

멜론 데이터 분석

F조



주제

우리나라 대표 음악 플랫폼인
'멜론'의 데이터 분석을 통해
국내 음악시장에 관한 전략을 제시

데이터 수집 절차

멜론에서 제공하는
곡 별 좋아요 수 파일을 기본적으로 이용하고

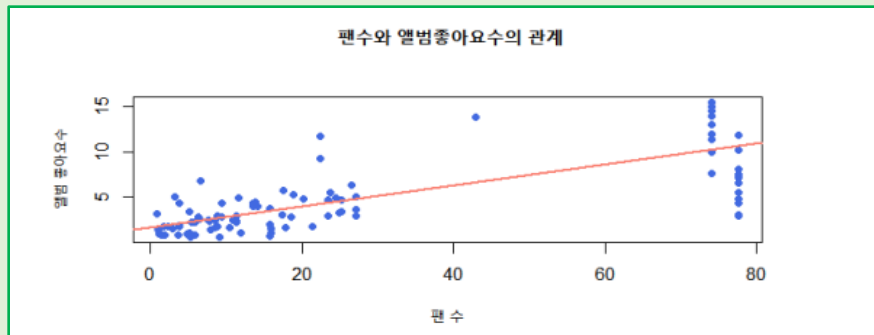
앨범 좋아요 수, 팬 수, 데뷔일, 최고순위, 1위 횟수, 뮤직
비디오 재생 수 변수를 추가적으로 등록

데이터 변수

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	순위	가수	제목	연도	성별	좋아요수	장르	구분	앨범	좋아요 수	팬 수	최고순위	1위 횟수	재생수
2	1	방탄소년단	봄날	2017	M	55.2353	랩	40만이상		15	74	2	0	876.5
3	2	아이유	밤편지	2017	F	45.0047	발라드	40만이상		5.5	77.6	1	13	365.5
4	3	방탄소년단	Dynamite	2020	M	44.1215	댄스	40만이상		10	74	1	75	497.7
5	4	방탄소년단	작은 것들을 위한	2019	M	43.8374	댄스	40만이상		13.9	74	1	8	596.6
			시 (Boy With Luv)(Feat. Halsey)											
6	5	폴킴	모든 날, 모든 순	2018	M	40.5269	발라드	40만이상		1.6	15.9	3	0	484.9
			간 (Every Day, Every Moment)											
7	6	방탄소년단	DNA	2017	M	39.6491	랩	30만이상		15.5	74	5	0	540
8	7	AKMU	어떻게 이별까지	2019	O	39.1856	발라드	30만이상		5.5	23.7	1	30	318.1
			사랑하겠어, 널 사 랑하는 거지											
9	8	Anne-Marie	2002	2018	F	38.6853	팝	30만이상		2.2	5.9	2	0	344.8
10	9	아이유	에잇(Prod.&Feat. SUGA of BTS)	2020	F	38.4012	록	30만이상		7.2	77.6	1	33	260.6
11	10	버스커 버스커	벚꽃 엔딩	2012	M	36.5374	기타	30만이상		6.8	6.6	3	0	302.9

좋아요 수 기준 1위 ~ 100위 까지

분석 - 좋아요 수



```
- likealbumfan<-lm(앨범좋아요수~팬수,data=likes)
summary(likealbumfan)
# p-value = <2e-16, 유의도, r-squared : 0.605
```

...

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	26.27296	0.86218	30.473	<2e-16 ***
앨범좋아요수	0.37745	0.20567	1.835	0.0695 .
팬수	0.01588	0.03056	0.520	0.6045

...

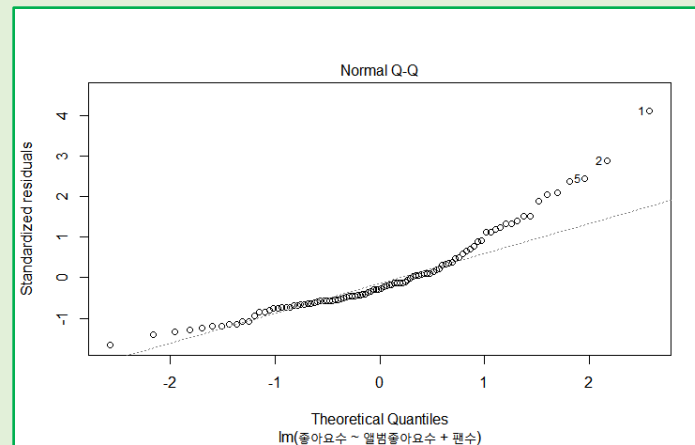
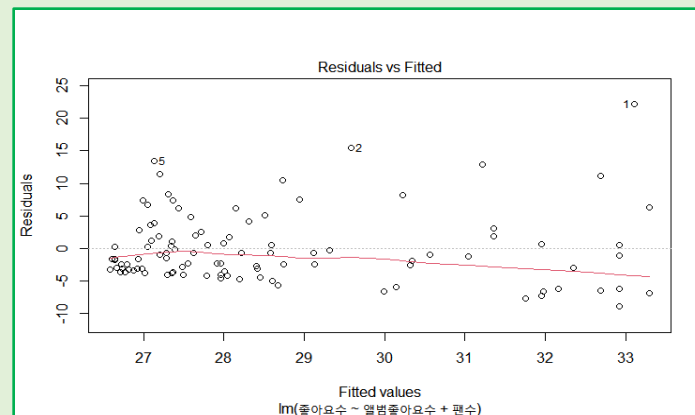
Multiple R-squared: 0.1191, Adjusted R-squared: 0.1009

F-statistic: 6.555 on 2 and 97 DF, p-value: 0.002137

like2.lm<- lm(좋아요수~앨범좋아요수+팬수,data=likes)

like2.lm

summary(like2.lm)



분석 - 좋아요 수

#tapply함수로 살펴본 장르별 평균 팬수, 평균 좋아요수

```
> levels(likes$장르)
[1] "랩"      "팝"      "발라드" "댄스"    "록"      "R&B"     "기타"
> tapply(likes$팬수, likes$장르, mean)
      랩      팝      발라드      댄스      록      R&B      기타
52.821429  7.608333 21.407407 41.893333 38.528571 29.186667 21.890000
> tapply(likes$좋아요수, likes$장르, mean)
      랩      팝      발라드      댄스      록      R&B      기타
31.29716 27.42409 28.01132 28.66009 30.65267 26.29296 30.41022
```

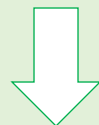
- 장르 별 곡 좋아요 수 : 유의미한 차이 x
- 장르 별 팬 수 : 랩 > 댄스 > ...
- 더미변수 회귀분석 결과, 장르별 팬 수 유의미한 차이 O
- 발라드, 알앤비, 타장르와 댄스, 랩장르가 큰 차이를 보임

#더미변수회귀분석

```
genlm=lm(팬수~장르, data=likes)
regen<-relevel(likes$장르, ref=6) #발라드 기준
regenlm<-lm(팬수~regen, data=likes)
summary(regenlm)
```

```
Coefficients:
(Intercept) 21.4074 5.1045 4.194 6.26e-05 ***
regenR&B      7.7793  8.5415  0.911 0.364777
regen기타     0.4826  9.8188  0.049 0.960905
regen댄스    20.4859  8.5415  2.398 0.018463 *
regen랩     31.4140  8.7354  3.596 0.000519 ***
regen록     17.1212 11.2498  1.522 0.131426
regen팝    -13.7991  9.2023 -1.500 0.137126
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.52 on 93 degrees of freedom
Multiple R-squared:  0.2212,    Adjusted R-squared:  0.171
F-statistic: 4.403 on 6 and 93 DF,  p-value: 0.0005829
```



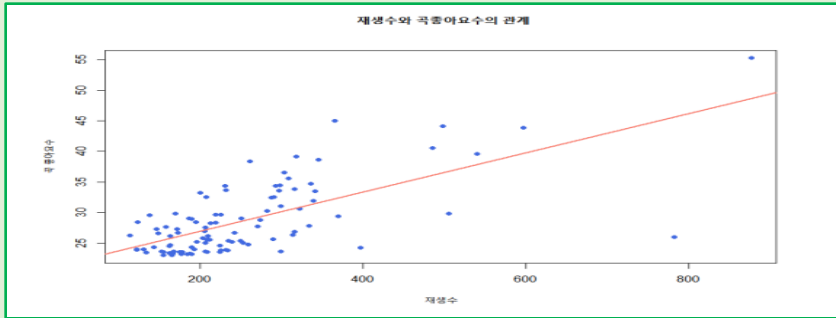
두터운 팬 층을 보유하지 못한 가수
댄스, 랩 장르를 제외한 발라드, R&B 와 같은 장르 타겟

03 분석

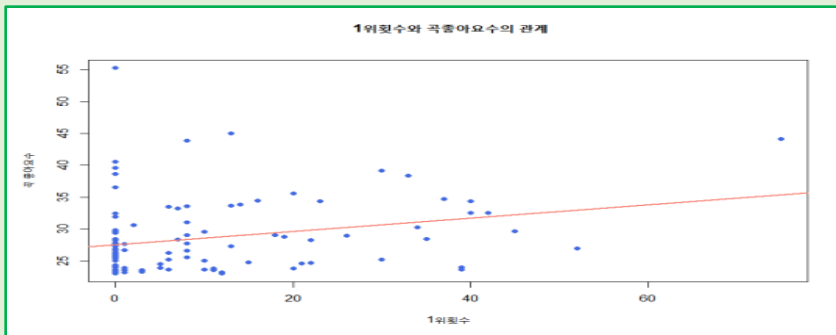
멜론 데이터 분석

분석 - 좋아요 수

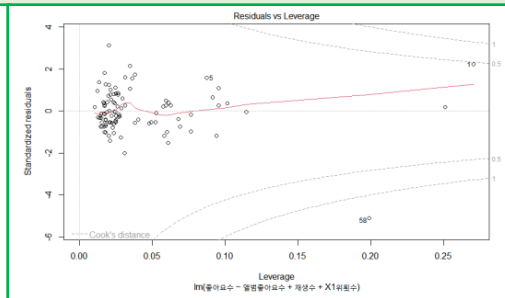
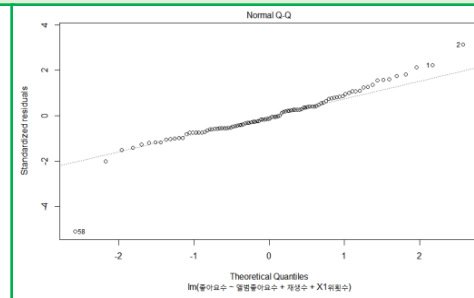
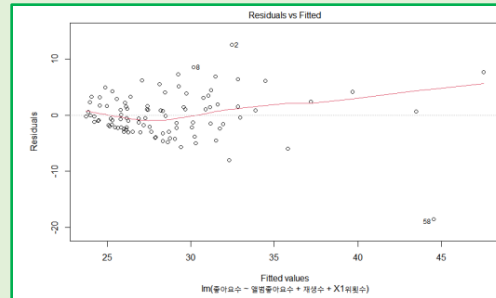
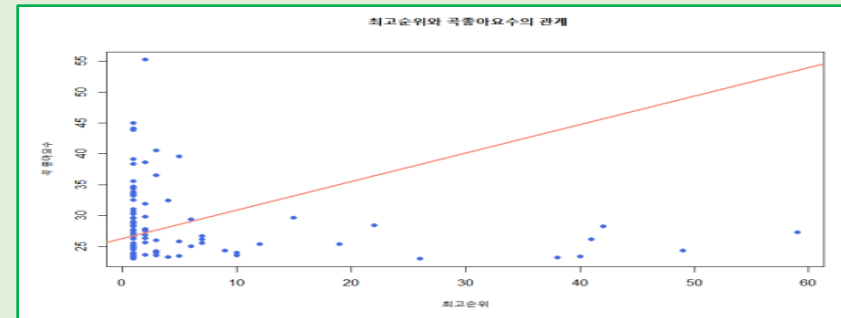
- `likestream<-lm(좋아요수~재생수,data=likes)`
`summary(likestream)`
p-value = 8.86e-15 , 유의o



- `like1st<-lm(좋아요수~X1위횟수,data=likes)`
`summary(like1st)`
p-value = 0.00876, 유의o



- `liketop<-lm(좋아요수~최고순위,data=likes)`
`summary(liketop)`
p-value = 0.0679 , 유의x



- 직선에 가까운 패턴 존재, 선형성 가정 충족
- 수평 추세선 관찰, 등분산 가정 충족
- 아주 일부 대각선 벗어남, 정규성 가정 충족

분석 - 좋아요 수

```
- like.lm<-lm(좋아요수~재생수+X1위횟수,data=likes)
summary(like.lm)
```

재생수 p-value = 4.16e-16

X1위횟수 p-value = 0.000247

R-squared : 0.5207

P-value : < 2.2e-16

```
> summary(likes.lm)
```

Call:

```
lm(formula = 좋아요수 ~ 앨범좋아요수 + 팬수 + 최고순위 + X1위횟수 +
    재생수, data = likes)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-17.7373	-2.4674	-0.4231	2.2532	10.0408

Coefficients:

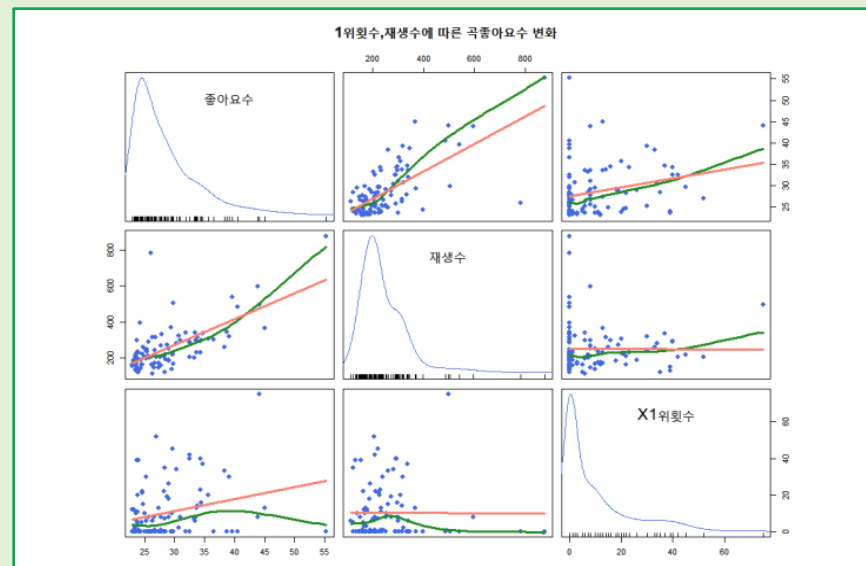
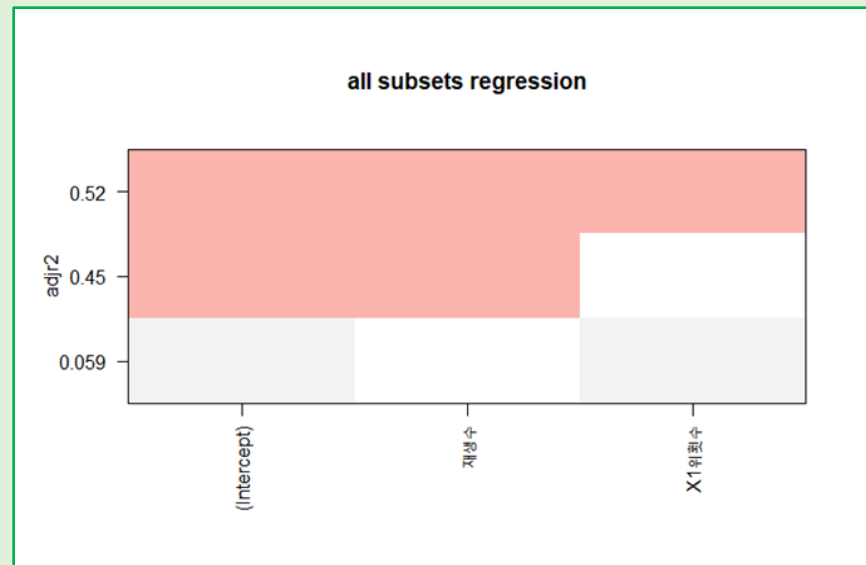
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18.619230	1.101667	16.901	< 2e-16 ***
앨범좋아요수	-0.213354	0.165761	-1.287	0.201213
팬수	0.052093	0.022504	2.315	0.022798 *
최고순위	0.025467	0.039206	0.650	0.517566
X1위횟수	0.108517	0.028886	3.757	0.000298 ***
재생수	0.032940	0.003743	8.801	6.51e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.979 on 94 degrees of freedom

Multiple R-squared: 0.5594, Adjusted R-squared: 0.536

F-statistic: 23.87 on 5 and 94 DF, p-value: 2.004e-15



분석 - 앨범 좋아요 수

```

      F      M      O
4.850000 5.213636 4.900000
> tapply(likes$앨범좋아요수, likes$장르, mean)
      랩      팝      발라드      댄스      록      R&B      기타
10.542857  2.175000  2.974074  7.946667  5.642857  3.986667  3.660000
> tapply(likes$앨범좋아요수, likes$연도, mean)
      2011      2012      2013      2014      2015      2016      2017      2018      2019      2020      2021      2022
0.900000  2.833333  4.300000  3.112500  1.250000  4.318182  5.160714  6.142857  5.556250  6.366667  6.233333  6.400000
> tapply(likes$앨범좋아요수, likes$성별, mean)
      F      M      O
4.850000 5.213636 4.900000

```

- 장르 별 평균 앨범 좋아요 수 : 랩 > 댄스 > 록 > R&B > 기타 > 발라드 > 팝
- 연도 별 평균 앨범 좋아요 수 : 2022 > 2020 > 2021 > 2018
- 성별 별 평균 앨범 좋아요 수 : 남성 > 혼성 > 여성
 but 눈에 띄는 차이x,
 (남성 가수가 여성 가수보다 근소하게 높은 것으로 나옴)
 (혼성 두 팀, 표본이 너무 적음)

분석 - 앨범 좋아요 수 (일원분산분석)

```
result<-aov(앨범좋아요수~성별, data=likes)
```

```
result
```

```
summary(result)
```

#p-value=0.926, 유의x, 성별 간 앨범 좋아요 수의 차이는 존재x

```
> summary(result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
성별	2	2.9	1.462	0.077	0.926
Residuals	97	1851.7	19.089		

```
result<-aov(앨범좋아요수~장르, data=likes)
```

```
result
```

```
summary(result)
```

#p-value=8.16e-10, 유의o, 장르 간 앨범 좋아요 수의 차이는 존재

```
> summary(result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
장르	6	802.4	133.73	11.82	8.16e-10 ***
Residuals	93	1052.2	11.31		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
result<-aov(앨범좋아요수~연도, data=likes)
```

```
result
```

```
summary(result)
```

#p-value=0.0142, 유의o, 연도 간 앨범 좋아요 수의 차이는 존재

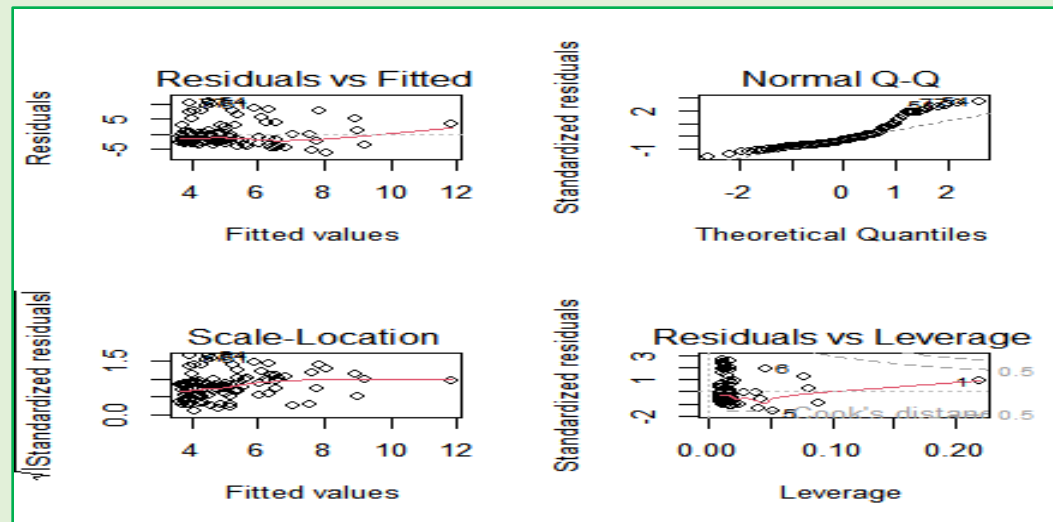
```
> summary(result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
연도	1	110.9	110.87	6.231	0.0142 *
Residuals	98	1743.7	17.79		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

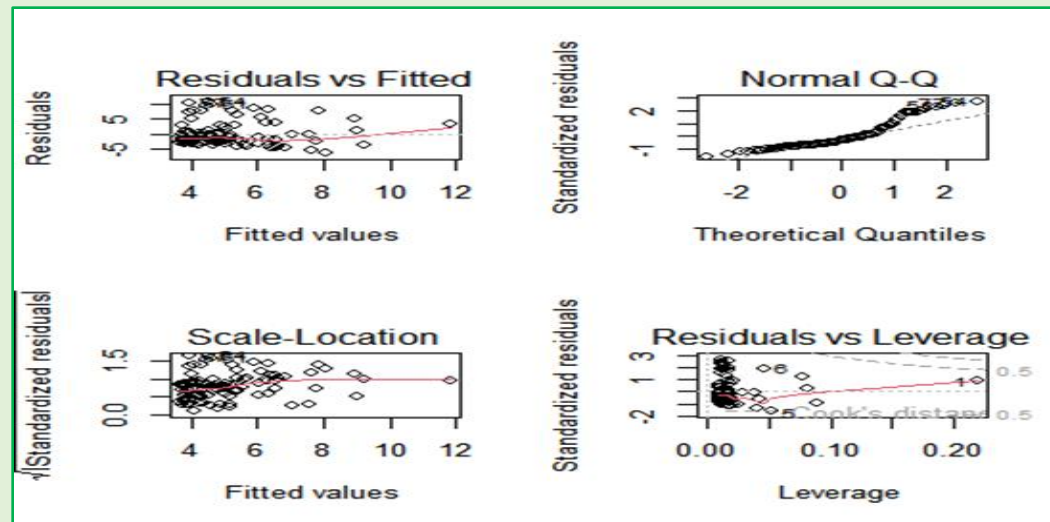
분석 - 앨범 좋아요 수 (상관분석, 회귀진단)

- `cor(likes$앨범좋아요수, likes$좋아요수)`
#상관계수 = **0.3414813**
- `with(likes, cor.test(앨범좋아요수, 좋아요수))`
#p-value = **0.000507**, 유의
좋아요수-앨범좋아요수 상관관계 존재
- `likes.lm<-lm(앨범좋아요수~좋아요수, data=likes)`
`plot(likes.lm)`
#검정결과 정규성 충족x -> 회귀분석 불가능



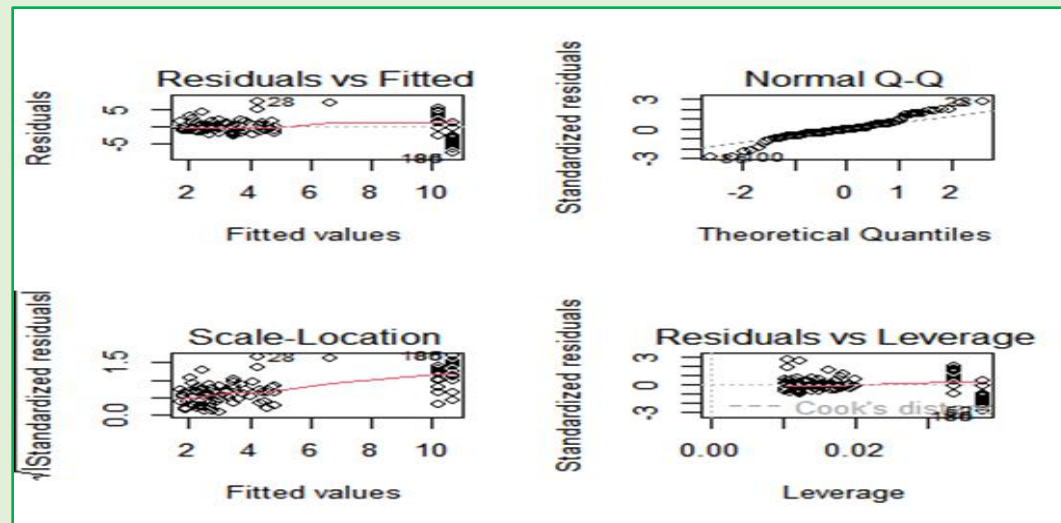
분석 - 앨범 좋아요 수 (상관분석, 회귀진단)

- `cor(likes$앨범좋아요수, likes$팬수)`
#상관계수 = 0.7803542
- `with(likes, cor.test(앨범좋아요수, 팬수))`
#p-value = 2.2e-16. 유의o
팬수-앨범좋아요수 상관관계 존재
- `likes.lm<-lm(앨범좋아요수~좋아요수, data=likes)`
`plot(likes.lm)`
#검정결과 정규성, 등분산성 충족x -> 회귀분석 불가능



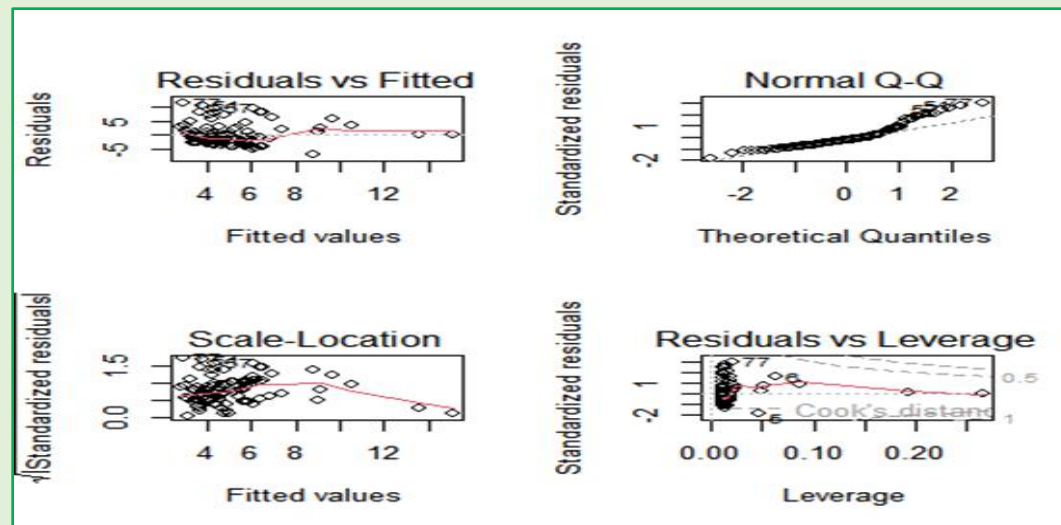
분석 - 앨범 좋아요 수 (상관분석, 회귀진단)

- `cor(likes$앨범좋아요수, likes$1위횟수)`
#상관계수 = -0.05071845
- `with(likes, cor.test(앨범좋아요수, 1위횟수))`
#p-value = 0.6163., 유의x
1위횟수-앨범좋아요수 상관관계 존재x
- `likes.lm<-lm(앨범좋아요수~1위횟수, data=likes)`
`plot(likes.lm)`
#검정결과 선형성, 등분산성 충족x -> 회귀분석 불가능



분석 - 앨범 좋아요 수 (상관분석, 회귀진단)

- `cor(likes$앨범좋아요수, likes$재생수)`
#상관계수 = 0.457656
- `with(likes, cor.test(앨범좋아요수, 재생수))`
#p-value = 1.689e-06, 유의x
1위횟수-앨범좋아요수 상관관계 존재x
- `likes.lm<-lm(앨범좋아요수~재생수, data=likes)`
`plot(likes.lm)`
#검정결과 정규성 충족x -> 회귀분석 불가능



분석 - 재생 수

```
> tapply(likes$재생수, likes$장르, mean)
R&B      기타      댄스      랩      록      발라드      팝
193.9600 228.5900 314.1533 321.7714 202.8000 259.0852 202.9583
> tapply(likes$재생수, likes$성별, mean)
F      M      O
214.5656 271.9424 224.2500
> tapply(likes$재생수, likes$연도, mean)
2011      2012      2013      2014      2015      2016      2017
154.3000 231.0667 241.7000 214.8625 177.7500 224.3818 304.4821
2018      2019      2020      2021      2022
265.8143 244.5312 241.2556 196.2833 122.6000
```

- 장르 별 평균 재생수 : 댄스 > 랩 > 발라드 > 기타 > 팝 > 록 > R&B
- 성별 별 평균 재생수 : 남성 > 혼성 > 여성
(혼성 두 팀, 표본이 너무 적음)

```
> cor(likes$재생수, likes$좋아요수 )
[1] 0.6783962
> cor(likes$재생수, likes$팬수 )
[1] 0.242034
> cor(likes$재생수, likes$앨범좋아요수 )
[1] 0.4576756
> cor(likes$재생수, likes$X1위횟수 )
[1] -0.005917267
```

- 상관계수 : 좋아요수 > 앨범좋아요수 > 팬수 > 1위횟수
but 단독적으로 사용 어려움.

분석 - 재생 수

```
> result<-aov(재생수~장르, data=likes)
> summary(result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
장르	6	229230	38205	2.75	0.0167 *
Residuals	93	1292195	13895		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#p-value=0.0167, 유의o, 장르 간 재생수의 차이는 존재x

```
> result<-aov(재생수~성별, data=likes)
> summary(result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
성별	2	72592	36296	2.43	0.0934 .
Residuals	97	1448833	14936		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#p-value=0.0934, 유의x, 성별 간 재생수의 차이는 존재

```
> result<-aov(재생수~연도, data=likes)
> summary(result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
연도	1	17	17	0.001	0.974
Residuals	98	1521408	15525		

#p-value=0.974, 유의x, 연도 간 재생수의 차이는 존재

분석 - 재생 수

```
> result<-aov(재생수~장르*성별, data=likes)
> summary(result)
          Df Sum Sq Mean Sq F value    Pr(>F)    
장르        6 229230   38205   3.228 0.00667 **
성별        2  59608   29804   2.518 0.08671 .  
장르:성별    7 238253   34036   2.875 0.00966 **
Residuals   84 994334   11837
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#p-value=0.00966, 유의o,
재생수는 장르와 성별의 관계에 의해서 영향을 받음

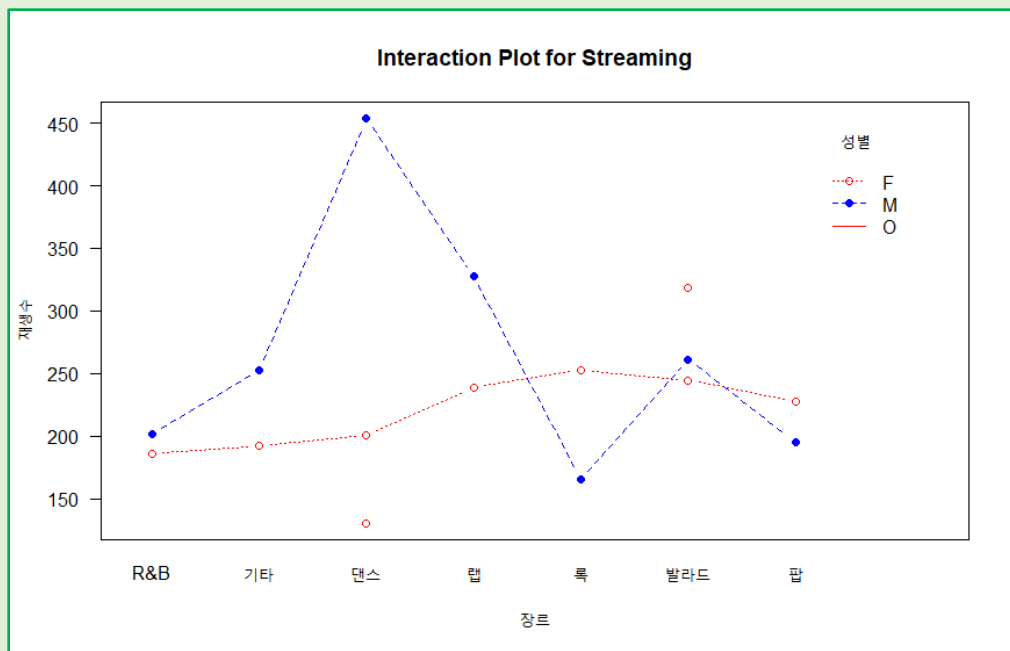
```
> result<-aov(재생수~장르*연도, data=likes)
> summary(result)
          Df Sum Sq Mean Sq F value    Pr(>F)    
장르        6 229230   38205   2.867 0.0136 *
연도        1   7513    7513   0.564 0.4548
장르:연도    6 138664   23111   1.734 0.1227
Residuals   86 1146017   13326
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#p-value=0.1227, 유의x

```
> result<-aov(재생수~성별*연도, data=likes)
> summary(result)
          Df Sum Sq Mean Sq F value    Pr(>F)    
성별        2  72592   36296   2.405 0.0958 .  
연도        1   3130    3130   0.207 0.6498
성별:연도    2  27272   13636   0.904 0.4086
Residuals   94 1418430   15090
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#p-value=0.4086, 유의x

분석 - 재생 수



```
> #남녀만 비교(혼성 가수의 경우 표본 수가 부족) - 유의o
> fm<-read.csv('좋아요수-최고순위-일위횟수추가-단위수정-재생수- 남여만.csv',fileEncoding="euc-kr")
> fm.aov<-aov(재생수~성별, data=fm)
> summary(fm.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
성별	1	70948	70948	4.759	0.0316 *
Residuals	96	1431217	14909		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#p-value=0.0316, 유의o, 성별에 따른 재생수 차이 존재

분석 - 재생 수

```
> likes.lm<-lm(formula = 재생수~장르, data=likes)
> summary(likes.lm)

Call:
lm(formula = 재생수 ~ 장르, data = likes)

Residuals:
    Min       1Q   Median       3Q      Max
-199.97  -69.37  -19.89   41.20  554.73

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   193.960     30.435   6.373 7.04e-09 ***
장르기타       34.630     48.122   0.720 0.47356
장르댄스      120.193     43.042   2.792 0.00635 **
장르랩       127.811     43.804   2.918 0.00442 **
장르록         8.840     53.956   0.164 0.87021
장르발라드    65.125     37.959   1.716 0.08956 .
장르팝         8.998     45.653   0.197 0.84418
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 117.9 on 93 degrees of freedom
Multiple R-squared:  0.1507,    Adjusted R-squared:  0.09587
F-statistic:  2.75 on 6 and 93 DF,  p-value: 0.01666
```

#댄스 p-value : 0.00635 , 랩 p-value : 0.00442
댄스, 랩 만 유의한 차이.

- 기준 범주 : R&B
- 앞서 분석한 장르별 평균 재생수와 비교 시 **가장 낮은 평균값 : R&B, 가장 높은 평균값 : 댄스, 랩**

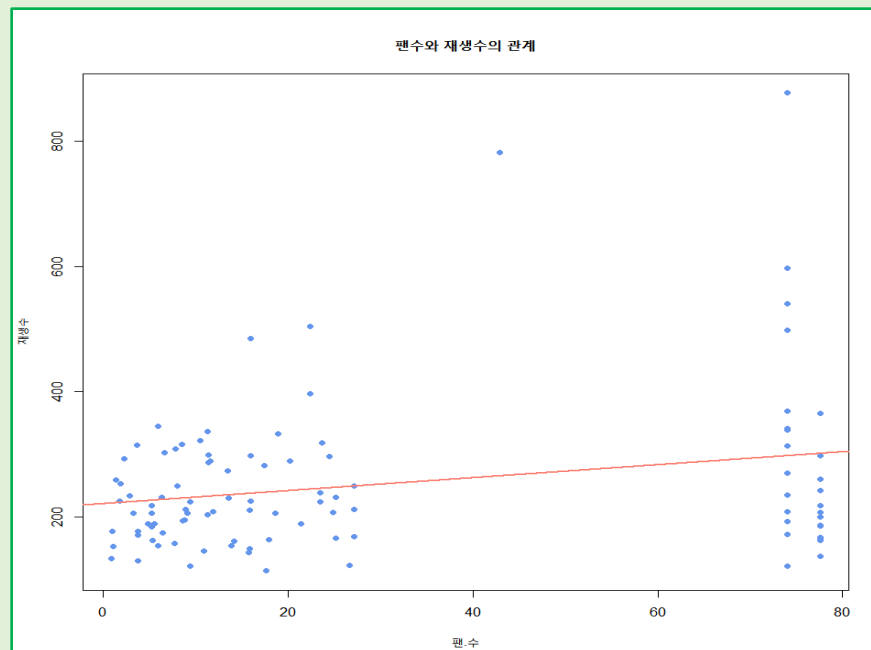
분석 - 재생 수

```
> result<-aov(재생수~팬수, data=likes)
> summary(result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
팬수	1	89126	89126	6.098	0.0153 *
Residuals	98	1432299	14615		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#p-value=0.0153, 유의o, 팬수에 따른 재생수 차이 존재



- 정(+)의 선형관계

분석 - 재생 수

```
> cor.test(~재생수+팬수, data=likes, subset=(성별=='F'))

Pearson's product-moment correlation

data:  재생수 and 팬수
t = 0.06814, df = 30, p-value = 0.9461
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3377192  0.3595740
sample estimates:
      cor 
0.0124397
```

#p-value=0.90401, 유의x
여성가수의 팬수 - 재생수 상관관계 존재x

```
> cor.test(~재생수+팬수, data=likes, subset=(성별=='M'))

Pearson's product-moment correlation

data:  재생수 and 팬수
t = 3.6603, df = 64, p-value = 0.0005119
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1935127 0.5978846
sample estimates:
      cor 
0.4160588
```

#p-value=0.0005119, 유의o
남성가수의 팬수 - 재생수 상관관계 존재

결론

- **곡좋아요수 변화**에 압도적으로 큰 영향을 미치는 변수는 **재생수**이며, 두번째는 **1위횟수**이다.
- **팬수가 적은 가수**여도 발라드, 알앤비 **장르**를 통해 대중들에게 어필한다면 타장르에 비해 비교적 **곡 좋아요수**를 많이 받을 수 있다.
- **앨범좋아요 - 팬수** : 강한 정의 선형관계 존재
-> **팬수가 증가할수록 앨범좋아요수** 또한 증가
- **여성가수의 곡재생수**에는 **팬수** 영향x
남성가수의 곡재생수에는 **팬수** 영향o
-> 남성가수 곡재생수에만 팬수가 영향을 미치는 것은 잘 알려진 '스밍' 문화의 영향일수도.
(남성가수는 여성가수와 달리 팬덤형 가수라는 인식이 강한 편)

감사합니다 😊