Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística

Primer Semestre 2012

Curso	:	**Probabilidad y Estadística**
Sigla	:	**EYP1113**
Profesor	:	**Ricardo Aravena (Sec. 1 y 3) y Ana María Araneda (Sec. 2 y 4)**
Ayudantes	:	**Carlos Cayuman Cofre, Felipe Ossa Monge, Juan Pablo Vigneaux Ariztía y Claudia Reyes Vizcarra**

Examen

- *Se permite el uso de calculadora científica básica.*
- *No se permite usar apuntes, correctores y cualquier aparato de transmisión electrónica (por ejemplo celulares y aparatos con bluetooth y wifi).*
- *Alumnos que escriban sus soluciones con lápiz mina renuncian a su derecho a re-corrección.*
- *El alumno que sea sorprendido copiando o en otras actividades reñidas con las normas de comportamiento académico, será calificado con nota 1.0 (uno.cero) en la interrogación y su caso será informado a la Dirección de Docencia de la Escuela de Ingeniería.*
- *En su lugar de trabajo Ud. debe tener solo lápices y sus cuadernillos.*
- *Recuerde poner su N° de lista en ambos cuadernillos.*

Problema 1

(a) **[4.0 Ptos.]** Según reporta el diario La Tercera: "Las murallas costeras del Maule y Biobío serían ineficaces ante nuevo tsunami". Esta conclusión se basa en la altura, Z, necesaria de un murallón para que éste sea eficaz ante un nuevo tsunami. Se postula que esta variable Z se relaciona con la velocidad de desplazamiento de la masa de agua, V, la profundidad de la zona, d, y la distancia a recorrer, F, a través de la siguiente ecuación:

$$Z = \frac{F \cdot V^2}{1400 \cdot d},$$

donde V corresponde a una variable aleatoria, mientras que F y d son constantes conocidas para cada uno de los sectores. Suponga que $V \sim \text{Exponencial}(\lambda)$. Determine la función de densidad de Z.

(b) **[2.0 Ptos.]** El desborde del Río Las Minas de Punta Arenas depende de la acumulación de nieve en invierno en los cerros adyacentes. La acumulación de nieve puede catalogarse como alta, media o baja. Si la acumulación es alta la probabilidad de desborde alcanza a un 90 %, mientras que si la acumulación es media la probabilidad de desborde cae a un 40 %. Por último, si la acumulación es baja la probabilidad de desborde es sólo de un 10 %. Basado en estadísticas históricas, en el 20 % de los inviernos la acumulación ha sido alta, y en uno de cada dos inviernos ha sido media. Basado en todos estos antecedentes, evalúe la probabilidad que el presente año el Río Las Minas se desborde.

Problema 2

Durante los últimos días, como se ha informado, se ha producido un colapso en los servicios de urgencia de varios hospitales de la Región Metropolitana. Usted, interesado en disponer de información para la toma de una decisión informada, propone estimar el tiempo medio de espera de los pacientes, a través de un intervalo de confianza. Para esto, considere X como la variable que representa el tiempo de espera en urgencia en minutos, y suponga que X distribuye Exponencial($\nu = 1/\lambda$).

(b) **[2.0 Ptos.]** Determine paso a paso un intervalo de confianza exacto al $(1-\alpha) \times 100\,\%$ para λ, considerando que $\dfrac{2}{\lambda} \displaystyle\sum_{i=1}^{n} X_i \sim \chi^2(2\,n)$.

(b) **[2.0 Ptos.]** Determine paso a paso un intervalo de confianza aproximado al $(1-\alpha)\times 100\,\%$ para λ, a partir del estadístico $T = \sum_{i=1}^{n} X_i$ y el Teorema de límite Central.

(c) **[2.0 Ptos.]** Considere que se obtuvo 10 observaciones tal que la suma de sus tiempos de espera fue de 240 minutos. Evalúe cada uno de los intervalos de confianza obtenidos en (a) y (b) para una confianza del 95 %.

Solución:

a) Utilizando el pivote entregado se puede escribir

$$P\left(\mathcal{X}^2_{2n,\alpha/2} \leq \frac{2\sum_i X_i}{\lambda} \leq \mathcal{X}^2_{2n,1-\alpha/2}\right) = 1-\alpha \qquad \textbf{[0.8pt]}$$

$$\Longleftrightarrow \qquad P\left(\frac{2\sum_i X_i}{\mathcal{X}^2_{2n,1-\alpha/2}} \leq \lambda \leq \frac{2\sum_i X_i}{\mathcal{X}^2_{2n,\alpha/2}}\right) = 1-\alpha. \qquad \textbf{[0.8pt]}$$

Luego, el intervalo de $(1-\alpha)\times 100\,\%$ de confianza para λ está dado por:

$$\left(\frac{2\sum_i X_i}{\mathcal{X}^2_{2n,1-\alpha/2}},\ \frac{2\sum_i X_i}{\mathcal{X}^2_{2n,\alpha/2}}\right). \qquad \textbf{[0.4pt]}$$

b) Tenemos $\mu = \mathrm{E}(X) = \lambda$, $\sigma = \sqrt{\mathrm{Var}(X)} = \lambda$. El Teorema del Límite Central dice que, para n grande,

$$\sum_i X_i \sim Normal(n\lambda, \sqrt{n}\lambda),$$

es decir

$$\frac{\sum_i X_i - n\lambda}{\sqrt{n}\lambda} \sim Normal(0,1). \qquad \textbf{[0.5pt]}$$

Podemos plantear

$$P\left(-z_{1-\alpha/2} \leq \frac{\sum_i X_i - n\lambda}{\sqrt{n}\lambda} \leq z_{1-\alpha/2}\right) = 1-\alpha \qquad \textbf{[0.5pt]}$$

$$\Longleftrightarrow \qquad P\left(-z_{1-\alpha/2}\lambda\sqrt{n} \leq \sum_i X_i - n\lambda \leq z_{1-\alpha/2}\lambda\sqrt{n}\right) = 1-\alpha$$

$$\Longleftrightarrow \qquad P\left(\frac{\sum_i X_i}{z_{1-\alpha/2}\sqrt{n}+n} \leq \lambda \leq \frac{\sum_i X_i}{-z_{1-\alpha/2}\sqrt{n}+n}\right) = 1-\alpha. \qquad \textbf{[0.5pt]}$$

Luego, el intervalo de $(1-\alpha)\times 100\,\%$ de confianza para λ está dado por:

$$\left(\frac{\sum_i X_i}{z_{1-\alpha/2}\sqrt{n}+n},\ \frac{\sum_i X_i}{-z_{1-\alpha/2}\sqrt{n}+n}\right). \qquad \textbf{[0.5pt]}$$

c) Tenemos $n = 10, \sum_i X_i = 240, \mathcal{X}^2_{20,0.025} = 9.59, \mathcal{X}^2_{20,0.975} = 34.17, z_{1-\alpha/2} = 1.96$. **[0.2pt] por cada cuantil**
El intervalo de confianza en a) queda:

$$\left(\frac{2\times 240}{34.17}, \frac{2\times 240}{9.59}\right) \Longleftrightarrow (14.04, 50.05). \qquad \textbf{[0.7pt]}$$

El intervalo de confianza en b) queda:

$$\left(\frac{240}{1.96\sqrt{10}+10}, \frac{240}{-1.96\sqrt{10}+10}\right) \Longleftrightarrow (14.82, 63.13). \qquad \textbf{[0.7pt]}$$

+ 1 pto. base

Problema 3

Con el objeto de realizar el modelamiento del desgaste, X, sufrido por rodamientos (de acuerdo a lo especificado en la norma ISO 280/1), se debe optar entre una distribución log-Normal(λ, ζ) o LEV(μ, σ) para este desgaste (medido como el número de revoluciones hasta la fractura) asumiendo un valor esperado de 15 y un coeficiente de variación del 40 % en ambos casos.

Para tomar la decisión se someten a pruebas 100 rodamientos, observando las siguientes duraciones:

Número de revoluciones (en millones)	número de rodamientos
< 4	30
4 - 10	33
10 - 20	20
> 20	17

¿Qué distribución recomienda usted?

Nota: Si $X \sim \text{LEV}(\mu, \sigma)$ se tiene que

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \text{y} \quad f_X(x) = \frac{1}{\sigma}\,\phi\left(\frac{x-\mu}{\sigma}\right)$$

donde

$$\Phi(z) = \exp[-\exp(-z)], \quad \phi(z) = \exp[-z-\exp(-z)], \quad x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \quad \sigma > 0$$
$$\text{E}(X) = \mu + \sigma\,\gamma, \quad \text{Var}(X) = \sigma^2\,\pi^2/6, \quad \gamma \approx 0.5772, \quad \pi \approx 3.1416$$

Problema 4

El grabado con plasma es esencial en los procesos de fabricación de semiconductores. El artículo "Ion Beam-assisted etching of aluminum with chlorine" (J.Electrochem. Soc., 1985, pp:2010-12) entrega datos de 9 experimentos relativos al flujo de cloro (X, medido en sccm) empleado en el mecanismo grabador, y la rapidez del grabado (Y, medida en 100A/min). Al analizar la gráfica se sugiere un modelo de regresión lineal simple. Algunos resultados son:

$$\sum_{i=1}^{9} x_i = 24, \quad \sum_{i=1}^{9} y_i = 312.5, \quad \sum_{i=1}^{9} x_i^2 = 70.5, \quad \sum_{i=1}^{9} x_i\,y_i = 902.25, \quad \sum_{i=1}^{9} y_i^2 = 11626.75.$$

(a) **[2.0 Ptos.]** Obtenga los parámetros estimados de la recta de regresión, y los estadísticos correspondiente, $s^2_{Y\,|\,X}$ y r^2.

(b) **[2.0 Ptos.]** Asumiendo normalidad, obtenga un intervalo de confianza de 90 % para la rapidez media de grabado cuando el flujo de cloro alcanza a 4 sccm.

(c) **[2.0 Ptos.]** Asumiendo normalidad, determine la probabilidad que la rapidez de grabado sea superior a 50 100A/min, cuando el flujo de cloro alcanza los 4 sccm.

Tiempo: 2 Horas

Distribuciones

Distribución	Densidad de Probabilidad	Θ_X	Parámetros	Esperanza y Varianza
Binomial	$\binom{n}{x}\, p^x\,(1-p)^{n-x}$	$x = 0, \ldots, n$	$n,\ p$	$\mu_X = n\,p$ $\sigma_X^2 = n\,p\,(1-p)$ $M(t) = [p\,e^t + (1-p)]^n, \quad t \in \mathbb{R}$
Geométrica	$p\,(1-p)^{x-1}$	$x = 1, 2, \ldots$	p	$\mu_X = 1/p$ $\sigma_X^2 = (1-p)/p^2$ $M(t) = p\,e^t/[1-(1-p)\,e^t], \quad t < -\ln(1-p)$
Binomial-Negativa	$\binom{x-1}{r-1}\, p^r\,(1-p)^{x-r}$	$x = r, r+1, \ldots$	$r,\ p$	$\mu_X = r/p$ $\sigma_X^2 = r\,(1-p)/p^2$ $M(t) = \left\{p\,e^t/[1-(1-p)\,e^t]\right\}^r, \quad t < -\ln(1-p)$
Poisson	$\dfrac{(\nu\,t)^x\,e^{-\nu\,t}}{x!}$	$x = 0, 1, \ldots$	ν	$\mu_X = \nu\,t$ $\sigma_X^2 = \nu\,t$ $M(t) = \exp\left[\lambda\left(e^t - 1\right)\right], \quad t \in \mathbb{R}$
Exponencial	$\nu\,e^{-\nu\,x}$	$x \geq 0$	ν	$\mu_X = 1/\nu$ $\sigma_X^2 = 1/\nu^2$ $M(t) = \nu/(\nu - t), \quad t < \nu$
Gamma	$\dfrac{\nu^k}{\Gamma(k)}\, x^{k-1}\,e^{-\nu\,x}$	$x \geq 0$	$k,\ \nu$	$\mu_X = k/\nu$ $\sigma_X^2 = k/\nu^2$ $M(t) = [\nu/(\nu - t)]^k, \quad t < \nu$
Normal	$\dfrac{1}{\sqrt{2\,\pi}\,\sigma}\,\exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$	$-\infty < x < \infty$	$\mu,\ \sigma$	$\mu_X = \mu$ $\sigma_X^2 = \sigma^2$ $M(t) = \exp(\mu\,t + \sigma^2\,t^2/2), \quad t \in \mathbb{R}$
Log-Normal	$\dfrac{1}{\sqrt{2\,\pi}\,(\zeta\,x)}\,\exp\left[-\dfrac{1}{2}\left(\dfrac{\ln\,x - \lambda}{\zeta}\right)^2\right]$	$x \geq 0$	$\lambda,\ \zeta$	$\mu_X = \exp\left(\lambda + \dfrac{1}{2}\,\zeta^2\right)$ $\sigma_X^2 = \mu_X^2\left(e^{\zeta^2} - 1\right)$ $E(X^r) = e^{r\,\lambda}\,M_Z(r\,\zeta)$, con $Z \sim$Normal(0,1)

Formulario

- Sea $X_1, \ldots, X_n$ una muestra aleatoria independiente e idénticamente distribuida con función de probabilidad p_X o de densidad f_X, determinada por un parámetro θ. Si $\hat{\theta}$ es el estimador máximo verosímil del parámetro θ, entonces:
  - $\mathrm{E}(\hat{\theta}) \to \theta$, cuando $n \to \infty$.
  - $\mathrm{Var}(\hat{\theta}) = \dfrac{1}{I_n(\theta)}$, con $I_n(\theta) = -\mathrm{E}\left[\dfrac{\partial^2}{\partial\,\theta^2}\,\ln L(\theta)\right]$.
  - $\hat{\theta} \,\dot{\sim}\, \text{Normal}\left(\theta,\ \sqrt{\dfrac{1}{I_n(\theta)}}\right)$, cuando $n \to \infty$.
  - El estimador máximo verosímil de $g(\theta)$ es $g(\hat{\theta})$, cuya varianza está dada por: $\mathrm{Var}[g(\hat{\theta})] = \dfrac{\left[g'(\theta)\right]^2}{I_n(\theta)}$.
- Sean $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas Normal$(\mu,\,\sigma)$, entonces

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0,1), \quad \frac{\overline{X}_n - \mu}{s/\sqrt{n}} \sim \text{t-student}(n-1), \quad \frac{s^2\,(n-1)}{\sigma^2} \sim \chi^2_{n-1}$$

  con $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.
- Para un modelo de regresión simple $y' = \alpha + \beta \cdot x$ se tiene que

$$\hat{\beta} = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2}, \quad \hat{\alpha} = \overline{y} - \hat{\beta}\cdot\overline{x}, \quad r^2 = 1 - \frac{s^2_{Y|X}}{s^2_Y}, \quad s^2_{Y|X} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - y_i')^2, \quad s^2_Y = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

$$\hat{\rho}^2 = 1 - \left(\frac{n-2}{n-1}\right) \cdot \frac{s^2_{Y|X}}{s^2_Y}, \quad s^2_X = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$< \mu_{Y \mid x_i} >_{1-\alpha} = \overline{y}_i \pm t_{(1-\alpha/2),\, n-2} \cdot s_{Y \mid x} \cdot \sqrt{\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum\limits_{j=1}^{n}(x_j - \overline{x})^2}}, \quad \overline{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i$$

Tablas de Percentiles p

Distribución Normal Estándar k_p

k_p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Distribución t-student $t_p(\nu)$

ν	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602
16	1.337	1.746	2.120	2.583
17	1.333	1.740	2.110	2.567
18	1.330	1.734	2.101	2.552
19	1.328	1.729	2.093	2.539
20	1.325	1.725	2.086	2.528
21	1.323	1.721	2.080	2.518
22	1.321	1.717	2.074	2.508
23	1.319	1.714	2.069	2.500
24	1.318	1.711	2.064	2.492
25	1.316	1.708	2.060	2.485
26	1.315	1.706	2.056	2.479
27	1.314	1.703	2.052	2.473
28	1.313	1.701	2.048	2.467
29	1.311	1.699	2.045	2.462
30	1.310	1.697	2.042	2.457
∞	1.282	1.645	1.960	2.326

Distribución Chi-Cuadrado $c_p(\nu)$

ν	$c_{0.025}$	$c_{0.05}$	$c_{0.10}$	$c_{0.90}$	$c_{0.95}$	$c_{0.975}$	$c_{0.99}$	$c_{0.995}$
1	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93