Esercizi di Statistica della 5^a settimana (Corso di Laurea in Biotecnologie, Università degli Studi di Padova).

Esercizio 1. In una grande città degli Stati Uniti in cui il 48% dei votanti alle ultime elezioni aveva votato repubblicano, un campione casuale di 200 votanti viene intervistato. Supponiamo che esattamente 100 dichiarino che alle prossime elezioni voteranno repubblicano.

1. Calcolare l'intervallo di confidenza al 95% della percentuale di repubblicani nella città.
2. Quanto grande bisognerebbe prendere il campione per avere un'incertezza massima dello 0.5% nel tasso di repubblicani?
3. Eseguire un test per vedere se la percentuale di repubblicani sia variata rispetto alle ultime elezioni: riportare un valore P.

Esercizio 2. Nelle precedenti elezioni comunali a Padova (2004), Flavio Zanonato è stato eletto sindaco al primo turno. Vogliamo ora vedere se questo risultato era prevedibile attraverso un sondaggio fatto 15 giorni prima. Dal sondaggio, condotto su 902 persone, è risultato che il 51.5% avrebbe votato Zanonato, e il 48.5% avrebbe votato altri candidati.

1. Calcolare l'intervallo di confidenza al 95% della proporzione di elettori che avrebbero votato Zanonato.
2. Che test bisognerebbe eseguire per escludere un ballottaggio in base al sondaggio? Ricordiamo che la vittoria al primo turno si ottiene solo se il candidato ha la maggioranza assoluta dei voti.
3. Eseguire il test del punto 1, calcolando il valore P.
4. Dire se in base a questo valore si poteva prevedere la vittoria di Zanonato al primo turno.

Esercizio 3. In un certo procedimento chimico, è di fondamentale importanza che il pH di uno dei reagenti sia esattamente 8.20. Supponiamo che 10 misurazioni indipendenti abbiano dato i seguenti valori:

$$8.18 \quad 8.16 \quad 8.17 \quad 8.22 \quad 8.19 \quad 8.17 \quad 8.15 \quad 8.21 \quad 8.16 \quad 8.18$$

1. Calcolare l'intervallo di confidenza al 95% del pH.
2. Effettuare un test per vedere se effettivamente il valore del pH si può considerare uguale a 8.20. Riportare limitazioni al valore P.

Supponiamo ora di voler verificare che il pH non differisca da 8.20 per più di 0.03 (in caso contrario la reazione non avrebbe luogo). Supponiamo anche che la vera deviazione standard sia circa uguale a quella stimata in precedenza.

3. Quanto numeroso dovrà essere il campione scelto perchè l'intervallo di confidenza sia largo al massimo 0.06?

4. Se con il nuovo campione avessimo $\bar{X} = 8.21$, che conclusione si trarrebbe?

Esercizio 4. Una compagnia farmaceutica produce un nuovo farmaco contro le emicranie con un principio attivo molto rapido ad entrare in circolo. Per convincere l'ente preposto al controllo dei nuovi medicinali che il tempo medio che il farmaco impiega a raggiungere il sangue è inferiore ai 10 minuti, questa ditta raduna un campione di persone soggette ad emicranie e conduce un esperimento.

1. Quale tipo di test bisogna effettuare e come vanno scelte l'ipotesi e l'alternativa?

2. Supponiamo di prendere un gruppo da 30 persone e di avere le seguenti statistiche cumulative:
$$\bar{X} = 9.267, \qquad s_X = 2.29$$
Supponendo che l'ente preposto al controllo voglia una significatività dell'1%, si può affermare che il farmaco raggiunga il sangue in meno di 10 minuti?

Esercizio 5. In una fabbrica è sempre stato registrato un livello di inquinamento acustico medio di 90 dB. Durante un successivo giorno lavorativo vengono effettuate 32 misure, con media $\bar{X} = 95.55$ e varianza $s_X^2 = 70.02$.

1. Calcolare l'intervallo di confidenza al 95% del livello medio di inquinamento acustico.

2. Quante misurazioni dovremmo prendere perchè l'intervallo di confidenza sia largo al massimo 10 dB?

3. Dire se in base al campione ci sono evidenze che il livello acustico nel giorno in questione sia stato maggiore di 90 dB e di 95 dB: usare per entrambi $\alpha = 0.05$.

Soluzioni su `http://www.math.unipd.it/~vargiolu/Statistica/`

Soluzioni

Esercizio 1.

1. Abbiamo $\hat{p} = \frac{100}{200} = 0.5$, e quindi l'intervallo di confidenza al 95% ha estremi

$$\hat{p} \pm \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{200}} \cdot z_{\alpha/2} = 0.5 \pm \sqrt{\frac{0.5 \cdot 0.5}{200}} \cdot z_{0.025} = 0.5 \pm \sqrt{0.00125} \cdot 1.96 = 0.5 \pm 0.0692$$

e quindi è $[0.4108; 0.5492]$.

2. Non sapendo qual è la vera proporzione, bisognerebbe usare la maggiorazione $\hat{p}(1-\hat{p}) \leq \frac{1}{4}$, e quindi l'incertezza è pari a

$$z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \geq \frac{z_{\alpha/2}}{\sqrt{4n}} = \frac{1.96}{2\sqrt{n}}$$

quindi per avere incertezza minore di $0.5\% = 0.005$ bisogna imporre

$$0.005 \geq z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \geq \frac{1.96}{2\sqrt{n}}$$

che risulta in

$$n \geq \left(\frac{1.96}{2 \cdot 0.005}\right)^2 = 196^2 = 38416$$

(ben più dei 200 del sondaggio).

3. Eseguiamo un test di ipotesi $H_0 : p = 0.48$ e alternativa $H_1 : p \neq 0.48$. Calcoliamo la statistica

$$Z = \frac{0.5 - 0.48}{\sqrt{\frac{0.48 \cdot (1-0.48)}{200}}} = 0.56$$

Poichè $F_Z(0.56) = 0.712$, abbiamo che il valore P è dato da

$$P = \mathbb{P}\{|Z| > 0.56\} = 2\mathbb{P}\{Z > 0.56\} = 2(1-F_Z(0.56)) = 2(1-0.712) = 2 \cdot 0.288 = 0.576$$

Con un valore P così alto, chiaramente accettiamo H_0 e concludiamo che il tasso di repubblicani non sembra essere variato significativamente.

Esercizio 2.

1. Abbiamo $\hat{p} = 0.515$ e

$$s_{\hat{p}} = \sqrt{\frac{0.5 \cdot 0.5}{902}} = 0.0166$$

quindi l'intervallo di confidenza al 95% ha estremi

$$\hat{p} \pm s_{\hat{p}} \cdot z_{\alpha/2} = 0.515 \pm 0.0166 \cdot z_{0.025} = 0.515 \pm 0.0166 \cdot 1.96 = 0.515 \pm 0.032$$

e quindi è uguale a $[0.483; 0.547]$.

2. Bisogna eseguire un test Z di ipotesi $H_0 : p \leq 1/2$ e alternativa $H_1 : p > 1/2$. Il test si può eseguire in quanto $np = 902 \cdot 1/2 = 451 > 5$.

3. Abbiamo
$$Z = \frac{0.515 - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{902}}} = 0.9036$$
Poichè $F(0.90) = 0.815$ e $F(0.91) = 0.818$, abbiamo che $q_{0.815} < 0.9036 < q_{0.818}$. Per un dato α, il valore critico del test sarebbe $q_{1-\alpha}$, poichè il test è unilatero, e quindi $0.182 = 1 - 0.818 < P < 1 - 0.815 = 0.185$.

4. Con un valore P così alto, rifiutando l'ipotesi ci sarebbe stata una probabilità molto alta (circa 0.18) di commettere un errore di prima specie. Pertanto la conclusione dal sondaggio è che non si sarebbe potuto escludere un eventuale ballottaggio.

Esercizio 3.

1. Calcolando media e varianza campionarie otteniamo $\bar{X} = 8.179$ e $s_X = 0.0223$, quindi $s_{\bar{X}} = s_X/\sqrt{10} = 0.00706$. Abbiamo poi $t_{n-1,\alpha/2} = t_{9;0.025} = 2.262$. Allora l'intervallo di confidenza ha estremi $\bar{X} \pm t_{n-1,\alpha/2} s_{\bar{X}} = 8.179 \pm 2.262 \cdot 0.00706$, e quindi risulta uguale a $[8.163; 8.195]$.

2. Facciamo un test t sulla media con ipotesi $H_0 : \mu = 8.20$ e alternativa $H_1 : \mu \neq 8.20$. Abbiamo
$$t = \frac{\bar{X} - 8.20}{s_{\bar{X}}} = \frac{8.179 - 8.20}{0.0706} = -2.973$$
I gradi di libertà sono $\nu = 10 - 1 = 9$. Siccome $t_{9;0.01} = 2.821 < |t| < t_{9;0.005} = 3.249$, possiamo riportare $0.01 < P < 0.02$. Questo significa che possiamo rifiutare H_0 e accettare H_1.

3. Per semplicità supponiamo che n venga "grande", così da poter usare, in prima approssimazione, i quantili di una normale. Dobbiamo imporre che
$$n \geq \left(\frac{2 z_{\alpha/2} \sigma}{0.06} \right)^2 = \left(\frac{2 \cdot 1.96 \cdot 0.0223}{0.06} \right)^2 = 2.12$$
quindi $n \geq 3$, numero per cui l'approssimazione normale non è assolutamente adatta. Proviamo quindi a rifare il calcolo con i quantili di una t di Student con $3 - 1 = 2$ gradi di libertà:
$$n \geq \left(\frac{2 t_{2;\alpha/2} \sigma}{0.06} \right)^2 = \left(\frac{2 \cdot 4.303 \cdot 0.0223}{0.06} \right)^2 = 10.23$$
che dà $n \geq 11$. Procedendo per approssimazioni successive, si trova che il valore che non dà contraddizioni è $n \geq 5$: difatti per $n = 5$ si ha
$$n \geq \left(\frac{2 t_{4;\alpha/2} \sigma}{0.06} \right)^2 = \left(\frac{2 \cdot 2.776 \cdot 0.0223}{0.06} \right)^2 = 4.25$$
quindi per avere un intervallo di confidenza largo al massimo 0.06 bisogna avere almeno 5 elementi nel campione.

4. Rifacciamo il test t con questi nuovi valori, supponendo che la deviazione standard sia circa uguale a quella calcolata nel punto 1:
$$t = \frac{\bar{X} - 8.20}{s_{\bar{X}}} = \frac{8.31 - 8.20}{\frac{0.0223}{\sqrt{15}}} = 19.07$$
I gradi di libertà stavolta sono $\nu = 15 - 1 = 14$. Siccome $t_{14;0.0005} = 4.140 < t$, possiamo riportare $P < 0.001$. Questo significa che possiamo rifiutare H_0 e accettare H_1.

Esercizio 4.

1. Bisogna fare un test unilatero sulla media di ipotesi $H_0 : \mu \geq 10$ e alternativa $H_1 : \mu < 10$.

2. Calcoliamo
$$t = \frac{\bar{X} - 10}{s_{\bar{X}}} = \frac{9.267 - 10}{0.4181} = -1.753$$
dove $s_{\bar{X}} = s_X/\sqrt{30} = 0.4181$. Il valore critico è dato da $t_{n-1;\alpha} = t_{29;0.01} = 2.462$. Siccome $t > -t_{n-1;\alpha}$, accettiamo H_0: non ci sono quindi abbastanza elementi per affermare che il farmaco raggiunga il sangue in meno di 10 minuti.

Esercizio 5.

1. Calcoliamo gli estremi dell'intervallo:
$$\begin{aligned}\bar{X} \pm \sqrt{\frac{s_X^2}{n}} \cdot t_{n-1;\alpha/2} &= 95.55 \pm \sqrt{\frac{70.02}{32}} \cdot t_{31;0.05} \simeq 95.55 \pm 1.47 \cdot t_{30;0.05} = \\ &= 95.55 \pm 1.47 \cdot 2.042 = 95.55 \pm 3.00\end{aligned}$$
e quindi l'intervallo di confidenza è $[92.55; 98.55]$.

2. Supponendo n abbastanza grande da applicare l'approssimazione normale, e che la vera varianza sia simile a quella stimata (cioè $\sigma^2 \simeq s_X^2 = 70.02$), bisogna prendere
$$n \geq \left(\frac{2z_{\alpha/2}\sigma}{10}\right)^2 \simeq \left(\frac{2 \cdot 1.96 \cdot \sqrt{70.02}}{10}\right)^2 = 10.75$$
quindi $n \geq 11$, numero per cui l'approssimazione normale non è adatta. Proviamo quindi a rifare il calcolo con i quantili di una t di Student con $11 - 1 = 10$ gradi di libertà:
$$n \geq \left(\frac{2t_{10;\alpha/2}\sigma}{10}\right)^2 \simeq \left(\frac{2 \cdot 2.228 \cdot \sqrt{70.02}}{10}\right)^2 = 13.90$$
che dà $n \geq 14$. Proviamo infine a rifare il calcolo con i quantili di una t di Student con $14 - 1 = 13$ gradi di libertà:
$$n \geq \left(\frac{2t_{13;\alpha/2}\sigma}{10}\right)^2 \simeq \left(\frac{2 \cdot 2.160 \cdot \sqrt{70.02}}{10}\right)^2 = 13.06$$
che dà $n \geq 14$, coerente con quanto ottenuto: quindi per avere un intervallo largo al massimo 10 dB bastano 14 misurazioni.

3. Per il primo valore, dobbiamo effettuare un test di ipotesi $H_0 : \mu_X = 90$ e alternativa $H_1 : \mu_X > 90$. Calcoliamo

$$t = \frac{\bar{X} - 90}{\sqrt{\frac{s_X^2}{32}}} = \frac{0.55}{1.47} = 3.718$$

Poichè $t_{31;0.05} \simeq t_{30;0.05} = 1.697 < t$, rifiutiamo H_0 e accettiamo H_1: ci sono quindi evidenze che il livello acustico nel giorno in questione sia stato maggiore di 90 dB.

Per il secondo valore, dobbiamo effettuare un test di ipotesi $H_0 : \mu_X = 95$ e alternativa $H_1 : \mu_X > 95$. Calcoliamo

$$t = \frac{\bar{X} - 95}{\sqrt{\frac{s_X^2}{32}}} = \frac{0.55}{1.47} = 0.3718$$

Poichè $t_{31;0.05} \simeq t_{30;0.05} = 1.697 > t$, accettiamo H_0; non ci sono quindi evidenze che il livello acustico nel giorno in questione sia stato maggiore di 95 dB.