Metodi Quantitativi per Economia, Finanza e Management

Lezione n°5

Test d’Ipotesi

Test per lo studio dell'associazione tra variabili

- Nella teoria dei test, il ricercatore fornisce ipotesi riguardo la distribuzione della popolazione; tali Hp sono *parametriche* se riguardano il valore di uno ò più parametri della popolazione conoscendone la distribuzione a meno dei parametri stessi; *non parametriche* se prescindono dalla conoscenza della distribuzione della popolazione.

- ***Obiettivo dei test***: come decidere se accettare o rifiutare un'ipotesi statistica alla luce di un risultato campionario. Esistono due ipotesi: H_0 e H_1, di cui la prima è l'ipotesi nulla, la seconda l'ipotesi alternativa la quale rappresenta, di fatto, l'ipotesi che il ricercatore sta cercando di dimostrare.

Test per lo studio dell’associazione tra variabili

Cosa è un’ipotesi?

- Un’ipotesi è una affermazione (assunzione) circa il parametro della popolazione:
  - media della popolazione

Esempio: In questa città, il costo medio della bolletta mensile per il cellulare è μ = $42

L’ipotesi Nulla, H_0 rappresenta l’ipotesi che deve essere verificata, l’Ipotesi Alternativa, H_1 è generalmente l’ipotesi che il ricercatore stà cercando di dimostrare

Test per lo studio dell'associazione tra variabili

- Si può incorrere in due tipologie di errore:

Possibili Risultati Verifica di Ipotesi		
	Stato di Natura	
Decisione	H_0 Vera	H_0 Falsa
Non Rifiutare $\mathbf{H_0}$	**No errore**	**Errore Secondo Tipo**
Rifiutare $\mathbf{H_0}$	**Errore Primo Tipo**	**No Errore**

Test per lo studio dell'associazione tra variabili

- **Errore di Primo Tipo**
  - Rifiutare un'ipotesi nulla vera
  - Considerato un tipo di errore molto serio

La probabilità dell'errore di primo tipo è α

- Chiamato livello si significatività del test
- Fissato a priori dal ricercatore

Test per lo studio dell'associazione tra variabili

- **Errore di Secondo Tipo**
  - Non rifiutare un'ipotesi nulla falsa

La probabilità dell'errore di secondo tipo è β

Test per lo studio dell'associazione tra variabili

Possibili Risultati Verifica di Ipotesi

	Stato di Natura	
Decisione	H_0 Vera	H_0 Falsa
Non Rifiutare **H_0**	**No errore** **(1 - α)**	**Errore Secondo Tipo** **(β)**
Rifiutare **H_0**	**Errore Primo Tipo** **(α)**	**No Errore** **(1 - β)**

Legenda:
Risultato
(Probabilità)

Test per lo studio dell’associazione tra variabili

- Errore di primo tipo ed errore di secondo tipo non si posso verificare contemporaneamente

  - Errore di primo tipo può occorrere solo se H_0 è vera
  - Errore di secondo tipo può occorrere solo se H_0 è falsa

Se la probabilità dell’errore di primo tipo (α) $\Uparrow$,

allora la probabilità dell’errore di secondo tipo (β) $\Downarrow$

Lettura di un test statistico (1)

Esempio:

1) Ipotesi

H_0: $b1 = b2 = \ldots = bk = 0$

H_1: $bi \neq 0$

2) Statistica test — Statistica F

3) p-value — Rappresenta la probabilità di commettere l'errore di prima specie.
Può essere interpretato come la probabilità che H_0 sia "vera" in base al valore osservato della statistica test

Lettura di un test statistico (2)

Il p-value:

- è la probabilità che H_0 sia “vera” in base al valore osservato della statistica test

- è anche chiamato livello di significatività osservato

- è il più piccolo valore di α per il quale H_0 può essere rifiutata

Lettura di un test statistico (3)

Regola di Decisione: confrontare il p-value con **α**

Se p-value piccolo ($< \alpha$) ⇨ RIFIUTO H_0

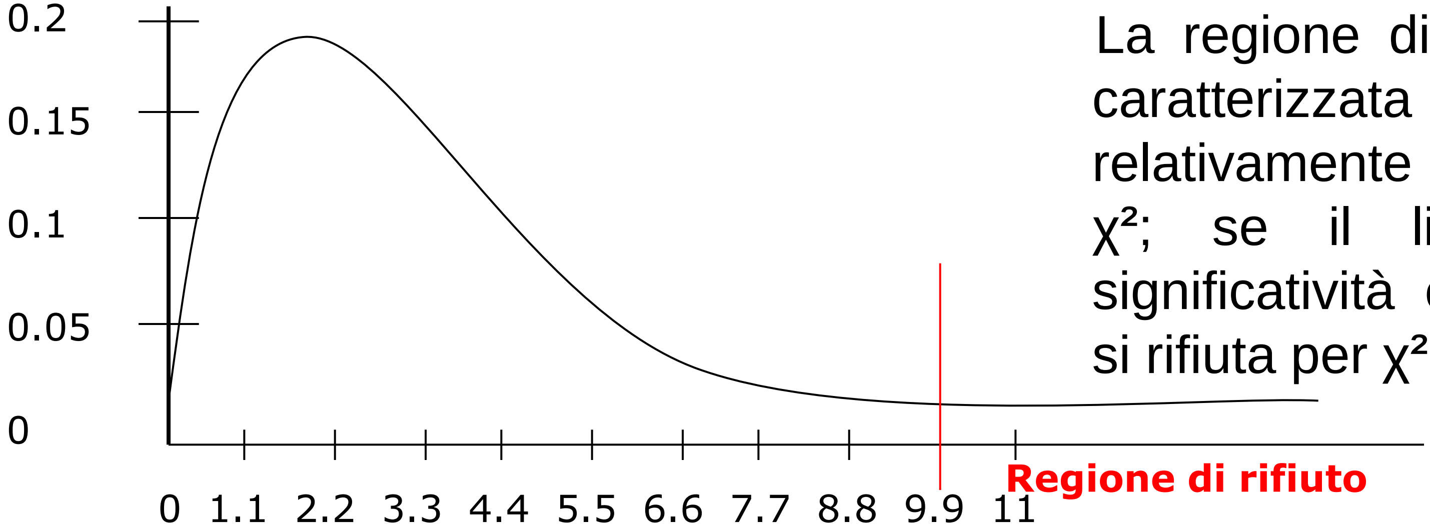
Altrimenti ($>= \alpha$) ⇨ ACCETTO H_0

Test χ^2 per l'indipendenza statistica

Si considera la distribuzione χ^2, con un numero di gradi di libertà pari a (k-1)(h-1), dove k è il numero di righe e h il numero di colonne della tabella di contingenza. Qui:

- H_0 :indipendenza statistica tra X e Y
- H_1 : dipendenza statistica tra X e Y

La regione di rifiuto cade nella coda di destra della distribuzione



La regione di rifiuto è caratterizzata da valori relativamente elevati di χ^2; se il livello di significatività è al 5%, si rifiuta per $\chi^2 > \chi^2_{0.95}$

Test χ² per l'indipendenza statistica

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.471[a]	3	.140
Likelihood Ratio	5.402	3	.145
N of Valid Cases	221		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 15.95.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	26.304[a]	8	.001
Likelihood Ratio	28.928	8	.000
N of Valid Cases	221		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.47.

Test χ^2 per l'indipendenza statistica
Esempio

H_0: assenza di associazione tra mano dominante e sesso (indipendenza statistica)

H_1: mano dominante non è independente dal sesso (dipendenza statistica)

Sesso	Mano dominante: Sinistra	Mano dominante: Destra	
Femmina	12	108	120
Maschio	24	156	180
	36	264	300

Se non c'è associazione, allora

P(Mancino | Femmina) = P(Mancino | Maschio) =P(Mancino)= 36/300= 0.12

Quindi ci aspetteremmo che Il 12% delle 120 femmine e Il 12% dei 180 maschi siano mancini…

Test χ^2 per l'indipendenza statistica
Esempio

- Se H_0 è vera, allora la proporzione di donne mancine dovrebbe coincidere con la proporzione di uomini mancini
- Le due proporzioni precedenti dovrebbero coincidere con la proporzione generale di gente mancina

Sesso	Mano dominante		
	Sinistra	Destra	
Femmina	**Osservate = 12** **Attese = 14.4**	**Osservate = 108** **Attese = 105.6**	**120**
Maschio	**Osservate = 24** **Attese = 21.6**	**Osservate = 156** **Attese = 158.4**	**180**
	36	**264**	**300**

$$E_{11} = \frac{(120)(36)}{300} = 14.4$$

Test χ² per l'indipendenza statistica
Esempio

La statistica test chi-quadrato è:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

con $g.d.l. = (r-1)(c-1)$

dove:

O_{ij} = frequenza osservate nella cella (*i, j*)

E_{ij} = frequenza attesa nella cella (*i, j*)

r = numero di righe

c = numero di colonne

Regola di Decisione:

confrontare il p-value con α

p-value = 0.32 > 0.05, quindi accettiamo H_0 e concludiamo che sesso e mano dominante non sono associate

Test t per l'indipendenza lineare

Questo test verifica l'ipotesi di indipendenza lineare tra due variabili, partendo dall'indice di correlazione lineare ρ. Si ha:

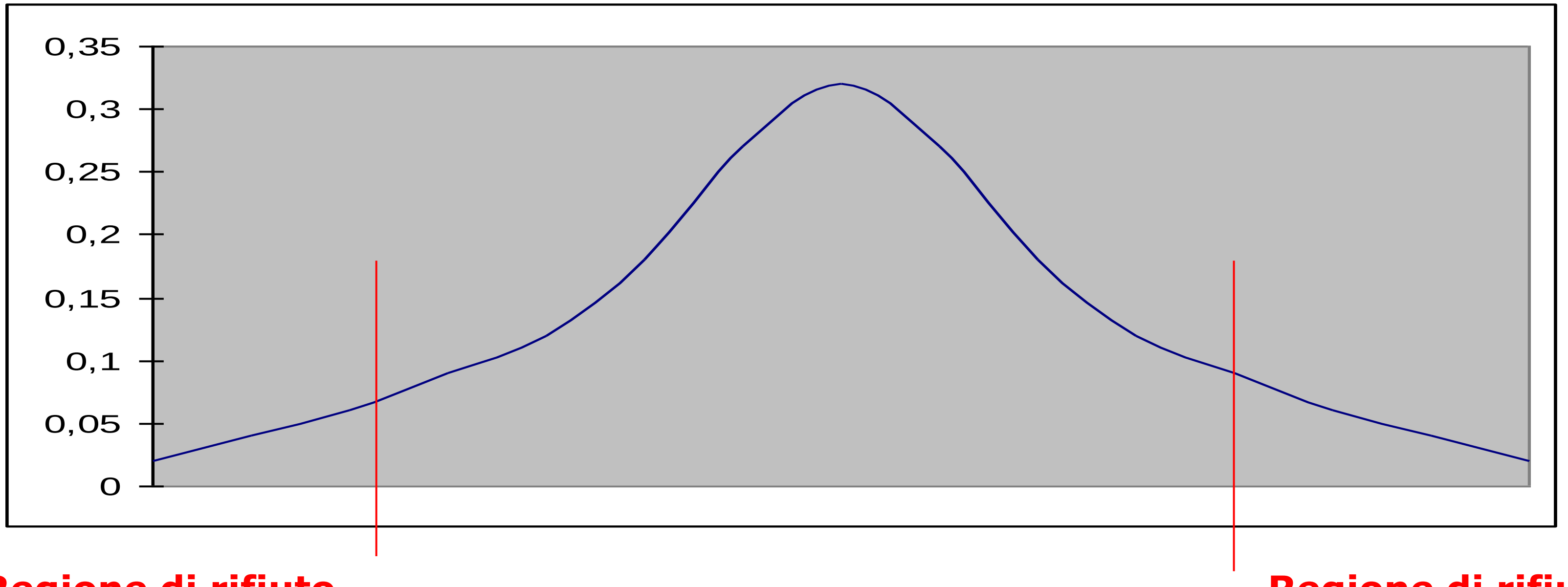
- H_0: indipendenza lineare tra X e Y ($\rho_{popolaz}=0$)
- H_1: dipendenza lineare tra X e Y ($\rho_{popolaz} \neq 0$)

La statistica test è distribuita come una t di Student con *n-2* gradi di libertà, e tende a crescere all'aumentare dell'ampiezza campionaria

$$t = \rho \sqrt{(n-2)/(1-\rho^2)}$$

Test t per l'indipendenza lineare

La regione di rifiuto è caratterizzata da valori relativamente elevati di t in modulo; se il livello di significatività è al 5%, si rifiuta per $|t| > t_{0,975}$



Regione di rifiuto **Regione di rifiuto**

Test t per l'indipendenza lineare

Correlations

		Qualità degli ingredienti	Genuinità	Leggerezza	Sapore/gusto
Qualità degli ingredienti	Pearson Correlation	1	.629**	.299**	.232**
	Sig. (2-tailed)		.000	.000	.001
	N	220	220	218	220
Genuinità	Pearson Correlation	.629**	1	.468**	.090
	Sig. (2-tailed)	.000		.000	.181
	N	220	220	218	220
Leggerezza	Pearson Correlation	.299**	.468**	1	.030
	Sig. (2-tailed)	.000	.000		.657
	N	218	218	219	219
Sapore/gusto	Pearson Correlation	.232**	.090	.030	1
	Sig. (2-tailed)	.001	.181	.657	
	N	220	220	219	221

**. Correlation is significant at the 0.01 level (2-tailed).

Test F per la verifica di ipotesi sulla differenza tra medie

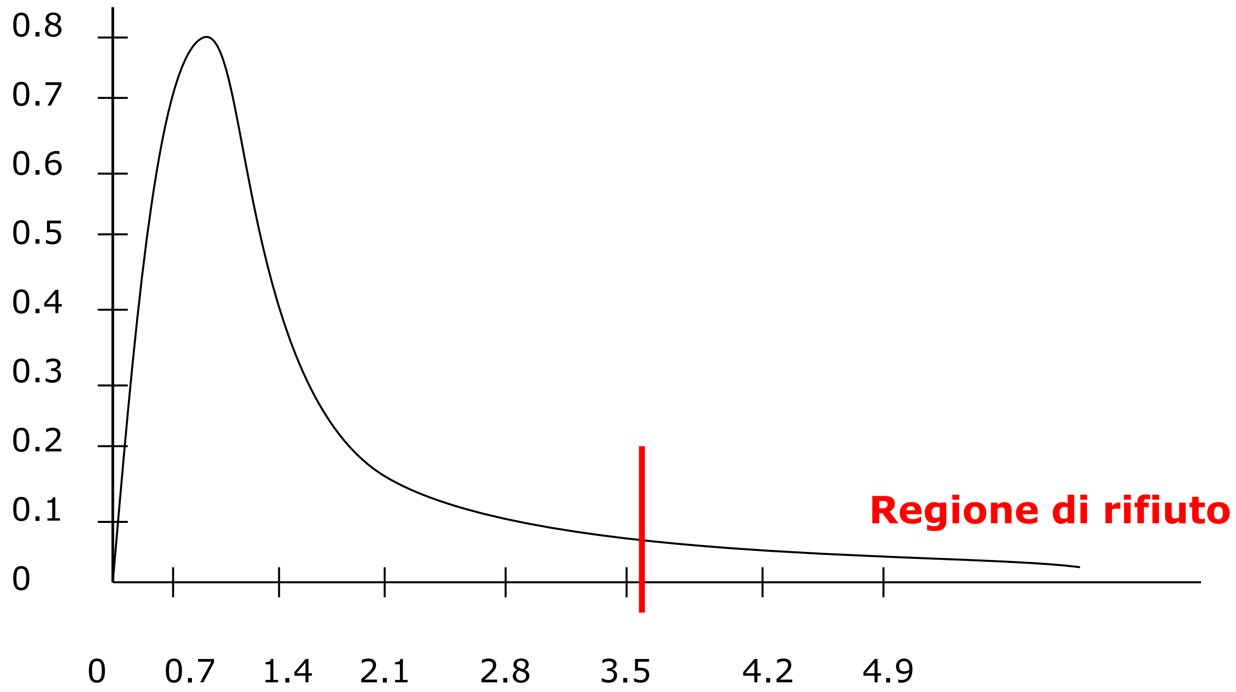
Si prende in considerazione la scomposizione della varianza; qui

- H_0: le medie sono tutte uguali tra loro
- H_1: esistono almeno due medie diverse tra loro

La statistica test da utilizzare, sotto l'ipotesi H_0, si distribuisce come una F di Fisher con (c-1,n-1) gradi di libertà. Tende a crescere all'aumentare della varianza tra medie e al diminuire della variabilità interna alle categorie. Cresce inoltre all'aumentare dell'ampiezza campionaria.

Test F per la verifica di ipotesi sulla differenza tra medie

La regione di rifiuto cade nella coda di destra della distribuzione, cioè è caratterizzata da valori relativamente elevati di F; se il livello di significatività è 5%, si rifiuta per F> $F_{0,95}$



Test F per la verifica di ipotesi sulla differenza tra medie

Report

Produzione artigianale

Età	Mean	N	Std. Deviation
18-25	5.01	78	2.224
26-35	5.53	55	2.609
36-50	6.00	41	2.098
Over 50	6.09	47	2.320
Total	5.55	221	2.352

Measures of Association

	Eta	Eta Squared
Produzione artigianale * Età	.191	.036

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Produzione artigianale * Età	Between Groups	(Combined)	44.296	3	14.765	2.733	.045
	Within Groups		1172.356	217	5.403		
	Total		1216.652	220			

Produzione artigianale

Età	Mean	N	Std. Deviation
18-25	5.01	78	2.224
26-35	5.53	55	2.609
36-50	6.00	41	2.098
Over 50	6.09	47	2.320
Total	5.55	221	2.352

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Produzione artigianale * Età	Between Groups	(Combined)	44.296	3	14.765	2.733	.045
	Within Groups		1172.356	217	5.403		
	Total		1216.652	220			

Report

Attenzione a bisogni specifici

Età	Mean	N	Std. Deviation
18-25	4.05	78	2.772
26-35	4.53	53	2.791
36-50	5.00	41	2.837
Over 50	5.83	47	8.168
Total	4.73	219	4.536

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Attenzione a bisogni specifici * Età	Between Groups	(Combined)	97.921	3	32.640	1.599	.191
	Within Groups		4387.641	215	20.408		
	Total		4485.562	218			