ANOVA

ANalysis Of VAriance

Federico Plazzi

12 Dicembre 2017

A che cosa serve?

A che cosa serve?

Applicazione

- L'ANOVA ha finalità simili al test t: confrontare campioni. Al contrario del test t, però, è in grado di confrontare più di due campioni alla volta.

A che cosa serve?

Applicazione

- L'ANOVA ha finalità simili al test *t*: confrontare campioni. Al contrario del test *t*, però, è in grado di confrontare più di due campioni alla volta.
- L'ANOVA può essere "One-Way" (un solo criterio di classificazione) o "Two-Way" (due criteri di classificazione).

A che cosa serve?

Applicazione

- L'ANOVA ha finalità simili al test t: confrontare campioni. Al contrario del test t, però, è in grado di confrontare più di due campioni alla volta.
- L'ANOVA può essere "One-Way" (un solo criterio di classificazione) o "Two-Way" (due criteri di classificazione).

Solite condizioni...

- **Scala di misura continua**.

A che cosa serve?

Applicazione

- L'ANOVA ha finalità simili al test t: confrontare campioni. Al contrario del test t, però, è in grado di confrontare più di due campioni alla volta.
- L'ANOVA può essere "One-Way" (un solo criterio di classificazione) o "Two-Way" (due criteri di classificazione).

Solite condizioni...

- **Scala di misura continua**.
- Campioni estratti **a caso** da una popolazione **a distribuzione normale**.

A che cosa serve?

A che cosa serve?

Condizioni particolari...

- **Varianze comparabili tra gruppi per dimensioni**: la varianza maggiore dovrebbe essere al massimo il 150% della minore.

A che cosa serve?

Condizioni particolari...

- **Varianze comparabili tra gruppi per dimensioni**: la varianza maggiore dovrebbe essere al massimo il 150% della minore.
- *Solo in caso di ANOVA per due campioni appaiati*: **sfericità delle correlazioni**, ossia coefficienti r positivi e comparabili per tutte le coppie di variabili.

A che cosa serve?

Condizioni particolari...

- **Varianze comparabili tra gruppi per dimensioni**: la varianza maggiore dovrebbe essere al massimo il 150% della minore.
- *Solo in caso di ANOVA per due campioni appaiati*: **sfericità delle correlazioni**, ossia coefficienti r positivi e comparabili per tutte le coppie di variabili.

E se le condizioni non sono verificate?

- **L'ANOVA è molto robusta**, soprattutto se i gruppi sono più o meno della stessa dimensione.

A che cosa serve?

Condizioni particolari...

- **Varianze comparabili tra gruppi per dimensioni**: la varianza maggiore dovrebbe essere al massimo il 150% della minore.
- *Solo in caso di ANOVA per due campioni appaiati*: **sfericità delle correlazioni**, ossia coefficienti r positivi e comparabili per tutte le coppie di variabili.

E se le condizioni non sono verificate?

- **L'ANOVA è molto robusta**, soprattutto se i gruppi sono più o meno della stessa dimensione.
- In caso di gruppi di dimensioni diverse presi da una popolazione la cui normalità è dubbia, esistono **varianti non parametriche**: il *test di Kruskal e Wallis* ed il *test di Friedman*.

One-Way ANOVA

One-Way ANOVA

I gruppi

- Dividiamo le nostre osservazioni in gruppi:

Tabella: Altezze per gruppi di interesse

Botanica	Ecologia	Geologia	Paleontologia	Zoologia	
167	168	147	173	172	182
170	177	160	155	177	192
165	165	178	166	187	166
161	160	168	180	171	165
170	165	185		158	181
170	162	179		160	169
	185			175	167
				170	168
				166	183
				175	170
				187	163
				163	170
				171	181
				180	

One-Way ANOVA

I conti

- Calcoliamo media e devianza dei singoli gruppi e del campione completo ("Generale" o "G"):

Tabella: Altezze per gruppi di interesse

	Botanica	Ecologia	Geologia	Paleontologia	Zoologia	Generale
μ	167,17	168,86	169,50	168,50	172,93	170,90
D	66,83	482,86	1001,50	341,00	2003,85	4154,50

One-Way ANOVA

I conti

- Calcolo la devianza *entro* gruppi:

$$D_{\text{entro}} = \sum_{i=1}^{k} D_i \tag{1}$$

dove D_i è la devianza dell'i-esimo gruppo e k è il numero dei gruppi.

Tabella: Altezze per gruppi di interesse

	Botanica	Ecologia	Geologia	Paleontologia	Zoologia	Generale
μ	167,17	168,86	169,50	168,50	172,93	170.90
D	66,83	482,86	1001,50	341,00	2003,85	4154,50
D_{entro}					3896,04	

One-Way ANOVA

I conti

- Calcolo la devianza *tra* gruppi come una specie di "devianza pesata" tra medie:

$$D_{\text{tra}} = \sum_{i=1}^{k} N_i \cdot (\mu_i - \mu_G)^2 \tag{2}$$

dove μ_i ed N_i sono la media e la dimensione dell'i-esimo gruppo e μ_G è la media generale (*non* la media delle medie!).

Tabella: Altezze per gruppi di interesse

	Botanica	Ecologia	Geologia	Paleontologia	Zoologia	Generale
μ	167,17	168,86	169,50	168,50	172,93	170.90
D	66,83	482,86	1001,50	341,00	2003,85	4154,50
D_{entro}					3896,04	
D_{tra}					258,46	

One-Way ANOVA

- Passiamo alle varianze *di popolazione*:

$$\sigma^2_{\text{entro}} = \frac{D_{\text{entro}}}{\sum_{i=1}^{k}(N_i - 1)} = \frac{D_{\text{entro}}}{N_G - k} \tag{3}$$

$$\sigma^2_{\text{tra}} = \frac{D_{\text{tra}}}{k - 1} \tag{4}$$

Tabella: Altezze per gruppi di interesse

	Botanica	Ecologia	Geologia	Paleontologia	Zoologia	Generale
μ	167,17	168,86	169,50	168,50	172,93	170.90
D	66,83	482,86	1001,50	341,00	2003,85	4154,50
D_{entro}					3896,04	
D_{tra}					258,46	
σ^2_{entro}					86,58	
σ^2_{tra}					64,61	

One-Way ANOVA

I conti

- Calcoliamo la statistica F:

$$F = \frac{\sigma^2_{\text{target}}}{\sigma^2_{\text{casuale}}} = \frac{\sigma^2_{\text{tra}}}{\sigma^2_{\text{entro}}} \tag{5}$$

Tabella: Altezze per gruppi di interesse

	Botanica	Ecologia	Geologia	Paleontologia	Zoologia	Generale
μ	167,17	168,86	169,50	168,50	172,93	170.90
D	66,83	482,86	1001,50	341,00	2003,85	4154,50
D_{entro}					3896,04	
D_{tra}					258,46	
σ^2_{entro}					86,58	
σ^2_{tra}					64,61	
F					0,75	

One-Way ANOVA

I conti

- Riassumendo:

Tabella: ANOVA

	Devianza	Gradi di libertà	Varianza	F	p
Entro gruppi ("casuale")	3896,04	45	86,58		
Tra gruppi ("target")	258,46	4	64,64	0,75	
Generale	4154,50	49			

One-Way ANOVA

La distribuzione di F

- Solito approccio: se conosciamo la distribuzione di F, possiamo valutare la significatività del nostro valore!

One-Way ANOVA

La distribuzione di F

- Solito approccio: se conosciamo la distribuzione di F, possiamo valutare la significatività del nostro valore!
- H_0**: non c'è differenza tra i gruppi.**

One-Way ANOVA

La distribuzione di F

- Solito approccio: se conosciamo la distribuzione di F, possiamo valutare la significatività del nostro valore!
- H_0: **non c'è differenza tra i gruppi.**
- Costruiamo una popolazione di 10.000 numeri completamente casuali;

One-Way ANOVA

La distribuzione di F

- Solito approccio: se conosciamo la distribuzione di F, possiamo valutare la significatività del nostro valore!
- H_0**: non c'è differenza tra i gruppi.**
- Costruiamo una popolazione di 10.000 numeri completamente casuali;
- estraiamo da questa popolazione 5 campioni delle dimensioni dei nostri e calcoliamo F;

One-Way ANOVA

La distribuzione di F

- Solito approccio: se conosciamo la distribuzione di F, possiamo valutare la significatività del nostro valore!
- H_0**: non c'è differenza tra i gruppi.**
- Costruiamo una popolazione di 10.000 numeri completamente casuali;
- estraiamo da questa popolazione 5 campioni delle dimensioni dei nostri e calcoliamo F;
- Ripetiamo l'operazione per 5.000 volte;

One-Way ANOVA

La distribuzione di F

- Solito approccio: se conosciamo la distribuzione di F, possiamo valutare la significatività del nostro valore!
- H_0**: non c'è differenza tra i gruppi.**
- Costruiamo una popolazione di 10.000 numeri completamente casuali;
- estraiamo da questa popolazione 5 campioni delle dimensioni dei nostri e calcoliamo F;
- Ripetiamo l'operazione per 5.000 volte;
- Al termine, avremo stimato la distribuzione di F **per 4 e 45 gradi di libertà**.

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

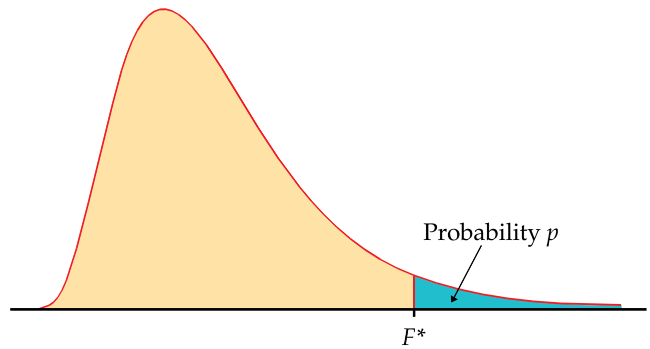

TABLE E

F critical values

		Degrees of freedom in the numerator								
	p	1	2	3	4	5	6	7	8	9
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

Degrees of freedom in the denominator

One-Way ANOVA

Risultati

- Inseriamo in tabella il p-value:

Tabella: ANOVA

	Devianza	Gradi di libertà	Varianza	F	p
Entro gruppi ("casuale")	3896,04	45	86,58		
Tra gruppi ("target")	258,46	4	64,64	0,75	0,56
Generale	4154,50	49			

One-Way ANOVA

Risultati

- Inseriamo in tabella il p-value:

Tabella: ANOVA

	Devianza	Gradi di libertà	Varianza	*F*	p
Entro gruppi ("casuale")	3896,04	45	86,58		
Tra gruppi ("target")	258,46	4	64,64	0,75	0,56
Generale	4154,50	49			

- Siccome il p-value è > 0.05 non possiamo rigettare l'ipotesi nulla: l'ANOVA non rivela nei nostri dati alcuna strutturazione in *questi* cinque gruppi.

Oltre l'One-Way ANOVA

Oltre l'One-Way ANOVA

Il test di Tukey

- Il test di Tukey è utile, in caso di valori significativi di F, per effettuare confronti a coppie.

Oltre l'One-Way ANOVA

Il test di Tukey

- Il test di Tukey è utile, in caso di valori significativi di F, per effettuare confronti a coppie.
- **Il test t non può essere utilizzato a questo scopo!**

Oltre l'One-Way ANOVA

Il test di Tukey

- Il test di Tukey è utile, in caso di valori significativi di F, per effettuare confronti a coppie.
- **Il test t non può essere utilizzato a questo scopo!**
- Per ogni coppia di campioni i e j, stimiamo la statistica Q:

$$Q = \frac{|\mu_i - \mu_j|}{\sqrt{\frac{\sigma^2_{\text{entro}}}{N}}} \tag{6}$$

dove μ_i e μ_j sono la media dell'i-esimo e del j-esimo campione e σ^2_{entro} la varianza entro gruppi.

Oltre l'One-Way ANOVA

Il test di Tukey

- Il test di Tukey è utile, in caso di valori significativi di F, per effettuare confronti a coppie.
- **Il test t non può essere utilizzato a questo scopo!**
- Per ogni coppia di campioni i e j, stimiamo la statistica Q:

$$Q = \frac{|\mu_i - \mu_j|}{\sqrt{\frac{\sigma^2_{\text{entro}}}{\overline{N}}}} \tag{6}$$

  dove μ_i e μ_j sono la media dell'i-esimo e del j-esimo campione e σ^2_{entro} la varianza entro gruppi.
- $\overline{N}$ è la dimensione dei campioni; se non è costante, usiamo la media armonica:

$$\overline{N} = \frac{k}{\sum_{i=1}^{k} \frac{1}{N_i}} \tag{7}$$

Oltre l'One-Way ANOVA

Il test di Tukey

- Al solito, la statistica Q ha la sua distribuzione, che ci permetterà di capire se il valore che abbiamo ottenuto è significativo oppure no.

Oltre l'One-Way ANOVA

Il test di Tukey

- Al solito, la statistica Q ha la sua distribuzione, che ci permetterà di capire se il valore che abbiamo ottenuto è significativo oppure no.
- H_0: **non c'è differenza tra i due gruppi** ($|\mu_i - \mu_j| = 0$)**.**

Table: Q scores for Tukey's method

$\alpha = 0.05$

df \ k	2	3	4	5	6	7	8	9	10
1	18.0	27.0	32.8	37.1	40.4	43.1	45.4	47.4	49.1
2	6.08	8.33	9.80	10.88	11.73	12.43	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

$\alpha = 0.01$

df \ k	2	3	4	5	6	7	8	9	10
1	90.0	135	164	186	202	216	227	237	246
2	13.90	19.02	22.56	25.37	27.76	29.86	31.73	33.41	34.93
3	8.26	10.62	12.17	13.32	14.24	15.00	15.65	16.21	16.71
4	6.51	8.12	9.17	9.96	10.58	11.10	11.54	11.92	12.26
5	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10
7	4.95	5.92	6.54	7.00	7.37	7.68	7.94	8.17	8.37
8	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21
11	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99
12	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54
15	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44
16	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09
24	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76
40	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60
60	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16

Oltre l'One-Way ANOVA

Il test di Tukey

- Al solito, la statistica Q ha la sua distribuzione, che ci permetterà di capire se il valore che abbiamo ottenuto è significativo oppure no.
- H_0: **non c'è differenza tra i due gruppi ($|\mu_i - \mu_j| = 0$).**

η^2

- η^2 stima il livello di correlazione presente tra i dati in generale, indipendentemente dalla linearità della regressione. È definito semplicemente come

$$\eta^2 = \frac{D_{\text{tra}}}{D_G} \tag{8}$$

Oltre l'One-Way ANOVA

Il test di Tukey

- Al solito, la statistica Q ha la sua distribuzione, che ci permetterà di capire se il valore che abbiamo ottenuto è significativo oppure no.
- H_0: **non c'è differenza tra i due gruppi ($|\mu_i - \mu_j| = 0$).**

η^2

- η^2 stima il livello di correlazione presente tra i dati in generale, indipendentemente dalla linearità della regressione. È definito semplicemente come

$$\eta^2 = \frac{D_{\text{tra}}}{D_G} \tag{8}$$

- Per esempio, un valore di η^2 pari a 0,80 indica che l'80% della variabilità è associato alla divisione in gruppi effettuata.

One-Way ANOVA per due campioni appaiati

Cosa cambia

- I primi step sono identici: si calcolano le devianze entro gruppi e tra gruppi.

One-Way ANOVA per due campioni appaiati

Cosa cambia

- I primi step sono identici: si calcolano le devianze entro gruppi e tra gruppi.
- In questo caso, possiamo calcolare anche una "devianza dovuta alle differenze individuali intrinseche":

$$D_{\text{ind}} = k \cdot \sum_{i=1}^{N} (\mu_i - \mu_G)^2 \tag{9}$$

dove μ_i è la media del singolo individuo i.

One-Way ANOVA per due campioni appaiati

Cosa cambia

- I primi step sono identici: si calcolano le devianze entro gruppi e tra gruppi.
- In questo caso, possiamo calcolare anche una "devianza dovuta alle differenze individuali intrinseche":

$$D_{\text{ind}} = k \cdot \sum_{i=1}^{N} (\mu_i - \mu_G)^2 \tag{9}$$

  dove μ_i è la media del singolo individuo i.
- Questa "devianza individuale" viene poi semplicemente sottratta dalla devianza *entro gruppi* prima del calcolo delle varianze.

One-Way ANOVA per due campioni appaiati

Cosa cambia

- I primi step sono identici: si calcolano le devianze entro gruppi e tra gruppi.
- In questo caso, possiamo calcolare anche una "devianza dovuta alle differenze individuali intrinseche":

$$D_{\text{ind}} = k \cdot \sum_{i=1}^{N} (\mu_i - \mu_G)^2 \tag{9}$$

  dove μ_i è la media del singolo individuo i.
- Questa "devianza individuale" viene poi semplicemente sottratta dalla devianza *entro gruppi* prima del calcolo delle varianze.
- La "devianza individuale" ha $N - 1$ gradi di libertà, per cui la varianza *entro gruppi* risultante avrà $(N_G - k) - (N - 1)$ gradi di libertà.

One-Way ANOVA per due campioni appaiati

Cosa cambia

- Si calcola e si interpreta F come nell'altro caso.

One-Way ANOVA per due campioni appaiati

Cosa cambia

- Si calcola e si interpreta F come nell'altro caso.
- Possiamo effettuare anche il test di Tukey come nell'altro caso: al posto della varianza entro gruppi useremo la varianza residua (dopo la sottrazione della "varianza individuale").

One-Way ANOVA per due campioni appaiati

Cosa cambia

- Si calcola e si interpreta F come nell'altro caso.
- Possiamo effettuare anche il test di Tukey come nell'altro caso: al posto della varianza entro gruppi useremo la varianza residua (dopo la sottrazione della "varianza individuale").
- Come accennato all'inizio, in questo caso abbiamo un'ulteriore assunzione, quella della **sfericità delle correlazioni**: calcoliamo i coefficienti di correlazione (r) tra tutti le possibili coppie e devono essere tutti comparabili.

One-Way ANOVA per due campioni appaiati

Cosa cambia

- Si calcola e si interpreta F come nell'altro caso.
- Possiamo effettuare anche il test di Tukey come nell'altro caso: al posto della varianza entro gruppi useremo la varianza residua (dopo la sottrazione della "varianza individuale").
- Come accennato all'inizio, in questo caso abbiamo un'ulteriore assunzione, quella della **sfericità delle correlazioni**: calcoliamo i coefficienti di correlazione (r) tra tutti le possibili coppie e devono essere tutti comparabili.

A cosa serve

- **Di fatto, se i campioni sono correlati, possiamo eliminare quella fetta di variabilità dovuta al fatto che non solo i trattamenti, ma anche gli individui stessi sono diversi tra loro!**

Two-Way ANOVA

Cosa cambia

- Le variabili sono ripartite secondo due tipi di categoria (per esempio, il sesso e l'ambito desiderato), producendo una tabella a doppia entrata.

Two-Way ANOVA

Cosa cambia

- Le variabili sono ripartite secondo due tipi di categoria (per esempio, il sesso e l'ambito desiderato), producendo una tabella a doppia entrata.
- Possiamo calcolare le solite devianze, ma possiamo anche calcolare la devianza delle righe (D_{righe}) e la devianza delle colonne (D_{colonne}) rispetto alla media totale, nel solito modo ($\sum N \cdot (\mu - \mu_G)^2$).

Two-Way ANOVA

Cosa cambia

- Le variabili sono ripartite secondo due tipi di categoria (per esempio, il sesso e l'ambito desiderato), producendo una tabella a doppia entrata.
- Possiamo calcolare le solite devianze, ma possiamo anche calcolare la devianza delle righe (D_{righe}) e la devianza delle colonne (D_{colonne}) rispetto alla media totale, nel solito modo ($\sum N \cdot (\mu - \mu_G)^2$).
- In questo caso, avremo anche una devianza dovuta all'interazione delle due variabili classificatorie, D_{inter}:

$$D_{\text{tra}} = D_{\text{righe}} + D_{\text{colonne}} + D_{\text{inter}} \tag{10}$$

Two-Way ANOVA

Cosa cambia

- Sotto l'ipotesi nulla, ci aspettiamo che la differenza tra la media di ogni gruppo (μ_i) e la media generale sia banalmente additiva.

$$\mu_{Ai} - \mu_G = (\mu_r - \mu_G) + (\mu_c - \mu_G) \quad (11)$$

$$\mu_{Ai} - \mu_G = \mu_r + \mu_c - 2\mu_G \quad (12)$$

$$\mu_{Ai} = \mu_r + \mu_c - \mu_G \quad (13)$$

dove μ_{Ai} è la media attesa dell'*i*-esimo gruppo e μ_r e μ_c sono la media della riga e della colonna a cui appartiene.

Two-Way ANOVA

Cosa cambia

- Sotto l'ipotesi nulla, ci aspettiamo che la differenza tra la media di ogni gruppo (μ_i) e la media generale sia banalmente additiva.

$$\mu_{Ai} - \mu_G = (\mu_r - \mu_G) + (\mu_c - \mu_G) \tag{11}$$

$$\mu_{Ai} - \mu_G = \mu_r + \mu_c - 2\mu_G \tag{12}$$

$$\mu_{Ai} = \mu_r + \mu_c - \mu_G \tag{13}$$

  dove μ_{Ai} è la media attesa dell'i-esimo gruppo e μ_r e μ_c sono la media della riga e della colonna a cui appartiene.

- Possiamo usare le medie attese per ogni gruppo sotto l'ipotesi nulla per calcolare la devianza di interazione:

$$D_{\text{inter}} = \sum_{i=1}^{k} N_i \cdot (\mu_{Ai} - \mu_i)^2 \tag{14}$$

Two-Way ANOVA

Cosa cambia

- Calcoliamo i gradi di libertà come $(R-1)\cdot(C-1)$, dove R e C sono il numero delle righe ed il numero delle colonne.

Two-Way ANOVA

Cosa cambia

- Calcoliamo i gradi di libertà come $(R-1)\cdot(C-1)$, dove R e C sono il numero delle righe ed il numero delle colonne.
- Dalle devianze e dai gradi di libertà possiamo calcolare le varianze.

Two-Way ANOVA

Cosa cambia

- Calcoliamo i gradi di libertà come $(R-1)\cdot(C-1)$, dove R e C sono il numero delle righe ed il numero delle colonne.
- Dalle devianze e dai gradi di libertà possiamo calcolare le varianze.
- A questo punto possiamo calcolare ben tre F di cui testare la significatività!

$$F_{\text{righe}} = \frac{\sigma^2_{\text{righe}}}{\sigma^2_{\text{entro}}} \tag{15}$$

$$F_{\text{colonne}} = \frac{\sigma^2_{\text{colonne}}}{\sigma^2_{\text{entro}}} \tag{16}$$

$$F_{\text{inter}} = \frac{\sigma^2_{\text{inter}}}{\sigma^2_{\text{entro}}} \tag{17}$$