
STUDIO DELLA POTENZA DEI TORNADO IN BASE ALLE LORO CARATTERISTICHE

Barjam Bardhi Matr 872396, Daniele Quattrocchi Matr 825418, Raffaele Moretti Matr 794537
Laurea Magistrale in Data Science
Report di Foundations of Probability and Statistics
Università degli Studi di Milano-Bicocca

Indice

1	INTRODUZIONE	2
1.1	Come si definisce un tornado?	2
1.2	Come si forma un tornado?	2
1.3	Come si misura la potenza di un tornado?	2
2	DATASET	3
2.1	Fonte e attributi	3
2.2	Preprocessing dei dati	4
3	STATISTICA DESCRITTIVA	4
3.1	Analisi univariata - Esplorazione dati	4
3.2	Analisi bivariata - Studio della magnitudine	12
3.3	Analisi delle dipendenze	15
3.3.1	Test Chi-quadro	15
3.3.2	Eta-quadro	16
3.3.3	Correlazione	16
4	STATISTICA INFERENZIALE	17
4.1	Campionamento casuale stratificato	17
4.2	Modello statistico: regressione lineare multipla	17
4.3	Previsione utilizzando il modello	18
5	Conclusioni	18

DISCLAIMER

Questo studio è stato fatto prendendo i dati dell'agenzia governativa degli Stati Uniti *National Weather Service* disponibili al pubblico sul proprio sito ufficiale e non ha alcun fine commerciale. Tutti i risultati ottenuti potrebbero differire da quelli pubblicati da tale agenzia governativa. Non ci assumiamo alcuna responsabilità per tale differimento. La ricerca è esclusivamente da intendersi come esercizio per migliorare le proprie abilità nel campo della data science.

ABSTRACT

In questo elaborato andremo ad analizzare la popolazione dei tornado che si sono verificati negli Stati Uniti tra il 1950 e il 2015. L'obiettivo del nostro studio è di individuare quali sono le caratteristiche di un tornado dalle quali dipende l'intensità di tale fenomeno. Dato che le caratteristiche principali di un tornado direttamente visibili sul campo di osservazione sono la distanza percorsa dal tornado e la sua larghezza, il momento e il luogo di accadimento, abbiamo deciso di capire se osservando un tornado sul campo sia possibile comprendere la sua intensità e dunque potenziale minaccia. A tal fine il prodotto della nostra ricerca sarà un modello che ci permetta di prevedere la magnitudine di un tornado, sia ipotizzando le caratteristiche che il fenomeno potrebbe avere in un certo luogo e in un certo momento, sia introducendo nel modello dati osservati realmente sul campo, senza aspettare che la potenza del tornado venga stabilita successivamente alla dispersione del tornado.

1 INTRODUZIONE

1.1 Come si definisce un tornado?

Il tornado è un vortice d'aria che sale all'interno di una nube. Nella fase iniziale o in quella più matura, tutti i temporali sono caratterizzati da correnti ascendenti, chiamate "updrafts". Queste correnti sono formate dall'aria calda e umida che alimenta i temporali, ma in alcuni casi la colonna di aria ascendente diventa un vortice assumendo la forma di imbuto. Spesso un tornado è localizzato sul bordo delle correnti ascendenti, vicino alle correnti discendenti contenenti pioggia o grandine. Il fenomeno tornadico, per quanto possa verificarsi ovunque escludendo le zone polari, si presenta in alcune zone geografiche particolari che presentano con regolarità le condizioni ottimali per il suo sviluppo. La più famosa è la cosiddetta "Tornado Alley" (Viale dei Tornado), negli Stati Uniti (il paese maggiormente colpito da questo fenomeno per frequenza e intensità). Questa zona comprende lo Stato del Texas, l'Arkansas, l'Oklahoma, il Nebraska, il Kansas, il Missouri, l'Iowa, il Dakota del Sud e l'Illinois. Lo stato maggiormente coinvolto è sicuramente il Texas, dove si registra la maggior densità di tornado.

1.2 Come si forma un tornado?

Il National Weather Service¹ (NWS), l'agenzia governativa statunitense che si occupa delle previsioni meteorologiche, definisce un tornado come "una violenta rotazione di una colonna d'aria in contatto con il suolo che scende da un temporale"². In altre parole un temporale è il primo stadio per la creazione di un tornado. Quando anche le altre condizioni sono favorevoli, la nube temporalesca crea uno o più tornado. Le condizioni richieste per lo sviluppo della nube temporalesca sono:

1. *Forte riscaldamento del suolo, con umidità nei bassi e medi livelli dell'atmosfera;*
2. *Aria più fredda e secca in quota;*
3. *Una forza ascendente. È necessario qualcosa che causi il sollevamento dell'aria; la più comune forza ascendente è la spinta di Archimede (tramite il riscaldamento dell'aria vicino al suolo da parte dei raggi solari, l'aria diventa più leggera);*

4. *Aria instabile. Ciò significa che l'aria continua a salire una volta che ha iniziato la sua ascesa. E questo è tanto più favorito quanto più l'aria in quota è fredda;*

5. *Un'altra "spinta" ascendente può essere provocata da un'irruzione di aria fredda la quale, incuneandosi con forza al di sotto della preesistente aria più calda, ne provoca l'ascesa (fronte freddo).*

1.3 Come si misura la potenza di un tornado?

La potenza dei tornado si misura odiernamente negli Stati Uniti e nel Canada attraverso la "Scala Fujita Avanzata"³, che prende il nome da T. Theodore Fujita, un professore dell'Università di Chicago, in quanto egli ha originariamente ideato tale scala di misurazione. Infatti, inizialmente, i tornado venivano classificati mediante "La Scala Fujita", che è stata poi migliorata nel 2007 dal comitato del National Oceanic And Atmospheric Administration (NOAA). Da ciò deriva il termine "avanzata".

La "Scala Fujita Avanzata", a differenza della "Scala Fujita", classifica i tornado inizialmente in funzione della velocità del vento e, successivamente, in base ai danni che essi hanno prodotto all'ambiente, agli edifici e ad altre infrastrutture o manufatti costruiti dall'uomo, quantificati mediante 28 indicatori di danno⁴, elaborati dal comitato del "NOAA"⁵. Si noti che la scala fujita originaria non presentava tali indicatori che dunque costituiscono l'innovazione principale nella odierna misurazione della potenza dei tornado. Infatti, la scala originale è stata fortemente criticata in quanto classificava soggettivamente i tornado, ossia esclusivamente in base ai danni che esso provocava. Questa classificazione mostra comunque la velocità del vento per le diverse potenze del tornado in quanto all'aumentare della velocità del vento aumenta la potenziale distruttività del tornado.

La velocità del vento viene calcolata mediante una rete radar ad effetto doppler introdotta dal National Weather Service nel "1990"⁶.

Di seguito viene mostrata una tabella che sintetizza le diverse intensità dei tornado per la Scala Fujita Avanzata.

¹<https://www.spc.noaa.gov/wcm/data>

²[http://www.centrometeo.com/articoli-reportage-approfondimenti/fisica-atmosferica/4301-fenomeno-distruttivo-tornado-par. Come si formano i tornado: una guida base](http://www.centrometeo.com/articoli-reportage-approfondimenti/fisica-atmosferica/4301-fenomeno-distruttivo-tornado-par.-Come-si-formano-i-tornado:-una-guida-base)

³<https://www.spc.noaa.gov/efscale/>

⁴<https://www.spc.noaa.gov/efscale/ef-scale.html>

⁵[https://www.spc.noaa.gov/efscale par. The Enchained Fujita Scale/](https://www.spc.noaa.gov/efscale-par.-The-Enchained-Fujita-Scale/)

⁶<https://www.spc.noaa.gov/efscale/>

The Enhanced Fujita Scale

Scale	Wind Speed (mph)	Damage
EF0	65 - 85	Light damage: Some damage to chimneys. Branches broken from trees and some trees blown over.
EF1	86 - 110	Moderate damage: Moving cars blown off roads, mobile homes overturned.
EF2	111 - 135	Considerable damage: Large trees snapped or uprooted, cars lifted off the ground. Mobile homes destroyed.
EF3	136 - 165	Severe damage: Trains overturned, most trees uprooted, walls of homes destroyed.
EF4	166 - 200	Devastating damage: Well built houses destroyed, large objects thrown.
EF5	> 200	Incredible damage: Cars thrown more than 100 metres, strong buildings swept away. Extensive destruction.

Figura 1: Enhanced Scala Fujita⁷

Benchè la "Scala Fujita Avanzata" classifichi i tornado riferendosi direttamente alla velocità del fenomeno, abbiamo deciso di riportare tale immagine per far capire a lettore i danni che mediamente il fenomeno genera nelle sue diverse potenze. La scala termina con un'intensità del fenomeno pari a 5 (magnitudine massima) in quanto il massimo potenziale teorico delle variabili atmosferiche, ossia le variabili che determinano la velocità del vento, non può superare il valore F5. Per fare un esempio, un ipotetico tornado di classe F12, corrisponderebbe una velocità del pari a circa 1235 km/h, ossia la velocità del suono⁸.

2 DATASET

2.1 Fonte e attributi

Il NWS mette a disposizione del pubblico, sul proprio sito⁹, il dataset di microdati relativo ai fenomeni dei tornado che si sono verificati tra il 1950 al 2015. Dunque, il tornado costituisce l'unità statistica del dataset in questione. Esso è composto da 60114 records e dai seguenti 22 attributi:

1. *om*(id tornado)
2. *yr* (year)
 - from 1950
 - to 2015
3. *mo*(month)
4. *dy*(day)
5. *date*
6. *time*
7. *tz*(time zone)
8. *st*(State)
9. *stf* (State federal code)
10. *stn*(number of tornado in a State)
11. *mag* (magnitude)
12. *inj* (injuries)
13. *fat* (fatalities)
14. *loss* (dollars loss)
15. *closs* (estimate clop loss)
16. *slat* (starting latitude)
17. *slon* (starting longitude)
18. *elat* (ending latitude)
19. *elon* (ending longitude)
20. *len* (tornado's path length)
21. *wid* (tornado's width)
22. *fc* (altered f-scale rating)
 - 0 = it isn't changed
 - 1 = it is changed

⁸https://it.wikipedia.org/wiki/Scala_Fujita, par. *Affidabilità delle misurazioni*

⁹<https://www.spc.noaa.gov>

2.2 Preprocessing dei dati

Del dataset originale abbiamo analizzato le seguenti colonne: *yr*, *mo*, *dy*, *st*, *stf*, *mag*, *inj*, *fat*, *slat*, *slon*, *len*, *wid*. Per quanto riguarda le colonne *slat* e *slon*, l'unità di misura è stata convertita in Kilometri. Successivamente si è proceduto a studiare ogni variabile considerata.

3 STATISTICA DESCRITTIVA

3.1 Analisi univariata - Esplorazione dati

yr	mo	dy	slat	slon
Min. :1950	Min. : 1.000	Min. : 1.00	Min. :18.13	Min. : -163.53
1st Qu.:1974	1st Qu.: 4.000	1st Qu.: 8.00	1st Qu.:33.24	1st Qu.: -98.60
Median :1991	Median : 6.000	Median :16.00	Median :37.09	Median : -93.95
Mean :1988	Mean : 5.972	Mean :15.88	Mean :37.16	Mean : -92.96
3rd Qu.:2003	3rd Qu.: 7.000	3rd Qu.:24.00	3rd Qu.:40.97	3rd Qu.: -86.87
Max. :2015	Max. :12.000	Max. :31.00	Max. :61.02	Max. : -64.90

Figura 2: Informazioni sul avvenimento del fenomeno

Considerati 65 anni di rilevazioni, per quanto riguarda il momento dell'accadimento del fenomeno tornadico possiamo dedurre che i tornado tendono a verificarsi vicino al 1988, tra la fine del mese di maggio e l'inizio del mese di giugno, in un intorno dei giorni 15 e 16 del mese. In particolare, date le medie delle coordinate di latitudine e longitudine, attraverso google maps, possiamo dedurre che l'area prediletta dal fenomeno tornadico ruota mediamente attorno all'indirizzo "1541-1543 State Hwy PP, Fordland, MO 65652, Stati Uniti". Tale risultato ottenuto è dovuto al fatto che il Missouri fa parte della "Tornado Alley" e per tale ragione, dimostra quanto affermato nell'introduzione.

mag	inj	fat	len	wid
Min. :0.0000	Min. : 0.000	Min. : 0.00000	Min. : 0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.: 0.000	1st Qu.: 0.00000	1st Qu.: 0.1609	1st Qu.:0.01189
Median :1.0000	Median : 0.000	Median : 0.00000	Median : 0.9656	Median :0.03658
Mean :0.7962	Mean : 1.561	Mean : 0.09687	Mean : 5.6017	Mean :0.09003
3rd Qu.:1.0000	3rd Qu.: 0.000	3rd Qu.: 0.00000	3rd Qu.: 4.8280	3rd Qu.:0.09144
Max. :5.0000	Max. :1740.000	Max. :158.00000	Max. :377.7121	Max. :4.18429

Figura 3: Informazioni sulle caratteristiche del fenomeno

Per quanto riguarda le caratteristiche con le quali il fenomeno si presenta, possiamo affermare che i tornado tendono a verificarsi con una bassa potenza, misurata mediante la Scala Fujita Avanzata, che si aggira intorno ai valori EF 0-1. Assumendo che i servizi meteorologici siano spesso in grado di prevedere con sufficiente anticipo il fenomeno tornadico, attraverso lo studio delle condizioni alla base del fenomeno, questo spiega il numero esiguo di feriti e morti dovuti ai tornado. Tuttavia la natura non è da sottovalutare in quanto si possono verificare fenomeni tornadici improvvisi che possono risultare estremamente violenti (ad esempio con le supercelle) oppure è possibile che i servizi meteorologici non riescano a predire con una sufficiente certezza la magnitudine del tornado. Questo potrebbe spiegare i valori massimi che si sono verificati per i feriti e i morti dovuti al fenomeno. L'ipotesi più intuitiva che potrebbe spiegare l'esiguo numero di morti e feriti è rappresentata dal fatto che il numero di morti e feriti dipende dalla densità di popolazione e dunque dall'area colpita. Il fatto che le medie di tali valori siano vicine allo 0 potrebbe essere spiegato dal fatto che la maggior parte dei tornado avvenga in zone poco abitate oppure deserte. Per quanto riguarda le caratteristiche fisiche dei tornado, essi tendono spesso a verificarsi con un diametro compreso entro il kilometro. In media, la lunghezza che il tornado percorre si aggira intorno tra i 5 e i 6 kilometri.

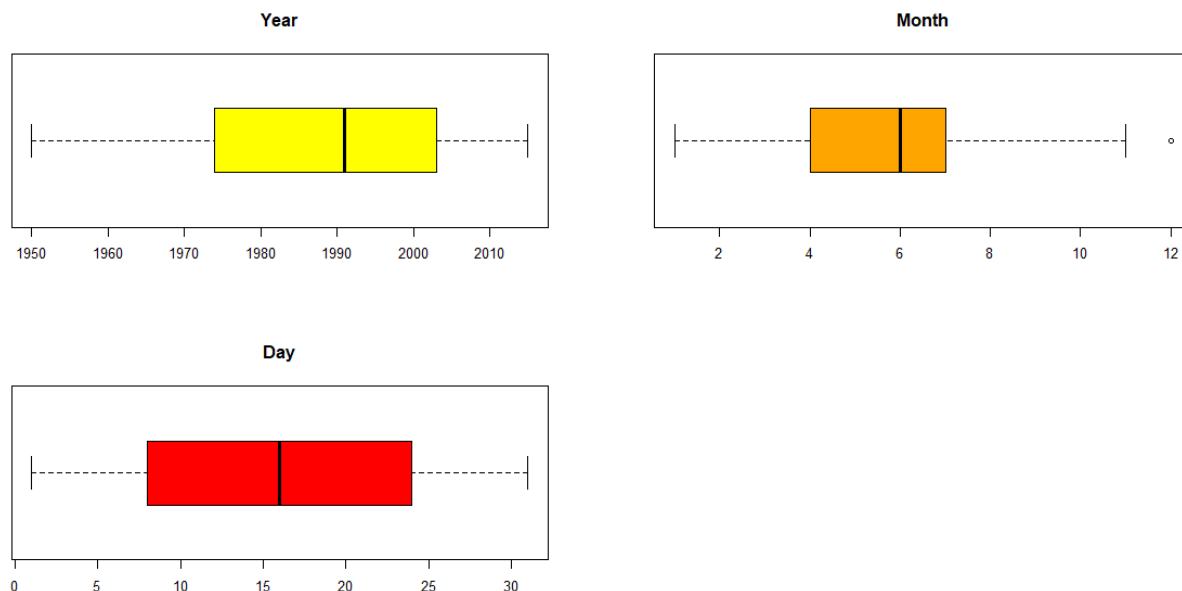


Figura 4: Distribuzione dei tornado nel tempo

Da tali grafici possiamo affermare che la frequenza dei tornado si distribuisce come segue:

1. Per quanto riguarda gli anni, il 50% dei valori centrali è compreso tra il 1974 e il 2004. La distanza dal valore minimo al valore mediano è di molto superiore rispetto alla distanza tra il valore mediano e il valore di massimo. Questo significa che i tornado hanno iniziato a concentrarsi negli anni a seguito del 1970, e poi questa concentrazione è fortemente aumentata negli anni successivi al 1990. Possiamo dunque concludere che il fenomeno tornadico è maggiormente concentrato e potrebbe essere aumentato negli ultimi 25 anni. In conclusione, considerati i risultati ottenuti, possiamo ipotizzare che la concentrazione dei tornado negli ultimi anni potrebbe essere causata un maggiore riscaldamento globale. In altri termini, maggiore è la temperatura dell'aria e del suolo, maggiore potrebbe essere la frequenza del fenomeno tornadico.

2. Per quanto riguarda i mesi, la mediana coincide con il valore medio ma il box non è simmetrico. Dato che il box si trova più vicino al valore minimo, possiamo concludere che i tornado tendono a verificarsi entro i primi 7 mesi dell'anno. Inoltre, il fatto che i tornado si verifichino nel mese di dicembre rappresenta un'anomalia. Questo risultato è supportato dai dati:

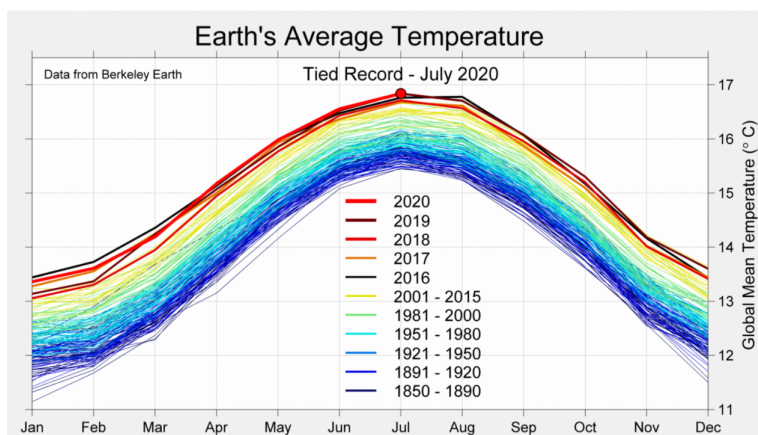


Figura 5: Image Temperature¹⁰

la temperatura mondiale media per i mesi si distribuisce come una normale ed aumenta da gennaio(picco minimo di aumento) a luglio(picco massimo della temperatura) e successivamente diminuisce fino a dicembre (picco minimo di diminuzione). In conclusione, se il riscaldamento dell'aria e del terreno è in aumento, la frequenza del fenomeno tornadico aumenta. Viceversa, se il riscaldamento dell'aria e del terreno è in diminuzione, la frequenza del fenomeno tornadico diminuisce.

3.Per quanto riguarda i giorni, il boxplot è simmetrico. Ciascun segmento che contiene il 25% dei valori osservati, mantiene la stessa variabilità, come dimostra la distanza tra il valore minimo e il primo quartile, la distanza tra il primo quartile e la mediana, la distanza tra la mediana e il terzo quartile ed infine, la distanza tra il terzo quartile e il valore massimo. Possiamo affermare che la frequenza del fenomeno tornadico si distribuisce equamente tra i giorni. In altri termini, dal risultato ottenuto, il tornado non predilige un particolare intervallo di giorni nel quale si verifica.

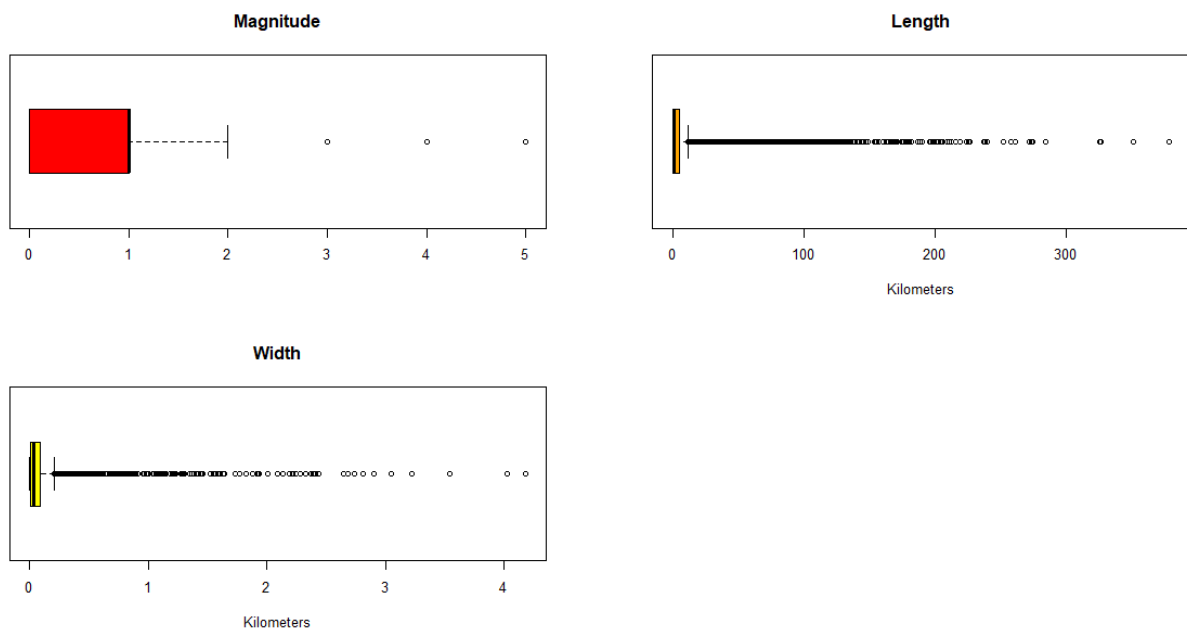


Figura 6: Distribuzione dei tornado per caratteristiche fisiche

Tali grafici mostrano le caratteristiche fisiche con le quali il fenomeno tornadico può manifestarsi. Esse si distribuiscono come segue:

1. Per quanto riguarda la magnitudine, dato che il valore minimo coincide con il primo quartile, la mediana con il terzo quartile, e data la presenza di valori anomali che non rendono il baffo superiore come il valore massimo nei dati, possiamo affermare che approssimativamente il 90/95% dei tornado osservati si manifesta con una potenza compresa tra 0 e 2 (minima-media intensità). I fenomeni tornadici che si manifestano con una magnitudine compresa tra 3 e 5 (elevata-estrema intensità) sono da considerare eventi estremamente rari (i.e. anomali). Questo risultato potrebbe essere spiegato dal fatto che più un tornado è potente, più il fenomeno necessita di una maggiore presenza di calore nel terreno e nell'aria. Infatti, a differenza dei comuni temporali, le supercelle con forma wall-cloud¹¹ ossia violenti temporali, possono dare origine a tornado estremamente potenti, tra cui quelli di potenza 4 e 5¹²

¹¹<http://www.centrometeo.com/articoli-reportage-approfondimenti/fisica-atmosferica/4301-fenomeno-distruttivo-tornado>, par. *Le Supercelle sono fondamentali per la formazione dei tornado*

¹²<http://www.fenomenitemporaleschi.it/tornado.htm>

2. Per quanto riguarda la lunghezza, dato che il valore minimo coincide con il primo quartile e la mediana, possiamo affermare che il 50% dei tornado osservati, quando si manifestano tendono a ruotare su se stessi senza procedere in alcuna direzione. Inoltre, considerando la parte dei valori restanti nel box, possiamo affermare che il 75% dei tornado non percorre più di 10 chilometri. Infine, considerando anche il baffo superiore del grafico, possiamo concludere che il 90/92% dei tornado osservati non percorrono approssimativamente più di 20 chilometri. Dal grafico si può notare che sono presenti una modesta quantità di valori anomali che potrebbe influenzare di molto la variabilità del fenomeno nella distanza percorsa.

3. Per quanto riguarda la larghezza del tornado, dato che il box si approssima al valore minimo, possiamo affermare che il fenomeno tornadico tende a manifestarsi con un diametro che è inferiore a 100 metri. Considerando il 25% dei valori rappresentati dalla distanza tra il terzo quartile e il baffo superiore, data la presenza di una modesta quantità di valori anomali, possiamo concludere che il 90/92% dei tornado osservati si manifesta con un diametro non superiore a 200 metri.

In conclusione, questi risultati ottenuti per la lunghezza e larghezza potrebbero essere spiegati dal grafico relativo alle magnitudini. L'elevata presenza di valori anomali nella lunghezza e larghezza è da attribuire alla bassa manifestazione di tornado di potenza compresa tra 3 e 5. Da questi risultati si potrebbe formulare l'ipotesi secondo la quale più un tornado è di bassa potenza, più basse sono le caratteristiche dei tornado in termini di lunghezza e larghezza. Viceversa, il contrario.

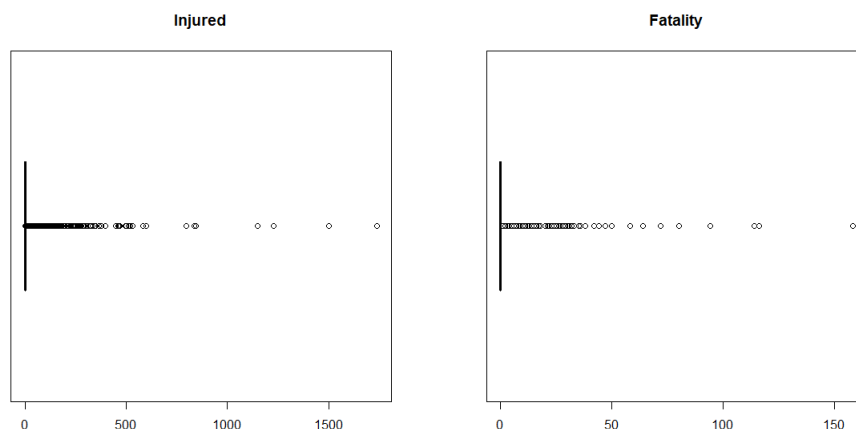


Figura 7: Distribuzione dei tornado rispettivamente per feriti e morti

Per quanto riguarda i danni che il fenomeno tornadico produce nei confronti delle persone, oltre il 92% dei tornado osservati, quando si manifestano non generano feriti e morti (i.e il valore minimo coincide con il baffo superiore). Questo accade perchè la formazione di un tornado è nella maggior parte dei casi visibile/udibile e perchè i servizi di meteorologia riescono ad avvertire per tempo le persone dell'imminente manifestazione di un tornado, avvisando di evacuare l'area colpita in funzione della velocità prevista del tornado.

Tuttavia, il grafico dei feriti, e in secondo piano quello relativo ai morti, permettono di individuare una modesta quantità di valori anomali. Questo si potrebbe spiegare con il fatto che i servizi di meteorologia hanno sbagliato la previsione del fenomeno (nella potenza e/o nel luogo di manifestazione), il fenomeno è accaduto all'improvviso con un'elevata potenza e le persone non si sono preparate, il fenomeno è accaduto di notte e quindi è stato poco visibile/udibile, il fenomeno è stato estremamente potente ¹³ e così via . ¹⁴

¹³<https://www.ilmeteo.it/notizie/meteo-storia-tremendo-tornado-oklahoma>

¹⁴<https://www.spc.noaa.gov/faq/tornado/ts-ital.pdf>

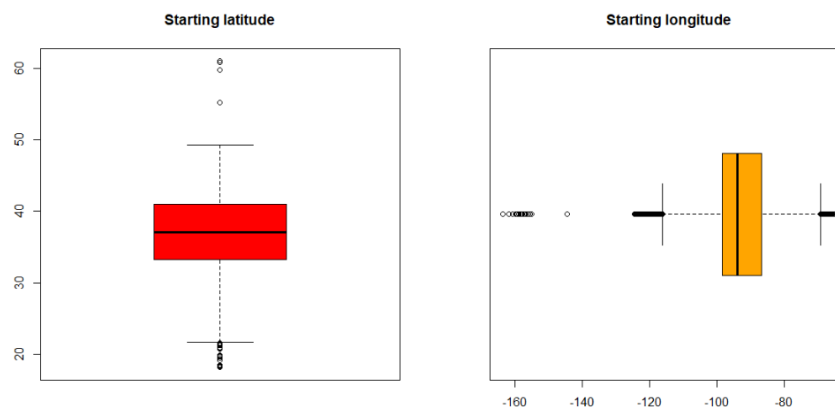


Figura 8: Distribuzione dei tornado per coordinate di inizio del fenomeno

Per quanto riguarda la distribuzione delle coordinate in cui si verificano i tornado possiamo affermare che:

1. Nel grafico relativo alle coordinate di latitudine di inizio del fenomeno, il box è simmetrico. Questo suggerisce che i tornado si distribuiscono uniformemente tra i valori compresi nel box. Inoltre, dato che la distanza tra la mediana e il baffo superiore è più piccola della distanza tra la mediana e il baffo inferiore, possiamo concludere che i tornado tendono verificarsi tra i valori di latitudine compresi tra 32 e 50, concentrandosi prevalentemente a partire dal valore di latitudine pari a 36 (valore mediano).
2. Invece, nel grafico relativo alle coordinate di longitudine di inizio del fenomeno, il box non è simmetrico e la mediana è spostata verso il primo quartile. Questo suggerisce che all'interno dei valori compresi nel box, il fenomeno tornadico tende a concentrarsi nella parte compresa tra il primo quartile e la mediana. Nonostante ciò, il grafico suggerisce che il 92/95% dei tornado osservati si distribuisce quasi uniformemente tra i valori compresi tra baffo inferiore e baffo superiore.

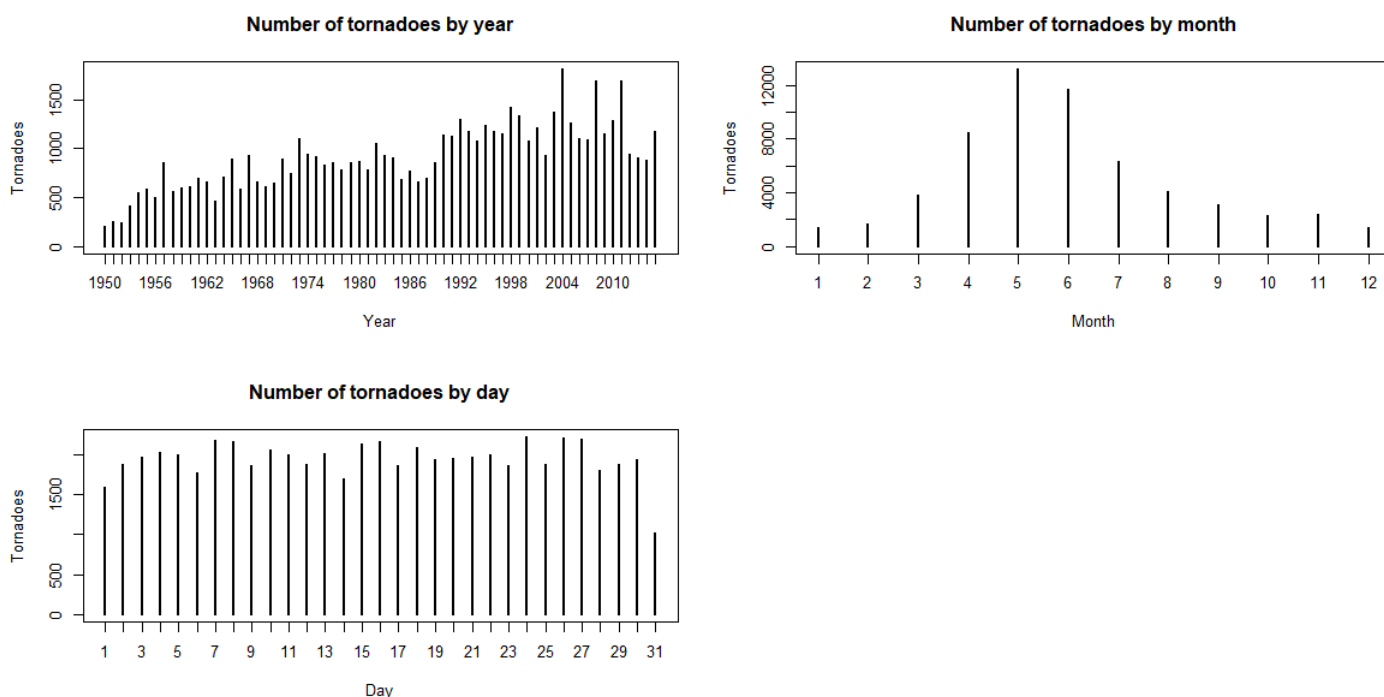


Figura 9

I grafici ottenuti dimostrano quanto affermato nell'analisi dei rispettivi box plot di riferimento. La chiave di lettura di tali grafici è da ricercare nel grafico relativo ai mesi. In questo grafico i tornado tra i mesi si distribuiscono come una normale gaussiana, allo stesso modo del grafico che mostra le temperature medie per i mesi nel mondo.

Riteniamo che questi risultati non siano dovuti a una casualità e dunque, affermiamo che l'aumento della temperatura è positivamente correlato all'aumento del fenomeno tornadico. Infatti, come ulteriore dimostrazione di tale affermazione, dal grafico pubblicato dal NOAA che mostra l'aumento marginale delle temperature nel mondo, le temperature sono aumentate significativamente dal 1985, che trova esatta corrispondenza nell'aumento dei tornado visibile nel grafico relativo agli anni.

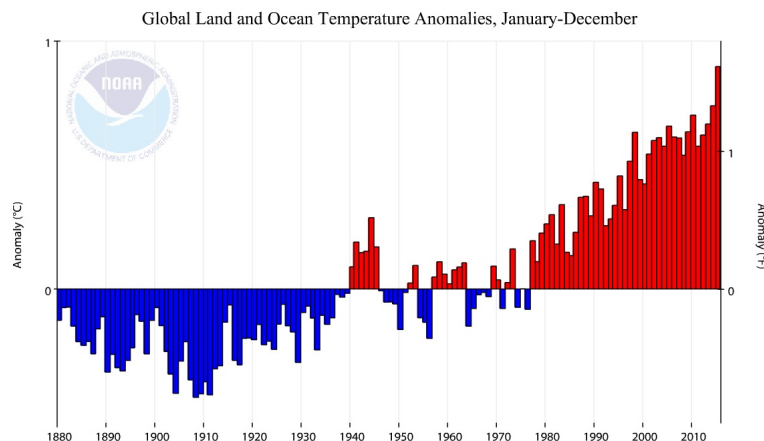


Figura 10: Variazione della temperatura nel mondo¹⁵

Tutti questi grafici ci permettono di avanzare l'ipotesi secondo la quale, all'aumentare del tasso di crescita della temperatura aumentano anche i tornado. In altri termini questi grafici suggeriscono che il fenomeno dell'aumento della temperatura sia positivamente correlato con l'aumento dei tornado.

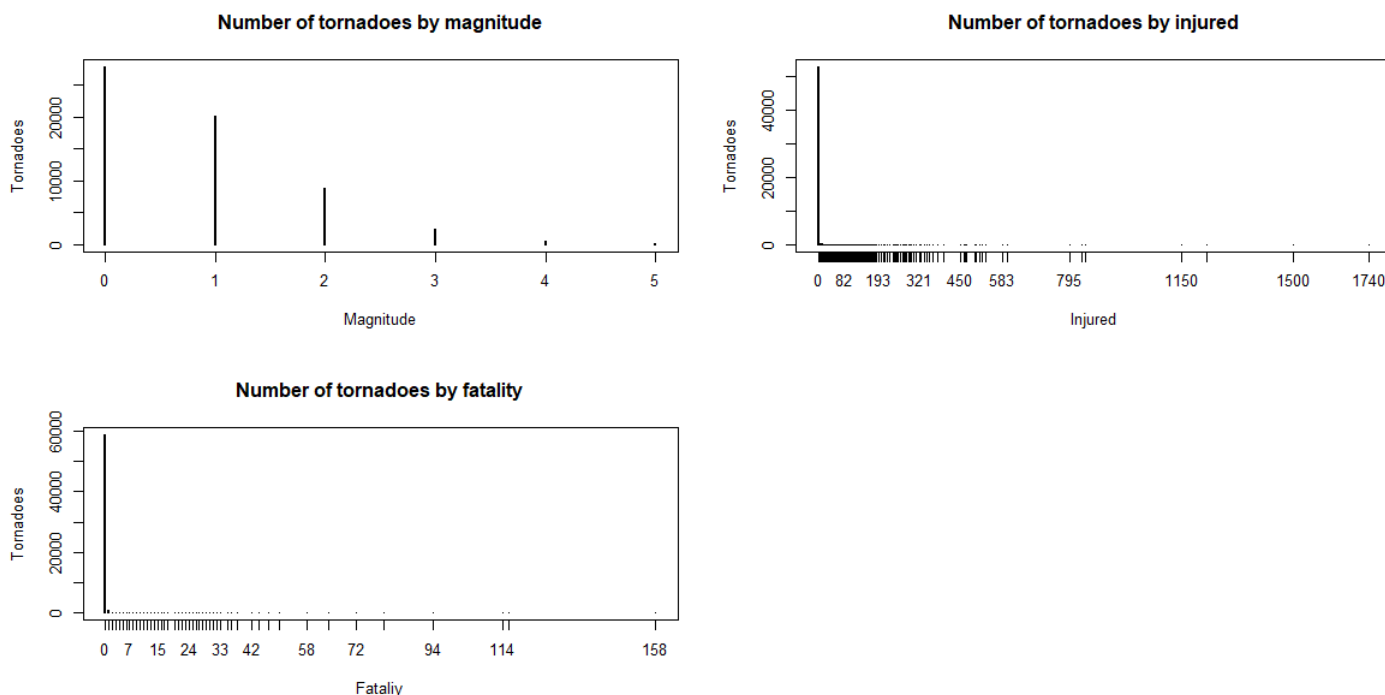


Figura 11

I grafici ottenuti dimostrano quanto affermato nell'analisi dei rispettivi box plot di riferimento.

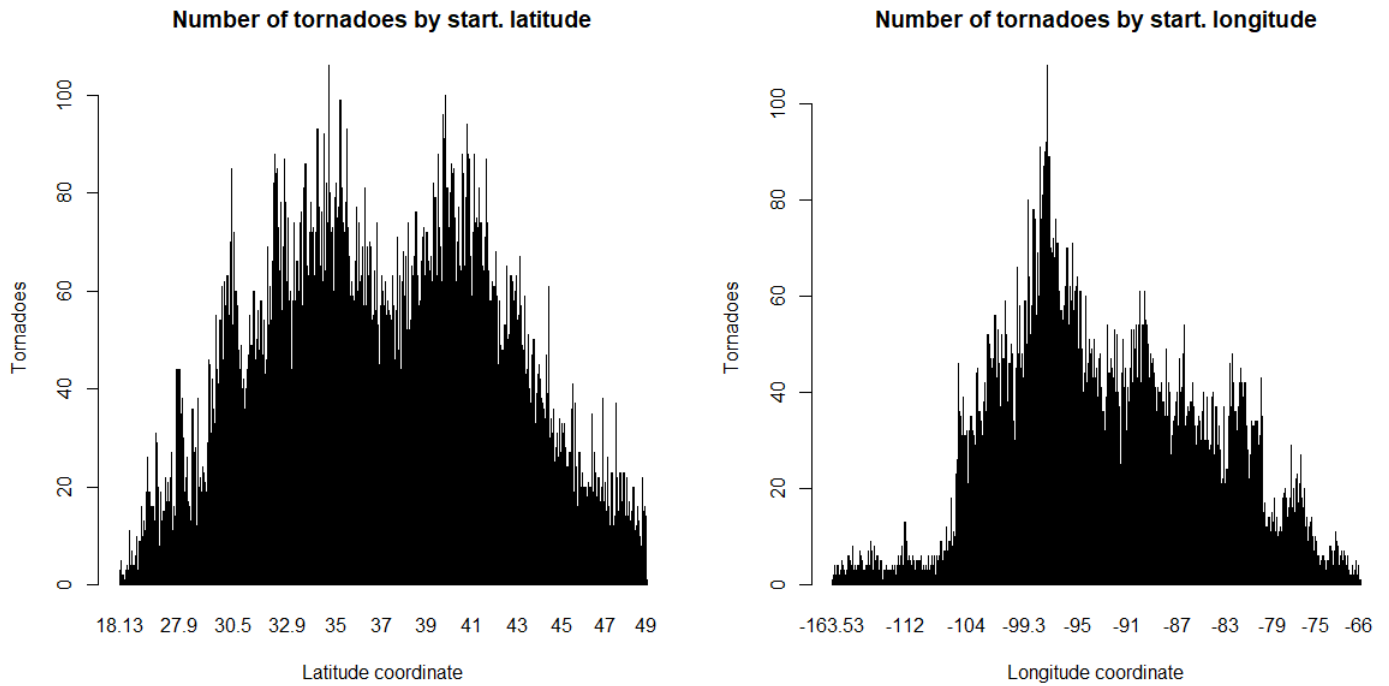


Figura 12

Questi grafici si distribuiscono approssimativamente come una funzione normale gaussiana con una particolare eccezione per il grafico relativo alle coordinate di latitudine in quanto sembrano verificarsi due valori di massimo. Dall'analisi dei rispettivi box plot di riferimento, non ci aspettavamo che i tornado si distribuissero come una normale gaussiana. Questo risultato mette in luce che vi sono dei luoghi in cui le condizioni alla base del fenomeno tornadico sono maggiormente favorevoli.

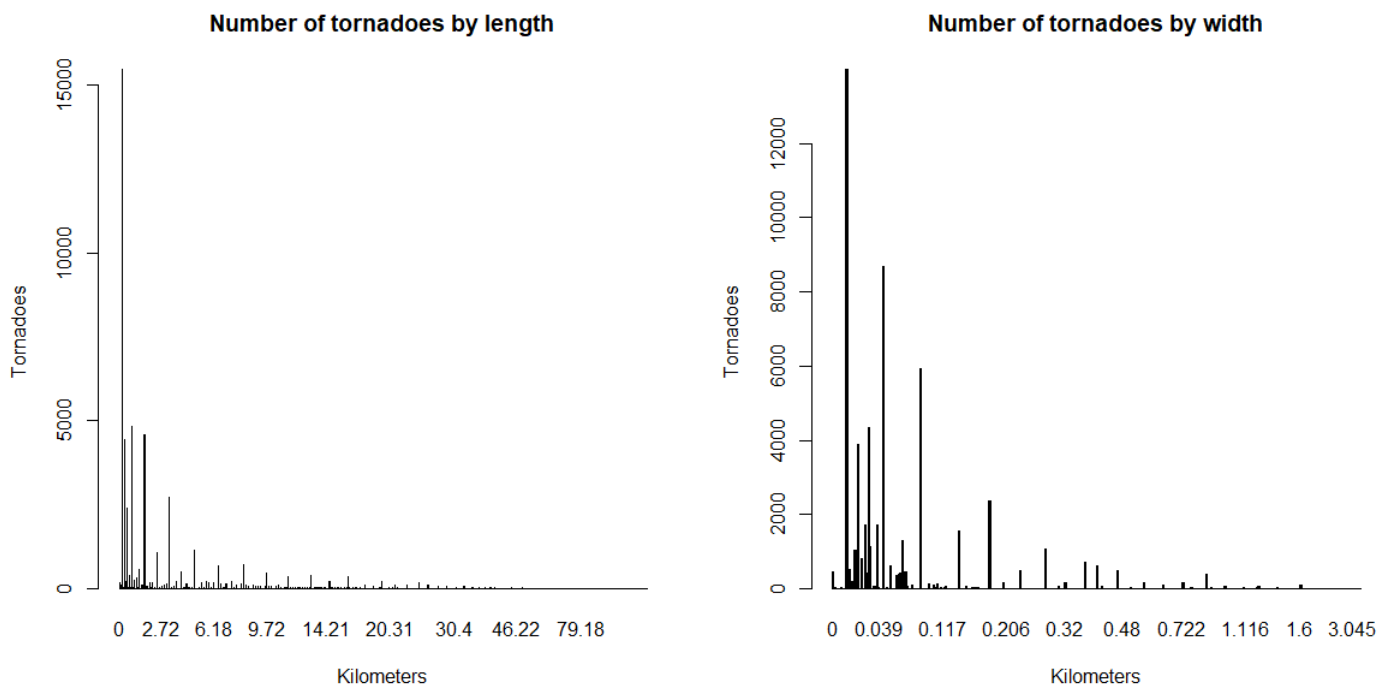


Figura 13

I grafici ottenuti dimostrano quanto affermato nell'analisi dei rispettivi box plot. Da questi grafici possiamo concludere che maggiore è la lunghezza percorsa dal tornado e maggiore è il diametro, minore è la frequenza del fenomeno. Possiamo ipotizzare che, se un tornado ha "dimensioni" maggiori, questo significa che necessita di maggiore calore/umidità. I temporali che si verificano in condizioni di

elevato calore ed umidità sono quelli più distruttivi. Quindi l'ipotesi del collegamento tra dimensioni e potenza è sempre più visibile: all'aumentare delle dimensioni, aumenta la potenza del tornado, il quale è meno frequente.

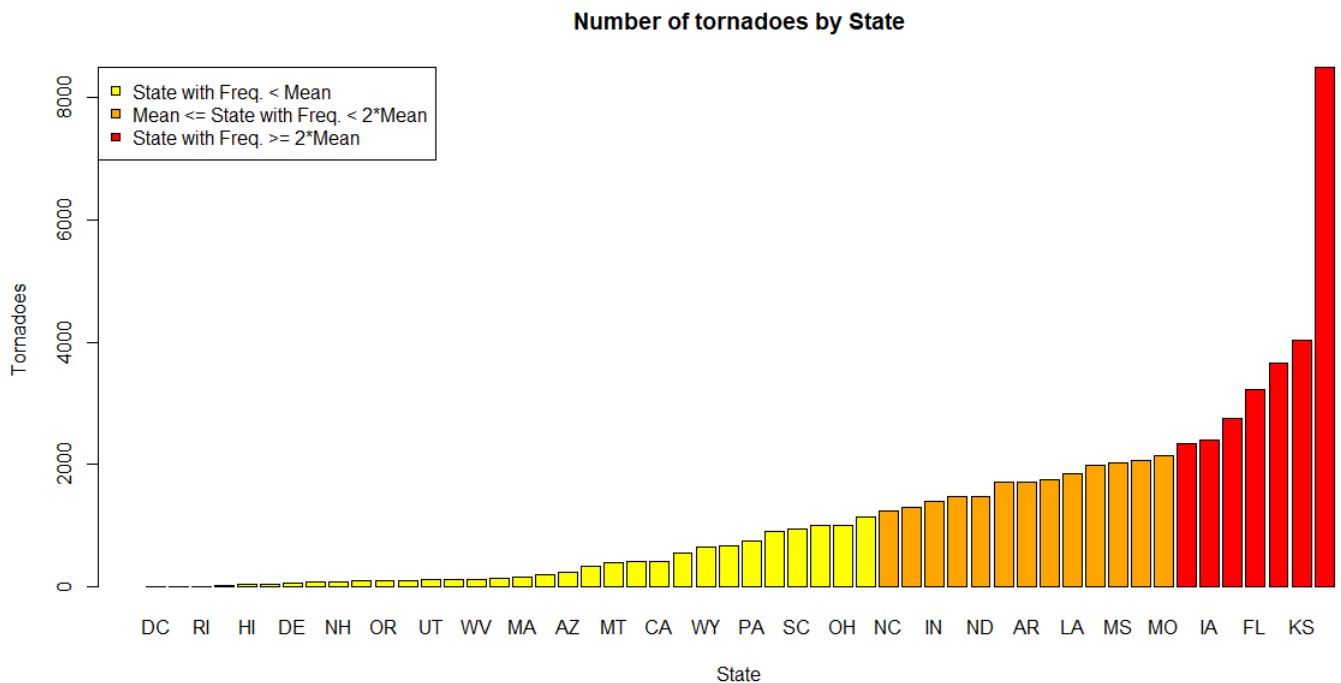


Figura 14

Per realizzare questo grafico, abbiamo innanzitutto ordinato gli Stati per ordine crescente della frequenza del fenomeno. Successivamente abbiamo calcolato la frequenza media del fenomeno per Stato e raggruppato gli Stati in funzione della loro frequenza dei tornado rispetto alla media. Abbiamo in questo modo individuato tre gruppi:

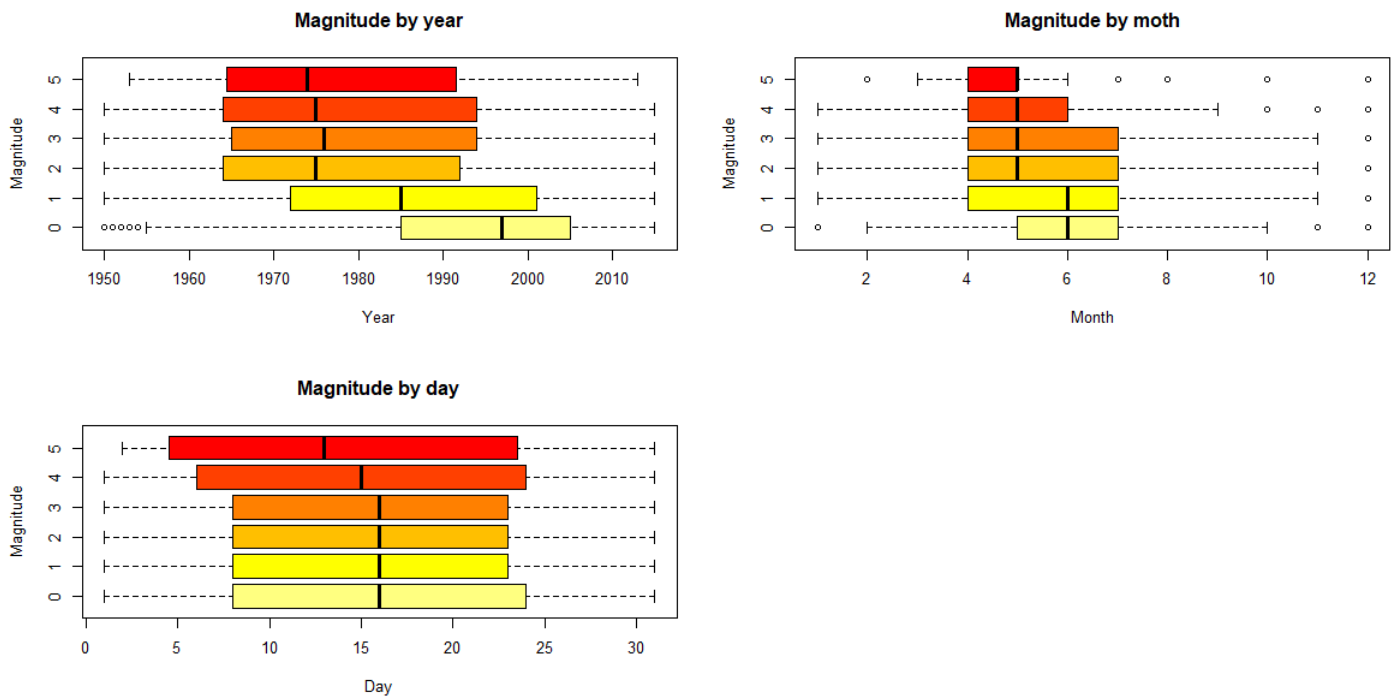
1. il primo gruppo, rappresentato dal colore giallo, comprende tutti gli stati la cui la frequenza di tornado è al di sotto della media. Tali Stati si possono definire come Stati poco/mediamente soggetti al fenomeno tornadico. La probabilità che un fenomeno si verifichi nella regione gialla è approssimativamente del 18,35%.
2. Il secondo gruppo, rappresentato dal colore arancione, comprende gli Stati la cui la frequenza del fenomeno è maggiore della media ma minore del doppio della media. Tali Stati si possono definire come Stati mediamente/molto soggetti al fenomeno tornadico. La probabilità che un fenomeno si verifichi nella regione arancione è approssimativamente del 36,88%.
3. Il terzo gruppo, rappresentato dal colore rosso, comprende gli Stati la cui la frequenza del fenomeno è maggiore di due volte la media. Tali stati si possono definire come Stati molto/estremamente soggetti al fenomeno tornadico. La probabilità che un fenomeno si verifichi nella regione rossa è approssimativamente del 44,77%.

Il nostro raggruppamento degli Stati per attività del fenomeno tornadico, seppur per qualche scostamento come ad esempio lo Stato del Missouri che non rientra nella regione rossa, è in linea con quanto dichiarato dal NOAA per quanto riguarda l'attività dei tornado negli Stati Uniti ed in particolare per quanto riguarda la definizione della "Tornado Alley", che possiamo approssimare con la regione rossa. Questo scostamento è dovuto da un diverso raggruppamento degli stati da parte della suddetta agenzia governativa che avviene per area geografica.

3.2 Analisi bivariata - Studio della magnitudine

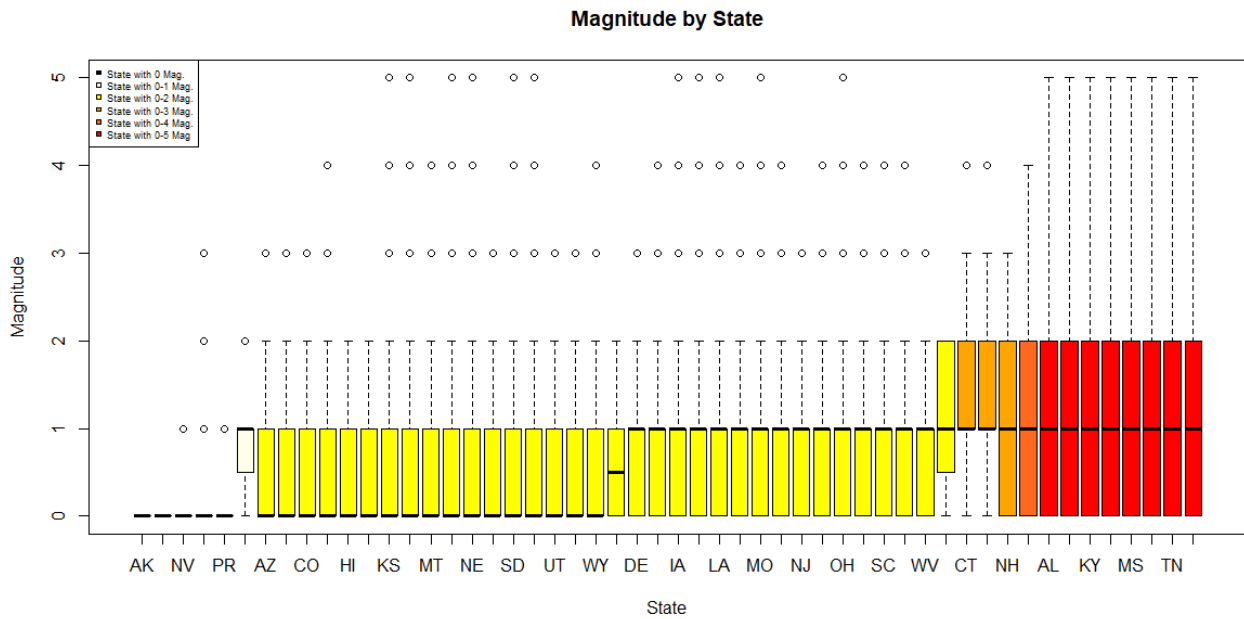
Al fine di verificare l'ipotesi secondo la quale la potenza del fenomeno tornadico potrebbe avere un legame con le sue caratteristiche fisiche (i.e. lunghezza percorsa e diametro) abbiamo deciso di riproporre lo studio univariato differenziandolo per le diverse potenze del fenomeno, cercando di capire se vi sono altre variabili rilevate che influiscono sulla magnitudine del tornado.

Figura 15



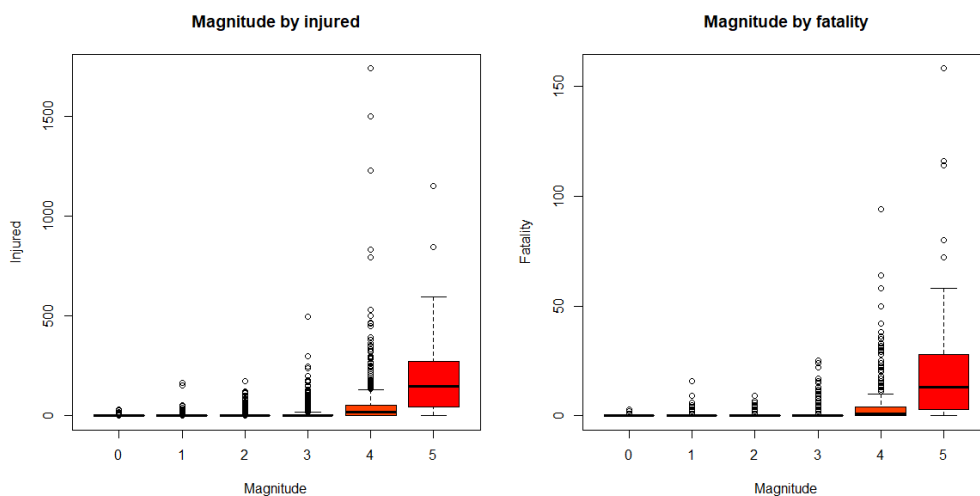
Per quanto riguarda il grafico relativo agli anni, partendo dalla magnitudo uguale a 5 possiamo affermare che il valore mediano tende a spostarsi verso il 2015 al diminuire della potenza, soprattutto in modo particolare per le potenze 0 ed 1. Questo ci permette di affermare che i tornado con potenza 0 ed 1 si concentrano significativamente negli ultimi 30 anni. Per quanto riguarda il grafico relativo ai mesi la distribuzione dei tornado per le potenze tende ad aumentare con il diminuire della potenza per poi aumentare nuovamente in corrispondenza della potenza zero. Per quanto riguarda il grafico relativo ai giorni i tornado con potenza compresa tra 0 e 3 tendono a distribuirsi uniformemente tra i mesi. Per quanto invece riguarda i tornado con potenza compresa tra 4 e 5, essi tendono a verificarsi verso la prima metà del mese, soprattutto in modo particolare per i tornado di potenza 5.

Figura 16



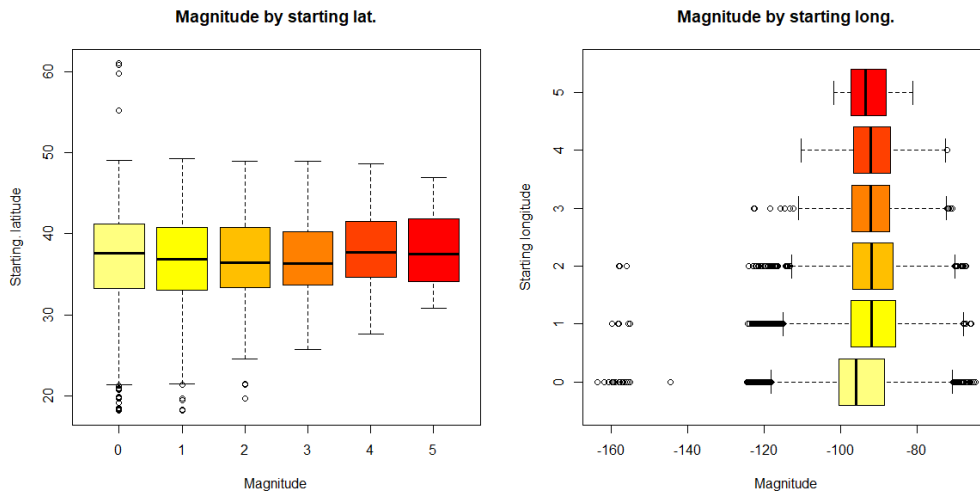
In questo grafico abbiamo deciso di raggruppare gli Stati in ordine crescente della distribuzione delle potenze al fine di individuare Stati con una medesima (o approssimativamente simile) distribuzione delle potenze. In particolare, abbiamo studiato gli Stati che hanno una distribuzione della magnitudine compresa tra 0 e 5. L'obiettivo di tale studio è stato quello di trovare una corrispondenza tra un'elevata frequenza dei tornado e una possibile elevata magnitudine del fenomeno. Confrontando questo grafico con il grafico che mostra la frequenza del fenomeno negli Stati, il risultato è stato il seguente: degli 8 stati che appartengono alla regione 0-5 di distribuzione delle potenze, 1 stato (OK) appartiene alla regione di estrema frequenza dei tornado, 4 stati (AL, IN, WI, MS) appartengono alla regione di medio-alta frequenza dei tornado e 3 stati (MI, KY, TN) appartengono alla regione di medio-bassa frequenza del fenomeno. Possiamo dunque affermare che non c'è un collegamento tra stati con estrema frequenza e stati con possibile estrema magnitudine.

Figura 17



Per quanto riguarda i grafici che mostrano i danni alle persone per potenza dei tornado possiamo affermare che all'aumentare della potenza aumenta la distribuzione dei feriti e dei morti e anche dei valori anomali ad essa associata. Per quanto riguarda i valori anomali, l'ipotesi principale che sosteniamo è relativa al fatto che il numero di feriti e morti dipende dalla densità della popolazione dell'area colpita dal tornado.

Figura 18

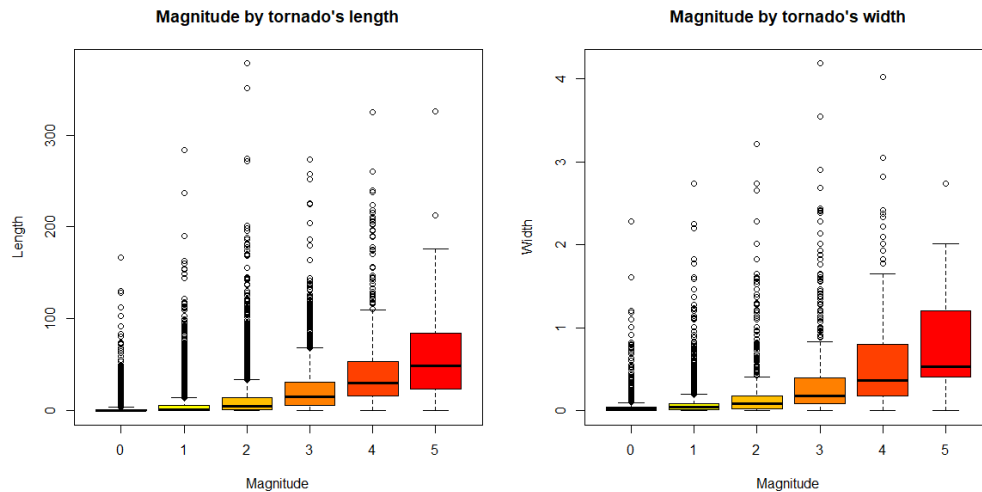


In questi grafici possiamo notare che le diverse potenze osservate dei tornado tendono a distribuirsi attorno al medesimo luogo mediano sia per coordinate di latitudine che di longitudine. Inoltre, è possibile osservare che all'aumentare della potenza del fenomeno, la distribuzione delle coordinate di latitudine e di longitudine si restringono verso il valore mediano. Possiamo quindi affermare che più ci avviciniamo al luogo mediano, dato dalle coordinate mediane di latitudine e longitudine, più è possibile osservare tornado di estrema potenza. Avendo individuato il luogo mediano, che approssimativamente è "State Hwy 97, Stotts City, MO 65756, Stati Uniti", analizzando esclusivamente i tornado di potenza 5 possiamo affermare quanto segue:

1. Per quanto riguarda le coordinate di longitudine, i tornado di potenza 5 tendono a concentrarsi in maniera significativa tra i valori -102 e -93.95 (ossia il valore mediano). Dato che la variabilità del fenomeno tra il valore minimo e il valore mediano è sufficientemente minore rispetto alla variabilità tra la mediana e il valore massimo, questo significa che se vogliamo trovare un intorno più ristretto di valori in cui è più probabile che si concentrino fenomeni di potenza 5, tale area è da ricercare in un intorno di valori a sinistra della mediana. In conclusione, per la longitudine consideriamo tale intervallo di valori.

2. Per quanto riguarda le coordinate di latitudine tale fenomeno tende a concentrarsi maggiormente in un intervallo di valori compresi tra 37.085 (ossia il valore mediano) e 31. Applicando in questo contesto lo stesso ragionamento effettuato per i valori della longitudine, considerando che la variabilità dei valori compresi tra la mediana e il valore minimo è chiaramente inferiore all'intervallo di valori al di sopra della mediana, ricerchiamo l'area a più alta concentrazione di fenomeni EF5 tra i valori al di sotto della mediana. Unendo i risultati ottenuti, considerando l'ampiezza di valori di longitudine e latitudine, l'area verso la quale è più probabile individuare fenomeni di estrema potenza è compresa tra le coordinate di longitudine comprese tra -102 e -93.95 e di latitudine tra 37.085 e 31. Dato che la variabilità del fenomeno tornadico diminuisce all'aumentare della potenza, possiamo concludere che più ci avviciniamo a tale area designata, maggiore è la probabilità di osservare fenomeni a più elevata potenza.

Figura 19



Nonostante l'elevata presenza di valori anomali per ciascuna potenza, possiamo affermare che maggiore è la potenza del fenomeno maggiore è la variabilità nelle sue dimensioni. Tuttavia, per ciascuna potenza la mediana tende a salire di poco sia per la lunghezza del percorso che per la larghezza del fenomeno.

3.3 Analisi delle dipendenze

In questo paragrafo continuiamo lo studio sulla potenza dei tornado, individuando quali sono le variabili connesse ad essa al fine di elaborare un modello predittivo per la magnitudine del fenomeno.

3.3.1 Test Chi-quadro

Variabile	Chi-quadro	Gradi di Libertà	P-Value
yr	8986.3	325	$< 2.2e - 16$
mo	2284.7	55	$< 2.2e - 16$
dy	391.71	150	$< 2.2e - 16$
st	4704.4	255	$< 2.2e - 16$
inj	62268	1025	$< 2.2e - 16$
fat	38004	235	$< 2.2e - 16$
slat	11254	11590	0.987
slon	20076	21165	1
len	69579	10655	$< 2.2e - 16$
wid	32324	1565	$< 2.2e - 16$

Eseguendo il test Chi-quadro sulle coppie di variabili che includono la magnitudine e ciascuna singola variabile precedentemente studiata, assumendo un valore alfa = 0.01:

1. Per le variabili 'yr', 'mo', 'dy', 'st', 'inj', 'fat', 'len' e 'wid' possiamo rifiutare l'ipotesi nulla di indipendenza dalla variabile 'mag' in quanto il valore del p-value è minore di alfa e soprattutto, è prossimo a zero. In altri termini, questo risultato indica che, per ciascuna coppia studiata, la probabilità che l'associazione sia dovuta al caso è approssimabile a zero, che la relazione è statisticamente significativa e dunque, che vi è dipendenza tra la magnitudine e le variabili analizzate. Per le variabili 'slat' e 'slon', non possiamo rifiutare l'ipotesi nulla di indipendenza in quanto il p-value è maggiore del valore di alfa e soprattutto, è prossimo a uno. Questo risultato significa che la relazione non è statisticamente significativa e dunque, è molto probabile che vi sia indipendenza tra magnitudine e rispettivamente

latitudine/longitudine di partenza. Questo risultato ottenuto per le variabili ‘slat’ ed ‘slon’ non ci convince, per questo motivo procederemo a continuare con lo studio della dipendenza di tutte le variabili analizzate dalla variabile magnitudine, calcolando l’indice Eta-quadro e successivamente l’indice di correlazione lineare di Pearson.

3.3.2 Eta-quadro

Variabile	Eta-quadro
yr	0.1173472
mo	0.03354339
dy	0.002654231
st	0.0556267
inj	0.2661365
fat	0.127997
slat	0.0565449
slon	0.1077428
len	0.3607021
wid	0.299345

Dai risultati ottenuti calcolando le medie della magnitudine condizionate da ciascun valore rispettivamente delle variabili fin ora analizzate, possiamo affermare che la magnitudine ha una dipendenza in media più forte per quanto riguarda la lunghezza del percorso e il diametro del tornado. Riteniamo che il valore di tali indici sia basso in quanto, all’aumentare della potenza, i tornado aumentano la variabilità nelle proprie “dimensioni”. Inoltre, il valore dell’indice Eta-quadro relativo alle variabili ‘slat’ e ‘slon’ ci permettono di affermare che nel rispettivo test Chi-quadro, si è verificato un errore del secondo tipo, ossia è stata accettata l’ipotesi nulla (i.e di indipendenza) che in realtà risulta essere falsa, in quanto se così non fosse, il valore di tali indici avrebbe dovuto essere prossimo allo zero. Invece, il valore dell’indice per la variabile ‘dy’ potrebbe indicare che nel test si sia verificato un errore del primo tipo (più grave rispetto all’errore ottenuto per ‘slat’ e ‘slon’) secondo il quale è stata rifiutata l’ipotesi nulla quando in realtà essa è vera.

3.3.3 Correlazione

	yr	mo	dy	mag	inj	fat	slat	slon	len	wid
yr	1.00000000	0.01196932	2.565690e-02	-0.30902726	-0.037733825	-0.03170588	1.636770e-02	0.022514179	-0.075686843	0.07242543
mo	0.01196932	1.00000000	-4.262939e-02	-0.05057423	-0.025623373	-0.02705212	4.396642e-02	0.037097451	-0.056399616	-0.05768498
dy	0.02565690	-0.04262939	1.000000e+00	-0.01087690	-0.009753834	-0.00367207	-6.473802e-05	-0.004350462	-0.006369090	0.01162349
mag	-0.30902726	-0.05057423	-1.087690e-02	1.00000000	0.223764980	0.19780667	-1.728982e-02	0.136227540	0.437393605	0.42487325
inj	-0.03773383	-0.02562337	-9.753834e-03	0.22376498	1.000000000	0.76039446	-1.102196e-02	0.034252156	0.252199938	0.19855164
fat	-0.03170588	-0.02705212	-3.672070e-03	0.19780667	0.760394463	1.00000000	-1.068019e-02	0.023289709	0.230122778	0.18341921
slat	0.01636770	0.04396642	-6.473802e-05	-0.01728982	-0.011021960	-0.01068019	1.000000e+00	-0.170971252	0.003195338	-0.01104266
slon	0.02251418	0.03709745	-4.350462e-03	0.13622754	0.034252156	0.02328971	-1.709713e-01	1.000000000	0.050416263	0.05358609
len	-0.07568684	-0.05639962	-6.369090e-03	0.43739360	0.252199938	0.23012278	3.195338e-03	0.050416263	1.000000000	0.36356835
wid	0.07242543	-0.05768498	1.162349e-02	0.42487325	0.198551639	0.18341921	-1.104266e-02	0.053586087	0.363568349	1.00000000

Figura 20

Dato che l’obiettivo è studiare la magnitudine in funzione delle variabili ad essa connesse, procediamo ad analizzare la colonna relativa alla magnitudine, i cui valori esprimono il legame lineare tra la magnitudine e ciascuna variabile considerata. Dunque, come per l’indice Eta-quadro di riferimento, i risultati più significativi sono dati dalle variabili ‘len’ e ‘wid’. Successivamente, come dimostra il grafico dell’indice Eta-quadro per la variabile ‘yr’, la magnitudine è correlata negativamente con gli anni.

4 STATISTICA INFERENZIALE

I risultati ottenuti ci hanno permesso di capire quali variabili utilizzare in un modello di regressione lineare multipla, che consentisse di prevedere il valore della potenza del tornado, inserendo i valori per determinate variabili considerate. A tal fine, le variabili che abbiamo deciso di utilizzare sono 'yr', 'slon', 'len', 'wid'. Abbiamo deciso di escludere le variabili 'inj' e 'fat' perché tali variabili sono una conseguenza non solo della potenza del tornado ma anche della regione in cui il tornado si verifica (i.e. maggiore è la densità della popolazione, maggiore è la probabilità che un tornado con una modesta potenza possa causare vittime e/o feriti). Inoltre, il nostro obiettivo è descrivere un modello che, partendo da caratteristiche certe, come l'anno di riferimento del fenomeno, e presumibili, come l'area in cui si potrebbe verificare un tornado e le dimensioni connesse a questa area, ci permettesse di prevedere la potenza del fenomeno.

4.1 Campionamento casuale stratificato

A tal fine, abbiamo deciso di non considerare tutta la popolazione dei tornado verificatisi tra il 1950 e il 2015 per due motivi:

1. I fenomeni aumentano la loro variabilità nelle caratteristiche fisiche all'aumentare della potenza.
2. Il modello di regressione lineare multipla che considera tutta la popolazione fornisce un R quadro aggiustato pari a 0.3806 che rappresenta un risultato troppo basso anche solo per poter descrivere e presentare il modello. Questo risultato non è dovuto solo ad una maggiore variabilità del fenomeno per le dimensioni quando la potenza aumenta, ma anche ad una maggiore variabilità del fenomeno per la sua localizzazione quando la potenza diminuisce. Per ridurre tale variabilità abbiamo deciso di considerare un campione stratificato di tornado per la potenza del fenomeno, senza ripetizione dei dati estratti, occupandoci di mantenere le stesse proporzioni delle diverse potenze che si presentano nella popolazione. Abbiamo verificato la rappresentatività del campione fissando un intervallo di confidenza al 95%. Come si può vedere, tale verifica ha avuto esito positivo e quindi abbiamo proceduto a elaborare il suddetto modello.

4.2 Modello statistico: regressione lineare multipla

Nonostante i nostri grandi sforzi per trovare un modello di regressione multipla che potesse spiegare, attraverso la variabilità degli attributi 'yr', 'slon', 'len', 'wid', un'elevata percentuale della variabilità totale della potenza dei tornado, siamo riusciti ad ottenere un R-quadro aggiustato pari a 0.6212. Dunque, il modello, data la variabilità delle suddette variabili, riesce a spiegare il 62,12% della variabilità totale della magnitudine. Rispetto al valore iniziale, siamo soddisfatti del modello presentato attraverso il campionamento stratificato.

```
call:
lm(formula = mag ~ yr + len + wid + slon, data = stratificato)

Residuals:
    Min       1Q   Median       3Q      Max
-1.4815 -0.4007 -0.1301  0.3913  2.6269

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  41.981801    3.231061   12.993 <2e-16 ***
yr           -0.020495    0.001626  -12.608 <2e-16 ***
len           0.022371    0.001942   11.520 <2e-16 ***
wid           2.373300    0.152931   15.519 <2e-16 ***
slon          0.008116    0.003456    2.348  0.0192 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6569 on 506 degrees of freedom
Multiple R-squared:  0.6242,    Adjusted R-squared:  0.6212
F-statistic: 210.1 on 4 and 506 DF,  p-value: < 2.2e-16
```

4.3 Previsione utilizzando il modello

yr	slon	len	wid
1968	128.264	0.805	-98.37
2021	128.264	0.805	-98.37

Il primo dato inserito nella tabella è stato osservato realmente, il tornado di riferimento è stato osservato in Nebraska in quelle coordinate di longitudine che abbiamo ipotizzato essere la sorgente dei tornado EF5. Di fatti, nel dataset la potenza del fenomeno in questione è 5. Abbiamo inserito tali dati nel modello da noi elaborato per testarlo. Il risultato ottenuto dal modello è pari a 5.629128. Nonostante tale valore si discosti dal valore 5 possiamo comunque affermare che il nostro modello ha predetto l'estrema intensità possibile del tornado.

Il secondo dato all'interno della tabella rappresenta il medesimo tornado analizzato precedentemente tuttavia ipotizzando che si manifesterà con le stesse caratteristiche nel 2021. Il modello per questo fenomeno genera una potenza pari a 4.542894.

Questi 2 risultati ci permettono di trarre le seguenti conclusioni.

5 Conclusioni

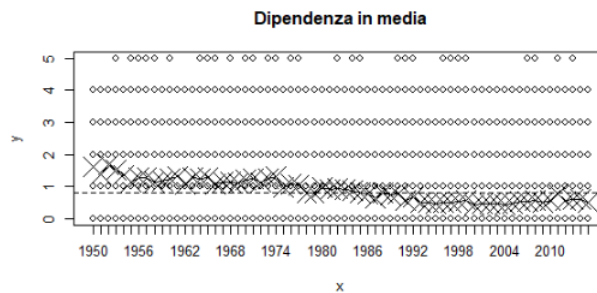


Figura 21

Questo grafico è il risultato della rappresentazione dell'indice Eta-quadro per la magnitudine, condizionata dai valori della variabile relativa all'anno. Insieme ai risultati ottenuti nella previsione, considerando il risultato ottenuto dal modello elaborato per l'indice R-quadro aggiustato, possiamo trarre le seguenti conclusioni:

1. Più ci avviciniamo al 2015 più la potenza media dei tornado diminuisce. Quanto affermato viene confermato dalla previsione del modello relativo al dato del 2021. Tuttavia, il valore previsionale pari 4.542894 ci dice che oggi i tornado EF4 hanno assunto le "dimensioni" (len-wid) dei tornado EF5 di ieri. Dunque, nonostante la potenza dei tornado sia diminuita possiamo avanzare l'ipotesi che il fenomeno tornadico in ciascuna delle possibili potenze abbia aumentato la variabilità nelle sue dimensioni. Questo significa che oggi i tornado risultano più imprevedibili.
2. Il modello spiega solamente il 62% della variabilità totale della magnitudine attraverso la variabilità delle variabili esplicative a causa del fatto che i tornado, all'aumentare della potenza, aumentano la propria variabilità nelle dimensioni. Questa variabilità è spiegata anche dall'enorme mole di valori anomali individuati per ciascuna potenza nei box plot relativi alle dimensioni del fenomeno. In definitiva, benché i tornado di bassa potenza (0,1,2) siano più prevedibili (i.e hanno una minore variabilità per le suddette caratteristiche), i tornado di potenza elevata (3,4,5) risultano essere molto più imprevedibili. Questo spiega perché il modello non raggiunge un R-quadro aggiustato prossimo all' 80/90%

Riferimenti bibliografici

<https://www.spc.noaa.gov/faq/tornado/ts-ital.pdf>

<https://www.spc.noaa.gov>

<https://climateandgeohazards.wordpress.com/tag/tornado/>

<http://www.centrometeo.com/articoli-reportage-approfondimenti/fisica-atmosferica/4301-fenomeno-distruttivo-tornado>

<http://www.fenomenitemporaleschi.it/tornado.htm>

<https://it.wikipedia.org>

<https://berkeleyearth.org/july-2020-temperature-update/>

<https://www.climate.gov/maps-data/dataset/global-temperature-anomalies-graphing-tool>