ESTADISTICA NO PARAMETRICA

La mayor parte de los procedimientos de prueba de hipótesis que se presentan en las unidades anteriores se basan en la suposición de que las muestras aleatorias se seleccionan de poblaciones normales. Afortunadamente, la mayor parte de estas pruebas aún son confiables cuando experimentamos ligeras desviaciones de la normalidad, en particular cuando el tamaño de la muestra es grande.

Tradicionalmente, estos procedimientos de prueba se denominan ***métodos paramétricos***. En esta sección se consideran varios procedimientos de prueba alternativos, llamados ***no paramétricos ó métodos de distribución libre***, que a menudo no suponen conocimiento de ninguna clase acerca de las distribuciones de las poblaciones fundamentales, excepto que éstas son continuas.

Los procedimientos no paramétricos o de distribución libre se usan con mayor frecuencia por los analistas de datos. Existen muchas aplicaciones en la ciencia y la ingeniería donde los datos se reportan no como valores de un continuo sino más bien en una escala ordinal tal que es bastante natural asignar rangos a los datos.

Un ejemplo donde se aplica una prueba no paramétrica es el siguiente, dos jueces deben clasificar cinco marcas de cerveza de mucha demanda mediante la asignación de un grado de 1 a la marca que se considera que tiene la mejor calidad global, un grado 2 a la segunda mejor, etcétera. Se puede utilizar entonces una prueba no paramétrica para determinar donde existe algún acuerdo entre los dos jueces.

Se debe señalar que hay varias desventajas asociadas con las pruebas no paramétricas. En primer lugar, no utilizan la información que proporciona la muestra, y por ello una prueba no paramétrica será menos eficiente que el procedimiento paramétrico correspondiente, cuando se pueden aplicar ambos métodos. En consecuencia, para lograr la misma potencia, una prueba no paramétrica requerirá la correspondiente prueba no paramétrica.

Como se indicó antes, ligeras divergencias de la normalidad tienen como resultado desviaciones menores del ideal para las pruebas paramétricas estándar. Esto es cierto en particular para la prueba t y la prueba F. En el caso de la prueba t y la prueba F, el valor P citado puede ser ligeramente erróneo si existe una violación moderada de la suposición de normalidad.
En resumen, si se puede aplicar una prueba paramétrica y una no paramétrica al mismo conjunto de datos, debemos aplicar la técnica paramétrica más eficiente. Sin embargo, se debe reconocer que las suposiciones de normalidad a menudo no se pueden justificar, y que no siempre se tienen mediciones cuantitativas.

PRUEBA DEL SIGNO

La prueba del signo se utiliza para probar la hipótesis sobre la mediana $\tilde{\mu}$ de una distribución continua. La mediana de una distribución es un valor de la variable aleatoria X tal que la probabilidad de que un valor observado de X sea menor o igual, o mayor o igual, que la mediana es 0.5. Esto es, $P(X \leq \tilde{\mu}) = P(X \geq \tilde{\mu}) = 0.5$.

Puesto que la distribución normal es simétrica, la media de una distribución normal es igual a la mediana. Por consiguiente, la prueba del signo puede emplearse para probar hipótesis sobre la media de una población normal.

Suponga que las hipótesis son:

$H_0; \tilde{\mu} = \tilde{\mu}_o$

$H_1; \tilde{\mu} < \tilde{\mu}_o$

Supóngase que X_1, X_2, . . . , X_n es una muestra aleatoria tomada de la población de interés. Fórmense las diferencias

$X_i - \tilde{\mu}_o, i = 1,2,\ldots n$

Ahora bien si la hipótesis nula $H_0; \tilde{\mu} = \tilde{\mu}_o$ es verdadera, cualquier diferencia $X_i - \tilde{\mu}_o$ tiene la misma probabilidad de ser negativa o positiva. Un estadístico de prueba apropiado es el número de estas diferencias que son positivas, por ejemplo R^+. Por consiguiente, la prueba de la hipótesis nula es en realidad una prueba de que el número de signos positivos es un valor de una variable aleatoria binomial con parámetro P = ½. Puede calcularse un valor P para el número observado de signos positivos r^+ directamente de la distribución binomial. Al probar la hipótesis que se muestra al principio, se rechaza H_0 en favor de H_1 sólo si la proporción de signos positivos es suficientemente menor que ½ (o de manera equivalente, cada vez que el número observado de signos positivos r^+ es muy pequeño).

Por tanto, si el valor P calculado P = P($R^+ \leq r^+$ cuando p = 1/2) es menor o igual que algún nivel de significancia seleccionado previamente, entonces se rechaza H_0 y se concluye que H_1 es verdadera.

Para probar la otra hipótesis unilateral

$H_0; \tilde{\mu} = \tilde{\mu}_o$

$H_1; \tilde{\mu} > \tilde{\mu}_o$

Se rechaza H_0 en favor de H_1 sólo si el número observado de signos más, r^+, es grande o, de manera equivalente, cada vez que la fracción observada de signos positivos es significativamente mayor que ½. En consecuencia, si el valor P calculado P = P($R^+ \geq r^+$ cuando p = 1/2) es menor que α, entonces H_0 se rechaza y se concluye que H_1 es verdadera.

También puede probarse la alternativa bilateral. Si las hipótesis son:

$$H_0; \tilde{\mu} = \tilde{\mu}_0$$

$$H_1; \tilde{\mu} \neq \tilde{\mu}_0$$

Se rechaza H_0 si la proporción de signos positivos difiere de manera significativa de ½ (ya se por encima o por debajo). Esto es equivalente a que el número observado de signos r^+ sea suficientemente grande o suficientemente pequeño. Por tanto, si $r^+ > n/2$ el valor P es

P=2P($R^+ \leq r^+$ cuando p = ½)
Y si $r^+ > n/2$ el valor P es
P=2P($R^+ \geq r^+$ cuando p = ½)

Si el valor P es menor que algún nivel preseleccionado α, entonces se rechaza H_0 y se concluye que H_1 es verdadera.

Ejemplos:

1. Un artículo informa cerca de un estudio en el que se modela el motor de un cohete reuniendo el combustible y la mezcla de encendido dentro de un contenedor metálico. Una característica importante es la resistencia al esfuerzo cortante de la unión entre los dos tipos de sustancias. En la siguiente tabla se muestran los resultados obtenidos al probar 20 motores seleccionados al azar. Se desea probar la hipótesis de que la mediana de la resistencia al esfuerzo cortante es 2000 psi, utilizando α = 0.05.

Solución:

Se mostrará la tabla del ejercicio y es función del investigador poner los signos con respecto a la mediana.

$$H_0; \hat{\mu} = 2000 psi$$

$$H_1; \hat{\mu} \neq 2000 psi$$

Observación	**Resistencia al esfuerzo cortante x_i**	**Signo de la diferencia x_i-2000**	**Observación**	**Resistencia al esfuerzo cortante x_i**	**Signo de la diferencia x_i-2000**
1	2158.70	**+**	11	2165.20	**+**
2	1678.15	**-**	12	2399.55	**+**
3	2316.00	**+**	13	1779.80	**-**
4	2061.30	**+**	14	2336.75	**+**
5	2207.50	**+**	15	1765.30	**-**
6	1708.30	**-**	16	2053.50	**+**
7	1784.70	**-**	17	2414.40	**+**
8	2575.10	**+**	18	2200.50	**+**
9	2357.90	**+**	19	2654.20	**+**
10	2256.70	**+**	20	1753.70	**-**

De la tabla se puede observar que el estadístico de prueba $r^+ = 14$.

Regla de decisión:

Si el valor de P correspondiente a $r^+=14$ es menor o igual que α =0.05 se rechaza H_0.

Cálculos:

Puesto que $r^+=14$ es mayor que n/2=20/2=10, el valor de P se calcula de

$P=2P(R^+ \geq 14$ cuando $p = ½)$

La P se calcula con la fórmula de la distribución binomial:

$$P = 2\sum_{r=14}^{20} {}_{20}C_r (0.5)^r (0.5)^{20-r} = 0.1153$$

Conclusión:

Como P=0.1153 no es menor que α =0.05, no es posible rechazar la hipótesis nula de que la mediana de la resistencia al esfuerzo constante es 2000 psi.

Otra manera de resolver el problema es con Aproximación normal:

Cuando p=0.5, la distribución binomial esta bien aproximada por la distribución normal cuando *n* es al menos 10. Por tanto, dado que la media de la distribución binomial es *np* y la varianza es *npq*, la distribución de R^+ es aproximadamente normal con media 0.5*n* y varianza 0.25*n*, cada vez que *n* es moderadamente grande. Por consiguiente las hipótesis pueden probarse con el estadístico:

$$Z = \frac{r^+ - 0.5n}{0.5\sqrt{n}}$$

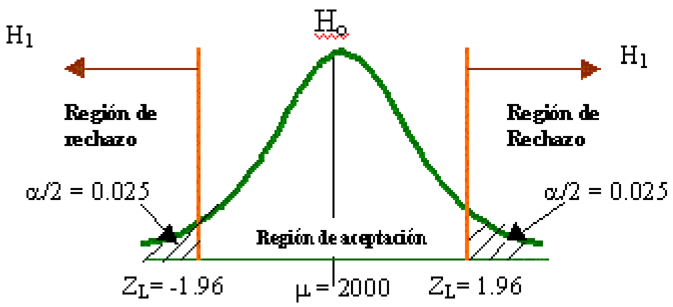
Las reglas de decisión se establecerán como cualquier ensayo en una distribución muestral en donde se utiliza la distribución normal.

Para resolver el problema anterior:

$$H_0 ; \hat{\mu} = 2000 psi$$

$$H_1 ; \hat{\mu} \neq 2000 psi$$

Como la es mayor que 10 se utilizará la aproximación normal.



Regla de Decisión:

Si $-1.96 \leq Z_R \leq 1.96$ No se rechaza H_o

Si $Z_R < -1.96$ ó si $Z_R > 1.96$ Se rechaza H_o

Cálculos:

$$Z = \frac{r^+ - 0.5n}{0.5\sqrt{n}} = \frac{14 - (0.5)(20)}{0.5\sqrt{20}} = 1.789$$

Decisión y Conclusión:

Como 1.789 esta entre –1.96 y 1.96, no se rechaza H_0 y se concluye con un α =0.05 que la mediana es de 2000 psi.

Prueba del Signo para Muestras Pareadas

También se puede utilizar la prueba de signo para probar la hipótesis nula $\tilde{\mu}_1 - \tilde{\mu}_2 = d_0$ para observaciones pareadas. Aquí se reemplaza cada diferencia, d_i, con un signo más o menos dependiendo si la diferencia ajustada, d_i-d_0, es positiva o negativa. A lo largo de esta sección suponemos que las poblaciones son simétricas. Sin embargo, aun si las poblaciones son asimétricas se puede llevar a cabo el mismo procedimiento de prueba, pero las hipótesis se refieren a las medianas poblacionales en lugar de las medias.

Ejemplo:

1. Una compañía de taxis trata de decidir si el uso de llantas radiales en lugar de llantas regulares con cinturón mejora la economía de combustible. Se equipan 16 automóviles con llantas radiales y se manejan por un recorrido de prueba establecido. Sin cambiar de conductores, se equipan los mismos autos con llantas regulares con cinturón y se manejan una vez más por el recorrido de prueba. Se registra el consumo de gasolina, en kilómetros por litro, de la siguiente manera:

Automóvil	**Llantas radiales**	**Llantas con cinturón**
1	4.2	4.1
2	4.7	4.9
3	6.6	6.2
4	7.0	6.9
5	6.7	6.8
6	4.5	4.4
7	5.7	5.7
8	6.0	5.8
9	7.4	6.9
10	4.9	4.9
11	6.1	6.0
12	5.2	4.9
13	5.7	5.3

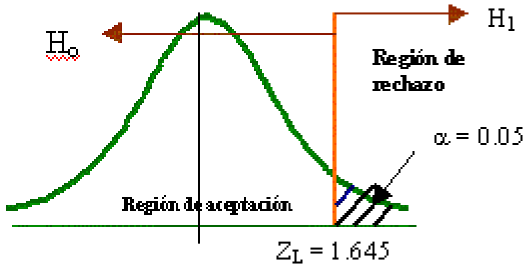
14	6.9	6.5
15	6.8	7.1
16	4.9	4.8

¿Se puede concluir en el nivel de significancia de 0.05 que los autos equipados con llantas radiales obtienen mejores economías de combustible que los equipados con llantas regulares con cinturón?

Solución:

$H_0; \tilde{\mu}_R - \tilde{\mu}_C = 0$

$H_1; \tilde{\mu}_R - \tilde{\mu}_C > 0$



Regla de decisión:

Si $z_R \leq$ 1.645 no se rechaza H_o.
Si $z_R >$ 1.645 se rechaza H_o.

Se procede a realizar las diferencias entre de los kilómetros por litro entre llantas radiales y con cinturón:

Automóvil	Llantas radiales	Llantas con cinturón	d
1	4.2	4.1	+
2	4.7	4.9	-
3	6.6	6.2	+
4	7.0	6.9	+
5	6.7	6.8	-
6	4.5	4.4	+

7	5.7	5.7	**0**
8	6.0	5.8	**+**
9	7.4	6.9	**+**
10	4.9	4.9	**0**
11	6.1	6.0	**+**
12	5.2	4.9	**+**
13	5.7	5.3	**+**
14	6.9	6.5	**+**
15	6.8	7.1	**-**
16	4.9	4.8	**+**

Al observar las diferencias se ve que sólo existe una n=14, ya que se descartan los valores de cero. Se tiene $r^+ = 11$

$$Z = \frac{r^+ - 0.5n}{0.5\sqrt{n}} = \frac{11-(0.5)(14)}{0.5\sqrt{14}} = 2.14$$

Decisión y conclusión:

Como 2.14 es mayor a 1.645 se rechaza H_0 y se concluye con un $\alpha = 0.05$ que las llantas radiales mejoran la economía de combustible.

PRUEBA DE RANGO CON SIGNO DE WILCOXON

Se puede notar que la prueba de signo utiliza sólo los signos más y menos de las diferencias entre las observaciones y μ_0 en el caso de una muestra, o los signos más y menos de las diferencias entro los pares de observaciones en el caso de la muestra pareada, pero no toma en consideración la magnitud de estas diferencias. Una prueba que utiliza dirección y magnitud, propuesta en 1945 por Frank Wilcoxon, se llama ahora comúnmente **prueba de rango con signo de Wilcoxon.**

Esta prueba se aplica en el caso de una distribución continua simétrica. Bajo esta condición se puede probar la hipótesis nula $\mu = \mu_0$. Primero se resta μ_0 de cada valor muestral y se descarta todas las diferencias iguales a cero. Se asigna un rango de 1 a la diferencia absoluta más pequeña, un rango de 2 a la siguiente más pequeña, y así sucesivamente. Cuando el valor absoluto de dos o más diferencias es el mismo, se asigna a cada uno el promedio de los rangos que se asignarían si las diferencias se distinguieran.

Por ejemplo, si la quinta y sexta diferencia son iguales en valor absoluto, a cada una se le asignaría un rango de 5.5. Si la hipótesis $\mu = \mu_0$ es verdadera, el total de los rangos que corresponden a las diferencias positivas debe ser casi igual al total de los rangos que corresponden a las diferencias negativas. Se representan esos totales como w_+ y w_-, respectivamente. Se designa el menor de w_+ y w_- con w.

Al seleccionar muestras repetidas esperaríamos que variarían w_+ y w_-, y por tanto w. De esta manera se puede considerar a w_+ y w_-, y w como valores de las correspondiente variables aleatorias W_+, W_-, y W. La hipótesis nula $\mu = \mu_0$ se puede rechazar a favor de la alternativa $\mu < \mu_0$ sólo si w_+ es pequeña y w_- es grande. Del mismo modo, la alternativa $\mu > \mu_0$ se puede aceptar sólo si w_+ es grande y w_- es pequeña. Para una alternativa bilateral se puede rechazar H_0 a favor de H_1 si w_+ o w_- y por tanto w son suficientemente pequeñas. No importa cuál hipótesis alternativa puede ser, rechazar la hipótesis nula cuando el valor de la estadística apropiada W_+, W_-, o W es suficientemente pequeño.

Dos Muestras con Observaciones Pareadas

Para probar la hipótesis nula de que se muestrean dos poblaciones simétricas continuas con $\mu_1 = \mu_2$ para el caso de una muestra pareada, se clasifican las diferencias de las observaciones paradas sin importar el signo y se procede como en el caso de una muestra. Los diversos procedimientos de prueba para los casos de una sola muestra y de una muestra pareada se resumen en la siguiente tabla:

Para probar H_0	Contra H_1	Calcular
$\mu = \mu_0$	$\mu < \mu_0$	w_+
	$\mu > \mu_0$	w_-
	$\mu \neq \mu_0$	w
$\mu_1 = \mu_2$	$\mu_1 < \mu_2$	w_+
	$\mu_1 > \mu_2$	w_-
	$\mu_1 \neq \mu_2$	w

No es difícil mostrar que siempre que $n<5$ y el nivel de significancia no exceda 0.05 para una prueba de una cola ó 0.10 para una prueba de dos colas, todos los valores posibles de w_+, w_- o w conducirán a la aceptación de la hipótesis nula. Sin embargo, cuando $5 \leq n \leq 30$, la tabla A.16 muestra valores críticos aproximados de W_+ y W_- para niveles de significancia iguales a 0.01, 0.025 y 0.05 para una prueba de una cola, y valores críticos de W para niveles de significancia iguales a 0.02, 0.05 y 0.10 para una prueba de dos colas. La hipótesis nula se rechaza si el valor calculado

w_+, w_- o w **es menor o igual que** el valor de tabla apropiado. Por ejemplo, cuando n=12 la tabla A.16 muestra que se requiere un valor de $w_+ \leq 17$ para que la alternativa unilateral $\mu < \mu_0$ sea significativa en el nivel 0.05.

Ejemplos:

1. Los siguientes datos representan el número de horas que un compensador opera antes de requerir una recarga: 1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2 y 1.7. Utilice la prueba de rango con signo para probar la hipótesis en el nivel de significancia de 0.05 que este compensador particular opera con una media de 1.8 horas antes de requerir una recarga.

Solución:

H_0; $\mu = 1.8$

H_1; $\mu \neq 1.8$

Se procederá a efectuar las diferencias y a poner rango con signo a los datos.

Dato	**d_i = dato - 1.8**	**Rangos**
1.5	-0.3	**5.5**
2.2	0.4	**7**
0.9	-0.9	**10**
1.3	-0.5	**8**
2.0	0.2	**3**
1.6	-0.2	**3**
1.8	0	**Se anula**
1.5	-0.3	**5.5**
2.0	0.2	**3**
1.2	-0.6	**9**
1.7	-0.1	**1**

Regla de decisión:

Para una n = 10, después de descartar la medición que es igual a 1.8, la tabla A.16 muestra que la región crítica es $w \leq 8$.

Cálculos:

$w_+ = 7 + 3 + 3 = 13$

$w_- = 5.5 + 10 + 8 + 3 + 5.5 + 9 + 1 = 42$

por lo que $w = 13$ (menor entre w_+ y w_-).

Decisión y Conclusión:

Como 13 no es menor que 8, no se rechaza H_0 y se concluye con un $\alpha = 0.05$ que el tiempo promedio de operación no es significativamente diferente de 1.8 horas.

2. Se afirma que un estudiante universitario de último año puede aumentar su calificación en el área del campo de especialidad del examen de registro de graduados en al menos 50 puntos si de antemano se le proporcionan problemas de muestra. Para probar esta afirmación, se dividen 20 estudiantes del último año en 10 pares de modo que cada par tenga casi el mismo promedio de puntos de calidad general en sus primeros años en la universidad. Los problemas y respuestas de muestra se proporcionan al azar a un miembro de cada par una semana antes del examen. Se registran las siguientes calificaciones del examen:

Par	**Con problemas de muestra**	**Sin problemas de muestra**
1	531	509
2	621	540
3	663	688
4	579	502
5	451	424
6	660	683
7	591	568
8	719	748
9	543	530
10	575	524

Pruebe la hipótesis nula en el nivel de significancia de 0.05 de que los problemas aumentan las calificaciones en 50 puntos contra la hipótesis alternativa de que el aumento es menor a 50 puntos.

Solución:

La prueba de rango con signo también se puede utilizar para probar la hipótesis nula $\mu_1 - \mu_2 = d_0$. En este caso las poblaciones no necesitan ser simétricas. Como con la prueba de signo, se resta d_0 de cada diferencia, se clasifican las diferencias ajustadas sin importar el signo y se aplica el mismo procedimiento.

En este caso $d_0 = 50$, por lo que se procede a calcular las diferencias entre las muestras y luego restarles el valor de 50. Se representara con μ_1 y μ_2 la calificación media de todos los estudiantes que resuelven el examen en cuestión con y sin problemas de muestra, respectivamente.

H_0; $\mu_1 - \mu_2 = 50$

H_1; $\mu_1 - \mu_2 < 50$

Regla de decisión:

Para n=10 la tabla muestra que la región crítica es $w_+ \leq 11$.

Cálculos:

Par	**Con problemas de muestra**	**Sin problemas de muestra**	**d_i**	**$d_i - d_0$**	**Rangos**
1	531	509	22	-28	**5**
2	621	540	81	31	**6**
3	663	688	-25	-75	**9**
4	579	502	77	27	**3.5**
5	451	424	27	-23	**2**
6	660	683	-23	-73	**8**
7	591	568	23	-27	**3.5**
8	719	748	-29	-79	**10**
9	543	530	13	-37	**7**

10	575	524	51	1	**1**

$w_+ = 6 + 3.5 + 1 = 10.5$

Decisión y Conclusión:

Como 10.5 es menor que 11 se rechaza H_0 y se concluye con un α = 0.05 que los problemas de muestra, en promedio, no aumentan las calificaciones de registro de graduados en 50 puntos.

Aproximación Normal para Muestras Grandes

Cuando n $\geq$ 15, la distribución muestral de W_+ ó W_- se aproxima a la distribución normal con media $\mu_w = \frac{n(n+4)}{4}$ y varianza $\sigma_w^2 = \frac{n(n+1)(2n+1)}{24}$.

Por tanto, cuando *n* excede el valor más grande en la tabla A.16, se puede utilizar la estadística

$$z = \frac{w_+ - \mu_w}{\sigma_w}$$

para determinar la región crítica de la prueba.

Problemas Propuestos

1. Se lanza 180 veces un dado con los siguientes resultados:

X	1	2	3	4	5	6
f	28	36	36	30	27	23

2. ¿Es un dado balanceado? Utilice un α = 0.01.
3. Se supone que una máquina mezcla cacahuates, avellanas, anacardos y pacanas a razón de 5:2:2:1. Se encuentra que una lata que contiene 500 de estas nueces mezcladas tiene 269 cacahuates, 112 avellanas, 74 anacardos y 45 pacanas. Al nivel de significancia de 0.05 pruebe la hipótesis de que la máquina mezcla las nueces a razón de 5:2:2:1.
4. Se seleccionan tres canicas de una urna que contiene cinco canicas rojas y tres verdes. Después de registrar el número *x* de canicas rojas, las canicas se reemplazan en la urna y el experimento se repite 112 veces. Los resultados que se obtienen son los siguientes:

x	0	1	2	3
f	1	31	55	25

5. Pruebe la hipótesis con un nivel de significancia de 0.05, de que los datos registrados se pueden ajustar a una distribución hipergeométrica.

6. Se lanza una moneda hasta que sale cara y se registra el número de lanzamientos x. Después de repetir el experimento 256 veces, se obtuvieron los siguientes resultados:

X	1	2	3	4	5	6	7	8
f	136	60	34	12	9	1	3	1

7. Pruebe la hipótesis con un nivel de significancia de 0.05 de que la distribución observada de x se puede ajustar por una distribución geométrica g(x;1/2), x = 1, 2, 3, …
8. Con los siguientes datos, pruebe la bondad de ajuste entre las frecuencias de clase que se observan y las frecuencias esperadas correspondientes de una distribución normal con μ = 65 y σ = 21, utilice un nivel de significancia de 0.05.

Límite de clase	**Frecuencia**
10 - 19	3
20 – 29	2
30 – 39	3
40 – 49	4
50 – 59	5
60 – 69	11
70 – 79	14
80 – 89	14
90 - 99	4

9. En un experimento para estudiar la dependencia de la hipertensión de los hábitos de fumar, se tomaron los siguientes datos de 180 individuos:

	No fumadores	**Fumadores moderados**	**Fumadores empedernidos**
Con hipertensión	21	36	30
Sin hipertensión	48	26	19

10. Pruebe la hipótesis de que la presencia o ausencia de hipertensión es independiente de los hábitos de fumar. Utilice un nivel de significancia de 0.05.

11. Una muestra aleatoria de 200 hombres casados, todos retirados, se clasifica de acuerdo con la educación y el número de hijos:

Número de hijos

Educación	**0 - 1**	**2 - 3**	**Más de 3**
Elemental	14	37	32
Secundaria	19	42	17
Universidad	12	17	10

Pruebe la hipótesis, con un nivel de significancia de 0.05, de que el tamaño de la familia es independiente del nivel de instrucción del padre.

12. Se comparan dos tipos de instrumentos para medir la cantidad de monóxido de azufre en la atmósfera en un experimento de contaminación atmosférica. Se registraron las siguientes lecturas diarias en un período de dos semanas:

Día	**Instrumento A**	**Instrumento B**
1	0.96	0.87
2	0.82	0.74
3	0.75	0.63
4	0.61	0.55
5	0.89	0.76
6	0.64	0.70
7	0.81	0.69
8	0.68	0.57
9	0.65	0.53
10	0.84	0.88
11	0.59	0.51
12	0.94	0.79
13	0.91	0.84

14	0.77	0.63

13. Con el uso de la aproximación normal a la distribución binomial, realice una prueba de signo para determinar si los diferentes instrumentos conducen a diferentes resultados. Utilice un nivel de significancia de 0.05.
14. Los siguientes datos representan el tiempo, en minutos, que un paciente tiene que esperar durante 12 visitas al consultorio de una doctora antes de ser atendido por ésta:

17 15 20 20

32 28 12 26

25 25 35 24

Utilice la prueba de rango con signo al nivel de significancia de 0.05 para probar la afirmación de la doctora de que la media del tiempo de espera para sus pacientes no es mayor que 20 minutos antes de entrar al consultorio.

15. Los pesos de cuatro personas antes de que dejan de fumar y cinco semanas después de dejar de fumar, en kilogramos, son los siguientes:

Individuo 1 2 3 4 5

Antes 66 80 69 52 75

Después 71 82 68 56 73

Utilice la prueba de rango con signo para observaciones pareadas para probar la hipótesis, en el nivel de significancia de 0.05, de que dejar de fumar no tiene efecto en el peso de una persona contra la alternativa del que el peso aumenta si deja de fumar.

16. Los siguientes son los números de recetas surtidas por dos farmacias en un período de 20 días:

Día	**Farmacia A**	**Farmacia B**
1	19	17
2	21	15
3	15	12
4	17	12
5	24	16
6	12	15
7	19	11
8	14	13
9	20	14

10	18	21
11	23	19
12	21	15
13	17	11
14	12	10
15	16	20
16	15	12
17	20	13
18	18	17
19	14	16
20	22	18

17. Utilice la prueba de rango con signo al nivel de significancia de 0.01 para determinar si las dos farmacias, en promedio, surten el mismo número de recetas contra la alternativa de que la farmacia A surte más recetas que la farmacia B.
18. Se afirma que una nueva dieta reducirá el peso de una persona 4.5 kilogramos, en promedio, en un período de dos semanas. Se registran los pesos de 10 mujeres que siguen esta dieta antes y después de un período de dos semanas, y se obtienen los siguientes datos:

Mujer	**Peso antes**	**Peso después**
1	58.5	60.0
2	60.3	54.9
3	61.7	58.1
4	69.0	62.1
5	64.0	58.5
6	62.6	59.9
7	56.7	54.4

8	63.6	60.2
9	68.2	62.3
10	59.4	58.7

19. Utilice la prueba de rango con signo al nivel de significancia de 0.05 para probar la hipótesis de que la dieta reduce la mediana del peso en 4.5 kilogramos contra la hipótesis alternativa de que la mediana de la diferencia en pesos es menor que 4.5 kilogramos.
20. Se toman 10 muestras de un baño de cultivo sobre placa utilizado en un proceso de fabricación de componentes electrónicos, y se mide el pH del baño. Los valores de pH medidos son 7.91, 7.85, 6.82, 8.01, 7.46, 6.95, 7.05, 7.35, 7.25, 7.42. Los ingenieros creen que el valor de la mediana del pH es 7.0. ¿ La muestra indica que esta proposición es correcta? Utilice la prueba del signo con α = 0.05 para investigar esta hipótesis. Encuentre el valor P de esta prueba.
21. Se mide de manera rutinaria el nivel de impurezas (en ppm) en un producto químico intermedio. En una prueba reciente se observan los datos siguientes:

2.4	2.5	1.7	1.6	1.9	2.6	1.3	1.9	2.0	2.5	2.6
2.3	2.0	1.8	1.3	1.7	2.0	1.9	2.3	1.9	2.4	1.6

¿Puede afirmarse que la mediana del nivel de impureza es menor que 2.5 ppm? Establezca y pruebe la hipótesis apropiada utilizando la prueba de signo con α = 0.05. ¿Cuál es el valor P de esta prueba?

Respuestas a los Problemas Propuestos

1. Región crítica $X^2 > 15.086$, $X^2 = 4.47$ por lo tanto no rechazar H_0, el dado está balanceado.
2. Región crítica $X^2 > 7.815$, $X^2 = 10.14$, rechazar H_0. Las nueces no están mezcladas en la proporción 5:2:2:1.
3. Región crítica $X^2 > 5.991$, $X^2 = 1.67$, no rechazar H_0. Los datos se ajustan a una distribución hipergeométrica.
4. Región crítica $X^2 > 11.07$, $X^2 = 2.57$, no rechazar H_0. Los datos se ajustan a una distribución geométrica.
5. Región crítica $X^2 > 12.592$, $X^2 = 12.78$, rechazar H_0. Los datos no se ajustan a una distribución normal.
6. Región crítica $X^2 > 5.991$, $X^2 = 14.6$, rechazar H_0. La presencia o ausencia de hipertensión y hábitos de fumar no son independientes.
7. Región crítica $X^2 > 9.488$, $X^2 = 7.54$, no rechazar H_0. El tamaño de la familia es independiente del nivel se educación del padre.
8. Región crítica $-1.96 \leq z \leq 1.96$, z= 2.67, rechazar H_0.
9. Región crítica $w_- \leq 11$ para una n=10, $w_- = 12.5$, no rechazar H_0.

10. Región crítica $w_+ \leq 1$ para $n = 5$, $w_+ = 3.5$, no rechazar H_0.
11. Región crítica $z>2.575$. $z= 2.80$, rechazar H_0, la farmacia A surte más recetas que la farmacia B.
12. Región crítica $w_+ \leq 11$ para una $n = 10$. $w_+ = 17.5$, no rechazar H_0.
13. $2P(R^+ \geq 8 / p = 0.5) = 0.109$, como no es menor a 0.05, no se rechaza H_0.
14. H_0; $\tilde{\mu} = 2.5$ H_1; $\tilde{\mu} < 2.5$ $P(R^+ \leq 2/ p = 0.5) = 0.0002$, se rechaza H_0.