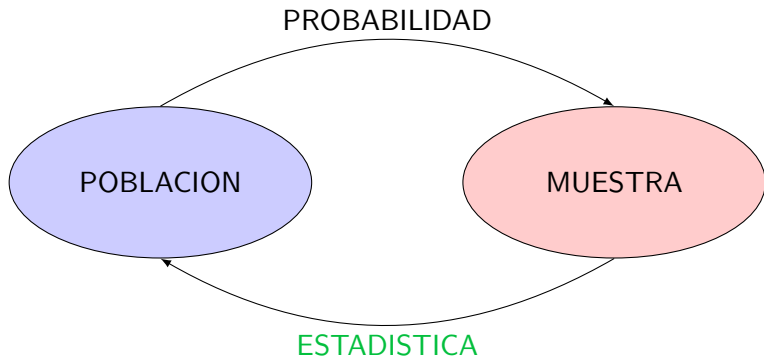
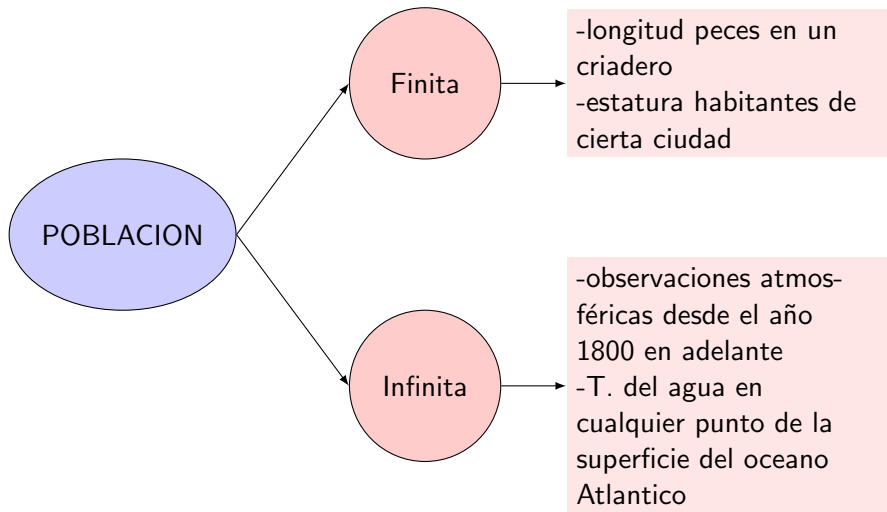


Distribuciones en el muestreo





Ejemplo: Calcular el tiempo de respuesta de un analgésico para calmar dolores musculares.

No es factible por costos/tiempos/eficacia probar el analgésico en toda la población \Rightarrow realiza un **muestreo** a fines de considerar un subconjunto de observaciones de la población.

Muestreo

- ▶ Una muestra es un subconjunto de una población
- ▶ Para que las inferencias que hagamos con ella sean válidas, debe ser aleatoria y representativa de la población bajo estudio.
- ▶ Se elige mediante un procedimiento de **muestreo**, que debe realizarse de manera muy cuidadosa respondiendo a preguntas tales como cuántos elementos? cómo? dónde?, etc.
- ▶ Distintas metodologías de muestreo (consultar bibliografía asignatura)

Continuación Ejemplo

Luego de tomar una **muestra aleatoria** de pacientes y de haber registrado el tiempo de respuesta del analgésico en cada uno de ellos, si denotamos por μ_T al valor esperado del tiempo de respuesta, podríamos arribar a conclusiones como:

- ▶ **Estimación Puntual:** $\hat{\mu}_T = 35 \text{ m}$
- ▶ **Estimación por Intervalos:** $\mu_T \in (32.3, 38.2)$
- ▶ **Test de hipótesis:** $\mu_T > 31 \text{ m}$

dependiendo del método a utilizar según lo que se quiera informar.

Distribuciones Muestrales

Definición 10.1:

Si X_1, X_2, \dots, X_n es una colección de n variables aleatorias tales que:

1. cada una de ellas tiene la misma función de probabilidad que la distribución de la población,
2. son independientes entre sí,

entonces se dice que X_1, \dots, X_n son *variables aleatorias independientes e idénticamente distribuidas* (v.a.i.i.d.)

y constituyen una **muestra aleatoria** de la población (muestra aleatoria de tamaño n).

Definiciones y notaciones

Definición 10.2

Un **estadístico** es una función de variables aleatorias que constituyen una o más muestras.

Ejemplo: \bar{X} = media muestral de la muestra X_1, X_2, \dots, X_n .
(Notar que un estadístico también es una variable aleatoria).

Definición 10.3: Distribución Muestral

La distribución de probabilidades de un estadístico se denomina **distribución muestral**

Observación

La **distribución muestral** de un estadístico depende de la distribución de la población, del tamaño de la muestra, etc.

Distribuciones muestrales

Distribuciones Muestrales

Sean X_1, X_2, \dots, X_n v.a.i.i.d. con $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $\forall i = 1, \dots, n$, $0 < \sigma^2 < \infty$.

1. $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ es la **media muestral** de la muestra aleatoria X_1, X_2, \dots, X_n .
2. $S_n = \sum_{i=1}^n X_i$ es el **total muestral** de la muestra X_1, X_2, \dots, X_n .

\bar{X}

- ▶ $\mu_{\bar{X}} = E(\bar{X}) = \mu$
- ▶ $\sigma_{\bar{X}}^2 = Var(\bar{X}) = \frac{\sigma^2}{n}$

S_n

- ▶ $\mu_{S_n} = E(S_n) = n\mu$
- ▶ $\sigma_{S_n}^2 = Var(S_n) = n\sigma^2$

Distribuciones muestrales

Sean X_1, X_2, \dots, X_n v.a.i.i.d. con $E(X_i) = \mu$, $Var(X_i) = \sigma^2$,
 $\forall i = 1, \dots, n$, $0 < \sigma^2 < \infty$.

- ▶ Si **no** conocemos la distribución de la población, **no** podemos, en general, calcular la distribución de los estadísticos contruidos a partir de X_1, X_2, \dots, X_n .
- ▶ **SI** podemos determinar la esperanza y varianza de los principales estadísticos en función de los parámetros de la distribución de la población.

Distribución de la media muestral \bar{X}

Sean X_1, X_2, \dots, X_n v.a.i.i.d. con $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $\forall i = 1, \dots, n$, $0 < \sigma^2 < \infty$.

Situaciones:

1. Si n es pequeño $\Rightarrow E(\bar{X}) = E(X)$ y $Var(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$
donde $E(X)$ y σ^2 son parámetros poblacionales.
2. Si $n > 30$ entonces podemos usar **TCL** y $\bar{X} \approx \mathcal{N}(\mu, \sigma^2/n)$
3. Si se asume que la población tiene distribución normal $\Rightarrow X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ y $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$

Ejemplo 1

El nivel de colesterol total en una población se distribuye normalmente con una media de 210 mg/dL y una varianza de 400 (mg/dL)². Se extrae una muestra aleatoria de tamaño 16 de esa población. Calcular la probabilidad que la media muestral supere a la media poblacional en por lo menos 8 mg/dL.

Continuación Ejemplo 1

Datos:

- ▶ La población se distribuye **normalmente**, esto es, $X \sim \mathcal{N}(\mu, \sigma^2)$ con $\mu = 210$, $\sigma^2 = 400$.
- ▶ Se toma una muestra X_1, \dots, X_n con $n=16$

Preguntas: Si $\bar{X} = \frac{\sum_{i=1}^{16} X_i}{16}$

1. $P(\bar{X} > \mu + 8) = ?$
2. ¿Cuál es la distribución de \bar{X} ?

Continuación Ejemplo 1

Respuestas:

2. Como la población es normal, y las X_i una muestra aleatoria de esa población, entonces $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n) = \mathcal{N}(200, 400/16)$

$$\begin{aligned} 1. P(\bar{X} > \mu + 8) &= P(\bar{X} > 210 + 8) = 1 - P(\bar{X} \leq 218) = \\ 1 - P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \leq \frac{218 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) &= 1 - \Phi(8/5) = 1 - \Phi(1.6) = 0.0548 \end{aligned}$$

Distribución de la proporción muestral

Supuestos

- ▶ Población finita, formada por elementos de dos clases.
- ▶ N = tamaño de la población, N_1 cantidad de elementos de clase 1, N_2 cantidad de elementos de clase 2, $N = N_1 + N_2$
- ▶ Se toma una muestra de tamaño n , interesa conocer cuántos elementos de la clase 1 (éxitos) hay en esa muestra
- ▶ Se define X = número de elementos de la clase 1 presentes en la muestra
- ▶ La **proporción poblacional** de éxitos es $p = \frac{N_1}{N}$ (parámetro)
- ▶ La **proporción muestral** de éxitos es $\hat{p} = \frac{X}{n}$ (estadístico)

$$P(\hat{p} = \frac{r}{n}) = P(\frac{X}{n} = \frac{r}{n}) = P(X = r)$$

Distribución de la proporción muestral

Si la extracción de la muestra se hace:

1. Con reemplazo $\Rightarrow \hat{p}$ tiene distribución Binomial, y
$$E(\hat{p}) = p = \frac{N_1}{N}, \quad Var(\hat{p}) = \frac{pq}{n}$$
2. Sin reemplazo $\Rightarrow \hat{p}$ tiene distribución hipergeométrica, y
$$E(\hat{p}) = p = \frac{N_1}{N}, \quad Var(\hat{p}) = \frac{pq}{n} \cdot \frac{N-n}{N-1}$$
3. Si $n \geq 100 \Rightarrow$ por Teo. DeMoivre-Laplace $\hat{p} \approx \mathcal{N}(\hat{p}, \frac{\hat{p}\hat{q}}{n})$

Ejemplo 2

En un grupo de 80 personas, 30 son hipertensas (el resto no). Se extrae una muestra aleatoria, sin reemplazo, de 8 personas.

1. encontrar la distribución de probabilidad de la proporción \hat{p} de hipertensos presentes en la muestra
2. calcular $P(0.5 \leq \hat{p} \leq 0.7)$
3. calcular $E(\hat{p})$ y $Var(\hat{p})$
4. Si la población hubiera sido de 800 personas, de las cuales 300 son hipertensas y tomaba una muestra de 100 personas, ¿cómo podría aproximar la distribución de \hat{p} ?

Continuación Ejemplo 2

Datos:

- ▶ $N=80$ personas, $N_1=30$ personas hipertensas, $N_2=50$ personas no hipertensas,
- ▶ Se extrae una muestra aleatoria de tamaño $n=8$, **sin** reemplazo

Sea X = número de personas hipertensas en la muestra (que se toma sin reemplazo)

$\Rightarrow X \sim h(n, N_1, N)$ y

$$P(\hat{p} = \frac{r}{n}) = P(X = r) = \frac{\binom{N_1}{r} \binom{N - N_1}{n - r}}{\binom{N}{n}} = \frac{\binom{30}{r} \binom{50}{8 - r}}{\binom{80}{8}}$$

1.

r	0	1	2	3	4	5	6	7	8
\hat{p}	$\frac{0}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{5}{8}$	$\frac{6}{8}$	$\frac{7}{8}$	$\frac{8}{8}$
$P(\hat{p} = \frac{r}{n})$	0.0185	0.1034	0.2384	0.297	0.218	0.096	0.025	0.0035	0.0002

2. $P(0.5 \leq \hat{p} \leq 0.7) = P(4 \leq X \leq 5.6) = P(X = 4) + P(X = 5) = 0.314$

3. $E(\hat{p}) = \frac{N_1}{N} = \frac{30}{80} = \frac{3}{8}$, $Var(\hat{p}) = \frac{(N_1/N) \cdot (1 - N_1/N)}{n} \cdot \frac{N - n}{N - 1} = 0.03$

4. Por De Moivre - Laplace, $\hat{p} \approx \mathcal{N}(\frac{3}{8}, 0.002)$

Distribución χ^2

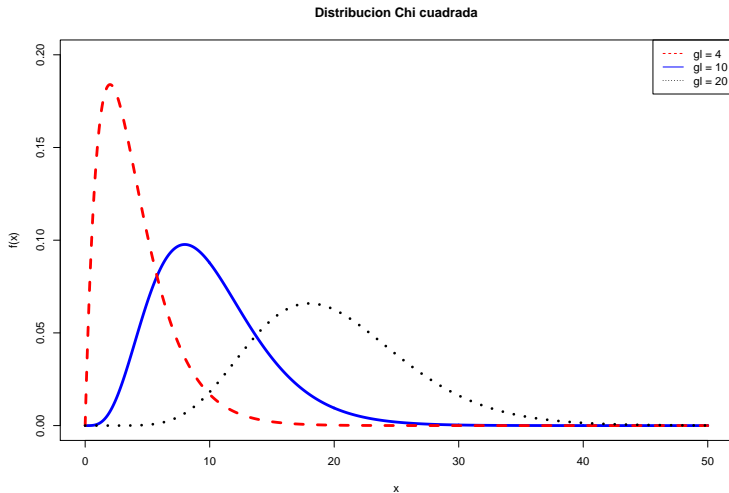
Sea X_1, \dots, X_n muestra aleatoria de una distribución normal con media μ y varianza σ^2 . Entonces $Z_i = \frac{X_i - \mu}{\sigma}$ son v.a. independientes con distribución normal estándar y la v.a definida como

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$$

tiene distribución χ^2 con n grados de libertad.

Si $Y \sim \chi_n^2 \Rightarrow E(Y) = n, \text{Var}(Y) = 2n$

Distribución χ^2



Distribución χ^2

Sea X_1, \dots, X_n muestra aleatoria de una distribución normal con media μ y varianza σ^2 . Si $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$, entonces

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)S^2}{\sigma^2}$$

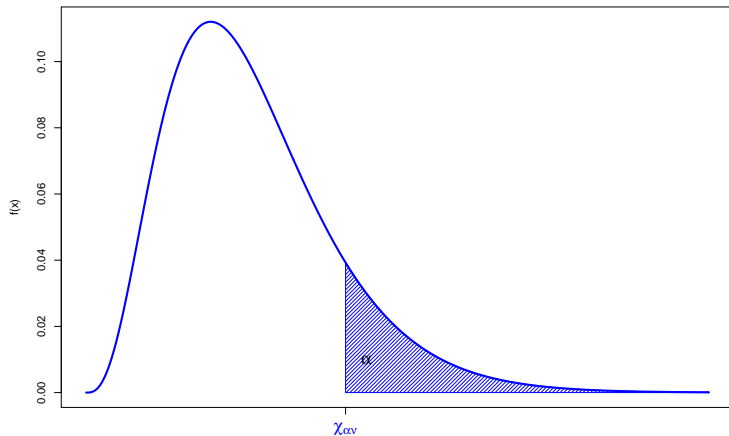
tiene distribución χ^2 con $(n-1)$ grados de libertad. Además \bar{X} y S^2 son v.a. independientes.

Cálculo de Probabilidades

Supongamos $\chi^2 \sim \chi_n^2$. Para calcular probabilidades debe recurrirse a tablas que dan valores aproximados. Fijado $0 < \alpha < 1$ se define el *valor de porcentaje* $\chi_{\alpha,n}^2$ como

$$P(\chi^2 > \chi_{\alpha,n}^2) = \alpha$$

Cálculo de Probabilidades χ^2



Uso de tabla χ^2

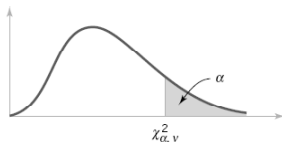


Table III Percentage Points $\chi^2_{\alpha, v}$ of the Chi-Squared Distribution

α v	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80

Uso de tabla χ^2

Ejemplos:

1. Encontrar los valores de los siguientes percentiles:
 $\chi^2_{0.05,10}$, $\chi^2_{0.025,5}$, $\chi^2_{0.95,8}$
2. Si $\chi^2 \sim \chi^2_{11}$, calcular $P(\chi^2 > 17.28)$
3. En R invocar la función `pchisq(q, df, lower.tail = FALSE)`

Distribución t

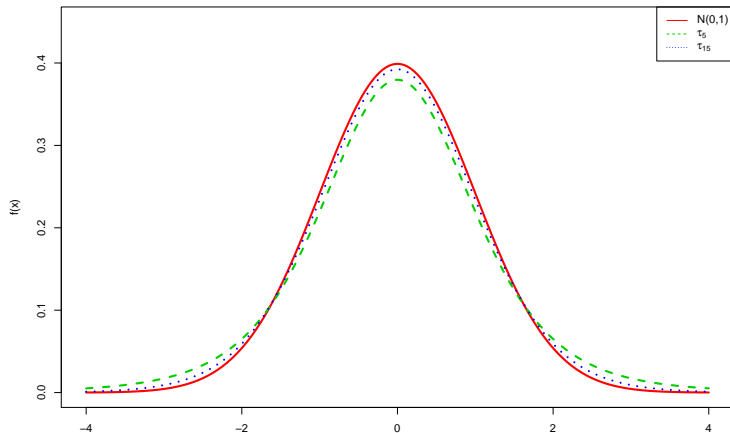
Sea Z v.a. normal estándar y sea χ^2 una v.a. chi cuadrada con ν grados de libertad. Entonces si Z y χ^2 son independientes, la v.a. definida como

$$T = \frac{Z}{\sqrt{\chi^2/\nu}}$$

tiene una **distribución t con ν grados de libertad**.

$$\mu = E(T) = 0, \text{ Var}(T) = \frac{\nu}{\nu - 2}$$

Distribución t



Distribución F de Fisher - Snedecor

Sean $Y \sim \chi_{n_1}^2$ y $W \sim \chi_{n_2}^2$ y supongamos que Y y W son independientes. Entonces se dice que la variable aleatoria definida como

$$F = \frac{Y/n_1}{W/n_2}$$

tiene **distribución F** con n_1 g.l en el numerador, n_2 g.l en el denominador. Se denota por

$$F \sim F_{n_1, n_2}$$

Distribución F

