The Data Driven Manager

3

Two Sample Independent Tests

Learning Objectives

- Discern between samples that are independent and dependent
- Perform a two sample test for differences in means (independent groups) when variances are equal
- Perform a two sample test for differences in means (independent groups) when variances are unequal
- Perform a two sample test for differences in variances (independent groups) when underlying distribution is normal

Learning Objectives

- Perform a two sample test for differences in proportions (independent groups)
- Perform a two sample test for differences in counts (independent groups)
- Perform a two sample nonparametric test (independent groups)

Introduction to Independent Tests

Two Sample Tests

- Comparing Parameters of Two Populations
  - Is the new design of a production part different from the old design?
  - Is the proportion of promotable employees at one facility different from that at another?
  - Did one group of experimental subjects react differently from the other?

How to Select the Appropriate Test for Two Samples

- Identify the type of data associated with the **Criterion Measure** of interest:
  - Nominal
  - Ordinal
  - Continuous

- Determine whether the samples come from Two **Independent** or **Dependent** Populations

How to Select the Appropriate Test for Two Samples

- **Independent** Samples
  - Each item within each sample is independent of each other item
  - All of the items in a given sample (group) are independent of each and every item in the other sample (group)
  - There is no linkage between any of items in each of the two samples (groups)

How to Select the Appropriate Test for Two Samples

- **Dependent** Samples
  - Each of the items within each sample are independent of every other item in the sample
  - Each item (specimen) in one group is linked or related to a corresponding item in the other sample
  - This linkage dependency can be due to
    - Repeated Measures
    - Matching
    - Pairing

How to Select the Appropriate Test for Two Samples

- **Repeated Measures**
  - The two sets of data represent repeated measures (pairs of observations) from a single sample (**dependent by nature**)

- **Matching / Pairing**
  - The two samples are **dependent by design**, based on paired or grouped testing through time, or upon a pretest or covariate.

Independent Examples

- An economist wishes to determine whether there is a difference in mean family income for households in two socioeconomic groups.
- An admissions officer of a small liberal arts college wants to compare the mean SAT scores of applicants educated in rural high schools & in urban high schools.

Dependent Examples

- An analyst for Educational Testing Service wants to compare the mean GMAT scores of students before & after taking a GMAT review course.
- Nike wants to see if there is a difference in durability of 2 sole materials. One type is placed on one shoe, the other type on the other shoe of the same pair.

File Formats

- In the Independent File format the data file should have as a minimum, a **Grouping Variable Column** (a column which contains the group number or designator), and at least one additional column which contains the Data Column, that is, the values of the Criterion Measure.

- In the Dependent File format, each group of Data values is in a separate column of their own with their own variable name.

File Formats

- In the Independent File format the data file should have as a minimum, a **Grouping Variable Column** (a column which contains the group number or designator), and at least one additional column which contains the Data Column, that is, the values of the Criterion Measure.

- In the Dependent File format, each group of Data values is in a separate column of their own with their own variable name.

File Formats

Independent to Dependent format:

```
transform.independent.format.to.dependent.format(fx = cm ~ group.var, data = data.frame.name)
```

Dependent to Independent format:

```
transform.dependent.format.to.independent.format(data = data.frame.name)
```

Two Sample Statistical Tests

Parameters / Condition	Means	Variances	Proportions	Counts (Poisson)
Independent	• Two Sample t-test for Independent Measures • Two Sample Approximate t-test	• F - test • Levene and ADM(n-1)	• Two Sample Exact Binomial Test (Fisher's Exact Test)	• Two Sample Poisson Test
Dependent	• Repeated Measures t-test • Matched Pairs t-test	• Matched Pairs t-test for Variances	• McNemar's Dependent Proportions Test • Sign Test	• Wilcoxon Signed Ranks Test

Two Independent Sample Tests

- Two Independent Sample tests for Means
  - Sigma Known
  - Sigma unknown and presumed equal
  - Sigma unknown and presumed unequal

- Testing Hypotheses for Variances and Dispersion
  - F Test
  - Levene and ADM(n-1)

Two Independent Sample Tests

- Two Independent Sample Proportion Tests
  - Fisher's Exact Test
  - χ^2 Test

- Two Independent Sample Count (Poisson Rate) Tests
  - Two Sample Poisson Test

- Wilcoxon-Mann-Whitney Test (Nonparametric)

Two Sample Hypothesis Tests for **Means**

Independent Conditions

IMPORTANT!

- For any two sample (or more) test for **continuous** data, you must test in this order:
  - **Normality** (to verify underlying assumptions for variance)
  - **Variance** (to verify underlying assumptions for means)
  - **Means**

Two Independent Sample Tests for Means

- **Unknown** σ^2_1 and σ^2_2 (t test)
  - σ^2_1 and σ^2_2 presumed **equal**
  - σ^2_1 and σ^2_2 presumed **unequal**

Two Sample Equal Variance

t Test for Means

Two Sample Equal Variance t Test for Means

- Underlying assumptions
  - The samples are randomly selected from two independent populations or processes.
  - The underlying processes are normally distributed.
  - Homogeneity of variance is assumed
    - $\sigma^2_1 = \sigma^2_2 = \sigma^2$
  - Population variances are **unknown**

Two Sample Equal Variance t Test for Means

- Test Statistic

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_P^2}{n_1} + \frac{s_P^2}{n_2}}}$$

- Has df = $n_1 + n_2 - 2$, where

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

Example t Test Problem

- A design engineer finds that for two different designs of the same motor, the brush box wear when tested appeared as follows:

$$\bar{X}_1 = 0.0060 \qquad \bar{X}_2 = 0.0090$$

$$s_1 = 0.0015 \qquad s_2 = 0.0013$$

$$n_1 = 25 \qquad n_2 = 30$$

- Based upon the two random samples, is it reasonable to assume that the average amount of wear is equal for the two design populations? Assume $\alpha = 0.05$.

Example t Test Problem

In RStudio

```
t.test.twosample.independent.simple()
```

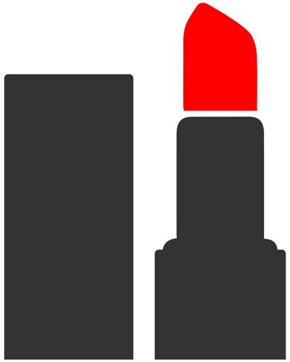
View the Full Hypothesis Testing Procedure here:

https://tinyurl.com/indttestequalvar

Your Turn!

- The characteristic “cap pull force” refers to the amount of effort required (measured in pounds) to remove a lipstick cap from an assembly base.

- A cap that pulls off too easily results in an assembly that may fall apart. A cap that is difficult to remove would tend to frustrate the end user.



Your Turn!

- All of the assembly components are made from injection molded plastic.
- There are two molds that make the cap.
- The plant manager wants to know if the average cap pull force was equal for caps produced on the two molds.

Your Turn!

- Two groups of random samples were drawn, one from each of the two production lines, with each batch representing a different cap mold.
- Appropriate procedures were followed in measuring cap pull force for the two batches.
- The resultant data are recorded below and are stored in the data file **CapPull2.txt**. The variable names are Mold and Pull. Test an appropriate hypothesis to address the plant managers concern. Assume $\alpha = 0.05$.

Two Sample Equal Variance t Test for Means

1. The hypothesis is about:
   a. μ
   b. σ^2
   c. π
   d. λ
   e. ρ

2. What type of data do we have (criterion measure)?
   a. Continuous
   b. Ordinal
   c. Nominal
   d. Absolute Scale

Two Sample Equal Variance t Test for Means

3. What is the proper test statistic?
   a. Z
   b. T
   c. F
   d. χ^2
   e. Exact

4. What is the p value of the proper test statistic?
   a. 0.709
   b. 0.647
   c. 0.657
   d. 0.065

Two Sample Equal Variance t Test for Means

5. Do you
   a. Reject the null hypothesis?
   b. Fail to reject the null hypothesis?

Two Sample Unequal Variance

t Test for Means

Two Sample Unequal Variance t Test for Means

- Underlying assumptions
  - The samples are randomly selected from two independent populations or processes.
  - The underlying processes are normally distributed.
  - The population or process variances are not equal.

Two Sample Unequal Variance t Test for Means

- Test Statistic

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Has df = df*

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

Example t Test Problem

- A production manager wants to compare two ultrasonic welders for average resistance to destruction for the parts made on each.

- History has shown that very different amounts of variation can occur between the two machines, and the parts are very expensive to test.

Example t Test Problem

- The randomly selected samples show:

$$\bar{X}_1 = 75 \qquad \bar{X}_2 = 82$$
$$s_1 = 20 \qquad s_2 = 9$$
$$n_1 = 12 \qquad n_2 = 12$$

- Test an appropriate hypothesis at an $\alpha = 0.10$.

Example t Test Problem

In RStudio

```
t.test.twosample.independent.simple()
```

View the Full Hypothesis Testing Procedure here:

https://tinyurl.com/indttestunequalvar

Your Turn!

- One of the many consumable supplies that a company's Purchasing Department is responsible for obtaining is cutter inserts for their lathe operations.

- The purchasing manager previously elected to buy an equal volume of cutters from two suppliers.

Your Turn!

- The lathe operators, through time, have been reporting to their department supervisor that there is a difference between the two vendors' cutter inserts in terms of cutting life.
- The supervisor decided to statistically test whether there was really a difference in the average cutting life between the two vendors' inserts.

Your Turn!

- They requested their operators to use the inserts from the two suppliers in random fashion across the parts being machined and to record the total cutting time (life) for each of the inserts.

- When sufficient data were compiled, the supervisor organized the obtained values and saved them in a file named **ToolLife.txt**

Your Turn!

- The insert life was recorded in minutes.
- Assist the supervisor in conducting the appropriate analysis required to answer the research question.
- Assume $\alpha = 0.05$.

Two Sample Unequal Variance t Test for Means

1. The hypothesis is about:
   a. μ
   b. σ^2
   c. π
   d. λ
   e. ρ

2. What type of data do we have (criterion measure)?
   a. Continuous
   b. Ordinal
   c. Nominal
   d. Absolute Scale

Two Sample Equal Variance t Test for Means

3. What is the proper test statistic?
   a. Z
   b. t
   c. F
   d. χ^2
   e. Exact

4. What is the p value of the proper test statistic?
   a. 0.000
   b. 0.001
   c. 0.002
   d. 0.308

Two Sample Equal Variance t Test for Means

5. Do you
   a. Reject the null hypothesis?
   b. Fail to reject the null hypothesis?

Two Sample Hypothesis Tests for **Dispersion**

Independent Conditions

Testing Hypotheses for Variances and Dispersion

- Purposes for dispersion testing
    - To determine which t test to use when testing hypotheses for means
    - To determine whether “treatments” applied to two groups differentially affect dispersion

Testing Hypotheses for Variances and Dispersion

- Test procedures
  - F test for variances
  - Levene test for dispersion (ADA)
  - ADM(n-1) procedure (extending Levene)

Two Independent Sample F Test for Variances

- This test is for testing variances specifically

- Underlying assumptions
  - The samples are randomly selected from two independent populations or processes.
  - The underlying processes are normally distributed.

Two Independent Sample F Test for Variances

- Test Statistic

$$F_{(n_1-1,\, n_2-1\ \mathrm{df})} = \frac{s_1^2}{s_2^2}$$

Example F Test Problem

- A design engineer finds that for two different designs of the same motor, the brush box wear when tested appeared as follows:

$$\bar{X}_1 = 0.0060 \qquad \bar{X}_2 = 0.0090$$

$$s_1 = 0.0015 \qquad s_2 = 0.0013$$

$$n_1 = 25 \qquad n_2 = 30$$

Example F Test Problem

- In preparation for testing an appropriate hypothesis about the equality of means, the design engineer has determined that they should first test for the equivalence of the variances.

- We will use a significance level of $\alpha = 0.05$ to test this hypothesis.

Example F Test Problem

In RStudio

```
variance.test.twosample.independent.simple()
```

View the Full Hypothesis Testing Procedure here:

https://tinyurl.com/indFtest

Your Turn!

- A manufacturing engineer is interested in determining whether the variability in the product produced in January is different from the variability of the product produced in February.
- The data for this study were properly collected and saved in a file named **Thick2.txt**. The variable names are Batch and Thick.
- Conduct an appropriate test for dispersion. Assume α= 0.05.

Two Independent Sample F Test for Variances

1. The hypothesis is about:
   a. μ
   b. σ^2
   c. π
   d. λ
   e. ρ

2. What type of data do we have (criterion measure)?
   a. Continuous
   b. Ordinal
   c. Nominal
   d. Absolute Scale

Two Independent Sample F Test for Variances

3. What is the proper test statistic?
   a. Z
   b. t
   c. F
   d. χ^2
   e. Exact

4. What is the p value of the proper test statistic?
   a. 0.043
   b. 0.122
   c. 0.116
   d. 0.000

Two Independent Sample F Test for Variances

5. Do you
    a. Reject the null hypothesis?
    b. Fail to reject the null hypothesis?

The Levene Test for Dispersion

- Not a test of variances; tests **A**bsolute **D**eviation from the **A**verage (ADA)
- ADA = $|X_i - \text{Average}_{group}|$ is generated for each score
- Uses the ANOVA (or a t test for 2 groups)

The Levene Test for Dispersion

Classify each variable as a factor	`as.factor()` `factor()`
Compute group dispersion ADA scores	`compute.group.dispersion.ADA()`
Perform t-test on the ADA scores by group	`t.test.twosample.independent.fx()`

The ADM$_{(n-1)}$ Procedure

- Used when non-normality prevents the use of the F test
- Not a test of variances, but tests for differences in dispersion around medians
- This procedure employs the **A**bsolute **D**eviations from the group **M**edians

The ADM$_{(n-1)}$ Procedure

- $ADM_{(n-1)} = |X_i - Median_{group}|$ is generated for each score
- The "n - 1" comes from dropping a redundant score (the middle value)
- Is much more robust to non-normality than the Levene test
- The preferred test when normality may not be assumed

The ADM$_{(n-1)}$ Procedure

- When there is an odd number of scores in a group, the median will be the exact middle value and its corresponding deviation score value will be zero.
- Since there is no information with regard to variability in a score of 0, it is dropped.

Procedure developed by M. Petrovich and R. Littlejohn

The ADM$_{(n-1)}$ Procedure

- When there is an even number of scores in a group, the median will be the average of the two middle scores. The absolute value of the corresponding deviation scores associated with those two middle scores will be identical.

- Since there is no additional dispersion information in the second identical score once the first one is used, it is redundant and therefore it is dropped.

Procedure developed by M. Petrovich and R. Littlejohn

The ADM$_{(n-1)}$ Procedure

Classify each variable as a factor	`as.factor()` `factor()`
Compute group dispersion ADMn-1 scores	`compute.group.dispersion.ADMn1()`
Perform t-test on the ADMn-1 scores by group	`t.test.twosample.independent.fx()`

Your Turn!

- Use the ToolLife data and both the Levene test and ADMn-1 Procedures to assess for differences in dispersion.
- Assume α= 0.01.

Pooled Variance

- What if we don't reject the null hypothesis? Can we provide a point estimate?

Two Sample Hypothesis Tests for **Proportions**

Independent Conditions

Two Independent Sample Proportion Tests

Examples:

- Is there a difference in the number (percent) of quarters vs. nickels rejected by a given type of vending machine?

- Is there a difference in the proportion of bottles that are not properly filled as related to an "old" versus "new" filler valve design?

Two Independent Sample Proportion Tests

- Several ways to perform this test
    - Fisher's Exact two sample proportion test (preferred)

Two Independent Sample Proportion Tests

- Underlying assumptions
    - The two processes from which the sample data are drawn are inherently independent in nature, and are both based upon the Bernoulli process
    - The samples are randomly selected from the underlying processes being investigated

Fisher's Exact Test

- p value

$$p = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,N!}$$

- Where a, b, c and d are frequencies (counts) in a 2x2 contingency table, and N is the total count

	Group 1	Group 2
Pass	a	b
Fail	c	d

Fisher’s Exact Test

	Group 1	Group 2	Row Total
Pass	a	b	a + b
Fail	c	d	c + d
Column Total	a + c	b + d	a + b + c + d = N

Fisher’s Exact Test Problem

- A systems engineer is anxious to determine whether two recently installed pieces of equipment are operating on an equivalent basis.

- The machines are blow molders, and the canisters they produce are assessed on an attribute basis.

Fisher's Exact Test Problem

- Specifically, each canister is evaluated only on a pass/fail basis.
- Nonconformities include leaks/doesn't leak and cracked/not cracked, etc.

Fisher's Exact Test Problem

- A random sample of 750 canisters is selected from the initial production run of each machine. The results were as follows.

$$p_1 = 0.18 \qquad p_2 = 0.12$$

$$n_1 = 750 \qquad n_2 = 750$$

- Test an appropriate hypothesis. Assume $\alpha = 0.01$.

Fisher's Exact Test Problem

In RStudio

```
proportion.test.twosample.exact.simple( )
```

View the Full Hypothesis Testing Procedure here:

https://tinyurl.com/Fishersexactproptest

Your Turn!

- In order to meet volume requirements, two injection molding machines are used to produce the same connector housing.
- There are several visual characteristics that are identified as critical to the customer.

Your Turn!

- The plant manager has requested that the department supervisor determine if the proportion nonconforming of all combined visual characteristics is different for the two machines.
- Note that a product is nonconforming if one or more nonconformities are found on the product.

Your Turn!

- Two random samples of 500 units each were selected from the production output from each machine. The parts were visually inspected and yielded the following results.
- The data are in a file named **visual.txt**. A value of 0 represents a conforming part, a value of 1 represents a non-conforming part.
- Test an appropriate hypothesis. Assume $\alpha = 0.05$.

Two Independent Sample Proportion Test

1. The hypothesis is about:
   a. μ
   b. σ^2
   c. π
   d. λ
   e. ρ

2. What type of data do we have (criterion measure)?
   a. Continuous
   b. Ordinal
   c. Nominal
   d. Absolute Scale

Two Independent Sample Proportion Test

3. What is the proper test statistic?
   a. Z
   b. t
   c. F
   d. χ^2
   e. Exact

4. What is the p value of the proper test statistic?
   a. 0.052
   b. 0.222
   c. 0.937
   d. 0.111

Two Independent Sample Proportion Test

5. Do you
    a. Reject the null hypothesis?
    b. Fail to reject the null hypothesis?

Pooled Proportion

- What if we don't reject the null hypothesis? Can we provide a point estimate?

Two Sample Hypothesis Tests for **Counts**

Independent Conditions

Two Independent Sample Poisson Rate Test

Underlying Assumptions

- The two processes from which the sample data are drawn are inherently independent in nature

- The data are discrete counts that follow a Poisson distribution.

Poisson Rate Test Problem

- A team was interested in determining whether their activities involving cleanliness have made a difference in the (average) population rate (λ) of minor Eddy Current indications.

- They used two different procedures for cleaning to see which method resulted in better cleanliness.

Poisson Rate Test Problem

- Based on the data collected by the team (in the file **Eddycur.txt)**, is there a difference in cleaning methods, and if so, which method should the team use?

- Assume α = 0.05.

Poisson Rate Test Problem

In RStudio

Test for the Poisson Distribution	`poisson.dist.test()`
Two Sample Poisson Test	`poisson.test.twosample.simple()`

Poisson Rate Test Problem

In RStudio

```
poisson.test.twosample.simple( )
```

View the Full Hypothesis Testing Procedure here:

https://tinyurl.com/2sampPoisTest

Your Turn!

- A quality improvement team is charged with reducing the number of customer complaints that the plant has been receiving.
- The company provides LCD screens for monitors that go into the display unit of a medical testing device.
- These screens have many thousands of pixels that make up the display and are either turned on (and in a certain color) or turned off producing the visible display on the screen.

Your Turn!

- When a pixel does not turn on or off at the right time, or is the wrong color, it is counted as a "bad pixel."
- The manufacturer of this equipment has indicated that although it is their desire to have displays with 100% of the pixels functioning, a few bad pixels can be tolerated.
- Nonetheless, the plant has had numerous displays rejected in the past few weeks for bad pixels and the customer is threatening to cut their orders if the problem is not fixed.

Your Turn!

- An appropriately staffed team worked diligently on the problem and found what they think is a fix. (It had to do with the characteristics of the liquid in the display at the time of assembly.)

- In order to determine whether or not the “fix” is effective, the team produced 12 displays in a laboratory setting under the conditions that they believe will improve the display by reducing the number of bad pixels.

Your Turn!

- The average number of bad pixels for these 12 new displays was 4.5.
- Does this constitute an improvement from an average of 7.2 bad pixels for the 10 displays in the other group that were assessed just before the modified procedure was implemented for the 12 test displays?

Two Independent Sample Proportion Test

1. The hypothesis is about:
   a. μ
   b. σ^2
   c. π
   d. λ
   e. ρ

2. What type of data do we have (criterion measure)?
   a. Continuous
   b. Ordinal
   c. Nominal
   d. Absolute Scale

Two Independent Sample Proportion Test

3. What is the proper test statistic?
   a. Z
   b. t
   c. F
   d. χ^2
   e. Exact

4. What is the p value of the proper test statistic?
   a. 0.008
   b. 0.109
   c. 0.004
   d. 2.635

Two Independent Sample Proportion Test

5. Do you
   a. Reject the null hypothesis?
   b. Fail to reject the null hypothesis?

Nonparametric Tests

Independent Conditions

Wilcoxon Mann Whitney Test

- Assesses whether the scores come from the same distribution
- Most sensitive to differences in location, the medians in particular
- Is a rank-based test (compares the differences between the sums of ranks)

Wilcoxon Mann Whitney Test

Underlying Assumptions

- The scores in the two groups are independent of one another within and between the groups
- The data are measured on at least an ordinal scale with sufficient resolution
- The underlying characteristic is continuous

Wilcoxon Mann Whitney Test Problem

- Consider the situation where a marketing team is interested in whether or not the color of a frosting can impact the assessment of sweetness of the product employing the following 6-point rating scale.

Not At All Sweet	Slightly Sweet	Mildly Sweet	Moderately Sweet	Very Sweet	Much Too Sweet
1	2	3	4	5	6

Wilcoxon Mann Whitney Test Problem

The marketing team has one group of people taste and evaluate the sweetness for product frosted with a light pink frosting.

A second group is asked to evaluate the same product and frosting, but for this group, the frosting is colored white.

The results of the ratings are in the data file named **Frosting.txt**. Use a Type 1 Error level of 0.05 to interpret results.

Wilcoxon Mann Whitney Test Problem

In RStudio

`median.test.twosample.independent.mann.whitney.fx( )`

View the Full Hypothesis Testing Procedure here:

https://tinyurl.com/WilcoxMannWhitney

Your Turn

Suppose we conducted a survey where we asked two groups of people, Group A and Group B, to rate a new restaurant on a scale of 1 to 5 (1 being "poor" and 5 being "excellent").

The ratings are as follows:

Group A (local residents): 5, 4, 3, 5, 4, 4, 5, 3, 4, 5

Group B (tourists): 3, 3, 4, 3, 3, 2, 3, 2, 4, 3

Your Turn

Do local residents and tourists rate a new restaurant differently in terms of overall dining experience?

The data are in a file named **dining.txt**.

Wilcoxon Mann Whitney Test

1. The hypothesis is about:
   a. μ
   b. M (distributions)
   c. π
   d. λ
   e. ρ

2. What type of data do we have (criterion measure)?
   a. Continuous
   b. Ordinal
   c. Nominal
   d. Absolute Scale

Wilcoxon Mann Whitney Test

3. What is the proper test statistic?
   a. Z
   b. t
   c. F
   d. U

4. What is the p value of the proper test statistic?
   a. 0.0027
   b. 0.0014
   c. 0.0000
   d. 0.9986

Wilcoxon Mann Whitney Test

5. Do you
   a. Reject the null hypothesis?
   b. Fail to reject the null hypothesis?