

Statistica Applicata

prof. Federico Plazzi

Corso di Laurea in Scienze Naturali

a. a. 2015/2016

PROVA DEL 25 GENNAIO 2016

Nome: _____

Cognome: _____

ALCUNE INDICAZIONI:

- La prova consiste in cinque esercizi; dopo ogni esercizio c'è lo spazio in cui scrivere la risposta o le risposte. In caso questo spazio non sia sufficiente, si può continuare a rispondere sul retro del foglio, avendo cura di indicare il numero dell'esercizio a fianco della continuazione della risposta.
- Alcuni esercizi richiedono semplici calcoli, per i quali è consentito l'uso di una calcolatrice ed eventualmente la consultazione di una o più delle tabelle allegate.
- Altri esercizi richiedono invece la lettura dei dati: verrà valutata in questo caso l'argomentazione che giustifica l'interpretazione fornita.
- La durata massima della prova è di 60 minuti.
- Si prega di non scrivere nulla sulle tabelle allegate.

1 Descrizione dei dati

Di seguito sono riportate alcune statistiche che si riferiscono al genoma mitocondriale di alcuni individui di molluschi bivalvi. Nella prima colonna (“Specie”) c’è il nome della specie, nella seconda (“Sottoclasse”) la sottoclasse alla quale la specie appartiene, nella terza (“Lunghezza (pb)”) la lunghezza del genoma mitocondriale (misurata in paia di basi) e nella quarta (“%UR”) la percentuale di queste paia di basi alle quali non siamo a tutt’oggi in grado di assegnare una funzione e che consideriamo perciò come *Unassigned Region* (*UR*, appunto).

2 Dati

Tabella 1: Lunghezza e percentuale di regioni senza funzione nota nei genomi mitocondriali dei bivalvi.

Specie	Sottoclasse	Lunghezza (pb)	%UR
<i>Laternula elliptica</i>	Anomalodesmati	14622	1.44
<i>Nucula nucleus</i>	Opponobranchi	13671	16.22
<i>Solemya velum</i>	Opponobranchi	15660	2.13
<i>Cuspidaria rostrata</i>	Anomalodesmati	28228	18.48
<i>Ostrea lurida</i>	Pteriomorfi	16344	10.39
<i>Crassostrea angulata</i>	Pteriomorfi	18225	9.82
<i>Acanthocardia tuberculata</i>	Eterodonti	16104	9.58
<i>Lamprotula tortuosa</i>	Paleoeterodonti	15722	6.66
<i>Mimachlamys senatoria</i>	Pteriomorfi	17383	11.11
<i>Crassostrea virginica</i>	Pteriomorfi	17244	12.67
<i>Margaritifera falcata</i>	Paleoeterodonti	16121	5.92
<i>Lamprotula gottschei</i>	Paleoeterodonti	15915	7.39
<i>Mytilus californianus</i>	Pteriomorfi	16190	5.3
<i>Pyganodon grandis</i>	Paleoeterodonti	17071	8.16
<i>Lasmigona compressa</i>	Paleoeterodonti	15903	4.31
<i>Arconaia lanceolata</i>	Paleoeterodonti	15782	10.07
<i>Quadrula quadrula</i>	Paleoeterodonti	16970	7.04
<i>Utterbackia imbecillis</i>	Paleoeterodonti	16103	4.49
<i>Hyriopsis cumingii</i>	Paleoeterodonti	15954	6.88
<i>Paphia undulata</i>	Eterodonti	18154	11.99
<i>Meretrix petechialis</i>	Eterodonti	19567	16.72
<i>Meretrix meretrix</i>	Eterodonti	19826	17.46
<i>Atrina pectinata</i>	Pteriomorfi	16811	12.25
<i>Perna viridis</i>	Pteriomorfi	16014	8.63
<i>Placopecten magellanicus</i>	Pteriomorfi	32115	51.04

3 Esercizi

3.1 Devianza, varianza e deviazione standard

Calcolare devianza, varianza e deviazione standard della lunghezza dei genomi mitocondriali delle specie appartenenti alla sottoclasse degli eterodonti.

3.2 Biodiversità

La tabella seguente mostra il numero di generi noti per le tre sottoclassi principali.

Tabella 2: Numero di generi noti per sottoclasse.

Sottoclasse	Generi
Eterodonti	1668
Paleoeterodonti	500
Pteriomorfi	933

Assumendo che il numero di specie sia nella stessa proporzione, verificare *utilizzando la tabella appropriata tra quelle fornite* se le tre sottoclassi sono rappresentate nelle quantità attese nel campione preso in esame.

In altre parole, nel campione in esame eterodonti, paleoeterodonti e pteriomorfi sono nella proporzione di 4:9:8 - quale test statistico serve per capire se questa situazione è significativamente diversa da quella che ci si potrebbe aspettare in base alla tabella qui sopra? Qual è il risultato?

3.3 Distribuzione dei dati

Di seguito è riportato l'output di R dopo aver effettuato un test di Shapiro e Wilk su tutte e 25 le lunghezze dei genomi e su tutte e 25 le percentuali di regioni senza funzione nota.

Shapiro-Wilk test

```
data: Lunghezza (pb)
W = 0.6195, p-value = 7.139e-07
```

```
data: %UR
W = 0.6629, p-value = 2.403e-06
```

Cosa si può concludere circa la distribuzione di queste due variabili? Perché?

3.4 Confronti

Dato che è importante stabilire se i genomi dei paleoeterodonti siano significativamente più corti dei genomi degli pteriomorfi, sono stati eseguiti tre diversi test, il cui risultato è riportato di seguito. I paleoeterodonti sono sempre per primi e tutti i test sono ad una coda.

Two Sample t-test

```
data: Lunghezza dei paleoeterodonti e lunghezza degli pteriomorfi
t = -1.3587, df = 7.105, p-value = 0.1079
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf 1025.09
sample estimates:
mean of x mean of y
16171.22 18790.75
```

Wilcoxon test

data: Lunghezza dei paleoeterodonti e lunghezza degli pteriomorfi
V = 5, p-value = 0.03906
alternative hypothesis: true location shift is less than 0

Mann-Whitney test

data: Lunghezza dei paleoeterodonti e lunghezza degli pteriomorfi
W = 10, p-value = 0.005553
alternative hypothesis: true location shift is less than 0

Quale di questi test è il più appropriato? Cosa ne possiamo concludere? Perché?

3.5 Correlazioni

Di seguito sono riportati alcuni risultati relativi al calcolo di una correlazione lineare tra la lunghezza del genoma la percentuale di regioni senza funzione nota.

Tabella 3: Modello di correlazione lineare tra lunghezza del genoma e percentuale di regioni a funzione non nota.

	Stima	p-value	r	R^2
Intercetta	-2.422e+01			
Pendenza	1.996e-03	8.779e-08	0.8479	0.7189

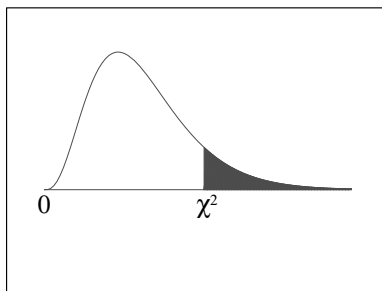
C'è correlazione tra le due variabili in questione? Perché?

TABLE A.2

t Distribution: Critical Values of t

<i>Degrees of freedom</i>	<i>Two-tailed test: One-tailed test:</i>	<i>Significance level</i>					
		10% 5%	5% 2.5%	2% 1%	1% 0.5%	0.2% 0.1%	0.1% 0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646
32		1.694	2.037	2.449	2.738	3.365	3.622
34		1.691	2.032	2.441	2.728	3.348	3.601
36		1.688	2.028	2.434	2.719	3.333	3.582
38		1.686	2.024	2.429	2.712	3.319	3.566
40		1.684	2.021	2.423	2.704	3.307	3.551
42		1.682	2.018	2.418	2.698	3.296	3.538
44		1.680	2.015	2.414	2.692	3.286	3.526
46		1.679	2.013	2.410	2.687	3.277	3.515
48		1.677	2.011	2.407	2.682	3.269	3.505
50		1.676	2.009	2.403	2.678	3.261	3.496
60		1.671	2.000	2.390	2.660	3.232	3.460
70		1.667	1.994	2.381	2.648	3.211	3.435
80		1.664	1.990	2.374	2.639	3.195	3.416
90		1.662	1.987	2.368	2.632	3.183	3.402
100		1.660	1.984	2.364	2.626	3.174	3.390
120		1.658	1.980	2.358	2.617	3.160	3.373
150		1.655	1.976	2.351	2.609	3.145	3.357
200		1.653	1.972	2.345	2.601	3.131	3.340
300		1.650	1.968	2.339	2.592	3.118	3.323
400		1.649	1.966	2.336	2.588	3.111	3.315
500		1.648	1.965	2.334	2.586	3.107	3.310
600		1.647	1.964	2.333	2.584	3.104	3.307
∞		1.645	1.960	2.326	2.576	3.090	3.291

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_{\alpha}$.

<i>df</i>	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Table entry for p is the critical value F^* with probability p lying to its right.

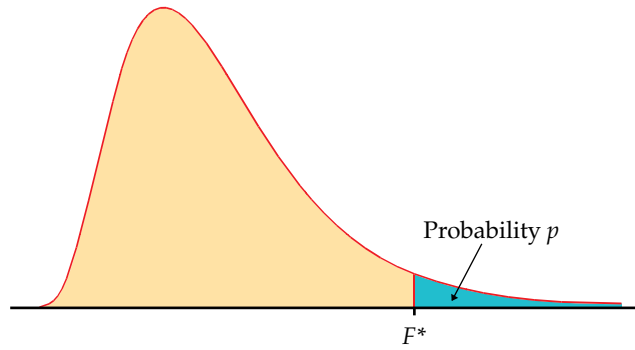


TABLE E

F critical values

		Degrees of freedom in the numerator									
p		1	2	3	4	5	6	7	8	9	
Degrees of freedom in the denominator	1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
		.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
		.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
		.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
		.001	405284	500000	540379	562500	576405	585937	592873	598144	602284
	2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
		.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
		.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
		.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
		.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
	3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
		.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
		.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
		.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
		.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
	4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
		.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
		.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
		.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
		.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
5	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24	
6	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69	
7	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	
	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33	