



Buenas prácticas docentes en el uso de herramientas digitales

28

Papel cero en las asignaturas de estadística: Python como lenguaje vehicular

Paper Zero in Statistics Courses: Python as a Vehicular Language



AUTOR

Eugenio Francisco Sánchez Úbeda

eugenio.sanchez@comillas.edu

Departamento de Organización Industrial

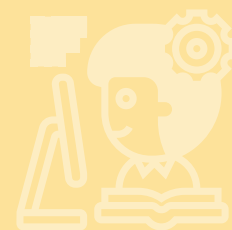
Instituto de Investigación Tecnológica (IIT). Escuela Técnica Superior de Ingeniería (ICAI)



PALABRAS CLAVE | KEYWORDS

Estadística, ingeniería, jupyter, probabilidad, Python.

Statistics, engineering, jupyter, probability, Python.



RESUMEN

Es muy habitual, dentro de las titulaciones de ingeniería, que las asignaturas de estadística básica se impartan con un enfoque tradicional basado en la resolución de problemas con lápiz y papel, con la ayuda de una calculadora. Sin embargo, en el ejercicio profesional de la ingeniería, siempre se utiliza algún tipo de software estadístico.

En la actualidad el uso de cuadernos Jupyter con lenguaje Python se ha convertido en un estándar en las empresas, siendo una competencia muy demandada.

¿No sería posible impartir estas asignaturas de Estadística con un enfoque innovador que permita a los alumnos desarrollar las competencias necesarias para poder resolver problemas reales sin descuidar el aprendizaje profundo de los conceptos estadísticos?

La asignatura anual de Probabilidad y Estadística, del reciente Grado en Ingeniería Matemática e Inteligencia Artificial (IMAT) en Comillas ICAI, se está impartiendo utilizando Python como lenguaje vehicular. El aprendizaje observado en los alumnos, y su opinión sugiere que ha sido un acierto.

Todos los exámenes se están realizando utilizando Python con cuadernos Jupyter. No hay papel. En muchos exámenes los alumnos se han enfrentado a problemas reales con miles de datos y trabajando con modelos estadísticos complejos, únicamente manejables con un entorno software como el utilizado.

ABSTRACT

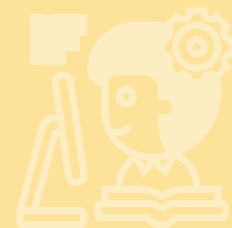
Within engineering programs, it is common for basic statistics courses to adopt a traditional teaching approach, focused on solving problems with pen and paper, aided by a calculator. However, in the professional practice of engineering, some form of statistical software is always used.

Currently, using Jupyter notebooks with the Python language has become a standard in the industry, and it is a highly sought-after skill.

Wouldn't it be possible to teach these statistics courses with an innovative approach that allows students to develop the necessary skills to solve real-world problems without neglecting a deep understanding of statistical concepts?

The annual Probability and Statistics course in the recent Bachelor's in Mathematical Engineering and Artificial Intelligence (IMAT) at Comillas ICAI is being taught using Python as the primary language. The learning observed in the students, as well as their feedback, suggests that this has been a successful approach.

All exams are being conducted using Python with Jupyter notebooks. There is no paper involved. In many exams, students have tackled real problems using thousands of data points and working with complex statistical models, which are only manageable with the software environment used.



1. Motivación y fundamentación

Es muy habitual, dentro de las titulaciones de ingeniería, que las asignaturas de estadística básica se impartan con un enfoque tradicional basado en la resolución de problemas en la pizarra, a mano, con la ayuda de una calculadora si fuese necesario. Son problemas preparados para poder ser resueltos de esta forma en un tiempo razonable. Esto se traslada directamente a los exámenes, en los que solo se permite utilizar lápiz y papel y una calculadora sencilla. Para conectar el pasado con el presente, se realizan prácticas de manera puntual en el ordenador, utilizando software estadístico para ofrecer a los alumnos una aproximación a lo que se puede hacer en la actualidad.

Sin embargo, en el ejercicio profesional de la ingeniería, siempre que se utilizan técnicas estadísticas para resolver problemas se hace mediante algún tipo de software estadístico.

Entonces, surge la pregunta ¿no sería posible impartir estas asignaturas con un enfoque innovador que permita a los alumnos desarrollar las competencias necesarias para poder resolver problemas reales sin descuidar el aprendizaje profundo de los conceptos?

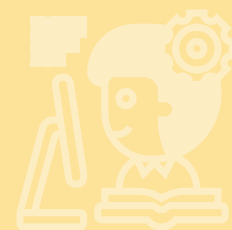
2. Objetivos

El objetivo fundamental es impartir una asignatura de Estadística en ingeniería utilizando el lenguaje Python como herramienta de trabajo básica para la resolución de los problemas. Utilizando Python como lenguaje vehicular se pretendía conseguir:

- ➔ Eliminar la limitación de solo poder realizar problemas que puedan ser resueltos con lápiz y papel y una calculadora sencilla, tanto en clase como en los exámenes.
- ➔ Dotar al alumno con una herramienta de trabajo potente que permita enfrentarse a conceptos complicados de forma empírica, empleando técnicas de estadística avanzada y fomentando el auto-aprendizaje.
- ➔ Dotar con conocimientos prácticos adicionales al alumno sobre temas colaterales a la asignatura de Estadística, como son el tratamiento y manejo de los datos o la generación de buenos gráficos para su análisis.

Además, se esperaba conseguir otros efectos beneficiosos en el aprendizaje de los estudiantes, que desarrollarían de forma inconsciente otras competencias muy interesantes:

- ➔ Fomentar el uso del lenguaje de programación Python como la herramienta de elección para resolver problemas en cualquier ámbito.
- ➔ Desarrollar competencias avanzadas sobre el uso de un entorno software de trabajo muy demandado en la actualidad por las empresas.



- ➔ Afianzar y ampliar los conocimientos de programación en Python adquiridos en la asignatura correspondiente.
- ➔ Abrir la mente al *machine learning*, objeto de otras asignaturas de cursos superiores y que hoy en día se trabaja utilizando este tipo de entornos software.

3. Contexto de aplicación

En el curso 2021-2022 comenzó a impartirse el nuevo Grado en Ingeniería Matemática e Inteligencia Artificial (IMAT) en Comillas ICAI. La asignatura anual de Probabilidad y Estadística, de primer curso, presentaba una gran oportunidad para poner en práctica el enfoque docente concebido.

Después de un primer curso académico en el que prepararon y rodaron todos los materiales específicos bajo este nuevo esquema, en el curso actual se considera que el planteamiento está completamente asentado. En el curso 2022-2023 un total de 80 alumnos, repartidos en dos grupos, han utilizado Python como herramienta de trabajo habitual para resolver todos los problemas planteados en la asignatura, tanto en clase como en los exámenes.

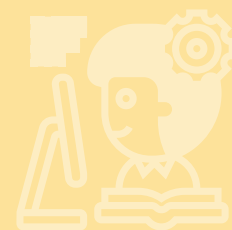
4. Descripción

La manera de evaluar condiciona, al final, tanto la forma en la que se imparte una asignatura como la manera en la que los alumnos la estudian. Por tanto, se utilizó como estrategia angular en el diseño de la asignatura una única restricción: exigir el uso de cuadernos de Jupyter con Python como único medio para hacer los exámenes. Al imponer esta restricción dura, todas las actividades a plantear surgieron de forma natural para asegurar la viabilidad del enfoque.

Una dificultad adicional es que los alumnos de Probabilidad y Estadística de 1º de IMAT aprenden Python durante el primer cuatrimestre, por lo que de facto hay que asumir que no adquieren los conocimientos suficientes de este lenguaje hasta el segundo cuatrimestre.

Para asegurar que los alumnos desarrollan realmente las competencias necesarias en estas herramientas informáticas para que su uso suponga un beneficio y no un problema durante los exámenes, ha sido necesario someter a los alumnos desde un primer momento a una inmersión completa pero gradual en esta forma de trabajo. En concreto, para conseguirlo ha sido necesario realizar las siguientes actividades:

- ➔ Preparar una “chuleta oficial” para cada tema con código ejemplo para poner en práctica los conceptos teóricos estadísticos y que se pueda utilizar en los exámenes.



- ➔ Crear un conjunto de problemas usando Python con una estructura claramente definida, similar a la que luego se emplea en los exámenes, para habituar a los alumnos a la forma de trabajo exigida.
- ➔ Preparar los exámenes para poder ser realizados en un entorno seguro que dificulte el compartir información entre alumnos o buscar otra ayuda diferente de la proporcionada.
- ➔ Acondicionar los problemas resueltos por los alumnos para facilitar su corrección digital y la realimentación posterior.

A continuación, se describe cada actividad con un poco más de detalle.

4.1. Preparación de chuletas oficiales como cuadernos de Jupyter

Este material es fundamental para ayudar a resolver todos los ejercicios y problemas, tanto aquellos realizados en clase como los problemas de examen. Por ello tiene que recoger toda la información sobre los detalles de programación en Python que no son objeto de evaluación en la asignatura pero que son necesarios para la resolución de problemas, evitando incluir información sobre los conceptos teóricos estadísticos que sí se quieren evaluar.

Se han preparado numerosos cuadernos Jupyter, organizados para que el alumno pueda encontrar cómodamente lo que considere necesario en cada momento (ver ejemplo de la Figura 1).

Figura 1. Ejemplo de una chuleta implementada como un cuaderno de Jupyter

1. Variable aleatoria continua

- 1.1 Crear una v.a. continua
- 1.2 Cálculos con pdf y cdf**
- 1.3 Medidas estadísticas fundamentales de la v.a.
- 1.4 Esperanza matemática de una función de una v.a.
- 1.5 Gráfica de la función de densidad (pdf)
- 1.6 Gráfica de la función de distribución (cdf)
- 1.7 Gráficas juntas de las funciones de densidad (pdf) y de distribución (cdf)

2. Distribución Uniforme $U(a, b)$

3. Distribución Triangular $T(a, c, b)$

1.2 Cálculos con pdf y cdf

```
[3]: # Densidad de probabilidad en un punto concreto
x=3.0
densidad = VA.pdf(x)
print(f'f({x}) =', densidad)

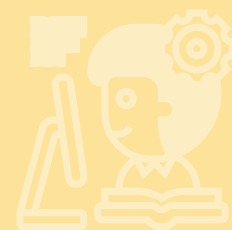
# Función de probabilidad acumulada
x=3
cdf_va = VA.cdf(x)
print(f'F({x}) = P(X<={x}) =', cdf_va)

# Soporte de la v.a.
soporte_va = VA.support()
print('Soporte(X) =', soporte_va)

f(3.0) = 0.00495750435332717
F(3) = P(X<=3) = 0.997521247823336
Soporte(X) = (0.0, inf)
```

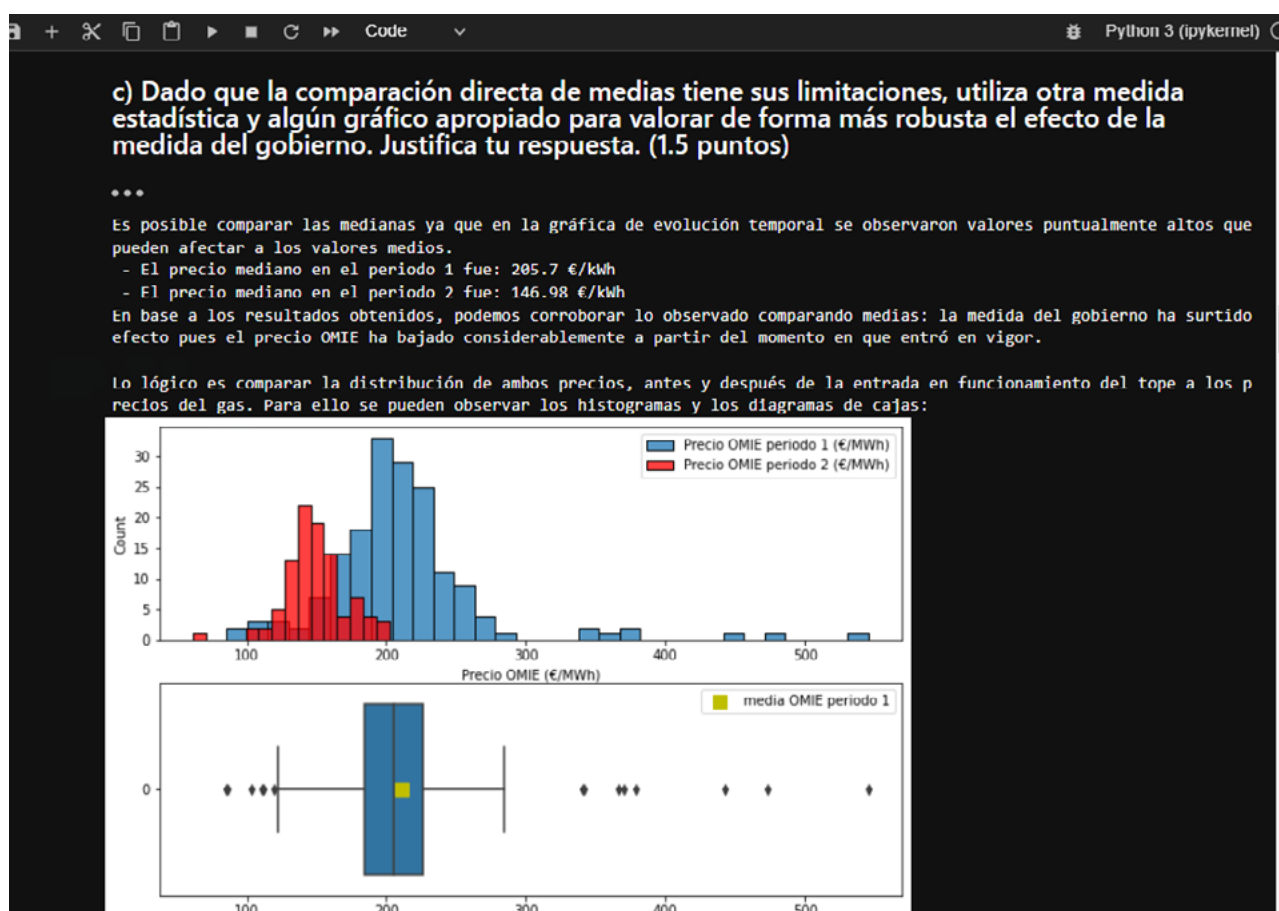
4.2. Uso en clase de ejercicios y problemas como cuadernos de Jupyter

En cada tema se dispone de un conjunto de ejercicios, pensados para que el alumno entienda para qué vale el código proporcionado en las chuletas. Los problemas suministrados, de mayor complejidad, requieren haber interiorizado los



conceptos estadísticos teóricos para poderlos resolver. Tanto los ejercicios como los problemas están desarrollados en cuadernos Jupyter, combinando enunciado y espacio para resolverlo de una forma amigable (Figura 2). Ambos se entregan sin solución inicialmente, algunos se realizan en clase y finalmente se entrega la solución de todos para que cada alumno pueda estudiarlos en detalle.

Figura 2. Ejemplo de un apartado de un problema implementado como un cuaderno de Jupyter



4.3. Realización de exámenes en un entorno seguro

Se dispone de dos entornos, el habitual de trabajo y el de examen (Figura 3). Ambos entornos son idénticos para el alumno, pudiéndose acceder únicamente al servidor de examen de manera puntual mediante *Safe Exam Browser*. Previamente al examen se elimina todo el contenido existente en la zona de trabajo del alumno, distribuyendo automáticamente tanto los enunciados como todas las chuletas. De esta forma se asegura que cada alumno trabajará únicamente con acceso al material oficial. Al finalizar el examen, se recoge automáticamente todo el contenido de la zona de trabajo del alumno, realizando capturas intermedias durante el examen por si ocurriese algún problema técnico.

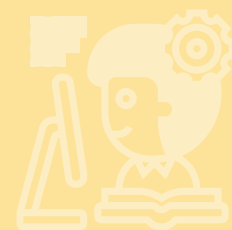


Figura 3. Solución tecnológica utilizada para implementar el entorno de trabajo y examen



4.4. Corrección digital de los exámenes

Para facilitar la corrección de los exámenes ha sido necesario generar automáticamente ficheros pdf a partir de los cuadernos Jupyter entregados por los alumnos con las respuestas. De esta forma se pueden corregir digitalmente en cualquier ordenador o tableta, permitiendo además una realimentación sencilla a los alumnos enviando correos electrónicos individualizados con las correcciones (ver ejemplo en la Figura 4).

Figura 4. Ejemplo de examen corregido digitalmente sobre el pdf generado a partir del cuaderno

d) Ayuda al grupo de ingenieros realizando un contraste de hipótesis paramétrico para el tiempo medio hasta el fallo. Plantea el contraste lógico a realizar, indicando las hipótesis H_0 y H_1 . Resuelve el contraste de forma detallada a partir de la muestra disponible, con un nivel de significación del 1%. Calcula el valor del estadístico de contraste en la muestra, los valores críticos y el p-valor asociado. Razona utilizando por separado los valores críticos y el p-valor para tener más seguridad en lo que haces y completa con una representación gráfica de toda la información. No olvides concluir claramente. (4 puntos)

```

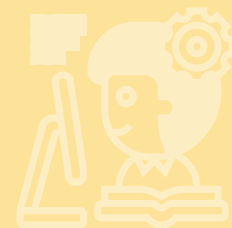
In [41]: # Respuesta
sigma2_pob = np.var(x)
# media poblacional que se quiere contrastar
mu_0 = 5
# calcule el estadístico de contraste con la muestra (var. conocida)
n = len(x)
t0 = (np.mean(x) - mu_0) / (sigma2_pob / np.sqrt(n))
# crea la distribución del estadístico de contraste
T0 = stats.norm(0,1)
# calcula el p-valor (más de 10000)
pvalor = 2 * (1 - T0.cdf(abs(t0)))
print(f'El valor observado del estadístico de contraste es (t0), con un p-valor')

# razonando con valores críticos
# nivel de significación
alfa = 0.01
  
```

Handwritten corrections in red ink: "3/", "NO!", "a la vez", and a signature.

Stack of PDF files and Outlook icon.

- **Realimentación sencilla a los alumnos**
 - Enviando correos individualizados con las correcciones



5. Resultados y valoración del trabajo realizado

Los resultados obtenidos han sido muy satisfactorios. Tras dos cursos académicos se puede confirmar que los alumnos están haciendo un buen uso de los recursos proporcionados para mejorar su aprendizaje sobre la Estadística y su aplicación en problemas reales. Todos los exámenes se están realizando utilizando Python como lenguaje vehicular para resolver los problemas. Los alumnos se han enfrentado a problemas realistas utilizando miles de datos y trabajando con modelos estadísticos complejos, únicamente manejables con un entorno software como el utilizado.

El desempeño mostrado por los alumnos, en general, ha sido muy bueno, confirmando que es posible aprender los conceptos abstractos más complejos de estadística con ayuda de un buen entorno de trabajo software, y desarrollar al mismo tiempo competencias prácticas adicionales de nivel profesional de forma cómoda.

6. Conclusiones y próximos pasos

Se ha conseguido implementar completamente el planteamiento inicial, siendo posible impartir la asignatura con el enfoque pretendido, utilizando Python como lenguaje vehicular. El aprendizaje observado en los alumnos, y su opinión sugiere que ha sido un acierto.

Está previsto seguir impartiendo la asignatura de Probabilidad y Estadística con el mismo planteamiento en los próximos cursos.

Por el momento solo se ha detectado una dificultad para la puesta en práctica del enfoque propuesto. El profesorado que imparte la asignatura tiene que conocer, no solo los conceptos teóricos estadísticos, sino además el lenguaje Python y el entorno de desarrollo basado en cuadernos de Jupyter. La preparación de los exámenes, la ejecución de los mismos y la corrección posterior también requiere de ciertas destrezas digitales adicionales.