



SLOVAKIA 2010 TRNAVA

IGIP-SEFI

*Joint International IGIP-SEFI Annual Conference 2010,
19th - 22nd September 2010, Trnava, Slovakia*

SELECTION OF BASIC TOPICS FOR THE KNOWLEDGE OF STATISTICS IN ENGINEERING

Esperanza Ayuga-Téllez,
Concepción González-García
M^a Ángeles Grande-Ortiz

Beatriz Reyes-Hernández & Claudia García-Ventura



POLITÉCNICA



Agencia de Calidad, Acreditación y Prospectiva
de las Universidades de Madrid



SLOVAKIA 2010 TRNAVA

IGIP-SEFI

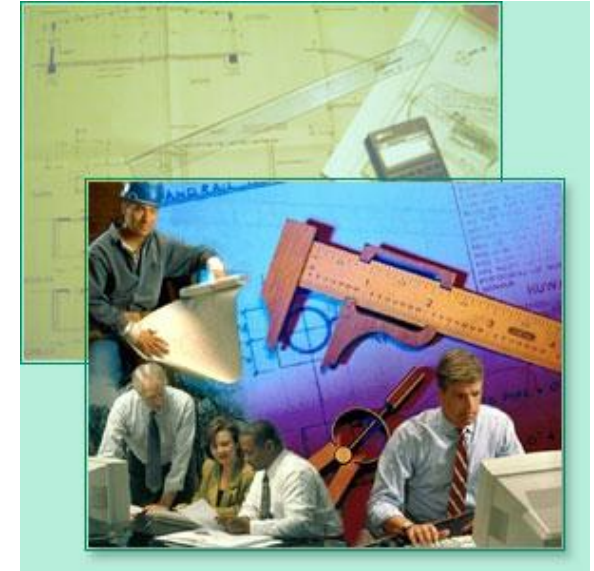


POLITÉCNICA



Agencia de Calidad, Acreditación y Prospectiva de las Universidades de Madrid

The subject of statistics is specially important in technical studies.



The knowledge of Statistics depends on the work they are expected to carry out.



SLOVAKIA 2010 TRNAVA

IGIP-SEFI

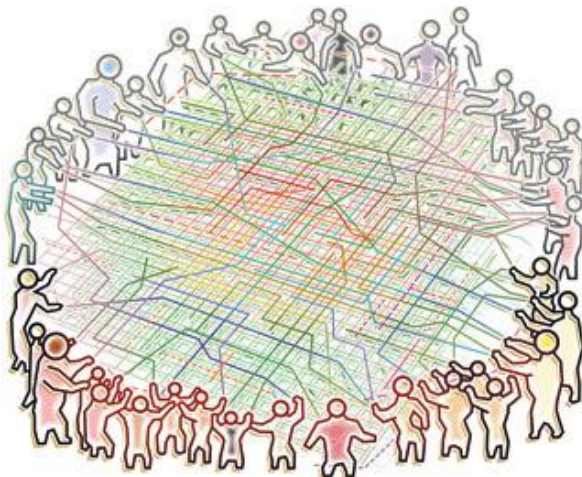


POLITÉCNICA



Agencia de Calidad, Acreditación y Prospectiva
de las Universidades de Madrid

EHEA makes it necessary to re-evaluate the knowledge, skills and abilities to acquire in the different subjects.

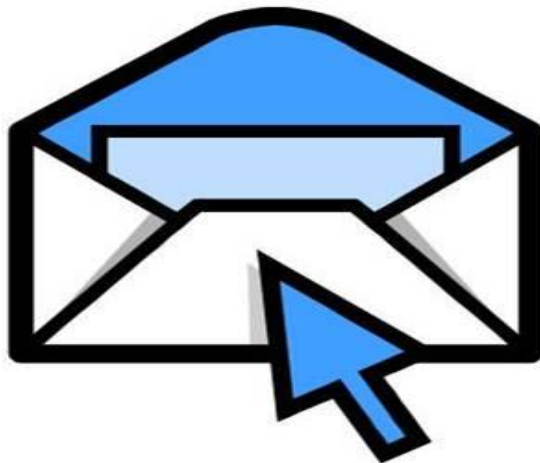


To establish the minimum common knowledge that must be learned in order to develop the activities.



Selecting the panel of experts

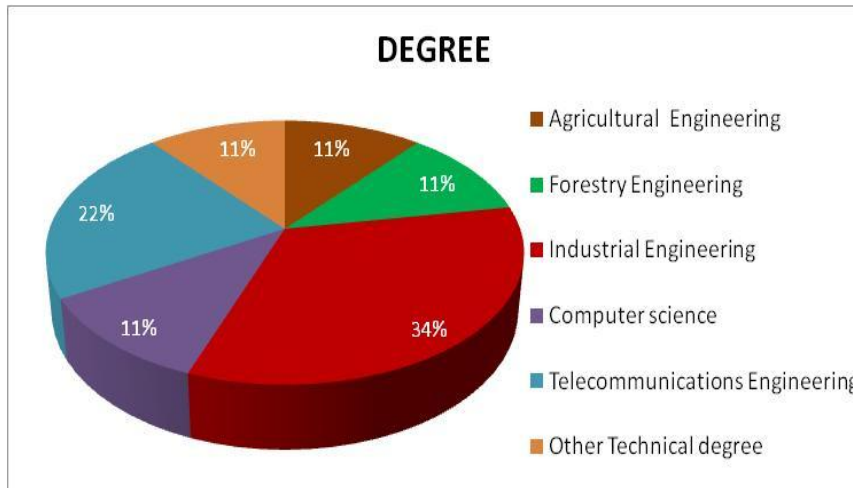
Teachers of Applied Statistics in various universities and engineering and technical courses, not including the Madrid Polytechnic University.



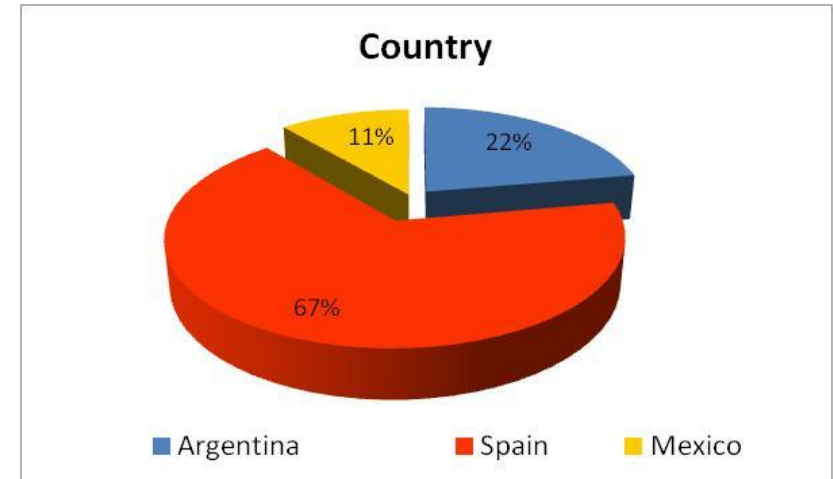
The questionnaires were sent by e-mail with a letter of introduction, in order to encourage the subjects participation.



Composition of the panel of experts



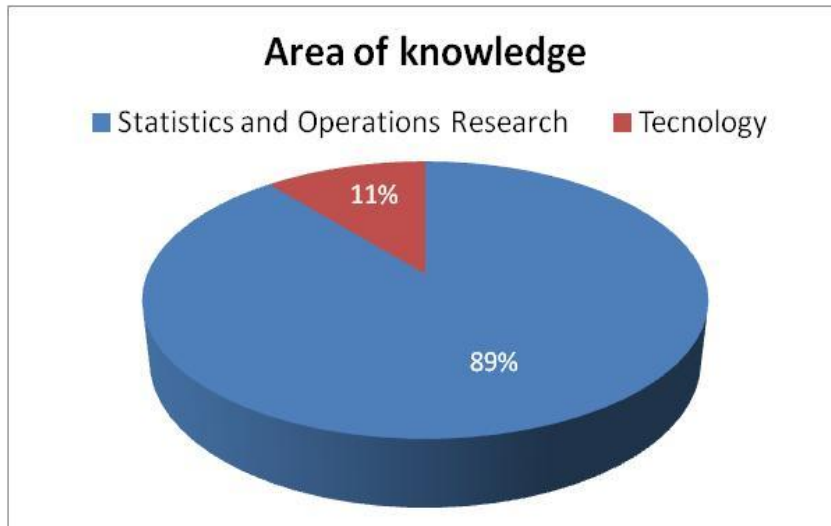
According to the degree in which they teach Statistics.



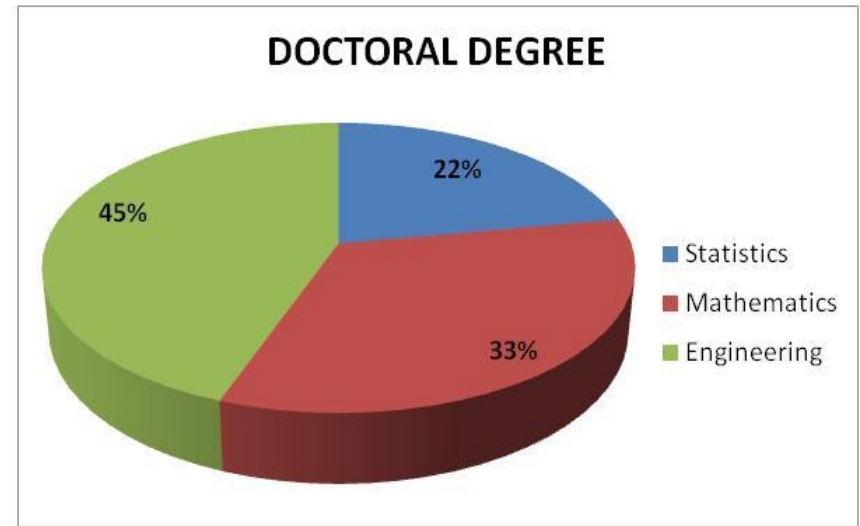
According to country



Composition of the panel of experts



According to the their area of knowledge.



According to the their doctoral degree



Design of the questionnaire



CUESTIONARIO PREVIO PARA EVALUAR LOS CONOCIMIENTOS EN LA MATERIA DE ESTADÍSTICA APLICADA DE ESTUDIANTES DE LA UPM

1.- Ordenar los ocho bloques temáticos que se relacionan a continuación de mayor a menor importancia para el desarrollo profesional del Ingeniero (1 el más importante, 8 el menos importante):

								Nº por orden de importancia
Estadística descriptiva	→	→	→	→	→	→	→
Probabilidad	→	→	→	→	→	→	→
Muestreo y estimación	→	→	→	→	→	→	→
Test de hipótesis	→	→	→	→	→	→	→
El modelo lineal general	→	→	→	→	→	→	→
Diseño de experimentos	→	→	→	→	→	→	→
Procesos	→	→	→	→	→	→	→
Análisis Multivariante	→	→	→	→	→	→	→

Dentro de cada bloque temático se plantean 10 cuestiones tipo test, señalar las cuatro que considere más adecuadas para evaluar los conocimientos de Estadística Aplicada que debería poseer un Ingeniero.

Estadística descriptiva

1.- En un estudio se mide la "profundidad de suelo", ¿qué tipo de variable estadística es?

- a) Discreta
- b) Continua
- c) Cualitativa



4.- Datos 1.- El primer cuartil es:

- a) 9,5
- b) 3,6
- c) 2,5

5.- Datos 1.- El percentil del 10% es:

- a) 6,8
- b) 3,6
- c) 2,5

Datos 2.- En una fábrica de motosierras se controla la longitud de la cuchilla y se han obtenido las siguientes mediciones, en cm.

Valores X	58,7	60,1	61,5	62,9	64,3
Frecuencia absoluta	8	14	8	8	2

6.- Datos 2.- El valor de la moda es:

- a) 60,1
- b) 61,5
- c) 62,9

7.- Datos 2.- El valor de la media es:

- a) 58,95
- b) 60,87
- c) 62,10

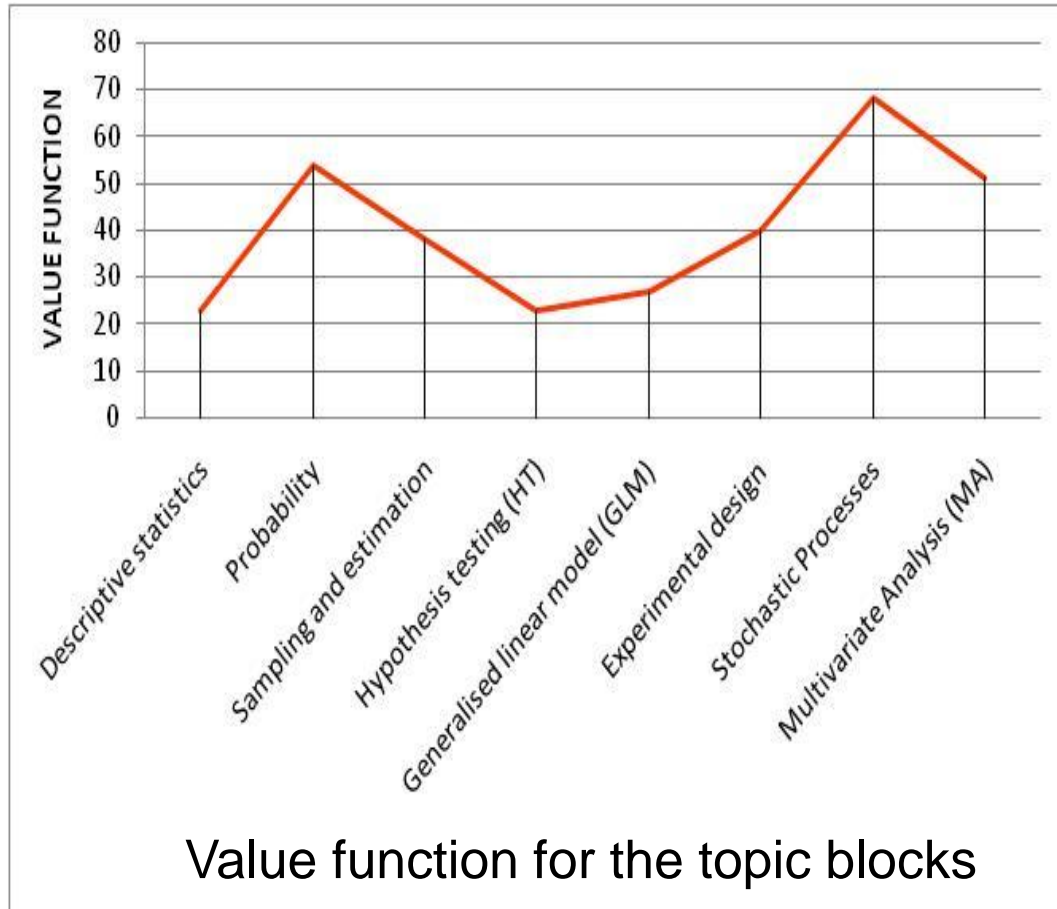
8.- Datos 2.- El rango de los valores es:

- a) 4,78
- b) 5,23
- c) 5,6

9.- En un estudio se mide la "profundidad de suelo" con un instrumento de frecuencia de



Results



• **Descriptive statistics**

• **Sampling and estimation**

• **Hypothesis testing**

• **Generalised linear model**



Results

Blocks / ITEMS	1	2	3	4	5	6	7	8	9	10
Descriptive	3	3	7	0	1	0	6	2	7	7
Probability	2	2	1	6	6	3	5	4	6	1
Estimation	5	0	1	6	8	4	3	0	4	5
HT	5	4	3	5	5	4	5	1	2	2
GLM	7	2	5	5	6	4	1	3	0	3
Design E.	6	6	2	3	3	3	2	3	1	7
S. Processes	6	3	5	2	3	3	1	2	6	5
MA	4	6	1	7	0	7	0	9	1	1

Number of experts who selected the items



QUESTIONARIO PARA EVALUAR LOS CONOCIMIENTOS EN LA MATERIA DE ESTADÍSTICA APLICADA DE ESTUDIANTES DE LA UPM

Sexo HOMBRE MUJER
 Año de nacimiento.....
 Nota obtenida en la asignatura de Estadística APROBADO NOTABLE SOBRESALIENTE MH
 Año en que curso la asignatura de Estadística HACE 3 AÑOS HACE 4 AÑOS HACE 5 AÑOS O MÁS
 Siguió el método de innovación educativa SI NO

1.- En un control de calidad se observa el "número de tornillos defectuosos" en cada lote fabricado, ¿qué tipo de variable estadística es?

- Discreta
- Continua
- Cualitativa

2.- En un proceso industrial se mide $X =$ tiempo (en minutos) de espera de llegada de piezas a la envasadora, con los siguientes resultados:

Valores X	99,5	99,1	53,4	23,8	9,5	3,6	2,5	0,87	0,78
Frec relativa acumulada	1	0,995	0,9	0,75	0,5	0,25	0,1	0,05	0,0

El valor de la mediana es:

- 23,8
- 16,4
- 9,5

3.- En una fábrica de motosierras se controla la longitud de la cuchilla y se han obtenido las siguientes mediciones, en cm:

Valores X	58,7	60,1	61,5	62,9	64,3
Frecuencia absoluta	8	14	8	8	2

El valor de la media:

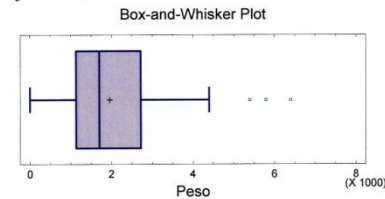
- 58,95
- 60,87
- 62,10

4.- En un estudio se mide la "profundidad de suelo", para representar las frecuencias de los valores de la variable ¿qué tipo de gráfico se emplearía?

- Diagrama de barras
- Gráfico de sectores
- Histograma



5.- En un estudio sobre el "peso de la carga de los camiones" que entran en fábrica se ha obtenido el siguiente gráfico. ¿Qué representa la línea vertical del interior de la caja?



- La media
- La mediana
- La moda

6. En el cálculo del tamaño n muestral, es necesario conocer,

- El error máximo de muestreo e que se está dispuesto a aceptar y el nivel de confianza α para la estimación.
- El error máximo de muestreo e que se está dispuesto a aceptar y la desviación típica de la variable a estudiar o algún valor estimado de la misma.
- La desviación típica de la variable a estudiar o algún valor estimado de la misma, el error máximo de muestreo e que se está dispuesto a aceptar y el nivel de confianza α para la estimación.

7. Sabemos que la tensión arterial de la población aumenta a partir de los 60 años. Para estimar la "presión arterial media" de los individuos de una gran ciudad, se debe elegir un diseño de muestreo, ¿cuál?:

- Aleatorio simple
- Estratificado, con diferentes estratos según la edad.
- Por conglomerados, con conglomerados según el barrio.

8. En un muestreo aleatorio simple se desea obtener el tamaño de muestra adecuado para estimar la "proporción de vidrio por kilo de basura reciclable". La expresión que se puede utilizar es:

$$n_0 = \left(\frac{1,96}{0,005} \right)^2 0,5 \cdot 0,5 = 38416 \text{ kilos}$$

- Con un 95% de confianza, error de estimación 0,5% y máxima varianza.
- Con un 95% de confianza, error de estimación 0,005 y mínima varianza
- Con un 90% de confianza error de estimación 0,5 y varianza 0,005



9. Para realizar un muestreo sistemático en un terreno:
 - a. Se eligen n puntos de muestreo al azar sobre el total de la superficie.
 - b. Se superpone una malla con $k \times n$ nodos y se escogen al azar n de ellos, de tal forma que, cada uno de los $k \times n$ nodos tenga la misma probabilidad de ser escogido.
 - c. Se superpone una malla con n nodos y se escoge al azar un punto de la superficie, superponiendo en ese punto un nodo de la malla. Los n nodos así colocados serán los puntos de la muestra.
10. Se tienen muestras de diámetros medios de dos especies de pinos medidos a diferentes edades: 5, 10, ..., 150 años. ¿Qué conviene utilizar para calcular un Intervalo de Confianza para la diferencia de diámetros entre las dos especies?
 - a. Un intervalo para diferencia de medias con muestras pareadas
 - b. Un intervalo para diferencia de medias con muestras independientes
 - c. Un intervalo para cada media diamétrica
11. Para aplicar un contraste de hipótesis, el estadístico del contraste debe medir:
 - a. La diferencia entre la H_0 y la H_1
 - b. La diferencia entre la muestra y la H_0
 - c. La diferencia entre la muestra y la H_1
12. El error de tipo I se comete cuando:
 - a. La H_0 es verdadera
 - b. La hipótesis alternativa es verdadera
 - c. La H_0 es menos probable
13. El nivel de significación del contraste es igual:
 - a. Al p-valor (nivel crítico)
 - b. A la probabilidad de cometer un error de tipo I
 - c. A la probabilidad de cometer un error de tipo II
14. Si dos variables cualesquiera siguen una distribución normal, el coeficiente adecuado para detectar su independencia es:
 - a. El de la τ de Kendall
 - b. El de Spearman
 - c. El de correlación de Pearson
15. Un contraste de bondad de ajuste con la χ^2 debe cumplir que:
 - a. Las frecuencias observadas deben ser mayores que 5
 - b. Las frecuencias esperadas deben ser menores que 5
 - c. Las frecuencias esperadas deben ser mayores que 5
16. ¿Cuál es la hipótesis nula empleada en el análisis de varianza?
 - a. La igualdad de varianzas entre grupos
 - b. La igualdad entre todas las medias de los grupos
 - c. La veracidad del modelo



17. Un modelo lineal simple sirve para:

- a. Minimizar la dispersión relativa de la variable dependiente.
- b. Modelizar relaciones y predecir valores de alguna de las variables relacionadas.
- c. Encontrar la relación causa-efecto entre dos variables

18. Tenemos dos modelos linealizables que pueden utilizarse para predecir el valor de Y. El modelo 1 tiene una $R^2=0,87$ y muestra heterocedasticidad residual. El modelo 2 tiene una $R^2=0,67$ pero los residuos cumplen todos los requisitos. Decido:

- a. Utilizar la recta de regresión, ya que ambos modelos son malos.
- b. Elegir el modelo 1 que tiene mayor R^2 .
- c. Elegir el modelo 2 que cumple todas las hipótesis

19. En un modelo de regresión se obtiene el siguiente análisis de varianza:

Analysis of Variance with Lack-of-Fit					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	678,745	1	678,745	2,08	0,1505
Residual	118709,0	363	327,023		
Lack-of-Fit	6952,76	29	239,751	0,72	0,8605
Pure Error	111756,0	334	334,6		
Total (Corr.)	119388,0	364			

Estos resultados se pueden interpretar de la siguiente forma:

- a. El modelo es malo ya que los p-valores del ANOVA son mayores de 0,05.
- b. El modelo explica poca variabilidad pero la relación entre x e y es lineal.
- c. El modelo explica mucha de la variabilidad pero la relación entre x e y no es lineal.

20. Los residuos del modelo deben ser: (valores de una variable aleatoria, V.a.)

- a. V.a. con distribución $N(\mu, 1)$ con $\mu=cte$ e incorrelados entre sí
- b. V.a. con distribución $N(0, \sigma)$ con $\sigma=cte$ e incorrelados entre sí
- c. V.a. con distribución $N(\mu, \sigma)$ con $\sigma=cte$ y media diferente

¿Ha realizado algún curso externo o Seminario relacionado con la Estadística?

¿Cuál es su opinión general sobre la Estadística? ¿Qué importancia cree que tiene la Estadística en la carrera que estudia? ¿Y posteriormente en su vida profesional?

Cree interesante añadir algo al cuestionario.



SLOVAKIA 2010 TRNAVA

IGIP-SEFI



POLITÉCNICA



Agencia de Calidad, Acreditación y Prospectiva
de las Universidades de Madrid

Conclusions

A questionnaire containing 20 questions divided into 4 topic blocks.

The 4 topic blocks were as follows: Descriptive Statistics, Sampling and Estimation, Hypothesis Testing, and Generalised Linear Model.

Each block contains the 5 questions selected from the initial questionnaire by the greatest number of experts on the panel.



SLOVAKIA 2010 TRNAVA

IGIP-SEFI



POLITÉCNICA



Agencia de Calidad, Acreditación y Prospectiva
de las Universidades de Madrid

esperanza.ayuga@upm.es
concepcion.gonzalez@upm.es
m.angeles.grande@upm.es



WEB GIE

<http://www2.montes.upm.es/gie/tcim/index.htm>