

*Project Report*

*On*

# **Statistical Analysis of Market Survey for New Mobile Application Launch**



**IE 6200 ENGINEERING PROBABILITY AND STATISTICS - FALL 18**

**By**

**Parth Umeshchandra Rana (NUID: 001417553)**

**Pratik Rathod (NUID: 001408374)**

**Ajith kumar venkatesulu (NUID: 001403356)**

## ABSTRACT

eRunner is real time mobile application for running competition with two or more players located anywhere in the world. Statistical analysis of pre launching market survey of this application is done to identify market segment and expected revenue. Acceptance of some features have been determined by this survey. Potential customers have been identified by comparing two different market segments and efficient marketing strategy is finalized. Simple random sampling is used for our survey results to identify the requirements of potential customers of eRunner application in this project.

## METHODOLOGY

Pre-launch mobile application survey is done using Google Form. Questions are made to collect information from various people regarding age, gender, running days, features, etc. Collected data are both qualitative and quantitative. Statistical analysis is done using software packages like MS Excel and MINITAB. Conclusion has been derived using this data.

Link for the Google Survey:

[https://docs.google.com/forms/d/e/1FAIpQLSfZUMzkfoabz9PeQTPRX26rxFFRVROpCnp2gCUgwuEQc0DClw/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSfZUMzkfoabz9PeQTPRX26rxFFRVROpCnp2gCUgwuEQc0DClw/viewform?usp=sf_link)

Link for Dataset:

[https://docs.google.com/spreadsheets/d/1r0UdcsIdv15V9xrakeHgfvDcGtel\\_d-ZSAYa2bas9hw/edit#gid=909391011](https://docs.google.com/spreadsheets/d/1r0UdcsIdv15V9xrakeHgfvDcGtel_d-ZSAYa2bas9hw/edit#gid=909391011)

- eRunner Mobile Application:



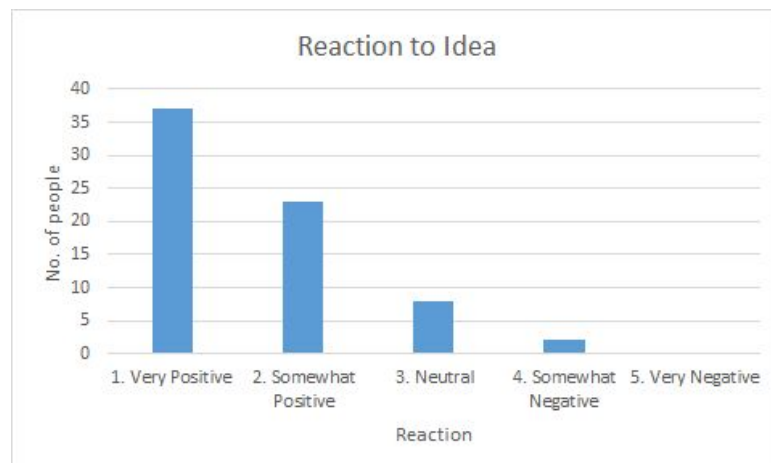
eRunner is real time platform for runners to compete with one or more players located anywhere in the world. People are becoming aware about mHealthCare and Runtastic Running App, Nike+ Run Club, etc. mobile applications are famous platforms. These platforms provide personal data like distance, calories burn, speed, etc. Present platforms are lacking in Fun and Competition. Running can be made more interesting by technology which can attract more people towards running and finally towards healthy life! Two or more players will start running at same time, eRunner mobile application is taking their GPS and gyroscope data and sending them to the cloud server. Using these data precise distance is calculated and by comparing these distances in real time, we can identify winner of the running race. eRunner is all about running, competition, health and fun.

- First Reaction after listening to idea of eRunner:

First reaction after listening / reading to idea of eRunner is capture in this survey. It is important parameter because it can show potential of the idea. Reactions are taken into 5 levels:

- 1) Very Positive
- 2) Somewhat Positive
- 3) Neutral
- 4) Somewhat Negative
- 5) Very Negative

Survey of total 70 people from India and USA is taken and their reactions can be understood from the graph below.



If Very Positive and Somewhat Positive are taken as potential customer probability of liking for overall population can be identified using binomial distribution. To consider this parameter as a binomial distribution following conditions should be fulfilled.

- 1) Fixed no. of trails: Let's say we are targeting 1000 people based on same background as people in this survey.
- 2) “n” trials are independent: Trails are independent because liking for application for one person is not depend on other people.
- 3) Outcomes: “success” or “failure”: Success is person will like the application and failure is person will not like the application.
- 4) Probability of success (p) remains constant from trial to trial.

Probability of success

$$\begin{aligned}
 p &= (\text{Feedback with Very Positive and Somewhat positive}) / \text{Total no. of feedback} \\
 &= (37+23) / 70 \\
 &= 0.8571
 \end{aligned}$$

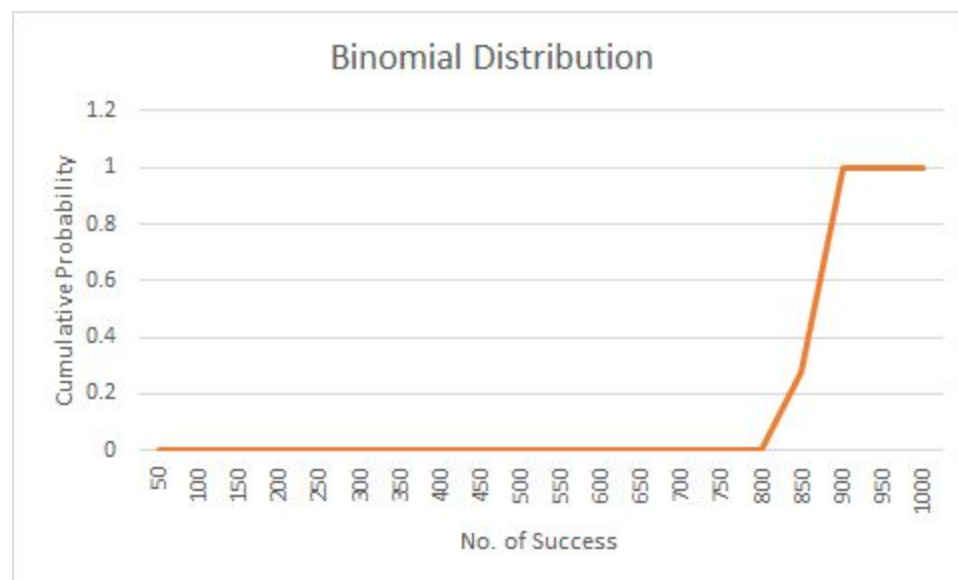
Here n = 1000

$$q = (1-p) = 0.1429$$

Binomial Distribution Formula:

$$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

Result:



Identification of Potential Market Segment using Conditional Probability:

By considering Gender:

Event M = Person is Man ; m = 56

Event W = Person is Woman ; w = 14

Event L = Person is liking the Application

$$\begin{aligned}\text{Probability of person is liking the app while he is man} &= P(L | M) = P(L \text{ and } M) / P(M) \\ &= (49/70) / (56/70) \\ &= 49/56 \\ &= 0.875\end{aligned}$$

$$\begin{aligned}\text{Probability of person is liking the app while he is woman} &= P(L | W) = P(L \text{ and } W) / P(W) \\ &= (11/70) / (14/70) \\ &= 11/14 \\ &= 0.785\end{aligned}$$

This statistics shows that market segment of men is slightly more potential than women. So marketing strategies can be generate accordingly. These results can be more accurate if no. of people surveyed will increase than 70.

## ● Statistical Analysis of Days of Running w.r.t. Age Groups:

Age is one of the factor to be considered for the success of the application because the data collected represents that people of age 21-35 are potentially who will determine the success of this application as they are of huge numbers who have replied to this survey and positively as well to this idea in majority. But we should also consider the data of other age groups to improve the application so that the application will be appropriate for them which will enhance the success of the application.

Age groups :

- Age 21 - 35.
- Age 35 - 50.
- Age > 50.

A survey consisting of 70 people were performed who are from U.S and India to arrive at various conclusions which are illustrated below.

The expected value of people to jog/run **upto 3 days** within the age group **21-35** is calculated below :

$$E(X) = \mu X = \sum [ x_i * P(x_i) ]$$

$$E(X) = 0*0+1*0.0606+2*0.303+3*0.6363=2.5755.$$

The expected value of people to jog/run **more than 3 days** within the age group **21-35** is calculated below

$$E(X) = \mu X = \sum [ x_i * P(x_i) ]$$

$$E(x)=0*0+1*0.076+3*0.23+4*0.307+5*0.3846=3.9.$$

The expected value of people to jog/run **upto 3 days** within the age group **50 above** is calculated below :

$$E(X) = \mu X = \sum [ x_i * P(x_i) ]$$

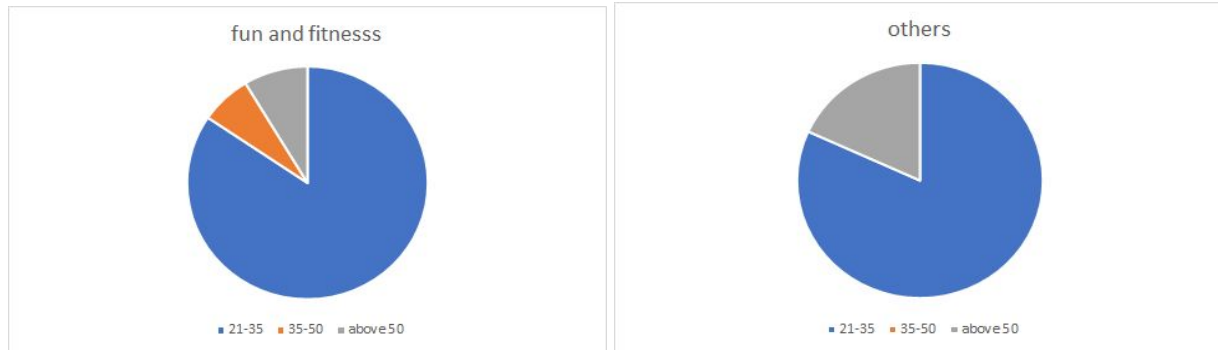
$$E(x) = 0*0+1*0.2+3*0.8 = 2.6.$$

The expected value of people to jog/run **more than 3 days** within the age group **50 above** is calculated below

$$E(X) = \mu X = \sum [ x_i * P(x_i) ]$$

$$E(x) = 4*0.44+5*0.55 = 4.51.$$

- Analysis of Various Objective For Running or Jogging:

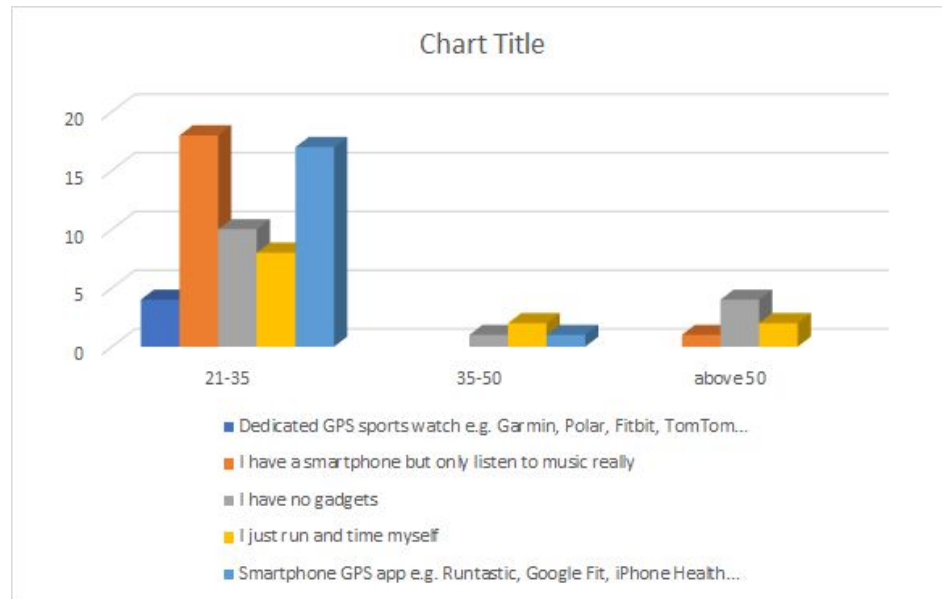


From the above pie charts it is inferred that majority of people who run/jog are concerned about their fitness and do it for fun.

Probability of Objective as Fun or Fitness:

$$\begin{aligned} p &= (\text{Feedback with fun and fitness}) / \text{Total no. of feedback} \\ &= (57) / 68 \\ &= 0.838 \end{aligned}$$

- Analysis of Gadgets used for Running:



Probability table for gadgets that are frequently used :

Age	Dedicated GPS sports watch	I have a smartphone but only listen to music really	I have no gadgets	I just run and time myself	Smartphone GPS app
21 - 35	0.0701	0.315	0.1754	0.14	0.298
36 - 50	0	0	0.25	0.5	0.25
Above 50	0	0	0.667	0.333	0

From the above data it can inferred that people who are between the age 21-35 tend to use smartphones while running or jogging have a probability of 0.298.

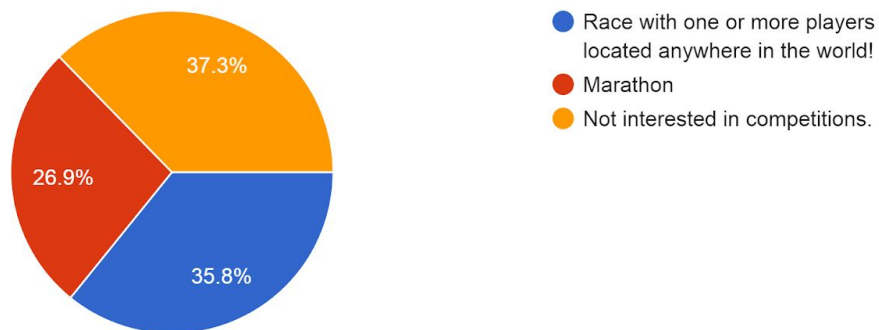


- Analysis for interest in types of competition:

Real time competition is a major fun factor of this application. In this survey is has been covered by one question. Responses can be observed in below graph.

Which of the mode for the race competition will you prefer on this application?

67 responses



Same as gender analysis for market segment, interest in various types of competition can be identified for different age groups using conditional probability concept.

Probability		Competitions		
		Race	Marathon	Not Interested
Age Groups	10 - 20	0	0.25	0.75
	21 - 35	0.41	0.27	0.32
	36 - 50	0	0.5	0.5
	Above 50	0.33	0	0.67

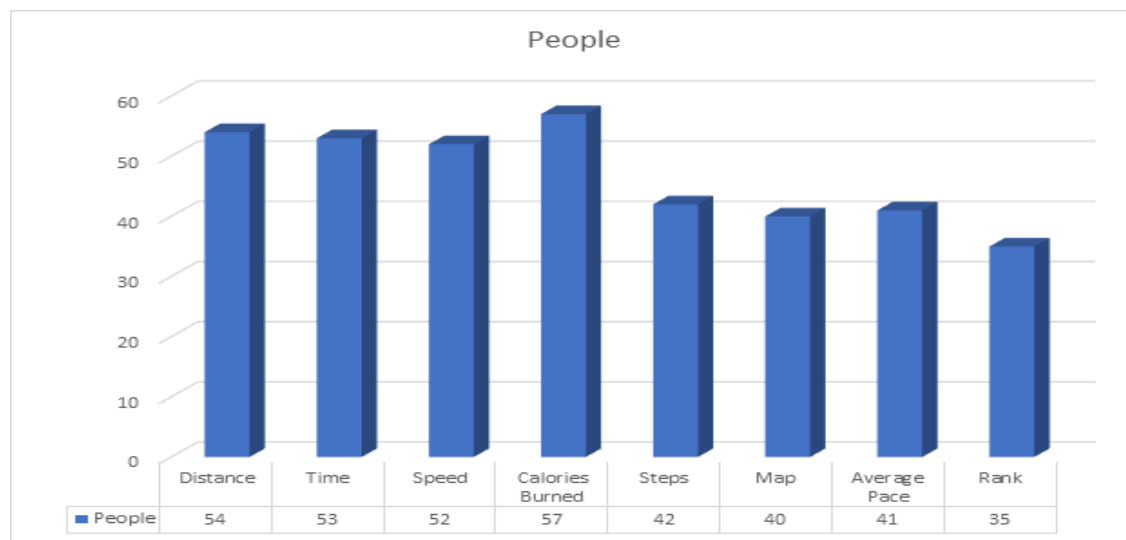
So interest of different age groups in various types of competition can be understood from this probability distribution. From this data marketing strategies can be made to target particular age group.

- Statistical Analysis of Features for Application:

Features of the application are the characteristics that describes its appearance, components, and capabilities. A application feature is one of the distinguishing characteristics of a application that helps boost its appeal to potential buyers, and might be used to formulate a product marketing strategy that highlights the usefulness of the product to targeted potential consumers. In pre launch survey of a eRunner application we have added some features to recognize customer needs which are as below.

1. Distance
2. Time
3. Speed
4. Calories burned
5. Steps
6. Map
7. Average Pace
8. Rank

We surveyed total 70 people by using google form. Everyone of them mentioned their requirement in the form. They had a choice to select multiple features. Most demanding feature was the calories burned. 57 out of 70 people choose that as their one of the required features.



By using variance formula of binomial distribution,

Here let's assume we take a population of 1000 people

Then,

$$n = 1000$$

$p =$  probability of success = Number of people who requested the people/ total number of people(70)

$$q = 1 - p$$

	Feature	People	Probability of success	Variance
1)	Distance	54	0.77	176.33
2)	Time	53	0.76	183.88
3)	Speed	52	0.74	191.02
4)	Calories Burned	57	0.81	151.22
5)	Steps	42	0.60	240.00
6)	Map	40	0.57	244.90
7)	Average Pace	41	0.59	242.65
8)	Rank	35	0.50	250.00

We can see that four features distance, time, speed, calories burned have the less variance which indicated that most of the people wants these four features in the application.

- Statistical Analysis of Money to Participate in the Race / Marathon

In order to participate in the race, you have to chip in some money. We targeted mainly to countries INDIA and USA. We surveyed total 70 people in which 44 decided to chip in money in the form of Indian RS while 26 decided to chip in money in the form of dollars.

By using formula of mean, variance and standard deviation

$$\text{variance} = \sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

$$\text{standard deviation } \sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

$\mu = \text{mean}$

For people who chip in money in dollars (\$) :

n = 26

mean = 31.80769231

variance = 1704.802

standard deviation = 41.28924

For people who chip in money in Indian Rs (₹) :

By using formula of mean, variance and standard deviation as mentioned above,

n = 44

Mean = 279.3181818

Variance= 360204.8732

Standard deviation= 600.1707

- Comparison between market of US and INDIA using confidence interval for differences of mean

Sample 1 = data of chipped money in dollars

Sample 2 = data of chipped money in rupees (equivalent dollars)

Using Minitab with 95% CI limit of  $\sigma_1 / \sigma_2$  is identified between 0.4691 and 11.5228 so  $\sigma_1 = \sigma_2$  can be assumed.

Hypothesis is solved using MINITAB to compare these two samples and identified which one is bigger.

2-Sample t: Dollar, Eq. Dollars

Method

$\mu_1$ : mean of Dollar

$\mu_2$ : mean of Eq. Dollars

Difference:  $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

#### Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Dollar	26	31.808	41.289	8.097
Eq. Dollars	44	6.912	22.6	3.407

#### Estimation for Difference

Difference	Pooled StDev	95% CI for Difference
24.9	30.8	(9.7,40.1)

#### Test

Null hypothesis  $H_0: \mu_1 - \mu_2 < 0$

Alternative hypothesis  $H_1: \mu_1 - \mu_2 > 0$

T-Value	DF	P-Value
3.27	68	0.0017

P value = 0.0017 < alpha = 0.05, so Null Hypothesis is rejected and  $\mu_1 - \mu_2 > 0$  is true.

So market of people who chipped in dollar is bigger than who chipped in rupees with 95% confidence interval for monetization purpose.

● Conclusion:

1. 85% of the people from the survey like this mobile application.
2. Market segment of men is slightly more potential than women.
3. 84% people have objective of fun and fitness from running or jogging and this application provide good combination of fun and fitness which is positive sign for application launch.
4. This application is based on mobile platform which is most recommended by age group of 21-35.
5. 41% of people from age group 21-35 prefer race as mode of competition.
6. Distance, time, speed and calories burned features have the less variance which indicates that most of the people want these four features in the application.
7. Average revenue per person is 32 dollars for USA market and 279 rupees for Indian market.
8. Using hypothesis it is identified that market of people who chipped in dollar is bigger than who chipped in rupees with 95% confidence interval for monetization purpose.