ESTADÍSTICA (Química)
PRÁCTICA 7 – Tests de Hipótesis

Comentario: En todos los ejercicios propuestos

a) defina las variables aleatorias y los parámetros involucrados.
b) de ser posible indique:
   i. la distribución de las variables aleatorias
   ii. el significado intuitivo de los parámetros.
c) plantee las hipótesis nula y alternativa, e indique el nivel que usará para el test.
d) elija un test, calcule el valor del estadístico, calcule o acote el valor *p* e indique la conclusión del test.

1. Una colonia de ratones de laboratorio tiene varios cientos de animales. El peso en gramos de los ratones adultos sigue una distribución normal con media igual a 30g y desvío estándar de 5g. Como parte de un experimento, se les pidió a algunos estudiantes que eligieran 25 ratones adultos, sin ninguna premisa. El peso promedio de estos 25 animales fue de 33g. ¿Muestran estos datos evidencia suficiente a un nivel del 5% para decir que seleccionar los animales de esta manera no es lo mismo que elegirlos al azar? Justifique.

2. Supongamos que la proporción de monóxido de carbono (CO) de un gas es de 70 ppm. Se realizan mediciones con un fotómetro cuyos errores de medición siguen una distribución N(0, σ^2), es decir que si el fotómetro está bien calibrado podemos suponer que las mediciones siguen una distribución N(70, σ^2). Cuando el fotómetro no está calibrado y se produce un error sistemático las mediciones siguen una distribución N(μ, σ^2) con $\mu \neq 70$. Para cada uno de los siguientes conjuntos de mediciones independientes plantee un test adecuado para ver si encuentra evidencia a nivel 5% de error sistemático.

   a) 71, 68, 79
   b) 71, 68, 79, 84, 78, 85, 69
   c) 71
   d) 71, 84

   En uno de los casos es imposible plantear el test, ¿cuál y por qué?

3. Se quiere calibrar otro fotómetro. En este caso no está claro si el modelo que supone errores con distribución normal se verifica.
   Los siguientes son conjuntos de mediciones independientes sobre el gas del ejercicio anterior:

   a) 71, 70, 72, 69, 71, 68, 93, 75, 68, 61, 94, 91
   b) 71, 73, 69, 74, 65, 67, 71, 69, 70, 75, 71, 68
   c) 71, 69, 71, 69, 71, 69, 71, 69, 71, 69, 71, 69

   En dos de los casos anteriores el modelo de errores normales no se satisface: ¿cuáles y por qué? Para el conjunto de datos para el que parece razonable suponer un modelo de normal, haga el test que considere más apropiado para estudiar si hay evidencia de error sistemático a nivel del 5%.

4. Se arroja un dado 100 veces. La suma de los resultados dio 368 en vez de los 350 esperados. ¿Encuentra evidencia a nivel 1% de que el dado está cargado o esta diferencia se puede explicar como una variación aleatoria?

5. Consideremos un procedimiento para medir el contenido de manganeso en un mineral. A este procedimiento se lo ha usado muchas veces y se sabe que las mediciones siguen una distribución normal cuya desviación estándar es 0.15. Se está estudiando si el método tiene error sistemático.
   a) Se hacen 8 determinaciones de un mineral preparado que tiene 7% de manganeso y se obtienen los siguientes resultados:

      6.90 7.10 7.20 7.07 7.15 7.04 7.18 6.95 ($\bar{x}$ =7.074)

   ¿Cuál es su conclusión si desea que la probabilidad de equivocarse al decir que el método tiene error sistemático cuando en realidad no lo tiene sea 0.05? (En este caso si no hay error sistemático las determinaciones siguen una distribución N(7, 0.15^2) y si hay error la distribución es N(μ, 0.15^2) con $\mu \neq 7$)
   b) Si el método tiene un error sistemático de 0.10 (o sea, si μ = 7.10), ¿cuál es la probabilidad de cometer error de tipo II?
   c) Se quiere aplicar un test estadístico de modo que, al igual que en el inciso a), la probabilidad de decir que hay error sistemático cuando no lo hay sea 0.05. Pero además se desea que si hay un error sistemático de 0.10, la probabilidad de detectarlo sea $\geq$ 0.80 (o lo que es equivalente, la probabilidad de error tipo II sea $\leq$ 0.20).
      i. El test del inciso a) ¿cumple con este requisito?
      ii. En caso contrario, ¿cuántas determinaciones habría que hacer como mínimo?

6. Sea $X_1,...,X_n$ una muestra aleatoria *Bi* (1, p).
   a) Proponga un test asintótico de nivel α para H_0: $p=p_0$ contra H_1: $p \neq p_0$.
   b) Se tiene la hipótesis de que en una población de insectos la proporción de machos y de hembras es la misma, es decir que la proporción de hembras y de machos es 0.5. Testear esta hipótesis a nivel 0.05 sabiendo que se tomaron 100 insectos obteniéndose 43 machos.
   c) ¿Qué test utilizaría si sospechara que el porcentaje de hembras es mayor que el de machos?

7. Este es un ejemplo en el que se desea estudiar si un cambio en las condiciones de un experimento afecta el resultado. Se está estudiando un procedimiento para la determinación de estaño en productos alimenticios. Para ello se tomaron 12 muestras del mismo producto. Se eligieron 6 de estas muestras al azar y se llevaron al punto de ebullición con HCl a reflujo durante 30 minutos.Con las 6 muestras restantes el tiempo de reflujo fue de 70 minutos. Los resultados fueron los siguientes:

Tiempo de reflujo (^{m})	Estaño encontrado (mg/Kg)	$\overline{X}$	s²
30	57, 57, 59, 56, 56, 59	57.00	2.80
70	57, 55, 58, 59, 59, 59	57.83	2.57

   Suponga que la cantidad de estaño encontrada cuando el tiempo de reflujo es de 30 m es una variable aleatoria con distribución $N(\mu_1, \sigma_1^2)$ y dicha distribución es $N(\mu_2, \sigma_2^2)$ cuando el tiempo de reflujo es de 70 m.

¿Encuentra evidencia a nivel 5% de que la media de estaño encontrada para los dos tiempos de ebullición es diferente?

8. Se comparó un nuevo método propuesto para la determinación de la demanda de oxígeno en aguas residuales contra el método standard. Se hicieron 10 determinaciones para cada método en una misma muestra de aguas residuales, obteniéndose los siguientes resultados (en mg/l):

Método standard	74.4	67.2	66.1	71.2	68.7	69.9	71.0	77.8	72.4	70.1
Método propuesto	71.6	71.4	71.3	74.5	71.9	72.6	69.1	73.4	69.5	70.2

Suponga que la diferencia de lo obtenido con los dos métodos en cada muestra de agua sigue una distribución normal. Ingresando los datos al Statistix se calcularon las medias y las DE de estos datos:

```
DESCRIPTIVE STATISTICS

VARIABLE          N       MEAN        SD
METODOS          10     70.880    3.4224
METODOP          10     71.550    1.6821
```

¿Tenemos información suficiente para decir que la precisión del método propuesto es significativamente mejor que la del método standard a un nivel del 5%?

9. Utilizando dos métodos de análisis se hicieron determinaciones del contenido de hierro de una muestra de un mineral. Se asume que los determinaciones correspondientes a ambos métodos tienen distribución normal. Los resultados obtenidos son los siguientes:

Método 1	n_1=12	$\bar{x}_1$=15.22%	s_1=0.10%
Método 2	n_2=11	$\bar{x}_2$=15.30%	s_2=0.12%

a) ¿Son significativamente diferentes las desviaciones estándares observadas con ambos métodos a un nivel del 5%?
b) ¿Son significativamente diferentes a nivel 5% las medias de ambos métodos? (para elegir el test, tenga en cuenta el resultado del inciso anterior).
c) Repita a) y b) pero con s_2=0.20% (en lugar de s_2=0.12%).

10. En una estación del INTA se divide un terreno en 30 parcelas homogéneas. En 15 de ellas elegidas al azar se utiliza el fertilizante A y en las restantes el B. En las 30 parcelas se cultiva la misma variedad de maíz. Se supone que los rendimientos con el tratamiento A son variables aleatorias $N(\mu_A,\sigma_A^2)$ y con el tratamiento B son $N(\mu_B,\sigma_B^2)$.
Los resultados obtenidos son:

PARCELAS CON EL TRATAMIENTO A														
238	237	235	220	233	203	228	220	221	215	218	217	232	225	209

PARCELAS CON EL TRATAMIENTO B														
253	227	241	245	237	248	250	218	239	243	257	208	215	240	229

a) Proponga un test para H_0: $\mu_A=\mu_B$ contra H_1:$\mu_A\neq\mu_B$.

b) Proponga un test para H_0: $\mu_A=\mu_B$ contra H_1: $\mu_A<\mu_B$. ¿Hubiera podido anticipar su conclusión (rechazar H_0 o no hacerlo) a partir de las cuentas realizadas en a)?
c) Verifique que se satisfagan los supuestos del test.
d) Construya un intervalo de confianza de nivel 0.95 para la diferencia de rendimiento promedio entre los fertilizantes.
e) Responda a las preguntas a) y b) si en realidad los datos corresponden a 15 parcelas cada una dividida en 2 de forma que en una mitad se usó el fertilizante A y en la otra el B y además se supone que la diferencia tiene distribución normal.

11. Se comparó un método espectroscópico de absorción atómica de llama para determinar antimonio en la atmósfera con el método colorimétrico recomendado.
Para seis muestras de atmósfera urbana se obtuvieron los siguientes resultados:

	Antimonio encontrado (mg/m^3)	
Muestra	Método standard	Método nuevo
1	25.0	23.8
2	19.5	19.0
3	16.6	15.9
4	21.3	20.4
5	20.7	19.6
6	16.8	15.8

a) ¿Difieren significativamente las medias de ambos métodos? Antes de aplicar un test, observe si es razonable hacer las suposiciones correspondientes.
b) En caso de que responda afirmativamente a la pregunta anterior, calcule un IC al 95% para la diferencia de las medias.
c) Si el valor obtenido con el nuevo método para la muestra 1 hubiese sido 20.0 (en lugar de 23.8), ¿cuál hubiese sido su respuesta al item a)?
d) Compare los resultados obtenidos en a) y en c).

12. Vuelva a responder las preguntas del ejercicio 4 de la Práctica 1, pero ahora cuando se le pida hacer una comparación, aplique un test de hipótesis que considere adecuado.

13. Vuelva a responder las preguntas del ejercicio 6 de la Práctica 1, pero ahora cuando se le pida hacer una comparación, aplique un test de hipótesis que considere adecuado.

14. (continuación del ejercicio 3) Para cada uno de los conjuntos de datos para los que no es razonable suponer un modelo de Gauss, aplique, cuando sea posible, un test para estudiar si hay evidencia de error sistemático.