

INSTRUCCIONES:

Material permitido: Fotocopia de Tablas estadísticas y calculadora no programable.

TEÓRICO-PRÁCTICA

Puntuación Total 4 Puntos:

Cada pregunta contestada correctamente puntúa 0,4 sobre diez, la contestada incorrectamente se penaliza con 0,2 y la no contestada no puntúa.

1. Indique la opción correcta:

- La función de distribución $F(x)$ se define como probabilidad acumulada en el caso de variables aleatorias discretas o continuas.
- La representación gráfica de la función de distribución $F(x)$ de una variable aleatoria, ya sea discreta o continua, siempre es decreciente.
- La función de distribución $F(x)$ para una variable aleatoria discreta tiene un valor constante igual a 1.
- La opción a) y b) son correctas.

2. La varianza...:

- Su valor tipificado está comprendido entre -1 y 1.
- Es una medida orientada a conocer la dependencia lineal entre variables.
- Sirve para calcular el valor de la media.
- No es correcta ninguna de las opciones.

3. La tipificación de una variable aleatoria.....

- La utilizamos cuando conocemos la media y la varianza de una distribución con el objetivo de analizar la covarianza.
- Es una transformación de una variable aleatoria mediante un cambio de origen y escala.
- Permite conocer el grado de eficiencia relativa.
- La opción b) y c) son correctas.

4. Indique la opción correcta:

- Para analizar la dispersión relativa de dos distribuciones podemos utilizar tanto la covarianza como la tipificación.
- Para cuantificar el apuntamiento entre dos variables aleatorias utilizamos el coeficiente de correlación lineal.
- No existe ninguna diferencia entre utilizar el coeficiente de correlación lineal y el coeficiente de variación para analizar la causalidad entre dos variables aleatorias.
- No es correcta ninguna de las opciones.

5. De las siguientes opciones indique la que está relacionada con una distribución binomial:

- Es una distribución continua que tiene forma de campana.
- Es una distribución caracterizada por su media y varianza, valores que siempre coinciden.
- La variable aleatoria X toma valores enteros comprendidos entre 0 y n , y nos indica el número repeticiones independientes de una prueba de Bernoulli.
- No es correcta ninguna de las opciones.

6. Dadas dos distribuciones normales $N(\mu_x, \sigma_x)$ y $N(\mu_y, \sigma_y)$ centradas en la misma posición $\mu_x = \mu_y$ pero con distinta $\sigma_x > \sigma_y$. Indique la opción correcta:

- $N(\mu_x, \sigma_x)$ no es simétrica respecto a su media.
- $N(\mu_x, \sigma_x)$ es más apuntada que $N(\mu_y, \sigma_y)$.
- $N(\mu_x, \sigma_x)$ tiene menor dispersión relativa que $N(\mu_y, \sigma_y)$.
- No es correcta ninguna de las opciones.

7. Cuál de las siguientes opciones no está relacionada con el concepto de muestra aleatoria simple:

- a) Está formada por variables aleatorias que son independientes.
- b) Debe ser representativa de la población de la que se ha extraído la muestra.
- c) Es un subconjunto del que tratamos de obtener información que podamos extrapolar a toda la población.
- d) Todas las opciones son correctas.

8. Indique la opción correcta para la varianza de la distribución del estadístico media muestral \bar{X} :

- a) Siempre coincide con la varianza poblacional independientemente del tamaño de la muestra.
- b) Siempre es superior a la varianza poblacional para tamaño de la muestra $n > 1$
- c) Es inferior a la varianza poblacional para tamaño de la muestra $n > 1$
- d) No es correcta ninguna de las opciones.

9. Indique la opción correcta para la potencia o poder de un test:

- a) Es la capacidad que tiene el contraste para reconocer correctamente que la hipótesis nula es falsa, y por tanto, rechazarla.
- b) Es $1 - \beta$
- c) Para asegurar la calidad de un proceso de contraste es conveniente encontrar test estadísticos que proporcionen regiones críticas con potencias grandes.
- d) Todas las opciones son correctas.

10. El contraste de rangos-signos de Wilcoxon.....:

- a) Es un contraste de localización.
- b) Es un contraste de bondad del ajuste.
- c) Es un contraste de aleatoriedad.
- d) No es correcta ninguna de las opciones.

PROBLEMA 1**Puntuación Total 3 Puntos:**

Cada pregunta contestada correctamente puntúa 1 sobre diez, la contestada incorrectamente se penaliza con 0,25 y la no contestada no puntúa.

PROBLEMA 1º:

Como encargados del departamento de RRHH de nuestra compañía, y sabiendo que el porcentaje anual de trabajadores que se cogen bajas laborales a lo largo del año es del 10%, queremos saber:

11. Si tomamos una muestra de 20 trabajadores, ¿qué probabilidad habría de que ningún trabajador tenga una baja laboral a lo largo del año?
- a) 0,5112
 - b) 0
 - c) 0,1216
 - d) No es correcta ninguna de las opciones
12. Y si tomásemos una muestra de 10 trabajadores, ¿qué probabilidad habría de que se cojan la baja más de 2?
- a) 0,0702
 - b) 0,9298
 - c) 0,4312
 - d) No es correcta ninguna de las opciones
13. Si tomamos una muestra de 200 trabajadores, ¿Cuál sería la probabilidad aproximada de que 15 o menos trabajadores se cojan una baja?
- a) 0,1190.
 - b) 0,8810.
 - c) 1,179.
 - d) No es correcta ninguna de las opciones.

OJO: PREGUNTA DE RESPUESTA OBLIGATORIA

Pregunta 14.

SELECCIÓN DE ITINERARIO. PREGUNTA DE RESPUESTA OBLIGATORIA. Si el alumno no marca ninguna de las opciones no se corregirá ni el problema 2 ni el problema 3.

- a) **Marcar la opción a) Alumnos que NO han realizado PEC o que han realizado PEC, y NO quieren que les puntúe.** Los alumnos realizarán el problema 2 y contestarán a sus respectivas preguntas: 15-16-17. Estos alumnos **NO** pueden contestar al problema 3.
- b) **Marcar la opción b) Alumnos que SÍ han realizado PEC y SÍ quieren que se les puntúe.** Los alumnos realizarán el problema 3 y contestarán a sus respectivas preguntas: 18-19-20. Estos alumnos **NO** pueden contestar al problema 2.

PROBLEMA 2: opción A pregunta 14 – Alumnos NO PEC.

Puntuación Total 3 Puntos:

Cada pregunta contestada correctamente puntúa 1 sobre diez, la contestada incorrectamente se penaliza con 0,25 y la no contestada no puntúa.

PROBLEMA 2º: Sólo alumnos que hayan marcado la opción a) en la pregunta 14.

Hemos recibido una oferta laboral como científico de datos, la oferta supone un salario bruto anual de 53.200€. Como creemos que este salario está por debajo de lo ofrecido en el sector, hacemos un pequeño estudio de mercado a través de una red social profesional y obtenemos una muestra aleatoria simple de 78 científicos de datos con perfiles similares al nuestro. El resultado obtenido del muestreo es que el salario bruto anual medio es de 58.600€ con una desviación típica de 31.630€. Observamos normalidad en la distribución de la muestra por lo que asumimos normalidad en el análisis. Con estos datos vamos a calcular:

15. Un intervalo de confianza aproximado para la media poblacional con un nivel de confianza del 95%

- a) $I_{\mu} = [46.180,47; 60.219,53]$
- b) $I_{\mu} = [51.580,47; 65.619,53]$
- c) $I_{\mu} = [52.405,19; 53.994,81]$
- d) No es correcta ninguna de las opciones.

16. Contrastar con un nivel de significación del 0,05 si el sueldo que nos ofrecen es similar al de mercado o es significativamente diferente según criterios estadísticos.

- a) El resultado es que $Z_{exp}=1,51$ que es menor que 1,96 por lo que se rechaza la Hipótesis nula.
- b) El resultado es que $Z_{exp}=13,32$ que es mayor que 1,96 por lo que se acepta la Hipótesis nula.
- c) El resultado es que $Z_{exp}=-1,51$ que es mayor que -1,96 por lo que se rechaza la Hipótesis nula.
- d) No es correcta ninguna de las opciones.

17. Obtener en términos de p-valor los resultados del apartado 15:

- a) el p-valor sería 0,066 por lo que el contraste es significativo y se rechaza la hipótesis nula
- b) el p-valor sería de 0,132 por lo que el contraste es significativo y se rechaza la Hipótesis nula
- c) el p-valor sería de 0,934 por lo que el contraste es significativo y no se rechaza la hipótesis nula
- d) No es correcta ninguna de las opciones.

PROBLEMA 3: opción B pregunta 14 – Alumnos SI PEC.

Puntuación Total 3 Puntos:

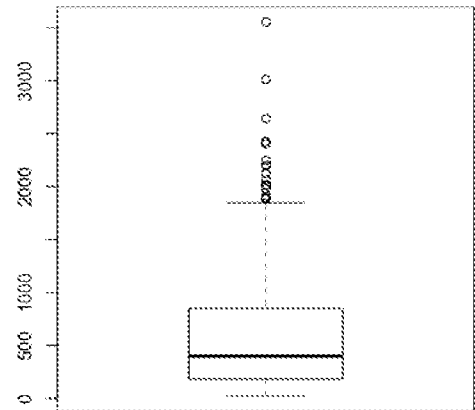
Cada pregunta contestada correctamente puntúa 1 sobre diez, la contestada incorrectamente se penaliza con 0,25 y la no contestada no puntúa.

PROBLEMA 3º: Sólo alumnos que hayan marcado la opción b) en la pregunta 14.

Se ha trabajado con una base de datos de una muestra de hospitales, la variable que vamos a analizar es la referente a los pacientes dados de alta, la variable se denomina outpatients.

18. Sobre el siguiente gráfico indicar cual es la respuesta correcta:

- a) Es un boxplot y lo utilizamos para estudiar la normalidad
- b) Es un boxplot y lo utilizamos para ver los valores atípicos
- c) Es un histograma y lo utilizamos para ver los valores atípicos
- d) Es un histograma y lo utilizamos para ver la normalidad



19. Indicar cual sería el comando de R que utilizaríamos para obtener la media de la variable outpatients:

- a) > outpatients(media)
- b) > mean(outpatients)
- c) > media(outpatients)
- d) > outpatients(mean)

20. Queremos contrastar la hipótesis nula de si la media de pacientes dados de alta en los hospitales puede ser igual a 500 al 5% de significación, así que generamos el contraste de hipótesis y obtenemos:

One Sample t-test

One Sample t-test

```
data: outpatients
t = 4.8699, df = 957, p-value = 1.306e-06
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
 546.6286 609.5739
sample estimates:
mean of x
 578.1013
```

Interpretando los resultados podemos concluir que:

- a) Se rechaza que los pacientes medios dados de alta puedan ser del entorno de 500.
- b) No podemos rechazar la H_0 ya que la media 578,10 está dentro del intervalo
- c) El p-valor no nos deja claro si se rechaza o se acepta la hipótesis nula
- d) No es correcta ninguna de las opciones.