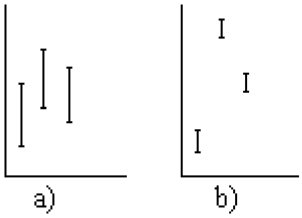
1.5. Estimar si un conjunto de muestras pertenecen a la misma población:

Objetivos:

1. Obtener un estimador insesgado de la varianza poblacional a partir de la media de las varianzas de un conjunto de muestras
2. Obtener un estimador insesgado de la varianza poblacional a partir de la varianza de las medias de un conjunto de muestras pertenecientes a la misma población.
3. Evaluar por análisis de varianza si un conjunto de muestras independientes pertenecen a la misma población.



Actividad 1.42. Si tenemos un conjunto de muestras independientes obtenidas por diferentes procedimientos, en caso de que estos procedimientos sean equivalentes la dispersión entre las muestras deberá ser proporcionada a la dispersión dentro de cada muestra (figura a). A fin de evaluarlo trabajaremos con m muestras de tamaño n y llamaremos:
X_{jk} al elemento k de la muestra j
$\overline{X}_j$ y $s(X_j)^2$ respectivamente a la media y la varianza de la muestra j
$\overline{X}$a la media de la muestra de tamaño m·n resultante de mezclar las m muestras de tamaño n.
Demostrar el
<u>Teorema 1.41:</u> $\overline{X} = \sum_j \overline{X}_j/m$.

Actividad 1.43. Llamaremos *varianza dentro de variables* a la media de las varianzas $s_w^2 = \sum_j s(X_j)^2/m$.
Supondremos que todas las muestras pertenecen a poblaciones por lo menos con la misma varianza σ^2 .
Demostrar el
<u>Teorema 1.42:</u> $\mu(s_w^2) = \sigma^2\cdot(n-1)/n$.
Llamaremos por lo tanto *varianza corregida dentro de variables* a $\hat{s}_w^2 = s_w^2\cdot n/(n-1)$, que será un estimador insesgado de la varianza poblacional σ^2 .

Actividad 1.44. Teniendo en cuenta el
<u>Teorema 1.43:</u> si las variables aleatorias independientes Y_1, Y_2 tienen distribución Ji-cuadrado con grados de libertad ν_1 y ν_2 respectivamente, entonces la variable aleatoria Y_1+Y_2 tiene distribución Ji-cuadrado con $\nu_1+\nu_2$ grados de libertad
demostrar el
<u>Teorema 1.44:</u> $mn\cdot s_w^2/\sigma^2$ tiene distribución Ji-cuadrado con $m\cdot(n-1)$ grados de libertad

Actividad 1.45. Llamaremos
$\mu_j=\mu(X_j)$ a la media de la población a la que pertenece la muestra j
$\mu=\mu(X)$ a la media de la población resultante de mezclar las poblaciones a las que pertenecen las m muestras
$\alpha_j=\mu_j-\mu$ para cada muestra j (naturalmente, valdrá 0 si todas las muestras pertenecen a la misma población).

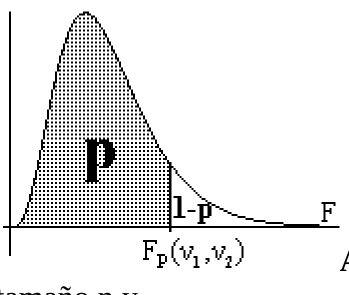
Demostrar que en cualquier caso se cumple el
Teorema 1.45: $\sum_j \alpha_j = 0$.

Actividad 1.46. Llamaremos *varianza entre variables* a la varianza de las medias
$s_b^2 = \sum_j (\bar{X}_j - \bar{X})^2/m = \sum_j \bar{X}_j^2/m - \bar{X}^2$.
Recordando de la Actividad 1.26 que $\sigma(\bar{X}_j)^2=\sigma^2/n$, $\sigma(\bar{X})^2=\sigma^2/(mn)$, demostrar
Teorema 1.46: $\mu(\bar{X}_j^2) = \sigma^2/n + (\mu+\alpha_j)^2$.
Teorema 1.47: $\mu(\bar{X}^2) = \sigma^2/(mn) + \mu^2$.
Teorema 1.48: $\mu(s_b^2) = \sigma^2(m-1)/(mn) + \sum_j \alpha_j^2/m$.

Actividad 1.47. Llamaremos varianza corregida entre variables a $\hat{s}_b^2 = s_b^2 \cdot nm/(m-1)$.
Demostrar el
Teorema 1.49: $\mu(\hat{s}_b^2) = \sigma^2 + n \cdot \sum_j \alpha_j^2/(m-1)$.
Por lo tanto, $\hat{s}_b^2$ será un estimador insesgado de la varianza poblacional σ^2 si y sólo si las poblaciones a las cuales pertenecen las diferentes muestras tienen todas la misma media (lo que llamamos *hipótesis nula* en la que todo $\alpha_j = 0$), cosa que naturalmente pasará si todas las muestras pertenecen a la misma población. En este caso, $F=\hat{s}_b^2/\hat{s}_w^2$ deberá ser próximo a la unidad. En otro caso $\mu(\hat{s}_b^2)>\sigma^2$, y por lo tanto se puede prever que F sea mayor que la unidad.

Actividad 1.48. Teniendo en cuenta que s_b^2 es la varianza de la muestra ($\bar{X}_1$, $\bar{X}_2$,... $\bar{X}_m$), de tamaño m, y que $\sigma(\bar{X}_j)^2=\sigma^2/n$, demostrar que si las muestras pertenecen a la misma población se cumple el
Teorema 1.50: $mn \cdot s_b^2/\sigma^2$ tiene distribución Ji-cuadrado con m-1 grados de libertad.

Actividad 1.49. Teniendo la cuenta el
Teorema 1.51: si las variables aleatorias independientes Y_1, Y_2 tienen distribución Ji-cuadrado con grados de libertad ν_1 y ν_2 respectivamente, entonces la distribución de $F=(Y_1/\nu_1)/(Y_2/\nu_2)$ entre 0 y ∞ es
$W_{\nu1,\nu2}(F) = K_{\nu1,\nu2} \cdot F^{\nu_1/2-1}/(1+\nu_1 \cdot F/\nu_2)^{(\nu_1+\nu_2)/2}$, que se denomina distribución *F de Snedecor* con grados de libertad ν_1 y ν_2 . $K_{\nu1,\nu2}$ se escoge de modo que $\int_0^\infty W_{\nu1,\nu2}(F)\, dF = 1$
demostrar el
Teorema 1.52: si tenemos m muestras independientes de tamaño n pertenecientes a la misma población, entonces
$F=\hat{s}_b^2/\hat{s}_w^2$ tiene una distribución F de Snedecor con grados de libertad ν_1=m-1, ν_2=m·(n-1) .



Actividad 1.50. Si tenemos m muestras independientes de tamaño n y
$F=\hat{s}_b^2/\hat{s}_w^2 > F_p(m-1 , m \cdot (n-1))$, siendo $F_p(\nu_1, \nu_2)$ el coeficiente crítico de la distribución F

de Snedecor con grados de libertad ν_1 y ν_2 tal que la probabilidad de un valor menor o igual a este coeficiente sea p, entonces podemos rechazar con un nivel de significación $\beta=1-p$ la hipótesis nula de que las muestras pertenezcan a la misma población.
Utilizaremos las tablas de la distribución F de Snedecor (inversa) para determinar el correspondiente coeficiente crítico.
Problema 1.22: anotar el número de calzado en varias muestras del mismo tamaño entre el alumnado asistente a clase y valorar si pertenecen a la misma población (a ser posible, procurar que alguna de las muestras esté formada únicamente por chicos y otra únicamente por chicas).

Actividad 1.51. ¿Cómo habríamos de interpretar el hecho que $F=\hat{s}_b^2/\hat{s}_w^2 << 1$? Aplicarlo a la resolución del siguiente
Problema 1.23: calcular el estadístico F correspondiente al siguiente par de muestras:
X_1=(24'2, 25'3, 25'4 , 26'2, 27'5)
X_2=(24'2, 25'3, 25'4 , 26'2, 27'4)
¿Se puede considerar que las muestras no cumplan alguna de las premises del Teorema 1.52?
Para valorarlo con un cierto nivel de significación podemos utilizar la relación $F_p(\nu_1, \nu_2)=1/F_{1-p}(\nu_2, \nu_1)$.

Trabajo 3 (para su realización en equipo)**:**
Comparar diferentes procedimientos para obtener algún preparado químico utilizando alguna variable aleatoria adecuada (cantidad del preparado, tiempo para su obtención, etc.). Obtener los datos de experiencias reales y utilizar un mínimo de 3 procedimientos aplicando como mínimo 5 veces cada procedimiento. Estimar por análisis de varianza si los diferentes procedimientos se pueden considerar equivalentes.