BANCA D'ITALIA
EUROSISTEMA

10 ESPERTI CON ORIENTAMENTO NELLE DISCIPLINE STATISTICHE
lett. C dell'art. 1 del bando del 16 novembre 2022

Testo n. 2

Due quesiti a scelta - tra tre proposti dalla Commissione – sulla seguente materia.

Statistica e probabilità

QUESITO N. 1

1. Un ricercatore vuole svolgere un'indagine sul reddito degli italiani per valutare l'entità della disuguaglianza nella popolazione. Si consideri, a titolo esemplificativo, un insieme di cinque unità statistiche $\{u_1 \dots u_5\}$ per le quali viene osservato il reddito annuo lordo (in migliaia di euro) riportato in Tabella 1:

Tabella 1: redditi in migliaia di euro osservati su cinque unità statistiche indipendenti

u_1	u_2	u_3	u_4	u_5
100	55	70	85	60

La candidata/il candidato definisca in quali condizioni si abbia massima concentrazione o equidistribuzione del reddito; derivi l'indice di concentrazione di Gini, rappresenti graficamente la curva di Lorenz e commenti i risultati.

2. Assumendo che il reddito nella popolazione si distribuisca come una densità di Pareto:

$$f(x;\alpha) = \frac{\alpha}{x^{\alpha+1}}$$

con $\alpha \in [0,\infty)$ e $x \geq 1$, la candidata/il candidato determini lo stimatore $\hat{\alpha}_{ML}$ di massima verosimiglianza per il parametro α e calcoli la stima di massima verosimiglianza sulla base dei valori osservati in Tabella 1.

3. Si consideri ora un campione casuale $\{X_1 \dots X_n\}$ i.i.d. con funzione di densità $f(x;\theta)$ e funzione di verosimiglianza $L(\theta;x)$ con un massimo in $\hat{\theta}_{ML}$. Sapendo che la funzione di verosimiglianza relativa è definita come $\bar{L}(\theta;x) = \frac{L(\theta;x)}{L(\hat{\theta}_{ML},x)}$, la candidata/il candidato:
   a. definisca sotto quali condizioni è possibile l'approssimazione Normale di $\bar{L}(\theta;x)$ e ne fornisca la dimostrazione;
   b. calcoli $\bar{L}(\theta;x)$ nel caso di un campione casuale Normale con media μ incognita e varianza σ^2 nota; calcoli inoltre l'intervallo di verosimiglianza per μ di livello $q \in (0,1)$.

QUESITO N. 2

1. Un'azienda sta considerando di incorporare una nuova componente elettronica nei suoi processi produttivi. Per valutare la convenienza di questa strategia, l'azienda acquista n unità di questa componente, sapendo a priori che l'1% delle unità acquistate risulterà difettosa e che la probabilità che la componente smetta di funzionare durante l'utilizzo quando difettosa è del 95%. La probabilità che la componente smetta di funzionare durante l'utilizzo quando invece non ha difetti è del 2%. La candidata/il candidato calcoli la probabilità che la componente fosse effettivamente difettosa dopo che ha smesso di funzionare.
2. Data la variabile aleatoria X uniforme nell'intervallo $[0, \theta]$, si supponga di aver osservato un valore x. Si vuole effettuare il seguente test di ipotesi:

$$\begin{cases} H_0 : \theta = 1{,}9 \\ H_1 : \theta = 2{,}4 \end{cases}$$

   L'ipotesi nulla H_0 viene rifiutata se $x \leq 0{,}3$.
   La candidata/il candidato definisca e calcoli la probabilità dell'errore di seconda specie e la potenza del test.

3. Si consideri ora una variabile aleatoria X assolutamente continua con densità $f(x)$, con $x \in (-\infty, +\infty)$. La candidata/il candidato:
   a. definisca la corrispondente famiglia di posizione-scala $g(x;\, \mu, \sigma)$ con parametri di posizione $\mu \in R$ e scala $\sigma \in R^+$;
   b. descriva le differenze tra le distribuzioni $f(x)$ e $g(x;\, \mu, \sigma)$;
   c. dimostri che $g(x;\, \mu, \sigma)$ è una densità di probabilità.

QUESITO N. 3

1. La candidata/il candidato illustri le proprietà desiderabili per gli stimatori di un parametro della popolazione, includendo anche quelle asintotiche.
2. Si considerino due campioni casuali indipendenti X e Y provenienti da popolazioni di Poisson con parametro θ, con funzioni di densità, rispettivamente:

$$f(x) = \frac{e^{-\theta}\theta^x}{x!}$$
$$f(y) = \frac{e^{-\theta}\theta^y}{y!}$$

   Supponendo di aver osservato i seguenti campioni: $x = \{0;\ 2;\ 6;\ 8;\ 0\}$ e $y = \{3;\ 3;\ 3;\ 3;\ 4\}$, la candidata/il candidato enunci il principio forte di verosimiglianza e lo verifichi sui campioni osservati.

3. Date X e Y due variabili casuali uniformi continue e indipendenti aventi supporto in $[0,1]$, la candidata/il candidato:
   a. derivi la funzione di densità per la variabile casuale $Q = X + Y$ e la rappresenti graficamente;
   b. definisca la distribuzione limite di una variabile casuale data dalla somma di n variabili casuali i.i.d. uniformi continue in $[0,1]$.

Un quesito a scelta - tra due proposti dalla Commissione – sulla seguente materia.

Econometria e *statistical learning*

QUESITO N. 4

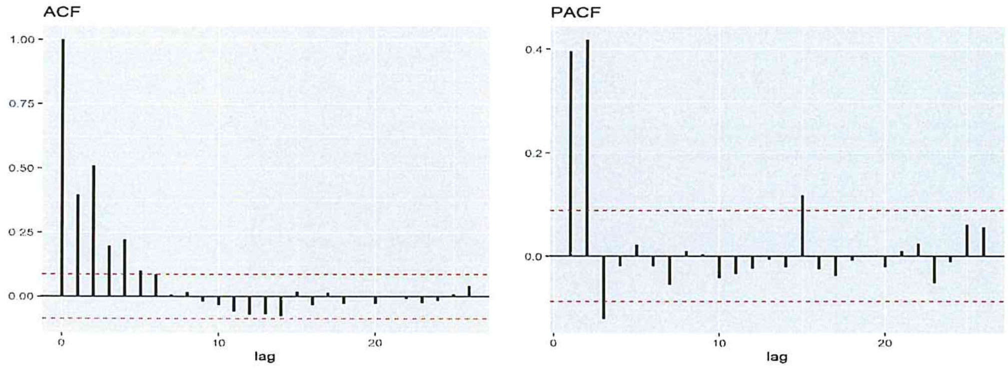
1. La candidata/il candidato illustri i concetti di stazionarietà debole e di ergodicità per dati in serie storica. Derivi inoltre la media, la varianza e le funzioni di autocovarianza e di autocorrelazione globale per il seguente processo $MA(1)$:

$$y_t = \mu + \varepsilon_t + \varphi_1 \varepsilon_{t-1} \text{ con } \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

Determini inoltre le condizioni di invertibilità di questo processo e ne verifichi la stazionarietà.

2. La serie storica $\{X_t\}$ è caratterizzata dalle funzioni di autocorrelazione globale (ACF) e autocorrelazione parziale (PACF) riportate in Figura 1.

Figura 1: funzioni di autocorrelazione globale e parziale



   a. Supponendo che la serie sia stazionaria in covarianza, quale modello nella classe dei modelli ARMA la candidata/il candidato suggerirebbe per la serie $\{X_t\}$?

   b. Sia y_t un processo stocastico I(1) per il quale Δy_t è rappresentabile tramite un modello ARMA(1,1). La candidata/il candidato utilizzi la scomposizione di *Beveridge-Nelson* per derivare da y_t la componente non stazionaria e quella I(0) responsabile delle oscillazioni di breve periodo.

3. La candidata/il candidato discuta della cointegrazione tra serie storiche ed esponga l'utilità del teorema di rappresentazione di *Granger* in questo contesto. Illustri inoltre come determinare il rango di cointegrazione con la procedura di *Johansen*.

QUESITO N. 5

1. Dato il modello di regressione multipla

$$y_i = \beta_0 + \sum_{j=1}^{K} x_{ij}\beta_j + \varepsilon_i$$

dove $(y_1, \dots, y_n)$ rappresenta una variabile numerica osservata su un campione di n individui e $\varepsilon_i \sim N(0, \sigma^2)$ con $i = 1, \dots, n$ e K il numero di variabili esplicative. La candidata/il candidato derivi lo stimatore **b** del vettore dei parametri $\boldsymbol{\beta}$ con il metodo dei minimi quadrati e ne descriva le proprietà.
Nell'ipotesi che $x_i = z_i\gamma + v_i$, con $Z = (z_1, \dots, z_n)$ variabile non osservabile, e che $E(\varepsilon|Z) = 0$, la candidata/il candidato indichi i limiti dello stimatore **b** e suggerisca possibili soluzioni.

2. Si consideri il seguente modello di regressione lineare del salario (*wage*) in funzione degli anni di scuola (*educ*), a meno di errori casuali:

$$wage_i = \beta_0 + \beta_1 educ_i + \varepsilon_i$$

La Tabella 1 riporta la stima dei parametri e i relativi errori standard del modello, sotto le usuali ipotesi.

Tabella 1: risultati del modello stimato

	Dependent variable:
	wage
Constant	0.578
	(0.603)
educ	0.553***
	(0.047)
Observations	753
R^2	0.156
Adjusted R^2	0.155
Residual Std. Error	3.890 (df = 751)
F Statistic	138.558*** (df = 1; 751)
Durbin-Watson D	1.8115***
Note:	*p<0.1; **p<0.05; ***p<0.01

La candidata/il candidato:

a. interpreti i risultati del modello;
b. discuta di eventuali problemi di endogeneità e di come potrebbe essere migliorato in questo caso lo stimatore OLS.

3. Su un campione di n lavoratrici, si vuole stimare la relazione esistente tra salario percepito e ore lavorate. A tal fine si consideri il seguente modello a due equazioni simultanee:

$$\begin{cases} hours_i = \beta_0 + \beta_1 \ln(wage)_i + \beta_2 educ_i + \beta_3 age_i + \beta_4 k5_i + \varepsilon_i \\ \ln(wage)_i = \gamma_0 + \gamma_1 hours_i + \gamma_2 educ_i + \gamma_3 educ_i^2 + v_i \end{cases}$$

con $i = 1, \dots, n$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$, $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$, $E(\varepsilon_i v_j) = \omega_{ij}$ con $j = 1, \dots, n$ e dove

- *hours* rappresenta le ore lavorate;
- *wage* rappresenta il salario percepito dalle donne che lavorano;
- *educ* rappresenta il livello di istruzione espresso in anni di scuola;
- *age* rappresenta l'età espressa in anni;
- *k5* rappresenta il numero di figli con età inferiore a 6 anni.

La candidata/il candidato suggerisca quale metodologia utilizzerebbe per stimare i parametri del modello di cui sopra.

Un quesito a scelta - tra due proposti dalla Commissione – sulla seguente materia.

Metodi di campionamento

QUESITO N. 6

Si consideri un comune suddiviso in diversi municipi. Per conoscere la proporzione di persone residenti in un municipio che utilizzano regolarmente un'auto in *car sharing* per i propri spostamenti, si esegue una rilevazione campionaria selezionando in modo casuale e senza ripetizione 1.000 individui tra i 50.000 residenti nel municipio.

La candidata/il candidato:

a. costruisca l'intervallo di confidenza per la proporzione dei residenti del municipio che utilizzano un'auto in *car sharing*, supponendo che la proporzione stimata nel campione sia pari al 20 per cento. Spieghi in che modo l'intervallo di confidenza cambia al variare della numerosità della popolazione del municipio e della proporzione stessa;
b. supponga di non possedere l'elenco esaustivo dei residenti del municipio, ma di avere a disposizione una delle seguenti liste da cui selezionare il campione:
   i. la lista dei residenti dell'intero comune, di numerosità pari a 300.000 unità, per le quali non è noto il municipio di residenza;
   ii. la lista dei soli occupati residenti nel municipio di interesse, di numerosità pari a 30.000 unità.

   Discuta quali problemi possono derivare dall'utilizzo di ciascuna delle due liste per selezionare le unità statistiche;
c. spieghi in che modo potrebbe condurre la rilevazione volendo stimare la proporzione di chi utilizza un'auto in *car sharing* per fasce d'età, supponendo di avere a disposizione la lista dei residenti del municipio e la distribuzione degli stessi per classi di età.

QUESITO N. 7

Si vuole condurre un'indagine sulle famiglie italiane volta a raccogliere informazioni sulle loro condizioni socio-demografiche ed economiche. Si adotta un disegno campionario stratificato a due stadi: nel primo stadio viene selezionato un campione di comuni e, in quello successivo, le famiglie sono estratte casualmente dai registri comunali dei comuni selezionati.

Il candidato/la candidata:

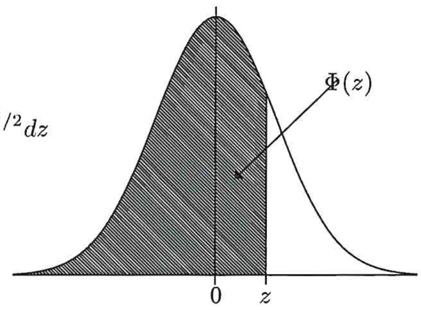
a. spieghi come selezionerebbe le unità di primo stadio nell'indagine considerata. Illustri in generale le caratteristiche di un disegno campionario stratificato a più stadi, descrivendo anche le probabilità di inclusione delle unità statistiche;
b. illustri le implicazioni sulle stime del reddito e della ricchezza che derivano dalla ridotta presenza nel campione di famiglie benestanti e le metodologie utilizzabili per mitigare questo problema;
c. sapendo che la rilevazione è condotta con cadenza biennale e che circa la metà del campione è di tipo *panel*, spieghi come si può utilizzare il dato osservato nella rilevazione precedente sulle unità facenti parte della componente *panel* per la stima (corrente) del reddito.

Prova in lingua inglese

Regular physical activity is proven to help maintain healthy body weight and can improve mental health, quality of life, and well-being. However, current global estimates show one in four adults do not do enough physical activity. Furthermore, as countries develop economically, levels of inactivity rise due to changing transport patterns, increased use of technology for work and recreation.
What changes would you recommend to reduce sedentary behaviours and increase levels of physical activity?

Tavola 1: Funzione di ripartizione della Variabile Casuale Normale Standardizzata

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Tavola 1 *(segue)*: Funzione di ripartizione della Variabile Casuale Normale Standardizzata

z	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
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Tavola 1a: Valori critici della Variabile Casuale Normale Standardizzata. $P(Z \geq z_\alpha) = \alpha$.

α	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001	
z_α	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905	3.7190	
α	0.00009	0.00008	0.00007	0.00006	0.00005	0.00004	0.00003	0.00002	0.00001
z_α	3.7455	3.7750	3.8082	3.8461	3.8906	3.9444	4.0128	4.1075	4.2649