Useful K-pop Industry Probability and Statistics

Alexis Petito

PROBABILITY AND APPLIED STATISTICS - CSCI-3327-001

Table of Contents

Introduction

The K-pop industry has massive global influence and is subject to continuous evolution, it is a complex ecosystem where various factors such as, height, weight and country of origin have strong influence on the trajectory of idols' careers. This report takes a nuanced examination of the correlation between these and various other factors in the industry. This dynamic ecosystem is characterized by a fusion of musical innovation, captivating performances, and a unique cultural phenomenon that transcends borders. It is imperative to recognize that various factors, beyond musical talent alone, contribute significantly to K-pop careers and groups.

Height, weight, and country of origin, seemingly unrelated to musical ability, are among the myriad factors that hold substantial influence. This report embarks on an in-depth exploration, offering a nuanced examination of the many variables within the K-pop industry, some that are not easily extracted without specific examination. The aim is to unravel the complexities that are underneath the success stories and challenges faced by K-pop idols and groups.

Some Heavily Weighted Factors Include:

1. Height:
In an industry where visuals are as important as musical talent, some say even more, an idol's height often becomes a defining characteristic. Whether this be their flaw or strength. The perceived ideal height, influenced by societal and industry standards, can influence an idol's opportunities, roles within a group, and overall marketability.

2. Weight:
The K-pop industry places a significant emphasis on appearance, and an idol's weight is a factor that permeates discussions around body image and beauty standards. This report explores how an idol's weight factors into the equation.

3. Country of Origin:
The global nature of the K-pop is ever increasing which means that an idol's country of origin holds unique significance. If a group would like to have more global appeal it would be wise to select a diverse range of idols. The most popular outside of Korea being China, Japan, and The United States. This paper will touch on how an idol's country of origin can influence other factors in the database.

Nuanced Correlations:
This report seeks to move beyond superficial observations, delving into the intricate web of relationships between these factors and the multifaceted nature of K-pop careers. By doing so, the aim is to provide a comprehensive understanding of the industry's dynamics, offering insights that extend beyond music to encompass the diverse elements that contribute to an idol's journey.

In the pages that follow, we explore the correlations, challenges, and triumphs within the K-pop industry, shining a light on the diverse factors that shape the narrative of its global impact.

Chapter 1

Mean, Median, Mode

Suppose a company is recruiting female trainees. It would be advisable to look at industry standards of what the average female K-pop idol's attributes are to give the company the greatest success. By getting the average weight, median height, and most common birthplace you can paint an image of what the most statistically successful traits a trainee might possess.

Average weight females
48.74786325 kg

Median height
165 cm

Most common birthplace
Seoul

Looking at this data it can be concluded that the trainee with the most success would likely be 165 cm tall, with a weight of 48.78 kg, and born in Seoul, South Korea.

Standard Deviation / Variance

Now we look at age as this is another important trait in the Kpop industry. To do this we begin by looking at the standard deviation of the age of male idols born in Japan and then compare them to the standard deviation of male idols born in South Korea.

Standard Deviation of Age Born in South Korea
5.865081276

Standard Deviation of Age Born in Japan
3.477580661

This data can provide valuable insights on the ages of idols in the industry. Male idols from Japan tend to have less outliers in age. Whereas male idols from South Korea are seen having more outliers in age. This could possibly be attributed to training if a foreign idol does not make satisfactory progress they may end up going back to their home country whereas an idol born in South Korea may have a better chance to debut at an older age.

Chapter 2

Set Theory:

Talent Pools Using Union, Intersection, and Compliment:

Non-Unique Stage Names (Set A):
Exploring idols with non-unique stage names reveals a pool large pool of talent integrates their first names into their stage personas. This set presents a mix of individuals, some exclusively committed to a single group, while others are engaged in collaborative groups. This could influence recruiting if an idol with a non-unique name wanted to be selected to create another group.

Artists in More Than One Group (Set B):
Idols engaged in more than one group may not be as attractive for recruiting for another collaborative group as the company may turn down the offer if they do not have the time in their schedule. This group was much sparser, but still has a notable number of idols that recruiters would potentially choose to skip or recruit.

Intersection

Intersection - A Recruitment guide:
The intersection of Set A and Set B highlights a unique group where idols possess both non-unique stage names and a penchant for engaging in multiple groups. If you were to want to recruit idols both these attributes it would be easiest to find the intersection of these two attributed. Seen below are the results of this operation.

Intersection = {Baekhyun, Bohyung, Bomi, Bora, Chaekyung, Changbin, Chenle, Chiayi, Chorong, Dayoung, Doah, Donghae, Donghyun, Dongwoo, Doyeon, Euijin, Eunwoo, Gayoon, Haseul, Heejin, Hyemi, Hyeyeon, Hyungseop, Hyunjin, Hyunseung, Hyunyoung, Jaekyung, Jaemin, Jeno, Jimin, Jinwoo, Jisook, Jisung, Jiwon, Jiyoon, Keumjo, Kyuhyun, Kyungri, Luda, Mina, Mina, Minhyun, Momo, Narae, Nayoung, Renjun, Ryeowook, Sana, Sebin, Seulgi, Seungah, Shotaro, Sihyun, Siwon, Sohee, Sojin, Somyi, Soobin, Sungchan, Sungjae, Sungjong, Sungmin, Sungwoon, Sungyeol, Taehyun, Taeil, Taemin, Taesung, Taeyeon, Taeyong, Woohee, Woohyun, Woojin, Woori, Yebin, Yeojin, Yoojung, Yoonhye, Youngmin, Yuna, Yuta}

Union

If a recruiter was looking to fill more than one position and wanted someone who either used their first name as a stage name or was in more than one group using the union would be the advisable operation to determine the available talent pool. Seen below are the results.

Union =
{Aeji,AhIn,Ahra,Ahyeon,,Ahyoon,Youngmin,Asa,Asahi,Bada,Dongho,Baekhyun,Bahiyyih,Chan,Minhyuk,Beomgyu,BoA,Boeun,Bohyung,Bokeun,Bom,Bomi,Bomin,Jiyeon,Bora,Boram,Byeongkwan,Byungchan,Chaehyun,Chaejeong,Chaekyung,Chaelin,Chaerin,Chaeryeong,Chaesol,Chaewon,Chaeyeon,Chaeyoung,Yoochan,Changbin,Changbum,Changgyu,Changsub,Changsun,Changuk,Chanmi,Chansung,Chanyeol,Chanyong,Chanyoung,Jongdae,Chenle,Cheongeum,Chiayi,Chihoon,Choa,Yerim,Chorong,Chowon,Choyeon,Chris,Chungha,Jiwoo,Daehwi,Daehyeon,Daehyun,Daeil,Daeun,Daeyeol,Dahye,Dahyun,Dain,Dajeong,Dakyung,Damhee,Damin,Dana,Daniel,Daseul,Dasom,Dawon,Dayeon,Dayoung,Dayun,Denise,Doah,Doha,Dohee,Dohwan,Dohyon,Dohyuk,Doi,Dokyun,Donggeon,Donghae,Donghan,Dongheon,Donghun,Donghyuk,Donghyun,Dongjun,Dongkyu,Dongmyeong,Dongpyo,Dongwan,Dongwon,Dongwoo,Dongwoon,Dongyun,Doojoon,

Doryun,Douhyun,Dowoon,Doyeon,Dongyoung,Doyoung,Duna,Euijin,Eunbi,Eunbin,Eunchae,Eunho,Hyukjae,Eunil,Eunjae,Eunji,Eunjin,Eunjo,Eunjun,Eunjung,Eunki,Eunkwang,Eunsang,Juyeon,Eunsung,Eunwoo,Eunyoung,Sojung,Gabin,Gaeul,Gaeun,Gahyeon,Gain,Gaon,Garam,Gayoon,Gayoung,Jiyong,Geumhee,Gihwan,Gijung,Gikwang,Giru,Giseok,Goeun,Gun,Gunmin,Gunwoo,Gyehyeon,Gyeongtae,Gyujin,Gyuri,Habin,Haebin,Haein,Haemin,Haena,Haeri,Haerin,Haeryeong,Haeun,Haewon,Haeyoon,Hajeong,Hakmin,Hamin,Jisung,Hana,Hanbyeol,Hanbyul,Haneul,Hangyeom,Hangyul,Hankook,Hanse,Hansol,Hanwoom,Haram,Harin,Harry-June,Haruna,Haruto,Haseul,Havit,Hayoon,Hayoung,Hayun,Heechan,Heechul,Heedo,Heejin,Heejoo,Heejun,Heeseok,Heeseung,Heesu,Heesun,Hikaru,Himchan,Hina,Hitomi,Hohyeon,Hojoon,Hojung,Hongbin,Honggi,Hongil,Hongjoong,Hongseob,Hongseok,Howon,Hoyoung,Hui,Huihyeon,Huijun,Huiyeon,Hwanhee,Hwanwoong,Hweseung,Hwi,Hyangsuk,Hyebin,Hyein,Hyejeong,Hyeji,Hyejin,Hyeju,Hyemi,Hyeongjun,Hyeongseok,Hyeran,Hyeri,Hyerim,Hyerin,Hyewon,Hyeyeon,Hyogyeong,Hyojin,Hyojung,Hyori,Hyosun,Hyoyeon,Hyuk,Sanghyuk,Hyukjin,Hyuna,HyunA,Hyunbin,Hyungseok,Hyungseop,Hyungsik,Hyungwon,Hyunho,Hyunji,Hyunjin,Hyunjoo,Hyunjun,Hyunmin,Hyunoh,Hyunseo,Hyunseong,Hyunseung,Hyunsik,Hyunsoo,Hyunsuk,Hyunuk,Hyunwoo,Hyunwook,Hyunwoong,Hyunyoung,Ilhoon,Inho,Inpyo,Inseong,Insoo,Intak,Juhyun,Jaechan,Jaehan,Jaehee,Jaeho,Jaehwan,Jaehyo,Jaehyuk,Jaehyun,Yoonoh,Jaejin,Jaejoong,Jaekyung,Jaemin,Jaeseok,Jaeun,Jaeyoon,Jaeyun,Jangjun,Jehyun,Jennie,Jeno,Jeonggyu,Jeonghan,Jeonghee,Jeongkyun,Jeongmin,Jeongseung,Jeonguk,Jeongwoo,Jeongyeon,Jewon,Jia,Jiae,Jiahn,Jian,Jibeom,Jieun,Jiheon,Jiho,Jihoon,Jihun,Jihyo,Jihyun,Jimin,Jingyu,Jinha,Jinho,Jinhong,Jinhwa,Jinju,Jinkwon,Jinsik,Jinsol,Jinsung,Jinwoo,Jinyoung,Jisoo,Jisook,Jisu,Jisun,Jiwon,Jiyoon,Jiyoung,Seo,Jongho,Jonghoon,Jonghwan,Jonghyeong,Jonghyun,Jongseob,Jongup,Joochan,Jooheon,Joohyoung,Jooyeon,Jueun,Jooeun,Juhyeon,Jungah,Junghoon,Junghun,Junghwa,Junghwan,Jungkook,Jungmo,Jungshin,Jungwon,Jungwoo,Junho,Junhyeok,Junhyuk,Junhyung,Junkyu,Junmin,Junseo,Junseok,Junseong,Juntae,Junwook,Junyoung,Juri,Kaede,Kaeun,Jongin,Kairi,Kangmin,Kangsung,Karin,Kazuha,Keita,Kelly,Kenta,Keonhee,Keonwoo,Keumjo,Kibum,Kichun,Kihyun,Jungeun,Sanggyun,Kiseop,Kisu,Kiwon,Kokoro,Kookheon,Kotone,Kun,Kwanghee,Kwangjin,Kwangmin,Kyeongheon,Kyuhwan,Kyuhyuk,Kyuhyun,Kyujin,Jieqiong,Kyungho,Kyungil,Kyungjun,Kyungmin,Kyungmun,Kyungri,Kyungyoon,Myungsoo,Leo,Taekwon,Euiwoong,Sooyoung,Longguo,Junghei,Luda,Marin,Lee,Mashiho,Mashiro,Meiqi,Mihee,Mijoo,Miko,Miku,Mimi,Mina,Minah,Minchan,Mingi,Mingyu,Minha,Minhee,Minho,Minhwan,Minhyun,Minjae,Minji,Minjoo,Minjun,Minjung,Minkyung,Minpyo,Minseo,Minsu,Minwoo,Mire,Miso,Miyeon,Moko,Momo,Momoka,Moonhee,Moonvok,Myungho,Myungji,Hakyeon,Naeun,Nahyun,Nako,Nakyung,Namjoo,Jinah,Nara,Narae,Nari,Narin,Nayeon,Nayoung,Nayun,Rakwon,Wonsik,Rayeon,Rei,Minki,Yebin,Renjun,Ria,Rinji,Ririka,Riwon,Rokhyun,Rui,Ruka,Ryeowook,Ryujin,Saebyeol,Saerom,Sakang,Sakura,Samuel,San,Sana,Sanga,Sangah,Sangdo,Sangho,Sanghoon,Sangil,Sangmin,Sangwoo,Sangyeon,Sanha,Sebin,Seeun,Segye,Sehun,Sehyung,Sejeong,Sejin,Sejun,Semi,Seoham,Seoho,Joohyun,Seokhwa,Hyunjung,Seolhyun,Seonghwa,Seonglee,Seongmin,Seongsoo,Seonyu,Seowon,Seoyeon,Seoyoung,Serim,Seulgi,Seungah,Seungchan,Seungeun,Seunghan,Seunghee,Seungheon,Seungho,Seunghun,Seunghwan,Seunghyub,Seunghyun,Seungjun,Seungkwan,Seungmin,Seungsik,Seungwoo,Seungyeon,Seungyeop,Seungyoon,Seyeon,Seyong,Seyoung,Shana,Shihyun,Shinwon,Shinyoung,Shotaro,Shuhua,Sieun,Sihyeon,Sihyoung,Sihyun,Simyeong,Sion,Siwan,Siwon,Siwoo,Siyeon,Siyoon,Siyoung,Sodam,Soeun,Sohee,Sohye,Sohyun,Sojin,Solbin,Solji,Somi,Somin,Somyi,Songhee,SongSun,Songyee,Soobin,Soodam,Soohyun,Soojin,Soojung,Soomin,Sora,Soyeon,Soyoung,Steven,Sua,Subin,Sugyeong,Suhye,Suil,Sujeong,Suji,Sujin,Sukjun,Sumin,Sungah,Sungchan,Sunghoon,Sunghyuk,S

unghyun,Sungjae,Sungjin,Sungjong,Sungjoo,Sungjun,Sungkyu,Sungmin,Sungoh,Sungwoon,Sungyeol,Sungyeon,Sunhwa,Sunmi,Sunwoo,Suri,Surin,Suwoong,Suyeon,Suyun,Taecyeon,Taedong,Taeeun,Taeha,Taeheon,Taehun,Taehyun,Taeil,Taekhyeon,Taemin,Taerae,Taeseon,Taesung,Taewoong,Taeyang,Taeyeon,Taeyong,Taeyoung,Takuya,Leechaiyapornkul,Miyoung,Tsuki,Tzuyu,Yukwon,Via,Viian,Wenhan,Wheein,Sicheng,Woncheol,Wondae,Wonjin,Wonjun,Wonpil,Wontak,Wonwoo,Wonyoung,Woobin,Woochul,Woodam,Woohee,Woohyun,Woojin,Woojoo,Wooju,Woolim,Woonggi,Woori,Wooseok,Wooyeop,Wooyoung,Minseok,Xuanyi,Yechan,Yedam,Yeeun,Yeham,Yehyeon,Yein,Yeji,Yeju,Yejun,Yena,Yeojin,Yeonhee,Yeonho,Yeonjae,Yeonje,Yeonji,Yeonjoo,Yeonjun,Yeonjung,Yeonkyu,Yeonseo,Yeontae,Jinsook,Yeori,Yeosang,Yeowool,Yeram,Yerin,Yeseo,Yeseul,Yeso,Yewang,Yewon,Yibo,Yijeong,Yiren,Yixuan,Yoel,Yohan,Yonggeun,Yongguk,Yongha,Yonghee,Yonghoon,Yonghwa,Yongseok,Yongseung,Yoochun,Yoohyeon,Yoojin,Yoojun,Yoojung,Yookyung,Yoomin,Yoon,Yoona,Yoonbin,Yoonhye,Yoonjo,Yoonsung,Yoosung,Yooyeon,Yooyoung,Yoseob,Youngbin,Youngeun,Youngheun,Youngho,Younghoon,Youngjae,Youngji,Youngseo,Yua,Yubeen,Yubin,Yugyeom,Yujeong,Yuji,Yujin,Yuju,Yujun,Yuku,Yukyung,Yulhee,Yuna,Yungyu,Yunho,Yunji,Yunjin,Yunju,Yunkyung,Yunseong,Yunsung,Yuqi,Yuri,Yurim,Yuta,Yuto,Yuuri,Yuvin,Jinye}

Compliment

More Useful Information
For recruiters exploring solo endeavors, you may want to recruit an artist that stands out. By recruiting an idol with a unique stage name and no other group affiliations this is effectively finding the compliment. Your recruitment pool would look like this where S represents the full database.

$$Compliement \ = \ S - Union$$

Factorial

Say we want to rearrange the database how many ways can we list the idols in the database?

$$1778!$$

If you put this into most calculators or programs you cannot get a result which speaks to the size of the database.

Combinations

Say there is an event, and a company wants to create we want to create a co-ed, boy group, or girl group of 4 randomly from the dataset. We want to create a co-ed, boy group, or girl group of 4 randomly from the dataset which means that we would be selecting from the full pool of idols. The total number of combinations is shown below.

Setup:

$$C(n,r) = \frac{n!}{r!\,(n-r)!}$$

Work:

$$C(Total\ Num\ of\ Idols, 4) = \frac{1778!}{4!\,(1778-4)!} = 4.150011131e11$$

Permutations

For this event the company must chose a lineup of 3 groups to perform on the main stage of the 8 attending. How many ways is it possible to arrange the selected groups for the performance? The result is show below.

Work:

$$P(n,r) = \frac{8!}{(8-3)!} = 336$$

There are 336 ways to arrange the selected groups for the performance.

Conditional and Bayes

Say we want to randomly sample female idols and determine their specific country of origin. What is the probability of an idol being born in China given they are a female.

$P(A) = Idol\ was\ born\ in\ China = \frac{46}{1778} = 0.026$

$P(A') = Idol\ was\ not\ born\ in\ China = \frac{1732}{1778} = 0.974$

$P(B|A) = The\ idol\ is\ female\ give\ they\ were\ born\ in\ china = \frac{20}{1778} = .011$

$P(B|A') = The\ idol\ is\ female\ given\ they\ were\ not\ born\ in\ china = \frac{869}{1778} = .488$

We Then Find:

$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')} = \frac{.011*.026}{.011*.026 + .488*.011} = .051$

This means there is a 5.1% chance an idol was born in China given they are a female.

Determining Independence

We now want to take a look at stage names and real names in regard to determining independence. The following sets were defined to determine this.

$A = Set\ of\ Stage\ Names$
$B = Set\ of\ Real\ First\ Names$
$A \cap B = the\ intersection\ of\ stage\ names\ and\ first\ names$

To calculate this, we begin by determining whether,

$$P(A \cap B) = P(A) * P(B)\ or\ P(A \cap B) \neq P(A) * P(B)$$

By plugging in the data from the dataset you can see,

$1 * 1 \neq .4308$

Thus, the events are dependent, indicating a connection between the occurrences of stage names and real first names. This suggests that the presence of a stage name is meaningfully associated with an idol's real first name in the dataset. The variables exhibit dependence, implying that changes in one are correlated with changes in the other. This provides valuable insights into stage names and real names in the idol dataset, highlighting a meaningful connection where alterations in one variable tend to coincide with alterations in the other.

Chapter 3

Binomial

Consider a broadcasting station where K-pop idol groups are being considered for a variety show randomly. What is the probability of exactly 2 groups being selected?

To begin we will define the variables,
$n = 288\ Groups$
$y = 2\ Chosen$
$P = \frac{1}{288}$
$q = \frac{287}{288}$

Work:

$$P(2) = P(Y = 2) = \binom{288}{2}\left(\frac{1}{288}\right)^{2}\left(\frac{287}{288}\right)^{286} = .18425$$

Geometric

If an agency were to pick randomly from the database what is the probability that the third idol from a random selection is one with a former group affiliation.

To begin we will define the variables,
$y = 3$
$p = \frac{1}{128}$
$q = \frac{127}{128}$

Work:

$$p(3) = P(Y = 3) = \left(\frac{127}{128}\right)^{3-1}\left(\frac{1}{127}\right) = .00775$$

Hyper Geometric

Suppose a company is forming a boy group of 6 idols and get lucky randomly with two born in The United States for international appeal, what is the probability of selecting this combination from the dataset?

To begin we will define the variables,
$N = 889\ idols$
$n = 6\ male\ idols\ in\ group$
$r = 9\ males\ born\ in\ USA$
$y = 2$

Work:

$$p(2) = P(Y = 2) = \frac{\binom{9}{2} * \binom{889 - 9}{6 - 2}}{\binom{889}{6}} = .001325$$

Negative Binomial

What is the probability of needing to select ten female idols before finding two with a former group affiliation?

To begin we will define the variables,
$r = 2\ idols\ with\ former\ group\ affiliation$
$y = 10\ trials$
$P = \frac{152}{889}$

Work:

$$p(10) = \binom{10 - 1}{2 - 1}\left(\frac{152}{889}\right)^{2}\left(\frac{737}{889}\right)^{10-2} = .0587$$

Poisson

Say we want to consider birth years over time in the database. In a given year, what is the probability of having exactly six K-Pop idols from the dataset born in the year 1998?

To begin we will define the variables,
$\lambda = \frac{Number\ of\ Idols}{Range\ of\ Birth\ Years} = \frac{1778}{2009 - 1977}$
$y = 6$

$$\sum_{k=0}^{6} \frac{55.5625^{k}}{k!} * e^{-55.5625} = 1.5714E - 20$$

Chebyshev

We want to find the ranges of female's weights in the Kpop industry. To do this you can apply Chebyshev's inequality to estimate the interval for which female idols weight is within one standard deviation of the mean.

To begin we will define the variables,
$\mu = 48.74786325$
$\sigma = 15.73985455$
$k = 1$

Work:
Upper Bound
$48.74786325 + 1(15.73985455) = 64.4877178$

Lower Bound
$48.74786325 + 1(15.73985455) = 33.0080087$

Interval Within One Standard Deviation of The Mean
$[33.0080087, 64.4877178]$

The large range within one standard deviation of the mean indicates that the majority of idols have weights that are not significantly different from the average, which is indicative of a certain standard or expectation within the industry. This analysis provides valuable insights into the general patterns of weights among female K-Pop idols, aiding in understanding the industry's standards and potential outliers in terms of body weight.

Chapter 4

The Probability Distribution for a Continuous Random Variable

We can also look at distributions that can help when battling scheduling conflicts. Say we have a pool of 6 idols, and they are being contacted one at a time to host Music Bank. Only one of them has no schedule conflicts. Let Y be the trial on which the idol with no schedule conflicts is selected.

We can find the probability function for Y this is represented by,

$$P(y = 1) = \frac{1}{6}$$
$$P(y = 2) = \frac{1}{5} * \frac{5}{6} = \frac{5}{30} = \frac{1}{6}$$
$$P(y = 3) = \frac{1}{4} * \frac{4}{5} * \frac{5}{6} = \frac{1}{6}$$
$$P(y = 4) = \frac{1}{3} * \frac{3}{4} * \frac{4}{5} * \frac{5}{6} = \frac{1}{6}$$
$$P(y = 5) = \frac{1}{2} * \frac{2}{3} * \frac{3}{4} * \frac{4}{5} * \frac{5}{6} = \frac{1}{6}$$

$$P(y=6) = 1 * \frac{1}{2} * \frac{2}{3} * \frac{3}{4} * \frac{4}{5} * \frac{5}{6} = \frac{1}{6}$$

We can then use this information to give the corresponding distribution function,

$$P(Y < 0) = 0$$
$$P(0 \leq Y < 1) = 0 + \frac{1}{6} = \frac{1}{6}$$
$$P(1 \leq Y < 2) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$
$$P(2 \leq Y < 3) = \frac{2}{6} + \frac{1}{6} = \frac{3}{6}$$
$$P(3 \leq Y < 4) = \frac{3}{6} + \frac{1}{6} = \frac{4}{6}$$
$$P(4 \leq Y < 5) = \frac{4}{6} + \frac{1}{6} = \frac{5}{6}$$
$$P(5 \leq Y < 6) = \frac{5}{6} + \frac{1}{6} = 1$$

We can use this distribution question to solve various questions. For example. What is the probability that you find the first idol without scheduling conflicts when you contact the third idol?

$$P(y=3) = \frac{1}{6}$$

Or what is the probability that you find the first idol without scheduling conflict in contacting less 3 idols?

$$P(y<3) = P(Y=0) + P(Y=1) + P(Y=2) + P(Y=3) = \frac{1}{2}$$

This can be helpful when attempting to determine probability depending on how many candidates a company plans to contact to fill a position of hosting music bank.

Expected Values for Continuous Random Variables

To make the example of filling the music bank host position even more useful value we can use the expected value. This allows us to know the expected number of idols that would need to be contacted to find one without scheduling conflicts.

This can be found by the sum of each possible value multiplied by its probability, so we would end up getting,

$$E(Y) = \sum_{i=1}^{6} y_i * P(Y = y_i) = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

So, you would expect to contact 3.5 idols before finding one without scheduling conflicts for hosting music bank. This can help evaluate expected scouting costs if a company is looking to

fill this position. Or if a scout is underperforming if it is taking them more than an average of 3.5 contacts to land a deal.

The Uniform Probability Distribution

Upon studying male idols weight in the Kpop industry, we find that the male idols have uniformly distributed weight between 60 kg and 64 kg.

Shown by,

$$f(y) = \begin{cases} \frac{1}{64-60}, & 60 < y < 64 \\ 0, & elsewhere \end{cases}$$

With this we can find out the probability that a male idol will fall below or above a certain weight. For Example,

Finding the probability that the weight of the male Kpop idol is below 61kg can be found by,

$$P(60 \leq x \leq 61) = \frac{61-60}{64-60} = \frac{1}{4} = .25$$

Or that the weight of the male Kpop idol is greater than 63 kg can be found by,

$$P(63 < x \leq 64) = \frac{64-63}{64-60} = \frac{1}{4} = .25$$

Chapter 5

Bivariate Probability

In a hypothetical situation we can help industry professionals calculate probabilities of certain more specific outcomes using bivariate probability. Contracts for five reality show gigs are randomly assigned to one or more of three idols, Hwasa, Felix, and HeeJin. Let x denote the number of contracts assigned to Hwasa and y the number of contracts assigned to Felix. Each idol can receive 1, 2, or 3 contracts.

So,
$x = Hwasa$
$y = Felix$

$S = \{(1,1),(1,1),(1,2),(1,3),(2,1),(2,2),(2,2),(2,3),(3,1),(3,2),(3,3),(3,3)\}$

Consider Hwasa as the columns and Felix as the rows we can display the joint probability function for Hwasa and Felix,

Table 1.

Hwasa/Felix	1	2	3

1	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$
2	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{2}{12}$
3	$\frac{2}{12}$	$\frac{2}{12}$	0

Using this we could find the likelihood of Hwasa getting 1 contract and Felix getting 2 using $F(1,2)$. Therefore, the probability of this happening is $\frac{2}{12}$ or $\frac{1}{6}$. Which means there is about a 17% chance of this outcome. This can be useful when trying to see how likely it 2 idols will be given a certain number of contracts out of a specific pool of idols.

Marginal and Conditional Probability Distributions

Marginal

We can use our table from the last example which pertains contract assignment between Hwasa, Felix, and HeeJin, to find the marginal probability distribution of Hwasa.

This can be displayed as a table,

Hwasa	1	2	3
p_1(Hwasa)	$\frac{6}{12}$	$\frac{6}{12}$	$\frac{4}{12}$

Conditional

Using this information to further our understanding of possible outcomes we can now ponder questions such as, given that Felix received 2 contracts, what is the probability that Hwasa received 1 contract?

So, we are looking for,

$$P(Hwasa = 1 \mid Felix = 2)$$

Which can be found by using our table,

$$P(Hwasa = 1 \mid Felix = 2) = \frac{F(1,2)}{\sum_{i=1}^{3} F(i,2)} = \frac{\frac{2}{12}}{\frac{2}{12} + \frac{2}{12} + \frac{2}{12}} = \frac{1}{3}$$

Thus, there is a 33.333% chance when given that Felix received 2 contracts, that Hwasa received 1 contract. This can be extremely useful if contract decisions have been finalized for one artist, but not another and there is uncertainty in business moves.

Independent Random Variables

To expand upon the data in table 1 further we found the marginal distribution function to be a binomial distribution. Now we can use this information to determine whether Felix and Hwasa's contracts are independent or dependent.

Let's look at the case where Felix gets 1 contract and Hwasa gets 1 contract,

$$P(1,1) = \left(\frac{2}{12}\right)$$

So, then we can determine independence using,
$\left(\frac{6}{12}\right) * \left(\frac{6}{12}\right) = \frac{1}{4} \neq \frac{2}{12}$

We can determine that Hwasa and Felix's contracts are dependent. Meaning that in this situation they influence each other. All these statistics can be useful when managing and predicting whether potential contracts will fall through or come to fruition.

Summary

In this paper many different hypotheticals using statistics and probability were explored using a dataset with real information about idols. These statistics can be used for many different benefits when considering recruiting, allocating resources, contracts, and entertainment opportunities. Below is a short summary of some of the findings.

1. **Talent Recruitment and Management:**
    - Exploration of mean, median, mode, and standard deviation for attributes like height, weight, and country of origin contributed valuable benchmarks for talent recruitment.
    - Understanding the average attributes of successful idols can guide talent agencies in selecting trainees with statistically successful traits.

2. **Country of Origin and Global Appeal:**
    - Analyzing the distribution of idols from different countries sheds light on the importance of diversity for global appeal in the K-pop industry.
    - Agencies can use this information to strategically form groups with members from different countries to boost the group's global appeal.

3. **Stage Name and Group Affiliation:**
    - Set theory operations (union, intersection, complement) on stage names and group affiliations offer insights into the talent pool and potential challenges in recruiting.
    - Recruiters can use this information to identify unique individuals or those with specific attributes, streamlining the recruitment process.

4. **Probabilistic Analysis:**

- Probability distributions (binomial, geometric, hypergeometric, etc.) provide tools to assess and predict various scenarios, such as the likelihood of group configurations or the probability of selecting idols with specific attributes.
- This aids in decision-making, event planning, and understanding the dynamics of idol selection and group formation.

5. **Industry Standards and Outliers:**
   - The analysis of weight distribution using Chebyshev's inequality offers insights into industry standards for female idols weight.
   - Identifying the range within one standard deviation helps recognize patterns and potential outliers, contributing to an understanding of the strictly enforced body image standards in the industry.

6. **Expected Values and Continuous Random Variables:**
   - Calculating expected values for continuous random variables provides a measure of central tendency, aiding in predicting the average number of trials needed for specific outcomes.
   - This is valuable for planning events, scheduling, and optimizing resource allocation.

7. **Contract Assignments and Probability Distributions:**
   - Bivariate probability distributions help in understanding the joint and conditional probabilities of contract assignments among idols.
   - This information is crucial for talent agencies when negotiating contracts and managing the participation of idols in various projects.

Resources and References

Mathematical Statistics with Applications 7th Edition
by Dennis Wackerly (Author), William Mendenhall (Author), Richard L. Scheaffer (Author)

1700+ K-Pop Idols Dataset
https://www.kaggle.com/datasets/nicolsalayoarias/all-kpop-idols