

STATISTICA SOCIALE
Corso di laurea in Scienze Turistiche, a.a. 2006/2007
Prova del 20 dicembre 2006

Esercizio 1 [13 pt]

La seguente Tabella 1 riporta la distribuzione degli esercizi ricettivi per tipologia e zona di una certa area geografica (dati fittizi).

Tabella 1 – Esercizi ricettivi per tipologia e zona

Tipologia	Zona	
	A	B
Albergo 5/4 stelle	1000	500
Albergo 2/3 stelle	3000	2500
Campeggi e villaggi turistici	300	250
Case in affitto	3500	1000
Bed and Breakfast	1500	1000
Altro	2700	2250

Si vuole effettuare un'indagine campionaria sul turismo nell'area considerata, intervistando tutti i clienti che soggiornano in un dato periodo di riferimento nelle strutture ricettive campionate.

1. Supponendo di estrarre un campione casuale semplice di esercizi ricettivi:
 - a. [2pt] Descrivere brevemente che cosa si intende per base di campionamento e quali problemi potrebbero presentarsi nella predisposizione della base di campionamento per questa indagine;
 - b. [2pt] Indicare la numerosità campionaria complessiva nel caso in cui si proceda all'estrazione di un campione con frazione di campionamento pari al 10%.
2. [5pt] Supponendo di voler estrarre un campione stratificato in base alla tipologia e alla zona, con estrazione di 100 esercizi in ciascuno strato: indicare il numero degli strati, la frazione di campionamento adottata in ciascuno strato e la numerosità campionaria complessiva.
3. [2pt] Come si dovrebbe procedere qualora si volesse formare un campione sistematico di 1000 esercizi?
4. [2pt] Quale tecnica di indagine utilizzereste per intervistare i clienti delle strutture ricettive selezionate? Motivare la risposta.

Esercizio 2 [28pt]

La Tabella 2 contiene la *nazione di provenienza* e il numero di *notti trascorse* presso l'albergo per 15 clienti dell'Albergo Bellavista.

- a) Per il carattere *Provenienza*:
 - [3pt] Costruire la distribuzione di frequenza relativa osservata;
 - [5pt] Calcolare l'indice di entropia e l'indice di entropia standardizzato, utilizzando i logaritmi naturali (base e);
 - [2pt] Costruire la distribuzione relativa di massima eterogeneità;
 - [3pt] Quali sono i valori minimo e massimo assumibili dall'indice di entropia e dall'indice di entropia standardizzato?
- b) Relativamente al carattere *Notti* (n. di notti trascorse):
 - [2pt] Calcolare campo di variazione
 - [5pt] Disegnare il box-plot

- [5pt] Calcolare varianza, deviazione standard e coefficiente di variazione
- [3pt] Indicando con X la variabile *Notti*, quanto valgono varianza, scarto quadratico medio e coefficiente di variazione della variabile $Y=2*X-1$?

Tabella 2 – Clienti Albergo Bellavista

Cliente	Provenienza	Notti
1	Italia	3
2	Francia	5
3	Spagna	4
4	Italia	2
5	Italia	2
6	Spagna	3
7	Francia	4
8	Italia	5
9	Italia	3
10	Spagna	2
11	Spagna	2
12	Italia	3
13	Francia	4
14	Italia	7
15	Italia	5

Esercizio 3 [17pt]

Rispondere brevemente ai seguenti quesiti:

1. [3pt] Dare la definizione di *visitatore* fornita dal WTO, distinguendo per tipologia di *turista* e di *escursionista*.
2. [3pt] Quali flussi compongono il turismo internazionale? In Italia, quali fonti statistiche ufficiali possono essere utilizzate per studiare i flussi turistici *internazionali*?
3. [3pt] Quali flussi compongono il turismo dei nazionali? Esistono in Italia fonti statistiche ufficiali che consentano di conoscere i flussi turistici *dei nazionali*? In caso affermativo, indicare quali.
4. [2pt] Che cosa si intende per turismo interno?
5. [3pt] Indicare la cadenza e le principali informazioni raccolte dall'indagine sui Viaggi in Italia e all'estero effettuata dall'ISTAT
6. [3pt] La Rilevazione sul movimento dei clienti negli esercizi ricettivi
 - a. è un'indagine campionaria o esaustiva?
 - b. quali entità rileva?
 - c. quali sono i principali problemi legati a questa indagine?

Soluzione

Esercizio 1

1. Supponendo di estrarre un campione casuale semplice di esercizi ricettivi:

- Descrivere brevemente che cosa si intende per base di campionamento e quali problemi potrebbero presentarsi nella predisposizione della base di campionamento per questa indagine;

1a La **base di campionamento** è una lista che elenca tutte le unità statistiche appartenenti alla popolazione di riferimento. In particolare, trattandosi di esercizi alberghieri, si potrebbe formare la lista facendo riferimento alla Registro delle imprese commerciali disponibile presso la Camera di Commercio (REC). Tuttavia possono sorgere problemi di sotto-copertura e sovra-copertura legati al mancato aggiornamento della lista: alcuni esercizi potrebbero non essere registrati al REC (specialmente nel caso degli alloggi in affitto), mentre alcuni degli esercizi registrati potrebbero non essere più in attività.

- Indicare la numerosità campionaria complessiva nel caso in cui si proceda all'estrazione di un campione con frazione di campionamento pari al 10%.

1b La numerosità complessiva della popolazione è pari a $N=19500$ esercizi, pertanto, applicando una frazione di campionamento $f=0.10$, si ottiene la numerosità campionaria $n=19500 \times 0.10=1950$.

2. Supponendo di voler estrarre un campione stratificato in base alla tipologia e alla zona, con estrazione di 100 esercizi in ciascuno strato: indicare il numero degli strati, la frazione di campionamento adottata in ciascuno strato e la numerosità campionaria complessiva.

- Il **numero degli strati** è pari al prodotto tra il numero di modalità di tipologia (6) e di zona (2), cioè numero di strati = $6 \times 2=12$
- La numerosità campionaria è pari a $n=12 \times 100=1200$
- Per **frazione di campionamento** si intende il rapporto tra numerosità campionaria e numerosità della popolazione, pertanto, per la frazione di campionamento per strato si ottiene nel modo seguente: $f_j=n_j/N_j$, dove $j=1,...,12$ indica lo strato, $n_j=n=100$ è la numerosità campionaria – costante – per strato mentre N_j è il numero di esercizi per strato.

Frazione di campionamento per strato

Tipologia	Zona	
	A	B
Albergo 5/4 stelle	0.10	0.20
Albergo 2/3 stelle	0.03	0.04
Campeggi e villaggi turistici	0.33	0.40
Case in affitto	0.03	0.10
Bed and Breakfast	0.07	0.10
Altro	0.04	0.04

3. Come si dovrebbe procedere qualora si volesse formare un campione sistematico di 1000 esercizi?

Il campionamento sistematico è una variante del campionamento casuale semplice molto efficiente da realizzare quando si disponga della lista delle unità statistiche della popolazione sotto forma di file elaborabile al computer. Il procedimento per la formazione di un campione sistematico di numerosità $n=1000$ è il seguente:

- Si estrae un numero a caso tra 1 e $N=19500$ e si inserisce nel campione l'unità corrispondente nella lista.
- Le unità successive sono scelte scorrendo la lista a partire dalla prima unità prescelta e selezionando nuove unità con un **passo** dato dal rapporto $k=N/n=19500/1000=19,5$. Si considera **$k=20$** , cioè il numero intero più vicino a 19.5 (anche 19 andava bene).
- Il procedimento deve essere tale che, una volta giunti in fondo alla lista delle N unità, occorre proseguire il conteggio a partire dall'inizio della lista.
- Il procedimento termina quando sono state selezionate tutte le n unità da campionare.

4. Quale tecnica di indagine utilizzereste per intervistare i clienti delle strutture ricettive selezionate? Motivare la risposta.

In questo caso è opportuno utilizzare un'**intervista diretta**, con l'ausilio di un intervistatore, o con questionario auto-compilato.

I principali vantaggi dell'intervista diretta, in questo caso, sono legati a:

- Maggiore possibilità di contattare e convincere il rispondente a collaborare
- Si identifica esattamente il rispondente.
- Possibilità di istruire il rispondente sul significato delle domande e sul modo corretto di fornire le risposte
- Interviste di maggiore durata

Tuttavia, dato il costo e lo sforzo organizzativo richiesti da questa tecnica, si potrebbe optare per un questionario auto-compilato, ricordando tuttavia che, in questo caso, tale tecnica potrebbe comportare problemi legati a minore capacità di ottenere la partecipazione all'indagine e maggiore difficoltà ad aiutare i rispondenti nella comprensione delle domande e nella compilazione del questionario (importanza della grafica).

Esercizio 2

a) Per il carattere *Provenienza*:

- a1 costruire la distribuzione di frequenza relativa osservata;

Provenienza	n_j	f_j	$\ln(f_j)$	$f_j \cdot \ln(f_j)$
Italia	8	0.53	-0.62861	-0.33526
Francia	3	0.20	-1.60944	-0.32189
Spagna	4	0.27	-1.32176	-0.35247
TOTALE	15	1.00		-1.00961

- *a2 Calcolare l'indice di entropia e l'indice di entropia standardizzato, utilizzando i logaritmi naturali (base e);*

Nella tabella delle frequenze relative sono stati inseriti anche i calcoli necessari per la costruzione dell'indice di entropia di Shannon H e l'indice normalizzato H' , calcolati utilizzando i logaritmi naturali delle frequenze relative $\ln(f_j)$ ¹.

$$-\sum_{j=1}^3 f_j \ln f_j = 1.009614$$

$$H' = H / \ln(3) = 0.91899$$

- *a3 Costruire la distribuzione relativa di massima eterogeneità;*

La distribuzione di massima eterogeneità corrisponde al caso di distribuzione uniforme delle frequenze, come riportato nella seguente tabella:

Provenienza	n_j	f_j
Italia	5	0.33
Francia	5	0.33
Spagna	5	0.33
TOTALE	15	1.00

- *a4 Quali sono i valori minimo e massimo assumibili dall'indice di entropia e dall'indice di entropia standardizzato?*

L'indice di entropia H varia tra 0, valore minimo, e $\ln(3)=1.098612$, valore massimo. Il valore minimo dell'indice standardizzato H' è sempre 0, mentre il valore massimo è 1.

H e H' assumono il valore minimo nel caso di massima omogeneità, ossia per variabile statistica degenerare. In tal caso le frequenze sono tutte concentrate in un'unica modalità del carattere. Il valore massimo viene invece raggiunto nel caso di massima eterogeneità, ossia nel caso di distribuzione uniforme delle frequenze.

b) Relativamente al carattere Notti (n. di notti trascorse)

Risulta utile, ma non necessario, per rispondere ai quesiti successivi, costruire la distribuzione di frequenze.

notti	n_j	N_j
2	4	4
3	4	8
4	3	11
5	3	14
7	1	15
Totale	15	

- *Calcolare campo di variazione: $R = x_{\max} - x_{\min} = 7 - 2 = 5$*
- *Disegnare il box-plot*

Per disegnare il box-plot è necessario calcolare i quartili $Q1$, *mediana* e $Q3$.

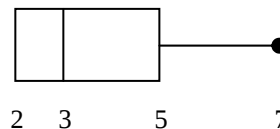
Se non si è costruita la tabella di frequenza, ricordarsi di ordinare i valori di *Notti* dal più piccolo al più grande. Quindi procedere nel modo seguente:

- ✓ Trovare la posizione occupata da $Q1$: $15 \cdot 0.25 = 3.75 \rightarrow 4^{\circ}$ posizione $\rightarrow Q1 = x_4 = 2$
- ✓ Trovare la posizione occupata dalla mediana M_e : $(15+1)/2 = 8 \rightarrow 8^{\circ}$ posizione $\rightarrow M_e = x_8 = 3$

- ✓ Trovare la posizione occupata da Q3: $15 \times 0.75 = 11.25 \rightarrow 12^\circ \text{ posizione} \rightarrow Q3 = x_{12} = 5$

Per trovare la lunghezza dei 'baffi' è inoltre necessario calcolare lo scarto interquartile e i due punti di troncamento, inferiore e superiore:

- ✓ Scarto interquartile: $Q3 - Q1 = 5 - 2 = 3$
- ✓ Troncamento inferiore: $\max[x_{\min}, (Q1 - 1.5 \times DI)] = 2$
- ✓ Troncamento superiore: $\min[x_{\max}, (Q3 + 1.5 \times DI)] = 7$



- Calcolare varianza, deviazione standard e coefficiente di variazione

- ✓ Media aritmetica: $M = \frac{1}{15} \sum_{i=1}^{15} x_i = \frac{1}{15} \sum_{j=1}^4 x_j n_j = 3.6$

- ✓ Media quadratica al quadrato $\frac{1}{15} \sum x_j^2 = \frac{1}{15} \sum x_j^2 n_j = 14.93$

- ✓ Varianza: $\sigma^2 = 14.93 - (3.6)^2 = 1.97$

- ✓ Deviazione standard: $\sigma = \sqrt{1.97} = 1.40$

- ✓ CV = $100 \times \sigma / M = 100 \times (1.40 / 3.6) = 39.02$

- Indicando con X la variabile Notti, quanto valgono varianza, scarto quadratico medio e coefficiente di variazione della variabile $Y = 2 \times X - 1$?

- ✓ Media di Y : $2 \times M - 1 = 3.6 \times 2 - 1 = 6.20$

- ✓ Varianza di Y : $V(Y) = V(2 \times X - 1) = 4 \times \sigma^2 = 7.89$

- ✓ Deviazione standard di Y : $2 \times \sigma = 2.80$

- ✓ CV di Y : $100 \times (2 \times \sigma) / (2 \times M - 1) = 100 \times (2.80 / 6.20) = 45.31$

Esercizio 3

1. Dare la definizione di visitatore fornita dal WTO, distinguendo per tipologia di turista e di escursionista.

Visitatore è ogni persona che si trasferisce in un luogo fuori da quelli solitamente frequentati per un periodo di durata inferiore ad un anno e per un motivo (principale) diverso dall'esercizio di attività remunerata nel luogo visitato.

Turista: visitatore che trascorre almeno una notte nel luogo visitato (e meno di 1 anno). I turisti possono praticare:

- Turismo di vacanza: durata di almeno 4 notti
- Breve turismo di vacanza: durata da 1 a 3 notti

Escursionista: visitatore che non trascorre neppure una notte nel luogo visitato. Gli escursionisti possono classificarsi in:

- veri (dal luogo di residenza)
- indiretti (dal luogo di vacanza)
- in transito (visita a Venezia e Firenze con fermata a Bologna)

- falsi: alloggiano in prossimità del luogo visitato (per motivi economici o di disponibilità)
2. *Quali flussi compongono il turismo internazionale? In Italia, quali fonti statistiche ufficiali possono essere utilizzate per studiare i flussi turistici internazionali?*
 - Il turismo internazionale è composto da due flussi: il flusso verso l'interno da parte degli stranieri (*ingoing* o *inbound*) e il flusso verso l'estero da parte dei nazionali (turismo dei nazionali all'estero o *outgoing* o *outbound*).
 - I flussi turistici internazionali sono rilevati alla frontiera dall'UIC. Una stima dei flussi *outgoing* si ha anche dall'Indagine ISTAT sui Viaggi e le vacanze degli italiani, mentre la rilevazione dei flussi di turisti stranieri in Italia che soggiornano in esercizi ricettivi registrati al REC si ha dalla rilevazione del Movimento dei turisti dell'ISTAT.
 3. *Quali flussi compongono il turismo dei nazionali? Esistono in Italia fonti statistiche ufficiali che consentano di conoscere i flussi turistici dei nazionali? In caso affermativo, indicare quali.*
 - Il turismo dei nazionali è composto da due flussi: il *flusso di turisti italiani in Italia* (turismo domestico) e il flusso verso l'estero da parte dei nazionali (turismo dei nazionali all'estero o *outgoing* o *outbound*).
 - Il turismo domestico è rilevato: (i) dal lato della domanda, dall'Indagine ISTAT sui Viaggi e le vacanze degli italiani; (b) dal lato dell'offerta, solo relativamente ai turisti italiani che soggiornano in esercizi ricettivi registrati al REC, dalla rilevazione del Movimento dei turisti dell'ISTAT.
 4. *Che cosa si intende per turismo interno?*
 - Il turismo interno è composto da due flussi: il *flusso di turisti italiani in Italia* (turismo domestico) e il flusso degli stranieri verso l'Italia (turismo *ingoing* o *inbound*).

NB: nel correggere il compito ho considerato corrette anche le risposte che definivano il turismo interno come coincidente con il turismo domestico, dato che in alcuni testi si ritrova questa definizione.
 5. *Indicare la cadenza e le principali informazioni raccolte dall'indagine sui Viaggi in Italia e all'estero effettuata dall'ISTAT*

L'indagine ha cadenza *trimestrale*, e i risultati sono resi noti con periodicità annuale. Le informazioni raccolte riguardano caratteristiche rilevanti di: *Viaggi, Notti* (trascorse in vacanza) e persone che hanno effettuato il viaggio.

Con riferimento ai viaggi, per ciascun viaggio effettuato vengono rilevati: il trimestre svolgimento del viaggio, la tipologia del viaggio (lavoro o vacanza), durata della vacanza (in notti), motivo prevalente del viaggio, il tipo della vacanza di piacere/svago (mare, montagna, ecc.), la destinazione del viaggio (provincia o regione, il mezzo di trasporto prevalente, l'alloggio prevalente e altre informazioni).

Con riferimento alle persone che hanno svolto il viaggio, l'indagine rileva: età e sesso, condizione professionale, posizione nella professione, ripartizione geografica di residenza, tipologia socio-demografica del comune di residenza.

6. *La Rilevazione sul movimento dei clienti negli esercizi ricettivi*
 - a. *è un'indagine campionaria o esaustiva?* È un'indagine esaustiva
 - b. *quali entità rileva?* Arrivi, presenze e permanenza media dei clienti negli alberghi, negli esercizi complementari e negli alloggi in affitto, per tipo, categoria di esercizio, tipologia di località, paese di residenza dei clienti stranieri e provincia e/o regione di provenienza dei clienti italiani.
 - c. *quali sono i principali problemi legati a questa indagine?*

- Poiché l'obbligatorietà di registrazione sussiste solo per arrivi, notti e presenze non sono affidabili;
- Sottostima dei flussi a causa dell'evasione
- Sfuggono clienti di esercizi non iscritti al REC (case in affitto, seconde case)
- I clienti degli esercizi ricettivi non sono tutti turisti. Poiché si conosce il MOTIVO della presenza, è difficile distinguere quali flussi sono turistici.
- L'indagine non rileva l'escursionismo.