Data Analysis & Pattern Recognition

Performance Evaluation. Type I and Type II Errors

Francesc Pozo

Escola d'Enginyeria de Barcelona Est (EEBE)
Universitat Politècnica de Catalunya (UPC)

Master's Degree in Chemical Engineering
Master's Degree in Interdisciplinary and Innovative Engineering

Motivating example

A structural damage detection indicator based on PCA and HT

IOP Publishing Smart Materials and Structures

Smart Mater. Struct. **23** (2014) 025014 (12pp) doi:10.1088/0964-1726/23/2/025014

A structural damage detection indicator based on principal component analysis and statistical hypothesis testing

L E Mujica[1], M Ruiz[1], F Pozo[1], J Rodellar[1,2] and A Güemes[3]

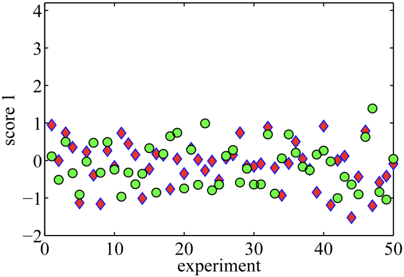
[1] CoDAlab[4] Departament de Matemàtica Aplicada III, Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona (EUETIB), Universitat Politècnica de Catalunya-BarcelonaTech (UPC), Comte d'Urgell, 187, E-08036 Barcelona, Spain
[2] CoDAlab[4], Departament de Matemàtica Aplicada III, Escola Tècnica Superior d'Enginyers de Camins, Canals i Ports de Barcelona (ETSECCPB), Universitat Politècnica de Catalunya-BarcelonaTech (UPC), Jordi Girona, 31, E-08034 Barcelona, Spain
[3] Center of Composites Materials and Smart Structures, Escuela Técnica Superior de Ingenieros Aeronáuticos, Universidad Politécnica de Madrid (UPM), Spain

E-mail: luis.eduardo.mujica@upc.edu, magda.ruiz@upc.edu, francesc.pozo@upc.edu, jose.rodellar@upc.edu and alfredo.guemes@upm.edu

Motivating example

A structural damage detection indicator based on PCA and HT

- In a classic application of the PCA strategy in the field of *structural health monitoring*, the *scores* allow a separation, clustering or visual grouping.
- However, in some cases, it can be clearly observed that a clustering, visual grouping, or separation cannot be performed.
- Therefore, more powerful and reliable tools are needed (HT, for instance!).



Hypothesis test

A hypothesis test is a decision criterion that allows to select between two complementary hypothesis.

Null hypothesis and alternative hypothesis

- Before conducting the hypothesis test, define the null hypothesis, H_0, which is assumed to be true prior to conducting the hypothesis test.
- The null hypothesis is compared to another hypothesis, called the alternative hypothesis, and denoted H_1.

Research hypothesis

- The alternative hypothesis is often called the research hypothesis since the theory or what is believed to be true about the parameter is specified in the alternative hypothesis.

A structural damage detection indicator based on PCA and HT

- In the motivating example:
  - If the null hypothesis (H_0) is accepted, the current structure is classified as healthy.
  - If the null hypothesis is rejected and the alternative hypothesis (H_1) is accepted, this would indicate the existence of some damage in the structure.

Parameter space

- H_0 and H_1 define complementary subsets of the parameter space Θ where the parameter θ is defined.
- The null hypothesis defines the region $[\theta \in \Theta_0]$ and the alternative hypothesis defines the region $[\theta \in \Theta_1]$

$$\Theta_0 \cap \Theta_1 = \emptyset$$
$$\Theta_0 \cup \Theta_1 = \Theta$$

Simple and composite hypothesis

- When a hypothesis uniquely specifies the distribution of the population from which the sample is taken, the hypothesis is said to be simple. For a simple hypothesis, $\Theta_0 = \{\theta_0\}$.
- Any hypothesis that is not a simple hypothesis is called a composite hypothesis.

Parameter space

Table 1: Form of hypothesis test

Null hypothesis	Alternative Hypothesis	Type of Alternative
	(A) $H_1 \ : \ \theta < \theta_0$	lower one-sided
$H_0 \ : \ \theta = \theta_0$	(B) $H_1 \ : \ \theta > \theta_0$	upper one-sided
	(C) $H_1 \ : \ \theta \neq \theta_0$	two-sided

Of the various combinations of hypothesis that could be examined, the case where H_0 is simple and H_1 is composite will be the focus of this lecture.

Example

If $H_0 \; : \; \pi = 0.4$ in a $b(\pi)$ (Bernoulli) distribution, the null hypothesis is simple since the hypothesis $H_0 \; : \; \pi = 0.4$ uniquely specifies the distribution as $b(0.4)$.

If $H_1 \; : \; \pi < 0.4$, the hypothesis is composite since π can take any value in the interval $[0, 0.4)$.

The goal in hypothesis testing

- The goal in hypothesis testing is to decide which one of the two hypothesis, H_0 and H_1, is true.
- To this end, split the sample space into two mutually exclusive subsets R and $\bar{R}$.
- R is the rejection region.
- $\bar{R}$ is the acceptance region.
- The critical value is the number that splits Θ into R and $\bar{R}$.
- To help decide, calculate a test statistic based on a sample.
- If the test statistic falls in the acceptance region, accept the null hypothesis.
- If the test statistic falls in the rejection region, reject the null hypothesis and accept the alternative hypothesis.

Example

The weight of a ball-bearing fluctuates between 1.5 g and 4.5 g. One wants to test whether the distribution of the weight for the ball bearing has a mean of either 2 g ($H_0 : \mu = 2$) or 2.5 g ($H_1 : \mu = 2.5$).

A random sample of size one is taken. If the weight of the ball bearing is greater than 2.3 g, the null hypothesis that the mean weight of the ball-bearing is 2 g is rejected, and the alternative hypothesis that the mean weight of the ball-bearing is 2.5 g is accepted.

Specify the sample space, the rejection region, the acceptance region and the critical value.

Example

The weight of a ball-bearing fluctuates between 1.5 g and 4.5 g. One wants to test whether the distribution of the weight for the ball bearing has a mean of either 2 g ($H_0 : \mu = 2$) or 2.5 g ($H_1 : \mu = 2.5$).

A random sample of size one is taken. If the weight of the ball bearing is greater than 2.3 g, the null hypothesis that the mean weight of the ball-bearing is 2 g is rejected, and the alternative hypothesis that the mean weight of the ball-bearing is 2.5 g is accepted.

Specify the sample space, the rejection region, the acceptance region and the critical value.

$$\boxed{\Theta = [1.5, 4.5],\ R = (2.3, 4.5],\ \bar{R} = [1.5, 2.3],\ cv = 2.3}$$

Type I and Type II Errors

"The decision one reaches using a hypothesis test is always subject to error."

To get a better grasp on the errors one might make with a hypothesis test, consider the following hypothetical legal situation.

In the United States' judicial system, as well as in Spain's, an individual is considered innocent until proven guilty of an offense.

Table 2: Possible outcomes and their consequences for a trial by jury

	True state of the defendant	
Jury's decision	H_0 true (innocent)	H_0 false (guilty)
Accept H_0 (not guilty)	Correct (A)	Error (B)
Reject H_0 (guilty)	Error (C)	Correct (D)

Jury's decision	True state of the defendant H_0 true (innocent)	H_0 false (guilty)
Accept H_0 (not guilty)	Correct (A)	Error (B)
Reject H_0 (guilty)	Error (C)	Correct (D)

Case A. Correct

If the defendant is innocent and the jury decides the defendant is not guilty of the charge, the jury's decision is correct.

Type I and Type II Errors

Jury's decision	True state of the defendant H_0 true (innocent)	H_0 false (guilty)
Accept H_0 (not guilty)	Correct (A)	Error (B)
Reject H_0 (guilty)	Error (C)	Correct (D)

Case B. Error

By failing to reject a false null hypothesis, an error has been made. In statistics, this error is called a type II error. In the legal scenario, a type II error is made when a guilty person is not convicted.

β

The probability of committing a type II error is β.

	True state of the defendant	
Jury's decision	H_0 true (innocent)	H_0 false (guilty)
Accept H_0 (not guilty)	Correct (A)	Error (B)
Reject H_0 (guilty)	Error (C)	Correct (D)

Case C. Error

By rejecting a true null hypothesis, an error has been made. In statistics, this type of error is called a type I error. In the legal example, a type I error would be to convict an innocent defendant.

α

The probability of committing a type I error is α.

Jury's decision	True state of the defendant H_0 true (innocent)	H_0 false (guilty)
Accept H_0 (not guilty)	Correct (A)	Error (B)
Reject H_0 (guilty)	Error (C)	Correct (D)

Case D. Correct

If the null hypothesis is false and it is rejected, the decision is correct. In the legal arena, this translates into a jury convicting a guilty defendant.

Level of significance

- The probability of committing a type I error is called the level of significance for a hypothesis test.
- The level of significance is also known as the size of the test and is denoted by α, where

$$\begin{aligned}\alpha &= \mathbb{P}(\text{type I error}) = \mathbb{P}(\text{reject } H_0 \mid H_0 \text{ is true}) \\ &= \mathbb{P}(\text{accept } H_1 \mid H_0 \text{ is true})\end{aligned}$$

- The probability of committing a type II error is β, where

$$\begin{aligned}\beta &= \mathbb{P}(\text{type II error}) = \mathbb{P}(\text{fail to reject } H_0 \mid H_0 \text{ is false}) \\ &= \mathbb{P}(\text{accept } H_0 \mid H_1 \text{ is true})\end{aligned}$$

Level of significance

Table 3: Relationship between type I and type II errors

Decision	Null hypothesis True	Null hypothesis False
Accept H_0	$\mathbb{P}(\text{accept } H_0 \mid H_0) = 1 - \alpha$	$\mathbb{P}(\text{accept } H_0 \mid H_1) = \beta$
Reject H_0	$\mathbb{P}(\text{reject } H_0 \mid H_0) = \alpha$	$\mathbb{P}(\text{reject } H_0 \mid H_1) = 1 - \beta$

The power of the test

$1 - \beta$ is also known as the power of the test.

Example

Given a normal distribution with unknown mean μ and known standard deviation $\sigma = 2$, one wishes to test

$$H_0 \ : \ \mu = 1 \quad \text{versus} \quad H_1 \ : \ \mu = 4$$

A sample of size one is taken where $R = (2, +\infty)$.

Determine α and β for this experiment.

Determine α

$$\alpha = \mathbb{P}(\text{reject } H_0 \mid H_0) = \mathbb{P}(X_1 > 2 \mid N(1,2)) = \mathbb{P}\left(\frac{X_1 - 1}{2} > \frac{2-1}{2}\right)$$
$$= \mathbb{P}(Z > 0.5) = 1 - \mathbb{P}(Z \leq 0.5), \quad Z \hookrightarrow N(0,1)$$

Type I and Type II Errors: Level of significance

Example

Given a normal distribution with unknown mean μ and known standard deviation $\sigma = 2$, one wishes to test

$$H_0 \ : \ \mu = 1 \quad \text{versus} \quad H_1 \ : \ \mu = 4$$

A sample of size one is taken where $R = (2, +\infty)$.

Determine α and β for this experiment.

Determine α

```
>>from scipy.stats import norm
>>1-norm.cdf(2, loc=1, scale=2)
  0.3085375387259869
```

Example

Given a normal distribution with unknown mean μ and known standard deviation $\sigma = 2$, one wishes to test

$$H_0 \;:\; \mu = 1 \quad \text{versus} \quad H_1 \;:\; \mu = 4$$

A sample of size one is taken where $R = (2, +\infty)$.

Determine α and β for this experiment.

Determine β

$$\beta = \mathbb{P}(\text{accept } H_0 \mid H_1) = \mathbb{P}(X_1 \leq 2 \mid N(4,2)) = \mathbb{P}\left(\frac{X_1 - 4}{2} \leq \frac{2-4}{2}\right)$$

$$= \mathbb{P}(Z \leq -1), \quad Z \hookrightarrow N(0,1)$$

Example

Given a normal distribution with unknown mean μ and known standard deviation $\sigma = 2$, one wishes to test

$$H_0 \; : \; \mu = 1 \quad \text{versus} \quad H_1 \; : \; \mu = 4$$

A sample of size one is taken where $R = (2, +\infty)$.

Determine α and β for this experiment.

Determine β

```
>>from scipy.stats import norm
>>norm.cdf(2, loc=4, scale=2)
  0.15865525393145707
```

Power function

- Given a composite alternative hypothesis $H_1 \ : \ \theta \in \Theta_1$, the power of the test, $\text{POWER}(\theta)$, is

$$\begin{aligned}\text{POWER}(\theta) &= \mathbb{P}(\text{reject } H_0 \mid H_0 \text{ is false}) = \mathbb{P}(\text{accept } H_1 \mid H_1) \\ &= 1 - \beta(\theta),\end{aligned}$$

  where $\beta(\theta)$ is the probability of a type II error at a given θ.
- The power of a test is the probability the test detects differences when differences exist.
- Note that $\text{POWER}(\theta)$ is a function of the parameter θ:

$$\begin{aligned}\text{POWER} : \Theta_1 &\mapsto [0, 1] \subset \mathbb{R} \\ \theta &\mapsto \text{POWER}(\theta)\end{aligned}$$

Power Function

Example

Given the density function

$$f(x;\theta) = \theta e^{-\theta x},\ x \geq 0,\ \theta > 0,$$

(a) Consider a test of hypothesis where $H_0 \ :\ \theta = 2$ versus $H_1 \ :\ \theta > 2$. Using a random sample of size one, find the critical value k such that the test is conducted at the $\alpha = 0.05$ level.

(b) Further, determine the power function of this test.

Determine the critical value k

$$\alpha = \mathbb{P}(X_1 > k \mid H_0) = \int_k^{+\infty} f(x_1;2)dx_1 = e^{-2k} = 0.05 \ \Rightarrow\ k = 1.4979$$

Example

Given the density function

$$f(x;\theta) = \theta e^{-\theta x},\ x \geq 0,\ \theta > 0,$$

(a) Consider a test of hypothesis where $H_0 : \theta = 2$ versus $H_1 : \theta > 2$. Using a random sample of size one, find the critical value k such that the test is conducted at the $\alpha = 0.05$ level.

(b) Further, determine the power function of this test.

Determine the critical value k

```
>>from scipy.stats import expon
>>expon.ppf(0.95, loc=0, scale=1/2)
  1.497866136776995
```

Example

Given the density function

$$f(x;\theta) = \theta e^{-\theta x},\ x \geq 0,\ \theta > 0,$$

(a) Consider a test of hypothesis where $H_0 \ : \ \theta = 2$ versus $H_1 \ : \ \theta > 2$. Using a random sample of size one, find the critical value k such that the test is conducted at the $\alpha = 0.05$ level.

(b) Further, determine the power function of this test.

Determine the power function of this test

$$\text{POWER}(\theta) = \mathbb{P}(\text{reject } H_0 \mid H_1) = \mathbb{P}(X_1 > k \mid H_1)$$
$$= \int_k^{+\infty} f(x_1;\theta)dx_1 = e^{-\theta k}$$

Example

Test the null hypothesis that for a certain age group the mean score on an achievement test (scores follow a normal distribution with $\sigma = 6$) is equal to 40 against the alternative that it is not equal to 40.

(a) Find the probability of type I error for $n = 9$ if the null hypothesis is rejected when the sample mean is less than 36 or greater than 44.

About the sample mean

If $X_i \hookrightarrow N(\mu, \sigma),\ i = 1, 2, \ldots, n$, then

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \hookrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Example

Test the null hypothesis that for a certain age group the mean score on an achievement test (scores follow a normal distribution with $\sigma = 6$) is equal to 40 against the alternative that it is not equal to 40.

(a) Find the probability of type I error for $n = 9$ if the null hypothesis is rejected when the sample mean is less than 36 or greater than 44.

Type I error

$$
\begin{aligned}
\alpha &= \mathbb{P}(\bar{X} < 36 \mid H_0) + \mathbb{P}(\bar{X} > 44 \mid H_0) \\
&= \mathbb{P}\left(\bar{X} < 36 \mid N(40, 6/\sqrt{9})\right) + \mathbb{P}\left(\bar{X} > 44 \mid N(40, 6/\sqrt{9})\right) \\
&= \mathbb{P}(Z < -2) + \mathbb{P}(Z > 2),\ Z \hookrightarrow N(0,1)
\end{aligned}
$$

Example

Test the null hypothesis that for a certain age group the mean score on an achievement test (scores follow a normal distribution with $\sigma = 6$) is equal to 40 against the alternative that it is not equal to 40.

(a) Find the probability of type I error for $n = 9$ if the null hypothesis is rejected when the sample mean is less than 36 or greater than 44.

Type I error

```
>>from scipy.stats import norm
>>norm.cdf(-2, loc=0, scale=1)+1-norm.cdf(2, loc=0, scale=1)
  0.04550026389635806
```

Example

Test the null hypothesis that for a certain age group the mean score on an achievement test (scores follow a normal distribution with $\sigma = 6$) is equal to 40 against the alternative that it is not equal to 40.

(b) Find the probability of type I error for $n = 36$ if the null hypothesis is rejected when the sample mean is less than 38 or greater than 42.

Type I error

$$\begin{aligned}\alpha &= \mathbb{P}(\bar{X} < 38 \mid H_0) + \mathbb{P}(\bar{X} > 42 \mid H_0) \\ &= \mathbb{P}\left(\bar{X} < 38 \mid N(40, 6/\sqrt{36})\right) + \mathbb{P}\left(\bar{X} > 42 \mid N(40, 6/\sqrt{36})\right) \\ &= \mathbb{P}(Z < -2) + \mathbb{P}(Z > 2),\ Z \hookrightarrow N(0,1)\end{aligned}$$

Example

Test the null hypothesis that for a certain age group the mean score on an achievement test (scores follow a normal distribution with $\sigma = 6$) is equal to 40 against the alternative that it is not equal to 40.

(b) Find the probability of type I error for $n = 36$ if the null hypothesis is rejected when the sample mean is less than 38 or greater than 42.

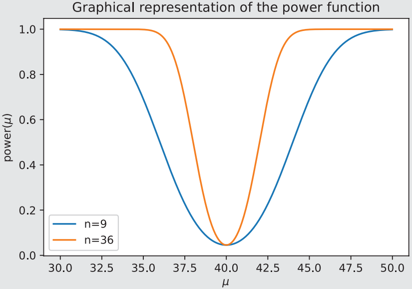
Type I error

```
>>from scipy.stats import norm
>>norm.cdf(-2, loc=0, scale=1)+1-norm.cdf(2, loc=0, scale=1)
  0.04550026389635806
```

Example

Test the null hypothesis that for a certain age group the mean score on an achievement test (scores follow a normal distribution with $\sigma = 6$) is equal to 40 against the alternative that it is not equal to 40.

(c) Plot the power functions for $n = 9$ and $n = 36$ for values of μ between 30 and 50. Note that $\text{POWER}(\mu_0) = \alpha$.



Python code

```
%matplotlib inline
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
def power9(mu):
    return norm.cdf((36-mu)/2, loc=0, scale=1)+(1-norm.cdf((44-
    mu)/2, loc=0, scale=1))
def power36(mu):
    return norm.cdf((38-mu)/1, loc=0, scale=1)+(1-norm.cdf((42-
    mu)/1, loc=0, scale=1))
rr = np.arange(30, 50, 0.01)
plt.xlabel('$\mu$')
plt.ylabel('power($\mu$)')
plt.title('Graphical representation of the power function')
plt.plot(rr, power9(rr),label='n=9')
plt.plot(rr, power36(rr),label='n=36')
plt.legend()
plt.savefig('plot.eps', dpi=300, bbox_inches='tight')
plt.show()
```

Power Function: Homework[1]

Example

Given a $N(\mu, 1)$ population from which one takes a simple random sample of size 1, test the null hypothesis $H_0 : \mu = 1$ versus the alternative hypothesis $H_1 : \mu = 2$.

Determine the significance level and the power of the test for the following rejection regions:

(a) $(2.036, +\infty)$

(b) $(1.100, 1.300) \cup (2.461, +\infty)$.

Remark

Note that tests with identical α values do not necessarily have identical power for a fixed sample size.

[1] To be handed in on or before the next lecture as a single PDF file. You can work in pairs.

p-Value or critical level

p-value

- The critical level or p-value is defined as the probability of observing a difference as extreme or more extreme that the difference observed under the assumption that the null hypothesis is true.
- The p-value is not fixed *a priori*, but rather is determined after the sample is taken.
- Given a fixed significance level α, reject H_0 whenever the p-value$< \alpha$.

p-value

Table 4: Calculation of p-values for continuous distributions

	p-value
$H_1 : \theta < \theta_0$	$\mathbb{P}(T < t_{obs} \mid H_0)$
$H_1 : \theta > \theta_0$	$\mathbb{P}(T > t_{obs} \mid H_0)$
$H_1 : \theta \neq \theta_0$	$2\min\{\mathbb{P}(T < t_{obs} \mid H_0), \mathbb{P}(T > t_{obs} \mid H_0)\}$

Confusion matrix (hypothesis testing)

	Null hypothesis	
Decision	True	False
Accept H_0	correct	type II error
Reject H_0	type I error	correct

Confusion matrix (predictive analytics)

	Actual class	
Predicted class	Positive	Negative
Positive	true positive (tp)	false positive (fp)
Negative	false negative (fn)	true negative (tn)

Confusion matrix (predictive analytics)

	Actual class	
Predicted class	Positive	Negative
Positive	true positive (tp)	false positive (fp)
Negative	false negative (fn)	true negative (tn)

Accuracy

- Accuracy is one metric for evaluating classification models.

$$\boxed{\text{acc} = \frac{\text{tp} + \text{fn}}{\text{tp} + \text{fp} + \text{fn} + \text{tn}}}$$

- Accuracy alone does not tell the full story when you are working with a class-imbalanced data set.

Confusion matrix (predictive analytics)

	Actual class	
Predicted class	Positive	Negative
Positive	true positive (tp)	false positive (fp)
Negative	false negative (fn)	true negative (tn)

Precision

- Precision (or positive predictive value (ppv) attempts to answer the following question: "What proportion of positive identifications was actually correct?"

$$\text{ppv} = \frac{\text{tp}}{\text{tp} + \text{fp}}$$

Confusion matrix (predictive analytics)

	Actual class	
Predicted class	Positive	Negative
Positive	true positive (tp)	false positive (fp)
Negative	false negative (fn)	true negative (tn)

Recall

- Recall (or sensitivity, hit rate, true positive rate (tpr)) attempts to answer the following question: "What proportion of actual positives was identified correctly?"

$$\boxed{\text{tpr} = \frac{\text{tp}}{\text{tp} + \text{fn}}}$$

Confusion matrix (predictive analytics)

Predicted class	Actual class: Positive	Actual class: Negative
Positive	true positive (tp)	false positive (fp)
Negative	false negative (fn)	true negative (tn)

$F1$ score

- $F1$ score is defined as the harmonic mean of precision (ppv) and recall (tpr):

$$F_1 = \left(\frac{\frac{1}{\text{ppv}} + \frac{1}{\text{tpr}}}{2} \right)^{-1} = \frac{2 \cdot \text{ppv} \cdot \text{tpr}}{\text{ppv} + \text{tpr}} = \frac{2\text{tp}}{2\text{tp} + \text{fn} + \text{fp}}$$

Confusion matrix (predictive analytics)

	Actual class	
Predicted class	Positive	Negative
Positive	true positive (tp)	false positive (fp)
Negative	false negative (fn)	true negative (tn)

Specificity

- Specificity (or selectivity, true negative rate (tnr)) attempts to answer the following question: "What proportion of negative identifications was actually correct?"

$$\boxed{\text{tnr} = \frac{\text{tn}}{\text{fp} + \text{tn}}}$$

Tests of Significance

- `Step 1`. **Hypotheses**. State the null and alternative hypotheses.
- `Step 2`. **Test Statistic**. Select and appropriate test statistic and determine the sampling distribution of the test statistic or the standardized test statistic under the assumption that the null hypothesis is true.
- `Step 3`. **Rejection Region Calculations**. Use the specified α level to compute the critical value and to determine the rejection region for the standardized test statistic. Then calculate the value of the statistic observed from the sample, $t(\mathbf{x}) = t_{\text{obs}}$.
- `Step 4`. **Statistical Conclusion**. Use the rejection region of the p-value to determine if the evidence warrants rejecting the null hypothesis.

Test for the population mean when sampling from a normal distribution with unknown population variance

Example

A random sample of size $n = 25$ is taken from a distribution known to be $N(\mu, \sigma)$. If the $\sum_{i=1}^{n} x_i = 100$ and the $\sum_{i=1}^{n} x_i^2 = 600$,

(a) Test the null hypothesis $H_0 \ : \ \mu = 2.5$ versus the alternative hypothesis $H_1 \ : \ \mu \neq 2.5$ at the $\alpha = 0.05$ significance level.

Step 1. Hypothesis

These are given in the problem as

$$H_0 \ : \mu = 2.5 \quad \text{versus}$$
$$H_1 \ : \mu \neq 2.5$$

Test for the population mean when sampling from a normal distribution with unknown population variance

Example

A random sample of size $n = 25$ is taken from a distribution known to be $N(\mu, \sigma)$. If the $\sum_{i=1}^{n} x_i = 100$ and the $\sum_{i=1}^{n} x_i^2 = 600$,

(a) Test the null hypothesis $H_0 \ : \ \mu = 2.5$ versus the alternative hypothesis $H_1 \ : \ \mu \neq 2.5$ at the $\alpha = 0.05$ significance level.

Step 2. Test Statistic

The test statistic chosen is $\bar{X}$ because $E[\bar{X}] = \mu$. The value for this test statistic is $\bar{x} = \left(\sum_{i=1}^{n} x_i\right)/n = 4$. The standardized test statistic and its distribution under the assumption H_0 is true are

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \hookrightarrow t_{25-1}.$$

Test for the population mean when sampling from a normal distribution with unknown population variance

Step 3. Rejection Region Calculations

Because the standardized test statistic is distributed t_{24}, and H_1 is a two-tailed hypothesis, the rejection region is

$$|t_{\text{obs}}| > t_{1-0.05/2;24} = t_{0.975;24} = 2.0639$$

where $t_{0.975;24}$ satisfies

$$\mathbb{P}(T < t_{0.975;24}) = 0.975, \ T \hookrightarrow t_{24}$$

```
>> from scipy.stats import t
>> t.ppf(0.975,24)
   2.0638985616280205
```

Test for the population mean when sampling from a normal distribution with unknown population variance

Step 3. Rejection Region Calculations

The value of the standardized test statistic is

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{4 - 2.5}{2.89/\sqrt{25}} = 2.5981$$

where the value for s is calculated as

$$s = \sqrt{\frac{\sum_{i=1}^{2} x_i^2 - n\bar{x}^2}{n-1}} = 2.8868$$

Test for the population mean when sampling from a normal distribution with unknown population variance

Step 4. Statistical Conclusion

From the rejection region, reject H_0 because

$$\boxed{|t_{\text{obs}}| = 2.5981 > 2.0639 = t_{0.975;24}}$$

The p-value is $2 \cdot \mathbb{P}(t_{24} \geq t_{\text{obs}}) = 0.015772855749001335 < \alpha$

```
>> from scipy.stats import t
>> 2*(1-t.cdf(2.598076211353316,24))
   2.0638985616280205
```

From the p-value, reject H_0 because the p-value is less than α.

There is evidence to suggest that the mean is not equal to 2.5.

Test for the population mean when sampling from a normal distribution with unknown population variance

Example

One-sample t-test: Fertilizers. A farmer wants to test if a new brand of fertilizer increases his wheat yields per plot. He puts the new fertilizer on 15 equal plots and records the subsequent yields for the 15 plots. If his traditional yield is two bushels per plot, conduct a test of significance for μ at the $\alpha = 0.05$ significance level assuming the data follow a normal distribution. The yields for the 15 yields are

2.5	3.0	3.1	4.0	1.2
5.0	4.1	3.9	3.2	3.3
2.8	4.1	2.7	2.9	3.7

The python function `stats.ttest_1samp` cannot be used in this example, since the alternative hypothesis is always two-tailed.

[2] To be handed in on or before the next lecture as a single PDF file. You can work in pairs.

Test for the difference in population means when sampling from independent normal distributions with known variances

- The null hypothesis for testing the difference between two means is

$$H_0 \; : \; \mu_X - \mu_Y = \delta_0,$$

and the standardized test statistic under the assumption that H_0 is true is

$$Z = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}} \hookrightarrow N(0, 1)$$

Test for the difference in population means when sampling from independent normal distributions with known variances

Example

A researcher wishes to see if it reasonable to believe that engineering majors have higher math SAT scores than English majors. She takes two random samples.

The first sample consists of 64 engineering majors' SAT math scores (X). Typically, these scores follow a normal distribution with a known standard deviation of $\sigma_X = 100$ but with an unknown mean.

The second sample consists of 144 observations of English majors' SAT scores (Y). These also follow a normal distribution with a standard deviation of $\sigma_Y = 108$ with an unknown mean as well.

(a) Test the null hypothesis of equality of means at the 10% significance level ($\alpha = 0.1$) knowing the difference in sample means is 20.

Test for the difference in population means when sampling from independent normal distributions with known variances

Step 1. Hypotheses

$$H_0 \;:\; \mu_X - \mu_Y = 0$$
$$H_1 \;:\; \mu_X - \mu_Y > 0$$

Step 2. Test Statistic

$$\frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}} \hookrightarrow N(0,1)$$

Test for the difference in population means when sampling from independent normal distributions with known variances

Step 3. Rejection Region Calculations

Since H_1 is an upper one-sided hypothesis, the rejection region is

$$z_{\text{obs}} > z_{1-\alpha} = z_{0.9} = 1.2816$$

where $z_{0.9}$ satisfies

$$\mathbb{P}(Z < z_{0.9}) = 0.9, \; Z \hookrightarrow N(0,1)$$

```
>> from scipy.stats import norm
>> norm.ppf(0.9, loc=0, scale=1)
   1.2815515655446004
```

Tests of Significance

Test for the difference in population means when sampling from independent normal distributions with known variances

Step 3. Rejection Region Calculations

The value of the standardized test statistic is

$$z_{\text{obs}} = \frac{\bar{x} - \bar{y} - \delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} = \frac{20 - 0}{\sqrt{\frac{100^2}{64} + \frac{108^2}{144}}} = 1.2985$$

Step 4. Statistical Conclusion

The p-value is $\mathbb{P}(Z \geq z_{\text{obs}}) = 0.0971 < 0.10 = \alpha$

```
>> from scipy.stats import norm
>> 1-norm.cdf(1.2985,loc=0,scale=1)
   0.09705778816669874
```

The evidence suggests engineering majors have a higher average math SAT score.

Test for the difference in means when sampling from independent normal distributions with variances that are unknown but assumed equal[3]

- The null hypothesis for testing the difference between two means is $H_0 \;:\; \mu_X - \mu_Y = \delta_0$ and the standardized test statistic under the assumption that H_0 is true is

$$T = \frac{\left[(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)\right]}{\sqrt{S_p^2 \left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \hookrightarrow t_{n_X + n_Y - 2},$$

where

$$S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}$$

[3] We consider two samples of size n_X and n_Y taken from two normal distributions $N(\mu_X, \sigma)$ and $N(\mu_Y, \sigma)$, where σ is unknown.

Test for the difference in means when sampling from independent normal distributions with variances that are unknown but assumed equal

Example

A questionnaire is devised by the Board of Governors to measure the level of satisfaction for graduates from two competing state schools.

Past history indicates the variance in satisfaction levels for both schools is equal.

The questionnaire is randomly administered to 11 students from State School X $\{69, 75, 76, 80, 81, 82, 86, 89, 91, 92, 97\}$ and 15 students from State School Y $\{59, 62, 66, 70, 70, 75, 75, 77, 78, 79, 81, 84, 84, 86, 94\}$.

(a) Test to see if there are significant differences between the mean satisfaction levels for graduates of the two competing state schools using a significance level of 5%.

Test for the difference in means when sampling from independent normal distributions with variances that are unknown but assumed equal

To compute the value of the standardized test statistic and its corresponding p-value with Python, key in

```
>> import numpy as np
>> from scipy import stats
>> x = np.array([69,75,76,80,81,82,86,89,91,92,97])
>> y = np.array([59,62,66,70,70,75,75,77,78,79,81,84,84,86,94])
>> stats.ttest_ind(x, y, equal_var = True)
   Ttest_indResult(statistic=2.07977794988124, pvalue
    =0.04839673294633766)
```

From the p-value, reject H_0 because the p-value is less than 0.05.

There is evidence to suggest the average satisfaction levels between State School X and State School Y are different.

Test for the difference in means when sampling from independent normal distributions with variances that are unknown and unequal[4]

- The null hypothesis for testing the difference between two means is $H_0 : \mu_X - \mu_Y = \delta_0$ and the standardized test statistic under the assumption that H_0 is true is

$$T = \frac{\left[(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)\right]}{\sqrt{\left(\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}\right)}} \overset{\approx}{\hookrightarrow} t_\nu,$$

where

$$\nu = \left\lfloor \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{(s_X^2/n_X)^2}{n_X - 1} + \frac{(s_Y^2/n_Y)^2}{n_Y - 1}} \right\rfloor$$

[4]We consider two samples of size n_X and n_Y taken from two normal distributions $N(\mu_X, \sigma_X)$ and $N(\mu_Y, \sigma_Y)$, where both σ_X and σ_Y are unknown.

Tests of Significance: Homework[5]

Test for the difference in means when sampling from independent normal distributions with variances that are unknown and unequal

Example

A bottled water company acquires its water from two independent sources X and Y.

The company suspects that the sodium content in the water from source X is less than the sodium content for water from source Y.

An independent agency measures the sodium content in 20 samples from source X and 10 samples from source Y.

Is there statistical evidence to suggest the average sodium content in the water from source X is less than the average sodium content in the water from source Y?

The measurements for the sodium values are mg/l. Use an α level of 0.05 to test the appropriate hypotheses.

[5] To be handed in on or before the next lecture as a single PDF file. You can work in pairs.

Test for the difference in means when sampling from independent normal distributions with variances that are unknown and unequal

Example

Source X: 84 73 92 84 95 74 80 86 80 77

86 72 62 54 77 63 85 59 66 79

Source Y: 78 79 84 82 80 85 81 83 79 81

The python function `stats.ttest_ind` cannot be used in this example, since in the python function the alternative hypothesis is always two-tailed.

Kolmogorov-Smirnov Goodness-of-Fit Test

The Kolmogorov-Smirnov goodness-of-fit test (K-S test) compares your data with a known distribution and lets you know if they have the same distribution. Although the test is nonparametric —it doesn?t assume any particular underlying distribution— it is commonly used as a test for normality to see if your data is normally distributed. It is also used to check the assumption of normality in analysis of variance.

Example

Test whether the observations $5, 6, 7, 8$, and 9 are from a normal distribution with $\mu = 6.5$ and $\sigma = \sqrt{2}$. That is, the hypothesized distribution is

$$F_0(x) \hookrightarrow N(6.5, \sqrt{2})$$

Kolmogorov-Smirnov Goodness-of-Fit Test

Example

Test whether the observations 5, 6, 7, 8, and 9 are from a normal distribution with $\mu = 6.5$ and $\sigma = \sqrt{2}$. That is, the hypothesized distribution is

$$F_0(x) \hookrightarrow N(6.5, \sqrt{2})$$

```
>> from scipy import stats
>> import numpy as np
>> x = np.array([5,6,7,8,9])
>> mu = 6.5
>> sigma = np.sqrt(2)
>> kstest1 = stats.kstest(x, 'norm', args=(mu,sigma))
>> print(kstest1)
   KstestResult(statistic=0.2555778168267576, pvalue
    =0.899636441822264)
```

Kolmogorov-Smirnov Goodness-of-Fit Test

Example

Test whether the observations 5, 6, 7, 8, and 9 are from a normal distribution with $\mu = \mu_X$ and $\sigma = \sigma_X$. That is, the hypothesized distribution is

$$F_0(x) \hookrightarrow N(\mu_X, \sigma_X)$$

```
>> from scipy import stats
>> import numpy as np
>> x = np.array([5,6,7,8,9])
>> mu = x.mean()
>> sigma = x.std()
>> kstest2 = stats.kstest(x, 'norm',args=(mu,sigma))
>> print(kstest2)
   KstestResult(statistic=0.16024993890652328, pvalue
    =0.9995301060544028)
```

Homework[6]

Example

Test whether the data in the examples in the previous slides are from a normal distribution:

- Fertilizers (on page 48);
- Satisfaction levels (on page 55);
- Sodium content (on page 58).

[6] To be handed in on or before the next lecture as a single PDF file. You can work in pairs.

Laboratory Session

Check and try to understand the Python code in the slides

- Page 19
- Page 21
- Page 24
- Page 28
- Page 30
- Page 32
- Page 45
- Page 45
- Page 47
- Pages 52 and 53
- Page 56
- Page 61
- Page 63

Use Python to solve the following problems

- Homework in page 33
- Homework in page 48
- The example in page 55 is solved using the Python function `stats.ttest_ind`. Solve the problem again without this function, finding the critical value, the p-value and t_{obs}
- Homework in page 58
- Homework in page 63