Livelli di Preferenza come Predittori di Performance

di giussani riccardo

# Summary

Nelle situazioni di conflitto ogni attore ha:

* Delle caratteristiche, comuni a tutti gli attori, in cui ogni singolo attore ha il proprio valore associato
* Un valore che identifica la propria performance nella situazione di conflitto contro gli altri

I valori delle caratteristiche non possono essere migliorati all’infinito, ma devono sussistere dei trade-off tra coppie di valori, per cui migliorandone uno ne peggioro un altro.

Ogni attore può essere visto, rispetto ad un trade-off, come un paniere, e pertanto si possono determinare dei livelli di preferenza, per cui alcuni attori sono buoni in questo trade-off mentre altri scarsi.

Ho voluto esaminare il problema:

come determinare quanto la bontà in un trade-off è significativa nel determinare la performance

Ho trovato una istanza del problema nel popolare videogioco ‘Super Mario Party’, ispirato al gioco dell’oca, in cui ogni personaggio si muove per mezzo di un proprio dado speciale.

L’attore è il personaggio (o meglio, il dado associato) che ha per caratteristiche le statistiche delle facce del proprio dado (medie e varianze); la performance è ricavata da un campione di 60 partite.

Per risolvere il problema ho voluto utilizzare la regressione dalle variabili indipendenti dei livelli di preferenza sui trade-off alla variabile dipendente della performance, sfruttando il t test per eliminare con backward elimination i trade-off non sufficientemente significativi.

Il risultato sta nel modello di regressione finale, in cui rimangono i trade-off significativi con i coefficienti associati, che quantificano numericamente questa significatività.

# Introduzione

Come caso di studio ho scelto un semplice videogioco ispirato al gioco dell’oca: ‘Super Mario Party’.

In questo gioco 4 personaggi si muovono su un tabellone le cui caselle determinano o il guadagno di monete, o la perdita di esse o il succedere di vari eventi speciali (come cambiamenti dei percorsi sul tabellone, ottenimento di strumenti utili ecc.). Alla fine di ogni turno i giocatori partecipano ad un minigioco che mette in palio delle monete. Il vincitore è determinato da colui che riesce ad acquistare più stelle, ottenibili in certe caselle lungo il percorso (che spesso cambiano posizione).

Nella più recente edizione del gioco è stata introdotta la possibilità per ogni personaggio di utilizzare il proprio dado speciale, diverso da quello di tutti gli altri personaggi.

## Risorse

Ogni dado ha delle caratteristiche fondamentali, che da adesso chiamerò ‘risorse’; queste sono:

* Il valore atteso di movimento del dado ad ogni lancio, ossia la media del numero di caselle indicato su ogni faccia
* Il valore atteso di monete ricevute/perse ad ogni lancio, ossia la media del numero di monete indicato su ogni faccia
* La varianza associata ad ognuna delle medie, quindi varianza di movimento e varianza di monete
* Il numero di ‘opzioni’ di movimento, cioè il numero delle facce di movimento diverse da zero e diverse tra loro. La ‘varietà’ è stata introdotta perché, essendoci sul percorso delle caselle speciali, il giocatore può trarre vantaggio nel muoversi di un numero specifico di caselle (non per forza il più alto possibile)

Di queste risorse:

* Media movimento, media monete e varietà sono risorse ‘favorevoli’ al giocatore, dunque si cercherà di massimizzarle
* Le varianze, che altro non sono che indicatori del rischio associato ad ogni valore atteso, possono essere viste in ottica risk seeker, che cerca di massimizzare la varianza per poter godere appieno dei ‘colpi di fortuna’ e in ottica risk adverse, che cerca di minimizzare la varianza per avere un valore atteso ‘certo’ ad ogni turno

I dadi hanno idealmente 7 funzioni obiettivo:

* Massimizzare media movimento
* Massimizzare media monete
* Massimizzare varietà
* Massimizzare / Minimizzare varianza movimento
* Massimizzare / Minimizzare varianza monete

Siccome in un gioco equilibrato non è possibile massimizzare/minimizzare tutte le risorse all’infinito, occorre operare dei trade-off.

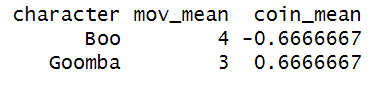
## Trade-off

Si ha un trade-off quando, date due quantità da massimizzare, arrivati ad un certo punto non è più possibile aumentare il valore di una quantità senza diminuire quello dell’altra.

Esempio

Prendiamo le risorse media movimento e media monete.

Consideriamo il dado classico (1,2,3,4,5,6) con media mov = 3.5 e media monete = 0 e prendiamo in esame due dei dadi del gioco:



‘Boo’ offre un aumento di 0.5 sul movimento a prezzo di 0.666 sulle monete;

‘Goomba’ offre al contrario un aumento di 0.666 sulle monete al prezzo di 0.5 sul movimento.

## Preferenza

Nei confronti di ogni coppia di risorse ogni dado può essere visto come un paniere e come tale può essere confrontato con tutti gli altri, andando a identificare panieri migliori e peggiori

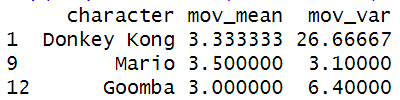
Esempio

Prendiamo le risorse media movimento e varianza movimento, in ottica risk adverse.

Come base per il ragionamento prendiamo il dado di ‘Yoshi’:



Questo lo confrontiamo con altri tre dadi disponibili, in ottica dei trade-off che propongono:



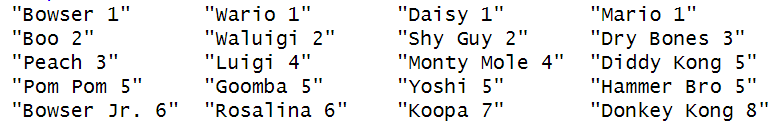
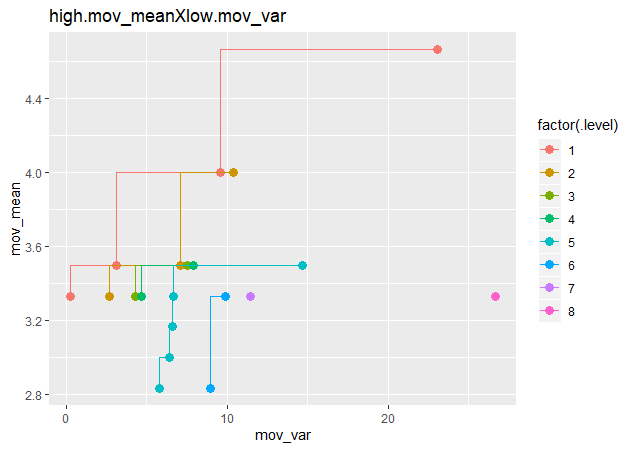
‘Goomba’ propone una diminuzione della varianza di 0.1666 al prezzo di una diminuzione della media di 0.1666: lo scambio sembra equo (non tanto perché i numeri sono identici quanto perché non è in grado di migliorare un valore senza peggiorarne un altro).

‘Mario’ propone contemporaneamente un aumento di 0.333 nella media e una diminuzione di 3.466 nella varianza: lo scambio è assolutamente vantaggioso!

‘Donkey Kong’ cerca invece di truffarci: per un aumento di 0.333 di media causa un’impennata della varianza.

La conclusione è questa: il dado di ‘Mario’ è da preferirsi a quello di ‘Yoshi’, che è pari a quello di ‘Goomba’ che è meglio di quello di ‘Donkey Kong’.

Estendendo questo tipo di ragionamento a tutti i dadi, si possono identificare dei livelli di preferenza, per cui tutti i dadi sullo stesso livello sono ‘indifferentemente buoni’ e sono migliori di quelli nei livelli sotto.



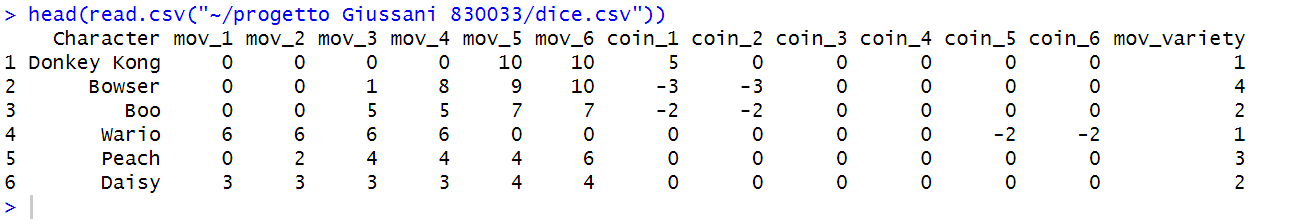
## Obiettivo

Preso un campione sufficientemente grande di partite, le frequenze di vittoria di ogni personaggio possono darci una graduatoria di quali dadi siano migliori e quali siano peggiori; questo in fiducia del fatto che, nonostante il gioco sia basato sull’imprevedibilità e ci siano molti eventi, ci siano dei dadi che hanno più probabilità di vincere in virtù del loro essere migliori su certi trade-off rispetto ad altri.

Presi i livelli di preferenza su ogni trade-off, voglio andare a determinare quanto l’essere migliore su un certo trade-off rende il dado effettivamente più forte degli altri.

# Data Acquisition

## dice.csv



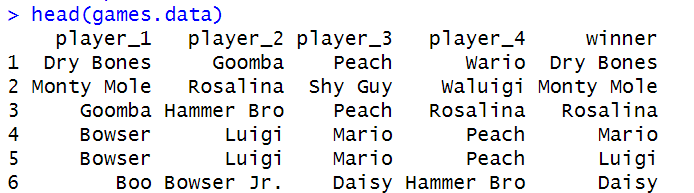
Direttamente dalla fonte principale (il videogioco), ho copiato i valori delle facce di ognuno dei 20 dadi.

Ogni dado ha 6 facce: queste o indicano una quantità di caselle della quale il personaggio può muoversi o una quantità di monete che riceve o gli viene tolta; le facce che indicano monete hanno movimento pari a zero.

Per ogni personaggio ho riportato le sei facce del suo dado, con numero di caselle di cui si può muovere e monete date o tolte. Infine, ho inserito una colonna con conteggiato il numero di facce di movimento diverse da zero e diverse tra loro. Questo valore è chiamato ‘varietà’ ed indica quante opzioni di movimento ha un dado.

Da questo dataset estraggo le ‘risorse’ di ogni dado.

## gamesResults.csv

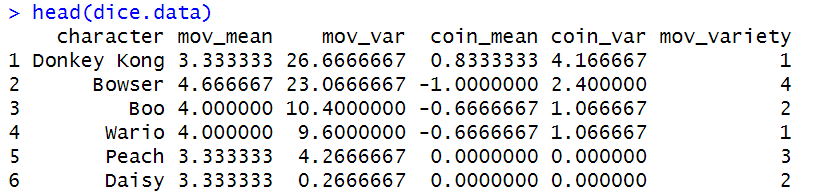


Navigando su Internet ho trovato 60 video di partite giocate tra 4 giocatori e per ognuna di queste ho riportato i 4 personaggi e il vincitore.

Di questo dataset sono interessato alla frequenza di vittorie di ogni personaggio, ossia al rapporto di partite vinte su partite giocate. La frequenza di vittorie viene considerata come un indicatore della forza del dado: più partite ha vinto più è forte e viceversa

# Exploratory Analysis

## dice.data



Da dice.csv ho ricavato questo dataset

### 

### Media Movimento

Rispetto al dado classico (1,2,3,4,5,6), con media 3.5, la media delle medie di movimento è leggermente più bassa. Questo indica che tendenzialmente questa risorsa viene ceduta in cambio di altre.

La deviazione standard suggerisce che non è sempre vero: infatti ci sono alcuni dadi con media pari a quella del dado standard e pochi casi in cui la media è più alta (anche di molto).

Il valore di Skewness identifica una direzione della deviazione standard rispetto alla media molto forte verso destra, più forte di quella di una distribuzione semi normale.

La Kurtosis indica che la distrubuzione è più appuntita al centro di una normale.

In sostanza, la media di movimento viene leggermente sacrificata in cambio di altre risorse, ma con alcuni casi eccellenti in cui questa viene aumentata di molto.

### Varianza Movimento

Rispetto al dado classico, con varianza 3.5, qui la varianza media è molto più alta, indicando probabilmente una tendenza al risk seeking generale.

Tuttavia, la deviazione standard da quel valore è molto alta; infatti si nota la presenza di tre casi ‘eccellenti’ con varianza altissima, ma anche che la maggior parte dei valori sono attorno alla media.

Il valore di Skewness avvalora quanto detto, essendo più alto di quello di una semi normale, così come quello di Kurtosis, che indica una distribuzione più appuntita di quella normale.

### Media coins

Come nel dado tradizionale anche qui il numero di monete è quasi nullo.

La deviazione standard non è molto alta: questo indica che ad ogni lancio di dado il valore atteso di monete non è mai più alto di +1 o più basso di -1.

Skewness indica una direzione della deviazione standard leggermente verso sinistra.

La Kurtosis è negativa (quindi funzione più piatta di una normale) nonostante il forte picco centrale a causa dei due massimi locali negli intervalli [-1, -0.5] e [0.5, +1].

### Varianza coins

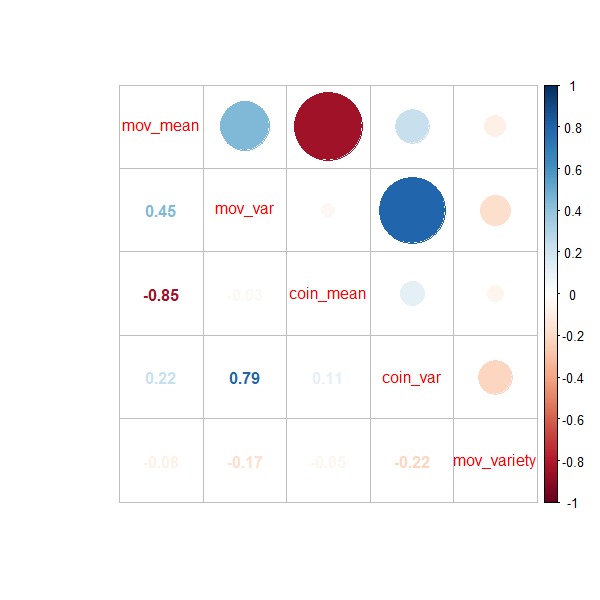
La varianza di monete ha delle features simili alla varianza sul movimento: infatti ha una deviazione standard relativamente alta a causa dei pochi ‘casi eccellenti’; è orientata verso destra molto più di una semi normale ed è molto più appuntita di una normale.

### Varietà

La varietà dei dadi è sempre minore di quella del dado standard (uguale a 6), con una media generale molto minore (meno della metà) e una deviazione standard modesta.

Rispetto alla media è più orientata a sinistra; la Kurtosis si avvicina molto a quella della distribuzione uniforme (uguale a -6/5). Questo suggerisce che questo valore sia distribuito uniformemente su tutti i dadi, senza particolari correlazioni.

### Correlazioni



Come precedentemente supposto, la varietà non è correlata a nessun’altra variabile.

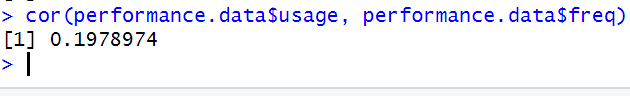
Esistono due forti correlazioni:

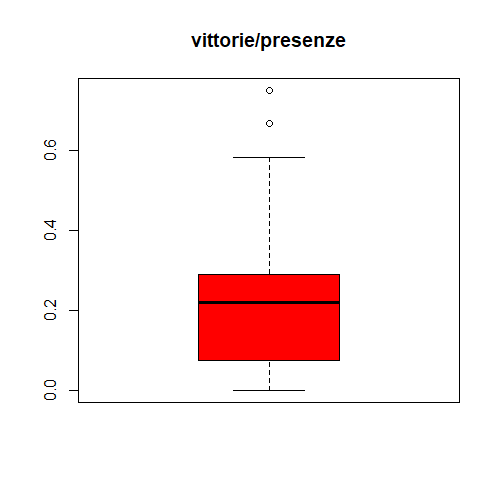
* Negativa tra media di movimento e media coins: nel gioco questi due beni desiderabili sono anche contrari; pertanto, il giocatore che vorrà un ritorno positivo di monete dal dado dovrà veder calare la propria media di movimento e il giocatore che vorrà aumentare la propria media movimento sarà costretto ad accettare facce che lo obbligano a perdere monete (e a muoversi di 0)
* Positiva tra le due varianze: ci sono dadi che sono rischiosi in tutto e per tutto (sia sul movimento che sulle monete) e dadi avversi ad entrambi i rischi

## games.data

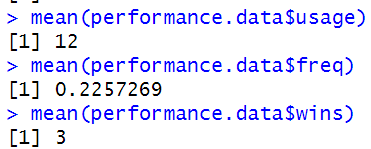
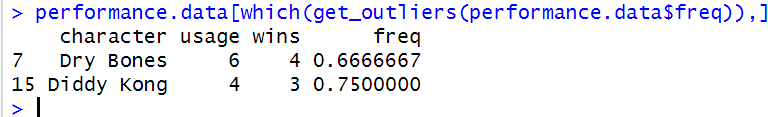
Preso il dataset delle partite ho calcolato la frequenza di vittorie (vittorie/partite giocate) per ognuno dei 20 personaggi.

Si nota che non esiste correlazione tra numero di partite giocate e vittorie. Pertanto, personaggi più e meno utilizzati possono essere messi sullo stesso piano.





L’utilizzo del boxplot evidenzia la presenza di due outliers:



I due personaggi outliers sono quelli che, giocando la metà o meno della metà di partite rispetto alla media, hanno vinto un numero di partite pari o superiore a quello medio di tutti i personaggi.

Prima di proseguire ho scelto di ‘penalizzare’ (ma non in maniera eccessiva) questi due personaggi, portando la loro frequenza al limite superiore del boxplot.

## levels.data

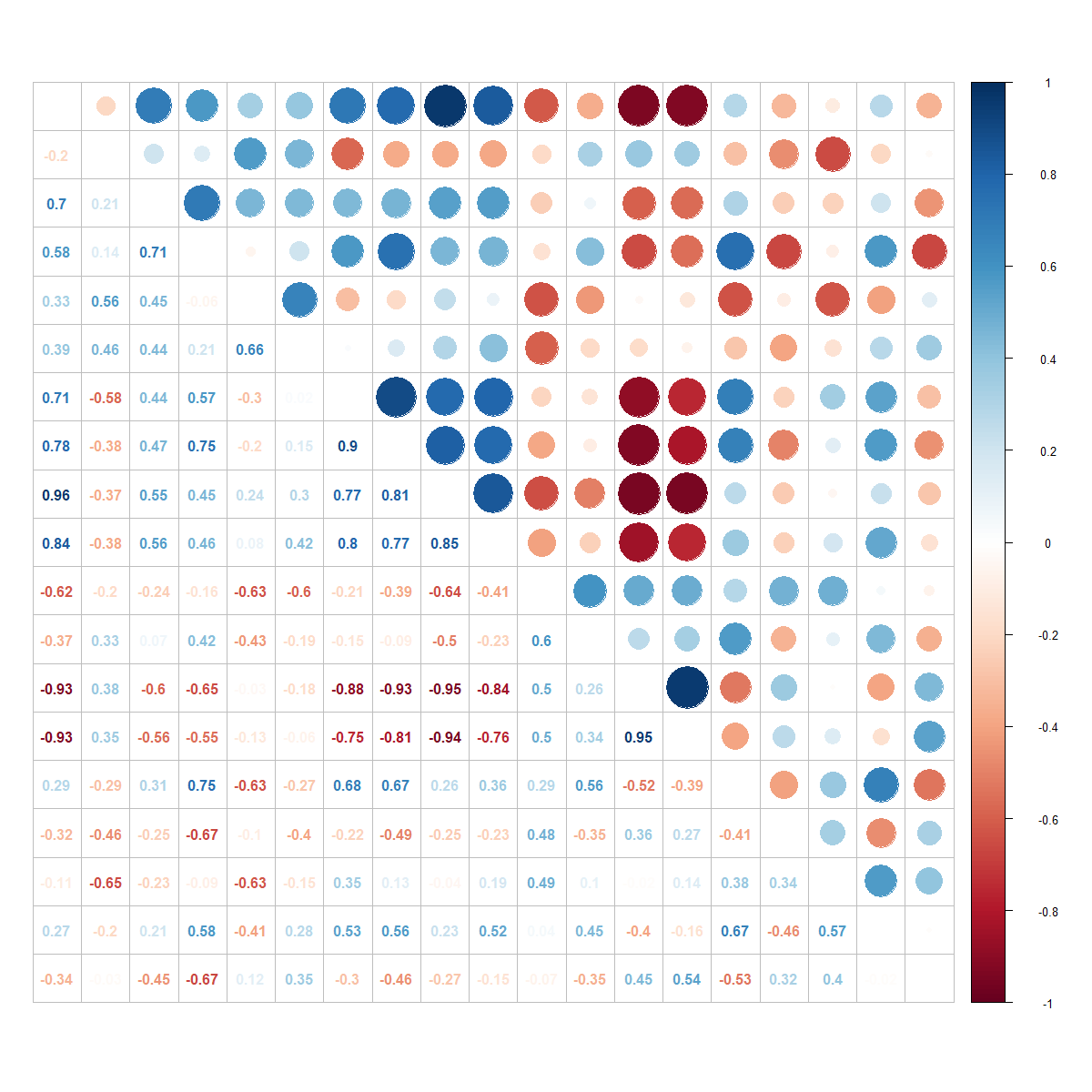
Attraverso la libreria rPref ho calcolato tutti i livelli di preferenza su tutti i trade-off possibili tra le risorse. Questi sono in tutto 19.

Avendo ogni trade-off un diverso numero di livelli, ho deciso di normalizzare ogni vettore dei livelli su un intervallo [0, 1] con questo significato:

* 0: il livello di preferenza uguale a 1 (quindi dado migliore)
* 1: il livello di preferenza più alto possibile (quindi dado peggiore)

### Correlazioni

Ora posso visualizzare le correlazioni:



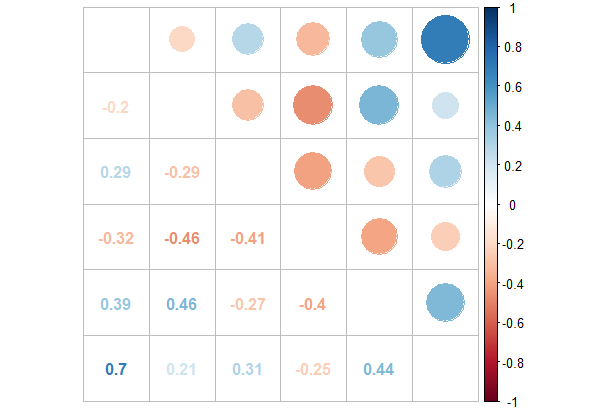
Quando due livelli sui trade-off sono altamente correlati (positivamente o negativamente) significa che uno dei due dà le stesse informazioni sulla forza dei dadi dell’altro. È quindi opportuno toglierne almeno uno dei due. Ho deciso di tagliare ad una soglia di correlazione > 0.7 (in valore assoluto).

In particolare, per l’operazione di scrematura, ho individuato dei trade-off ‘fondamentali’ e sono andato ad eliminare quelli correlati con questi.

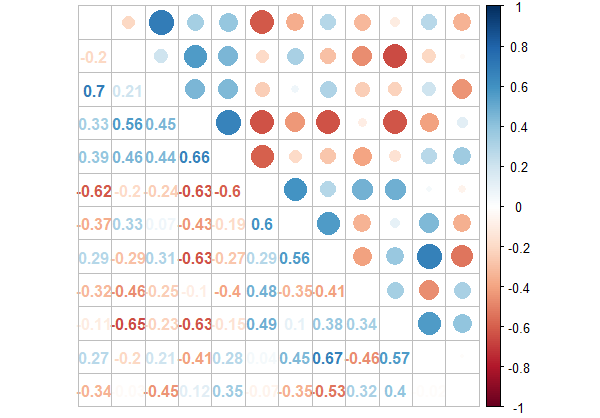
Con trade-off ‘fondamentali’ intendo quelli che umanamente appaiono avere più significato:

* I trade-off delle medie con le proprie varianze (sia p.o.v. risk seeker che risk adverse)
* Il trade-off tra media movimento e varietà movimeto
* Il trade-off tra le due medie

Prima di procedere ho controllato che questi non fossero correlati tra loro:



Alla selezione sono sopravvissuti 12 trade-off, che sono tutti entro la soglia di taglio.



In ordine dall’alto verso il basso questi sono:

* Media movimento – varianza movimento (p.o.v. risk seeker)
* Media movimento – varianza movimento (p.o.v. risk adverse)
* Media movimento – media monete
* Media movimento – varianza monete (p.o.v. risk adverse)
* Media movimento – varietà movimento
* Media monete – varianza movimento (p.o.v. risk adverse)
* Varianza movimento (p.o.v. risk adverse) – varianza monete (p.o.v. risk seeker)
* Media movimento - varianza monete (p.o.v. risk seeker)
* Media movimento - varianza monete (p.o.v. risk adverse)
* Media movimento – varietà movimento
* Varietà movimento - varianza monete (p.o.v. risk seeker)
* Varietà movimento - varianza monete (p.o.v. risk adverse)

### Performance

I livelli sono su un intervallo [0, 1] in cui:

Se a < b allora a è più forte di b

Mentre la frequenza di vittorie, sempre su [0, 1], vede:

Se a < b allora a è più debole di b

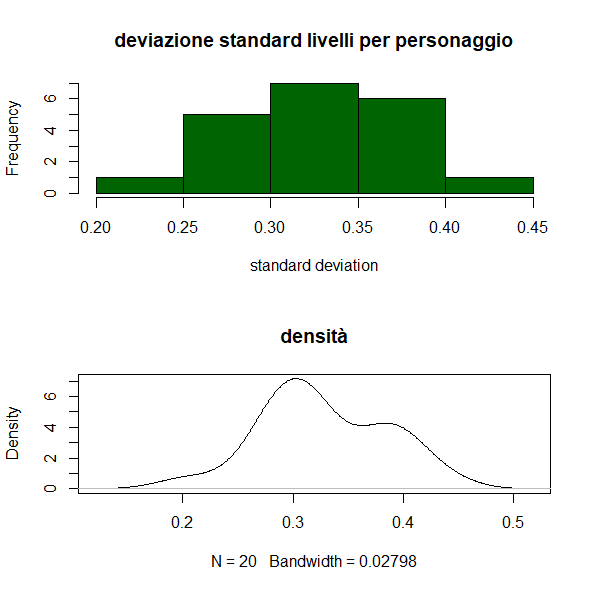
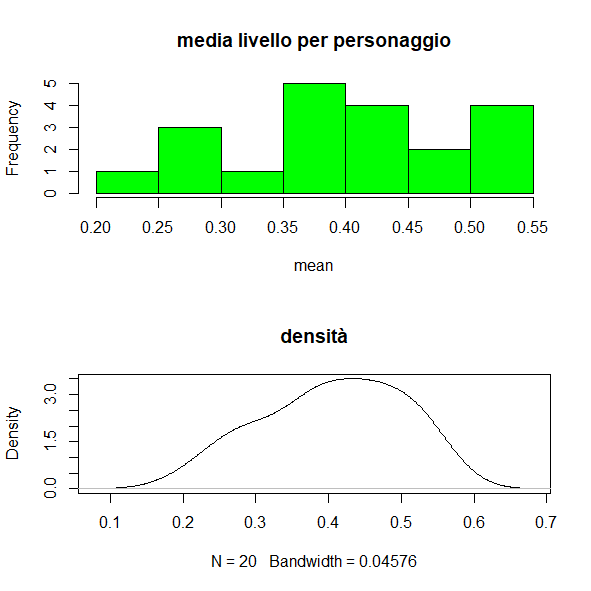
(avendo vinto meno partite)

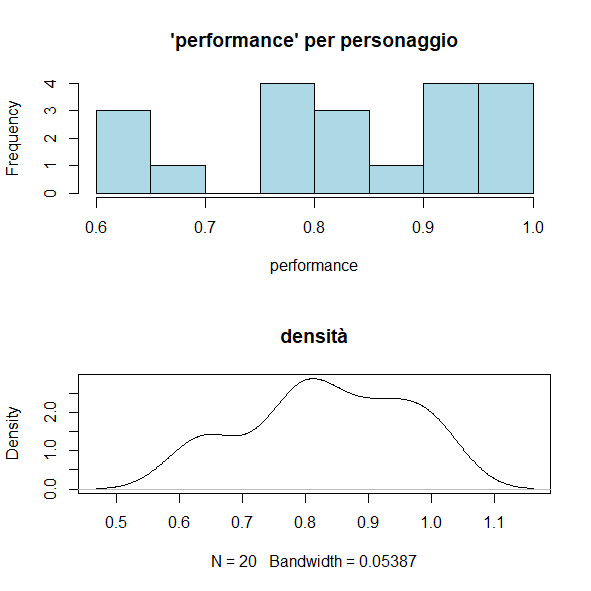
Ho deciso di considerare, in luogo della frequenza, una sua funzione, affinché il suo valore abbia lo stesso significato del valore del livello (ossia: Se a < b allora a è più forte di b).

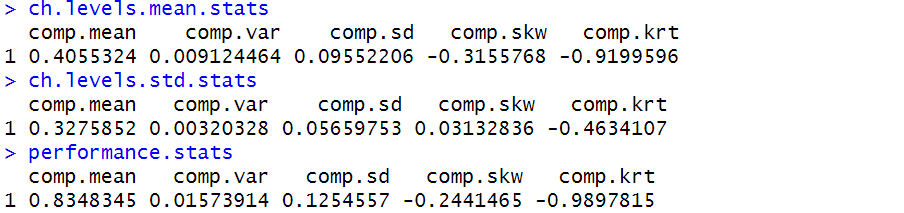
Ho scelto in particolare la funzione che mappa la frequenza sull’intervallo [0.5, 1] con significato:

* 0.5: la situazione ideale in cui il personaggio ha il 100% di vittorie (dunque dado migliore)
* 1: la situazione in cui il personaggio non ha mai vinto (quindi dado peggiore)

Adesso si considera per ogni personaggio la media dei propri livelli, la deviazione standard e la performance







### Media livelli per personaggio

In media, i personaggi hanno un livello di 0.4.

La media di ogni personaggio ha una deviazione standard rispetto alla media complessiva di 0.09, quindi relativamente esigua.

Questo evidenzia il fatto che i personaggi tendono a comportarsi in maniera complessivamente ‘buona’ senza troppi casi ‘eccellenti’.

La Skewness negativa indica che tendenzialmente i personaggi hanno livelli medi minori della media complessiva e dalla Kurtosis abbiamo che la media ha una distribuzione che assomiglia più ad una uniforme (essendo vicina a -6/5) che non ad una normale.

### Deviazione standard livelli per personaggio

Le deviazioni standard hanno una media alta (0.32) ed una esigua deviazione standard (0.06).

Quindi, sebbene come detto prima, i personaggi tendono ad avere una media ‘buona’, questa deriva dal fatto che i dadi sono ottimi in alcuni trade-off e scarsi in altri.

Dalla Skewness e dalla Kurtosis abbiamo che questa funzione ha simmetria simile ad una normale, ma è più piatta.

### Performance per personaggio

La media è molto alta, così come la deviazione standard: mediamente i personaggi hanno una cattiva performance ma con tanti casi sia migliori che peggiori.

La Skewness indica che in generale le performance sono migliori di questa media.

La Kurtosis è simile a quella della distribuzione uniforme.

# Regression

Al fine di trovare quali livelli sui trade-off sono veramente descrittivi della forza di un personaggio utilizziamo questi livelli come regressori della performance.

Questo perché la regressione ci dice quali trade-off sono statisticamente significativi e tramite i coefficienti permette di confrontare l’utilità che si ha nell’essere forti in un trade-off piuttosto che in un altro.

Ricordo i significati dei valori dei livelli e della peformance:

* Livelli
  + 0: il livello di preferenza uguale a 1 (quindi dado migliore)
  + 1: il livello di preferenza più alto possibile (quindi dado peggiore)
* Performance
  + 0.5: la situazione ideale in cui il personaggio ha il 100% di vittorie (dunque dado migliore)
  + 1: la situazione in cui il personaggio non ha mai vinto (quindi dado peggiore)

I coefficienti saranno interpretati secondo questo significato:

* Maggiori di zero: la performance cresce più velocemente al crescere del valore del livello e cresce più lentamente al decrescere del livello. Tanto più questo valore sarà grande tanto più sarà ‘importante’ avere un livello prossimo allo 0 per tenere la performance bassa. Questi trade-off sono quelli i cui livelli sono descrittivi.
* Minore di zero: la performance decresce più velocemente al crescere del valore del livello e decresce più lentamente al decrescere del livello. Tanto più questo valore sarà piccolo tanto più sarà ‘conveniente’ avere un livello prossimo a 1 per abbassare la performance. Questi trade-off sono addirittura dannosi, perché dadi forti su questo trade-off beneficiano meno dei dadi deboli.

## Esempio

Supponiamo di aver ricavato i seguenti coefficienti:

* Trade-off media movimento – varianza movimento risk adverse: 0.1
* Trade-off media movimento – varianza movimento risk seeker: 0.9

Ora vogliamo scegliere il migliore tra due dadi con livelli:

* Dado RiskSeeker: liv risk adverse = 1; liv risk seeker =0
* Dado RiskAdverse: liv risk adverse = 0; liv risk seeker = 1

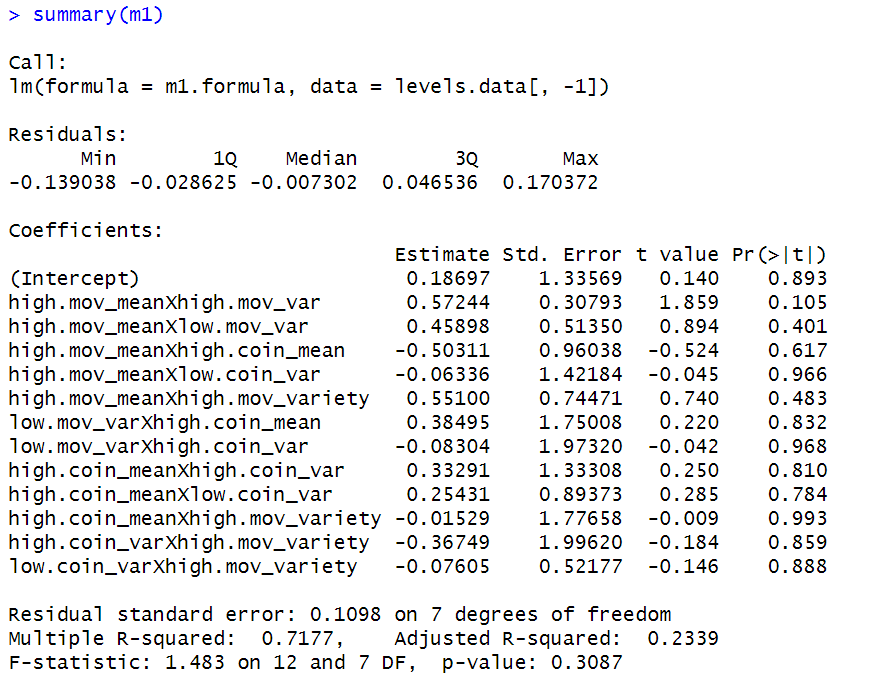
Quale scegliere?

Il dado RiskSeeker è migliore, infatti la sua performance cresce di 0.1 contro lo 0.9 del RiskAdverse. Quindi il trade-off media-varianza risk seeker descrive meglio la forza effettiva del dado rispetto al media-varianza risk adverse.

Se come coefficienti avessi trovato gli opposti (risk-adverse = -0.1, risk-seeker = -0.9) sarebbe meglio il dado RiskAdverse che beneficia maggiormente della negatività del coefficiente risk-seeker.

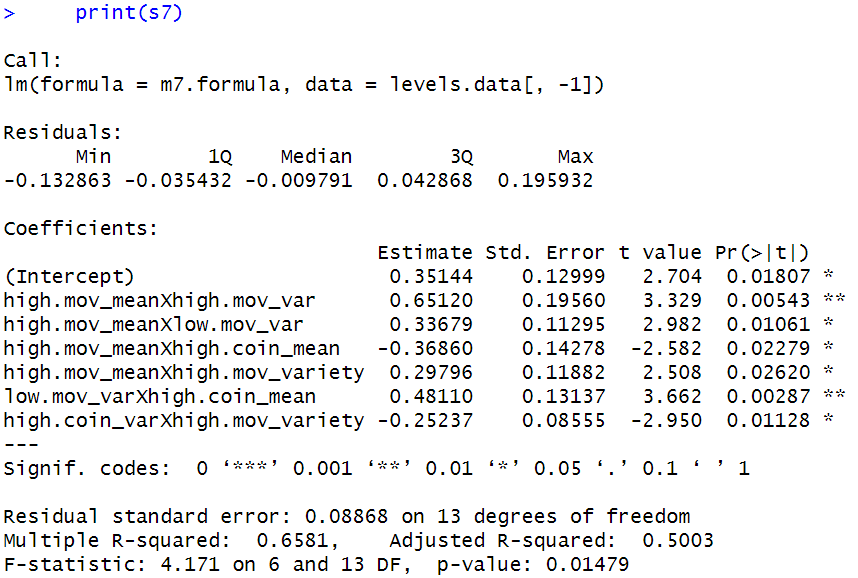
## Backward Elimination

Si parte dal modello di regressione con tutte le variabili e poi si tolgono progressivamente quelle meno significative.



Il primo modello non ha variabili statisticamente significative. Per questo, ho deciso di essere molto cauto nell’eliminare le variabili: ad ogni passo ne tolgo solo una, fino ad arrivare al modello con tutte variabili significative (almeno 0.01).

Il modello finale ha sei variabili:



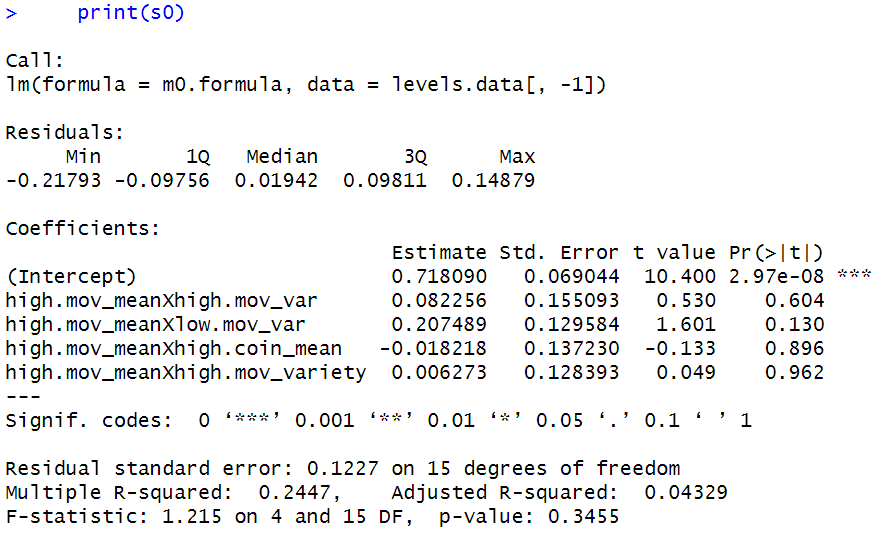
Il coefficiente indica che il 66% delle variazioni di performance sono ben descritte dalle variazioni nei livelli dei trade-off.

Nonostante il modello di partenza avesse il 72%, questo modello è migliore perché:

* è maggiore nel modello finale, indicando che nel togliere predittori ho migliorato il modello
* è minore di 0.05: si può rigettare l’ipotesi nulla che questo modello abbia le stesse prestazioni del modello senza predittori

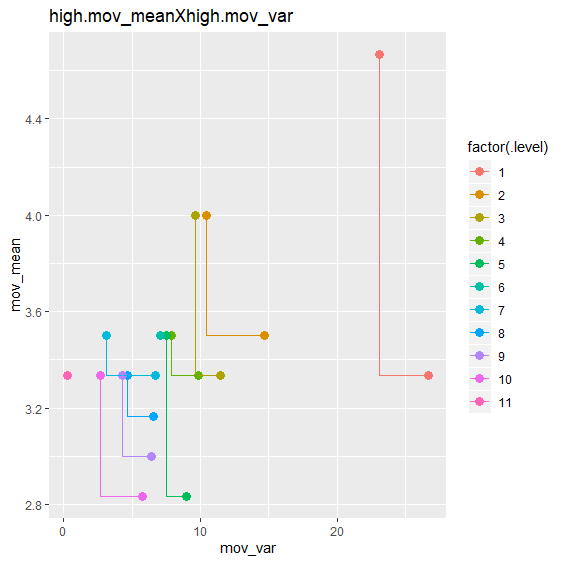
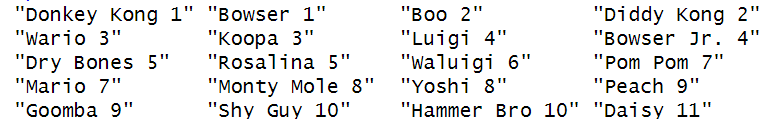
Essere molto cauto nell’eliminazione delle variabili è stato positivo. Infatti, due trade-off del modello finale avevano, nel modello iniziale, p-value del test t maggiore di 0.8.

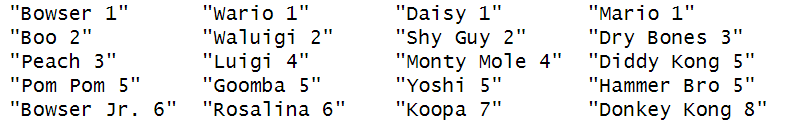
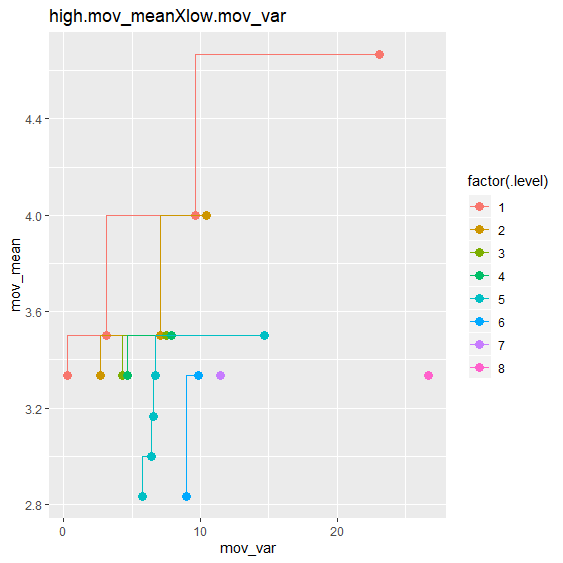
Se li togliessi dal modello finale avrei:

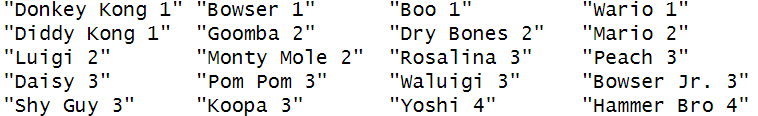
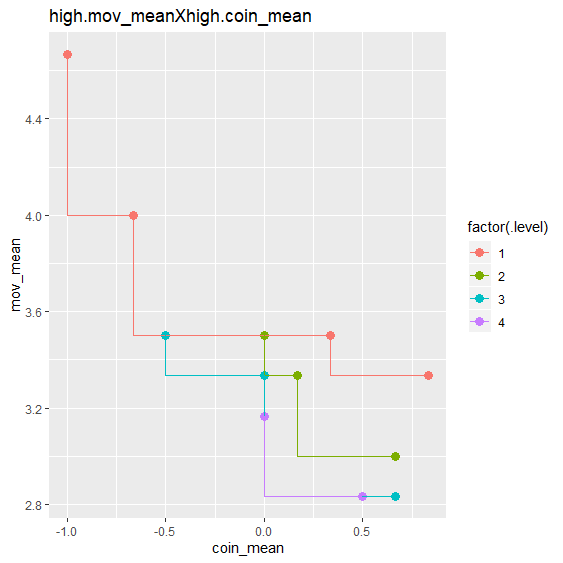


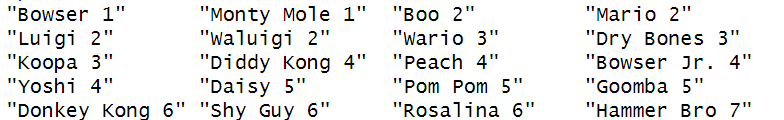
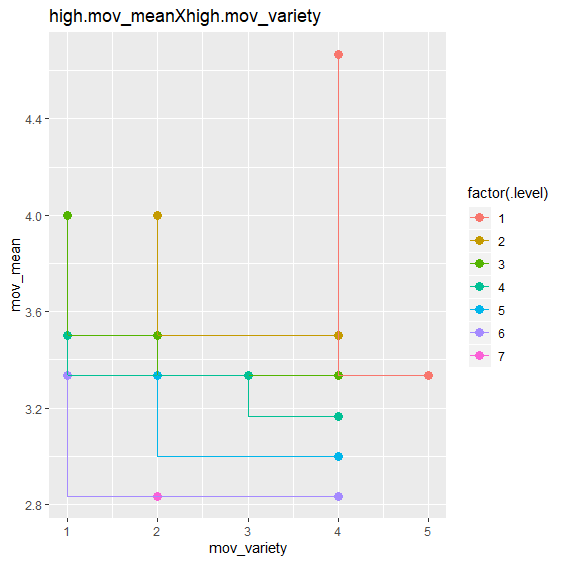
Che denota un netto peggioramento.

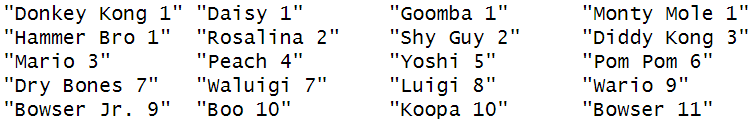
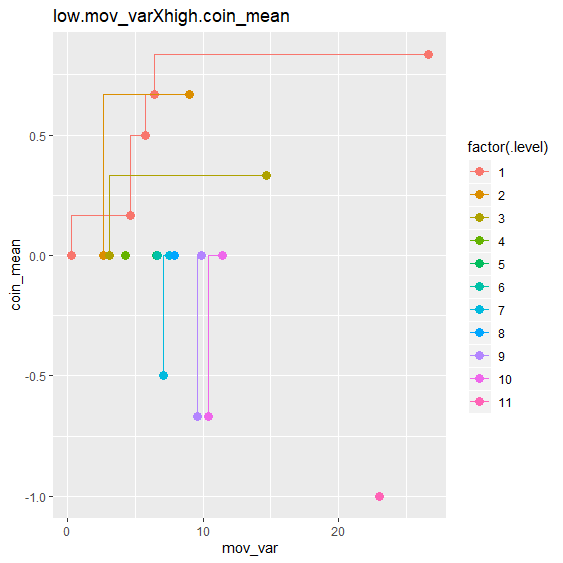
I sei trade-off del modello sono i seguenti:

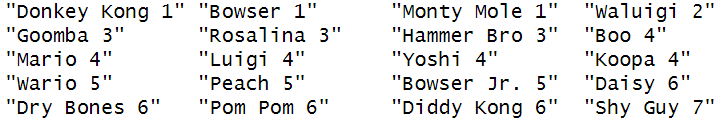
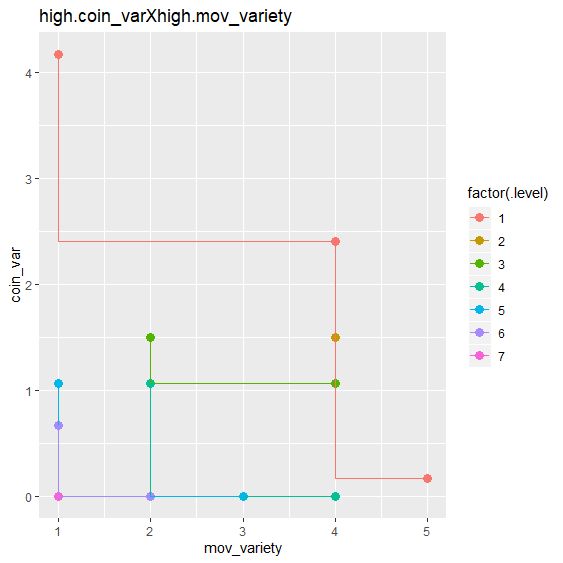
 



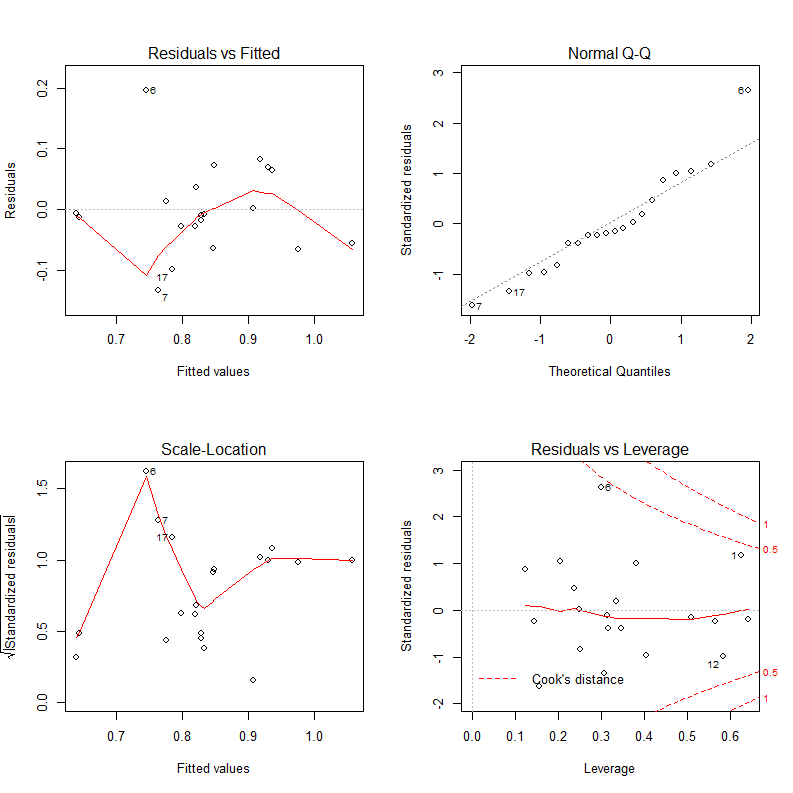




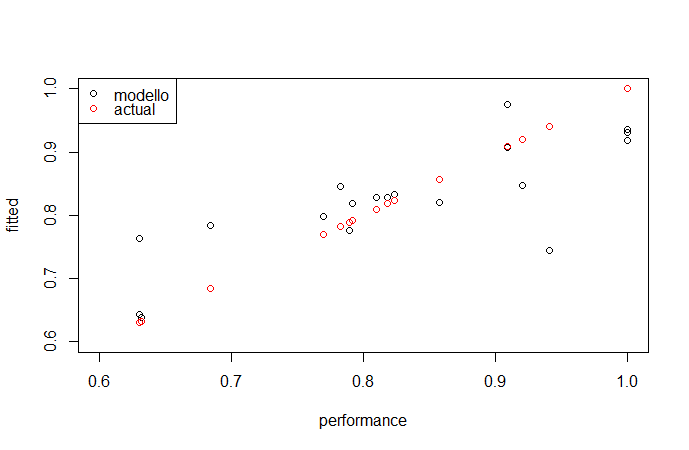




## Analisi del modello

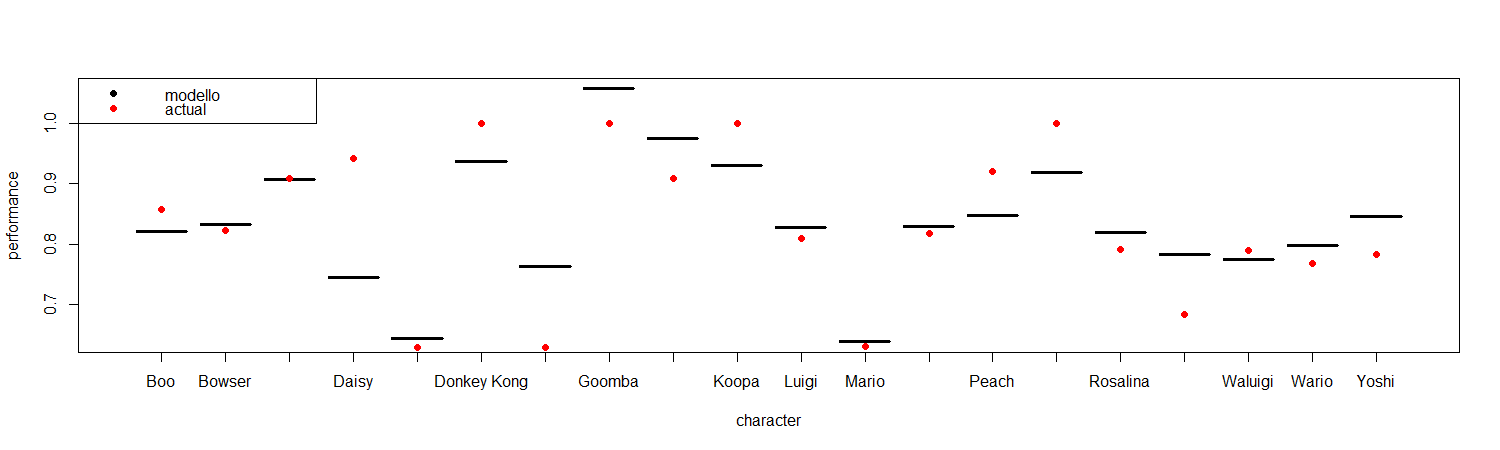


* I residui non mostrano di avere un pattern non lineare
* Nel QQ plot i residui si adattano abbastanza bene alla linea retta, dunque possiamo assumere che i residui siano normalmente distribuiti
* I residui appaiono distribuiti in maniera random sui valori predetti
* Tutti i punti sono entro la distanza di Cook, pertanto non ci sono ‘influential cases’ che, se esclusi, darebbero luogo ad un diverso modello predittivo

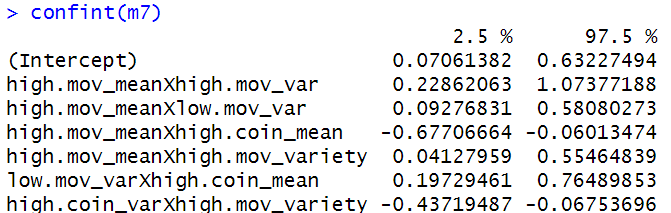
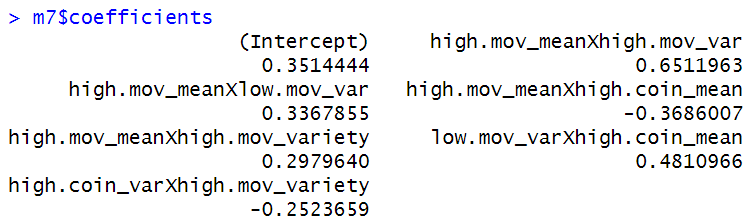


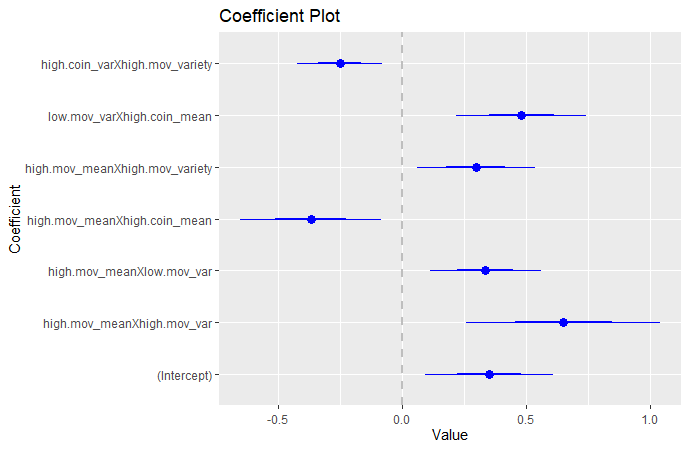
Notiamo che nella parte centrale (indicativamente tra 0.7 e 0.9) il modello predice abbastanza bene le performance; al di fuori, il modello può discostarsi di molto dal valore reale, con casi in cui ha un’ottima prestazione.

La classifica dei personaggi più forti secondo il modello:



# Results





Osservazioni:

* Il trade-off media-varianza è presente in entrambe le versioni risk seeker e risk adverse con coefficiente positivo. Questo indica che entrambi i trade-off sono buoni, ma il maggiore coefficiente del risk seeker suggerisce che il gioco premi la ricerca del rischio.
* Il trade-off tra le due medie ha coefficiente negativo. Pertanto, il modello sembra suggerire che scambiare movimento con monete è dannoso, e quindi potrebbe essere meglio avere tutte e 6 le caselle di movimento.

Tuttavia, ha coefficiente positivo il trade-off tra media monete e varianza movimento (risk adverse). Questo è interpretabile come: è positivo aumentare la media monete per aumentare la varianza di movimento, ma è negativo se si vuole diminuire la media movimento.

* Nei confronti della varietà si può ricavare un ragionamento simile: è positivo diminuire la varietà per aumentare la media movimento, ma è negativo per aumentare la varianza di monete.

Dai coefficienti si evince una forte presenza di trade-off sulla media di movimento nel determinare la forza dei dadi (il che è comprensibile) ma sorprendentemente la media di monete non è altrettanto importante, dato che è in due trade-off, uno positivo e uno negativo.

Essendo però i livelli di confidenza relativamente ampi e in gran parte sovrapposti, nessuna delle osservazioni precedenti può essere vista come regola assoluta.

# Conclusione

Applicando lo stesso metodo su un numero molto più ampio di partite può essere possibile trarre vere conclusioni sulla composizione dell’‘offerta dei dadi’ e intervenire per migliorarne la qualità per una ipotetica futura versione del gioco (ad esempio, decidere di migliorare i dadi forti sul trade-off tra medie per rendere il trade-off più rilevante).

L’idea è che un simile ragionamento possa essere applicato in molti altri casi. In ambito marketing, se ho un prodotto con numerose ‘risorse’, può essere utile capire quali trade-off tra risorse sono veramente importanti nel determinare il successo dello stesso.

Per esempio:

Un’automobile ha numerose caratteristiche (capienza bagagliaio, cilindrata, consumi ecc.) tra le quali, in fase di costruzione, avvengono necessariamente dei trade-off. Ogni auto, confrontata con tutte le altre sul mercato, si pone su un certo livello di preferenza su ogni trade-off. Se si riuscisse a capire quali livelli sono importanti nel determinare il numero di vendite, potrei, l’anno successivo, puntare su diverse risorse.

È comunque un processo che rimane nell’ambito del ruolo esecutivo del marketing, in quanto le ‘preferenze’ consistono in un calcolo teorico e non tengono conto dei processi decisionali, oltre che di fattori esterni al prodotto (es. pubblicità per le auto) né di variabili qualitative (es. il colore delle auto).