

Main Effect between Risk and Utterance Number

Chunyen

2024/3/13

Introduction

觀察自殺風險等級，與求助者/接線員逐字稿語句長度的關係

使用直方圖、變異數分析、迴歸分析，發現語句長度越長，自殺風險等級可能越高。且求助者與接線員的語句長度差，也可

Data

準備後續分析的資料結構如下

```
## spc_tbl_ [754 x 7] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ dialogue_id      : num [1:754] 2.01e+13 2.01e+13 2.01e+13 2.01e+13 2.01e+13 ...
## $ risk_level       : num [1:754] 2 2 2 1 5 1 1 4 2 2 ...
## $ risk3            : num [1:754] 0 0 0 0 2 0 0 2 0 0 ...
## $ risk4            : num [1:754] 1 1 1 0 3 0 0 3 1 1 ...
## $ risk5            : num [1:754] 1 1 1 0 4 0 0 3 1 1 ...
## $ utNum_caller     : int [1:754] 87 77 103 106 120 98 63 94 87 52 ...
## $ utNum_responser  : int [1:754] 93 77 106 111 125 109 63 113 89 52 ...
## - attr(*, "spec")=
## .. cols(
## ..   dialogue_id = col_double(),
## ..   risk_level = col_double(),
## ..   risk3 = col_double(),
## ..   risk4 = col_double(),
## ..   risk5 = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

求助者x語句長度

依照語句長度分成短、中、長三組

cut_off1	cut_off2
149	316

取 150、300 作為分組的切節分數
各組數量

utNum_caller_group	n
long	270
medium	227
short	257

接線員x語句長度

依照語句長度分成短、中、長三組

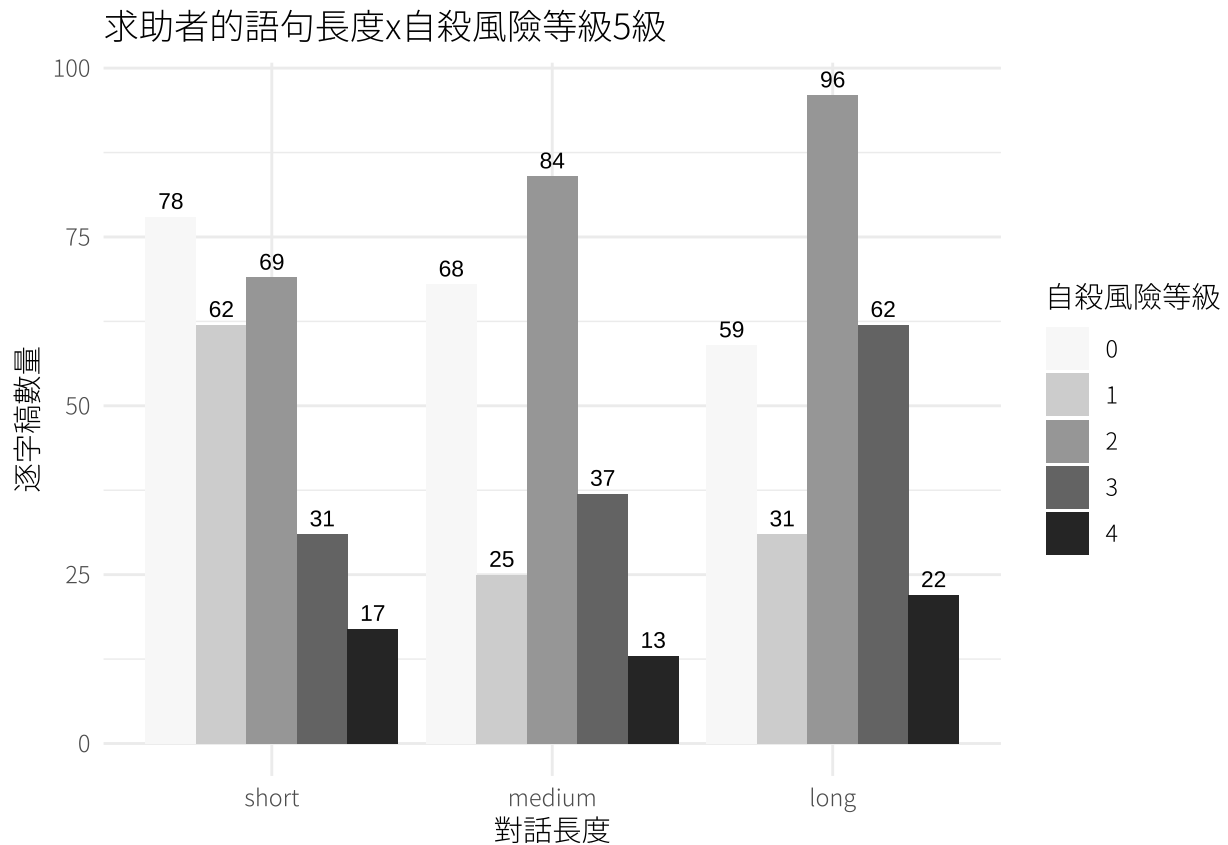
cut_off1	cut_off2
163	349

取 175、350 作為分組的切節分數
各組數量

utNum_responser_group	n
long	249
medium	226
short	279

求助者的語句長度x自殺風險等級5級

圖



統計檢定

ANOVA: utNum_caller ~ risk5

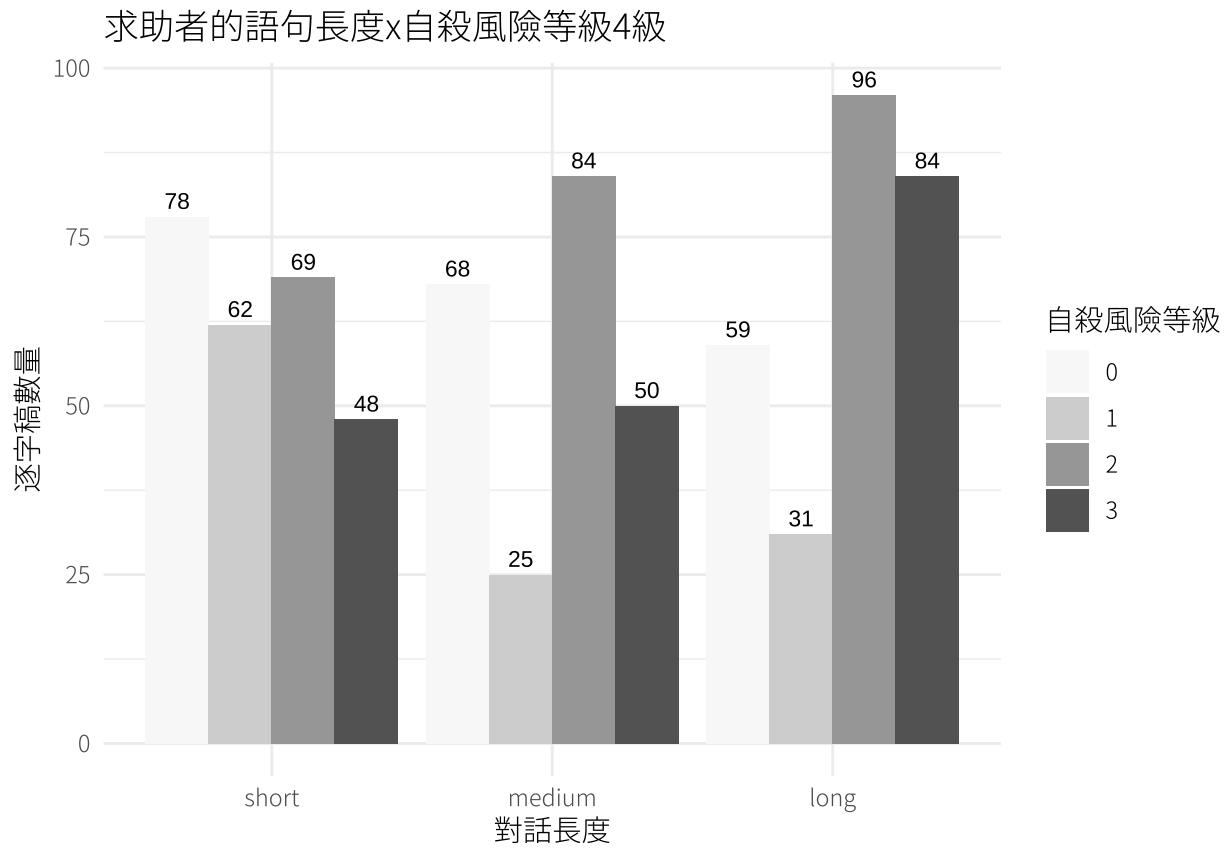
term	df	sumsq	meansq	statistic	p.value
risk5	4	1330848	332712.08	8.227308	1.7e-06
Residuals	749	30289536	40439.97	NA	NA

Regression: risk5 ~ utNum_caller

term	df	sumsq	meansq	statistic	p.value
utNum_caller	1	27.86813	27.868130	18.48868	1.93e-05
Residuals	752	1133.49527	1.507308	NA	NA

求助者的語句長度x自殺風險等級4級

圖



統計檢定

ANOVA: utNum_caller ~ risk4

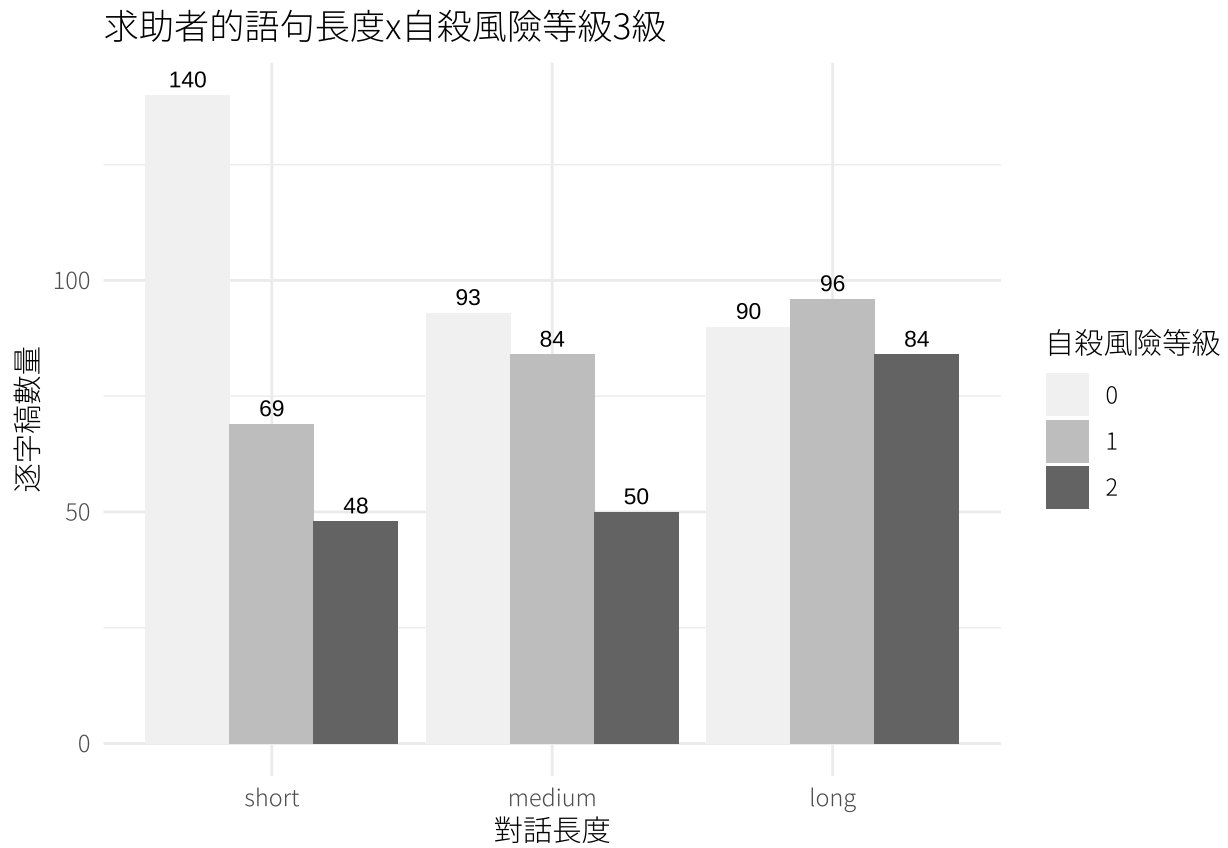
term	df	sumsq	meansq	statistic	p.value
risk4	3	1276408	425469.19	10.51615	9e-07
Residuals	750	30343976	40458.64	NA	NA

Regression: risk4 ~ utNum_caller

term	df	sumsq	meansq	statistic	p.value
utNum_caller	1	25.58383	25.583832	20.5624	6.7e-06
Residuals	752	935.64163	1.244204	NA	NA

求助者的語句長度x自殺風險等級3級

圖



統計檢定

ANOVA: utNum_caller ~ risk3

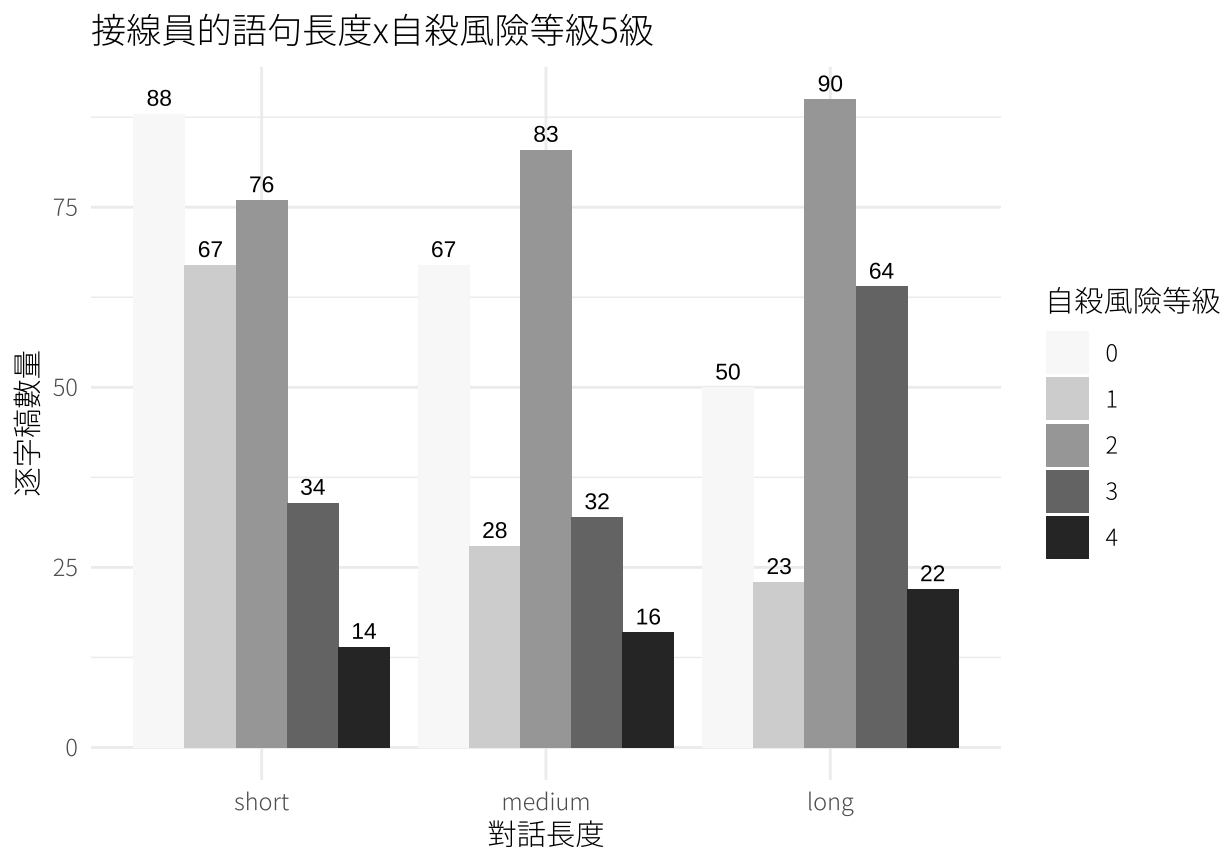
term	df	sumsq	meansq	statistic	p.value
risk3	2	1134463	567231.61	13.97337	1.1e-06
Residuals	751	30485921	40593.77	NA	NA

Regression: risk3 ~ utNum_caller

term	df	sumsq	meansq	statistic	p.value
utNum_caller	1	16.75144	16.7514446	27.27343	2e-07
Residuals	752	461.88118	0.6142037	NA	NA

接線員的語句長度x自殺風險等級5級

圖



統計檢定

ANOVA: utNum_responser ~ risk5

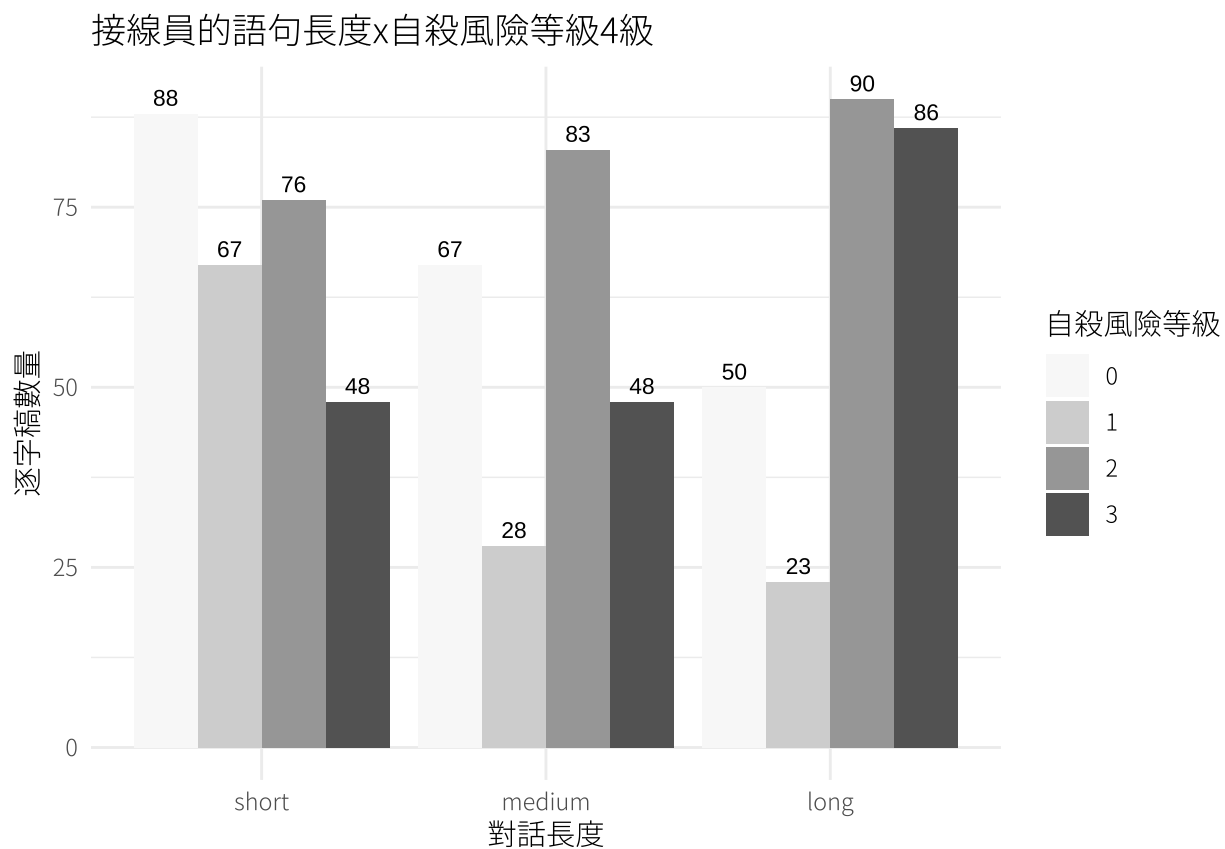
term	df	sumsq	meansq	statistic	p.value
risk5	4	2123725	530931.36	11.73271	0
Residuals	749	33893937	45252.25	NA	NA

Regression: risk5 ~ utNum_responser

term	df	sumsq	meansq	statistic	p.value
utNum_responser	1	49.20893	49.208934	33.27336	0
Residuals	752	1112.15446	1.478929	NA	NA

接線員的語句長度x自殺風險等級4級

圖



統計檢定

ANOVA: utNum_responser ~ risk4

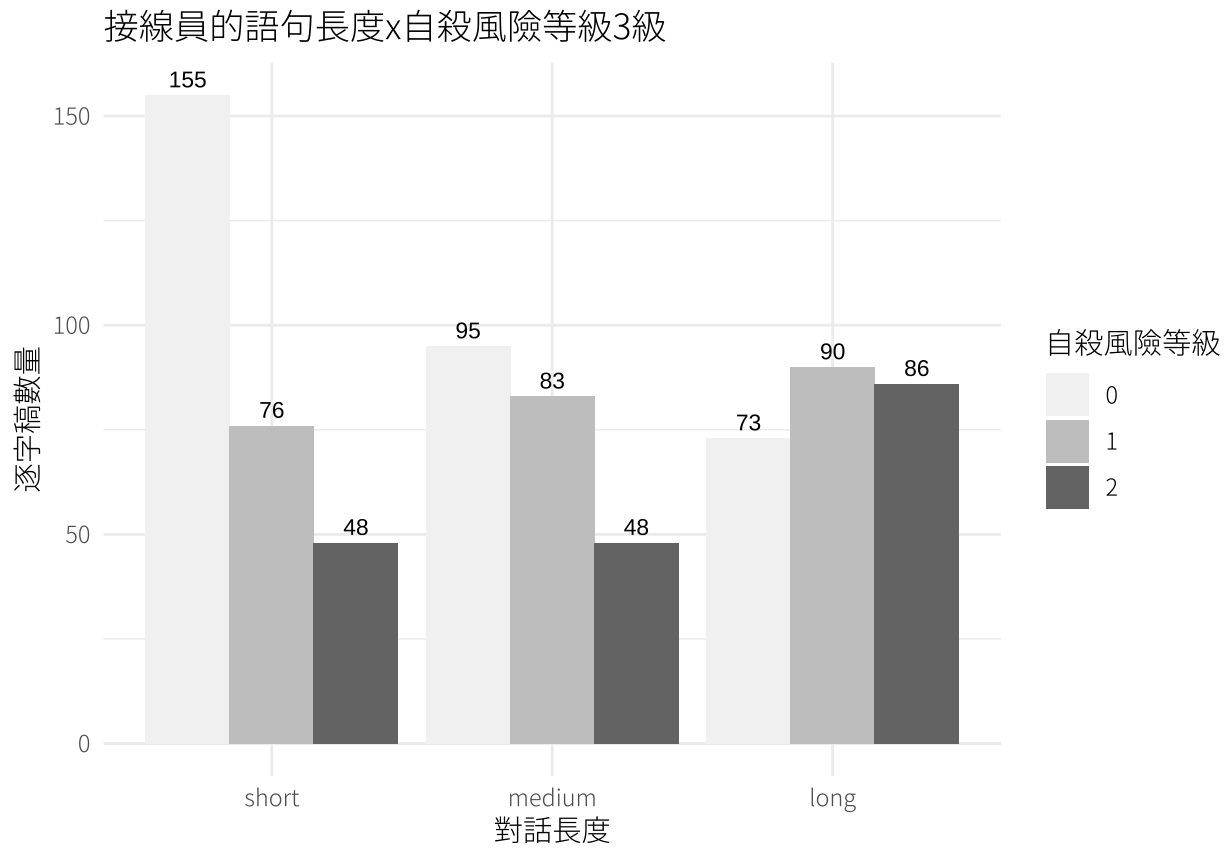
term	df	sumsq	meansq	statistic	p.value
risk4	3	2112776	704258.79	15.5787	0
Residuals	750	33904886	45206.52	NA	NA

Regression: risk4 ~ utNum_responser

term	df	sumsq	meansq	statistic	p.value
utNum_responser	1	42.28302	42.283019	34.60155	0
Residuals	752	918.94245	1.221998	NA	NA

接線員的語句長度x自殺風險等級3級

圖



統計檢定

ANOVA: utNum_responser ~ risk3

term	df	sumsq	meansq	statistic	p.value
risk3	2	1993902	996951.07	22.00551	0
Residuals	751	34023761	45304.61	NA	NA

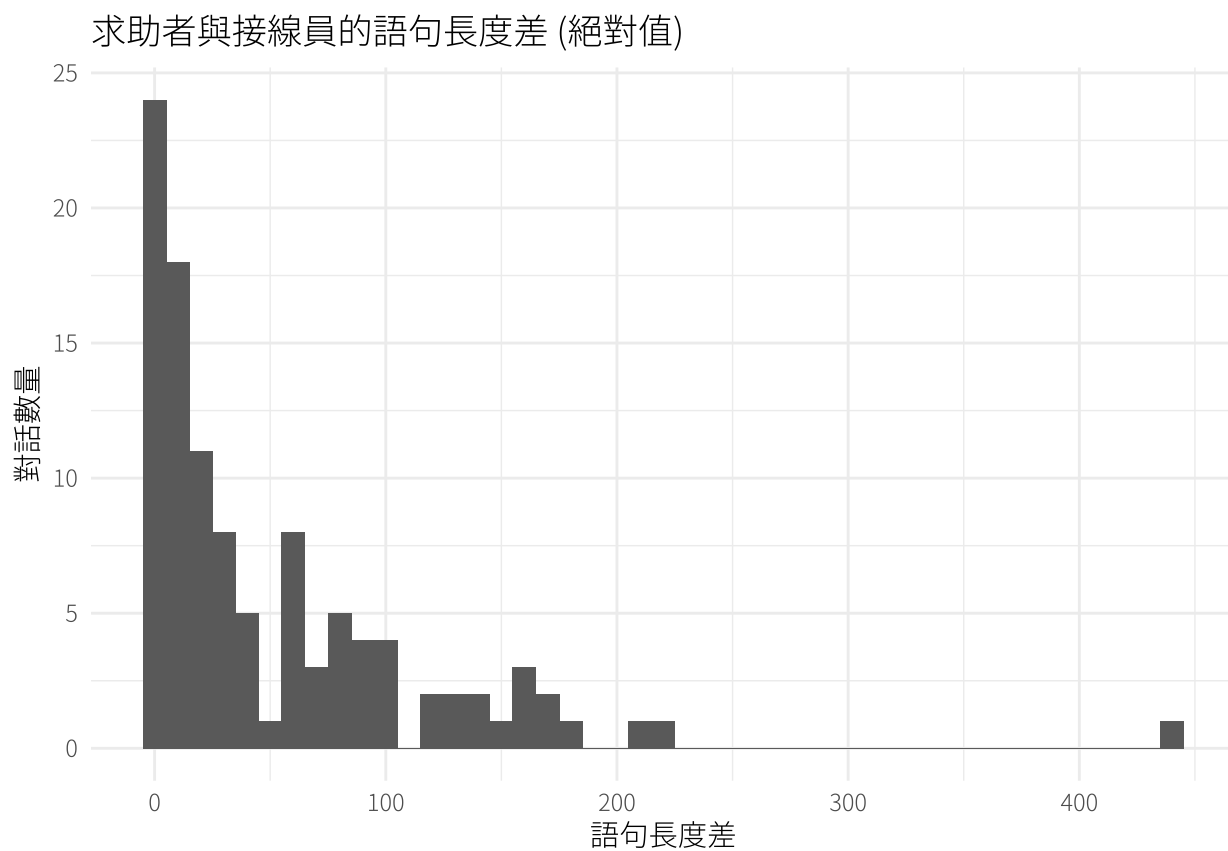
Regression: risk3 ~ utNum_responser

term	df	sumsq	meansq	statistic	p.value
utNum_responser	1	26.24177	26.2417738	43.62116	0
Residuals	752	452.39085	0.6015836	NA	NA

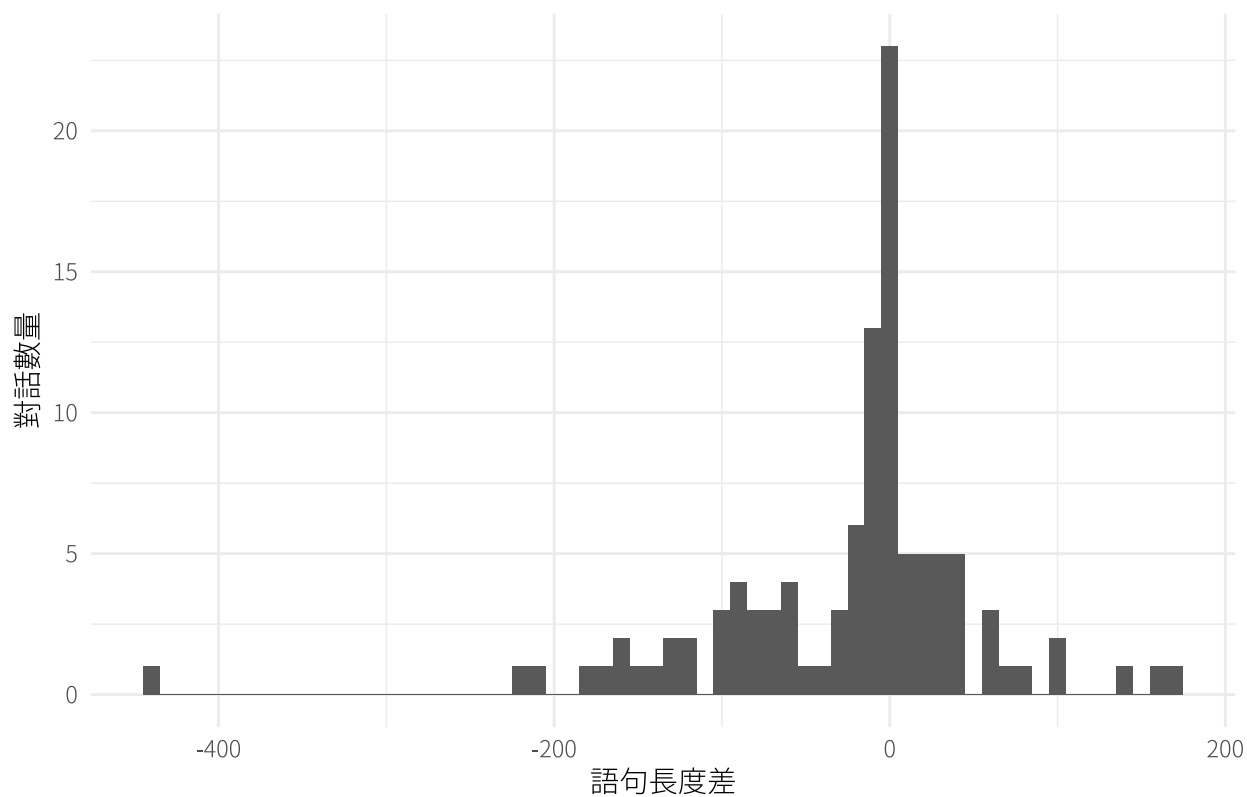
求助者的語句長度x接線員的語句長度

圖

以下繪製的為求助者與接線員被分在不同語句長度組別中，兩者的語句長度差的分佈，共有 107 篇。



求助者與接線員的語句長度差 (求助者-接線員)



語句差x自殺風險等級

統計檢定

ANOVA: utNum_diff ~ risk5

term	df	sumsq	meansq	statistic	p.value
risk5	4	113354.9	28338.73	10.8783	0
Residuals	749	1951197.2	2605.07	NA	NA

ANOVA: utNum_diff ~ risk4

term	df	sumsq	meansq	statistic	p.value
risk4	3	108784.9	36261.63	13.90566	0
Residuals	750	1955767.2	2607.69	NA	NA

ANOVA: utNum_diff ~ risk3

term	df	sumsq	meansq	statistic	p.value
risk3	2	107958.7	53979.346	20.71891	0
Residuals	751	1956593.4	2605.317	NA	NA

Regression: risk5 ~ utNum_diff

term	df	sumsq	meansq	statistic	p.value
utNum_diff	1	55.33059	55.330591	37.61968	0
Residuals	752	1106.03280	1.470788	NA	NA

Regression: risk4 ~ utNum_diff

term	df	sumsq	meansq	statistic	p.value
utNum_diff	1	41.92623	41.926230	34.29626	0
Residuals	752	919.29923	1.222472	NA	NA

Regression: risk3 ~ utNum_diff

term	df	sumsq	meansq	statistic	p.value
utNum_diff	1	24.86511	24.8651142	41.20737	0
Residuals	752	453.76751	0.6034142	NA	NA

印出長度差最多的前十篇逐字稿

dialogue_id	utNum_diff
20151204205509	438
20140424034404	220
20150930214759	211
20160426081414	179
20150711153732	175
20150112215927	174
20131231071118	164
20180907172744	162
20140908150411	159
20140606092155	148

去掉長度差最多的逐字稿，再分析一次

ANOVA: utNum_diff ~ risk5

term	df	sumsq	meansq	statistic	p.value
risk5	4	99919.43	24979.858	10.23273	0
Residuals	728	1777173.39	2441.172	NA	NA