Lista de fórmulas - Parcial 2

Estadística General - CM0244

1. Muestra Aleatoria: Sean las variables aleatorias $X_1, X_2, \cdots, X_n$, se dice que estas conforman una muestra aleatoria si:

   i) Las n variables son independientes entre si.

   ii) Cada una de estas variables tiene la misma distribución de probabilidad, $f(x)$ con media μ y varianza σ^2.

2. Distribución Chi-Cuadrado: Sean las variables $Z_1, Z_2, \cdots, Z_n$ una muestra aleatoria de la distribución $N(0,1)$. Ahora, consideremos la variable:

$$Y = \sum_{i=1}^{n} Z_i^2.$$

   Se dice que Y tiene distribución Chi-Cuadrado con n grados de libertad, denotado como $Y \sim \chi^2(n)$. Además, Y cumple las siguientes propiedades:

   i) Su función de densidad de probabilidad es:

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\, y^{\frac{n}{2}-1}\, e^{-\frac{y}{2}}, \text{ para } y > 0$$

   donde $\Gamma(\cdot)$ es la función gamma.

   ii) $\mathbb{E}(Y) = n$

   iii) $\mathbb{V}(Y) = 2n$

3. Distribución t-Student: Sean Z y Y variables aleatorias independientes entre si, tales que $Z \sim N(0,1)$ y $Y \sim \chi^2(n)$. Ahora, consideremos la variable:

$$T = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

   Se dice que T tiene distribución t-Student con n grados de libertad, denotado como $T \sim t(n)$. Además, T cumple las siguientes propiedades:

   i) Su función de densidad de probabilidad es:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \text{ para } n > 0,\, t \in \mathbb{R}$$

   ii) $\mathbb{E}(T) = 0$, para $n > 1$

   iii) $\mathbb{V}(T) = \dfrac{n}{n-2}$, para $n > 2$

4. Distribución F: Sean W_1 y W_2 variables aleatorias independientes, tales que $W_1 \sim \chi^2(n_1)$ y $W_2 \sim \chi^2(n_2)$. Ahora, consideremos la variable:

$$X = \frac{W_1/n_1}{W_2/n_2}$$

   Se dice que la variable X tiene distribución F con n_1 grados de libertad en el numerador y n_2 grados de libertad en el denominador, denotada $X \sim F(n_1, n_2)$. Además, F cumple las siguientes propiedades:

   i) Su función de densidad de probabilidad es:

$$f(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{x^{\frac{n_1-2}{2}}}{\left(1 + \frac{n_1 x}{n_2}\right)^{\frac{n_1+n_2}{2}}},$$

   para $n_1, n_2, x > 0$

   ii) $\mathbb{E}(X) = \frac{n_2}{n_2-2}$, para $n_2 > 1$

   iii) $\mathbb{V}(X) = \dfrac{2n_2^2\,(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$, para $n_2 > 4$

5. Distribuciones muestrales: Sea $X_1, X_2, \cdots X_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2. Entonces:

   i) $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

   ii) $\dfrac{(n-1)\,S^2}{\sigma^2} \sim \chi^2(n-1)$

   iii) $\sqrt{n}\left(\dfrac{\bar{X}-\mu}{S}\right) \sim t(n-1)$

   iv) $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

6. Teorema del Límite Central: Sea $X_1, X_2, \cdots X_n$ una muestra aleatoria de una distribución cualquiera con media μ y varianza σ^2. Si n es grande ($n \geq 30$), aproximadamente se cumple que:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ y } T = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

   Lo cual equivale a decir que:

$$\frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma} \sim N(0,1) \quad \text{y} \quad \frac{T-n\mu}{\sqrt{n}\sigma} \sim N(0,1)$$

7. Ley de los Grandes Números: Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria de una distribución cualquiera, con media μ. Se tiene que, cuando $n \to \infty$, $\bar{X} \longrightarrow \mu$, es decir, cuando n tiende al infinito, se tiene que $\bar{X}$ converge (probabilísticamente) a μ.

8. Propiedades del Valor Esperado y la Varianza: Sean a y b constantes; X, Y y W variables aleatorias, con X y W independientes. Entonces:

    i) $\mathbb{E}(a) = a$
    ii) $\mathbb{E}(aX + bY) = \mathbb{E}(aX) + \mathbb{E}(bY) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y)$
    iii) $\mathbb{V}(X) = \mathbb{E}\left[(X - \mu_X)^2\right] = \mathbb{E}(X^2) - \mu_X^2$
    iv) $\mathbb{E}(X^2) = \mu_X^2 + \sigma_X^2$
    v) $\mathbb{V}(a) = 0$
    vi) $\mathbb{V}(aX + bW) = \mathbb{V}(aX) + \mathbb{V}(bW) = a^2\,\mathbb{V}(X) + b^2\,\mathbb{V}(W)$

9. Propiedades de estimadores puntuales: Sea $\hat{\theta}$ un estimador de un parámetro θ. Entonces:

    i) Se dice que $\hat{\theta}$ es insesgado, si $\mathbb{E}(\hat{\theta}) = \theta$.
    ii) El sesgo del estimador se define como $b(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$.
    iii) Se dice que $\hat{\theta}$ es consistente si $\lim_{n\to\infty} \mathbb{E}(\hat{\theta}) = \theta$ y $\lim_{n\to\infty} \mathbb{V}(\hat{\theta}) = 0$.
    iv) Entre todos los estimadores insesgados de un parámetro, se prefiere el de menor varianza.
    v) El error cuadrático medio (ECM) de $\hat{\theta}$ se define como: $ECM(\hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right] = \mathbb{V}(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$. Entre múltiples estimadores de un parámetro, se prefiere el de menor ECM.

10. Intervalo de Confianza para la Media con Muestras Grandes: Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria, con n grande ($n \geq 30$), de una distribución cualquiera con media μ y varianza σ^2. Un intervalo de confianza aproximado al $100(1-\alpha)\%$ para μ es:

$$\left(\bar{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\,,\ \bar{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

    Si la varianza de la distribución es desconocida, en la fórmula anterior se puede remplazar σ por S.

11. Determinación del Tamaño de Muestra: El mínimo tamaño de muestra que garantiza que $P(|\bar{X} - \mu| \leq \varepsilon) = 1 - \alpha$, es:

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}}\,\sigma}{\varepsilon}\right)^2$$

    Si la varianza de la distribución es desconocida, en el resultado anterior se puede remplazar σ por S.

12. Intervalo de Confianza para una Proporción: Sea X una variable aleatoria, tal que $X \sim Bin(n,p)$, con n grande ($n \geq 30$) y p desconocido. Un intervalo de confianza aproximado al $100(1-\alpha)\%$ para p es:

$$\left(\hat{p} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,,\ \hat{p} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

13. Intervalos de Confianza para la Media Bajo Normalidad: Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2.

    i) Si σ^2 es conocida, un intervalo de confianza al $100(1-\alpha)\%$ para μ es:

$$\left(\bar{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\,,\ \bar{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

    ii) Si σ^2 es desconocida, un intervalo de confianza al $100(1-\alpha)\%$ para μ es:

$$\left[\bar{X} - t\left(n-1, 1-\frac{\alpha}{2}\right) \frac{S}{\sqrt{n}}\,,\ \bar{X} + t\left(n-1, 1-\frac{\alpha}{2}\right) \frac{S}{\sqrt{n}}\right]$$

14. Prueba de Hipótesis para la media con muestras grandes: Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria, con n grande ($n \geq 30$), de una distribución cualquiera con media μ y varianza σ^2. Se plantea:

    - Hipótesis nula, $H_0 : \mu = \mu_0$
    - El estadístico de prueba y su distribución aproximada bajo H_0 cierta:

$$Z_c = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Hipótesis Alternativa	Región de Rechazo	Valor p
$H_1 : \mu \neq \mu_0$	$\lvert Z_c\rvert \geq Z_{1-\frac{\alpha}{2}}$	$2P(Z > \lvert Z_c\rvert)$
$H_1 : \mu > \mu_0$	$Z_c > Z_{1-\alpha}$	$P(Z > Z_c)$
$H_1 : \mu < \mu_0$	$Z_c < -Z_{1-\alpha}$	$P(Z < Z_c)$

    Si σ es desconocida, el único cambio es en el estadístico de prueba reemplazar dicho valor por la desviación estándar muestral, S.

15. Prueba de Hipótesis para la media bajo normalidad: Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2 desconocidas. Se plantea:

    - Hipótesis nula, $H_0 : \mu = \mu_0$
    - El estadístico de prueba y su distribución bajo H_0 cierta:

$$T_c = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

Hipótesis Alternativa	Región de Rechazo	Valor p
$H_1 : \mu \neq \mu_0$	$\lvert T_c\rvert \geq t\left(n-1, 1-\frac{\alpha}{2}\right)$	$2P(t(n-1) > \lvert T_c\rvert)$
$H_1 : \mu > \mu_0$	$T_c > t(n-1, 1-\alpha)$	$P(t(n-1) > T_c)$
$H_1 : \mu < \mu_0$	$T_c < -t(n-1, 1-\alpha)$	$P(t(n-1) < T_c)$

Si σ es conocida, el procedimiento de prueba de hipótesis es igual al caso de muestras grandes.

16. Prueba de Hipótesis para una proporción: Sea X una variable aleatoria tal que $X \sim Bin(n,p)$, con p desconocida y n grande ($n \geq 30$). Se plantea:

- Hipótesis nula, $H_0 : p = p_0$
- El estadístico de prueba y su distribución aproximada bajo H_0 cierta:

$$Z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

Hipótesis Alternativa	Región de Rechazo	Valor p
$H_1 : p \neq p_0$	$\lvert Z_c\rvert \geq Z_{1-\frac{\alpha}{2}}$	$2P(Z > \lvert Z_c\rvert)$
$H_1 : p > p_0$	$Z_c > Z_{1-\alpha}$	$P(Z > Z_c)$
$H_1 : p < p_0$	$Z_c < -Z_{1-\alpha}$	$P(Z < Z_c)$