INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Curso: IN3401 – Estadística para le Economía y Gestión
Profesores: Marcelo Henríquez - Sebastián Maldonado

Pauta Control Nº 1

Una empresa productora de manzanas presenta la siguiente problemática: La producción se clasifica en cuatro categorías de calidad: 1 (exportación), 2 (Premium nacional), 3 (regular nacional), 4 (baja nacional). Las personas que inspeccionan las manzanas para su clasificación realizan adecuadamente su trabajo, sin embargo, se equivocan una fracción *p* de las veces. La clasificación la realizan dos personas por separado, y si ambas están en desacuerdo se llama a un experto que nunca se equivoca. Se asume que la proporción real de manzanas en cada clase es ¼.

1. Las manzanas para exportación deben pasar por un proceso de selección antes de su venta, el cual funciona de la siguiente manera: El experto extranjero toma aleatoriamente 5 manzanas de un cajón seleccionado también aleatoriamente. Si el experto encuentra 2 o más manzanas que no representan esta categoría (bajo los mismos criterios que sigue la empresa), los empresarios extranjeros rechazan el cargamento y las manzanas deberán venderse a un precio menor.

   El gerente de la empresa productora, preocupado, lo contrata a Ud. para que lo ayude a medir el riesgo del cargamento.

   a) Si se asume que la probabilidad de equivocarse en cualquiera de las categorías es la misma, ¿Cuál es la probabilidad de rechazo del cargamento?

R: La probabilidad de rechazo se distribuye binomial con parámetro p_{err}=P(manzana no es de exportación | manzana clasificada como exportación).

$$\boldsymbol{P(rechazo) = \sum_{j=2}^{5} \binom{5}{j} (p_{err})^j (1-p_{err})^{5-j}}$$

Para obtener p_{err} necesitamos el Teorema de Bayes. Se definen los sucesos:

- **Ne: Manzana seleccionada no es de exportación**
- **E: Manzana seleccionada es de exportación**
- **Ce: Manzana clasificada como exportación**

$$\boldsymbol{p_{err} = P(Ne|Ce) = \frac{P(Ce\,|\,Ne)P(Ne)}{P(Ce|Ne)P(Ne) + P(Ce\,|\,E)P(E)} = \frac{\frac{p^2}{9}\cdot\frac{3}{4}}{\frac{p^2}{9}\cdot\frac{3}{4} + (1-p^2)\cdot\frac{1}{4}} = \frac{p^2}{3-2p^2}}$$

b) Con el fin de reducir el riesgo de rechazo, Ud. sugiere tomar una muestra de tamaño *n* sobre las manzanas clasificadas para exportación. Asumiendo varianza máxima, un error máximo permitido de 0,05 y para un nivel de significancia del 5%, ¿Cuál es el valor de *n*?[1]

R:

$$n = \frac{Z^2 p(1-p)}{e^2} = \frac{1.96^2 0.5(1-0.5)}{0.05^2} = 384\ manzanas$$

c) Siguiendo su consejo, el gerente junto a su experto realiza un muestreo de *n* manzanas, con *n* el número sugerido por Ud., de las cuales sólo 1 manzana resultó mal clasificada para exportación. El gerente considera que un 1% de manzanas mal clasificadas como exportación es muy riesgoso para el negocio y decidiría venderlas junto con las Premium en el mercado nacional, ¿Qué le recomienda Ud.?, fundamente su respuesta con un test de hipótesis ad hoc con un nivel de significancia del 1%, definiendo claramente sus hipótesis[2]. Entregue adicionalmente el nivel de significancia más bajo al cual se puede rechazar la hipótesis nula.

R: H_0: $\pi \geq 0.01$, H_a: $\pi < 0.01$, $p = \frac{1}{384} = 0.0026$

$$z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.0026 - 0.01}{0.0026} = -2.844$$

Como z*=-2.33, se rechaza la hipótesis nula. La proporción poblacional del error de clasificación es estadísticamente menor al 1%. El gerente puede estar tranquilo. Finalmente, el valor p es 0.0023.

---

[1] Asuma que la muestra es grande

[2] Asuma que la muestra es grande

2. El Ministerio de Desarrollo Social desea recopilar información sobre el impacto del terremoto en los microempresarios de las distintas comunas afectadas. Sabiendo que el número de microempresarios afectados es muy grande y heterogéneo en las N comunas, desea aplicar alguna técnica de muestreo para obtener esta información. Para ello responda:

   a) En este caso particular, ¿recomendaría utilizar un muestreo aleatorio simple o sugeriría un muestreo más sofisticado? ¿de qué depende esta decisión?

   R: Se recomienda utilizar una técnica de muestreo que permita por un lado manejar la heterogeneidad (obtener muestras representativas de microempresarios en cada comuna) y por otro lado reducir los altos costos que implicaría realizar un muestreo aleatorio simple en las zonas afectadas.

   b) Diseñe un método de muestreo que permita cumplir con el objetivo del Ministerio, realizando los supuestos que estime necesario y utilizando toda la información disponible. ¿Qué puede decir del método propuesto cuando es estrictamente necesario contar con representantes de cada comuna?

   R: Considerando la información disponible, se recomienda considerar las comunas como unidades muestrales y utilizar un muestreo por etapas para reducir costos. El muestreo por etapas, sin embargo, NO asegura que todas las comunas se encuentren representadas, ya que realiza un muestreo aleatorio de los conglomerados. Se sugiere modificar el método para que el muestreo sea estratificado a nivel comunas (de esta manera se garantiza que todas las comuna se encuentren representadas) y por etapas dentro de las comunas, con el fin de reducir costos.

3. Dieciséis hombres adultos se aproximan a un ascensor para tomarlo. El ascensor dice “peso máximo 1300 kg”. Sabiendo que la media y desviación estándar (poblacionales) de los hombres adultos es 77 kg y 7 kg, respectivamente, responda:

   a) Utilizando el Teorema Central del Límite, calcule la probabilidad de que el ascensor falle, es decir, que las 16 personas excedan el peso máximo.

   R: Podemos expresar el problema como un promedio muestral. Debemos encontrar la probabilidad de que la media muestral de 16 adultos jóvenes exceda 1300/16=81.25. Por el TCL la distribución se aproxima a una normal con media 77 y desviación estándar $7/\sqrt{16}$=1.75.

$$\boldsymbol{z = \frac{81.25 - 77}{1.75} = 2.43}$$

La probabilidad solicitada es 0.0078 (valor p)

b) Una de las personas sabe que pesa 100 kilos y advierte a los demás que es bastante probable que se supere el peso máximo, por lo que resulta riesgoso que se suban todos al ascensor. ¿Qué le respondería? Fundamente estadísticamente.

R: Si bien es bastante probable que una persona supere el peso promedio, es poco probable que el promedio de 16 adultos supere los 81.25 kg, incluso si se sabe uno de los pesos:

Debemos encontrar ahora la probabilidad de que la media muestral de 16 adultos jóvenes exceda 1200/15=80. Por el TCL la distribución se aproxima a una normal con media 77 y desviación estándar 7/$\sqrt{15}$=1.81.

$$\boldsymbol{z = \frac{80 - 77}{1.81} = 1.66}$$

La probabilidad solicitada es 0.0485 (valor p). Notar que esto se cumple si las personas son independientes.