

**Statistica Applicata**  
**Corso di Laurea in Scienze Naturali**  
**a. a. 2018/2019**

prof. Federico Plazzi

6 Settembre 2019

Nome: \_\_\_\_\_

Cognome: \_\_\_\_\_

Matricola: \_\_\_\_\_

**Alcune indicazioni:**

- La prova è costituita da 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 Dati

I bruchi delle specie *Ephestia elutella* e *Plodia interpunctella* (Lepidoptera: Pyralidae) richiedono spesso disinfestazioni negli stabilimenti alimentari: si nutrono infatti di cereali, semi, frutta secca, pasta e altri fonti alimentari simili, fino a divenire infestanti. Possono essere trattati con pesticidi (un metodo in qualche modo “tradizionale”), ma esiste anche la tecnica della cosiddetta *mating disruption*, ossia l’intervento sulle capacità delle farfalle di localizzare un partner per l’accoppiamento, tipicamente per mezzo di speciali feromoni di sintesi che confondono gli insetti adulti. La tabella 1 presenta il numero di individui catturati nel corso delle ricognizioni mensili in due stabilimenti alimentari: la *mating disruption* è stata introdotta in entrambi gli stabilimenti a partire da Gennaio 2015 (compreso).

Tabella 1: Numero di catture nei due stabilimenti. La *mating disruption* è stata introdotta a partire da Gennaio 2015 (compreso).

Mese	Stabilimento 1	Stabilimento 2
gen-14	39	0
feb-14	55	18
mar-14	30	47
apr-14	34	205
mag-14	24	289
giu-14	46	213
lug-14	50	529
ago-14	54	574
set-14	0	0
ott-14	0	133
nov-14	20	127
dic-14	23	17
gen-15	22	0
feb-15	0	4
mar-15	14	13
apr-15	16	40
mag-15	26	12
giu-15	16	205
lug-15	10	19
ago-15	3	162
set-15	0	17
ott-15	8	49
nov-15	4	13
dic-15	4	0

## 2 Esercizi

### 2.1 Statistiche di base

Calcolare media, devianza, varianza e deviazione standard del numero di catture nello stabilimento 2 nel 2015.

<i>Media</i>	<i>Devianza</i>	<i>Varianza</i>	<i>Deviazione standard</i>
44,5	49655	4137,917	64,32664

### 2.2 Distribuzione dei dati

Per indagare la distribuzione dei dati, viene eseguito il test di Shapiro e Wilk su tutte le 48 catture; i risultati sono i seguenti.

Shapiro-Wilk test

```
data: catture
W = 0.56168, p-value = 9.485e-11
```

Cosa si può concludere sulla distribuzione dei dati? Perché?

*L'ipotesi nulla del test di Shapiro e Wilk è la normalità della distribuzione dei dati. Visto il valore molto basso del p-value, si può rifiutare l'ipotesi nulla e considerare non normale la distribuzione dei dati.*

### 2.3 Effetto della *mating disruption*

Per verificare l'effetto della *mating disruption* rispetto alle tecniche tradizionali, si eseguono diversi test statistici sui numeri di catture prima e dopo l'introduzione della nuova tecnica, i cui risultati sono mostrati di seguito. Qual è il test corretto? Perché? Cosa si può concludere sull'efficacia della *mating disruption*? Ha effettivamente ridotto il numero di animali (e quindi di catture)?

- Test *t* a campioni indipendenti a due code:

```
Two Sample t-test
```

```
t = 2.3175, df = 27.622, p-value = 0.20811
```

```
alternative hypothesis: true difference in means is not equal to 0
```

- Test  $t$  a campioni appaiati a due code:

Paired t-test

$t = 2.8253$ ,  $df = 23$ ,  $p\text{-value} = 0.079594$   
alternative hypothesis: true difference in means is not equal to 0

- Test  $t$  a campioni indipendenti a una coda:

Two Sample t-test

$t = 2.3175$ ,  $df = 27.622$ ,  $p\text{-value} = 0.01406$   
alternative hypothesis: true difference in means is greater than 0

- Test  $t$  a campioni appaiati a una coda:

Paired t-test

$t = 2.8253$ ,  $df = 23$ ,  $p\text{-value} = 0.004797$   
alternative hypothesis: true difference in means is greater than 0

- Test di Mann e Whitney a due code:

Wilcoxon rank sum test with continuity correction

$W = 429$ ,  $p\text{-value} = 0.00368$   
alternative hypothesis: true location shift is not equal to 0

- Test di Wilcoxon a due code:

Wilcoxon signed rank test with continuity correction

$V = 241.5$ ,  $p\text{-value} = 0.135984$   
alternative hypothesis: true location shift is not equal to 0

- Test di Mann e Whitney a una coda:

Wilcoxon rank sum test with continuity correction

$W = 429$ ,  $p\text{-value} = 0.41184$   
alternative hypothesis: true location shift is greater than 0

- Test di Wilcoxon a una coda:

Wilcoxon signed rank test with continuity correction

$V = 241.5$ ,  $p\text{-value} = 0.0001001$   
alternative hypothesis: true location shift is greater than 0

La variabile non è a distribuzione normale. I campioni non sono da considerarsi appaiati, perché i numeri di catture vanno confrontati nel loro insieme prima e dopo l'introduzione della nuova tecnica. La direzione della preferenza, inoltre, non è nota a priori. Tutto questo porta a decidere per il test di Mann e Whitney a due code. Il p-value è inferiore a 0,05 (0,00368) e quindi le differenze tra i due campioni sono da ritenersi significative. Dopo l'introduzione della mating disruption il numero di catture è effettivamente diminuito.

## 2.4 Andamento delle catture nel tempo

Un altro modo per rispondere alla domanda dell'esercizio precedente è verificare se esista una correlazione negativa tra lo scorrere del tempo (mese per mese, visto che la ricognizioni sono periodiche) e in numero di catture. Di seguito vengono mostrati grafico e parametri di un modello di correlazione lineare. Si può parlare di un calo significativo delle catture nel tempo? Perché?

Tabella 2: Modello di correlazione lineare.

	Stima	p-value	$r$	$R^2$
Intercetta	123,882			
Pendenza	-4,604	0,04925	-0,2645075	0,06996

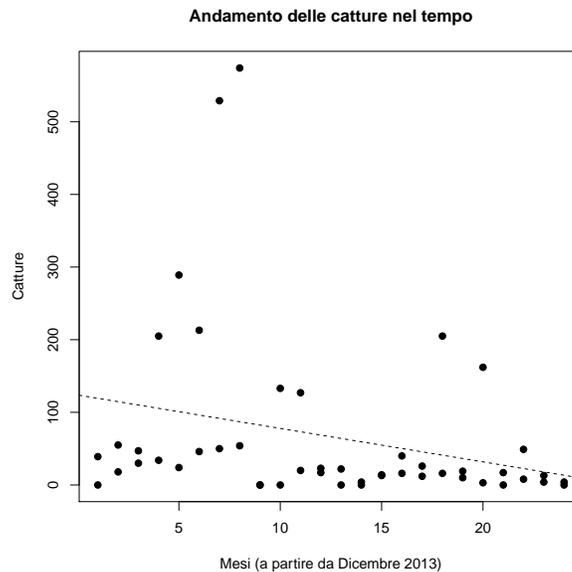


Figura 1: Catture nel tempo.

*La correlazione è significativa, ancorché di pochissimo ( $p = 0,04925$ ) e debolissima ( $R^2 = 0,06996$ ). Si tratta effettivamente di una correlazione negativa ( $r = -0,2645075$ ), per cui si può effettivamente affermare che c'è un calo delle catture nel tempo. D'altra parte, l'introduzione della mating disruption dovrebbe aver fatto calare abbastanza rapidamente le catture (come si vede dall'esercizio precedente) a cavallo tra Dicembre 2014 e Gennaio 2015, per cui in effetti un regolare calo lineare non è il modello più prevedibile in questo caso.*

## 2.5 Confronto tra stabilimenti

È possibile verificare se esistano differenze tra i due stabilimenti, ossia se ve ne sia uno più soggetto a infestazione dell'altro? Una possibilità è eseguire un test per confrontare due campioni: quale sarebbe il test corretto? A quante code? Perché?

Un'altra possibilità prevede invece di partire da una One-Way ANOVA: è un approccio ragionevole? È preferibile rispetto a un test per confrontare due campioni?

*È vero che la One-Way ANOVA permette di confrontare quattro campioni (due stabilimenti  $\times$  due trattamenti) anziché due, ma richiederebbe una distribuzione normale dei dati. I quattro gruppi verrebbero effettivamente della stessa dimensione e in queste circostanze l'ANOVA si dimostra robusta, ma bisognerebbe comunque verificare l'omoschedasticità: d'altra parte, un test non parametrico a campioni appaiati tiene comunque conto della dimensione temporale e quindi del cambio di trattamento in Gennaio 2015, perché confronta ogni dato di uno stabilimento con il dato corrispondente dell'altro. Il test di Wilcoxon è quindi l'approccio migliore: non avendo informazioni sui due stabilimenti, né quindi sulle relative facilità di infestazione, conviene effettuare un test a due code. Per la cronaca, il risultato sarebbe  $V = 49,5$ , con un  $p$ -value di  $0,007429$ , per cui si concluderebbe che effettivamente si osserva una differenza tra i due stabilimenti nel numero di catture.*