AMSB 9/1/2017 Test #1
Si ricorda nell'effettuare un test di specificare ogni volta le ipotesi nulla e alternativa.

Es 1 (tutti).
Si considerino i seguenti dati

-3.8 -2.7 7.4 -0.5 0.8 4.8 7.3 6.5 0.6 -2.2

1) Si costruisca l'istogramma suddividendo in 3 classi [-5, 0[, [0,5[, [5,10]. Se ne faccia il grafico. Si stimino la media pesata e la varianza pesata
2) Si riprendano i dati di partenza, si stimi la grandezza di dispersione che si ritiene più adeguata nel caso in cui i dati non derivino da una distribuzione gaussiana.
3) Si stimi un parametro che descriva la simmetria del dato.
4) In riferimento al punto 3) tale indice avrebbe senso nel caso di distribuzione dei dati di tipo gaussiano?

Per questo punto fornirò solo le indicazioni a mio avviso rilevanti. Si puntualizza che la media pesata deve tenere conto delle frequenze relative e utilizzare i punti centrali degli intervalli (-2.5 2.5 7.5).
Punto 4). Se siamo sicuri che i dati sono gaussiani, allora questa parametro perde interesse. In generale può essere calcolato, ma comunque ci aspettiamo un valore con aspettazione nulla. Per questo punto bastava osservare questa proprietà.

Es 2 (tutti)
Un filtro chimico viene applicato ad un sistema per la produzione di un componente. Prima della sua installazione il livello medio di impurezza misurato su un campione di $n_1 = 8$ elementi è pari $y_1 = 12.5$ con una varianza $s_1^2 = 101.7$. Dopo l'installazione si trovano i seguenti valori su un secondo campione di $n_2 = 9$ elementi, $y_2 = 10.2$ con una varianza $s_2^2 = 94.73$. Dopo aver verificato se le varianze sono uguali ($\alpha = 0.05$) si dica se il filtro ha portato o meno benefici ($\alpha = 0.05$). Si discuta brevemente la rilevanza del test per l'uguaglianza delle varianze.

Una possibile soluzione prevede la verifica dell'uguaglianza delle varianze con un test di Fisher. Infatti è importante per verificare la condizione di omogeneità delle varianze per una corretta esecuzione del test successivo
In particolare poteva essere usato un t test per dati indipendenti note le varianze campionarie e non quelle delle popolazioni.

Es 3 (tutti)
In un test si misura il livello di connettività tra due regioni cerebrali in funzione di due stimoli differenti in uno studio molto ampio che coinvolge 219 soggetti. Su tutti i soggetti sono stati studiati gli effetti dei due stimoli e si sono viste singolarmente le differenze. Con lo stimolo 1 si trova una connettività significativa in 119 soggetti, con lo stimolo 2 in la significatività si ha in 158 soggetti. Analizzando in dettaglio le variazioni si vede che 84 soggetti hanno ottenuto una risposta significativa sia nel primo che nel secondo caso, mentre su 26 soggetti non è stata osservata significatività in nessuno dei due stimoli. Si può dire che i due stimoli producano un effetto differente ($\alpha = 0.05$)?

Siamo in presenza di dati appaiati: è fondamentale usare un test di Mc Nemar e non uno del chi quadro. Questa confusione porta ad una penalizzazione significativa nel punteggio. Il test di Mc Nemar si esegue a partire da questa tabella.

		Secondo Stimolo		
		Signif.	Non. Signif	
Primo Stimolo	Signif.	84	35	119
	Non Signif.	74	26	100
		158	61	219

Es 4 **Versione 1** (tutti)
Un parametro per la misura della variabilità del parlato la deviazione standard della Frequenza fondamentale (F0) dei suoni vocalizzati. Tale parametro può variare in funzione dell'umore del parlante. Si considerano due gruppi di soggetti, il primo gruppo di soggetti in fase maniacale il secondo in fase depressiva e si misurano i seguenti valori. È noto che tale parametro non possiede una distribuzione di tipo gaussiano. Si dica se tale parametro è differente nelle due fasi ($\alpha = 0.05$).

Man	Dep
20	14
12	15
16	18
23	13
30	12
15	10
12	9
11	21

In questo caso si esegue il test di Mann Whitney. Si può fare l'approssimazione per grandi numeri.

4 **Versione 2** (tutti)
Un parametro per la misura della variabilità del parlato la deviazione standard della Frequenza fondamentale (F0) dei suoni vocalizzati. Tale parametro può variare in funzione dell'umore del parlante. Si considerano un gruppo di 8 soggetti affetti da disturbo bipolare. In una prima misura i soggetti sono in fase maniacale e in una seconda in fase depressiva. È noto che tale parametro non possiede una distribuzione di tipo gaussiano. Si dica se tale parametro è differente nelle due fasi ($\alpha = 0.05$).

(stessa tabella)

In questo caso si esegue il test di Wilcoxon con segno. Si può fare l'approssimazione per grandi numeri.
In entrambi i casi l'importante è la determinazione del tipo di test. Nei casi nei quali il test non è stato individuato correttamente, la valutazione è stata significativamente diminuita. L'individuazione dell'intervallo critico non ha pesato nella valutazione (la tabella fornita andava bene per l'intervallo delle somme dei ranghi positivi). È stata anche accettata l'approssimazione con la distribuzione normale.

Es 5 (tutti)
Il volume espiratorio forzato in un secondo, è stato misurato in tre ospedali differenti su soggetti affetti da malattia coronarica. I tre gruppi hanno uguale numerosità (n=23 soggetti in ciascun ospedale).
Verificare se esiste una differenza significativa tra le misure effettuate nei tre ospedali ($\alpha = 0.05$). Discutere come effettuare un'analisi post hoc e perché potrebbe essere di interesse

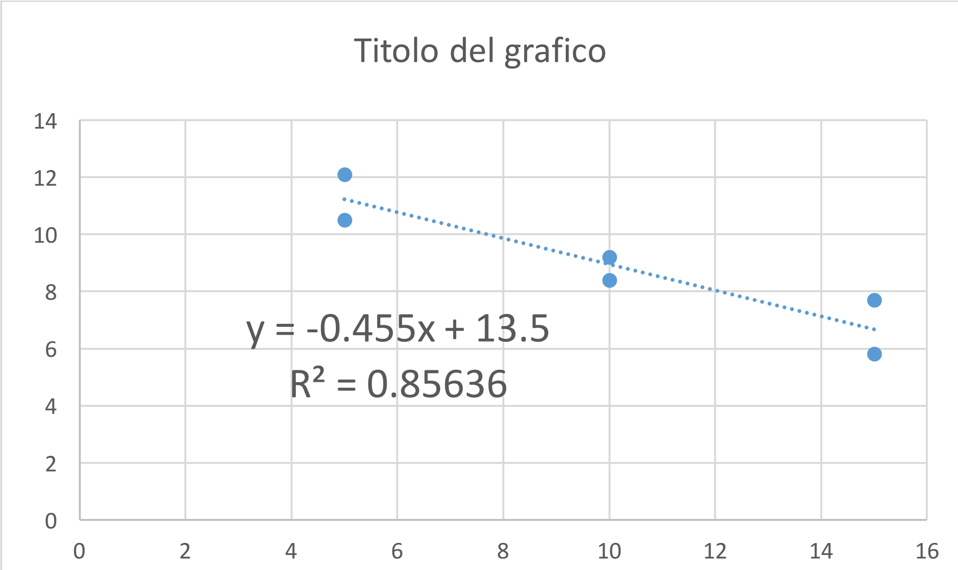
	Osp. 1	Osp. 2	Osp. 3
Media	2.63	3.03	2.88
Dev. Standard	0.496	0.523	0.498

Non verranno fornite indicazioni metodologiche, ampiamente documentate. Chi non ha specificato informazioni sull'analisi post-hoc ha subito una penalizzazione (motivazioni, tipo di test, significatività).

Es 6 (2016-2017)
Nella tabella seguente sono mostrare le misure in un campione biologico di un inquinante in funzione della dose ambientale. Si vuole stimare un modello di regressione che leghi la dose ambientale alla concentrazione (C) nel campione. Si verifichi se il coefficiente che lega variabile dipendente e indipendente sia significativamente diverso da zero ($\alpha = 0.05$).

		Dose		
		5	10	15
misure	1	10,5	8,4	7,7
	2	12,1	9,2	5,8

È molto importante impostare correttamente il modello. La dose è la variabile indipendente. Per ciascun valore di x, in questo caso sono riportati due valori della variabile indipendente.



Nel caso di significatività del modello (altra versione) poteva essere calcolata anche la statistica F=MSR/MSE. Nel seguito alcune statistiche di interesse. Sono state ottenute con excel. Sono segnate in grassetto quelle che sono state trattate durante il corso

OUTPUT RIEPILOGO

Statistica della regressione	
R multiplo	0.925397145
R al quadrato	**0.856359876**
R al quadrato corretto	0.820449845
Errore standard	**0.931732258**
Osservazioni	6

ANALISI VARIANZA

	gdl	*SQ*	*MQ*	*F*	*Significatività F*
Regressione	**1**	**20.7025**	**20.7025**	**23.85**	**0.0081**
Residuo	**4**	**3.4725**	**0.8681**		
Totale	**5**	**24.1750**			

	Coefficienti	*Errore standard*	*Stat t*	*Valore di significatività*
Intercetta	**13.5**	1.0064	13.4143	0.000179
Variabile X 1	**-0.455**	**0.0932**	**-4.8834**	**0.008141**

D1 (tutti)
Discutere perché la distribuzione normale assume un ruolo rilevante nell'analisi statistica e come l'assunzione di normalità incida sulla scelta di un test statistico.

D2 (tutti)
Descrivere come vengono determinati i valori critici nel test di Mann-Whitney.

D3 (2016-2017)
Definire la funzione di distribuzione e suo legame con la densità di probabilità. Fare il grafico approssimato della funzione di distribuzione di una variabile gaussiana con media pari a 1 e deviazione standard pari a 3 e confrontarlo con quella di una variabile gaussiana con stessa media e deviazione standard pari a 1.

D4 (2016-2017)
Enunciare il teorema di Bayes e il teorema delle probabilità totali. Discutere con un esempio l'utilizzo di uno o di entrambi i teoremi.