Estadística Aplicada
Resumen y notas de clases
Temas del primer parcial

Ultima modificación:11 de noviembre de 2004

En la ciencia estadística se utiliza la inferencia.
Se parte de lo particular a lo general.
Se elabora una "ley" a partir de casos de la realidad.
Básicamente, "lee el problema, y elabora vos el método".
La estadística aplicada se basa en el estudio de los errores que se producen al aplicar modelos aproximados a la realidad.

En todo proceso decisorio, se pueden cometer dos (2) errores.
Analizo los costos de los errores.
Se "predice" el futuro al conocer los posibles estados futuros.

Objetivo general de la Estadística:
Definir características de un conjunto
Ese conjunto se llama "población" (todos los individuos)
Tamaño "N" de la población.

- Si N es conocido, entonces, población finita.
- Si N es muy grande, cuantioso y difícil de conocer, entonces es población infinita.

Se desea definir un modelo/ley que defina como se comporta la población en su totalidad.
Las características se llaman características poblacionales, y son únicas a esa población y se las llama parámetros ; son desconocidas.
Los parámetros de una población nunca son conocidos , ya que para conocerlos hay que conocer toda la población.
A veces uno se enfrenta a un problema y no se conoce el modelo, y aun conociendo el modelo, no se conocen los parámetros.

Características poblacionales parámetros	**Características muestrales o estimadores puntuales** (son variables aleatorias)
μ	$\bar{x}$ buen estimador si n >= 2 es variable aleatoria normal
ν^2	S^2 buen estimador con n-1 solamente es distribución gamma.
p	$\hat{p}$ buen estimador es variable aleatoria binomial
λ	L

Como no se puede analizar toda la población, se toma una muestra de referencia de la población.
Tomando una muestra, veo como se comporta, y trato de obtener los parámetros para la población.
Estas muestras son variables aleatorias, ya que cada muestra da diferentes resultados.
(Al tomar varias muestras diferentes, obtengo diferentes resultados).
Con esta muestra/s se hace una estimación de la población, sujeta a errores.
Este error se llama error muestral, y se puede medir.

Para poder estimar, los estimadores puntuales deben ser buenos.

Propiedades para que un estimador sea bueno:

Tiene que ser **insesgado**:
El sesgo es vicio o tendencia.
El estimador no debe tener vicio o ser tendencioso.
Para evitar esta problemática, la muestra debe ser tomada al azar.
Cada caso particular tiene un procedimiento adecuado (normativa/norma) para tomar la muestra. Si el procedimiento no existe, hay que desarrollarlo.
La prueba de que el estimado es insesgado, es si usted encuentra que la esperanza matemática del estimador es exactamente igual al parámetro que estima.

Tiene que ser **potente:**
Es potente si la varianza del estimador es mas chico que la varianza de la población.

Si se cumplen **ambas** propiedades, el estimador es bueno.

Compruebo que estimadores son buenos

Si $\bar{x}$ sale de muestras >= 2, es bueno. Por teorema central del limite, $\bar{x}$ responde al modelo de variable aleatoria normal.

S^2 es buen estimador, si se hace con n-1. Su comportamiento: jamas da valores negativos, es una distribución gamma.

$\hat{p}=\frac{r}{n}$ Aca r es la variable aleatoria (oscila entre 0 y n, y tiene comportamiento de v.a.binomial)

Es buen estimador, es insesgado y potente.

Estimación de parámetros

Estimación puntual:
Siempre da errores, y nunca sabemos cuanto vale ese error, se usa cuando no hay mas remedio.

Estimación para máxima verosimilitud. (no se ve en este curso)

Estimación intervalo de confianza, o seguridad

Información que se dice con esto:

a <= parámetro <= b (es el nivel de seguridad)

Se dice que el parámetro esta entre a y b con un nivel de seguridad.

Elementos:
A y B: limites, cuando mas alejados están los limites, mas imprecisos son. Si uno se omite, el omitido es el valor máximo/mínimo de la variable.

$1-\alpha$ Es el nivel de seguridad o nivel de confianza ; es una **probabilidad**, la probabilidad de acertar, de que el parámetro se encuentre en esos limites, $P(A \leq parametro \leq B)=1-\alpha$

α es la probabilidad de equivocarse, de que el parámetro caiga fuera de los limites. Se lo llama **nivel de significación, o nivel de error, o riesgo.**

Notas:

Usualmente, prefiero precisión en vez de exactitud.
La precisión inexacta se puede corregir.
Cuando no se puede tener ambas, cede alguna de acuerdo a la situación.
Usualmente, prefiero precisión.

Estimación de media poblacional μ con ν **conocido**

$P(A \leq \mu \leq B) = 1-\alpha$; α es el riesgo deseado.

Tomo el mejor estimador de μ, que es $\bar{x}$

$\nu_{\bar{x}} = \frac{\nu}{\sqrt{n}}$; $Z = \frac{\bar{x}-\mu_{\bar{x}}}{\nu_{\bar{x}}}$ => $Z = \frac{\bar{x}-\mu}{\frac{\nu}{\sqrt{n}}}$ esto significa que puedo estandarizar $\bar{x}$

de aquí surge que:

$$P\left[\left(\bar{x}-Z(1-\frac{\alpha}{2})\frac{\nu}{\sqrt{n}}\right) \leq \mu \leq \left(\bar{x}+Z(1-\frac{\alpha}{2})\frac{\nu}{\sqrt{n}}\right)\right] = 1-\alpha$$

O sea que:

$A ; B = \bar{x} \pm Z(1-\frac{\alpha}{2})\frac{\nu}{\sqrt{n}}$ donde $e = Z(1-\frac{\alpha}{2})\frac{\nu}{\sqrt{n}}$ es el error muestral “e” (cuando mas chico, mas preciso soy)

Esto tiene un problema, se precisa conocer “v”

El error muestral depende de $\alpha \; y \; n$ (alfa = riesgo que se asume, n = tamaño muestra)

Para disminuir el error “e” puedo:
- Aumentar el riesgo (alfa)
- Aumentar la muestra (n)

Para sacar el tamaño óptimo de muestra para un “e” deseado:

$n = \left[\frac{Z\left(1-\frac{\alpha}{2}\right)\nu}{e}\right]^2$ (acordarse de restarle al n la muestra que ya teníamos)

Estimación de media poblacional μ con ν **desconocido**

$P(A \leq \mu \leq B) = 1-\alpha$; α es el riesgo deseado.

$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ } S^2 es variable aleatoria gamma; S (desvío muestral) es raíz de v.a gamma

$t = \frac{\frac{\bar{x}-\mu}{S}}{\sqrt{n}}$ } William Gosset, simetría centrada en 0, mas dispersa que la normal, responde a la distribución **T DE STUDENT.**

Distribución T de Student

Tiene un solo parámetro, el numero de grados de libertad del sistema. Si quiero buscar un fractil de T en la tabla, preciso el fractil y el numero de grados de libertad.
Ejemplo : T(0,95 ; 3) = 2.3534

Numero de grados de libertad del sistema

Es el numero de variables aleatorias independientes (de variacion voluntaria) que tiene un sistema matematico.

En el sistema anterior : $t=\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$ donde $S=\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$

Entonces, sin conocer el desvío poblacional (ν), hago:

$B\,;A=\bar{x}\pm\left[t\left(1-\frac{\alpha}{2};n-1\right)\cdot\frac{S}{\sqrt{n}}\right]$ (entre [ ] el error muestral) ; t (...) se busca en la tabla de T student.

Si deseo calcular n, puedo realizar: (desventajas: preciso una muestra previa para conocer S, además no conozco n, ya que preciso n-1, y estoy averiguando n.

$n=\left[\frac{t\left(1-\frac{\alpha}{2};n-1\right)S}{e}\right]^2$ e = error deseado, lo defino yo.

Es un método matemático **iterativo**. Se van probando valores de n hasta que el resultado me da el mismo n que yo puse. Se resuelve en máximo 4 pasos.
Nota: n se redondea hacia arriba, **siempre**! Ej: 6,01 = 6,98 = **7**

Atención: las expresiones anteriores son validas **solo** para poblaciones infinitas.

Población finita, N conocido

N mayúscula = tamaño población -- n minúscula = tamaño muestra

Saco una muestra: $\bar{x}=v.a.\,normal\,donde\begin{cases}\mu_{\bar{x}}=\mu\\ \nu_{\bar{x}}=\frac{\nu}{\sqrt{n}}\cdot\sqrt{\frac{N-n}{N}}\end{cases}$

Notar como se agrego el factor de corrección $\sqrt{\frac{N-n}{N}}$, otras formulas son iguales, pero agregando este factor de corrección por población finita.

$P(A\leq\mu\leq B)=1-\alpha$

$B\,;A=\bar{x}\pm\left[Z\left(1-\frac{\alpha}{2}\right)\frac{\nu}{\sqrt{n}}\cdot\sqrt{\frac{N-n}{N}}\right]$ } con v conocido. (lo que esta entre [ ] es el error muestral)

$B\,;A=\bar{x}\pm\left[t\left(1-\frac{\alpha}{2};n-1\right)\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}}\right]$ } con v desconocido.

$n=\frac{n_\infty\cdot N}{n_\infty+N}$ Donde n_∞ es en n de las expresiones anteriores, usados para poblaciones infinitas.
Usar la apropiada para v conocido o desconocido.

Ejemplo:
$n_\infty=80\,piezas$ de población infinita.
N = 78 cajón de población finita.
$n=\frac{80\cdot 78}{80+78}=40$; tomo una muestra de 40 piezas del cajón finito.

Si conozco el tamaño de población N, tengo una ventaja.

Nota: esto sirve en problemas donde, p/ej, dice “estimar el peso **total** de ...”

$$\mu=\frac{\sum_{i=1}^{n} x_i}{N} \quad \text{Y} \quad \sum_{i=1}^{N} x_i$$ es el total, y total = N * u ; solo para poblaciones finitas.

El **total** es otro parámetro, con letra griega “tao” τ

$$P(A\leq\tau\leq B)=1-\alpha$$

$$B\,;A=N\cdot\bar{x}\pm\left[Z\left(1-\frac{\alpha}{2}\right)N\frac{\nu}{\sqrt{n}}\sqrt{\frac{N\cdot n}{N}}\right]$$ si conozco v (entre [ ] el error muestral)

$$B\,;A=N\cdot\bar{x}\pm\left[t\left(1-\frac{\alpha}{2};n-1\right)N\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}}\right]$$ si no conozco v

El e muestral de tao sobre N es e_u $\frac{e_\tau}{N}=e_\mu$

Ensayos de hipótesis

Hipótesis: Afirmación de la realidad de la cual no se sabe si es V o F.

El ensayo de hipótesis es el procedimiento que se sigue para tratar de averiguar si una hipótesis planteada es V o F. Ver si la hipótesis se encuentra apoyada en la evidencia experimental (la muestra). A la hipótesis que se plantea se la llama hipótesis nula (H0). Se llama nula porque nadie va a poner duda sobre la hipótesis. La evidencia se busca para **refutar** la hipótesis, no para convalidarla. Existe una hipótesis alternativa (H1) ; es el complemento (opuesto) de H0 ; es decir, cuando busco pruebas, busco pruebas a favor de H1.

Curso de acción / Estado de la naturaleza	Rechazar H0	No rechazar H0
H0 es **verdadera**	Error de tipo 1 (ETI), α	Decisión **correcta**. $1-\alpha$ } nivel de **confianza**
H0 es **falsa**	Decisión **correcta** $1-\beta$ } **potencia** de ensayo	Error de tipo 2 (ETII) $\beta=p(ETII)$

Las hipótesis no se aceptan, se rechazan o no se rechazan.

ET I: Error al rechazar una H0 verdadera, se cuantifica con **alfa = p(ETI)**

El alfa de ETI se setea al nivel deseado, de acuerdo a cuento cuesta cometer ese error.

ET II error al no rechazar una H0 falsa. Generalmente no se puede estipular ni cuantificar a priori, solo se puede calcularlo como una probabilidad “beta” ; **beta = p de cometer error de tipo 2.**

La H0 se debe plantear de manera que el ETI sea mas grave que el ETII.

Criterios generales para plantear la hipótesis:

Dos posturas:
- pesimista : para problemas de inversión con alto riesgo económico, $, etc.
- optimista: para procesos de control periódicos, auditorías rutinarias, etc.

Pasos para la realización de un ensayo de hipótesis

- Planteo de la H0 adecuada al problemas
- Cuantificación del error de tipo 1 por intermedio de alfa (zona de rechazo)
- Fijar o calcular el tamaño de muestra adecuado.
- Sacar la muestra y calcular el estimador, a veces este paso se puede obviar.
- Estipular la condición de rechazo y la regla de decisión.
- Si el estimador pertenece a la zona de rechazo,, la hipótesis nula debe rechazarse porque se encontraron pruebas suficientes para hacerlo.
- Si el estimador no pertenece a la zona de rechazo, la H0 no puede rechazarse, porque no hay pruebas suficientes para hacerlo.

Planteos de hipótesis:

Ensayo de hipótesis con v **conocido**

Caso 1

$\left\{\begin{matrix} H_0)\ \mu \le \mu_0 \\ H_1)\ \mu > \mu_0 \end{matrix}\right\} \alpha, n$ defino alfa y n, y saco estimador $\bar{x}$

Luego, calculo un **x critico**: $\frac{\bar{x}_c - \mu_0}{\frac{\nu}{\sqrt{n}}} = Z(1-\alpha)$ => $\bar{x}_c = \mu_0 + Z(1-\alpha)\frac{\nu}{\sqrt{n}}$ <- x critico.

Condición de rechazo: si $\bar{x} > \bar{x}_c$ rechazo H0

Regla de Decisión: explicar en términos del problema que significa la condición de rechazo (explicarlo para que un operario lo entienda.)

Puedo calcular β (E.T.II) si conozco μ_1 => $\beta = \Phi\left(\frac{\bar{x}_c - \mu_1}{\frac{\nu}{\sqrt{n}}}\right)$

$Si\, \mu_0 = \mu_1 \rightarrow \beta = 1-\alpha =$ error max tipo 2
$Si\, \mu_1 = \bar{x}_c \rightarrow \beta = 0.5$

Curvas: $1-\beta$ es la **potencia** de **ensayo**. (todos los casos)

Caso 2

$\left\{\begin{matrix} H_0)\ \mu \ge \mu_0 \\ H_1)\ \mu < \mu_0 \end{matrix}\right\} \alpha, n$ => $\bar{x}_c = \mu_0 - Z(1-\alpha)\frac{\nu}{\sqrt{n}}$ => $\beta = 1 - \Phi\left(\frac{\bar{x}_c - \mu_1}{\frac{\nu}{\sqrt{n}}}\right)$

Como calcular n si no lo tengo ; sirve para caso #1 y #2 $n = \left[\frac{[Z(1-\alpha)+Z(1-\beta)]\nu}{\mu_1 - \mu_0}\right]^2$

Caso 3 – Hipótesis bilaterales

$\left\{\begin{matrix} H_0)\ \mu = \mu_0 \\ H_1)\ \mu \ne \mu_0 \end{matrix}\right\} \alpha, n$; pongo 2 limites de control, simétricos, $\bar{x}_{c_1}$, $\bar{x}_{c_2}$

$\bar{x}_{c_1} = \mu_0 - Z(1-\frac{\alpha}{2})\frac{\nu}{\sqrt{n}}$ y $\bar{x}_{c_2} = \mu_0 + Z(1-\frac{\alpha}{2})\frac{\nu}{\sqrt{n}}$

Rechazo si $\bar{x} < \bar{x}_{c_1}$ o $\bar{x} > \bar{x}_{c_2}$

Caso 3 [continua]

Beta, error de tipo 2: $\beta= \Phi\left(\frac{\bar{x}_{c_2}-\mu_1}{\frac{\nu}{\sqrt{n}}}\right) - \Phi\left(\frac{\bar{x}_{c_1}-\mu_1}{\frac{\nu}{\sqrt{n}}}\right)$

Tamaño de muestra: $n=\left[\frac{[Z(1-\frac{\alpha}{2})+Z(1-\beta)]\nu}{\mu_1-\mu_0}\right]^2$

Si el error de tipo 2 es peor que el error de tipo 1, intercambio H0 con H1, entonces

Caso 4

$\left.\begin{cases} H_0)\ \mu\neq\mu_0 \\ H_1)\ \mu=\mu_0 \end{cases}\right\}\alpha, n$ este caso no tiene solución matemática, solo filosófica.

Para aproximar la solución, puedo usar el caso 3 con alfa muy grande.

Con v **desconocido**

Caso 1

$\left.\begin{cases} H_0)\ \mu\leq\mu_0 \\ H_1)\ \mu>\mu_0 \end{cases}\right\}\alpha, n\rightarrow\bar{x}, S$ (sigue siendo para poblaciones infinitas, sino ajustar con $\sqrt{\frac{N-n}{N}}$

$\bar{x}_c=\mu_0+t(1-\alpha, n-1)\cdot\frac{S}{\sqrt{n}}$; rechazo H0 si $\bar{x}>\bar{x}_c$

No podemos calcular el tamaño de la muestra ; beta se puede calcular cuando se conoce v.

Caso 2

$\left.\begin{cases} H_0)\ \mu\geq\mu_0 \\ H_1)\ \mu<\mu_0 \end{cases}\right\}\alpha, n\rightarrow\bar{x}, S$

$\bar{x}_c=\mu_0-t(1-\alpha, n-1)\cdot\frac{S}{\sqrt{n}}$; rechazo H0 si $\bar{x}<\bar{x}_c$

Caso 3

$\left.\begin{cases} H_0)\ \mu=\mu_0 \\ H_1)\ \mu\neq\mu_0 \end{cases}\right\}\alpha, n\rightarrow\bar{x}, S$

$\bar{x}_{c_1}=\mu_0-t(1-\frac{\alpha}{2}, n-1)\cdot\frac{S}{\sqrt{n}}$ y $\bar{x}_{c_2}=\mu_0+t(1-\frac{\alpha}{2}, n-1)\cdot\frac{S}{\sqrt{n}}$

Rechazo si $\bar{x}<\bar{x}_{c_1}$ o $\bar{x}>\bar{x}_{c_2}$

Formulas útiles: $\bar{x}=\frac{\sum x_i}{n}$ y $S=\sqrt{\frac{\sum (x_i-\bar{x})^2}{n-1}}$

Como utilizar la calculadora (**mi** calculadora, Casio fx-82TL) para hallar estos valores:

[MODE] [SD] -- entrar en modo estadístico
[SHIFT] [AC] -- limpiar memoria (los datos quedan incluso con la calculadora apagada)

Ingresar los valores e ir apretando [M+] luego de cada uno, para cargarlos en memoria.

[RCL] [HYP] da la cantidad de datos ingresados (para verificar que ingresamos OK)

[SHIFT] [1] saca promedio $\bar{x}$

[SHIFT] [3] saca desvío **muestral** S (este es el que usamos en esta materia)

[SHIFT] [2] saca el desvío poblacional (no lo usamos)

$$\nu^2=varianza \rightarrow S^2=\frac{\sum (x_i-\bar{x})^2}{n-1}\rightarrow var.\ aleatoria.\ gamma$$; **(v es desvío)**

Gamma estandard (es adimensional)

$$X^2=(n-1)\frac{S^2}{\nu^2}$$ <- **chi-cuadrado o ji-cuadrado** (gamma estandarizada)

Parámetros: grados de libertad del sistema (n-1)

$$P\left[\frac{(n-1)S^2}{X^2\left(1-\frac{\alpha}{2};n-1\right)} \le \nu^2 \le \frac{(n-1)S^2}{X^2\left(\frac{\alpha}{2};n-1\right)}\right]=1-\alpha$$ esto es $P(A\le\nu^2\le B)=1-\alpha$

y así estimamos la **varianza poblacional...**

Error muestral : $e=\frac{B-A}{2}$ **Relación entre limites:** $R=\frac{B}{A}$; lo ideal es R = 1 (B=A)

A medida que la relación tiende a 1, el error tiende a 0.

Método de Roberto Garcia para hallar los grados de libertad (y sumándole + 1, hallamos n también)

$$n-1 = \frac{2}{9}\left(a+\sqrt{a^2+1}\right)^2$$ donde $a=\frac{Z\left(1-\frac{\alpha}{2}\right)(\sqrt[3]{R}+1)}{2(\sqrt[3]{R}-1)}$ | **atento!** $R=\frac{B}{A}$; si tenemos A' y B' es $R=(R')^2$

Relación de desvíos

$$P(A'\le\nu\le B')=1-\alpha \qquad A'=\sqrt{A} \qquad B'=\sqrt{B} \qquad R'=\sqrt{R}$$

Ensayos de hipótesis usando **chi cuadrado**

Caso 1

$$\left.\begin{cases} H_0)\ v^2 \le v_0^2 \\ H_1)\ v^2 > v_0^2 \end{cases}\right\} \alpha, n \rightarrow S^2$$

Cond. Rechazo: Si $\dfrac{(n-1)S^2}{v_0^2} > X^2(1-\alpha, n-1)$ rechazo H0 ; se llama condición de rechazo pivotal , y preciso conocer S^2 de antemano para poder realizarlo.

Otra manera: se puede calcular un valor critico:

$S_c^2 = \dfrac{X^2(1-\alpha, n-1)v_0^2}{n-1}$ con esto no preciso conocer S^2, sirve cuando no conozco la muestra todavia.

Rechazo H0 si $S^2 > S_c^2$

Caso 2

$$\left.\begin{cases} H_0)\ v^2 \ge v_0^2 \\ H_1)\ v^2 < v_0^2 \end{cases}\right\} \alpha, n \rightarrow S^2$$

Condición rechazo pivotal : $\dfrac{(n-1)S^2}{v_0^2} < X^2(\alpha, n-1)$

Condición de rechazo con valor critico: $S_c^2 = \dfrac{X^2(\alpha, n-1)v_0^2}{n-1}$

Rechazo H0 si $S^2 < S_c^2$

Caso 3

$$\left.\begin{cases} H_0)\ v^2 = v_0^2 \\ H_1)\ v^2 \ne v_0^2 \end{cases}\right\} \alpha, n \rightarrow S^2$$

Si no se rechaza, no se puede afirmar, este caso solo se usa cuando hay fuertes sospechas de poder rechazar H0.

Condición rechazo pivotal : $\dfrac{(n-1)S^2}{v_0^2}$ es menor que $X^2(\dfrac{\alpha}{2}, n-1)$ o mayor que $X^2(1-\dfrac{\alpha}{2}, n-1)$

Condición de rechazo con valor critico: $S_{c_1}^2 = \dfrac{X^2(\dfrac{\alpha}{2,} n-1)v_0^2}{n-1}$ y $S_{c_2}^2 = \dfrac{X^2(1-\dfrac{\alpha}{2,} n-1)v_0^2}{n-1}$

Rechazo H0 si $S^2 > S_{c_2}^2 \quad o \quad S^2 < S_{c_1}^2$

NOTA

Si en vez de con V^2 (varianza), quiero con v (desvío), se hacen unas modificaciones mínimas, se sacan los cuadrados y se ponen raíces a los S críticos.

Ejemplo: $\left.\begin{cases} H_0)\ v \le v_0 \\ H_1)\ v > v_0 \end{cases}\right\} \alpha, n \rightarrow S$ y $S_c = \sqrt{\dfrac{X^2(1-\alpha, n-1)v_0^2}{(n-1)}}$ (agregue raíz y saque el cuadrado a S_c)

Nota: Otra: $P(A' \leq \nu \leq B') = 1-\alpha \quad y \quad P(A \leq \nu^2 \leq B) = 1-\alpha$

donde $A' = \sqrt{\dfrac{(n-1)S^2}{X^2(1-\frac{\alpha}{2}, n-1)}} \qquad B' = \sqrt{\dfrac{(n-1)S^2}{X^2(\frac{\alpha}{2}, n-1)}}$

Esto sirve cuando te piden máximo y mínimo a la vez, se toma alfa/2 en la formula de A y B

Si pide mínimo o máximo, se toma alfa solo en la formula de A o B
$P(A \leq \nu^2)\, o\, P(\nu^2 \leq B)$ se toma alfa en la formula de A y B en X(1-alfa,...) o X(alfa, ...)

Relación de error muestral: R' = A' / B' ; R = (R')^2

Inferencia en los procesos de Bernoulli – Tema 4 de la guia

Al extraer elementos, la probabilidad de éxito es constante.
P = R / N ; este P es desconocido, en todas las situaciones.

Al sacar una muestra, puedo saber la proporción de éxitos de la muestra $\hat{p} = \dfrac{r}{n}$ <- variable aleatoria (población P = ?)

Como realizar un muestreo : hay 2 maneras :

- Fijo tamaño de n, los saco y cuento los éxitos “r” (variable aleatoria) (MUESTREO A LA BINOMIAL)
- Fijo “r” y veo cuantos “n” preciso (variable aleatoria) (MUESTREO A LA PASCAL, se comporta como una inversa de pascal.

Puedo entregar 0% de defectuosos, pero no puedo fabricar 0% de defectuosos.-

Ventajas de la binomial: se cuando termina ; en pascal no se cuando termina

Ventajas de pascal: sirve para buscar defectuosos, en la binomial puedo no sacar defectuosos y eso no indica que haya 0% defectuosos.
El muestreo pascal no se usa en la practica por el costo ($)
Nosotros usualmente usamos la binomial.

$\hat{p} = \dfrac{r}{n}$ Variable aleatoria binomial => $\mu\, \hat{p} = p \qquad \nu\, \hat{p} = \sqrt{\dfrac{p(1-p)}{n}}$

Datos importantes para proceso de Bernoulli

p_0 = AQL = limite de calidad aceptable

α = Riesgo del proveedor

p_1 = LTPD = porcentaje defectuosos tolerado en el lote

β = Riesgo del comprador

Estimando p

$$P(A \leq p \leq B) = 1-\alpha$$

$Gb(r/n, A) = \dfrac{\alpha}{2}$ y $Fb(r/n, B) = \dfrac{\alpha}{2}$; para solucionarlo, busco en la tabla un valor A, B que satisfaga la ecuacion, ya que r,n y alfa son conocidos.
Para hacer esta búsqueda, busco entre que valores se aproxima a alfa/2, y supongo una linealidad entre ambos valores, hago la interpolación entre ambos valores.

Recordar que en la tabla se pueden hacer complementos : $\begin{matrix} F_b(r/n, p) = 1 - G_b(r+1/n, p) \\ G_b(r/n, p) = 1 - F_b(r-1/n, p) \end{matrix}$ o sea, buscaríamos 1-alfa/2 = Gb(...) o 1-alfa/2 = Fb(...), depende de que tengamos en la tabla.

Nota: En la próxima pagina, hay un método para calcular exactamente los limites.

Calculo **exacto** de los limites

Relacion con Fisher y Snedecor, "F de Snedecor"

$F=\frac{X_N^2}{X_N^2}$ $\quad Parametros:\frac{(N\ numerador)}{(N\ denominador)}$ esos 2 parámetros son los grados de libertad de los Chi Cuadrado.

En la tabla: $F(1-\alpha, N_{numerador}, N_{denominador})$

$$A=\frac{1}{1+\left[\frac{(n-r+1)}{r}\right]\cdot F\left(1-\frac{\alpha}{2}\ ;\ 2(n-r+1)\ ;\ 2r\right)}$$

$$B=\frac{1}{1+\frac{(n-r)}{(r+1)F\left(1-\frac{\alpha}{2}\ ;\ 2r+2\ ;\ 2(n-r)\right)}}$$

Calculo de F "manual"(aproximación)

$$F(1-\alpha, N_{numerador}, N_{denominador}) \simeq \left[\frac{\left(t_1\cdot t_2+Z(1-\alpha)\sqrt{\frac{2}{N_n}\cdot(t_2)^2 + \frac{2}{N_d}\cdot(t_1)^2\cdot\frac{4Z^2(1-\alpha)}{Nn\cdot Nd}}\right)}{\left((t_2)^2-\frac{2Z^2(1-\alpha)}{N_d}\right)}\right]^{\mathbf{3}}$$

Donde $t_1=3-\frac{2}{3N_n}$ y $t_2=3-\frac{2}{3N_d}$

Criterio de Mermoz

Sirve para saber que aproximación usar : $\hat{n}_0=\frac{0{,}23\cdot(1-\hat{p})^2}{\hat{p}^3}$

Si $n>\hat{n}_0\rightarrow$usar NORMAL

Si $n<\hat{n}_0\rightarrow$usar POISSON

Calculo de limites por aproximación de Poisson

Usar Solo si p < 0,1!!!

$A=\frac{X^2\left(\frac{\alpha}{2}\ ;\ 2r\right)}{2n}$ | $B=\frac{X^2\left(1-\frac{\alpha}{2}\ ;\ 2r+2\right)}{2n}$

de aca, $R=\frac{B}{A}$ = $\frac{X^2\left(1-\frac{\alpha}{2}\ ;\ 2r+2\right)}{X^2\left(\frac{\alpha}{2}\ ;\ 2r\right)}=N=r$; lo que permite hacer muestreo de Pascal

Aproximación por distribución normal

Usar solo si: $n\cdot p>10$ **y** $n\cdot(1-p)>10$

$$z=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \quad \Rightarrow \quad A\ ;\ B\ =\ \hat{p}\ \pm\ \left[Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\right]$$ (la parte entre [ ] es el error muestral, "e")

Tamaño de muestra: $n=\left[\left(\frac{Z\left(1-\frac{\alpha}{2}\right)}{e}\right)^2(\hat{p}(1-\hat{p}))\right]+1$; notar que precisamos un muestreo previo para conocer $\hat{p}$;

podemos obviar este muestreo previo si reemplazamos $(\hat{p}(1-\hat{p}))$ por 0,25

Proceso de Bernoulli – **Casos**

Caso 1:

$$\left.\begin{matrix} H_0)\ p\leq p_0 \\ H_1)\ p>p_0 \end{matrix}\right\}\alpha, n$$

$$n\rightarrow r\}\ \hat{p}=\frac{r}{n}$$

Condición de rechazo: $si\ \hat{p}>\hat{p}_c$ o $r\geq r_c\rightarrow$ rechazo H_0

$$\hat{p}_c=\frac{r_c}{n}$$

Regla de decisión: curso de acción a seguir al rechazar H0

Como calcular los limites $\hat{p}_c$ y r_c

$G_b=(r_c/n\ \ ;p_0)=\alpha$ } incógnita: r_c ; se busca en la tabla, o se aproxima (si se puede aplicar)

Poisson

$G_{p0}(r_c/n\cdot p_0)=\alpha$ } de aca sale r_c

Normal

$$\hat{p}_c=p_0+Z(1-\alpha)\sqrt{\frac{p_0(1-p_0)}{n}}$$

Error de tipo II para hipótesis alternativa (hay un beta por cada hipótesis alternativa

$$\beta=F_b(r_c-1/n;p_1)$$

Poisson $\beta=F_{po}(rc-1/n\cdot p_1)$

Normal $\beta\ =\ \Phi\left(\frac{\hat{p}_c-p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)$

Calcular tamaño de muestra

$$n=? \quad ; \quad p_0=\alpha \quad ; \quad p_1=\beta$$

$$\left.\begin{array}{r} G_b(r_c/n \quad ; \quad p_0)=\alpha \\ F_b(r_c-1/n \quad ; \quad p_1)=\beta \end{array}\right\}$$ 2 ecuaciones con 2 incógnitas

Aproximación por Poisson

$$\left.\begin{array}{r} G_{p0}(r_c/n\cdot p_0)=\alpha \\ F_{p0}(r_c-1/n\cdot p_1)=\beta \end{array}\right\}$$ 2 ecuaciones, 2 incógnitas

Aproximación por Normal

$$n = \left[\frac{Z(1-\alpha)\sqrt{p_0(1-p_0)} + Z(1-\beta)\sqrt{p_1(1-p_1)}}{p_1-p_0}\right]^2$$ <- para verificar, si n*p > 10 => OK ; normalmente se puede usar esta siempre para calcular n, p'q' el tamaño se adapta para que la distro. aproxime la normal OK.

Caso 2:

$$\left\{\begin{array}{l} H_0)\ p\geq p_0 \\ H_1)\ p<p_0 \end{array}\right\}\alpha , n$$

$$n\rightarrow r\}\ \hat{p}=\frac{r}{n}$$

Condición de rechazo: $si\ \hat{p}<\hat{p}_c$ o $r\leq r_c\rightarrow$ rechazo H_0

$$\hat{p}_c=\frac{r_c}{n}$$ - Regla de decisión: curso de acción a seguir al rechazar H0

$$F_b(rc/n \quad ; \quad p_0)=\alpha$$ } saco rc

Poisson: $F_{p0}(rc/n\ p_0)=\alpha$

Normal: $\hat{p}_c=p_0-Z(1-\alpha)\sqrt{\frac{p_0(1-p_0)}{n}}$

Beta: $\beta\rightarrow p_1$

$$\beta=G_b(r_c+1/n \quad ; \quad p_1)$$

Poisson

$$\beta=G_{p0}(r_c+1/n\cdot p_1)$$

Normal

$$\beta=1-\Phi\left(\frac{\hat{p}_c-p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}}\right)$$

Sigue Caso 2

Averiguar n (tamaño de muestra)

$$n=? \ ; \ p_0=\alpha \ ; \ p_1=\beta$$

$$\left.\begin{array}{l} F_b(r_c/n \ ; \ p_0)=\alpha \\ G_b(r_c+1/n \ ; \ p_1)=\beta \end{array}\right\} \text{ 2 ecuaciones con 2 incógnitas}$$

Aproximación por Poisson

$$\left.\begin{array}{l} F_{p0}(r_c/n\cdot p_0)=\alpha \\ G_{p0}(r_c+1/n\cdot p_1)=\beta \end{array}\right\} \text{ 2 ecuaciones, 2 incógnitas}$$

Aproximación por Normal

IDEM a CASO 1

Caso 3

p = p_0

No tiene aplicación practica ; no se da en esta cátedra.

Notas útiles:

S = desvío muestral

$\bar{x}$ = media/promedio muestral

$\alpha \ \ valor\, estandard \ =0{,}05$

μ = media/promedio poblacional

ν = desvío muestral

Cuando usar aproximaciones ; esto también sirve para luego verificar si usamos la aproximación OK.

Si $(n\cdot p)>10 \ \ y \ \ n(1-p)>10$ usar Normal

Si $p<0{,}1$ usar Poisson

Es mejor usar el Criterio de Mermoz si se puede:

$n_0=\dfrac{0{,}23(1-p)^2}{p^3}$ si n > n0 usar NORMAL, si n < n0, usar POISSON