

# Affari Tuoi è truccato?

Pietro Parenti

2024-12-07

## Obiettivo

Il gioco televisivo affari tuoi è stato al centro di molte critiche alcuni anni fa con l'accusa di truccare le puntate per manipolare la trasmissione al fine di portare alla fine della puntata i pacchi dal valore più alto, il tutto per aumentare gli ascolti. Il gioco subì quindi una grande rivoluzione e il problema sembrò essere stato risolto. Col ritorno in auge del programma negli ultimi mesi, sono tornate voci su possibili irregolarità. Lo studio ha lo scopo di andare a verificare possibili anomalie presenti nelle puntate della primavera 2024. Le analisi si concentreranno sulla possibile associazione tra il valore del pacco e varie variabili, in particolare la variabile ordine, che evidenzia, attraverso un numero, l'ordine di uscita del pacco nel corso della puntata.

## Descrizione del gioco

Il gioco prevede per ogni puntata che il concorrente scelga un pacco tra i 20 che gli si presentano davanti. Ciascun pacco è caratterizzato da un numero, una regione e un valore in denaro. Le prime due variabili sono note, l'ultima è ignota, fino all'apertura del pacco. Il concorrente infatti, durante la puntata deve aprire ciascuno dei 20 pacchi, aprendo per ultimo il proprio. Il valore che è contenuto nel proprio pacco è il montepremi che il concorrente si porterà a casa.

## Analisi descrittive

Il dataset è formato da 5 variabili e 920 osservazioni. Queste 920 osservazioni fanno riferimento a tutte le estrazioni effettuate nel programma televisivo dal 5 marzo 2024 al 23 aprile 2024. Tra queste due date sono andate in onda un totale di 46 puntate e ciascuna delle quali è composta da 20 estrazioni.

```
skim_without_charts(DatiAffariTuoi)
```

Table 1: Data summary

Name	DatiAffariTuoi
Number of rows	920
Number of columns	5
Column type frequency:	
character	1
numeric	3
POSIXct	1

Group variables

None

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
regione	0	1	5	10	0	20	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
pacco	0	1	10.50	5.77	1	5.75	10.5	15.25	2e+01
soldi	0	1	40298.05	76623.85	0	42.50	2750.0	35000.00	3e+05
ordine	0	1	10.50	5.77	1	5.75	10.5	15.25	2e+01

**Variable type: POSIXct**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
data	0	1	2024-03-05	2024-04-23	2024-03-29	46

Per ciascuna estrazione sono stati rilevate 5 variabili:

- il numero del pacco che è stato aperto
- i soldi contenuti nel pacco
- la regione di appartenenza del pacco
- la data della puntata a cui si riferisce
- il numero ordinale di estrazione del pacco.

I valori medi delle variabili osservate sembrano essere totalmente coerenti con le logiche del gioco, dunque non si ritiene che ci siano errori nella raccolta dei dati. Non sono presenti dati mancanti.

## Analisi statistiche

### Verifica indipendenza tra le variabili soldi e ordine

Per fare ciò vengono utilizzati due metodi. Nel primo metodo la variabile soldi sarà trasformata in categoriale. In questo caso si segue la logica del programma, dunque la variabile viene divisa in pacchi cosiddetti blu (corrispondenti ai 10 pacchi con valore più basso) e rossi (corrispondenti ai 10 con valore più elevato). La variabile ordine dei pacchi invece viene invece divisa in 4 categorie di uguale dimensione.

```
DatiAffariTuoio$soldi_cat <- cut(DatiAffariTuoio$soldi, breaks = c(-1, 505, 300001),
                                labels = c("Blu", "Rosso" ))
skim_without_charts(DatiAffariTuoio$soldi_cat)
```

Table 5: Data summary

Name	DatiAffariTuoi\$soldi_cat
Number of rows	920
Number of columns	1
Column type frequency: factor	1
Group variables	None

**Variable type: factor**

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
data	0	1	FALSE	2	Blu: 460, Ros: 460

```
DatiAffariTuoi$ordine_cat <- cut(DatiAffariTuoi$ordine, breaks = c(-1, 5.5, 10.5, 15.5, 20.5),
                                labels = c("Inizio", "Inizio-metà", "Metà-fine", "Fine" ))
skim_without_charts(DatiAffariTuoi$ordine_cat)
```

Table 7: Data summary

Name	DatiAffariTuoi\$ordine_cat
Number of rows	920
Number of columns	1
Column type frequency: factor	1
Group variables	None

**Variable type: factor**

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
data	0	1	FALSE	4	Ini: 230, Ini: 230, Met: 230, Fin: 230

Si esegue quindi un test di indipendenza chi2. Viene creata prima la tabella di contingenza.

```
ordine_soldi <- table(DatiAffariTuoi$ordine_cat, DatiAffariTuoi$soldi_cat)
print(ordine_soldi)
```

```
##
##          Blu Rosso
## Inizio    125   105
## Inizio-metà 111   119
## Metà-fine  111   119
## Fine      113   117
```

```
chisq.test(ordine_soldi)
```

```
##
## Pearson's Chi-squared test
##
## data:  ordine_soldi
## X-squared = 2.3652, df = 3, p-value = 0.5001
```

Le due variabili non sono associate a nessun livello di significatività.

Si verifica ora l'associazione andando a valutare la media della variabile soldi negli ultimi 3 pacchi rimanenti in gioco. La media non dovrebbe essere significativamente distante dal valore della media generale dei pacchi che, come abbiamo visto in precedenza, è pari a 40298.05.

```
t.test((DatiAffariTuoisoldi[DatiAffariTuoisordine >= 18]),
      (DatiAffariTuoisoldi[DatiAffariTuoisordine < 18]),
      var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data:  (DatiAffariTuoisoldi[DatiAffariTuoisordine >= 18]) and (DatiAffariTuoisoldi[DatiAffariTuoisordine < 18])
## t = 0.53764, df = 918, p-value = 0.591
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -10084.89 17695.23
## sample estimates:
## mean of x mean of y
## 43532.44 39727.27
```

Il t-test a due medie non evidenzia una significativa differenza tra la media degli ultimi tre pacchi e quella degli altri.

## Verifica indipendenza tra le variabili regione/pacco e soldi

Per questa verifica utilizziamo nuovamente il test di associazione chi2 mantenendo la divisione precedente in pacchi blu e pacchi rossi.

### Associazione regione-soldi

```
regione_soldi <- table(DatiAffariTuoisregione, DatiAffariTuoisoldi_cat)
print(regione_soldi)
```

```
##
##          Blu Rosso
## abruzzo    31   15
## basilicata  17   29
## calabria   22   24
## campania   24   22
```

```
## emilia-r      23    23
## friuli        19    27
## lazio         25    21
## liguria       24    22
## lombard       22    24
## marche        23    23
## molise        20    26
## piemonte      20    26
## puglia        21    25
## sardegna      30    16
## sicilia       27    19
## toscana       23    23
## trentino      26    20
## umbria        23    23
## valle d'a     13    33
## veneto        27    19
```

```
chisq.test(regione_soldi)
```

```
##
## Pearson's Chi-squared test
##
## data:  regione_soldi
## X-squared = 29.217, df = 19, p-value = 0.06264
```

L'associazione tra le due variabili ha un p-value di 0.06, il che porta a valutazioni incerte. Le due variabili infatti non risultano associate a un livello di significatività del 5%, tuttavia risultano associate a un livello del 10%. In questo studio tuttavia il livello che è stato considerato anche per le altre variabili è il 5%, dunque questi dati non ci portano a rifiutare l'ipotesi nulla: le due variabili sono dunque indipendenti.

### Associazione pacco-soldi

```
pacco_soldi <- table(DatiAffariTuoipacco, DatiAffariTuoisoldi_cat)
print(pacco_soldi)
```

```
##
##      Blu Rosso
## 1    27    19
## 2    24    22
## 3    25    21
## 4    21    25
## 5    19    27
## 6    22    24
## 7    26    20
## 8    24    22
## 9    21    25
## 10   22    24
## 11   20    26
## 12   26    20
## 13   27    19
## 14   23    23
```

```
## 15 20 26
## 16 29 17
## 17 19 27
## 18 26 20
## 19 21 25
## 20 18 28
```

```
chisq.test(pacco_soldi)
```

```
##
## Pearson's Chi-squared test
##
## data:  pacco_soldi
## X-squared = 16.522, df = 19, p-value = 0.6222
```

Le due variabili sono indipendenti.

## Analisi montepremi vinti

I montepremi vinti sono coerenti con le logiche del gioco? La media dei pacchi aperti per ultimi (ordine=20) corrispondono al premio vinto dal concorrente. La media dei pacchi con ordine=20 dovrebbe essere non significativamente diversa da quella degli altri pacchi.

```
t.test((DatiAffariTuoisoldi[DatiAffariTuoisordine >= 20]),
      (DatiAffariTuoisoldi[DatiAffariTuoisordine < 20]),
      var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data:  (DatiAffariTuoisoldi[DatiAffariTuoisordine >= 20]) and (DatiAffariTuoisoldi[DatiAffariTuoisordine < 20])
## t = 0.51273, df = 918, p-value = 0.6083
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -16811.72  28702.65
## sample estimates:
## mean of x mean of y
## 45946.24 40000.78
```

La media dei montepremi vinti è in linea con gli altri pacchi.

## Conclusioni

L'indipendenza tra tutte le coppie di variabili prese in considerazione è stata confermata dai dati. E' dunque corretto ritenere che il gioco sia assolutamente regolare, almeno per i fattori tenuti in considerazione in questo studio.