Probabilitat i Estadística 1
Problemas Tema 5. Población y Muestra

1. Se sabe que los habitantes de un determinado país en promedio viven 76 años y que las edades de fallecimiento presentan una desviación típica de 10 años alrededor de ese valor medio. Se toma una muestra de 100 personas recientemente fallecidas en ese país y se anotan sus edades al morir.

   (a) ¿Qué valor esperado tiene la media muestral de esas 100 edades de fallecimiento?

   (b) ¿Qué desviación típica tiene la media?

   (c) ¿Cuál es la distribución aproximada de la media muestral?

   (d) Indica, aproximadamente, la probabilidad de que el valor medio de las 100 edades de fallecimiento sea inferior a 75 años.

   Solució:

   (a) El valor esperado de la media muestral es la media poblacional: 76 años.

   (b) La desviación típica de la media muestral es la poblacional dividida por $\sqrt{n}$: $10/\sqrt{100} = 1$.

   (c) Es aproximadamente normal con los parámetros antes calculados: $N(76, 1)$.

   (d) Usaremos las propiedades de las distribuciones normales. Sea Z una variable aleatoria con distribución $N(0, 1)$. Entonces,

   $$P(\bar{X} < 75) = P\left(\frac{\bar{X} - 76}{1} < \frac{75 - 76}{1}\right) \approxeq P(Z < -1) = 0.84.$$

   En la última igualdad se han usado las tablas de la distribución normal estándar.

2. Se lanza una moneda no trucada n veces. ¿Con qué probabilidad se obtendrán entre un 40% y un 60% de caras, para $n = 10$, $n = 100$ y $n = 1000$?

   Solució:
   Si $n = 10$ el 40% de n es 4 y el 60% es 6. Así que se pregunta por la probabilidad de que el número de caras sea 4, 5 o 6. Sea X número de caras obtenidas. Esta variable aleatoria sigue una distribución binomial de parámetros $n = 10$ y $p = 1/2$: $X \sim B(n, 1/2)$. Por lo tanto

   $$P(4 \leq X \leq 6) = P(X = 4) + P(X = 5) + P(X = 6) =$$

   $$\binom{10}{4}(1/2)^4(1/2)^6 + \binom{10}{5}(1/2)^5(1/2)^5 + \binom{10}{6}(1/2)^6(1/2)^4 =$$

   $$\frac{210 + 252 + 210}{2^{10}} = \frac{672}{1024} = 0.65625.$$

   Un razonamiento análogo se podría realizar con $n = 100$ y $n = 1000$, pero los cálculo con la función de probabilidad de la binomial son excesivamente laboriosos. Usaremos el hacho de que la proporción muestral es aproximadamente normal con parámetros conocidos:

   $$\hat{p} \approx N(p, \sqrt{p(1-p)/n}).$$

   Como $p = 1/2$, se tiene que $\hat{p} \approx N(1/2, (1/2)/\sqrt{n})$.
   Así, si $n = 100$ y $Z \sim N(0, 1)$,

   $$P(0.4 \leq \hat{p} \leq 0.6) = P\left(\frac{0.4 - 0.5}{0.05} \leq \frac{\hat{p} - p}{(1/2)/\sqrt{n}} \leq \frac{0.6 - 0.5}{0.05}\right) \approxeq$$

   $$P(-2 \leq Z \leq 2) = 0.955.$$

   Si $n = 1000$, entonces

   $$P(0.4 \leq \hat{p} \leq 0.6) = P\left(\frac{0.4 - 0.5}{0.016} \leq \frac{\hat{p} - p}{(1/2)/\sqrt{n}} \leq \frac{0.6 - 0.5}{0.016}\right) \approxeq$$

   $$P(-6.3 \leq Z \leq 6.3) \approxeq 1.$$

3. El número de palabras por página en los libros de una colección se distribuye según una normal de media 350 y desviación típica 30. Se toma una muestra de 25 páginas de uno de esos libros.

   (a) ¿Cuál es la distribución del número medio de palabras por página en la muestra?

   (b) ¿Con qué probabilidad será la media muestral menor de 300 palabras?

   (c) Se sabe que el número de líneas por página es exactamente 32. ¿Cuál sería un buen estimador del número medio de palabras por línea? ¿Qué distribución tiene ese estimador?

   Solució:

   (a) El número medio de palabras $\bar{X}_{25}$ es aproximadamente $N(350, 30/\sqrt{25})$, es decir, $N(350, 6)$.

   (b) Sea $Z \sim N(0, 1)$. Se tiene que

$$P(\bar{X}_{25} < 300) = P\left(\frac{\bar{X}_{25} - 350}{6} < \frac{300 - 350}{6}\right) \approxeq P(Z < -8.33) \approxeq 0.$$

   (c) Un buen estimador sería el número medio de palabras por página dividido entre 32, el número de líneas. Por las propiedades de la distribución normal se tiene que

$$\frac{\bar{X}_{25}}{32} \approx N\left(\frac{350}{32}, \frac{6}{32}\right) \Rightarrow \frac{\bar{X}_{25}}{32} \approx N(10.94,\ 0.1875).$$

4. Se sabe que el tiempo de funcionamiento de un determinado tipo de bombillas sigue una distribución de probabilidad con desviación típica de 200 horas.

   (a) Para una muestra de 36 bombillas, ¿con qué probabilidad excederá la media muestral a la media poblacional en más de 50 horas?

   (b) Para una muestra de 36 bombillas, ¿con qué probabilidad distará más de 80 horas la media muestral de la media poblacional?

   Solució:

   (a) Sabemos que la media muestral $\bar{X}_n$ es aproximadamente $N(m, 200/\sqrt{n})$. Así que si $n = 36$,

$$P(\bar{X}_n > m + 50) = P\left(\frac{\bar{X}_n - m}{200/6} > \frac{50}{200/6}\right) \approxeq P(Z > 1.5) = 0.06681.$$

   (b)

$$P(|\bar{X}_n - m| > 80) = 1 - P(-80 < \bar{X}_n - m < 80) \approxeq$$

$$1 - P\left(\frac{-80}{200/6} < Z < \frac{80}{200/6}\right) = 1 - P\left(-2.4 < Z < 2.4\right) = 1 - 0.9836 = 0.0164$$

5. El gasto de los clientes del restaurante *Comidas y cenas* sigue una distribución normal de media 18 euros y una desviación típica de 3 euros. Por su parte, los clientes del restaurante *Cenas y comidas* tienen un gasto normal de media 22 euros y desviación típica 5 euros. Se toman al azar 10 facturas de cada restaurante.

   (a) ¿Cuál es la distribución de la diferencia entre los importes medios de las facturas de uno y otro restaurantes?

   (b) ¿Con qué probabilidad la diferencia será, en valor absoluto, mayor que 5 euros?

   (c) ¿Con qué probabilidad el valor absoluto de la diferencia será inferior a 1 euro?

   Solució:

   (a) $X \sim N(18, 3)$, $Y \sim N(22, 5)$, $n = 10$, así que

$$\bar{X}_n - \bar{Y}_n \approx N\left(-4, \sqrt{\frac{9}{10} + \frac{25}{10}}\right) \Rightarrow$$

$$\bar{X}_n - \bar{Y}_n \approx N(-4,\ 1.844).$$

(b)

$$P(|\bar{X}_n - \bar{Y}_n| > 5) = 1 - P(-5 \leq \bar{X}_n - \bar{Y}_n \leq 5) \approxeq$$
$$1 - P\left(\frac{-5+4}{1.844} \leq Z \leq \frac{5+4}{1.844}\right) = 1 - P(-0.54 \leq Z \leq 4.88) = 0.7054.$$

(c)

$$P(|\bar{X}_n - \bar{Y}_n| < 1) = P(-1 \leq \bar{X}_n - \bar{Y}_n \leq 1) \approxeq$$
$$P\left(\frac{-1+4}{1.844} \leq Z \leq \frac{1+4}{1.844}\right) = P(5.532 \leq Z \leq 9.22) \approxeq 0.$$

6. Un estudio médico investiga la relación entre el estrés y la aparición de tumores. Se toma una muestra de 50 pacientes aquejados de tumores y se les pasa un cuestionario que pretende dilucidar si han estado sometidos o no a situaciones estresantes de forma continuada. Simultáneamente se toma una muestra de 100 individuos sanos a los que se les pasa el mismo cuestionario. Supongamos que realmente no hay relación entre el estrés y el desarrollo de la enfermedad, es decir, que la proporción de personas sometidas al estrés es la misma entre el grupo de enfermos que entre los sanos. Supongamos además que esta proporción es $p = 0.4$.

   (a) ¿Cuál es el valor esperado de la diferencia de proporciones de personas estresadas en una y otra muestra?

   (b) ¿Cuál es la desviación típica de la diferencia de la proporciones?

   (c) ¿Con qué probabilidad, aproximadamente, la proporción de personas estresadas entre los enfermos superará en 0.1 a la proporción entre los sanos?

   Solució:
   Definimos la variable aleatoria X que vale 1 si una persona enferma estuvo sometida a estrés y vale 0 si no lo estuvo. Así, el valor medio de X es p y su desviación típica es $\sigma_x = \sqrt{p(1-p)}$. Cuando tomamos una muestra de $n = 50$ personas enfermas y anotamos el valor de X en cada una, la media muestral $\bar{X}_n$ es la proporción muestral de enfermos que sufrieron estrés. Del mismo modo se puede definir Y para los individuos sanos. Su media es p y su desviación típica es $\sigma_y = \sqrt{p(1-p)}$. Llamaremos $\bar{Y}_m$ a la proporción muestral de los que sufrieron estrés en la muestra de tamaño $m = 100$ de individuos sanos.

   (a) El valor esperado de la diferencia de proporciones de personas estresadas en una y otra muestra es la diferencia de los valores medios de $\bar{X}_n$ y $\bar{Y}_m$: $p - p = 0$.

   (b) Será

   $$\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} = \sqrt{2p(1-p)\left(\frac{1}{n} + \frac{1}{m}\right)} = 0.0693.$$

   (c) La distribución de la diferencia de proporciones muestrales es, al fin y al cabo, un diferencia de dos medias muestrales independientes (por serlo las dos muestras) por lo que su distribución será aproximadamente normal (recordar que la suma de normales independientes es normal). Así,

   $$P(\hat{p}_x > \hat{p}_y + 0.1) = P\left(\frac{\hat{p}_x - \hat{p}_y}{0.0693} > \frac{0.1}{0.0693}\right) \approxeq P(Z > 1.4434) = 0.0745.$$

7. En una empresa trabajan 2000 hombres y 500 mujeres. El encargado de velar por la igualdad de oportunidades decide tomar una muestra estratificada de 200 hombres y 200 mujeres. A cada integrante de la muestra se le pregunta: "¿Crees que en esta empresa, en general, las mujeres ganan menos que los hombres con igual cualificación?" Responden afirmativamente 180 de las 200 mujeres y 60 de los 200 hombres.

   (a) ¿Qué probabilidad tiene cada mujer de formar parte de la muestra? ¿Y cada hombre?

   (b) Dado que de las 400 personas de la muestra, 240 respondieron afirmativamente, el encargado concluyó que "según los resultados de la muestra, el 60% de los trabajadores de esta empresa considera que las mujeres están discriminadas salarialmente". ¿Te parece correcta esta conclusión? ¿Por qué?

   Solució:

(a) La probabilidad de cada mujer de ser encuestada es

$$\frac{200}{500} = 0.4,$$

mientras que los hombres tienen una probabilidad de ser encuestados de

$$\frac{200}{2000} = 0.1.$$

(b) La conclusión no es correcta porque el muestreo que se hizo fue estratificado (había dos estratos: hombres y mujeres) y la conclusión que hace el encargado corresponde a la que se haría de un muestreo aleatorio simple.

Una forma de estimar la proporción de trabajadores que considera que las mujeres están discriminadas salarialmente acorde con el muestreo estratificado es la siguiente. Estimemos primero por separado las proporciones de mujeres y de hombres que piensan así:

$$\hat{p}_M = \frac{180}{200} = 0.9;\ \hat{p}_H = \frac{60}{200} = 0.3.$$

Ahora se combinan ambas proporciones de acuerdo a los pesos w_M y w_H que cada estrato tiene en la población:

$$w_M = \frac{500}{2500} = 0.2;\ w_H = \frac{2000}{2500} = 0.8,$$

$$\hat{p} = w_M\hat{p}_M + w_H\hat{p}_H = 0.2 \cdot 0.9 + 0.8 \cdot 0.3 = 0.42.$$