Examen de Estadística II. Módulo teórico. 07-09-2001.

1. $X \sim N(\mu, \sigma^2)$.

Para una muestra de tamaño n proveniente de una población Normal con varianza conocida se tiene que $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Utilizando este resultado se demuestra que el error de muestreo o de estimación de la media poblacional es $\epsilon = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, donde $z_{\frac{\alpha}{2}}$ es el cuantil de la distribución $N(0,1)$ que deja a su derecha una probabilidad igual a $\frac{\alpha}{2}$. Consecuentemente, sólo depende del nivel de confianza requerido y del tamaño muestral, de modo que si los dos investigadores utilizan idénticos parámetros, ambos han de proporcionar el mismo error de muestreo.

2. Para la comparación de estimadores se utiliza el error cuadrático medio del estimador cuya expresión es $E.C.M._{\hat{\theta}}(\theta) = Var(\hat{\theta}) + b^2(\theta)$, donde $b(\theta) = E(\hat{\theta}) - \theta$ es el sesgo del estimador. Debemos escoger aquel estimador de los propuestos que tenga asociado un menor error cuadrático medio. Ello nos garantizaría que es el mejor entre los comparados y por tanto sería preferible a los demás. Sin embargo, no existen garantías de que exista uno mejor que el resto pues puede ocurrir que los errores cuadráticos medios no sean comparables al depender de parámetros desconocidos.

3. X, exceso de tiempo para la realización del trabajo de rutina en minutos.
$\mu = 5, \sigma = 4$.

$\sum_{i=1}^{120} X_i$, exceso de tiempo acumulado en toda la semana en minutos.
Suponiendo equidistribución e independencia para las variables aleatorias X_i y que $n = 120$ es un tamaño muestral suficientemente grande, podemos aplicar el teorema central de límite para obtener que $\sum_{i=1}^{120} X_i \dot{\sim} N((120)(5) = 600, (120)(4)^2 = 1920)$.

4. El intervalo de confianza del 99% para la diferencia de los niveles medios de renta de las dos comunidades es $[600 - 700 = -100, 1300 = 600 + 700]$. Planteado el correspondiente contraste de igualdad de medias poblacionales, y dado que el cero pertenece al intervalo anterior, se aceptará para un nivel de significación 0.01 que las diferencias entre ambas no son significativas.

5. Para la resolución del contraste de independencia entre el tratamiento y el grado de enfermedad se tendría que seleccionar una muestra aleatoria simple de enfermos de dicha dolencia que se hallen bajo tratamiento de tamaño n suficientemente grande, a los que se mediría ambas variables. La variable tratamiento tiene $r = 3$ categorías, así como la variable grado de enfermedad con $s = 3$ categorías. Posteriormente, debería calcularse el estadístico chi-cuadrado de Pearson,

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - \frac{n_{i\cdot}\, n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}\, n_{\cdot j}}{n}} \dot{\sim} \chi^2_{(r-1)(s-1)},$$

utilizando para ello las frecuencias conjuntas y marginales. Especificando como hipótesis nula la independencia o no relación entre las dos variables, el test estadístico que resuelve el contraste es rechazar $H_0 \Leftrightarrow q \geq \chi^2_{(r-1)(s-1),\alpha}$, donde $\chi^2_{(r-1)(s-1),\alpha}$ es el cuantil de la distribución $\chi^2_{(r-1)(s-1)}$ que deja a su derecha una probabilidad igual a α.