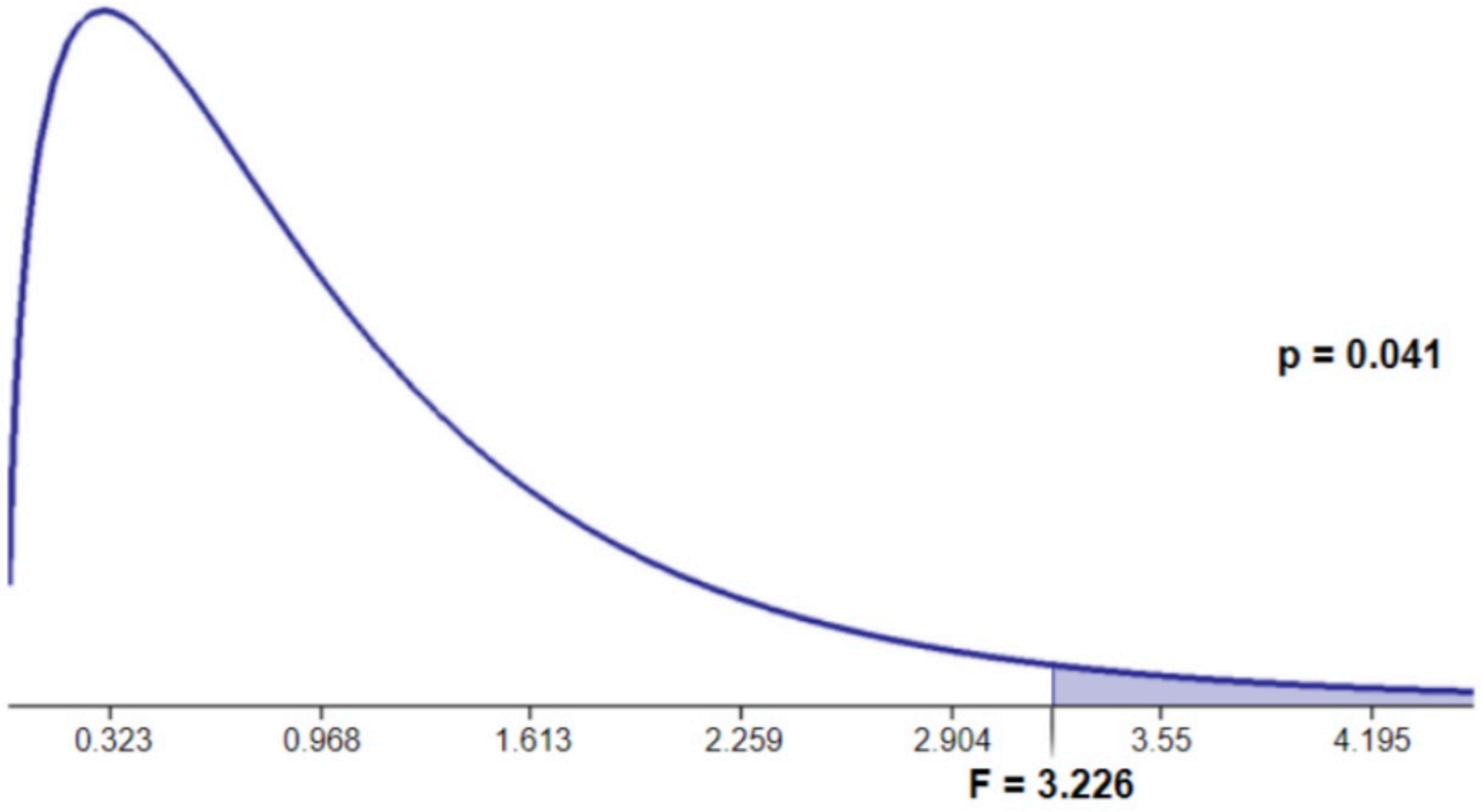
Step 4 : P-Value

```
In [1]: import scipy.stats as stats
```

```
In [2]: 1 - stats.f.cdf(3.226, dfn=3, dfd=23)
```

```
Out[2]: 0.04120786265996445 6
```

- F statistic = 3.226
- P-value = 0.041



Step 4 : P-Value

One line code

	Tetrahydrocortisone	Pregnanetriol	Type
a1	3.1	11.70	a
a2	3.0	1.30	a
a3	1.9	0.10	a
a4	3.8	0.04	a
a5	4.1	1.10	a

```
df.Type.unique()
```

```
array(['a', 'b', 'c', 'u'], dtype=object)
```

```
f_oneway((df.Tetrahydrocortisone[df.Type=="a"]), (df.Tetrahydrocortisone[df.Type=="b"]),
         (df.Tetrahydrocortisone[df.Type=="c"]), (df.Tetrahydrocortisone[df.Type=="u"]))
```

```
F_onewayResult(statistic=3.2257394791378426, pvalue=0.0412182793672776)
```

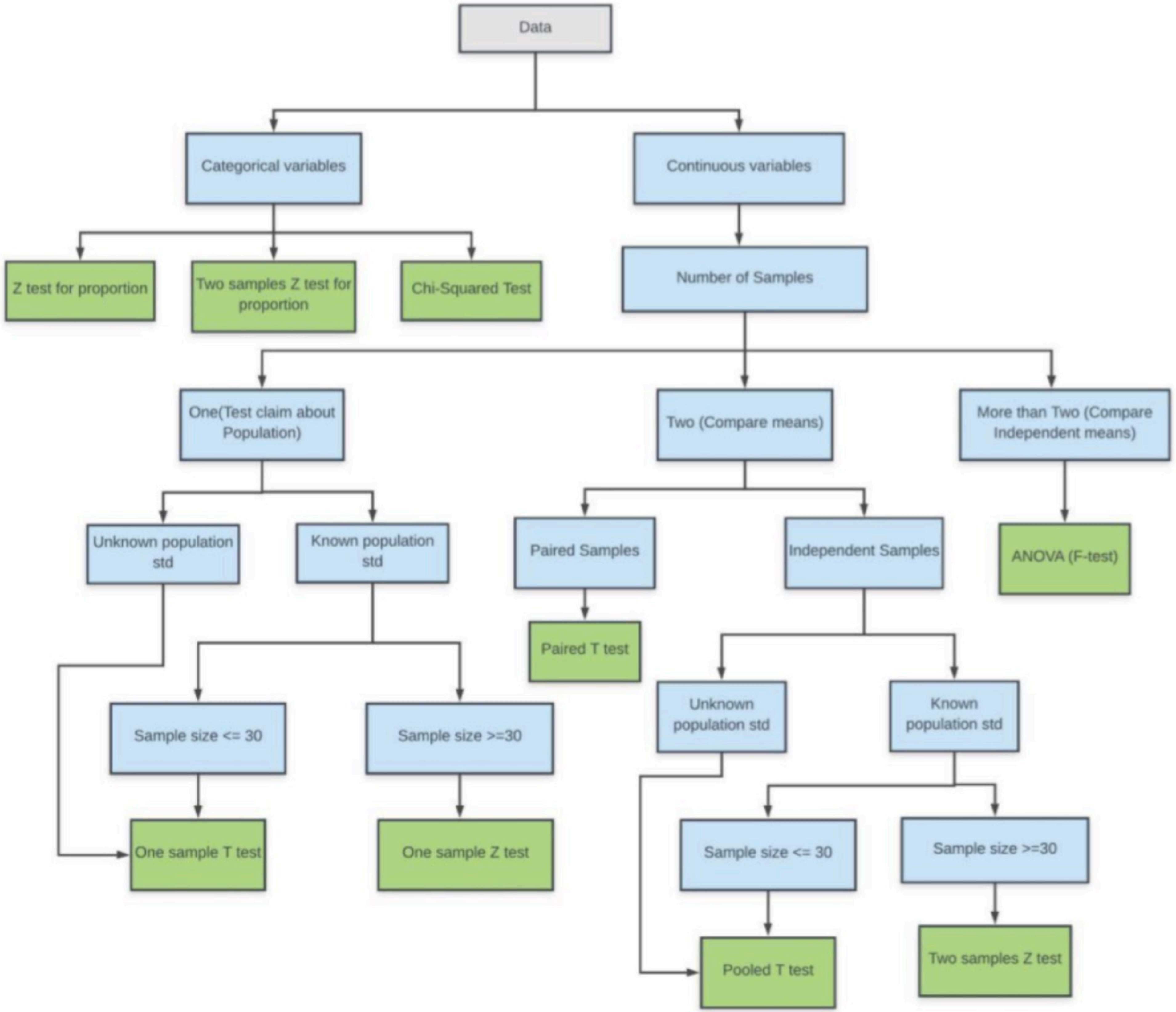
Step 5 : Conclusion

- P-value < α
- We reject the null hypothesis.

- P-value = 0.041
- Significance Level (α) = 0.05
- P-value < α

Reject the Null

- Therefore, we can reject H_0 at 0.05 significance level and conclude that the differences among group means are statistically significant

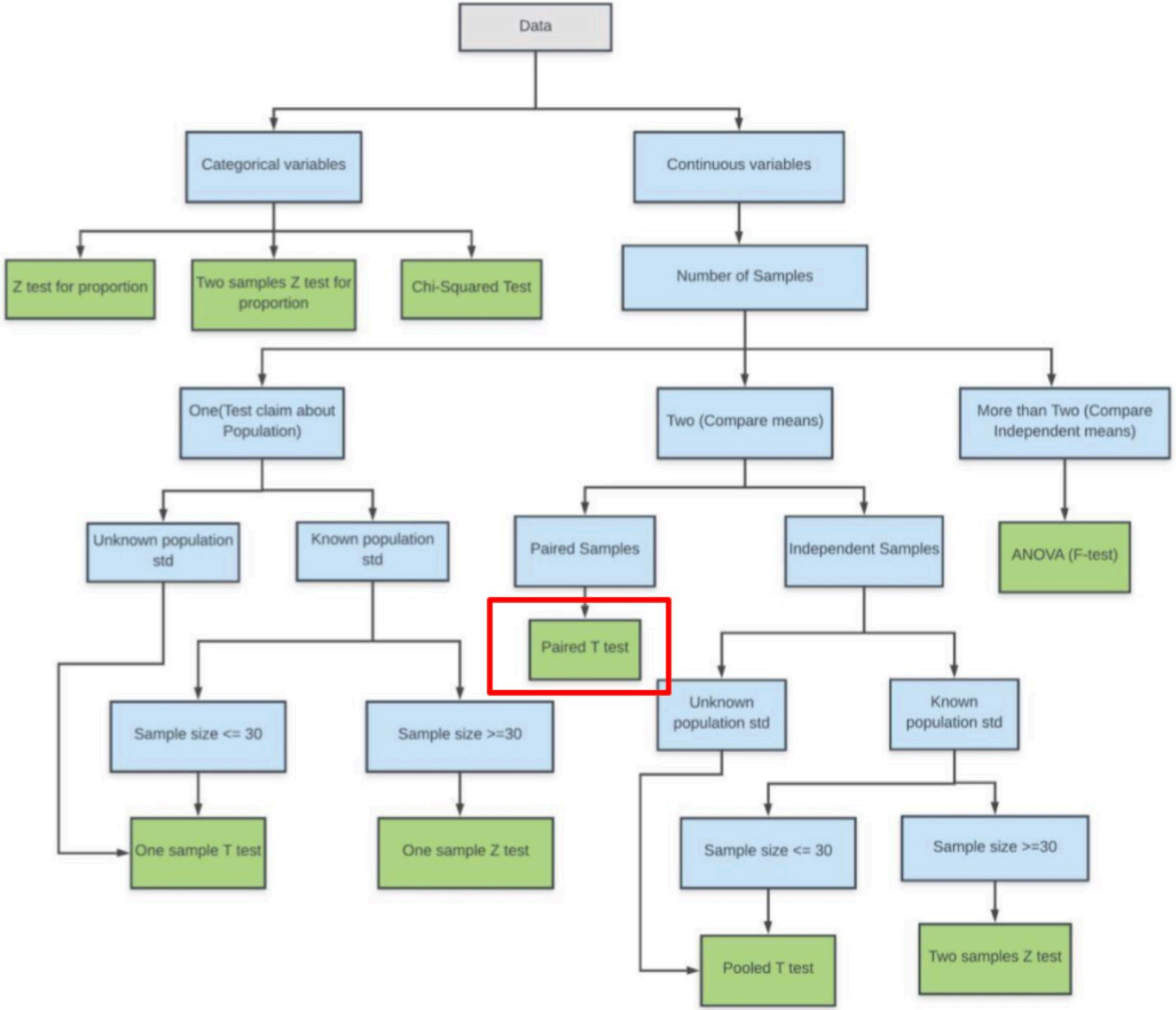


Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

- **(B)** Is there a significant mean difference between Wednesday and Saturday gas prices if we check the same 20 stations on both days?

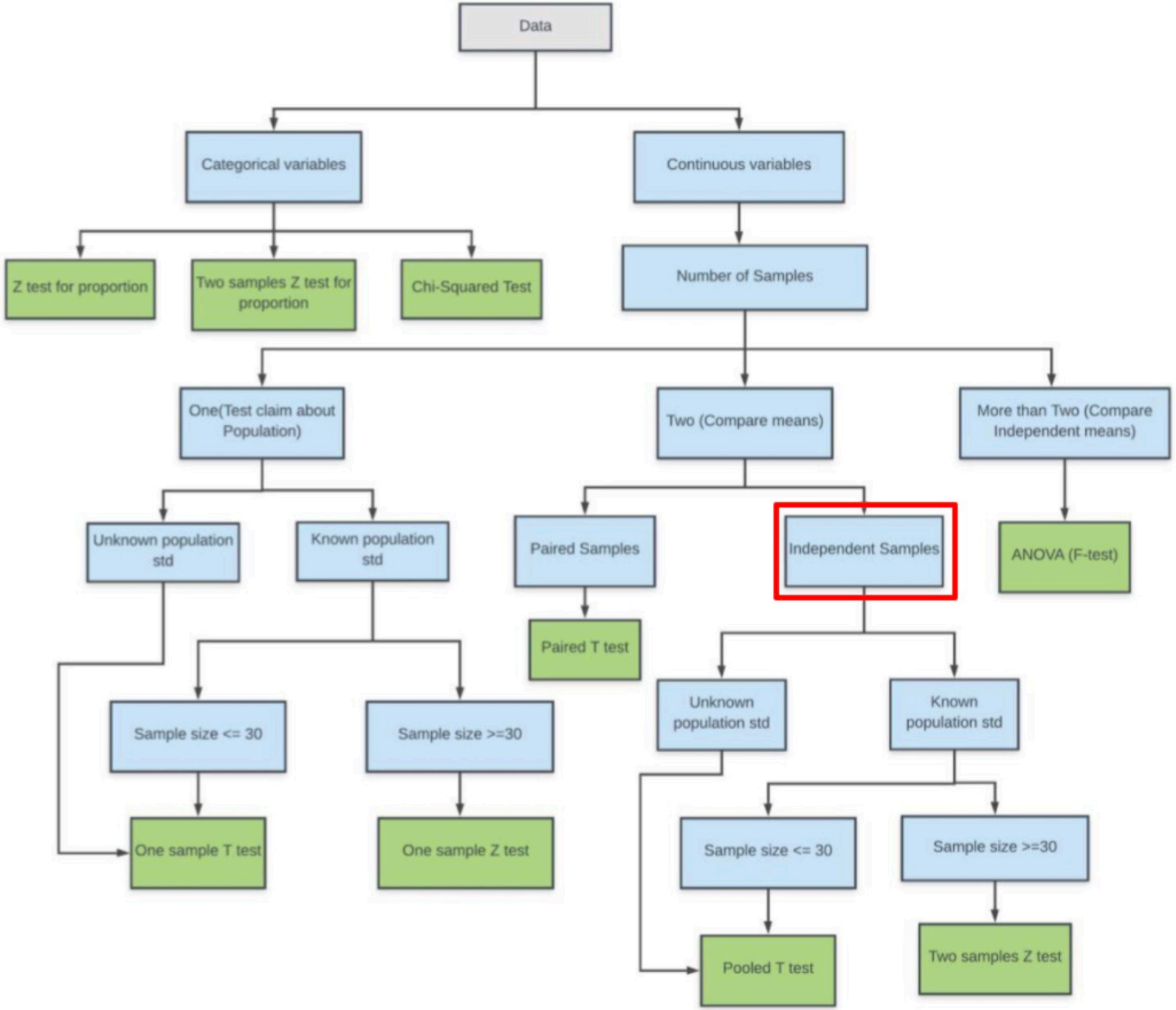
S O L U T I O N


Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

- **(C)** Is there a significant difference between the average Boston gas price and the average New York gas price today?

S O L U T I O N

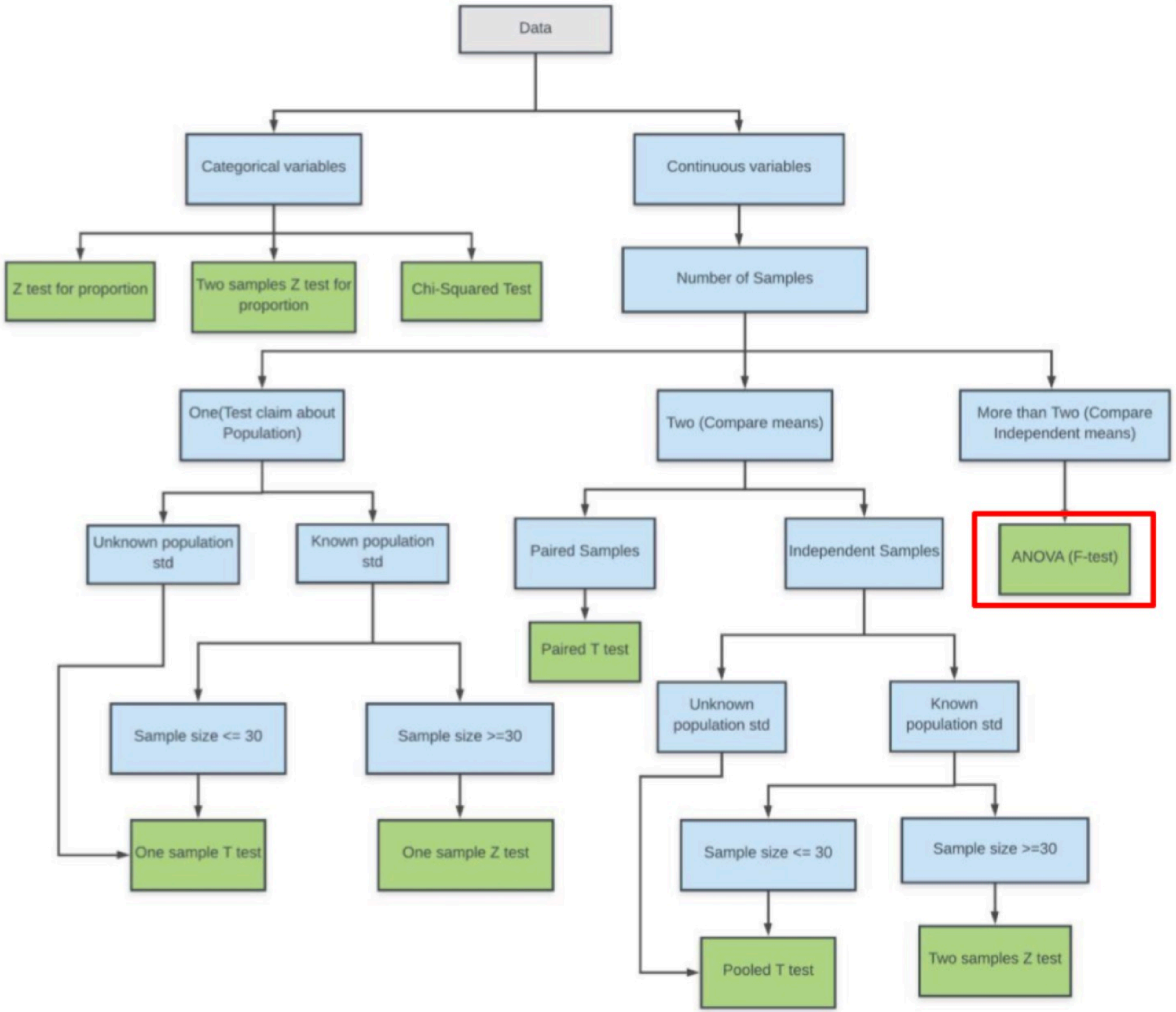


Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

- **(H)** Is there a difference in the mean mileage for the four populations of vehicle size (small cars, large cars, trucks, and SUVs)?

S O L U T I O N



Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

- **(C)** Are the average delays for two airlines are *different?*

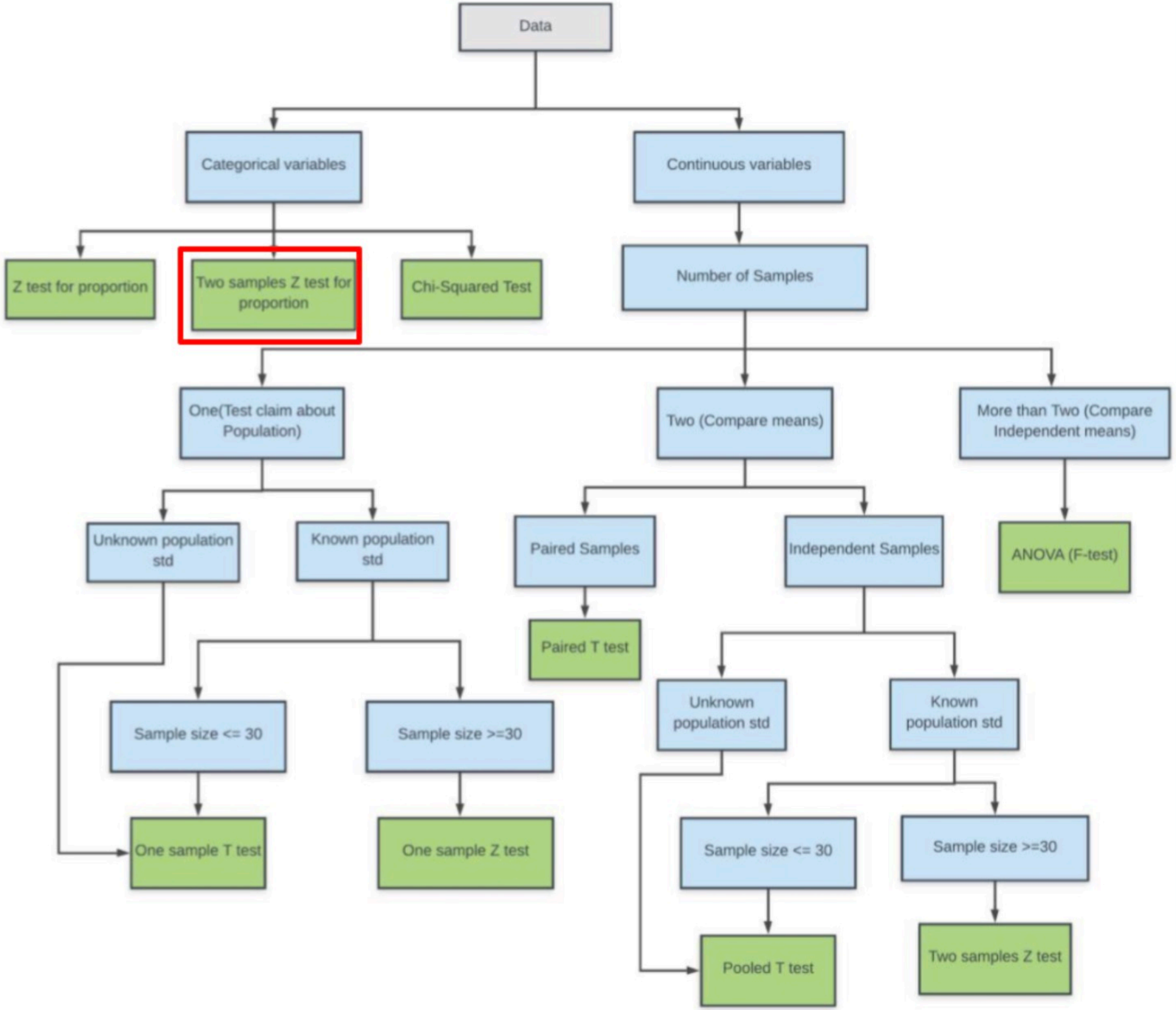
S O L U T I O N


Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Unknown population std
Known population std
Sample size <= 30
Sample size >=30
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

- **(E)** Is there a difference in proportions of male and female students who missed at least one discussion during this semester?

SOLUTION

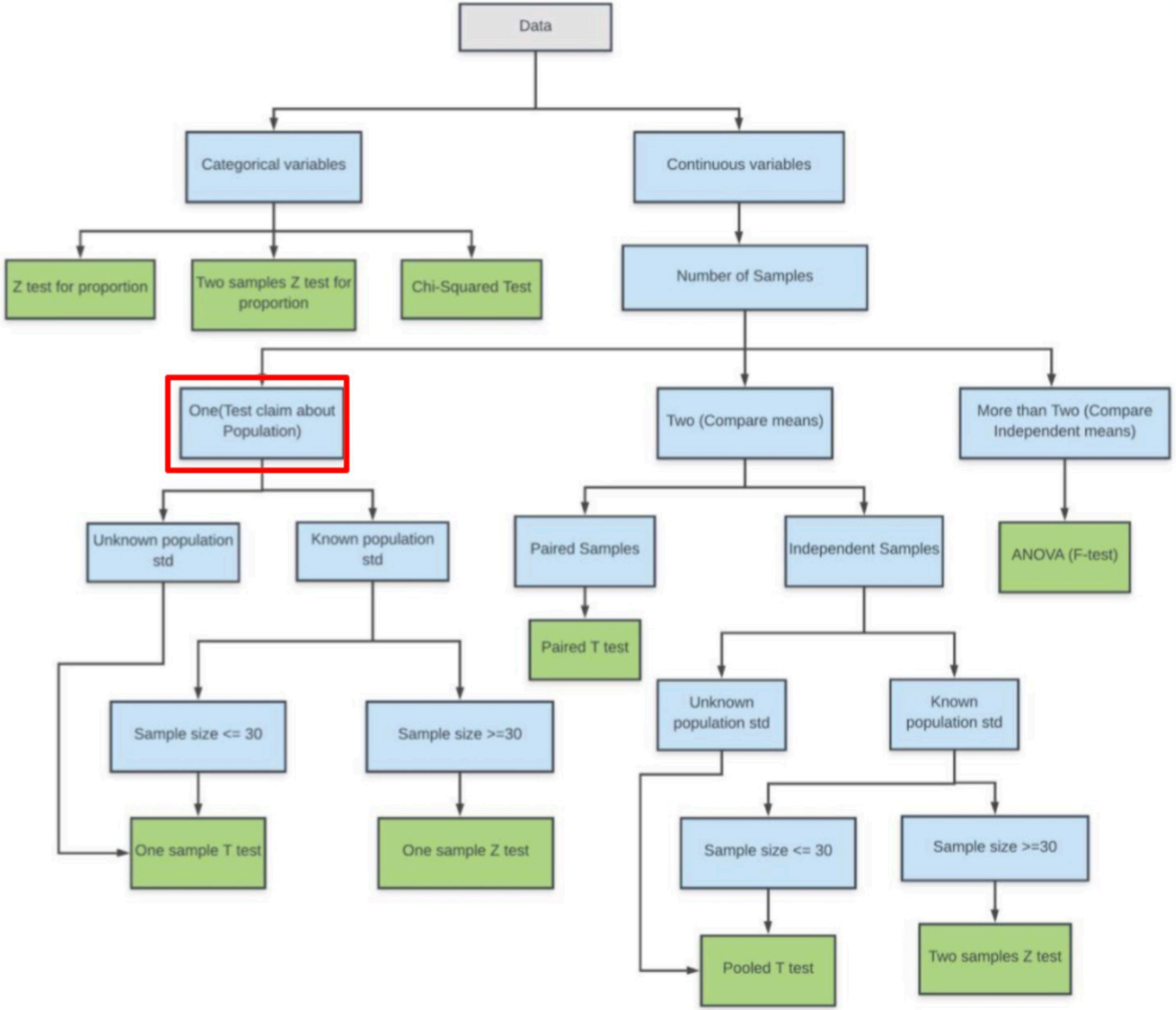

Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

Q: Ted Copper of the local police unit is concerned about the safety of drivers along Interstate 93. He would like your help in deciding which statistical test might be used for each of his questions. Read each of the research questions in parts (a) through (e) below and then write the appropriate statistical test:

- The police are interested in seeing how fast drivers typically go. The speed limit is 75 mph, but the officer suspects that they go faster.
- A large sample of male and female drivers is taken and each driver is asked the maximum mph over the speed limit they have ever gone. The police suspect that males will have driven faster than females, but want to test their theory.
- A sample of drivers is asked how fast they drove on Friday (max speed) and how fast they drove on Thursday. The police want to investigate if drivers head to their destinations faster on the weekend.
- A sample of 60 drivers is taken and each person is asked if they have ever used a cell phone while driving. Police are curious as to whether or not the rate has increased over a 60% rate from two years ago.
- A large sample of young, middle and older drivers is taken and each driver is asked how fast they drove. The police suspect that young drivers are faster than the others, but want to test their theory.

S O L U T I O N

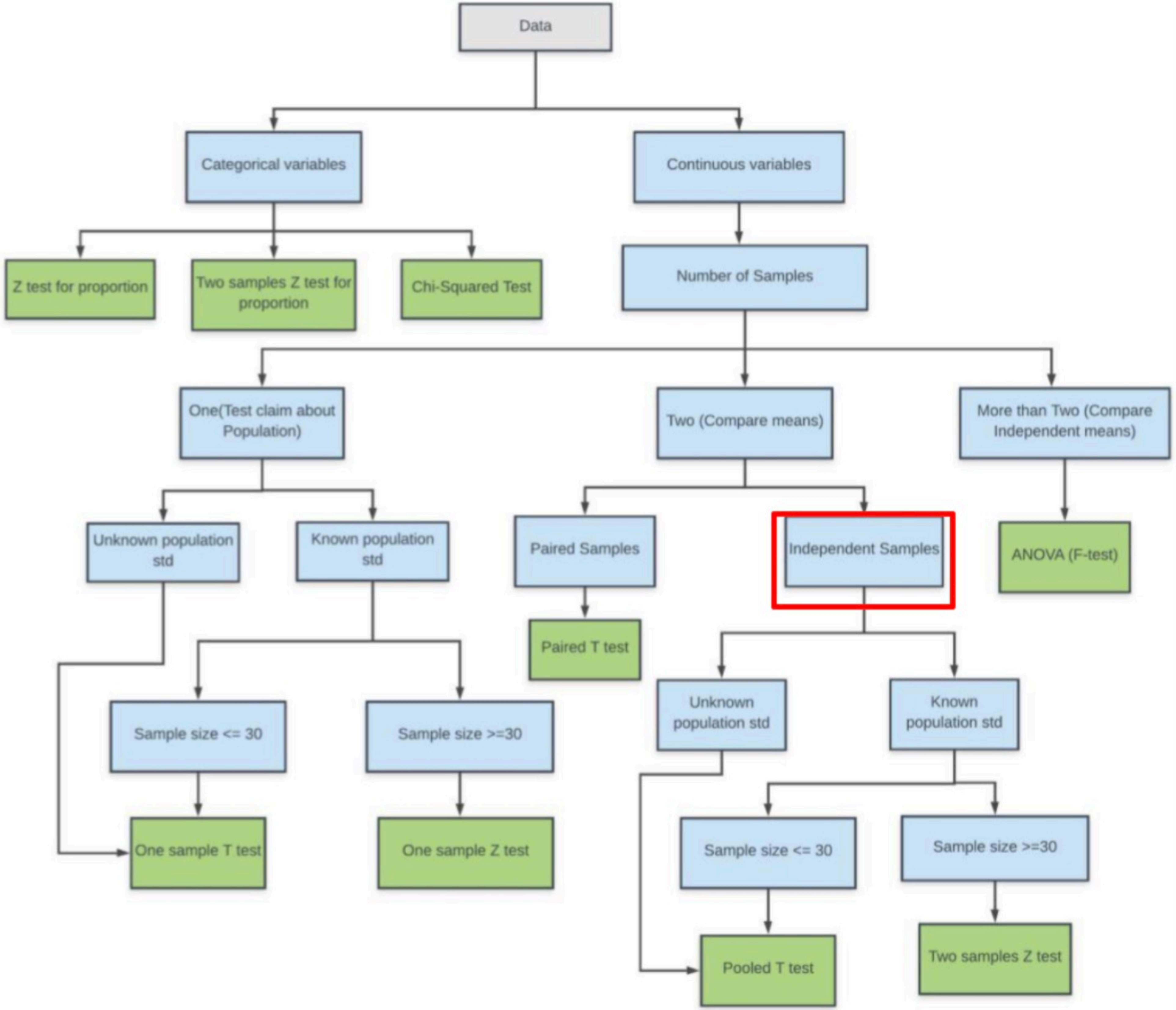

Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

Q: Ted Copper of the local police unit is concerned about the safety of drivers along Interstate 93. He would like your help in deciding which statistical test might be used for each of his questions. Read each of the research questions in parts (a) through (e) below and then write the appropriate statistical test:

- The police are interested in seeing how fast drivers typically go. The speed limit is 75 mph, but the officer suspects that they go faster.
- A large sample of male and female drivers is taken and each driver is asked the maximum mph over the speed limit they have ever gone. The police suspect that males will have driven faster than females, but want to test their theory.
- A sample of drivers is asked how fast they drove on Friday (max speed) and how fast they drove on Thursday. The police want to investigate if drivers head to their destinations faster on the weekend.
- A sample of 60 drivers is taken and each person is asked if they have ever used a cell phone while driving. Police are curious as to whether or not the rate has increased over a 60% rate from two years ago.
- A large sample of young, middle and older drivers is taken and each driver is asked how fast they drove. The police suspect that young drivers are faster than the others, but want to test their theory.

S O L U T I O N

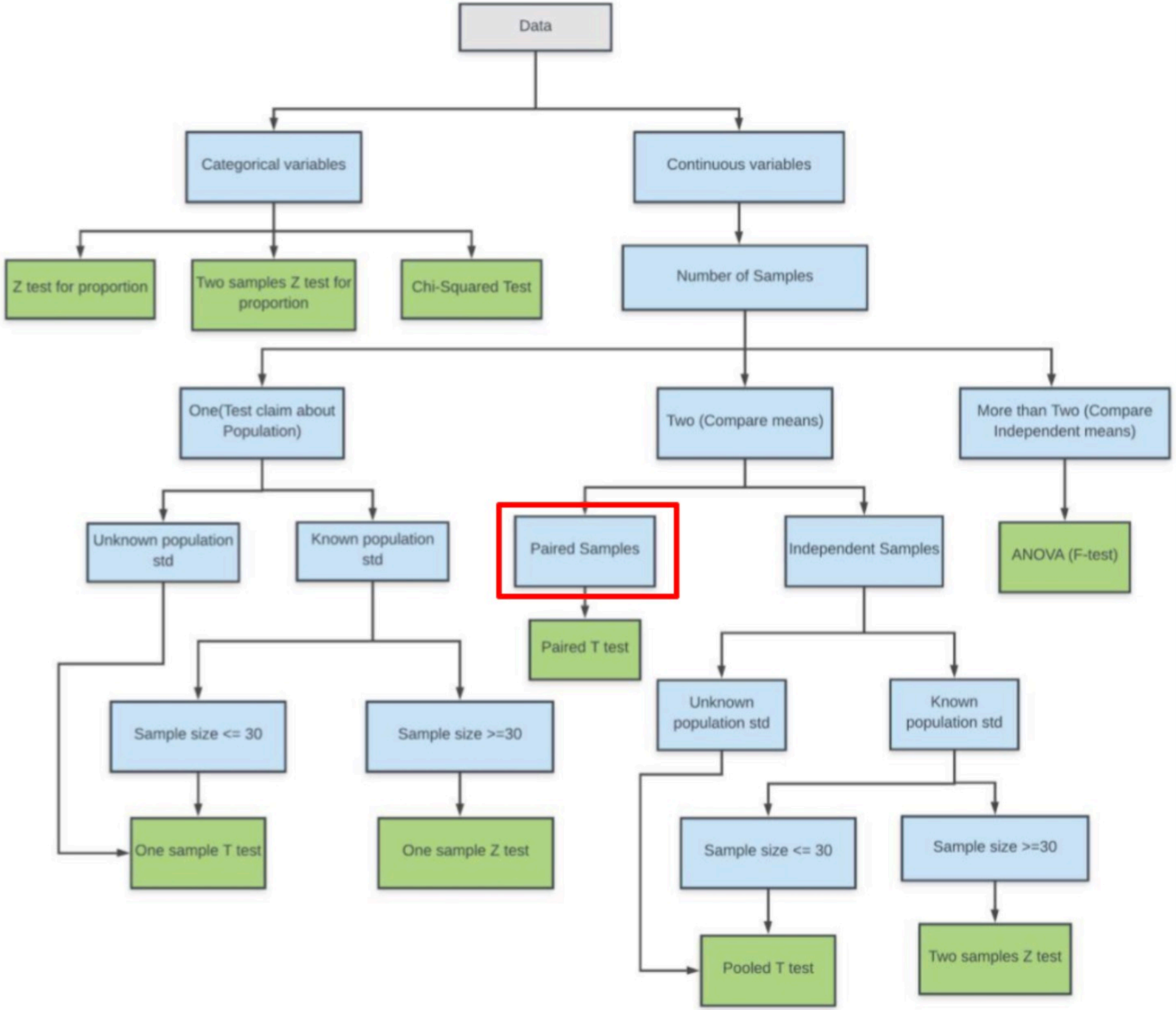

Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

Q: Ted Copper of the local police unit is concerned about the safety of drivers along Interstate 93. He would like your help in deciding which statistical test might be used for each of his questions. Read each of the research questions in parts (a) through (e) below and then write the appropriate statistical test:

- The police are interested in seeing how fast drivers typically go. The speed limit is 75 mph, but the officer suspects that they go faster.
- A large sample of male and female drivers is taken and each driver is asked the maximum mph over the speed limit they have ever gone. The police suspect that males will have driven faster than females, but want to test their theory.
- A sample of drivers is asked how fast they drove on Friday (max speed) and how fast they drove on Thursday. The police want to investigate if drivers head to their destinations faster on the weekend.
- A sample of 60 drivers is taken and each person is asked if they have ever used a cell phone while driving. Police are curious as to whether or not the rate has increased over a 60% rate from two years ago.
- A large sample of young, middle and older drivers is taken and each driver is asked how fast they drove. The police suspect that young drivers are faster than the others, but want to test their theory.

SOLUTION

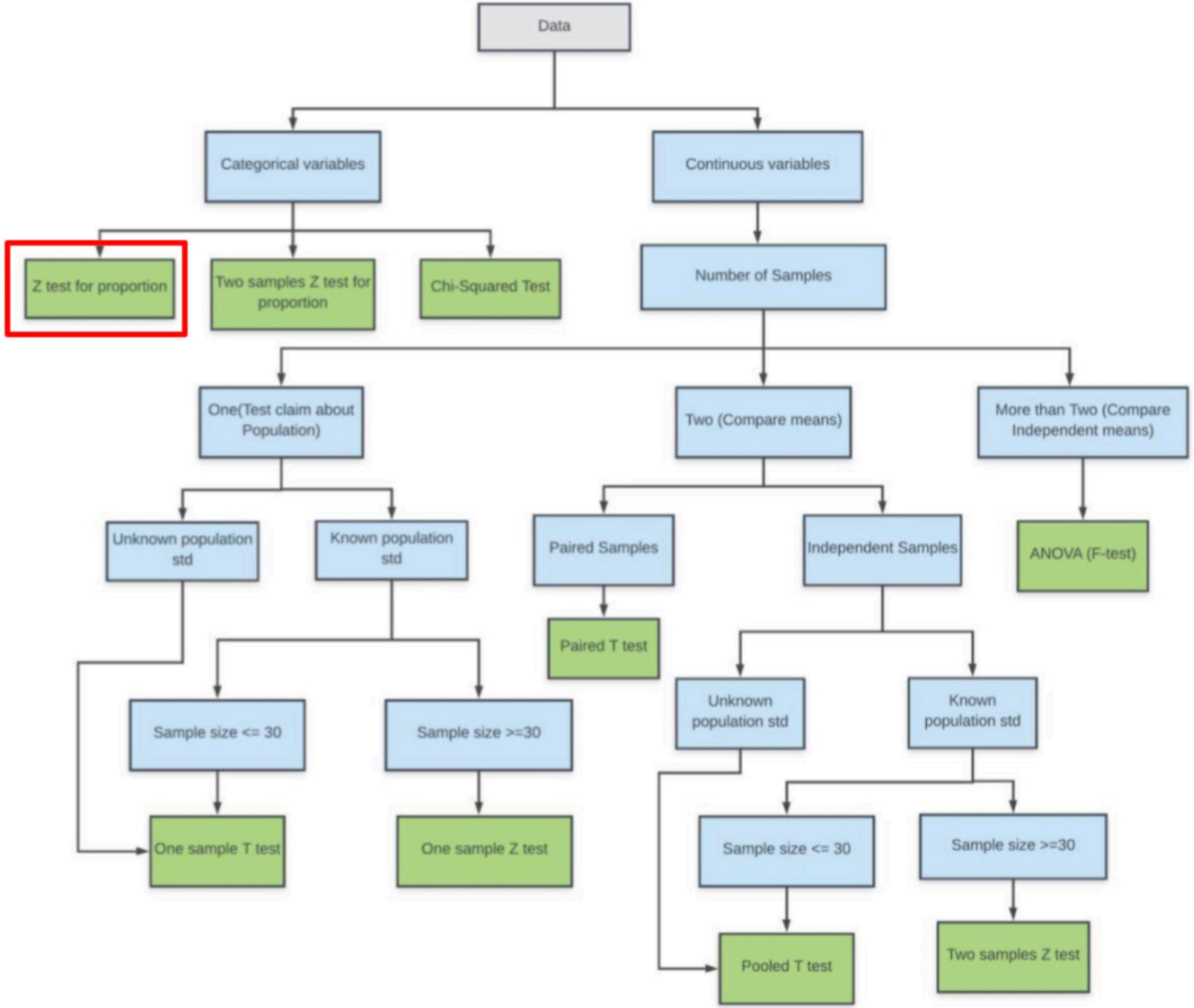

Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

LETS PRACTICE

Q: Ted Copper of the local police unit is concerned about the safety of drivers along Interstate 93. He would like your help in deciding which statistical test might be used for each of his questions. Read each of the research questions in parts (a) through (e) below and then write the appropriate statistical test:

- The police are interested in seeing how fast drivers typically go. The speed limit is 75 mph, but the officer suspects that they go faster.
- A large sample of male and female drivers is taken and each driver is asked the maximum mph over the speed limit they have ever gone. The police suspect that males will have driven faster than females, but want to test their theory.
- A sample of drivers is asked how fast they drove on Friday (max speed) and how fast they drove on Thursday. The police want to investigate if drivers head to their destinations faster on the weekend.
- A sample of 60 drivers is taken and each person is asked if they have ever used a cell phone while driving. Police are curious as to whether or not the rate has increased over a 60% rate from two years ago.
- A large sample of young, middle and older drivers is taken and each driver is asked how fast they drove. The police suspect that young drivers are faster than the others, but want to test their theory.

S O L U T I O N



Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Unknown population std
Known population std
Sample size <= 30
Sample size >=30
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test

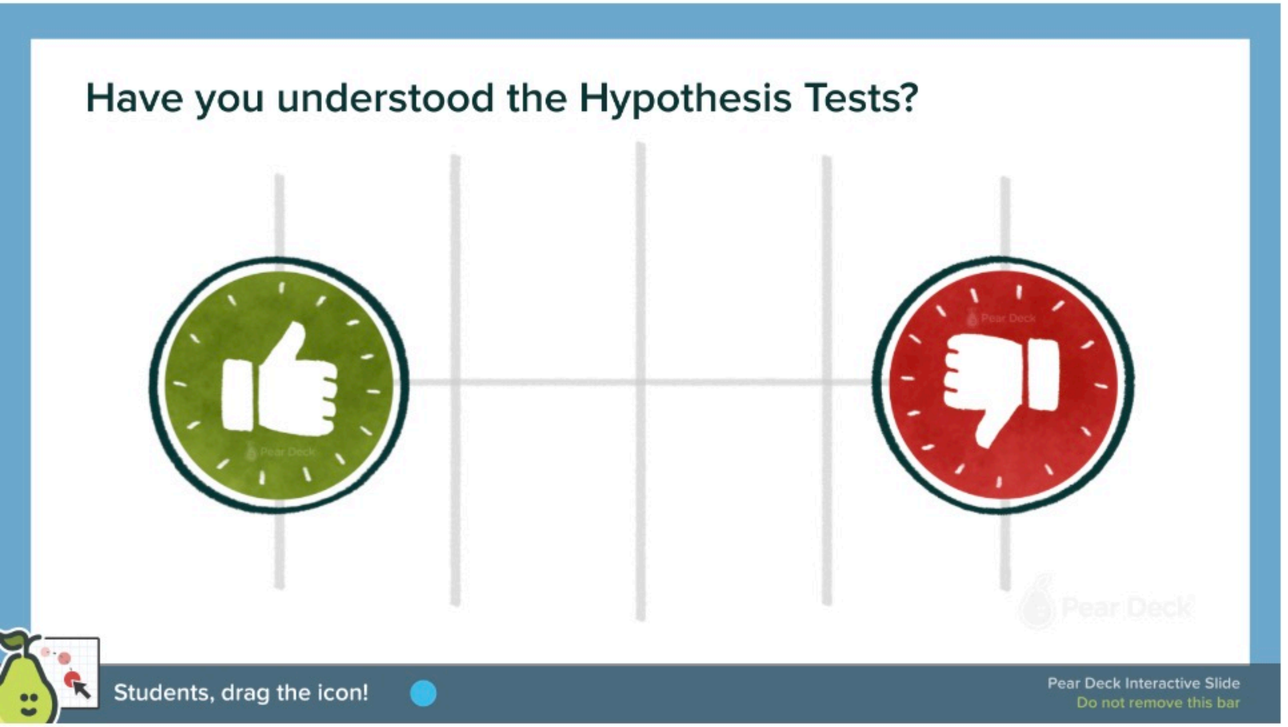
LETS PRACTICE

Q: Ted Copper of the local police unit is concerned about the safety of drivers along Interstate 93. He would like your help in deciding which statistical test might be used for each of his questions. Read each of the research questions in parts (a) through (e) below and then write the appropriate statistical test:

- The police are interested in seeing how fast drivers typically go. The speed limit is 75 mph, but the officer suspects that they go faster.
- A large sample of male and female drivers is taken and each driver is asked the maximum mph over the speed limit they have ever gone. The police suspect that males will have driven faster than females, but want to test their theory.
- A sample of drivers is asked how fast they drove on Friday (max speed) and how fast they drove on Thursday. The police want to investigate if drivers head to their destinations faster on the weekend.
- A sample of 60 drivers is taken and each person is asked if they have ever used a cell phone while driving. Police are curious as to whether or not the rate has increased over a 60% rate from two years ago.
- A large sample of young, middle and older drivers is taken and each driver is asked how fast they drove. The police suspect that young drivers are faster than the others, but want to test their theory.

S O L U T I O N


Data
Categorical variables
Continuous variables
Z test for proportion
Two samples Z test for proportion
Chi-Squared Test
Number of Samples
One(Test claim about Population)
Two (Compare means)
More than Two (Compare Independent means)
Unknown population std
Known population std
Paired Samples
Independent Samples
ANOVA (F-test)
Paired T test
Sample size <= 30
Sample size >=30
Unknown population std
Known population std
One sample T test
One sample Z test
Sample size <= 30
Sample size >=30
Pooled T test
Two samples Z test



NEXT SLIDE

No Draggable™ Response

You didn't answer this question

THANKS!

Any questions?

You can find me at:

- jason@clarusway.com