Statistica I - A.A. 2016-2017

Prova scritta - 11 gennaio 2017

Exercise 1 *(pt 18, $\sim$ 3+4+4+4+3) L'ufficio vendite di un'azienda del marmo vuole fare un'analisi statistica. Relativamente alle ultime 20 settimane, ha registrato le quantità vendute (in tonnellate) trovando una media empirica pari a 5.4 ed una deviazione standard empirica pari a 2.1. [Può essere utile ricordare che la v.a.* $(n-1)\frac{S^2}{\sigma^2}$ *è una chi quadro a* $n-1$ *gradi di libertà e la v.a.* $\frac{\overline{X}-\mu}{S}\sqrt{n}$ *è una t di Student a* $n-1$ *gradi di libertà.]*

1. Con quale precisione conosciamo la media vera settimanale, a livello di confidenza 90%?

2. L'ufficio è preoccupato di vendere troppo poco. Stimare la probabilità di vendere, la prossima settimana, meno di 4 tonnellate. Svolgere l'esercizio sia utilizzando i dati a disposizione nel modo più immediato, sia presupponendo per la media la stima più pessimistica, al 90%.

3. Ripetere il calcolo del punto 2 utilizzando per la media il valore 5.4 e per la varianza la stima più pessimistica, al 90%; per pessimistica si intenda qui "la varianza più grande possibile" e, nella ricerca dell'intervallo di confidenza per la varianza, lo si cerchi bilaterale. Per lo svolgimento di questo esercizio non sono ammesse formule a memoria prefatte ma si deve partire dal fatto ricordato sopra.

4. Supponiamo che, prima di eseguire le osservazioni, si ritenesse che le vendite del passato avevano una media di almeno 6 tonnellate alla settimana e che l'indagine sia stata svolta a causa del dubbio che le vendite recenti fossero in calo rispetto al passato. Scrivere ipotesi nulla ed alternativa ed una regione di rifiuto al 95% per testare l'ipotesi, e stabilire se 5.4 (con deviazione empirica 2.1) è una media compatibile con l'ipotesi.

5. Stimare la probabilità di vendere, cumulativamente nelle prossime 30 settimane, meno di $30 \cdot 4$ tonnellate (usare come parametri i valori empirici iniziali).

Exercise 2 *(pt 12) Sia* X *una v.a. di densità* $f(\theta, x) = C_\theta \exp\left(-\frac{|x|}{\theta}\right)$*, dove il parametro* θ *varia nell'insieme dei numeri reali positivi.*

1. Trovare la costante C_θ per cui la funzione $f(\theta, x)$ sia davvero una densità di probabilità.

2. Posto $Y = |X|$, trovare la funzione di ripartizione $F_Y(\theta, y)$ e riconoscere che densità di probabilità $f_Y(\theta, y)$ ha la v.a. Y.

3. Trovare lo stimatore di massima verosimiglianza del parametro θ.

4. Calcolare $P(X > 2 | X > 1)$. Vale la proprietà di assenza di memoria?

1 Soluzioni

Esercizio 1.

1. $$\mu = 5.4 \pm \frac{2.1 \cdot t_{19,0.05}}{\sqrt{20}} = 5.4 \pm \frac{2.1 \cdot 1.729}{\sqrt{20}} = 5.4 \pm 0.81189.$$

2. $$P\left(V \le 4\right) = P\left(\frac{V-\mu}{\sigma} \le \frac{4-\mu}{\sigma}\right) = \Phi\left(\frac{4-\mu}{\sigma}\right).$$

   L'approssimazione più immediata è

   $$\Phi\left(\frac{4-5.4}{2.1}\right) = \Phi\left(-0.66\right) = 1 - \Phi\left(0.66\right) = 1 - 0.74537 = 0.25463.$$

   Quella più elaborata è

   $$\Phi\left(\frac{4-5.4+0.81189}{2.1}\right) = \Phi\left(-0.28\right) = 1 - \Phi\left(0.28\right) = 1 - 0.61026 = 0.38974.$$

3. $$P\left(\chi^2_{19,0.95} \le (n-1)\frac{S^2}{\sigma^2} \le \chi^2_{19,0.05}\right) = 0.9.$$

   Invertendo,

   $$(n-1)\,S^2/\chi^2_{19,0.05} \le \sigma^2 \le (n-1)\,S^2/\chi^2_{19,0.95}.$$

   Il valore peggiore è

   $$(n-1)\,S^2/\chi^2_{19,0.95} = \frac{19 \cdot 2.1^2}{10.12} = 8.2796.$$

   Su questa base, l'approssimazione è

   $$\Phi\left(\frac{4-5.4}{\sqrt{(n-1)\,S^2/\chi^2_{19,0.95}}}\right) = \Phi\left(\frac{4-5.4}{\sqrt{8.2796}}\right) = \Phi\left(-0.48\right)$$
   $$= 1 - \Phi\left(0.48\right) = 1 - 0.68439 = 0.31561.$$

4. $\mathcal{H}_0$) media≥ 6, $\mathcal{H}_1$) media< 6; regione di rifiuto:

   $$\frac{\overline{X}-6}{S}\sqrt{n} < t_{n-1,0.95} = -t_{19,0.05}$$

   ovvero esplicitamente $\frac{\overline{X}-6}{S}\sqrt{20} < -1.729$; Per i valori empirici vale

   $$\frac{\overline{x}-6}{s}\sqrt{20} = \frac{5.4-6}{2.1}\sqrt{20} = -1.2778$$

   quindi il test non è significativo.

5. Dette $V_1, ..., V_{30}$ le vendite delle prossime 30 settimane, la vendita cumulativa è $V_1 + ... + V_{30}$; la domanda chiede di stimare

$$P\left(V_1 + ... + V_{30} \leq 30 \cdot 4\right) = P\left(\frac{V_1 + ... + V_{30} - 30 \cdot 5.4}{\sqrt{30}2.1} \leq \frac{30 \cdot 4 - 30 \cdot 5.4}{\sqrt{30}2.1}\right)$$

che è o uguale (se si assume la gaussianità) o approssimativamente uguale (per il TLC) a

$$\Phi\left(\frac{30 \cdot 4 - 30 \cdot 5.4}{\sqrt{30}2.1}\right) = \Phi\left(-3.6515\right) = 1 - \Phi\left(3.6515\right) = 1 - 0.99987 = 0.00013.$$

Esercizio 2.

1.

$$1 = \int_{-\infty}^{\infty} C_\theta \exp\left(-\frac{|x|}{\theta}\right) dx = 2\int_0^{\infty} C_\theta \exp\left(-\frac{x}{\theta}\right) dx$$
$$= -2C_\theta\left[\theta \exp\left(-\frac{x}{\theta}\right)\right]_0^{\infty} = 2C_\theta\theta\,, C_\theta = \frac{1}{2\theta}.$$

2.

$$F_Y\left(\theta, y\right) = P\left(Y \leq y\right) = P\left(|X| \leq y\right)$$

è zero per $y < 0$, mentre per $y \geq 0$

$$= P\left(-y \leq X \leq y\right) = \int_{-y}^{y} \frac{1}{2\theta} \exp\left(-\frac{|x|}{\theta}\right) dx = \frac{1}{\theta}\int_0^{y} \exp\left(-\frac{x}{\theta}\right) dx$$
$$= -\left[\exp\left(-\frac{x}{\theta}\right)\right]_0^{y} = 1 - \exp\left(-\frac{y}{\theta}\right)$$

La densità è esponenziale di parametro $\frac{1}{\theta}$.

3.

$$L\left(\theta, x_1, ..., x_n\right) = \frac{1}{2^n\theta^n}\exp\left(-\frac{|x_1| + ... + |x_n|}{\theta}\right)$$

$$\log L\left(\theta, x_1, ..., x_n\right) = .n\log 2 - n\log\theta - \frac{|x_1| + ... + |x_n|}{\theta}$$

$$-\frac{n}{\theta} + \frac{|x_1| + ... + |x_n|}{\theta^2} = 0$$

$$\frac{|x_1| + ... + |x_n|}{\theta} = n$$

$$\widehat{\theta} = \frac{|x_1| + ... + |x_n|}{n}.$$

4.

$$P\left(X > 2 | X > 1\right) = \frac{P\left(X > 2, X > 1\right)}{P\left(X > 1\right)} = \frac{P\left(X > 2\right)}{P\left(X > 1\right)}.$$

In generale, per $a > 0$,

$$P(X > a) = \int_a^{\infty} \frac{1}{2\theta} \exp\left(-\frac{x}{\theta}\right) dx = -\frac{1}{2}\left[\exp\left(-\frac{x}{\theta}\right)\right]_a^{\infty} = \frac{1}{2}\exp\left(-\frac{a}{\theta}\right).$$

Quindi

$$P(X > 2 | X > 1) = \frac{\exp\left(-\frac{2}{\theta}\right)}{\exp\left(-\frac{1}{\theta}\right)} = \exp\left(\frac{1}{\theta}\right).$$

Non vale la proprietà di assenza di memoria, perché dovrebbe essere uguale a $P(X > 2 - 1) = \frac{1}{2}\exp\left(\frac{1}{\theta}\right)$.