



CAPÍTULO 11

Inferencias acerca de varianzas poblacionales

CONTENIDO

LA ESTADÍSTICA EN
LA PRÁCTICA:
LA GENERAL ACCOUNTING
OFFICE DE ESTADOS UNIDOS

11.1 INFERENCIAS ACERCA
DE UNA VARIANZA
POBLACIONAL
Estimación por intervalos
Pruebas de hipótesis

11.2 INFERENCIAS ACERCA
DE DOS VARIANZAS
POBLACIONALES



LA ESTADÍSTICA *en* LA PRÁCTICA

LA GENERAL ACCOUNTING OFFICE DE ESTADOS UNIDOS* WASHINGTON, D. C.

La General Accounting Office de Estados Unidos (GAO) es una organización de auditoría independiente no política perteneciente al área legislativa del gobierno federal. Los auditores de la GAO determinan la efectividad tanto de los programas federales existentes como de los propuestos. Para realizar su labor, los auditores deben ser competentes en revisión de documentos, investigación legislativa y técnicas de análisis estadístico.

En una ocasión los auditores de la GAO estudiaron un programa del Department of Interior que tenía por objeto limpiar los ríos y lagos del país. Como parte de este programa se otorgaron subvenciones a las ciudades pequeñas de Estados Unidos. El Congreso pidió a la GAO que determinara la eficiencia con la que operaba este programa. Con tal objeto, la GAO revisó documentos y visitó varias de las plantas de tratamiento de desechos.

Uno de los objetivos de la auditoría de la GAO era verificar en las plantas que las aguas residuales (desechos tratados) cumplieran determinadas normas. En las auditorías se revisaba, entre otras cosas, datos muestrales sobre contenido de oxígeno, pH y cantidad de sólidos en suspensión en las aguas residuales. Un requisito del programa era que en cada planta diario se realizaran diversas pruebas y que los datos obtenidos se enviaran al departamento de ingeniería del estado. Los datos de la investigación servían para determinar si las características de las aguas residuales se encontraban dentro de límites aceptables.

Así, por ejemplo, examinaron con cuidado los valores promedio de pH. También analizaron la varianza en los valores del pH de las aguas residuales. La prueba de hipótesis acerca de la varianza del pH en la población de aguas residuales fue la siguiente.

$$\begin{aligned} H_0: \sigma^2 &= \sigma_0^2 \\ H_a: \sigma^2 &\neq \sigma_0^2 \end{aligned}$$

En esta prueba, σ_0^2 corresponde a la varianza poblacional esperada en los valores de pH de una planta funcionando

*Los autores agradecen a William R. Fawle, director de aseguramiento de la calidad de la empresa Colgate-Palmolive por proporcionarles este artículo para *La estadística en la práctica*.



Las aguas residuales de esta planta deben estar dentro de un determinado intervalo de valores de pH. © John Boykin/CORBIS.

adecuadamente. Para una de las plantas la hipótesis nula fue rechazada. Análisis más cuidadosos indicaron que en esa planta la varianza del pH era significativamente menor de lo normal.

Los auditores visitaron la planta para revisar el equipo de medición y analizar los resultados con el director de la planta. Encontraron que el equipo no se usaba para la medición de pH debido a que el operador no conocía su funcionamiento. Un ingeniero había informado al operador de los valores de pH aceptables y éste sólo anotaba valores aceptables sin realizar ninguna medición. La varianza inusualmente baja de los datos de esta planta hicieron que se rechazara H_0 . La GAO pensó que era probable que otras plantas tuvieran problemas similares y recomendó un programa de capacitación para los operadores, con objeto de mejorar la recolección de datos para el programa de control de la contaminación.

En este capítulo se verá cómo hacer inferencias estadísticas acerca de las varianzas de una o de dos poblaciones. También se presentarán dos nuevas distribuciones, la distribución chi-cuadrada y la distribución F , útiles para obtener estimación por intervalos y realizar pruebas de hipótesis para la varianza poblacional.

En los cuatro capítulos anteriores se vieron métodos de inferencia estadística para medias y proporciones poblacionales. En este capítulo se extiende dicho estudio a las varianzas poblacionales. Un ejemplo en que la varianza brinda una información importante para tomar una decisión es el caso de un proceso en el que se llenan recipientes con un detergente líquido. La máquina de llenado se ajusta de manera que logre un llenado medio de 16 onzas por envase. Aunque la media de llenado es importante, la varianza en los pesos de llenado también es relevante.

En muchos procesos de fabricación controlar la varianza del proceso es de suma importancia para conservar la calidad.

Es decir, aun cuando la máquina de llenado tenga un ajuste adecuado para una media de llenado de 16 onzas, no es de esperar que todos los envases tengan exactamente 16 onzas. Para calcular la varianza muestral de la cantidad de onzas en cada envase se toma una muestra de envases llenos. El valor de la varianza muestral sirve como estimación de la varianza en la población de envases que están siendo llenados en el proceso de producción. Si la varianza muestral es moderada, el proceso continúa. Pero, si la varianza muestral es grande, puede estar ocurriendo por exceso o defecto de llenado, aunque la media sea la correcta. En este caso habrá que reajustar la máquina de llenado con objeto de reducir la varianza de los envases.

En la primera sección se verán inferencias acerca de la varianza de una sola población. Después, procedimientos para inferencias acerca de varianzas de dos poblaciones.

11.1

Inferencias acerca de una varianza poblacional

La varianza muestral

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \quad (11.1)$$

es el estimador puntual de la varianza poblacional σ^2 . Cuando se hacen inferencias acerca de la varianza poblacional mediante la varianza muestral, la distribución muestral de la cantidad $(n - 1)s^2/\sigma^2$ es de utilidad. Esta distribución muestral se describe como sigue.

DISTRIBUCIÓN MUESTRAL DE $(n - 1)s^2/\sigma^2$

Siempre que de una población normal se tome una muestra aleatoria simple de tamaño n , la distribución muestral de

$$\frac{(n - 1)s^2}{\sigma^2} \quad (11.2)$$

será una distribución chi-cuadrada con $n - 1$ grados de libertad.

La distribución chi-cuadrada parte del muestreo de una población normal.

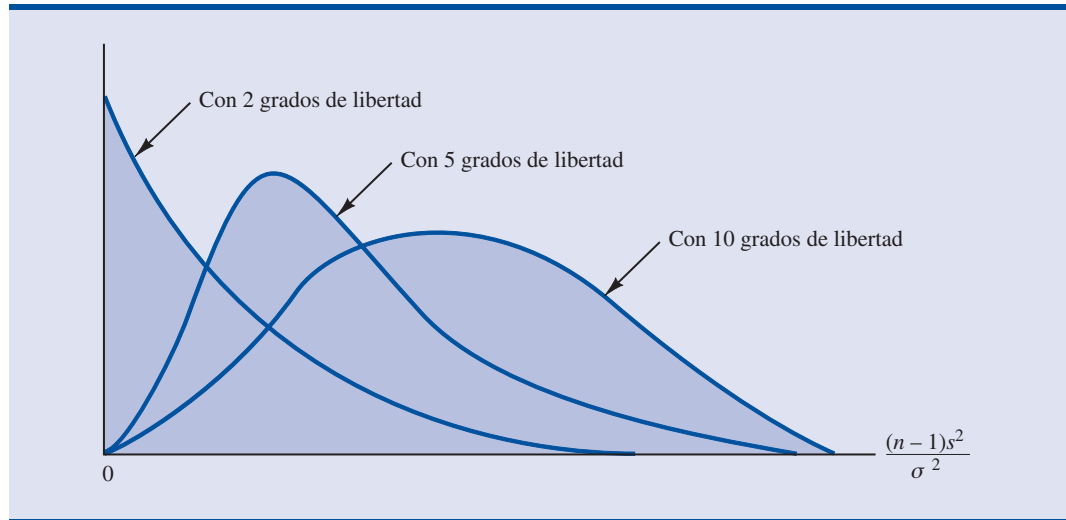
En la figura 11.1 se muestran algunas de las formas que puede tomar la distribución $(n - 1)s^2/\sigma^2$.

Como ya sabe, siempre que tome una muestra aleatoria simple de tamaño n de una población normal, la distribución muestral de $(n - 1)s^2/\sigma^2$ será una distribución chi-cuadrada, la cual sirve para obtener una estimación por intervalos y realizar pruebas de hipótesis para la varianza poblacional.

Estimación por intervalos

Con el fin de exponer cómo usar la distribución chi-cuadrada para obtener una estimación de la varianza poblacional σ^2 mediante un intervalo de confianza, suponga que desea estimar la varianza poblacional del proceso de llenado citado al comienzo de este capítulo. En una muestra de 20 envases encuentra que la varianza muestral de las cantidades de llenado es $s^2 = 0.0025$. Sin embargo, sabe que no puede esperar que la varianza de 20 envases corresponda al valor exacto de la varianza de toda la población de envases que se llenan en este proceso de producción. Así, deseará obtener una estimación por intervalo para la varianza poblacional.

FIGURA 11.1 EJEMPLOS DE DISTRIBUCIONES MUESTRALES DE $(n - 1)s^2/\sigma^2$ (DISTRIBUCIONES CHI-CUADRADA)



La notación χ^2_α denota el valor de la distribución chi-cuadrada que proporciona un área o probabilidad α a la *derecha* del valor χ^2_α . Por ejemplo, en la figura 11.2, aparece la distribución chi-cuadrada con 19 grados de libertad, en la que $\chi^2_{0.025} = 32.852$ significa que 2.5% de los valores chi-cuadrada se encuentran a la derecha de 32.852 y $\chi^2_{0.975} = 8.907$ significa que 97.5% de los valores chi-cuadrada se encuentran a la derecha de 8.907. Existen tablas que dan las áreas o probabilidades de la distribución chi-cuadrada. Consulte la tabla 11.1 y verifique que los valores de chi-cuadrada con 19 grados de libertad (renglón 19 de la tabla) son correctos. En la tabla 3 del apéndice B se encuentra una tabla más completa con valores chi-cuadrada.

En la gráfica de la figura 11.2 se ve que 0.95 o 95% de los valores chi-cuadrada se encuentran entre $\chi^2_{0.975}$ y $\chi^2_{0.025}$. Es decir, hay un 0.95 de probabilidad de obtener un valor χ^2 tal que.

$$\chi^2_{0.975} \leq \chi^2 \leq \chi^2_{0.025}$$

FIGURA 11.2 UNA DISTRIBUCIÓN CHI-CUADRADA CON 19 GRADOS DE LIBERTAD

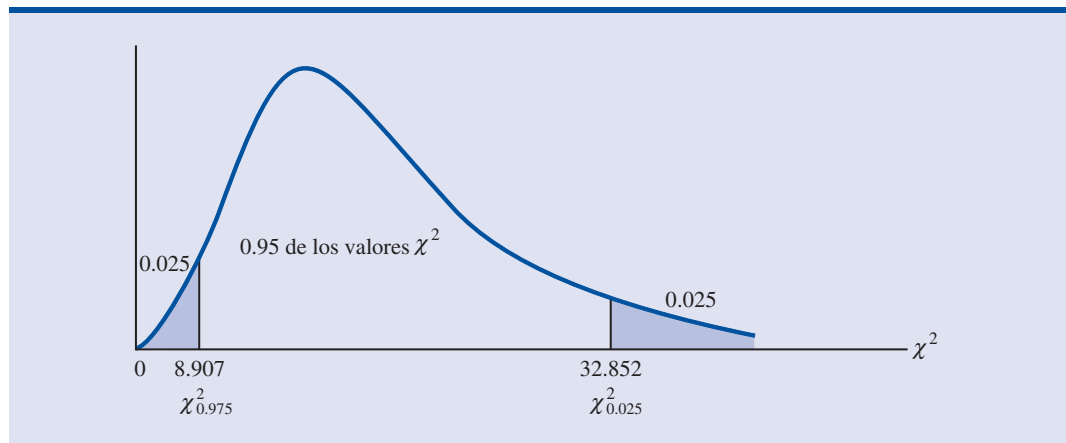
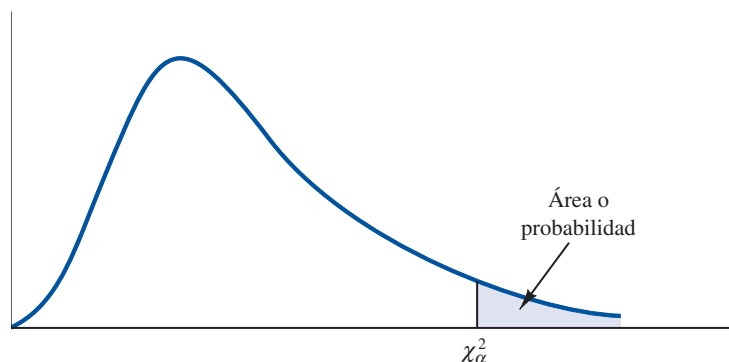


TABLA 11.1 ALGUNOS VALORES DE LA TABLA DE LA DISTRIBUCIÓN CHI-CUADRADA*

Grados de libertad	Área en la cola superior							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329
100	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807

*Nota: En la tabla 3 del apéndice B se encuentra una tabla más completa.

En la expresión (11.2) se dijo que $(n - 1)s^2/\sigma^2$ sigue una distribución chi-cuadrada, por tanto, se puede sustituir χ^2 por $(n - 1)s^2/\sigma^2$ y escribir

$$\chi_{0.975}^2 \leq \frac{(n - 1)s^2}{\sigma^2} \leq \chi_{0.025}^2 \quad (11.3)$$

En efecto, la expresión (11.3) proporciona una estimación por intervalo en la que 0.95 o 95% de todos los valores que puede tomar $(n - 1)s^2/\sigma^2$ están en el intervalo que va de $\chi_{0.975}^2$ y $\chi_{0.025}^2$. Ahora es necesario realizar algunas manipulaciones algebraicas a la expresión (11.3) para obtener una estimación por intervalo para la varianza poblacional s^2 . Modificando la desigualdad de la izquierda, se tiene

$$\chi_{0.975}^2 \leq \frac{(n - 1)s^2}{\sigma^2}$$

Por tanto

$$\sigma^2 \chi_{0.975}^2 \leq (n - 1)s^2$$

o

$$\sigma^2 \leq \frac{(n - 1)s^2}{\chi_{0.975}^2} \quad (11.4)$$

Con manipulaciones algebraicas similares a la desigualdad de la derecha de la expresión (11.3), se tiene

$$\frac{(n - 1)s^2}{\chi_{0.025}^2} \leq \sigma^2 \quad (11.5)$$

Los resultados de las expresiones (11.4) y (11.5) se combinan para obtener

$$\frac{(n - 1)s^2}{\chi_{0.025}^2} \leq \sigma^2 \leq \frac{(n - 1)s^2}{\chi_{0.975}^2} \quad (11.6)$$

Como la expresión (11.3) es verdadera para 95% de los valores $(n - 1)s^2/\sigma^2$, la expresión (11.6) proporciona una estimación por intervalo de confianza de 95% para la varianza poblacional σ^2 .

Ahora, de regreso con el problema de dar una estimación por intervalo para la varianza poblacional de las cantidades de llenado, recuerde que en la muestra de 20 envases, la varianza muestral fue $s^2 = 0.0025$. Dado que el tamaño de la muestra es 20, se tienen 19 grados de libertad. Como se muestra en la figura 11.2 ya se determinó que $\chi_{0.975}^2 = 8.907$ y $\chi_{0.025}^2 = 32.852$. Con estos valores en la expresión (11.6) se obtiene la siguiente estimación por intervalo para la varianza poblacional.

$$\frac{(19)(0.0025)}{32.852} \leq \sigma^2 \leq \frac{(19)(0.0025)}{8.907}$$

Con el fin de obtener un intervalo de confianza para la desviación estándar poblacional, sólo hace falta sacar la raíz cuadrada del límite inferior y del límite superior del intervalo de confianza para la varianza poblacional.

o

$$0.0014 \leq \sigma \leq 0.0053$$

Al sacar la raíz cuadrada de estos valores se obtiene el siguiente intervalo de confianza para la desviación estándar poblacional.

$$0.0380 \leq \sigma \leq 0.0730$$

De esta manera se ha ilustrado el proceso del uso de la distribución chi-cuadrada para obtener una estimación por intervalo para la varianza poblacional y para la desviación estándar poblacional. Observe, que, como se usaron $\chi^2_{0.975}$ y $\chi^2_{0.025}$, el coeficiente de confianza de la estimación por intervalo es 0.95. Extendiendo la expresión (11.6) al caso general, con cualquier coeficiente de confianza, se tiene la siguiente estimación por intervalo para la varianza poblacional.

ESTIMACIÓN POR INTERVALO PARA LA VARIANZA POBLACIONAL

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}} \quad (11.7)$$

donde los valores χ^2 están basados en una distribución chi-cuadrada con $n-1$ grados de libertad y donde $1-\alpha$ es el coeficiente de confianza.

Pruebas de hipótesis

Con σ_0^2 para denotar el valor hipotético de la varianza poblacional, las tres formas de una prueba de hipótesis para la varianza poblacional son las siguientes:

$$\begin{array}{lll} H_0: \sigma^2 \geq \sigma_0^2 & H_0: \sigma^2 \leq \sigma_0^2 & H_0: \sigma^2 = \sigma_0^2 \\ H_a: \sigma^2 < \sigma_0^2 & H_a: \sigma^2 > \sigma_0^2 & H_a: \sigma^2 \neq \sigma_0^2 \end{array}$$

Estas tres formas son semejantes a las tres formas, en los capítulos 9 y 10, para las pruebas de hipótesis de una cola y de dos colas para medias poblacionales y proporciones poblacionales.

En una prueba de hipótesis para la varianza poblacional se emplean el valor hipotético de la varianza poblacional σ_0^2 y la varianza muestral s^2 para calcular el valor del estadístico de prueba χ^2 . Si la población tiene una distribución normal, el estadístico de prueba es el siguiente:

ESTADÍSTICO DE PRUEBA EN PRUEBAS DE HIPÓTESIS PARA LA VARIANZA POBLACIONAL

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (11.8)$$

donde χ^2 tiene una distribución chi-cuadrada con $n-1$ grados de libertad.

Una vez calculado el valor del estadístico de prueba χ^2 , para determinar si se rechaza la hipótesis nula se emplea el método del valor- p o el método del valor crítico.

Considere ahora el siguiente ejemplo. La St. Louis Metro Bus Company de Estados Unidos, desea dar una imagen de confiabilidad haciendo que sus conductores sean puntuales en los horarios de llegada a las paradas. La empresa desea que haya poca variabilidad en dichos tiempos. En términos de la varianza de los tiempos de llegada a las paradas, la empresa desea que la varianza sea de 4 minutos o menos. Se formula la siguiente prueba de hipótesis para que la empresa pueda determinar si la varianza poblacional en los tiempos de llegada a las paradas es demasiado grande.

$$\begin{array}{l} H_0: \sigma^2 \leq 4 \\ H_a: \sigma^2 > 4 \end{array}$$



Suponer, tentativamente, que H_0 sea verdadera, es admitir que la varianza poblacional en los tiempos de llegada a las paradas se encuentra dentro de los lineamientos establecidos por la empresa. La H_0 se rechaza si las evidencias muestrales indican que la varianza poblacional excede los lineamientos de la empresa. En tal caso habrá que tomar medidas para reducir la varianza poblacional. Esta prueba de hipótesis se realiza usando como nivel de significancia $\alpha = 0.05$.

Asuma que en una muestra aleatoria de 24 llegadas a cierta parada en una intersección en el centro de la ciudad, la varianza muestral encontrada es $s^2 = 4.9$. Si la distribución poblacional de los tiempos de llegada a las paradas es aproximadamente normal, el valor del estadístico de prueba es el siguiente:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(24-1)(4.9)}{4} = 28.18$$

En la figura 11.3 se muestra la distribución chi-cuadrada con $n-1 = 24-1 = 23$ grados de libertad. Como ésta es una prueba de la cola superior, el área bajo la curva a la derecha del valor del estadístico de prueba $\chi^2 = 28.18$ es el valor- p de la prueba.

Como ocurre con las tablas de la distribución t , las tablas de la distribución chi-cuadrada no son suficientemente detalladas para permitir determinar con exactitud el valor- p . Sin embargo, con las tablas de la distribución chi-cuadrada se obtiene un intervalo en el que se encuentra el valor- p . Por ejemplo, usando la tabla 11.1 se encuentra la información siguiente para la distribución chi-cuadrada con 23 grados de libertad.

Área en la cola superior	0.10	0.05	0.025	0.01
Valor χ^2 (23 gl)	32.007	35.172	38.076	41.638

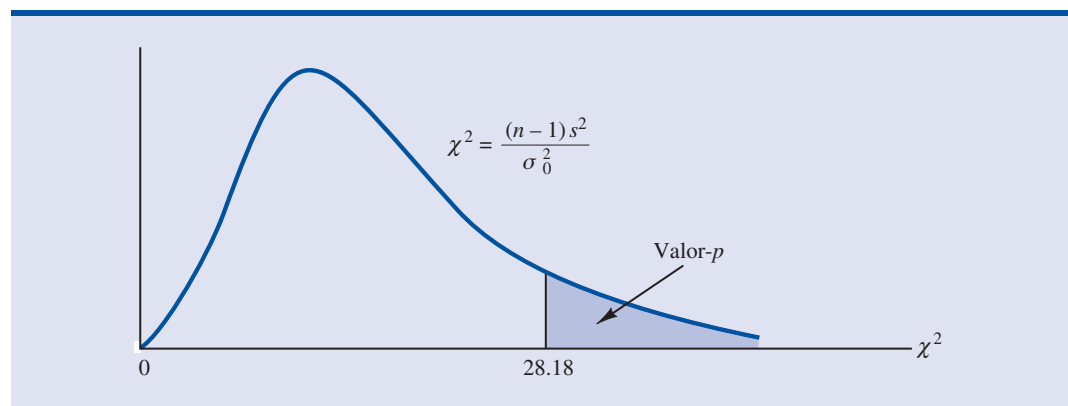
\uparrow
 $\chi^2 = 28.18$

Como $\chi^2 = 28.18$ es menor que 32.007, el área en la cola superior (el valor- p) es mayor que 0.10. Como el valor- p es $> \alpha = 0.05$, no se puede rechazar la hipótesis nula. La muestra no lleva a la conclusión de que la varianza en los tiempos de llegada a las paradas sea demasiado grande.

Dada la dificultad para determinar con exactitud el valor- p empleando las tablas de la distribución chi-cuadrada, es útil emplear un paquete de software como Minitab o Excel. En el apéndice F que se encuentra al final del libro se describe cómo calcular los valores- p . En el apéndice se muestra que el exacto valor- p correspondiente a $\chi^2 = 2.18$ es 0.2091.

Como ocurre con los otros procedimientos para pruebas de hipótesis, aquí también es posible emplear el método del valor crítico para obtener la conclusión de la prueba de hipótesis. Así,

FIGURA 11.3 DISTRIBUCIÓN CHI-CUADRADA PARA EL EJEMPLO DE ST. LOUIS METRO BUS



$\alpha = 0.05$, $\chi^2_{0.05}$ proporciona el valor crítico en la cola superior para esta prueba. Empleando la tabla 11.1 y 23 grados de libertad, $\chi^2_{0.05} = 35.172$. De manera que la regla de rechazo para los tiempos de llegada a las paradas es la siguiente:

$$\text{Rechazar } H_0 \text{ si } \chi^2 \geq 35.172$$

Como el valor del estadístico de prueba es $\chi^2 = 28.18$, no se puede rechazar la hipótesis nula.

En la práctica, las pruebas de hipótesis para la varianza poblacional que se encuentran con más frecuencia son pruebas de hipótesis de la cola superior, como la aquí presentada. En situaciones que implican tiempos de llegada, tiempos de producción, pesos de llenado, dimensiones de piezas, etc., varianzas pequeñas son deseables, mientras que varianzas grandes son inaceptables. Al establecer la varianza poblacional máxima permitida, es posible probar la hipótesis nula de que la varianza poblacional es menor o igual que el valor máximo permitido, contra la hipótesis alternativa de que la varianza poblacional es mayor que el valor máximo permitido. Con esta estructura de la prueba, deberán tomarse medidas correctivas, siempre que se rechace la hipótesis nula, lo que indica la presencia de una varianza poblacional demasiado grande.

Como ocurre con la media y la proporción poblacionales, también se realizan pruebas de hipótesis que tienen otras formas. A continuación se presenta una prueba de dos colas para la varianza poblacional considerando una situación que suele presentarse en las oficinas que otorgan licencias para conducir vehículos de motor en Estados Unidos. Históricamente, la varianza en las puntuaciones de los exámenes presentados por las personas que solicitan una licencia para conducir ha sido $\sigma^2 = 100$. Ahora se ha elaborado un nuevo examen con preguntas nuevas. Los administradores de dicha oficina desean que la varianza en las puntuaciones del examen permanezca en los niveles históricos. Para estudiar la varianza en las puntuaciones del nuevo examen se propone la siguiente prueba de hipótesis de dos colas.

$$H_0: \sigma^2 = 100$$

$$H_a: \sigma^2 \neq 100$$

El rechazo de H_0 indicará que la varianza ha cambiado y que será necesario revisar algunas de las preguntas del nuevo examen para que la varianza en las puntuaciones en este examen sea parecida a la varianza en las puntuaciones del examen anterior. El nuevo examen será aplicado a los integrantes de una muestra de 30 solicitantes de licencia de conducir. En esta prueba se usará como nivel de significancia $\alpha = 0.05$.

En este caso la varianza muestral de las puntuaciones de 30 exámenes fue $s^2 = 162$. El valor del estadístico de prueba chi-cuadrada es el siguiente:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(30-1)(162)}{100} = 46.98$$

Ahora, queda calcular el valor- p . Mediante la tabla 11.1 y $n-1 = 30-1 = 29$ grados de libertad (gl) se encuentra lo siguiente.

Área en la cola superior	0.10	0.05	0.025	0.01
Valor χ^2 (29 gl)	39.087	42.557	45.722	49.588

$\chi^2 = 46.98$

De manera que el valor del estadístico de prueba $\chi^2 = 46.98$ corresponde a un área entre 0.025 y 0.01 en la cola superior de la distribución chi-cuadrada. Al duplicar este valor, se tiene que el

valor- p está entre 0.05 y 0.02. Con Excel o Minitab se encuentra el exacto valor- $p = 0.374$. Como el valor- $p \leq \alpha = 0.05$, se rechaza H_0 y se concluye que en el nuevo examen las puntuaciones presentan una varianza poblacional distinta a la varianza histórica de $\sigma^2 = 100$. Un resumen de las pruebas de hipótesis para la varianza poblacional se presenta en la tabla 11.2.

TABLA 11.2 PRUEBAS DE HIPÓTESIS PARA LA VARIANZA POBLACIONAL

	Prueba de la cola inferior	Prueba de la cola superior	Prueba de dos colas
Hipótesis	$H_0: \sigma^2 \geq \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$
Estadístico de prueba	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
Regla de rechazo: método del valor-p	Rechazar H_0 si valor- $p \leq \alpha$	Rechazar H_0 si valor- $p \leq \alpha$	Rechazar H_0 si valor- $p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $\chi^2 \leq \chi_{(1-\alpha)}^2$	Rechazar H_0 si $\chi^2 \geq \chi_{\alpha}^2$	Rechazar H_0 si $\chi^2 \leq \chi_{(1-\alpha/2)}^2$ o si $\chi^2 \geq \chi_{\alpha/2}^2$

Ejercicios

Métodos

- En la tabla 11.1 o en la tabla 3 del apéndice B encuentre los valores siguientes de la distribución chi-cuadrada.
 - $\chi_{0.05}^2, gl = 5$
 - $\chi_{0.025}^2, gl = 15$
 - $\chi_{0.975}^2, gl = 20$
 - $\chi_{0.01}^2, gl = 10$
 - $\chi_{0.95}^2, gl = 18$
- En una muestra de 20 elementos la desviación estándar muestral es 5.
 - Calcule una estimación por intervalo de confianza de 90% para la varianza poblacional.
 - Calcule una estimación por intervalo de confianza de 95% para la varianza poblacional.
 - Calcule una estimación por intervalo de confianza de 95% para la desviación estándar poblacional.
- En una muestra de 16 elementos la desviación estándar muestral es 9.5. Pruebe la hipótesis siguiente usando $\alpha = 0.05$. ¿A qué conclusión llega? Use tanto el método del valor- p como el del valor crítico.

$$H_0: \sigma^2 \leq 50$$

$$H_a: \sigma^2 > 50$$

Aplicaciones

- En la industria farmacéutica la varianza en los pesos de los medicamentos es trascendental. Considere un medicamento cuyo peso está dado en gramos y una muestra de 18 unidades de este medicamento, la varianza muestral es $s^2 = 0.36$.
 - Dé un intervalo de 90% de confianza para estimar la varianza poblacional de los pesos de este medicamento.
 - Proporcione un intervalo de 90% de confianza para estimar la desviación estándar poblacional.

5. A continuación se presentan los precios de las rentas de un automóvil por día en ocho ciudades.

Ciudad	Renta de un automóvil por día (\$)
Atlanta	47
Chicago	50
Dallas	53
New Orleans	45
Phoenix	40
Pittsburgh	43
San Francisco	39
Seattle	37

- Calcule la varianza y la desviación estándar de estos datos.
 - Dé la estimación por intervalo de confianza de 95% por día para la varianza poblacional de los precios de renta de un automóvil por día.
 - Dé la estimación por intervalo de confianza de 95% para la desviación estándar poblacional.
6. La Fidelity Growth & Income recibe fondos mutualistas de tres estrellas, o neutrales, clasificados por Mornigstar. A continuación se presentan los rendimientos porcentuales trimestrales en el periodo de cinco años que va de 2001 a 2005 (*Mornigstar Funds 500*, 2006).



	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
2001	-10.91	5.80	-9.64	6.45
2002	0.83	-10.48	-14.03	5.58
2003	-2.27	10.43	0.85	9.33
2004	1.34	1.11	-0.77	8.03
2005	-2.46	0.89	2.55	1.78

- Calcule media, varianza y desviación estándar de estos rendimientos trimestrales.
 - Los analistas financieros suelen usar la desviación estándar como una medida del riesgo de acciones y fondos mutualistas. Dé un intervalo de 95% de confianza para la desviación estándar poblacional del rendimiento trimestral de los fondos mutualistas de Fidelity Growth & Income.
7. Para analizar el riesgo o la volatilidad al invertir en las acciones comunes de Chevron Corporation se toma una muestra de rendimiento porcentual total mensual. A continuación se presentan los rendimientos de los 12 meses de 2005 (*Compustat*, 24 de febrero de 2006). El rendimiento total es el precio más cualquier dividendo pagado.

Mes	Rendimiento (%)	Mes	Rendimiento (%)
Enero	3.60	Julio	3.74
Febrero	14.86	Agosto	6.62
Marzo	-6.07	Septiembre	5.42
Abril	-10.82	Octubre	-11.83
Mayo	4.29	Noviembre	1.21
Junio	3.98	Diciembre	-0.94

- Calcule la varianza muestral y la desviación estándar muestral como medidas de la volatilidad del rendimiento mensual total de Chevron.
 - Dé un intervalo de 95% de confianza para la varianza poblacional.
 - Proporcione un intervalo de 95% de confianza para la desviación estándar poblacional.
8. Un grupo de 12 analistas de seguridad proporcionó estimaciones, para el año 2001, de las ganancias por acción de Qualcomm, Inc. (*Zacks.com*, 13 de junio de 2000). Los datos son los siguientes:

1.40 1.40 1.45 1.49 1.37 1.27 1.40 1.55 1.40 1.42 1.48 1.63

Autoexamen

- a. Calcule la varianza muestral de las estimaciones de ganancia por acción.
 - b. Calcule la desviación estándar muestral de las estimaciones de ganancia por acción.
 - c. Dé una estimación por intervalo de confianza de 95% para la varianza poblacional y para la desviación estándar poblacional.
9. Una pieza para automóviles debe fabricarse con medidas de tolerancia muy estrechas para que sea aceptada por el cliente. Las especificaciones de producción indican que la varianza máxima en la longitud de la pieza debe ser 0.0004. Suponga que en 30 piezas la varianza muestral encontrada es $s^2 = 0.0005$. Use $\alpha = 0.05$ para probar si se está violando la especificación para la varianza poblacional.
 10. La desviación estándar promedio del rendimiento anual de fondos mutualistas de acciones de capital grande es 18.2% (*The Top Mutual Funds*, AAIL, 2004). La desviación estándar muestral en una muestra de 36 fondos mutualistas Vanguard PRIMECAP es 22%. Realice una prueba de hipótesis para determinar si la desviación estándar de los fondos Vanguard es mayor que la desviación estándar promedio de los fondos mutualistas de capital grande. Con nivel de significancia 0.05, ¿cuál es la conclusión?
 11. Las tasas de interés en hipotecas para vivienda a 30 años con plazos fijos varían en Estados Unidos. En el verano de 2000, los datos de varias partes del país indicaban que la desviación estándar de las tasas de interés era 0.096 (*The Wall Street Journal*, 8 de septiembre de 2000). La varianza correspondiente sería $(0.096)^2 = 0.009216$. En un estudio realizado en 2001, las tasas de interés en préstamos a 30 años con plazo fijo en una muestra de 20 instituciones de préstamo mostraron una desviación estándar muestral de 0.114. Realice una prueba de hipótesis usando $H_0: \sigma^2 = 0.009216$ para determinar si los datos muestrales indican que la variabilidad en las tasas de interés ha cambiado. Use $V = 0.05$ y dé la conclusión.
 12. En un estudio de *Fortune* se encontró que la varianza en la cantidad de vehículos que poseen o rentan los suscriptores de la revista *Fortune* es 0.94. Suponga que en una muestra de 12 suscriptores de otra revista se encuentran los datos siguientes sobre la cantidad de vehículos que poseen o rentan dichos suscriptores: 2, 1, 2, 0, 3, 2, 2, 1, 2, 1, 0 y 1.
 - a. Calcule la varianza muestral de la cantidad de vehículos que poseen o rentan estos 12 suscriptores.
 - b. Pruebe la hipótesis $H_0: \sigma^2 = 0.94$ para determinar si la varianza del número de vehículos que poseen o rentan los suscriptores de la otra revista difiere de la propia de *Fortune*, que es $\sigma^2 = 0.94$. Con un nivel de significancia 0.05, ¿cuál es la conclusión?

11.2

Inferencias acerca de dos varianzas poblacionales

En algunas aplicaciones estadísticas interesa comparar las varianzas de las calidades de producto obtenido mediante dos métodos de producción diferentes, o las varianzas de tiempos de fabricación empleando dos métodos diferentes, o las varianzas de las temperaturas que se tienen con dos dispositivos distintos de calentamiento. Para comparar dos varianzas poblacionales, se emplean datos obtenidos de dos muestras aleatorias independientes, una de la población 1 y otra de la población 2. Para hacer las inferencias acerca de las dos varianzas poblacionales s_1^2 y s_2^2 se usan las dos varianzas muestrales σ_1^2 y σ_2^2 . Cuando las varianzas de dos poblaciones normales son iguales ($\sigma_1^2 = \sigma_2^2$), la distribución muestral de la proporción entre las dos varianzas muestrales s_1^2/s_2^2 es la siguiente.

DISTRIBUCIÓN MUESTRAL DE s_1^2/s_2^2 CUANDO $\sigma_1^2 = \sigma_2^2$

Cuando se toman muestras aleatorias simples independientes de tamaños n_1 y n_2 de dos poblaciones normales con varianzas iguales, la distribución muestral de

$$\frac{s_1^2}{s_2^2}$$

(11.9)

La distribución F se basa en muestras de dos poblaciones normales.

es una distribución F con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador; s_1^2 es la varianza muestral de la muestra aleatoria de n_1 elementos tomados de la población 1, y s_2^2 es la varianza muestral de la muestra aleatoria de n_2 elementos tomados de la población 2.

La figura 11.4 es una gráfica de la distribución F con 20 grados de libertad tanto en el numerador como en el denominador. Como se ve en esta gráfica, la distribución F no es simétrica y los valores F no pueden ser negativos. La forma de cada distribución F depende de los grados de libertad en el numerador y de los grados de libertad en el denominador.

Para denotar el valor F correspondiente a un área o probabilidad α en la cola superior de la distribución, se usa la notación F_α . Por ejemplo, como aparece en la figura 11.4, $F_{0.05}$ corresponde a un área de 0.05 en la cola superior de la distribución F con 20 grados de libertad en el numerador y 20 grados de libertad en el denominador. El valor de $F_{0.05}$ se encuentra en la tabla de la distribución F , parte de la cual se presenta en la tabla 11.3. Usando 20 grados de libertad en el numerador, 20 grados de libertad en el denominador y el renglón correspondiente a un área de 0.05 en la cola superior de la distribución, se encuentra $F_{0.05} = 2.12$. Observe que la tabla sirve para hallar valores F correspondientes a áreas de 0.10, 0.05, 0.025 y 0.01 en la cola superior. En la tabla 4 del apéndice B se encuentra una tabla más completa de la distribución F .

Ahora se verá cómo usar la distribución F para realizar una prueba de hipótesis para las varianzas de dos poblaciones. Se empieza con una prueba sobre la igualdad de las dos varianzas poblacionales. Las hipótesis son las siguientes:

$$\begin{aligned} H_0: \sigma_1^2 &= \sigma_2^2 \\ H_a: \sigma_1^2 &\neq \sigma_2^2 \end{aligned}$$

De manera tentativa se supone que las varianzas poblacionales son iguales. Si se rechaza H_0 se llega a la conclusión de que las varianzas poblacionales no son iguales.

Para realizar esta prueba de hipótesis se requieren dos muestras aleatorias independientes, una de cada población. Se calculan las dos varianzas muestrales. A la población en la que se encuentre la *mayor* varianza muestral se le considera la población 1. De manera que el tamaño de muestra n_1 y la varianza muestral s_1^2 corresponden a la población 1 y el tamaño de muestra n_2 y

FIGURA 11.4 DISTRIBUCIÓN F CON 20 GRADOS DE LIBERTAD EN EL NUMERADOR Y 20 GRADOS DE LIBERTAD EN EL DENOMINADOR

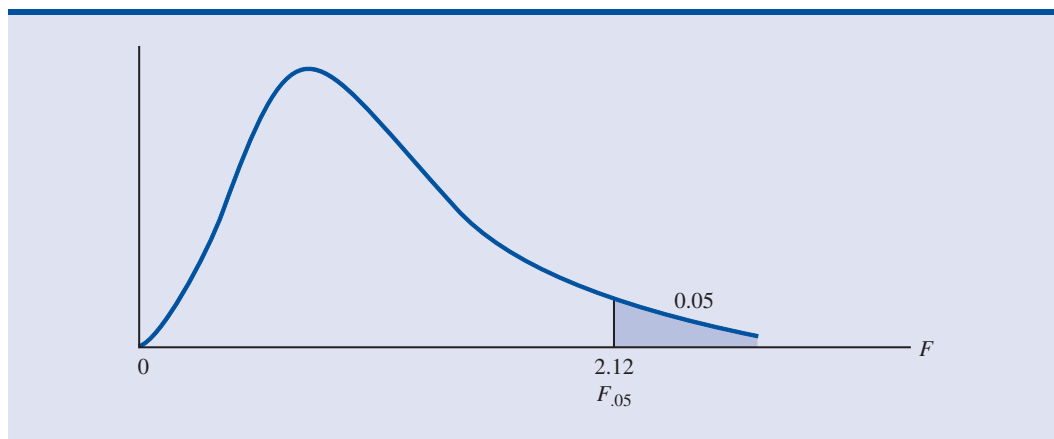
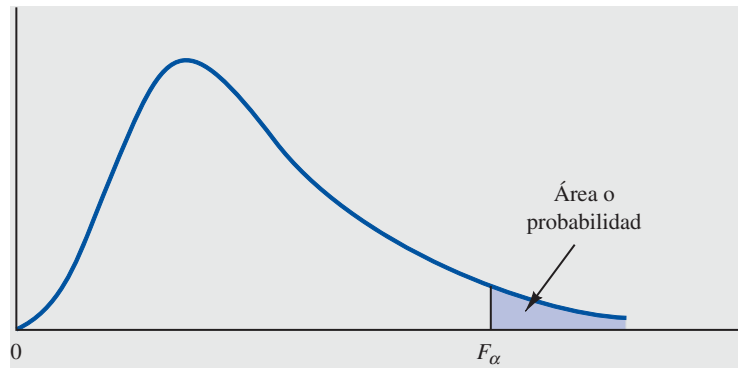


TABLA 11.3 ALGUNOS VALORES DE LA TABLA DE LA DISTRIBUCIÓN F^* 

Grados de libertad en el denominador	Área en la cola superior	Grados de libertad en el numerador				
		10	15	20	25	30
10	0.10	2.32	2.24	2.20	2.17	2.16
	0.05	2.98	2.85	2.77	2.73	2.70
	0.025	3.72	3.52	3.42	3.35	3.31
	0.01	4.85	4.56	4.41	4.31	4.25
15	0.10	2.06	1.97	1.92	1.89	1.87
	0.05	2.54	2.40	2.33	2.28	2.25
	0.025	3.06	2.86	2.76	2.69	2.64
	0.01	3.80	3.52	3.37	3.28	3.21
20	0.10	1.94	1.84	1.79	1.76	1.74
	0.05	2.35	2.20	2.12	2.07	2.04
	0.025	2.77	2.57	2.46	2.40	2.35
	0.01	3.37	3.09	2.94	2.84	2.78
25	0.10	1.87	1.77	1.72	1.68	1.66
	0.05	2.24	2.09	2.01	1.96	1.92
	0.025	2.61	2.41	2.30	2.23	2.18
	0.01	3.13	2.85	2.70	2.60	2.54
30	0.10	1.82	1.72	1.67	1.63	1.61
	0.05	2.16	2.01	1.93	1.88	1.84
	0.025	2.51	2.31	2.20	2.12	2.07
	0.01	2.98	2.70	2.55	2.45	2.39

**Nota:* La tabla 4 del apéndice B es una tabla más completa.

la varianza muestral s_2^2 corresponden a la población 2. Con base en la suposición de que las dos poblaciones tengan una distribución normal, la relación entre las varianzas muestrales proporciona el siguiente estadístico de prueba F .

ESTADÍSTICO DE PRUEBA PARA PRUEBAS DE HIPÓTESIS ACERCA DE
VARIANZAS POBLACIONALES CON $\sigma_1^2 = \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2} \quad (11.10)$$

Al denotar como población 1 a la población que tiene mayor varianza muestral, el estadístico de prueba tiene una distribución F con $n_1 - 1$ grados de libertad en el numerador y con $n_2 - 1$ grados de libertad en el denominador.

Como el estadístico de prueba F se construye con la varianza muestral más grande s_1^2 en el numerador, el valor del estadístico de prueba se encontrará siempre en la cola superior de la distribución F . Por tanto, las tablas de la distribución F , como la presentada en la tabla 11.3 y en la tabla 4 del apéndice B, únicamente necesitan proporcionar áreas o probabilidades en la cola superior. Si no se construyera de este modo el estadístico de prueba, serían necesarias áreas o probabilidades en la cola inferior. En tal caso se necesitarían más cálculos o tablas más extensas para la distribución F . A continuación se presenta un ejemplo de una prueba de hipótesis para la igualdad de dos varianzas poblacionales.

Dullus County Schools está por renovar el contrato del servicio de autobús para el año entrante y debe decidirse entre dos empresas que prestan el servicio, la empresa Milbank y la empresa Gulf Park. Como medida de la calidad del servicio se emplea la varianza en los tiempos en que llega a recoger/dejar a las personas. Poca varianza indica un mejor servicio, un servicio de mayor calidad. Si las varianzas de las dos empresas son iguales, la escuela elegirá la empresa que ofrezca mejores condiciones financieras. Pero si hay una diferencia significativa en las varianzas, la escuela preferirá la empresa con la menor varianza o mejor servicio. Las hipótesis en este caso son las siguientes:

$$\begin{aligned} H_0: \sigma_1^2 &= \sigma_2^2 \\ H_a: \sigma_1^2 &\neq \sigma_2^2 \end{aligned}$$

Si se rechaza H_0 , se concluirá que los servicios no tienen la misma calidad. Para realizar esta prueba de hipótesis se usa como nivel de significancia $\alpha = 0.10$.

En una muestra de 26 tiempos de llegada de la empresa Milbank la varianza muestral es 48 y en una muestra de 16 tiempos de llegada de la empresa Gulf Park la varianza muestral es 20. Como la varianza en la muestra de Milbank es la mayor, Milbank será la población 1. Usando la ecuación (11.10) se encuentra el valor del estadístico de prueba:

$$F = \frac{s_1^2}{s_2^2} = \frac{48}{20} = 2.40$$

La distribución F es la que tiene $n_1 - 1 = 26 - 1 = 25$ grados de libertad en el numerador y $n_2 - 1 = 16 - 1 = 15$ grados de libertad en el denominador.

Igual que en las otras pruebas de hipótesis, para llegar a una conclusión en esta prueba de hipótesis se puede emplear el método del valor- p o el método del valor crítico. En la tabla 11.3 se encuentran las siguientes áreas en la cola superior correspondientes a los valores F de una distribución F con 25 grados de libertad en el numerador y 15 grados de libertad en el denominador.

Área en la cola superior	0.10	0.05	0.025	0.01
Valor F ($gl_1 = 25, gl_2 = 15$)	1.89	2.28	2.69	3.28
			\uparrow $F = 2.40$	

Como $F = 2.40$ está entre 2.28 y 2.69, el área en la cola superior de la distribución está entre 0.05 y 0.025. Como se trata de una prueba de dos colas, se duplica el área de la cola superior, y

se obtiene un valor- p entre 0.10 y 0.05. Como se eligió $\alpha = 0.10$ como nivel de significancia, el valor- $p < \alpha = 0.10$. Por tanto, se rechaza la hipótesis nula. Esto lleva a la conclusión de que los dos servicios de autobús difieren en términos de la varianza de los tiempos en que llegan a recoger/dejar a las personas. Se le recomienda a la escuela el servicio de menor varianza o el servicio mejor que es el ofrecido por la empresa Gulf Park.

Usando Excel o Minitab se encuentra que el estadístico de prueba $F = 2.40$, corresponde a un valor- $p = 0.0811$. Como $0.0811 < \alpha = 0.10$, se rechaza la hipótesis nula de que las dos varianzas poblacionales son iguales.

Para usar el método del valor crítico en una prueba de hipótesis de dos colas con $\alpha = 0.10$, se toman los valores críticos correspondientes a un área $\alpha/2 = 0.10/2 = 0.05$ en cada cola de la distribución. Como el valor del estadístico de prueba calculado con la ecuación (11.10) está siempre en la cola superior, basta determinar el valor crítico en la cola superior. En la tabla 11.3 se encuentra que $F_{0.05} = 2.28$. Así, aun cuando se trata de una prueba de dos colas, la regla de rechazo es la siguiente:

$$\text{Rechazar } H_0 \text{ si } F \geq 2.28$$

Como el estadístico de prueba es $F = 2.40$, mayor que 2.28, se rechaza H_0 y se concluye que los dos servicios difieren en términos de la varianza en los tiempos en que llegan a recoger/dejar a las personas.

También se pueden realizar pruebas de una cola para dos varianzas poblacionales. En estos casos se usa la distribución F para determinar si una varianza poblacional es significativamente mayor que la otra. Una prueba de hipótesis para dos varianzas poblacionales se formula siempre como una prueba de la *cola superior*:

Una prueba de hipótesis de una cola para dos varianzas poblacionales siempre se formula como una prueba de la cola superior. Esto elimina la necesidad de tener valores F de la cola inferior.

$$\begin{aligned} H_0: \sigma_1^2 &\leq \sigma_2^2 \\ H_a: \sigma_1^2 &> \sigma_2^2 \end{aligned}$$

En esta forma de una prueba de hipótesis, el valor- p y el valor crítico siempre se encuentran en la cola superior de la distribución F . De esta manera, sólo se necesitan los valores F de la cola superior, lo cual simplifica tanto los cálculos como la tabla de la distribución F .

A continuación se ilustra el uso de la distribución F para realizar una prueba de una cola para las varianzas de dos poblaciones, empleando una encuesta sobre opinión pública. Para estudiar las actitudes frente a los asuntos políticos actuales se emplea una muestra de 31 hombres y otra de 41 mujeres. Al investigador que realiza el estudio le interesa saber si los datos muestrales indican que entre las mujeres hay mayor variación en las actitudes respecto de los asuntos políticos que entre los hombres. En la forma de la prueba de hipótesis dada arriba, las mujeres serán la población 1 y los hombres la población 2. La prueba de hipótesis se plantea como sigue:

$$\begin{aligned} H_0: \sigma_{\text{mujeres}}^2 &\leq \sigma_{\text{hombres}}^2 \\ H_a: \sigma_{\text{mujeres}}^2 &> \sigma_{\text{hombres}}^2 \end{aligned}$$

Rechazar H_0 dará al investigador el respaldo estadístico necesario para concluir que las mujeres muestran mayor variación en las actitudes respecto a los asuntos políticos.

Con la varianza muestral de las mujeres en el numerador y la varianza muestral de los hombres en el denominador, la distribución F tendrá $n_1 - 1 = 41 - 1 = 40$ grados de libertad en el numerador y $n_2 - 1 = 31 - 1 = 30$ grados de libertad en el denominador. En esta prueba de hipótesis se usa $\alpha = 0.05$ como nivel de significancia. Como resultado de la prueba se encontró una varianza muestral para las mujeres $s_1^2 = 120$ y una varianza muestral para los hombres $s_2^2 = 80$. El estadístico de prueba es el siguiente.

$$F = \frac{s_1^2}{s_2^2} = \frac{120}{80} = 1.50$$

TABLA 11.4 RESUMEN DE LAS PRUEBAS DE HIPÓTESIS ACERCA DE DOS VARIANZAS POBLACIONALES

Hipótesis	Prueba de la cola superior	Prueba de la cola inferior
	$H_0: \sigma_1^2 \leq \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \neq \sigma_2^2$ Nota: La población 1 tiene la varianza muestral más grande
Estadístico de prueba	$F = \frac{s_1^2}{s_2^2}$	$F = \frac{s_1^2}{s_2^2}$
Regla de rechazo: método del valor- p	Rechazar H_0 si valor- $p \leq \alpha$	Rechazar H_0 si valor- $p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $F \geq F_\alpha$	Rechazar H_0 si $F \geq F_{\alpha/2}$

En la tabla 4 del apéndice B la distribución F con 40 grados de libertad en el numerador y 30 grados de libertad en el denominador da $F_{0.10} = 1.57$. Como el estadístico de prueba, $F = 1.50$, es menor que 1.57, el área en la cola superior debe ser mayor que 0.10. Por ende, el valor- p es mayor a 0.10. Usando Excel o Minitab se encuentra que el valor- $p = 0.1256$. Como el valor- $p > \alpha = 0.05$, no se puede rechazar H_0 . Por tanto, los resultados muestrales no favorecen la conclusión de que entre las mujeres haya mayor variación en la actitud frente a los asuntos políticos que entre los hombres. En la tabla 11.4 se presenta un resumen de las pruebas de hipótesis para dos varianzas poblacionales.

NOTAS Y COMENTARIOS

Las investigaciones confirman el hecho de que para usar la distribución F es importante suponer que las poblaciones tienen una distribución normal. La distribución F no se puede usar a menos que sea

razonable suponer que ambas poblaciones tienen una distribución por lo menos aproximadamente normal.

Ejercicios

Métodos

13. En la tabla 4 del apéndice B halle los valores siguientes de la distribución F .
- a. $F_{0.05}$ con 5 y 10 grados de libertad

b. $F_{0.025}$ con 20 y 15 grados de libertad

c. $F_{0.01}$ con 8 y 12 grados de libertad

d. $F_{0.10}$ con 10 y 20 grados de libertad
14. En una muestra de 16 elementos de la población 1 la varianza muestral es $s_1^2 = 5.8$ y en una muestra de 21 elementos de la población 2 la varianza muestral es $s_2^2 = 2.4$. Pruebe las hipótesis siguientes usando 0.05 como nivel de significancia

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$H_a: \sigma_1^2 > \sigma_2^2$$

- a. Dé la conclusión a la que se llega usando el método del valor- p .
- b. Repita la prueba usando el método del valor crítico.

Autoexamen

15. Considere la prueba de hipótesis siguiente

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_a: \sigma_1^2 \neq \sigma_2^2$$

- ¿A qué conclusión se llega si $n_1 = 21$, $s_1^2 = 8.2$, $n_2 = 26$ y $s_2^2 = 4.0$? Use $\alpha = 0.05$ y el método del valor- p .
- Repita la prueba usando el método del valor crítico.

Aplicaciones

- Media Metrix and Jupiter Communications recogieron datos sobre la cantidad de tiempo que pasan conectados a Internet, por mes, adultos y jóvenes (*USA Today*, 14 de septiembre de 2000). Se concluyó que, en promedio, los adultos pasan más tiempo conectados a Internet que los jóvenes. Suponga que para confirmar esto se realiza otro estudio para el que se toma una muestra de 26 adultos y otra de 30 jóvenes. Las desviaciones estándar de las cantidades de tiempo que pasan conectados a Internet son 94 y 58 minutos, respectivamente. ¿Estos resultados muestrales favorecen la conclusión de que en el caso de los adultos la varianza del tiempo que pasan conectados a Internet es mayor que en el caso de los jóvenes? Use $\alpha = 0.01$. ¿Cuál es el valor- p ?
- La mayor parte de los individuos saben que el gasto anual medio en reparaciones de un automóvil depende de la antigüedad del automóvil. Un investigador desea saber si la varianza de los gastos anuales que se hacen en reparación también aumenta con la antigüedad del automóvil. En una muestra de 26 automóviles de 4 años de antigüedad la desviación estándar muestral en los gastos anuales en reparación fue \$170 y en una muestra de 25 automóviles de 2 años de antigüedad la desviación estándar muestral en los gastos anuales en reparación fue \$100.
 - Dé las hipótesis nula y alternativa para la investigación de que la varianza en los gastos anuales de reparación es mayor entre más viejos son los automóviles
 - Empleando 0.01 como nivel de significancia, ¿cuál es la conclusión? ¿Cuál es el valor- p ? Analice lo razonable de sus hallazgos.
- En 10 empresas de la industria aérea la desviación estándar en las ganancias a 12 meses por acción fue 4.27 y en 7 empresas de la industria automotriz la desviación estándar en las ganancias a 12 meses por acción fue 2.27 (*BusinessWeek*, 14 de agosto de 2000). Realice una prueba para varianzas iguales con $\alpha = 0.05$. ¿Cuál es la conclusión acerca de la variabilidad de las ganancias por acción en la industria aérea y en la industria automotriz?
- La varianza en un proceso de producción es un indicador importante de la calidad del proceso. Las varianzas grandes representan una oportunidad para mejorar un proceso, hallando maneras de reducir esa varianza. Realice una prueba estadística para determinar si existe una diferencia significativa entre las varianzas de los pesos de las bolsas procesadas con dos máquinas diferentes. Use 0.05 como nivel de significancia. ¿Cuál es la conclusión? ¿Representa alguna de las dos máquinas una oportunidad para mejorar la calidad?

Autoexamen



Máquina 1	2.95	3.45	3.50	3.75	3.48	3.26	3.33	3.20
	3.16	3.20	3.22	3.38	3.90	3.36	3.25	3.28
	3.20	3.22	2.98	3.45	3.70	3.34	3.18	3.35
	3.12							
Máquina 2	3.22	3.30	3.34	3.28	3.29	3.25	3.30	3.27
	3.38	3.34	3.35	3.19	3.35	3.05	3.36	3.28
	3.30	3.28	3.30	3.20	3.16	3.33		

- De acuerdo con datos obtenidos en un estudio, en las empresas de contadores públicos la varianza de los salarios anuales de los empleados de mayor antigüedad es aproximadamente 2.1 y la varianza de los salarios anuales de los gerentes es alrededor de 11.1. Estos datos están dados en miles de dólares. Si estos datos se obtuvieron de muestras de 25 empleados de mayor antigüedad y 26 gerentes, pruebe la hipótesis de que las varianzas poblacionales de estos dos salarios son iguales. Con 0.05 como nivel de significancia, ¿cuál es su conclusión?

21. Fidelity Magellan es un fondo mutualista de capital grande y Fidelity Small Cap Stock es un fondo mutualista de capital pequeño (*Morningstar Funds 500*, 2006). La desviación estándar de ambos fondos se calculó empleando muestras aleatorias de tamaño 26. La desviación estándar muestral de Fidelity Magellan fue 8.89% y la desviación estándar muestral de Fidelity Small Cap Stock fue 13.03%. Los analistas financieros suelen usar la desviación estándar como una medida del riesgo. Realice una prueba de hipótesis para determinar si los fondos de capital pequeño son más riesgosos que los fondos de capital grande. Use $\alpha = 0.05$ como nivel de significancia.
22. Una hipótesis de investigación sostiene que la varianza de las distancias de frenado de los automóviles sobre pavimento húmedo es mayor que la varianza de las distancias de frenado de los automóviles sobre pavimento seco. En un estudio a 16 automóviles que iban a una misma velocidad se les hizo frenar sobre pavimento húmedo y después sobre pavimento seco. En pavimento húmedo la desviación estándar de las distancias de frenado fue 32 pies. Sobre pavimento seco la desviación estándar es 16 pies.
 - a. Con 0.05 como nivel de significancia, ¿los datos muestrales justifican la conclusión de que en las distancias de frenado sobre pavimento húmedo la varianza es mayor que sobre pavimento seco? ¿Cuál es el valor- p ?
 - b. ¿Qué significan las conclusiones estadísticas de este estudio en términos de las recomendaciones para la seguridad al manejar?

Resumen

En este capítulo se presentaron los procedimientos estadísticos que se usan en las inferencias acerca de varianzas poblacionales. Se introdujeron dos distribuciones de probabilidad nuevas: la distribución chi-cuadrada y la distribución F . La distribución chi-cuadrada se usa en estimación por intervalos y pruebas de hipótesis para la varianza de una población normal.

Se ilustró el uso de la distribución F en pruebas de hipótesis para las varianzas de dos poblaciones normales. En particular, se mostró que si se tienen muestras aleatorias simples independientes de tamaños n_1 y n_2 , tomadas de dos poblaciones normales que tienen varianzas iguales $\sigma_1^2 = \sigma_2^2$, la distribución muestral de la razón entre las dos varianzas muestrales s_1^2/s_2^2 tiene una distribución F con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador.

Fórmulas clave

Estimación por intervalo para una varianza poblacional

$$\frac{(n-1)s^2}{\chi_{\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{(1-\alpha/2)}^2} \quad (11.7)$$

Estadístico de prueba en una prueba de hipótesis para la varianza poblacional

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (11.8)$$

Estadístico de prueba en una prueba de hipótesis para varianzas poblacionales con $\sigma_1^2 = \sigma_2^2$

$$F = \frac{s_1^2}{s_2^2} \quad (11.10)$$

Ejercicios complementarios

23. Por cuestiones de personal, los administradores de un hotel desean conocer la variabilidad en la cantidad de habitaciones ocupadas por día en una determinada temporada del año. En una muestra de 20 días la media muestral es 290 habitaciones ocupadas por día y la desviación estándar es 30 habitaciones.
 - a. Dé la estimación puntual de la varianza poblacional.
 - b. Dé una estimación por intervalo de 90% de confianza para la varianza poblacional.
 - c. Proporcione una estimación por intervalo de 90% de confianza para la desviación estándar poblacional.
24. Las ofertas públicas iniciales (OPI) de acciones suelen estar subvaluadas. La desviación estándar mide la dispersión o variación del indicador subvaluación-sobrevaluación. En una muestra de 13 OPI canadienses, que fueron después negociadas en la bolsa de cambio de Toronto, esta desviación estándar fue de 14.95. Dé una estimación por intervalo de confianza de 95% para la desviación estándar poblacional del indicador subvaluación-sobrevaluación.
25. A continuación se presentan los costos estimados de manutención por día de un ejecutivo que viaja a varias ciudades importantes. Las estimaciones comprenden una habitación individual en un hotel de cuatro estrellas, bebidas, desayuno, taxis y costos incidentales.

Ciudad	Costo de manutención por día	Ciudad	Costo de manutención por día
Bangkok	\$242.87	Madrid	\$283.56
Bogotá	260.93	Mexico City	212.00
Bombay	139.16	Milan	284.08
Cairo	194.19	Paris	436.72
Dublín	260.76	Rio de Janeiro	240.87
Frankfurt	355.36	Seoul	310.41
Hong Kong	346.32	Tel Aviv	223.73
Johannesburg	165.37	Toronto	181.25
Lima	250.08	Warsaw	238.20
Londres	326.76	Washington, D.C.	250.61



- a. Calcule la media muestral.
 - b. Calcule la desviación estándar muestral.
 - c. Calcule un intervalo de 95% de confianza para la desviación estándar poblacional.
26. La variabilidad es crucial en la fabricación de cojinetes de bolas. Una varianza grande en el tamaño de estos cojinetes ocasiona que no trabajen bien y que se desgasten rápidamente. Las normas de producción exigen una varianza máxima de 0.0001 en la medida de los cojinetes dada en pulgadas. En una muestra de 15 cojinetes, la desviación estándar muestral fue 0.014 pulgadas.
 - a. Use $\alpha = 0.10$ para determinar si la muestra indica que se ha excedido la varianza máxima indicada.
 - b. Dé una estimación por intervalo de confianza de 90% para la varianza poblacional de los cojinetes.
27. La varianza en el llenado de las cajas de cereal debe ser 0.02 o menos. En una muestra de 41 cajas de cereal la desviación estándar muestral es 0.16 onzas. Use $\alpha = 0.05$ para determinar si la varianza en el llenado de las cajas de cereal ha excedido la especificación.
28. Una empresa de transporte de carga asegura tiempos uniformes en sus entregas. En una muestra de 22 entregas la varianza muestral fue 1.5. Realice una prueba de hipótesis para determinar si se puede rechazar $H_0: \sigma^2 \leq 1$. Use $\alpha = 0.10$.
29. En una muestra de 9 días de los últimos seis meses se encontró que un dentista había tratado los siguientes números de pacientes: 22, 25, 20, 18, 15, 22, 24, 19 y 26. Si el número de pacientes atendidos por día tiene una distribución normal, ¿un análisis de estos datos muestrales permitiría

- rechazar la hipótesis de que la varianza de la cantidad de pacientes atendidos por día es 10? Use un nivel de significancia de 0.10. ¿Cuál es la conclusión?
30. La desviación estándar muestral del número de pasajeros que toman un determinado vuelo de una línea aérea es 8. Una estimación por intervalo de confianza de 95% para la desviación estándar poblacional es el que va de 5.86 pasajeros a 12.62 pasajeros.
- ¿El tamaño de la muestra usado en este análisis estadístico fue 10 o 15?
 - Suponga que la desviación estándar muestral $s = 8$ se obtuvo de una muestra de 25 vuelos. ¿Cuál sería el cambio en el intervalo de confianza para la desviación estándar poblacional? Calcule una estimación por intervalo de confianza de 95% para s con un tamaño de muestra de 25.
31. En los principales mercados de acciones diario existe un grupo de principales ganadoras en precio (acciones que registran las mayores alzas). Un día la desviación estándar del cambio porcentual en una muestra de 10 de las principales ganadoras que forman parte del NASDAQ fue 15.8. Ese mismo día la desviación estándar del cambio porcentual en una muestra de 10 de las principales ganadoras que forman parte del NYSE fue 7.9 (*USA Today*, 14 de septiembre de 2000). Realice una prueba para varianzas poblacionales iguales para saber si es posible concluir que existe diferencia en la volatilidad de las ganadoras principales de los dos grupos. Use $\alpha = 0.10$. ¿Cuál es la conclusión?
32. En los promedios de calificaciones de 352 estudiantes que terminaron un curso de contabilidad financiera, la desviación estándar es 0.940. En los promedios de calificaciones de 73 estudiantes que no aprobaron el mismo curso la desviación estándar es 0.797. ¿Estos datos indican alguna diferencia entre las varianzas de los promedios de las calificaciones de quienes terminaron el curso y de los que no lo aprobaron? Use 0.05 como nivel de significancia. *Nota:* $F_{0.025}$ con 351 y 72 grados de libertad es 1.466.
33. El departamento de contabilidad analiza la varianza de los costos unitarios semanales en los informes de dos departamentos de producción. En una muestra de 16 informes de costos de cada uno de los departamentos, las varianzas de los costos fueron 2.3 y 5.4, respectivamente. ¿La muestra es suficiente para concluir que los dos departamentos difieren en términos de la varianza en los costos unitarios? Use $\alpha = 0.10$
34. Al probar dos métodos de fabricación se da el tiempo requerido por cada uno de ellos. Use $\alpha = 0.10$ para probar la igualdad de las dos varianzas poblacionales.

	Método A	Método B
Tamaño de la muestra	$n_1 = 31$	$n_2 = 25$
Variación muestral	$s_1^2 = 25$	$s_2^2 = 12$

Caso problema Programa de capacitación para la Fuerza Aérea

En la fuerza aérea, en un curso introductorio sobre electrónica, se emplea un sistema personalizado en el que cada estudiante ve una clase grabada en una videocinta y después se le da un texto de enseñanza programada. Los estudiantes trabajan con el texto en forma independiente hasta que terminan y aprueban un examen. Aquí preocupan los diferentes ritmos en que los estudiantes realizan esta etapa de su capacitación. Algunos asimilan el texto de enseñanza programada relativamente pronto, mientras que otros necesitan mucho más tiempo. Entonces los primeros deben esperar hasta que los estudiantes más lentos estén listos y todo el grupo pueda pasar a otra etapa de la capacitación.

Se ha propuesto un sistema alternativo en el que se emplea enseñanza asistida por computadora. Este método consiste en que todos los estudiantes vean la misma clase grabada y después a cada uno se le asigne una terminal de computadora para continuar con la capacitación. La computadora guía al estudiante, quien trabaja en forma independiente, a través de esta parte de la capacitación.

Para comparar estos dos métodos, el propuesto y el actual, a los integrantes de un nuevo grupo de 122 estudiantes se les asignó en forma aleatoria uno de los métodos de capacitación. Un grupo de 61 estudiantes usó el método del texto programado y el otro grupo de 61 estudiantes usó el método de enseñanza asistida por computadora. Se registró el tiempo, en horas, que necesitó cada estudiante. Los datos que se presentan a continuación se encuentran en el archivo Training del disco compacto que se distribuye con el libro.



Horas necesarias para terminar el curso empleando el método actual										
76	76	77	74	76	74	74	77	72	78	73
78	75	80	79	72	69	79	72	70	70	81
76	78	72	82	72	73	71	70	77	78	73
79	82	65	77	79	73	76	81	69	75	75
77	79	76	78	76	76	73	77	84	74	74
69	79	66	70	74	72					

Horas necesarias para terminar el curso empleando el método de enseñanza asistida por computadora										
74	75	77	78	74	80	73	73	78	76	76
74	77	69	76	75	72	75	72	76	72	77
73	77	69	77	75	76	74	77	75	78	72
77	78	78	76	75	76	76	75	76	80	77
76	75	73	77	77	77	79	75	75	72	82
76	76	74	72	78	71					

Informe administrativo

1. Use la estadística descriptiva adecuada para resumir las horas que se necesitaron con cada método. ¿Qué semejanzas y diferencias observa entre estos datos muestrales?
2. Utilice los métodos del capítulo 10 para comentar las diferencias entre las medias poblacionales de los dos métodos.
3. Calcule la desviación estándar y la varianza de los datos obtenidos con cada método. Realice una prueba de hipótesis para la igualdad de las varianzas poblacionales en los datos obtenidos con los dos métodos. Explique sus hallazgos.
4. ¿Qué conclusión obtiene acerca de las diferencias entre los dos métodos? ¿Qué recomienda? Explique.
5. ¿Sugiere otros datos o pruebas que sean de utilidad, antes de decidir qué programa de capacitación usar?

Apéndice 11.1 Varianzas poblacionales con Minitab

Aquí se describe cómo usar Minitab para realizar una prueba de hipótesis para dos varianzas poblacionales.

Use los datos de la sección 11.2 del estudio para la elección del servicio de autobús escolar. Los tiempos correspondientes a la empresa Milbank se encuentran en la columna C1 y los tiempos correspondientes a la empresa Gulf Park en la columna C2. Para realizar la prueba de hipótesis $H_0: \sigma_1^2 = \sigma_2^2$ y $H_a: \sigma_1^2 \neq \sigma_2^2$, se sigue el procedimiento que se describe a continuación.

- Paso 1.** Seleccionar el menú **Stat**
- Paso 2.** Elegir **Basic Statistics**



Paso 3. Elegir **2-Variances****Paso 4.** Cuando aparezca el cuadro de diálogo 2-Variances:

Seleccionar **Samples in different columns**

Ingresar C1 en el cuadro **First**

Ingresar C2 en el cuadro **Second**

Clic en **OK**

En la zona con el título F-Test (normal distribution) se desplegará la información sobre la prueba, dando el estadístico de prueba $F = 2.40$ y el valor- $p = 0.81$. Con este procedimiento de Minitab se realiza una prueba de dos colas para la igualdad de las varianzas poblacionales. Por tanto, cuando se usa esta rutina de Minitab para una prueba de una cola, debe recordar que el área en una cola es la mitad del área del valor- p para dos colas, entonces será fácil calcular el valor- p para la prueba de una cola.

Apéndice 11.2 Varianzas poblacionales con Excel

Aquí se describe cómo usar Excel para realizar una prueba de hipótesis para dos varianzas poblacionales.



Use los datos de la sección 11.2 del estudio para la elección del servicio de autobús escolar. En la hoja de cálculo de Excel aparece en la celda A1 el rótulo Milbank y el rótulo Gulf Park en la celda B1. Los datos muestrales de Milbank se encuentran en las celdas A2:A27 y los datos muestrales de Gulf Park se encuentran en las celdas B2:B17. Los pasos para realizar la prueba de hipótesis $H_0: \sigma_1^2 = \sigma_2^2$ y $H_a: \sigma_1^2 \neq \sigma_2^2$ se presentan a continuación:

Paso 1. Seleccionar el menú **Herramientas****Paso 2.** Elegir **Análisis de datos****Paso 3.** Cuando aparezca el cuadro de diálogo Análisis de datos:

Elegir **Prueba F para varianza de dos muestras**

Clic en **Aceptar**

Paso 4. Cuando aparezca el cuadro de diálogo Prueba F para varianza de dos muestras:

Ingresar A1:A27 en el cuadro **Rango para la variable 1**

Ingresar B1:B17 en el cuadro **Rango para la variable 2**

Seleccionar **Rótulos**

Ingresar 0.05 en cuadro **Alfa**

(Nota: En este procedimiento Excel usa alfa como área en la cola superior.)

Seleccionar **Rango de salida** e ingresar C1 en el cuadro correspondiente

Clic en **Aceptar**

En los resultados, en la celda $P(F \leq f) = 0.0405$ se da el área en una cola correspondiente al estadístico de prueba $F = 2.40$. Por tanto, el valor- p para dos colas es $2(0.0405) = 0.081$. Si se trata de una prueba de hipótesis de una cola, el área en una cola que aparece en la celda rotulada $P(F \leq f)$ proporciona la información necesaria para determinar el valor- p de la prueba.