Chapter 7–1A: Estimating a Population Proportion

Proportions

Websters Dictionary lists **one** of the possible definitions of **proportion** as ”the ratio of one part to the whole with respect to a magnitude or quantity.” Statistics uses the term **proportion** to describe the ratio of the **quantity of part of the population divided by the total quantity of the population**

Population proportion

The **population proportion is** the ratio of the number of data bits with the same characteristic divided by the total number of data bits in the population. We denote the **population proportion with the symbol** p. p can be expressed as a decimal, fraction or percent.   **We will express the value of a proportion proportion p as a decimal.** The value of p will be a decimal less than or equal to 1.

What Proportion of M&M's are Blue ?

M&M's are candy-coated pieces of milk chocolate with the letter "m" printed on them, produced by Mars, Incorporated. Forrest Mars, Sr., founder of the Mars Company, got the idea for the confection in the 1930s. Hershey Chocolate Company president William F.R. Murrie had a 20 percent interest in the product, thus the M and M.   In 1950 a black "M" was imprinted on the candies. It was changed to white in 1954.

First produced in 1941, the candies were originally manufactured in brown, yellow, orange, red, green, and violet. Tan replaced violet in 1949. Blue replaced tan in 1995. In 2002, Mars solicited votes to add a new color from three choices. The choices were aqua, pink, and purple.   Purple won.

What Proportion of M&M's are Blue ?

Population Proportion p.

This question implies that we want to know the true proportion of the **entire population** of all M&M's. We call this the **Population Proportion** and denote this with the symbol p.

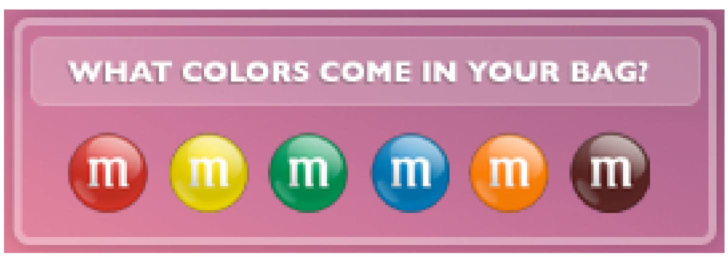
$$p\left(\text{blue M \& M's}\right) = \frac{\text{number of blue M \& M's in the population}}{\text{total number of M \& M's in the population}}$$

To find the true **Population Proportion** p we would have to be able to count every item in the population. We often cannot know or cannot afford to find out the true **Population Proportion** p.

The 2001 product Web Page for Mars stated that the population of Milk Chocolate M&M's contain 13% red, 14% yellow, 16% green. 24% blue, 20% orange and 14% brown M&M's.

The fact that 24% of the population of all M and M's are blue is a statement a lay person would use. In statistics we call the ratio of the number of items in the population that have a given characteristic compared to number of total items in the entire population the Population Proportion. We state the population proportion as a decimal and not as a fraction or a percent.

13% red, 14% yellow, 16% green. 24% blue, 20% orange and 13% brown M&M's.



p = .13 .14 .16 .24 .20 .15

If 13% of all M&M's are red, then $p(\text{red}) = .13$

If 14% of all M&M's are yellow, then $p(\text{yellow}) = .14$

If 16% of all M&M's are green, then $p(\text{green}) = .16$

If 24% of all M&M's are blue, then $p(\text{blue}) = .24$

If 20% of all M&M's are orange, then $p(\text{orange}) = .20$

If 14% of all M&M's are brown, then $p(\text{brown}) = .13$

One may think that Mars makes a given number of each color and then mixes them together before individual bags are filled. This would allow them to know the exact number of Blue M&M's in the whole population. The number of Blue M&M's in the batch divided by the total number of M&M's in the batch would be the population proportion of Blue M&M's for that batch. If every batch stays the same then Mars could know the true proportion of Blue M&M's in the population.

Sample Proportion $\hat{p}$

If we select a single bag of M&M's and count the total number of the number of blue M&M's in the bag and the total number of M&M's in the bag we could find the proportion of blue M&M's **in that sample.** We call this proportion the **Sample Proportion** and denote it with the symbol $\hat{p}$.

$$\hat{p}\left(\text{blue M \& M's}\right) = \frac{\text{number of blue M \& M's in the sample}}{\text{total number of M \& M's in the sample}}$$

The results of sampling 4 bags of M&M's

Sample bag 1

Sample size of the bag n = 50
number of blue M&M's = 13

$$\hat{p} \text{ of blue} = \frac{13}{50} = .26$$

Sample bag 2

Sample size of the bag n = 60
number of blue M&M's = 15

$$\hat{p} \text{ of blue} = \frac{15}{60} = .25$$

Sample bag 3

Sample size of the bag n = 31
number of blue M&M's = 7

$$\hat{p} \text{ of blue} = \frac{7}{31} \cong .23$$

Sample bag 4

Sample size of the bag n = 75
number of blue M&M's = 18

$$\hat{p} \text{ of blue} = \frac{18}{75} = .24$$

Every individual bag of M&M's is a single sample of the entire population of M&M's

Random mixing in the packaging process will result in each bag being a random sample of the entire population. Each individual bag of M&M's is a **different sample** of the entire population. Each of the examples above had a different value for the sample proportion $\hat{p}$ of blue M&M's. We would not expect to find the same sample proportion $\hat{p}$ of blue M&M's in each of the bags we sampled. All we can say at this point is that if the true population proportion p of blue M&M's is .24 as the Mars company reports then we would expect that **most of the sample proportions** $\hat{p}(\text{blue})$ would be **"close"** to the exact population proportion of $p(\text{blue}) = .24$

How can we determine the exact Population Proportion p?

The only way to find the exact value of a Population Proportion p is to count the entire population. Due to cost, population size, time, the changing nature of the population or other factors **it is often not possible to find the exact value of a Population Proportion p.**

If we cannot say what the exact population Proportion is p
can we just use a **Sample Proportion $\hat{p}$ and claim it is the** Population Proportion p **?**

While we cannot count the entire population and determine the exact value of the population proportion p, we can take **a sample of the population** and find the sample proportion $\hat{p}$. This Sample Proportion $\hat{p}$ **is based on only part of the population. You cannot expect a sample to produce the true Population Proportion p.** It is reasonable to say that the sample proportion $\hat{p}$ should be **"close"** to the true population proportion p.

Point Estimate $\hat{p}$

If we select just **one bag of M&M's** and find the sample proportion $\hat{p}$ of blue M&M's in the sample we would at least have one value for $\hat{p}$ of blue. This is not the exact value for the population proportion of blue but it is **the best single guess** about the exact value of the **Population Proportion p.** We call a **single sample proportion $\hat{p}$ the Point Estimate for the Population Proportion p.**

The Point Estimate $\hat{p}$
is the best single estimate of the
exact value of Population Proportion p

Can we do better than using a single point estimate $\hat{p}$
to estimate the exact value of Population Proportion p?

Yes we can. We know that $\hat{p}(\text{blue})$ **is not guaranteed to be equal to** $p(\text{blue})$ but it should be **"close"** to the true population proportion $p(\text{blue})$.

Confidence Interval for p.

We can refine the meaning of **"close"** by creating a number below $\hat{p}$, and a number above $\hat{p}$, to create a **range of values** $\text{number}_1 < p < \text{number}_2$ that we feel **somewhat confident** will contain the exact value of the **Population Proportion p. This range of values is a good estimate of the exact value of the Population Proportion p.**

$$\text{number}_1 < p < \text{number}_2$$

We call this range of probable values of p a **Confidence Interval for p.**

Confidence Interval for p.

$.20 < p < .28$ says that we feel **somewhat confident** that p lies within this interval

$.11 < p < .15$ says that we feel **somewhat confident** that p lies within this interval

What do we mean by **"somewhat confident"**

Confidence Level

What do we mean by **"somewhat confident".** We call the level of confidence we have that the exact population proportion p is contained in the Confidence Interval the **Confidence Level.** It is very common to require confidence levels of 90%, 95% or 99%. We will learn to create confidence intervals in the next section. This process will be based on **the value of the sample proportion** $\hat{p}$ and the **confidence level** required by the problem. A 95% confidence level says that the process we use will produce an interval that does contain the Population Proportion 95% of the time.

Confidence Interval Statements about the Population Proportion

There are two parts to a statement about the **true Population Proportion p.** The Confidence Level that is required by the problem and the sample proportion $\hat{p}$ from taking one sample. Using these values we will be able to make statements concerning the **true Population Proportion** p. Each confidence statement gives a range of values for the Population Proportion p and the level of confidence that you have that the interval does contain the Population Proportion.

Example 1: I am 99% confident that the true Population Proportion p of Blue M&M's **lies within the interval** $.20 < p < .28$

Example 2: I am 95% confident that the true Population Proportion p of Red M&M's **lies within the interval** $.11 < p < .15$

Example 3: I am 90% confident that the true Population Proportion p of Green M&M's **lies within the interval** $.10 < p < .20$

What can I do to get a more narrow Confidence Interval ?

$.10 < p < .20$ $.13 < p < .17$ $.14 < p < .16$

If you cannot find the true value for the population proportion p then it would seem that every effort should be made to get a very narrow range for the Confidence Interval for the Population Proportion.

This can be done in two ways.

1. Using a **larger sample size n** will produce a smaller range for the confidence interval if the confidence level stays the same. As the sample size gets larger (without replacement) the more the sample becomes like the population and there is less chance for a large difference between $\hat{p}$ and p.

2. Using a smaller **Confidence Level** will produce a smaller range for the confidence level if the confidence level stays the same.

The value for the Confidence Level effects the range of the Confidence Interval. The effect is not intuitive however. The larger the **Confidence Level the wider the interval for** p.

This may not seem true to you. Think of it as shooting a basketball from the free throw line. Lets say that you are 90% confident that you can make a free throw. You could you increase your confidence level to 95% if the opening of the rim were made wider. You could you increase your confidence level even higher to 99% if the opening of the rim were made extremely wide. In fact you could almost guarantee a made free throw if the rim were widened to the size of the Grand Canyon. (The canyon rim is almost 100 miles long). To become more confident in making the shot the rim needs to become wider. **To become more confident that the confidence interval contains the population proportion the confidence interval needs to become wider.**

The higher the confidence level the larger the size of the confidence interval

The lower the confidence level the smaller the size of the confidence interval

Larger Confidence Level <-----------------------------------> Smaller Confidence Level

Larger Range for p **<-----------------------------------> Smaller Range** p

$.10 < p < .20$	$.13 < p < .17$	$.14 < p < .16$
99%	**95%**	**90%**

This is the best we can do

The best way to know the value for p is to use the entire population. A sample cannot determine the exact value for p. The best that we can do in trying to determine what the true Population Proportion p without the entire population is to use a Sample Proportion $\hat{p}$. We use the sample proportion $\hat{p}$ as the starting point. We then create a range of values around $\hat{p}$ that we are somewhat confident will contain the true Population Proportion p. The width of the interval will be determined by the level of confidence you desire.

This process of creating a confidence interval for the population proportion with a given confidence level is the topic of the next section.

