Chapter 1

Point Estimators

Humility is to make a right estimate of one's self.

Charles Spurgeon

Statistical Inference

Why do we care about sample statistics so much? Since we want to use their values as *proxies* for corresponding characteristics of population. Generally, we can divide Statistics in two parts – **descriptive** statistics and **inferential** statistics. The first describes, presents and summarizes the observed data. The second is a special part of statistics theory that tries to reach conclusions extended beyond the data in place to make wider inferences.

Inference is a conclusion derived on the basis of some logical fact. In Statistics it usually refers to acceptance of some numerical assumptions about population based on the analysis of sample data. As we cannot reach the whole population the only way to learn its characteristics is the analysis of the available part of it, the sample. Then, an inference is done by *generalization* of sample results to the whole population. This explains why we pay so much attention to collecting a right sample and evaluating sampling errors. Researchers How well these steps are done influences the possibility to make correct conclusions about population.

Estimation

What does it mean to find a proxy? It means to find a numeric approximation for a population characteristic of interest. This process is called estimation. By characteristic of interest we mean some population parameter, like mean salary of financial specialists in Moscow. We cannot simply *calculate* the true mean, since the population data is not available, therefore we *estimate* the value of population parameter based on sample data. Estimation is a tool to defeat the problem of incomplete data.

Approximation is done by the use of estimators. **Estimator** of a population parameter is the formula or rule which produces a numerical value to approximate this parameter. An example can be a sample mean $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. It is an estimator for population mean μ and is a function of sample observations X_1, X_2, ..., X_n. An estimator aims at guessing the true value of the parameter of interest. A particular value produced by an estimator is called a point estimate.

Estimators versus population parameters

It is time now to learn how to distinguish between population parameters and estimators, which aim at approximating them. A population parameter is a value calculated from all the values in the population. It is always a *constant*, which is not available to us. An estimator is a random *variable* calculated on sample data to approximate that unknown constant. Examples of parameters and their corresponding estimators are provided in the table below. As you can see the right column with estimators actually $\bar{X}$, $\hat{p}$, s which are sample statistics already familiar to you.

Population parameters (θ)	Examples of estimators ($\hat{\theta}$)
Population mean (μ)	Sample mean ($\bar{X}$)
Population proportion (p)	Sample proportion ($\hat{p}$)
Population standard deviation (σ)	Sample standard deviation (s)

So, now you are going to look at the same things at a little bit different angle. You can think of population parameter and its estimator as of a target and a gun. Population parameter is the target we want to know. It is an unknown constant number. Estimator is the instrument to “hit” this target. A population parameter is denoted by θ and its estimator by $\hat{\theta}$. The letter θ is pronounced as “theta”. Hint: If you forgot how to draw this letter, recall that it looks like the old “Pepsi” sign.

Properties of Estimators

As any random variable, an estimator can be described by its probability distribution. It also has its expected value and variance. What makes up a good estimator? Strictly speaking there exist infinitely many different estimators for any population parameter. How to choose an estimator for a parameter? The answer is there are three characteristics that constitute a good estimator:

1. Unbiasedness
2. Efficiency
3. Consistency

Unbiasedness

So, an unknown constant number is our target. We said that estimators are instruments to hit this target. An estimator is said to be unbiased if, *on average*, it hits the target. Formally, an estimator is said to be unbiased if its *expectation* equals the *parameter*, namelly:

$$E(\hat{\theta}) = \theta$$

This expression is equivalent to $E(\hat{\theta}) - \theta = 0$, the difference between the true value of parameter and the mean of estimator is zero. If there exists a difference, then estimator is "not well tuned" and systematically provides error. Such an estimator is said to be **biased** and this difference is called bias.

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Bias is a tendency of an estimator to systematically over- or underestimate population parameter. It is like an "average error". The sign of bias indicates the direction of the average error. If bias of an estimator is equal to zero, consequently, an estimator is unbiased. Unbiased estimators, on average, produce the true value of parameter.

Consider the following example. Below are provided four different estimators for parameter $\theta = \mu$. Are they unbiased? If not, find their bias. Assume sample consists of 5 elements $\{X_1, X_2, X_3, X_4, X_5\}$.

- $\hat{\theta}_1 = \frac{X_2 + X_4}{2}$
- $\hat{\theta}_2 = X_3$
- $\hat{\theta}_3 = \frac{X_1 + X_3 + X_5}{5}$
- $\hat{\theta}_4 = \bar{X} = \frac{\sum_1^5 X_i}{5}$

To check whether the estimators are unbiased we have to find their means or, in other words, their *expectations*. You've studied the notion of expectation of a random variable in Chapter 2.

- $E(\hat{\theta}_1) = E(\frac{X_2+X_4}{2}) = \frac{1}{2}E(X_2) + \frac{1}{2}E(X_4) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$
- $E(\hat{\theta}_2) = E(X_3) = \mu$
- $E(\hat{\theta}_3) = E(\frac{X_1+X_3+X_5}{5}) = \frac{1}{5}E(X_1) + \frac{1}{5}E(X_3) + \frac{1}{5}E(X_5) = \frac{1}{5}\mu + \frac{1}{5}\mu + \frac{1}{5}\mu = \frac{3}{5}\mu$
- $E(\hat{\theta}_4) = E(\frac{\sum_1^5 X_i}{5}) = \frac{1}{5}E(X_1) + \ldots + \frac{1}{5}E(X_5) = 5\frac{1}{5}\mu = \mu$

All estimators are unbiased except for $\hat{\theta}_3$. Bias of $\hat{\theta}_3$ is equal to: $\text{Bias}(\hat{\theta}_3) = E(\hat{\theta}_3) - \theta = \frac{3}{5}\mu - \mu = -\frac{2}{5}\mu$ This estimator systematically underrepresents μ.

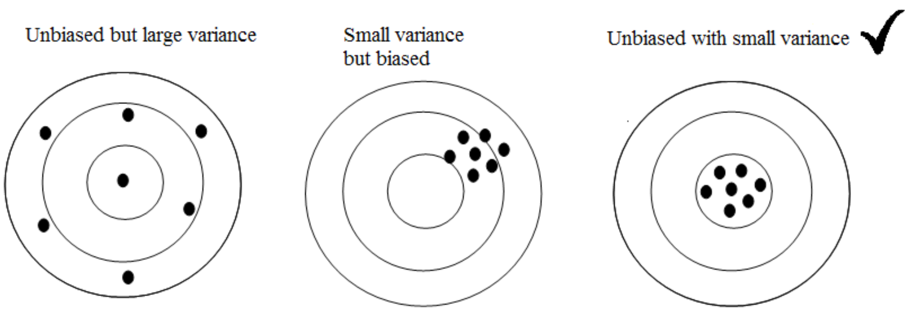
Efficiency
Efficiency of an estimator is measured by **mean squared error** (MSE):

$$MSE(\hat{\theta}) = \text{Bias}^2(\hat{\theta}) + Var(\hat{\theta})$$

It consists of systematic error, represented by bias, and random error, represented by variance. Recall that any random variable estimator may be characterized by its variance. Size of variance shows how big could be the dispersion of your estimates.

Bias and variance address two different questions. Bias shows how far $E(\hat{\theta})$ is from θ, and variance shows how far $\hat{\theta}$ could be from $E(\hat{\theta})$. For bias we can approximately predict the direction of the deviation from target while for variance the direction of deviation is random. They are both bad. In other words, bias is the average undershoot or overshoot of the target due to systematic error, and variance is the random mishoot due to natural deviation.

Bias is the backsight (прицел) of a gun and variance is the diameter of the backsight. Bias is the difference between the point you intended to shoot and the point at which you have actually shot. Variance is how much the shots deviate from each other while you're shooting at the same point (разброс попаданий).



Estimator with the least MSE is considered to be the most efficient.

Imagine that some politician says "I think that average salary in Russia is 75,000 rub". Such an estimator will have zero variance since it is a constant, but possibly has a large positive bias, since politician may try to overestimate the true figure.

Consider the same estimators as before. Which of the them is the most efficient? If an estimator is unbiased, its $MSE(\hat{\theta})$ equals $Var(\hat{\theta})$, since $bias(\hat{\theta})$ equals 0.

- $MSE(\hat{\theta}_1) = Var(\hat{\theta}_1) = Var(\frac{X_2+X_4}{2}) = \frac{1}{4}Var(X_2) + \frac{1}{4}Var(X_4) = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{1}{2}\sigma^2$

- $MSE(\hat{\theta}_2) = Var(\hat{\theta}_2) = Var(X_3) = \sigma^2$
- $MSE(\hat{\theta}_3) = \text{Bias}^2(\hat{\theta}_3) + Var(\hat{\theta}_3) = (-\frac{2}{5}\mu)^2 + Var(\frac{X_1+X_3+X_5}{5}) = (\frac{2}{5}\mu)^2 + \frac{1}{25}Var(X_1) + \frac{1}{25}Var(X_3) + \frac{1}{25}Var(X_5) = (\frac{2}{5}\mu)^2 + \frac{3}{25}\sigma^2$
- $MSE(\hat{\theta}_4) = Var(\frac{\sum_1^5 X_i}{5}) = \frac{1}{25}(Var(X_1) + \ldots + Var(X_5)) = \frac{1}{25}5\sigma^2 = \frac{1}{5}\sigma^2$

The most efficient estimator is $\hat{\theta}_4$.

Efficiency is not "yes or no" characteristic. It measures the degree of quality and allows to choose between several estimators.

Another way to compare estimators is relative efficiency defined as:

$$e(\hat{\theta}_1, \boldsymbol{\hat{\theta}_2}) = \frac{MSE(\boldsymbol{\hat{\theta}_2})}{MSE(\hat{\theta}_1)}$$

Please note that term 2 is in numerator!

We compare $e(\hat{\theta}_1, \hat{\theta}_2)$ with 1. If e is bigger than 1, it means that $MSE(\hat{\theta}_2)$ is higher than $MSE(\hat{\theta}_1)$, so $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$.

$e(\boldsymbol{\hat{\theta}_1}, \boldsymbol{\hat{\theta}_2}) = \frac{\sigma^2}{1/2\sigma^2} = 2 > 1$. Thus, $\boldsymbol{\hat{\theta}_1}$ is more efficient than $\hat{\theta}_2$.

Consistency

It is additional property not studied in this course although needed to know. Consistency means the distribution of estimates becomes more and more concentrated around the true value of the parameter being estimated.

Estimator is said to be consistent if: $\hat{\theta} \to \theta$ as $n \to \infty$ or, which is the same, $\lim\limits_{n\to\infty} MSE\left(\hat{\theta}\right) = 0$

Why $\bar{X}$, $\hat{p}$ and s?

So far we were using sample mean ($\bar{X}$), sample proportion ($\hat{p}$) and sample standard deviation (s) as proxies for μ, p and σ correspondingly. Why are they the best candidates? Some other estimators can also be suggested for these parameters. Let us look at them more precisely to answer why we always work with these particular estimators.

Estimating population mean (μ)
Population mean is a constant and is denoted by μ (e.g. average salary in a region). Consider an estimator $\hat{\theta} = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$:

$E(\frac{X_i}{n}) = \frac{1}{n}E(X_1 + X_2 + \ldots + X_n) = \frac{1}{n}(E(X_1) + E(X_2) + \ldots + E(X_n))$

Since each random variable X_i has the same distribution and mean μ, we know each expectation:
$\frac{1}{n}(E(X_1) + E(X_2) + \ldots + E(X_n)) = \frac{1}{n}(\mu + \mu + \ldots + \mu) = \frac{n\mu}{n} = \mu$

It is unbiased!

$Var(\bar{X}) = Var(\frac{X_i}{n}) = \frac{1}{n^2}Var(X_1 + X_2 + \ldots + X_n) = \frac{1}{n^2}(Var(X_1) + Var(X_2) + \ldots + Var(X_n)) = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \ldots + \sigma^2) = \frac{1}{n^2}n\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

It is the most efficient. We proved it in our example among 4 estimators and for all other estimators it holds too.

It is consistent as $MSE(\bar{X}) = Var(\bar{X}) = \frac{\sigma^2}{n} \to 0$ as n approaches infinity.

Sample mean $\bar{X}$ is a good estimator for population mean μ.

Estimating population proportion (p)
Population proportion is a constant and is denoted by p (e.g. the proportion of beer party supporters).

Consider an estimator $\hat{\theta} = \hat{p} = \frac{m}{n}$:

$E(\hat{p}) = E(\frac{m}{n}) = \frac{1}{n}E(m) = \frac{1}{n}np = \frac{np}{n} = p$

It is unbiased!

$Var(\hat{p}) = Var(\frac{m}{n}) = \frac{1}{n^2}Var(m) = \frac{1}{n^2}np(1-p) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$

It is proven to be the most efficient.

It is consistent as $MSE(\hat{p}) = Var(\hat{p}) = \frac{p(1-p)}{n} \to 0$ as n approaches infinity.

Sample proportion $\hat{p}$ is a good estimator for population proportion p.

Estimating population variance (σ^2)
Population standard deviation is a constant and is denoted by σ.

Consider an estimator $s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$:

We also showed in chapter 5 that there exists a more convenient version of the formula: $s^2 = \frac{\sum x^2 - n\bar{X}^2}{n-1}$

$E(s^2) = E(\frac{\sum x^2 - n\bar{X}^2}{n-1}) = \frac{1}{n-1}\left[E(\sum(x_i^2)) - E(n\bar{X}^2)\right] = \frac{1}{n-1}\left[\sum E(x_i^2) - nE(\bar{X}^2)\right]$

$Var(X) = E(X^2) - EX^2$ thus $Var(\bar{X}) = E(\bar{X}^2) - E\bar{X}^2$

So $E(\bar{X}^2) = Var(\bar{X}) + E\bar{X}^2 = \frac{\sigma^2}{n} + \mu^2$. $E(X^2) = Var(X) + E(X)^2 = \sigma^2 + \mu^2$

$E(s^2) = \frac{1}{n-1}\left[\sum(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)\right] = \frac{1}{n-1}\left[n(\sigma^2 + \mu^2) - \sigma^2 - n\mu^2\right] = \frac{1}{n-1}\left[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right] = \frac{1}{n-1}[(n-1)\sigma^2] = \sigma^2$.

It is unbiased!

This is why we need to divide by $n-1$ for estimator s^2 to be unbiased! If calculation was made on *population*, the formula would be $\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$, as was pointed out in Chapter 5. So, before choosing the formula check whether you work with the whole population or with the sample. In the first case you can calculate μ and the true population standard deviation σ using the respective formula. In the second case you can only calculate $\bar{X}$ and the sample standard deviation s as an estimator for σ. Then, use the formula with $(n-1)$ in the denominator.

So, yes, these estimators are the best candidates. We've just shown that $\bar{X}$, $\hat{p}$ and s are unbiased, efficient and consistent estimators for μ, p and σ. That is why they are so popular and frequently used.

In conclusion, estimators and statistics do not need to be different things though they are different concepts. Estimators are them because of your intent to estimate a quantity. Statistics are values calculated on samples. As it turns out some statistics are good estimators. There are some statistics which are not estimators, like test-statistics explained in Chapter 10. It is also possible to think that there exist some estimators which are not statistics, like constant numbers, though they are probably not at all good at estimating.

You MUST BE ABLE TO REPRODUCE even being drunk

- estimator is a random variable
- population parameter is a constant
- estimator approximates population parameter
- estimator is unbiased if $E(\hat{\theta}) = \theta$
- $MSE(\hat{\theta}) = \text{Bias}^2(\hat{\theta}) + Var(\hat{\theta})$
- estimator with the smallest MSE is the most efficient
- relative efficiency $e(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}$

AP Practice Problems with solutions

Problem 1. AP 2012 №6 Two students at a large high school, Peter and Rania, wanted to estimate m, the mean number of soft drinks that a student at their school consumes in a week. A complete roster of the names and genders for the 2,000 students at their school was available. Peter selected a simple random sample of 100 students. Rania, knowing that 60 percent of the students at the school are female, selected a simple random sample of 60 females and an independent simple random sample of 40 males. Both asked all of the students in their samples how many soft drinks they typically consume in a week.

Peter and Rania conducted their studies as described. Peter used the sample mean $\bar{X}$ as a point estimator for μ. Rania used $\bar{X}_{\text{overall}} = 0.6\bar{X}_{\text{female}} + 0.4\bar{X}_{\text{male}}$ as a point estimator for μ, where $\bar{X}_{\text{female}}$ is the mean of the sample of 60 females and $\bar{X}_{\text{male}}$ is the mean of the sample of 40 males.

Summary statistics for Peter's data are shown in the table below.

Variable	N	Mean	Standard Deviation
Number of Soft drinks	100	5.32	4.13

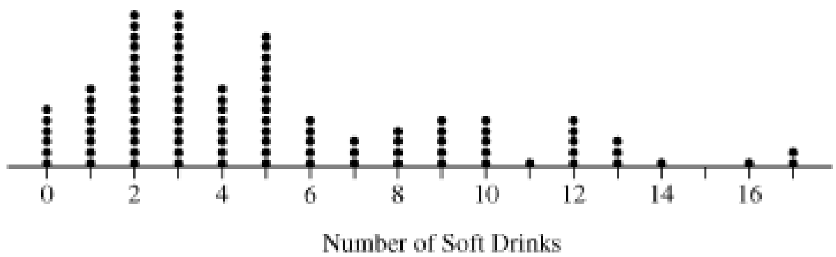
(b) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution (sometimes called the standard error) of Peter's point estimator $\bar{X}$.

Summary statistics for Rania's data are shown in the table below.

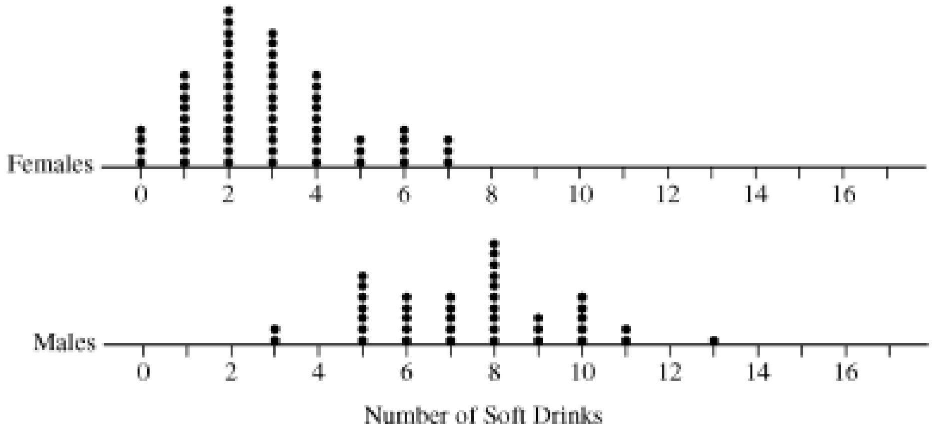
Variable	Gender	N	Mean	Standard Deviation
Number of Soft drinks	male	60	2.90	1.80
	female	40	7.45	2.22

(c) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution of Rania's point estimator $\bar{X}_{\text{overall}} = 0.6\bar{X}_{\text{female}} + 0.4\bar{X}_{\text{male}}$.

A dotplot of Peter's sample data is given below.



Comparative dotplots of Rania's sample data are given below.



(d) Using the dotplots above, explain why Rania's point estimator has a smaller estimated standard deviation than the estimated standard deviation of Peter's point estimator.

Solution

(b) The estimated standard deviation of the sampling distribution of the sample mean is: $\frac{s}{\sqrt{n}} = \frac{4.13}{\sqrt{100}} = \mathbf{0.413}$

(c) The variance of Rania's estimator is:
$Var(\bar{X}_{\text{overall}}) = 0.6^2 Var(\bar{X}_{\text{female}}) + 0.4^2 Var(\bar{X}_{\text{male}})$
$Var(\bar{X}_{\text{female}}) = \frac{\sigma_f^2}{n_f}$ $Var(\bar{X}_{\text{male}}) = \frac{\sigma_m^2}{n_m}$
Here we use the respective sample standard deviations s_f and s_m for the population parameters. Rania's estimate of standard deviation is calculated as:
$Var(\bar{X}_{\text{overall}}) = 0.6^2 \frac{s_f^2}{n_f} + 0.4^2 \frac{s_m^2}{n_m} = 0.6^2 \frac{1.80^2}{60} + 0.4^2 \frac{2.22^2}{40} = 0.01944 + 0.01972 = 0.03916$
$s_{\text{Rania}} = \sqrt{Var(\bar{X}_{\text{overall}})} = \sqrt{0.03916} = \mathbf{0.198}$

(d) The comparative dotplots from Rania's data reveal that the distribution of the number of soft drinks for females appears to be quite different from that of males. In particular, the centers of the distributions appear to be significantly different. Additionally, the variability of values around the center *within* gender in each of Rania's dotplots appears to be considerably smaller than the variability displayed in the dotplot of Peter's data. Rania's estimator takes advantage of the decreased variability within gender because her data were obtained by sampling the two genders separately. Peter's estimator has more variability because his data were obtained from a simple random sample of all the high school students.

Problem 2. AP 2008 №4
An experiment was conducted to study the effect of temperature on the reliability of an electronic device used in an undersea communications system. The experiment

was done in a laboratory where tanks of seawater were maintained at either. After the electronic devices were submerged in the tanks for 5,000 hours, each device was inspected to determine if it was still working. The following table provides information on the number of devices tested at each temperature and the number of working devices at the end of the 5,000-hour test.

Seawater Temperature	10°C	30°C	50°C	70°C
Number of working devices	29	42	21	12
Number of devices tested	30	50	30	20

You may assume that the result for any single device is not influenced by the result for any other device.

(c) An estimate of the proportion of devices that would work after 5,000 hours of submersion in 40°C seawater can be obtained by averaging the estimates at 30°C and 50°C. Compute this estimate and the associated standard error.

Solution

(c) $\hat{p}_{30} = \frac{42}{50}$, $\hat{p}_{50} = \frac{21}{30}$.
The estimated proportion is the linear combination $\hat{p}_{40} = \frac{1}{2} \cdot \hat{p}_{30} + \frac{1}{2} \cdot \hat{p}_{50} = \frac{1}{2}\frac{42}{50} + \frac{1}{2}\frac{21}{30} = \mathbf{0.77}$

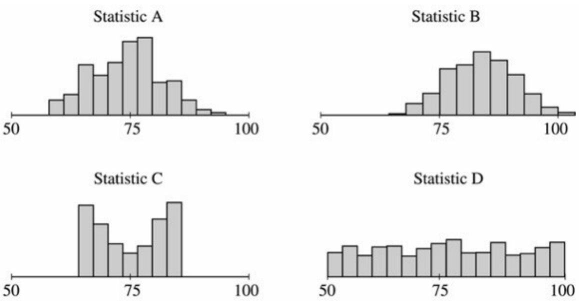
The associated standard error could be found by the following way. Recall $\hat{p} \sim N(p, \frac{p(1-p)}{n})$.

$Var(\hat{p}_{40}) = Var(\frac{1}{2} \cdot \hat{p}_{30} + \frac{1}{2} \cdot \hat{p}_{50}) = (\frac{1}{2})^2 Var(\hat{p}_{30}) + (\frac{1}{2})^2 Var(\hat{p}_{50}) = (\frac{1}{2})^2 \cdot (\frac{\frac{42}{50}\frac{8}{50}}{50}) + \frac{1}{2}^2 \cdot (\frac{\frac{21}{30}\frac{9}{30}}{30}) = (\frac{1}{2})^2(0.0027 + 0.007)$

$s = \sqrt{Var(\hat{p}_{40})} = \frac{1}{2}\sqrt{0.0027 + 0.007} = \frac{1}{2} \cdot 0.0984 = \mathbf{0.0492}$

Problem 3. AP 2008 Form B №2

Four different statistics have been proposed as estimators of a population parameter. To investigate the behavior of these estimators, 500 random samples are selected from a known population and each statistic is calculated for each sample. The true value of the population parameter is 75. The graphs below show the distribution of values for each statistic.



(a) Which of the statistics appear to be unbiased estimators of the population parameter? How can you tell?

(b) Which of statistics A or B would be a better estimator of the population parameter? Explain your choice.

(c) Which of statistics C or D would be a better estimator of the population parameter? Explain your choice.

Solution

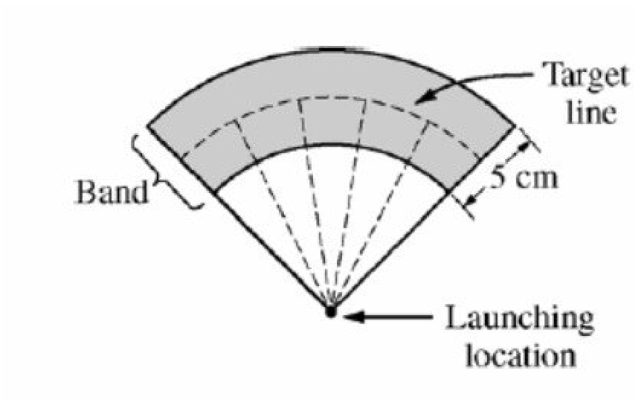
(a) Statistics A, C, and D appear to be unbiased. This is indicated by the fact that the mean of the estimated sampling distribution for each of these statistics is about 75, the value of the true population parameter.

(b) Statistic A would be a better choice because it appears to be unbiased. Although the variability of the two estimated sampling distributions is similar, statistic A would produce estimates that tend to be closer to the true population parameter value of 75 than would statistic B.

(c) Statistic C would be a better choice because it has smaller variability. Although both statistic C and statistic D appear to be unbiased, statistic C would produce estimates that tend to be closer to the true population parameter value of 75 than would statistic D.

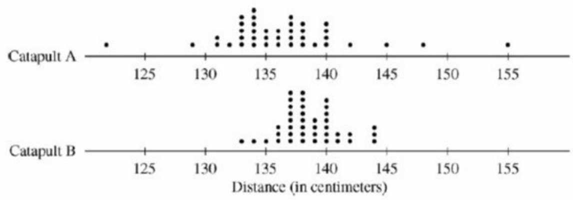
Problem 4. AP 2006 №1

Two parents have each built a toy catapult for use in game at an elementary school fair. To play the game the students will attempt to launch Ping-Pong balls from the catapults so that the balls land within a 5-centimeter band. A target line will be drawn through the middle of the band, as shown in the figure below. All points on the target lie are equidistant from the launching location.



If a ball lands within the shaded band, the student will win a prize.

The parents have constructed the two catapults according to slightly different plans. They want to test these catapults. Under identical conditions, the parents launch 40 Ping-Pong balls from each catapult and measure the distance that the ball travels before landing. Distances to the nearest centimeter are graphed in the dot plots below.



(b) If the parents want to maximize the probability of having the Ping-Pong balls land within the band, which one of the two catapults, A or B, would be better to use than the other? Justify your choice.

(c) Using the catapult that you chose in part (b), how many centimeters from the target line should this catapult be placed? Explain why you choose this distance.

Solution

(a) Topic Graphical Representation.

(b) It would be better to use catapult B because the distances vary much less around the center. If the catapult B is properly placed, the balls launched by it will have a higher probability of landing within the required target band.

(c) The catapult should be placed approximately 138 cm from the target line. The chosen distance resulted in a high proportion of the Ping-Pong balls in the sample landing within the target band, namely 2.5 cm around the target line. As could be seen from the dot plot this proportion is 30/40=0.75. By the way, 138 cm is the median of the distribution. Since the distribution is fairly symmetric and bell-shaped, the median is a good representation of center. Placing catapult at this location would result in a high proportion of balls from this sample landing in the target band.

A nice ICEF Problem

At the parliament election in some country the ruling “bear” party used its influence and power for election fraud. On each polling district it has carried out a throw-in of 100 fictitious bulletins for voting. Unfortunately (for “bear” party) some fictitious bulletins were filled incorrectly: from each 100 fake voices 30 by mistake have been given for “beer” party instead of “bear” party. Such bulletins have been recognized, so in fact only 70 bulletins from each 100 fake voices were given for ruling “bear” party.

Suppose that at some typical polling station with n registered voters the number of the voters for the “bears” people (without fraud) is m (where m has binomial distribution with probability of voting for “bears” p and total number of trials n). So in case of fair elections the estimator for the proportion of the votes for ‘bears’ is $\hat{p}_1 = \frac{m}{n}$. In case of unfair elections the estimator becomes $\hat{p}_2$.

(a) Define the $\hat{p}_2$.

(b) Are these estimators unbiased? Why or why not?

(c) (*) Are they consistent? Why or why not? Why at the real election the consistency would not help in case of unfair elections (in the other words, why at the real election consistency cannot make unfair elections at least a little "more fair")?

Solution

(a) The estimator for the proportion of "bears" party votes in case of *unfair* elections is the following:
$\hat{p}_2 = \frac{m+70}{n+100}$.

(b) We have $m \sim B(n, p)$, $E(m) = np$, $V(m) = np(1-p)$. For the "fair" estimator: $E(\hat{p}_1) = E(\frac{m}{n}) = \frac{1}{n}E(m) = \frac{np}{n} = p$. For the "fraud" estimator: $E(\hat{p}_2) = E(\frac{m+70}{n+100}) = \frac{1}{n+100}E(m+70) = \frac{np+70}{n+100} \neq p$. Hence, the estimator p_1 is unbiased while $\hat{p}_2$ is biased.

(c) (*) For the "fair" estimator by the law of large numbers we have $\hat{p}_1 = \frac{m}{n} \to \pi$ as $n \to +\infty$ For the 'fraud' estimator

$$\hat{p}_2 = \frac{m+70}{n+100} = \frac{\frac{m}{n} + \frac{70}{n}}{1 + \frac{100}{n}} \to \pi$$

as $\frac{70}{n} \to 0$ and $\frac{100}{n} \to 0$ as $n \to +\infty$. Hence both estimators are consistent. At the real elections the property of consistency would not help as for each polling station the number of registered voters is fixed and thus limited. It would help only while n approached infinity, thus only if the number of registered voters was unlimitedly infinite.