Hypothesis Testing

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Recall the basic assumptions and concepts related to hypothesis testing

What is a Hypothesis?

- An assumption related to a process or population.

Hypothesis Testing

- A procedure which uses sample statistic(s) to make inferences about a population.

Statistical Significance

- Refers to the assumption that the observed difference or association/phenomenon represents a significant departure from what might be expected by chance alone.

Testing Statistical Hypotheses

- Statistical hypotheses are generated in pairs, representing all possible outcomes
  - Null hypothesis
  - Alternative hypothesis

The Null Hypothesis

- Symbol: H_0
- The hypothesis that states that no difference or relationship exists.
- Examples:

$$H_0: \mu = 50 \qquad H_0: \sigma_1^2 = \sigma_2^2$$

The Alternative Research Hypothesis

- Symbol: H_1
- The hypothesis statement that represents the decision if the null hypothesis is rejected.
- Examples: $H_1: \mu \neq 50$ $\quad$ $H_1: \sigma_1^2 \neq \sigma_2^2$

The Directional Research Hypothesis

- Directional hypotheses state not only that the null hypothesis is not true, but that there is a specific direction involved

$$H_1: \mu > 50$$

The Directional Research Hypothesis

- Therefore, the appropriate H_0 / H_1

$$H_0: \mu \leq 50$$

$$H_1: \mu > 50$$

Observations and Cautions

- We either accept or reject H_0; we have never proven that a difference exists
- In essence, we have found that we do, or do not, have sufficient statistical evidence to accept or reject a hypothesis, respectively

Observations and Cautions

- Development of the hypotheses takes place prior to the collection of the data

Observations and Cautions

- An alternative hypothesis will lead to a two-tailed hypothesis test

- A directional hypothesis leads to a one-tailed hypothesis test

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Significance Level and Risk

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Interpret significance level and risk

Alpha Level and Risk

- Alpha, α, is a selection of risk that you are willing to take

- Given a true null hypothesis, α is the probability the null hypothesis could be rejected

Alpha Level and Risk

- The smaller the selected level of α , the smaller the probability of rejecting a true null hypothesis

- The researcher selects this risk value

p Value

- The significance level, or p-value, is the probability that an observed statistic, or one that is more extreme, could have occurred by chance, given a true null hypothesis

p Value

- The p-value is generated from calculation in statistical tests and is directly compared to α
- We will reject a null hypothesis if the p-value is less than or equal to the selected level of α

Test Statistics

In hypothesis testing we:

- Take samples
- Calculate sample statistics
- Calculate test statistics
- Calculate probabilities (significance) using the test statistics

Test Statistics

- Some important test statistics include:
- z, t, χ2 , and F

- Probabilities of test statistics can be approximate or in some cases exact, and require underlying assumptions are met

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

One and Two Tailed Tests

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

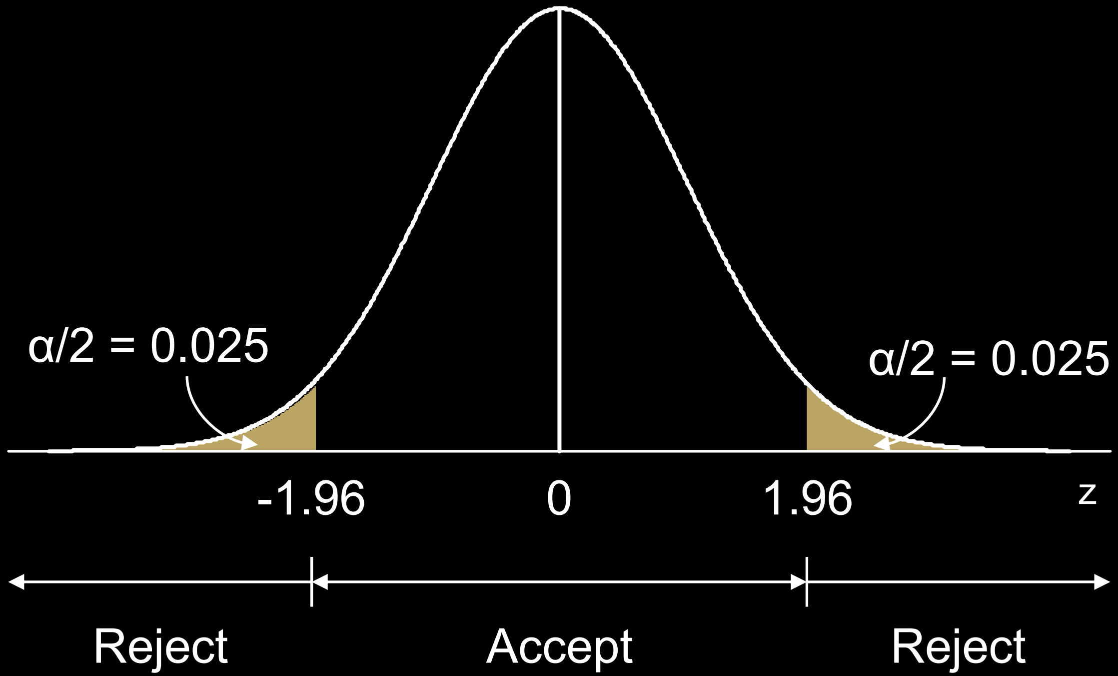
Learning objective:

Discriminate between one tail and two tailed tests

Two Tailed Tests

- In most instances, the researcher will not be able to make a prediction as to the direction of a possible change
- In this case, an alternative H structure is appropriate, and the test will have 2 rejection areas.

Rejection Regions
Two Tailed Tests



One Tailed Tests

- In some cases, an investigator will be able (or forced) to make a prediction based upon a theoretical rationale or prior research

One Tailed Tests

- In this case, a one-tailed test with a directional hypothesis may be appropriate.

- This choice would result in a single rejection region.

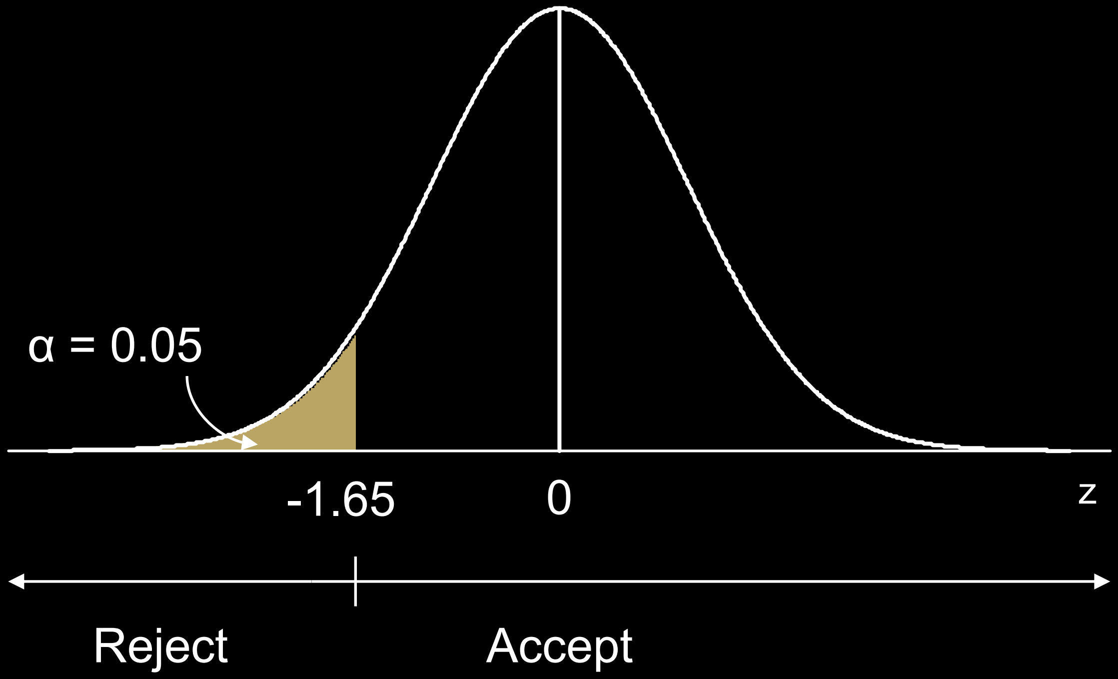
One Tailed Test Example

- Suppose that we do not wish to ship a production lot of washing machines unless, based upon a randomly drawn sample from the lot and spin tested, the mean (μ) of the lot may be reasonably assumed to be at least 2,000 rpm to failure

One Tailed Test Example

- The hypotheses to be tested are:
  - H_0: $\mu \geq 2{,}000$
  - H_1: $\mu < 2{,}000$

Rejection Region
One Tailed Test



Observations and Cautions

- The one-tailed test will cause the rejection of the H_0 more often than a two-tailed test if the direction predicted is the same as the direction of the difference or association

Observations and Cautions

- Note: there is no possibility of rejecting the null hypothesis if there is a difference in the opposite direction!

- A directional test should be employed only when thoroughly justified

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Type I and Type II Error

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Differentiate between Type I and Type II Error

Type I and Type II Error

- When we are testing hypotheses, we can make errors with respect to our conclusions.
- These errors are referred to as Type I and Type II errors.

Type I Error

- Symbol: α
- The probability of rejecting a true null hypothesis
- Also referred to as a false positive, or producer’s risk

Type II Error

- Symbol: β
- The probability of accepting a false null hypothesis
- Also referred to as a false negative, or consumer’s risk

Power

- Symbol: 1-β
- The probability of rejecting a false null hypothesis
- The ability of the test to correctly reject a false null hypothesis

Confidence

- Symbol: $1-\alpha$
- The probability of accepting a true null hypothesis

Experimental Outcomes

	TRUE	FALSE
Accept H_0	$1-\alpha$ (Confidence)	β (Type II Error)
Reject H_0	α (Type I Error)	$1-\beta$ (Power)

Example

Decision	Actual Situation or Reality - H_0	
	No Police with Radar	Police with Radar
Find No Police	Confidence (No False Signal)	Type II Error: (Something Missed)
Find Police	Type I Error: (False Signal)	Real Power Ability to Detect

Observations

- $\alpha + \beta$ will never equal 1. They are conditional probabilities based upon different conditions.
- Specifically, α is based upon the premise that H_0 is true, β is predicated on the assumption that H_0 is false.

Observations

- Both α and β represent risk.
- They are an expression of the researcher's willingness to commit an error in their inference.

Observations

- Power, the ability of the test to correctly reject a false H_0, must be “purchased” with sample size and with the selection of an appropriate experimental design.

Observations

- α is not “always more important” than β. For example:
- A drug company wishes to test the safety of a new drug formulation. The hypotheses tested are:

Observations

- H_0: The drug is safe
- H_1: The drug is not safe

- In this case, which type of error is of most concern?

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Beta and Power

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Interpret the relationship between Beta and Power

Beta and Power

- Power is the probability of correctly rejecting a False H_0
- Correct decision when H_0 is false
- Power is designated as $1 - \beta$

Beta and Power

- Power is used in determining how well a test is working and likely to detect a true effect (or difference)

Beta and Power

- Affected by
  - True value of the population parameter
  - Significance level, α
  - Standard deviation, s (or σ)
  - Sample size, n

Strategies for Considering Power in Experiments

Strategy 1

- Fix the probability of committing a Type I error (α).

Strategies for Considering Power in Experiments

Strategy 1

- Select a sample size large enough so that β is acceptably small, and testing is not too expensive or time consuming to conduct.

Strategies for Considering Power in Experiments

Strategy 2

- Consider the Null and Research Hypotheses and select the α and β pair which best represents your wishes related to the research,

Calculating Beta and Power for Means

Strategy 2

- Then, calculate the sample size required to maintain the selected risk levels

Example

Assumptions

- The Central Limit Theorem is applicable
- The RSD’s employed will be approximately normal

Example

An engineer has been studying the effects of modifying a component. They want to know if the change significantly reduces an output measure called response time, measured in milliseconds.

Example

Historical data indicates that the response time average, μ, is 20 ms with a standard deviation of 3.5 ms.

Example

Further, let us assume that a reduction in the mean of at least 4 ms is necessary before the modification becomes cost-effective. This is referred to as the effect size (Δ) of an experiment.

Example

In this case, the hypotheses tested might be as follows.

$$H_0: \mu \geq 20$$
$$H_1: \mu < 20$$

Example

Based on the original data, we recall that the process μ has been 20; therefore,

if H_0 is true, $\mu \geq 20$ and $\sigma = 3.5$

if H_1 is true, $\mu \leq 16$ and $\sigma = 3.5$

Given a Δ of 4 ms

Example

Assume further that α was selected at a 0.05 level

Note also that a one-tailed test has been employed for the hypotheses to be tested, which corresponds to a z of -1.645.

Example

We will assume that $n = 9$.

What would be the power to detect a difference of 4 units?

Calculating β and Power for Means

Step 1:

Determine the critical value in the H_0 RSD of means corresponding to the z value for the given value of α

Calculating β and Power for Means

$$z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$-1.645 = \frac{\bar{X} - 20}{\frac{3.5}{\sqrt{9}}} = \frac{\bar{X} - 20}{1.167}$$

$$\bar{X} = 18.081$$

Calculating β and Power for Means

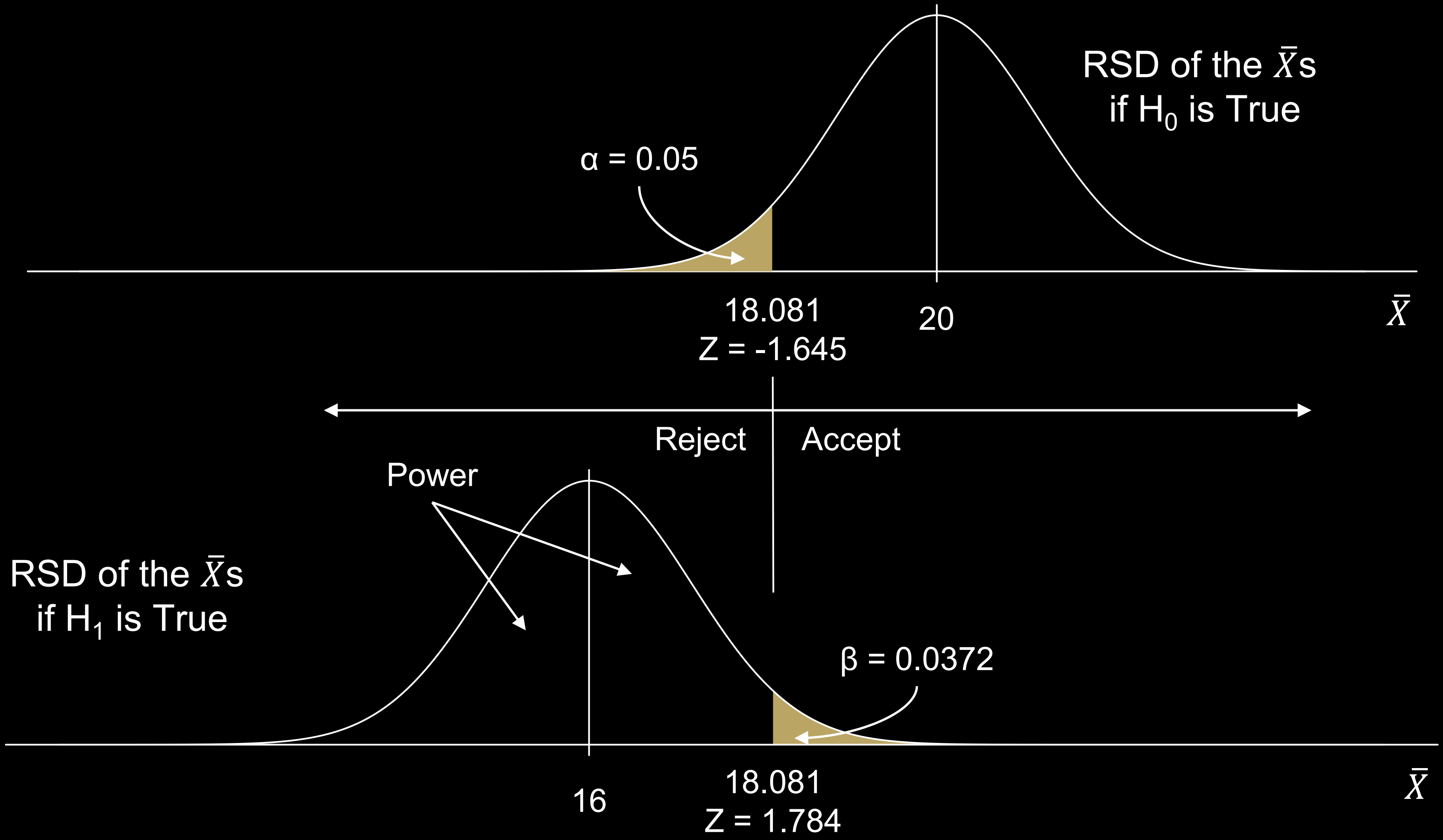
Step 2:

Calculate the z value and area corresponding to the calculated on the "H_1 is true" curve

Calculating β and Power for Means

$$z = \frac{18.081 - 16}{1.167} = 1.784$$

$$z = 1.784$$
$$\beta = 0.0372$$


RSD of the $\bar{X}$s
if H_0 is True
α = 0.05
18.081
Z = -1.645
20
$\bar{X}$
Reject
Accept
Power
RSD of the $\bar{X}$s
if H_1 is True
β = 0.0372
16
18.081
Z = 1.784
$\bar{X}$

Summary Observations

- The larger the value of $\mu_0 - \mu_1$ (Δ delta), the larger power ($1 - \beta$) will become

- Generally, both α and β should be small. In industry, studies planned without an initial regard for β generally result in low power or high β values

Summary Observations

- It is not possible to commit Type I and Type II errors at the same time
- Had a two-tailed test been employed, the power of the test would have been the sum of the two areas falling beyond α on the H_1 distribution
- Increasing α will generally reduce β

Summary Observations

- Increasing n will generally increase the power of the test

- Increasing the power of the test can be accomplished by reducing the standard error through design modifications (for example, matched groups and stratified sampling)

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Calculating Power

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Calculate Power for changes in means and variance for a non-directional test

Example 1

- The Product Design and Marketing Departments have agreed to consider changing the material for the handle of a trenching shovel

Example 1

- The material presently used is hickory wood
- A potential supplier has offered a handle of the same design but made of a composite material, at a significantly lower price

Example 1

- One of the critical product characteristics for the handle is handle strength

- The method of measuring this characteristic is destructive in nature

Example 1

- You have been given the following process values for the existing material:
  - μ = 440 lbs, σ = 10 lbs
  - $\gamma_3 = 0.0$, and $\gamma_4 = 0.0$ (Normally distributed)

Example 1

- A sample of the composite material handle is to be tested

- The following values are assumed to be appropriate for the test: $\Delta\mu$ = 10 lbs, n = 9, and level of confidence = 95%.

Example 1

- Under these conditions, what are β and power (assuming no change has occurred in the dispersion or variability of the process)?

Beta and Power for Changes in Means

- In Rstudio

```
> power.mean.t.onesample
```

Example 2

- Using the same data, let's now suppose that the following values are assumed to be appropriate for the test: $\Delta\sigma = 2$ lbs, n = 9, and level of confidence = 95%.

Beta and Power for Changes in Variance

- In Rstudio

```
> power.variance.onesample
```

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Calculating Sample Size

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Calculate sample size for means and variance

Sample Size Calculations

- For the industrial researcher, proper sample size is not an opinion

Factors to Considered

- The minimum effect size (Δ) to be detected

- That is, the smallest degree of shift in the parameter that the researcher wishes to identify

Factors to Considered

- The number of treatment levels (or groups)

- The population variance (σ^2)

Factors to Considered

- The probability of committing a Type I error (α)

- The probability of committing a Type II error (β)

Calculating Sample Size for Two Sample Tests of Means

- Assumptions:
  - σ unknown
  - Continuous Data, Independent Samples
  - Two Normal distributions
  - Non-directional test

Calculating Sample Size for Two Sample Tests of Means

- When σ is unknown, hypothesis tests for means will use the t distribution
- Unfortunately, the t distribution is based upon degrees of freedom, which is determined by sample size

Calculating Sample Size for Two Sample Tests of Means

- As such, sample size must be solved iteratively, where the sample size is determined to be the smallest n that satisfies the following formula

Calculating Sample Size for One Sample Tests of Means

- Formula for non-directional hypotheses

$$n \geq (t_{\alpha/2,(n-1)df} + t_{\beta,(n-1)df})^2 \frac{\sigma^2}{\Delta^2}$$

Example

- If the requirements of a pull test are to be $\alpha = 0.05$, $\beta = 0.02$, $\Delta = 1$ lbs, and $\sigma = 2$, what would the appropriate minimum sample size be for a non-directional test for means?

Example

- In Rstudio

```
> sample.size.mean.t.onesample
```

Calculating Sample Size for One Sample Tests of Variance

- Formula for non-directional hypotheses

$$\chi^2 = \frac{s^2(n-1)}{\sigma_0^2}$$

Calculating Sample Size for One Sample Tests of Variance

- For a non-directional test, we must consider two cases
  - One in which the variance increases
  - One in which the variance decreases

Example

- If the requirements of a pull test are to be $\alpha = 0.05$, $\beta = 0.02$, $\Delta\sigma = 1$ lbs, and $\sigma = 2$, what would the appropriate minimum sample size be for a non-directional test for variances?

Example

- If the variance increases, $\sigma = 3\ (2 + 1)$
- If the variance decreases, $\sigma = 1\ (2 - 1)$

Example

- In Rstudio

```
> sample.size.variance.onesample
```

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Independent vs Dependent Samples

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Discern between samples that are independent and dependent

Two Sample Tests

- Comparing Parameters of Two Populations
- Is the new design of a production part different from the old design?
- 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

How to Select the Appropriate Test for Two Samples

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

Independent Samples

- Each item within each sample is independent of each other item

- All the items in each sample (group) are independent of each and every item in the other sample (group)

Independent Samples

- There is no linkage between any of items in each of the two samples (groups)

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

Dependent Samples

- This linkage dependency can be due to
  - Repeated Measures
  - Matching
  - Pairing

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 Example

- An admissions officer of a small college wants to compare the mean standardized test scores of applicants educated in rural high schools & in urban high schools

Dependent Example

- An analyst for Educational Testing Service wants to compare the mean GMAT scores of students before and after taking a GMAT review course

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Two Sample Hypothesis Tests for Means - Independent Groups

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Perform a statistical test for differences in means (independent groups) when variances are equal and unequal

Two Independent Sample Tests for Means

- Unknown σ_1^2 and σ_2^2 (t test)
  - σ_1^2 and σ_2^2 presumed equal
  - σ_1^2 and σ_2^2 presumed unequal

Two Sample Equal Variance t Test for Means – Assumptions

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

Two Sample Equal Variance t Test for Means – Assumptions

- Homogeneity of variance is assumed
- $\sigma_1^2 = \sigma_2^2 = \sigma^2$
- Population variances are unknown

Two Sample Equal Variance t Test for Means

- Hypotheses

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

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 1 - 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:

Example 1 - t Test Problem

$\bar{X}_1$	$= 0.0060$	$\bar{X}_2$	$= 0.0090$
s_1	$= 0.0015$	s_2	$= 0.0013$
n_1	$= 25$	n_2	$= 30$

Example 1 - t Test Problem

- 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 α = 0.05.

Example 1- t Test Problem

- In RStudio

```
> t.test.twosample.independent
> t.test.twosample.independent.simple
```

Example 2 - t Test Problem

- 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

Example 2 - t Test Problem

- A cap that is difficult to remove would also tend to frustrate the end user
- The assembly components are made from injection molded plastic
- There are two molds that make the cap

Example 2 - t Test Problem

- The plant manager wanted to know if the average cap pull force was equal for caps produced on the two molds

- Two groups of random samples were drawn, one from each of the two production lines, with each batch representing a different cap mold

Example 2 - t Test Problem

- 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.dat**
- Assume $\alpha = 0.05$

Two Sample Unequal Variance t Test for Means – 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

- Hypotheses

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

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 3 - 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 3 - t Test Problem

- The randomly selected samples show

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

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

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Two Sample Hypothesis Tests for Means - Dependent Groups

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Perform a statistical test for differences in means (dependent groups) with repeated measures and with matched pairs

Repeated Measures t Test for Means

- Used to compare means of repeated measures or paired groups
- Tests the following hypotheses:

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2$$

$$\text{or} \qquad \text{or}$$

$$H_0: \mu_D = 0 \qquad H_1: \mu_D \neq 0$$

Repeated Measures t Test for Means – Assumptions

- The n pairs of scores are independent of one another
- The population for difference scores is normally distributed, as are the populations for each group

Repeated Measures t Test for Means

- Test Statistic $$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n} - 2r\frac{s_1}{\sqrt{n}}\frac{s_1}{\sqrt{n}}}} = \frac{\bar{D}}{s_d/\sqrt{n}}$$

Where n is the sample size for the pairs of scores

Example 1 – Repeated Measures t Test Problem

- A Black Belt is attempting to solve a noise level problem with a type of wiper motor
- They suspect that a major source of noise problems may be traced to the material from which the bearing is made

Example 1 – Repeated Measures t Test Problem

- To determine whether the average noise level of the motors can be decreased by changing to a new bearing material, they randomly select ten motors from the current production line

Example 1 – Repeated Measures t Test Problem

- A noise level reading is taken on each of the ten motors.

- Next, the bearings are replaced with bearings made from the new material and the motors are retested

Example 1 – Repeated Measures t Test Problem

- The noise level for the each of the modified motors is recorded as was previously done
- All testing is conducted under the same essential conditions as for the first set of measurements. The data were collected and entered in the file **Noise.dat**

Example 1 – Repeated Measures t Test Problem

- Test an appropriate hypothesis to determine whether it is reasonable to assume that the new bearing material will effectively reduce the initial noise level of the motors if implemented across the process. Assume a significance level of 0.05.

Example 1 – Repeated Measures t Test Problem

- In RStudio

```
> t.test.twosample.dependent
> t.test.twosample.dependent.simple.dbar
```

Matched Pairs t Test for Means - Assumptions

- The specimens in the two samples are independent (by nature)
- The population for difference scores is normally distributed, as are the populations for each group

Matched Pairs t Test for Means - Assumptions

- Homogeneity of variance is assumed (Not critical if sample sizes are equal)
- The units or specimens in the two samples are dependent by design.

Matched Pairs t Test for Means

- Test Statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n} - 2r\frac{s_1}{\sqrt{n}}\frac{s_1}{\sqrt{n}}}}$$

Where n is the sample size for the pairs of scores

Example 2 – Matched Pairs t Test Problem

- A production manager wishes to determine whether two secondary blanking presses are producing raw plates with equal average flatness, a smaller number is better

- It is known that hardness of the material is a variable that can affect the flatness of the plates

Example 2 – Matched Pairs t Test Problem

- Therefore, 60 plates are randomly selected from the primary blanking operation and tested for hardness

Example 2 – Matched Pairs t Test Problem

- After rank ordering and pairing the plates according to their tested hardness (i.e., lowest with the next lowest, and so on up the measured hardness scale), they randomly assign the plates from each pair to one of the two groups (i.e., one plate from each pair to the group going to press 1 and one plate from each pair to the group going to press 2)

Example 2 – Matched Pairs t Test Problem

- After the assignments are complete, the manager verifies that there are no significant differences between the two groups in terms of plate hardness on the basis of their variances and means. Why did they do this?

Example 2 – Matched Pairs t Test Problem

- After the assignments are complete, the manager verifies that there are no significant differences between the two groups in terms of plate hardness based on their variances and means.
- Then, the two groups are run concurrently as pairs through the presses, and the resultant flatness data are recorded.

Example 2 – Matched Pairs t Test Problem

- The summary statistics for the two groups of data are as follows.

$\bar{X}_1 = 35.24$ $\quad$ $\bar{X}_2 = 38.02$

$s_1 = 5.18$ $\quad$ $s_2 = 5.63$

$r_{12} = 0.60$

- Test an appropriate hypothesis at an $\alpha = 0.05$

Example 2 – Matched Pairs t Test Problem

- In RStudio

```
> cor.pearson.r.onesample.simple
> t.test.twosample.dependent.simple.meandiff
```

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Two Sample Hypothesis Tests for Variances

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Perform a statistical test for differences in variances for both independent and dependent groups when the underlying distribution is normal

Testing Hypotheses for Variances - Purpose

- 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

- F Test for Variance (Independent Groups)
- The Dependent Sample t Test for Variances (Dependent Groups)

Two Independent Sample F Test for Variances

- 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

- Hypotheses

$$\mathrm{H}_0: \sigma_1^2 = \sigma_2^2$$

$$\mathrm{H}_1: \sigma_1^2 \neq \sigma_2^2$$

- Test Statistic

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

Example 1 - 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:

Example 1 - F Test Problem

$\bar{X}_1 = 0.0060$

$s_1 = 0.0015$

$n_1 = 25$

$\bar{X}_2 = 0.0090$

$s_2 = 0.0013$

$n_2 = 30$

Example 1 - F Test Problem

- In RStudio

```
> variance.test.twosample.independent
> variance.test.twosample.independent.simple
```

The Dependent Sample t Test for Variances

- Hypotheses

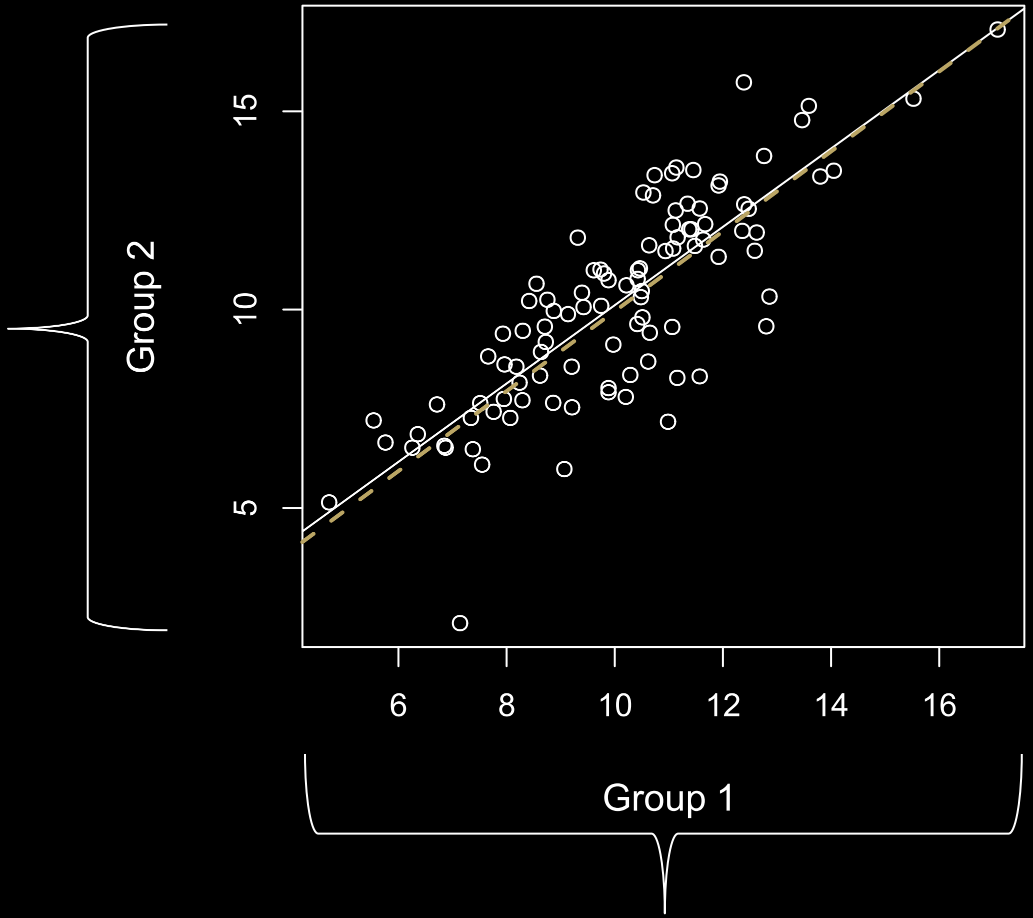
$$\mathrm{H}_0\colon \sigma_1^2 = \sigma_2^2$$

$$\mathrm{H}_1\colon \sigma_1^2 \neq \sigma_2^2$$

- Test Statistic

$$t = \frac{s_1^2 - s_2^2}{2s_1 s_2\sqrt{\frac{1-r^2}{n-2}}}$$

ISO Plot



Equal variation in both directions in the null case

The Dependent Sample t Test for Variances – Assumptions

- The pairs of scores are independent of one another (critical)
- The sample data are either dependent by nature, or dependent by design (critical, therefore you may be required to test the correlation)
- The underlying process distributions are normally distributed (not critical if n is large)

Example 2 – t Test for Variances

- It is quite possible that the two blanking presses (see Sample Problem - Dependent by Design) could be producing product which have equivalent means, but are different in terms of dispersion

Example 2 – t Test for Variances

- Assume you would like to know whether the processes are different in terms of piece-to-piece variability, and test an appropriate hypothesis for that data

Example 2 – t Test for Variances

- The summary statistics for the two groups of data are as follows.

$\bar{X}_1$	= 35.24	$\bar{X}_2$	= 38.02
s_1	= 5.18	s_2	= 5.63
r_{12}	= 0.60		

- Test an appropriate hypothesis at $\alpha = 0.05$

Example 2 – t Test for Variances

- In RStudio

```
> cor.pearson.r.onesample.simple
> variance.test.twosample.dependent.simple
```

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Two Sample Hypothesis Tests for Proportions

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Perform a statistical test for differences in proportions for both independent and dependent groups

Two Independent Sample Proportion Tests

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

Testing Hypotheses for Proportions

- Fisher’s Exact two sample proportion test (Independent Groups)
- McNemar’s Test for Change (Dependent Groups)

Fisher's Exact Test – 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

- Hypotheses

$$\mathrm{H}_0 : \pi_1 = \pi_2$$

$$\mathrm{H}_1 : \pi_1 \neq \pi_2$$

- p value

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

Fisher’s Exact Test

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

	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.

$p_1 = 0.18$ $p_2 = 0.12$

$n_1 = 750$ $n_2 = 750$

- Test an appropriate hypothesis. Assume α= 0.01.

Fisher’s Exact Test Problem

- In RStudio

```
> proportion.test.twosample.exact.simple
```

McNemar’s Test for Change

- Suppose we test 100 randomly-selected units of product and find that 20% are defective. Then, imagine that we apply some type of treatment to the units; and on a post-test, we find again that 20% are defective.

McNemar’s Test for Change

- It is possible that the 20 units that were defective originally were still defective.

- But it is also possible that the 20 units that were defective on the second test were a completely different set of 20 units! It makes a difference.

McNemar's Test for Change

- McNemar's Test employs two unique features for testing the difference between two dependent sample proportions:
  - A special fourfold (2x2) contingency table
  - A special-purpose chi-square (χ^2) test statistic (the approximate test).

McNemar’s Test for Change

- McNemar’s Test employs two unique features for testing the difference between two dependent sample proportions:
  - A special fourfold (2x2) contingency table
  - A special-purpose chi-square (χ^2) test statistic (the approximate test).

McNemar's Test for Change

H_0: $Pass_1 Fail_2 = Fail_1 Pass_2$

		After Condition	
		Pass	Fail
Before Condition	Pass	a	**b**
	Fail	**c**	d

where

(a+b) + (c+d) = (a+c) + (b+d) = n = number of pairs of units evaluated and where df = 1

McNemar’s Test for Change

- Test Statistic

$$\chi^2 = \frac{\{ABS(b - c) - 1\}^2}{b + c}$$

McNemar’s Test for Change Example

- An operations manager in a manufacturing plant wishes to determine whether a new maintenance procedure is likely to improve the repeatability of a particular test at a test station.

McNemar’s Test for Change Example

- They select a random sample of 120 electronically tuned radios, which contain nonconforming as well as conforming units at the same level as daily production levels.

McNemar’s Test for Change Example

- They select a random sample of 120 electronically tuned radios, which contain nonconforming as well as conforming units at the same level as daily production levels.
- The entire sample is then tested.

McNemar’s Test for Change Example

- The maintenance procedure is performed and the test is repeated on the same sample of 120 radios.

McNemar's Test for Change Example

- In both tests, the radios are tested in a random order. They are also numbered with a unique identifier so the results of the two tests may be recorded for the proper units. Note that this is a repeated assessment on the same radios.

McNemar's Test for Change Example

- The summary data from the study appear as follows.

Number of Units	Status Before Maintenance	Status After Maintenance
4	Fail	Fail
4	Pass	Fail
56	Fail	Pass
56	Pass	Pass

McNemar’s Test for Change Example

- Place these data in the proper cells of the 2x2 contingency table before we demonstrate the test.

		After Condition	
		Pass	Fail
Before Condition	Pass	56	**4**
	Fail	**56**	4

McNemar’s Test for Change Example

Create a vector of the frequencies (counts)	`ct<-(a,c,b,d)`
Create a 2x2 contingency table	`matrix(ct,nrow = 2)`
Perform McNemar’s Test	`proportion.test.mcnemar.simple` `mcnemar.test`

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982

Two Independent Sample Hypothesis Test for Poisson Counts

Data Science for Quality Management: Two Sample Hypothesis Testing

with **Wendy Martin**

Learning objective:

Perform a statistical test for differences in counts (independent groups)

Two Independent Sample Poisson Count (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.

Poisson Rate Test Problem

- Based on the data collected by the team (in the file **Eddycur.dat**), can the team feel confident that their efforts have changed the number of minor indications per bar? Assume $\alpha = 0.05$.

Poisson Rate Test Problem

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

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982