

# CS 237: Probability in Computing

Wayne Snyder  
Computer Science Department  
Boston University

---

## Lecture 15:

- Review: Confidence Intervals
- Sample Variance vs Population Variance
- Hypothesis Testing
- Two-Tailed Tests
- One-Tailed Tests, Upper and Lower

# Sampling When the Population Parameters are Unknown

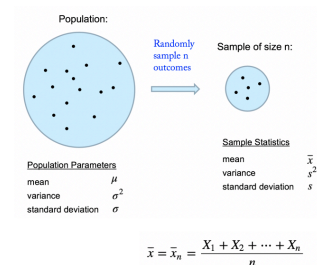
Remember that all this time we have made an **absurd assumption**, that we knew what the population standard deviation was:

## Sampling Theory When Population Parameters are Known

This is a warm-up to the real situation.....

Suppose (humor me!) that you have the actual height data about all BU students, including the mean and standard deviation, but then you LOSE all the data, but somehow you remember that the standard deviation is

$$\sigma = 3 \text{ inches.}$$



Furthermore, you need to know the **mean height**, but you don't have a lot of time, and in any case you only need an approximation (an **estimate**) of the true mean  $\mu$ .

What to do? **Sample 100 randomly-selected students (one trial) and use the sample mean as your estimate!** (Think, polling: you ask 100 random people who they voted for.)

When you report your result, you have an estimate, and you can use the CLT to give precise information about how accurate your estimate is. This is called a Confidence Interval...

This is universally done in statistics books to "warm-up" to the more realistic situation.

Now that we have applied the basic technique of sampling to confidence intervals it is time to take the training wheels off....

# Sampling When the Population Parameters are Unknown

When you don't know the standard deviation of the population, there are three cases where you can still proceed to use the CLT:



(1) When the population has a standard deviation which is **related to the mean by a formula** (e.g., all we studied except the Normal Distribution), **you can simply use the formula with the calculated mean of the sample.**

Example: (Proportions) Yes/No polls assume a Bernoulli population, so the standard deviation is:

$$s = \sqrt{\bar{x} \cdot (1.0 - \bar{x})}$$

(Bernoulli populations are called “Sampling with Proportions” – this is the most common case where we have a formula for the standard deviation.)

# Confidence Intervals: Sampling with Proportions

When the population is Bernoulli (Yes/No, 1/0, etc.)  
we can use the formula  $s = p(1-p)$  for the variance:

Then:

1. Choose a sample size  $n$ ;
2. Calculate the standard deviation of the sample mean;
3. Choose a confidence level CL (e.g., 95%);
4. Calculate the multiplier  $k$  for  $s$  corresponding to  $CL = P(\mu - k \cdot s_{\bar{x}} \leq \bar{x} \leq \mu + k \cdot s_{\bar{x}})$
5. Perform random sampling and calculate  $\bar{x}$  and  $s_{\bar{x}}$ ;
6. Report your results using the confidence interval corresponding to CL:

“The mean of the population is  $\bar{x} \pm k \cdot s_{\bar{x}}$  with a confidence of CL.”


$$\bar{x} = \frac{X_1 + \dots + X_n}{n}$$
$$s^2 = \bar{x} \cdot (1.0 - \bar{x})$$
$$s = \sqrt{s^2}$$
$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

```
In [3]: 1 norm.interval(alpha=0.95,loc=0,scale=1)
```

```
Out[3]: (-1.959963984540054, 1.959963984540054)
```

# Confidence Intervals for Proportions

Example -- Poll of BU Students: Should the MA ban on Vaping be continued?  
(1 = Yes and 0 = No)

1. Choose a sample size  $n = 50$ ;
2. Choose a confidence level  $CL = 90 \%$ ;
3. Calculate the multiplier  $k = 1.64$ ;
4. Perform random sampling of students: 
5. Calculate the percentage of sample who support the ban:  $\bar{x} = 0.4667$
6. Calculate the sample standard deviation and the standard deviation of the sample mean:  
$$s = \sqrt{\bar{x} * (1 - \bar{x})} = 0.4989; \quad s_{\bar{x}} = 0.4989 / \sqrt{50} = 0.0706$$
6. Report your results using the confidence interval corresponding to CL:

Sample =  
[ 1, 0, 1, 0, 0, 1, 0, 1, 0, 0,  
1, 0, 1, 0, 1, 0, 0, 1, 0, 0,  
0, 1, 1, 1, 0, 1, 1, 0, 0, 1... ]

“Of 30 BU students polled, 46.67 % +/- 11.61 % support a continued ban on vaping products, with a confidence of 90 %.”

# Sampling When the Population Parameters are Unknown

When you don't know the standard deviation of the population, there are three cases:

(1) (Formula) When the population has a standard deviation which is related to the mean by a formula (e.g., all we studied except the Normal Distribution), you can simply use the formula.

➔ (2) (Large Samples) When the population is large (typically,  $n > 30$ ), by the CLT the distribution of the sample mean is approximately normal, and we can use the sample standard deviation, with one small correction to the formula:

$$\bar{x} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
$$s^2 = \frac{(X_1 - \bar{x})^2 + (X_2 - \bar{x})^2 + \cdots + (X_n - \bar{x})^2}{n}$$
$$s = \sqrt{s^2}$$

We will  
change this

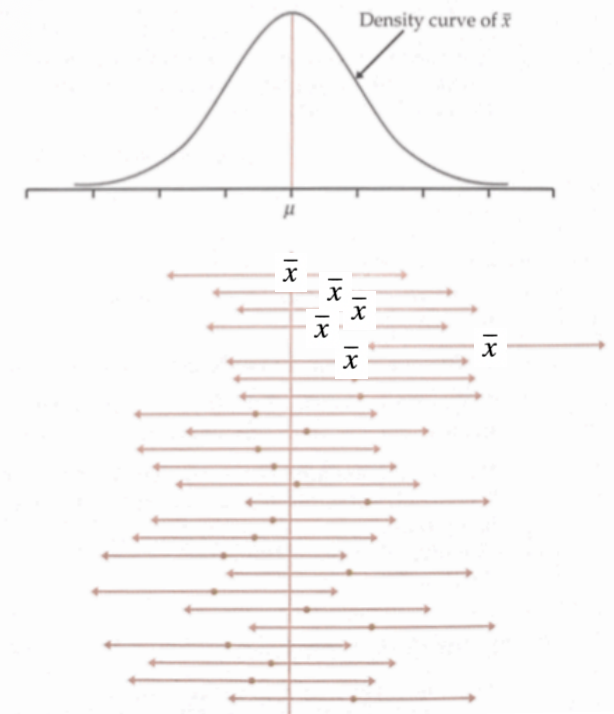
# Bessel's Correction for Sample Standard Deviation

The formula for the standard deviation has a bias: it **under-estimates** the true standard deviation when applied to samples, because it is an estimate (standard deviation **s** of sample) based on an estimate (mean  $\bar{x}$  of the sample):

$$\bar{x} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
$$s^2 = \frac{(X_1 - \bar{x})^2 + (X_2 - \bar{x})^2 + \cdots + (X_n - \bar{x})^2}{n}$$
$$s = \sqrt{s^2}$$

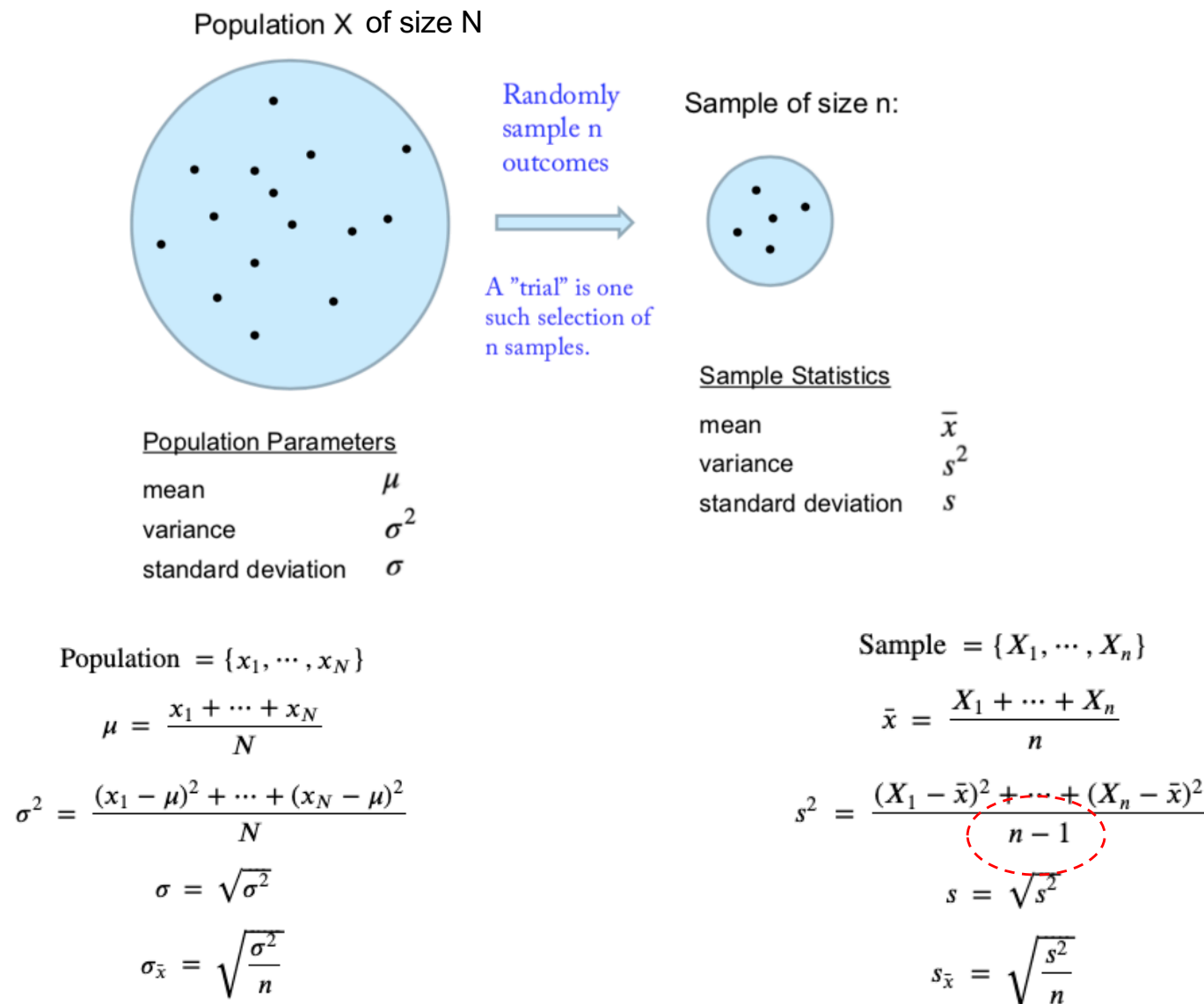
Intuition: Since the **sample** variance  $s^2$  is being measured against a random value  $\bar{x}$ , which varies as the sample changes, it is less than the **population** variance calculated from the mean  $\mu$ , which is a constant for the duration of the experiment.

Mathematically, it can be shown that by changing the denominator to  $n - 1$ , we eliminate the bias of the value calculated.



# Bessel's Correction for Standard Deviation

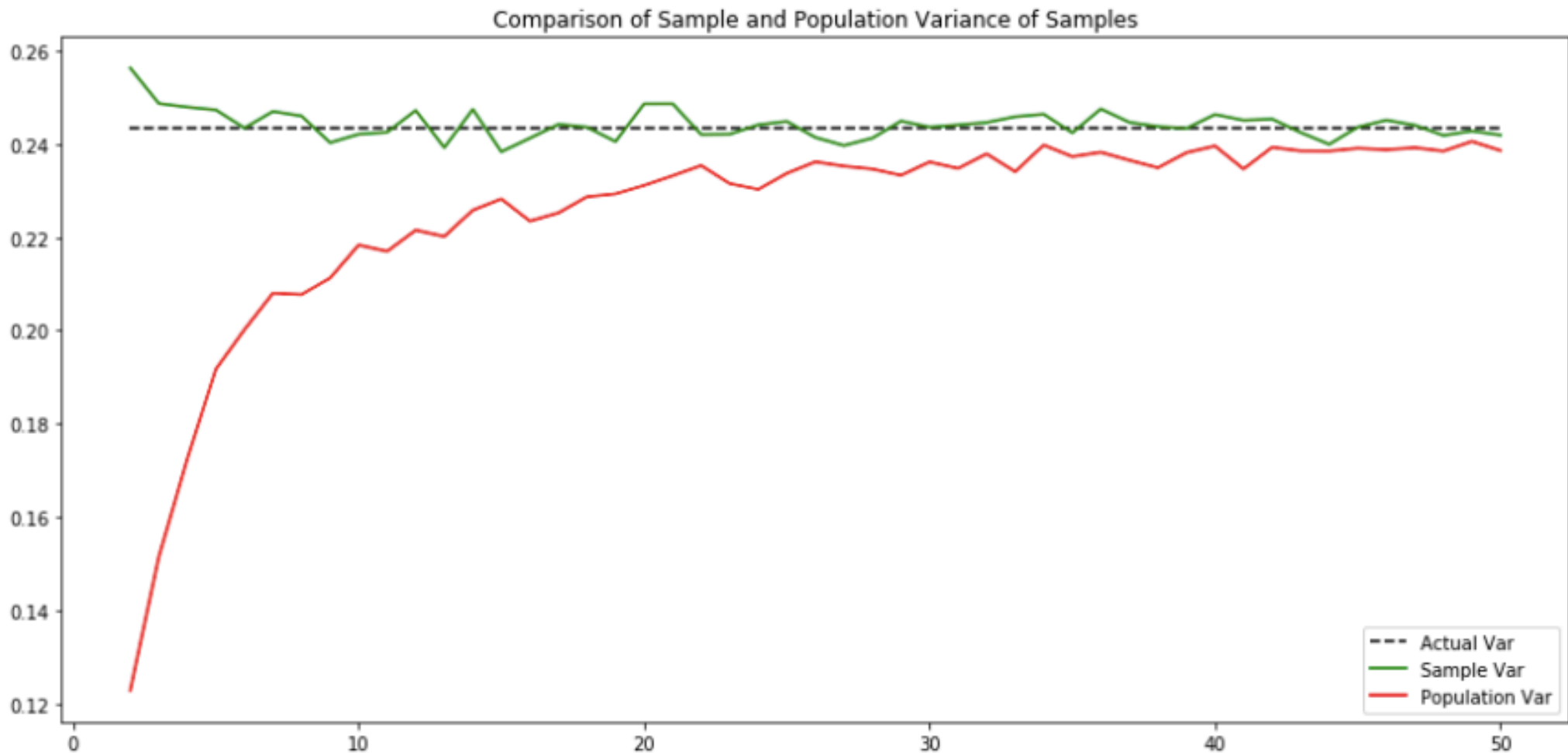
So, there are TWO different formula for the standard deviation:





# Bessel's Correction for Standard Deviation

As you will see in the last problem on the current homework, this improves the estimate!



# Confidence Intervals: Summary of the Procedure for Large Samples

## Confidence Intervals Using the Sample Standard Deviation when $n > 30$ .

We will use **s** as the standard deviation of the sample, calculated using Bessel's Correction (divide by  $n-1$ ):

Then:

1. Choose a sample size  $n$ ;
2. Choose a confidence level CL (e.g., 95 %);
3. Calculate the multiplier  $k$  corresponding to  $CL = P(\mu - k \cdot s_{\bar{x}} \leq \bar{x} \leq \mu + k \cdot s_{\bar{x}})$ ;
4. Perform random sampling and calculate  $\bar{x}$ , **s**, and  $s_{\bar{x}}$  ;
5. Report your results using the confidence interval corresponding to CL:

“The mean of the population is  $\bar{x} \pm k \cdot s_{\bar{x}}$  with a confidence of CL.”

```
In [3]: 1 norm.interval(alpha=0.95, loc=0, scale=1)
```

```
Out[3]: (-1.959963984540054, 1.959963984540054)
```

$$\text{Sample} = \{X_1, \dots, X_n\}$$

$$\bar{x} = \frac{X_1 + \dots + X_n}{n}$$

$$s^2 = \frac{(X_1 - \bar{x})^2 + \dots + (X_n - \bar{x})^2}{n - 1}$$

$$s = \sqrt{s^2}$$

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n}}$$

# Confidence Intervals Example

## Example -- Height of BU Students:

1. Choose a sample size  $n = 100$ ;
2. Choose a confidence level  $CL = 95.45\%$ ;
3. Calculate the multiplier  $k = 2$ ;
4. Perform random sampling of 100 students and calculate  $\bar{x} = 66.13$  and the sample standard deviation  $s = 3.45$  inches, and then

$$s_{\bar{x}} = \frac{3.45}{\sqrt{100}} = 0.345$$

5. Report your results using the confidence interval corresponding to CL:

“The mean height of BU students is 66.13 +/- 0.69 inches with a confidence of 95.45%.”

# Sampling When the Population Parameters are Unknown

When you don't know the standard deviation of the population, there are three cases:

- (1) (Formula) When the population has a standard deviation which is related to the mean by a formula (e.g., all we studied except the Normal Distribution), you can simply use the formula.
- (2) (Large Samples) When the population is large (typically,  $n > 30$ ), by the CLT the distribution of the sample mean is approximately normal, and we can use the sample standard deviation, with one small correction to the formula:
- ➔ (3) **Third**, when sampling with  $n \leq 30$  from a population known to be Normal, but with unknown mean and standard deviation, you use the sample standard deviation and a slightly different distribution, called the **T-Distribution**. (Not covered in CS 237.)

# Hypothesis Testing

**Hypothesis Testing** is a probabilistic version of a Refutation of a mathematical hypothesis, or a Proof by Contradiction.

Example of a **Refutation**:

**Hypothesis:** Any number with four occurrences of the digit 1, two occurrences of 4, two occurrences of 8, and no occurrences of 2 or 6, is a prime number.

**Refutation:** Nope!  $1,197,404,531,881 = 1,299,827 * 921,203$

Example of **Proof by Contradiction**:

**Theorem:** For all integers  $n$ , if  $n^2$  is odd, then  $n$  is odd.

**Proof:** Suppose we assume the negation of the theorem:

**Hypothesis:**  $\exists n$  such that  $n^2$  is odd and  $n$  is even.

Nope! Because then  $\exists k$  such that  $n = 2k$  and so  $n^2 = (2k)^2 = 4(k)^2$  and hence  $n^2$  is divisible by 2 and even. Therefore, the hypothesis is false, and the theorem (the inverse of the hypothesis) must be true. Q.E.D.

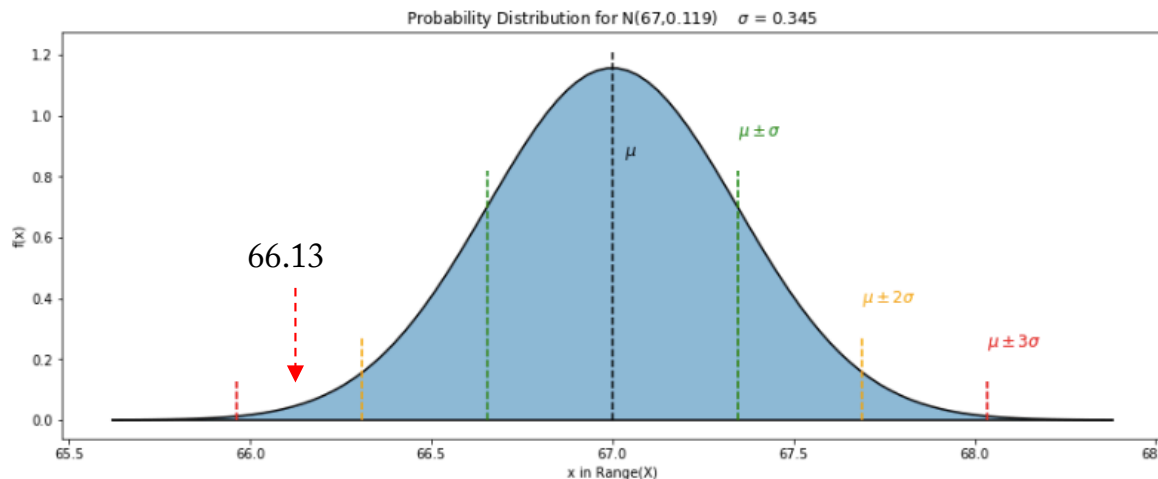
# Hypothesis Testing

When we test a hypothesis probabilistically, instead of absolutely refuting it, we show that the hypothesis is extremely unlikely given the result of our sampling experiment.

Now suppose you had a previous hypothesis about the heights of BU students:

**Hypothesis:** BU students have a mean height of 67 inches.

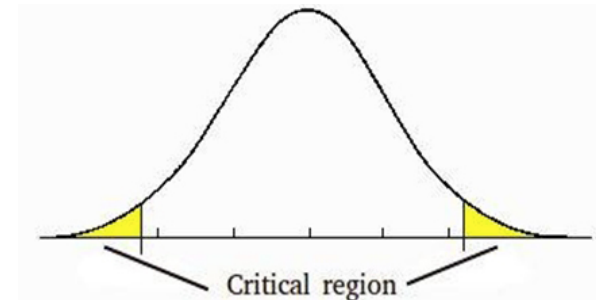
Now we estimated the standard deviation of the population as  $s = 3.45$  inches, and when we do our sample mean with  $n = 100$  students, our hypothesis implies that this sample mean should have the following distribution:



But our experiment gives a value of 66.13, which is unlikely! So our hypothesis is very likely to be wrong, and we should reject it. **But how to decide? How unlikely is this?**

# Hypothesis Testing: Two-Tailed Tests

When the extreme values could be in either direction (low or high): your hypothesis could be rejected because it is too low, OR because it is too high.



- BU students have a mean height of 68
  - Sam Adams Boston Lager contains 4.75% alcohol

In this case, you state a **Null Hypothesis** about the mean of a population **X**:

$$H_0 = \mu_X = k. \quad \leftarrow \text{This is the hypothesis to reject or not.}$$

And you state (or leave implicit) the **Alternative Hypothesis**:

$$H_1 = \mu_X < k \text{ or } k < \mu_X \quad \text{or, more simply,} \quad H_1 = \mu_X \neq k.$$

You Reject  $H_0$  if your sample mean is much larger or much smaller than  $k$  :

$$\bar{x} \ll k \quad \text{or} \quad \bar{x} \gg k$$

# Hypothesis Testing: Two-Tailed Tests

## Hypothesis Two-Sided Test:

**Step One:** State a **Null Hypothesis** making a claim about the mean of a population **X**:

$$H_0 = \mu_X = k. \quad (\text{and } H_1 = \mu_X \neq k.)$$

You will either **Reject** this hypothesis or do nothing (**Fail to Reject**).

**Step Two.** Determine how willing you are to be wrong, i.e., define the **Level of Significance**  $\alpha$  of the test:

$\alpha$  = probability you are wrong if you Reject  $H_0$  when it is actually correct.

## Example:

1.  $H_0$ : BU students have a mean height of 67 inches ( $k = 67$ ).
2.  $\alpha = 0.01$  (I am willing to be wrong 1% of the time)



# Hypothesis Testing: Two-Tailed Tests

## Hypothesis Two-Sided Test:

**Step Three.** Do the sampling experiment to find a sample mean  $\bar{x}$  and the standard deviation of the sampling distribution  $s$ .

### Example:

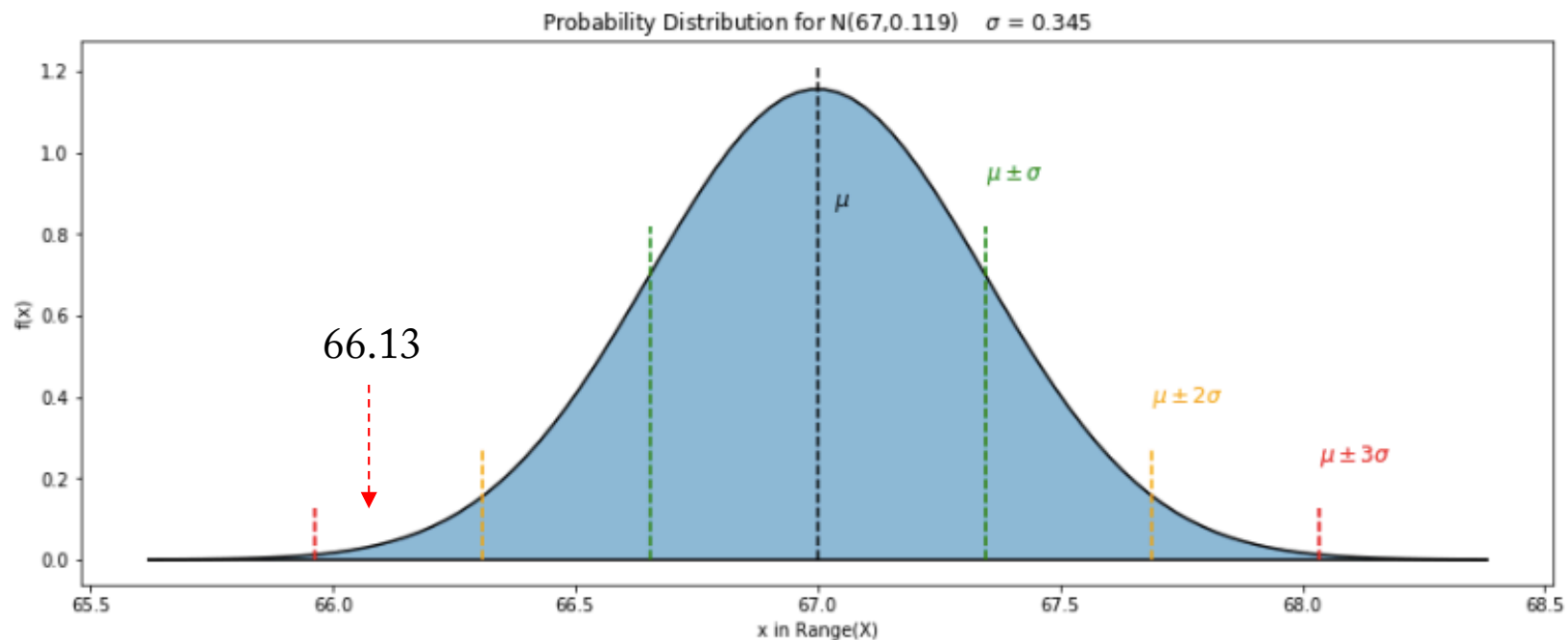
3. We perform the sampling experiment for  $n = 100$ , and find  $\bar{x} = 66.13$  and  $s = 3.45$ .

# Hypothesis Testing: Two-Tailed Tests

## Hypothesis Two-Sided Test:

Now, at this point, using the hypothesis that the mean should be 67 inches, and the fact that the standard deviation of the sampling distribution is  $s = 0.345$ , according to the hypothesis, we **should** have a sampling distribution of

$$\bar{X} = N(67, 0.345^2)$$



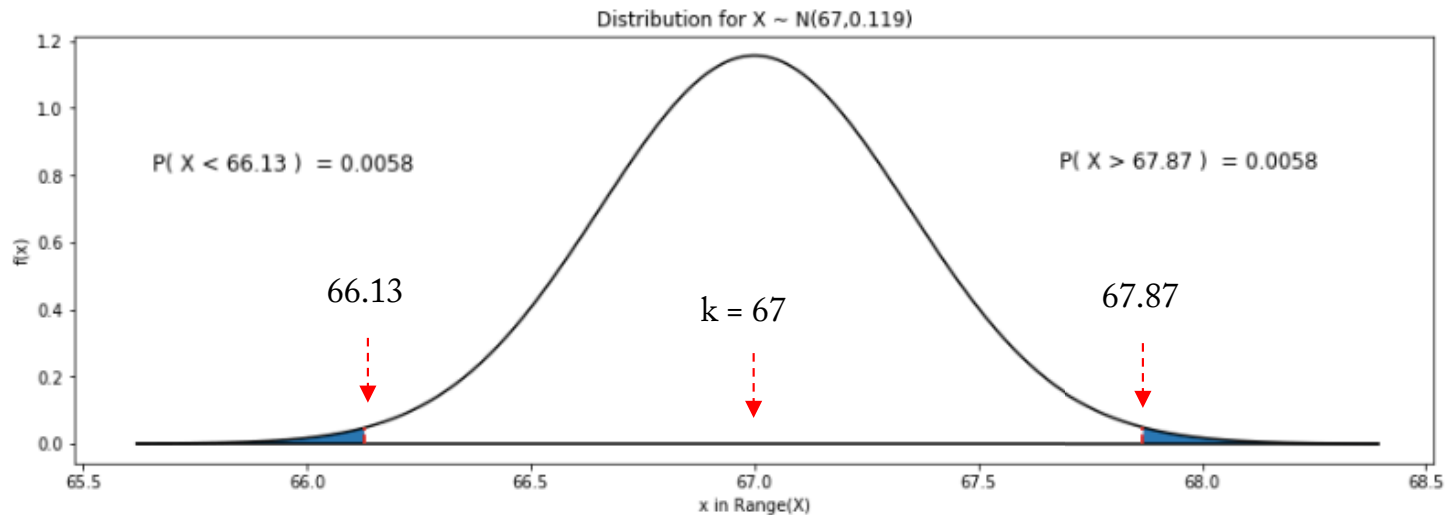
The question is, of course, how likely our actual value of 66.13 is under this assumption!

# Hypothesis Testing: Two-Tailed Tests

## Hypothesis Two-Sided Test:

**Step Four:** Calculate the **p-value** of the sample mean  $\bar{x}$ , the probability that the random variable  $\bar{X}$  would be farther away from  $k$  (our hypothesis value for the mean) than  $\bar{x}$  is:

$$P(|\bar{X} - k| > |\bar{x} - k|)$$



The p-value is the probability of seeing the value  $\bar{x}$  or a value even more unlikely, if  $H_0$  were true. Because we have a two-tailed test, we have to calculate how far  $\bar{x}$  is from the hypothesized value  $k$  and multiply by 2:

$$2 * P(\bar{X} < \bar{x}) \text{ if } \bar{x} < k$$

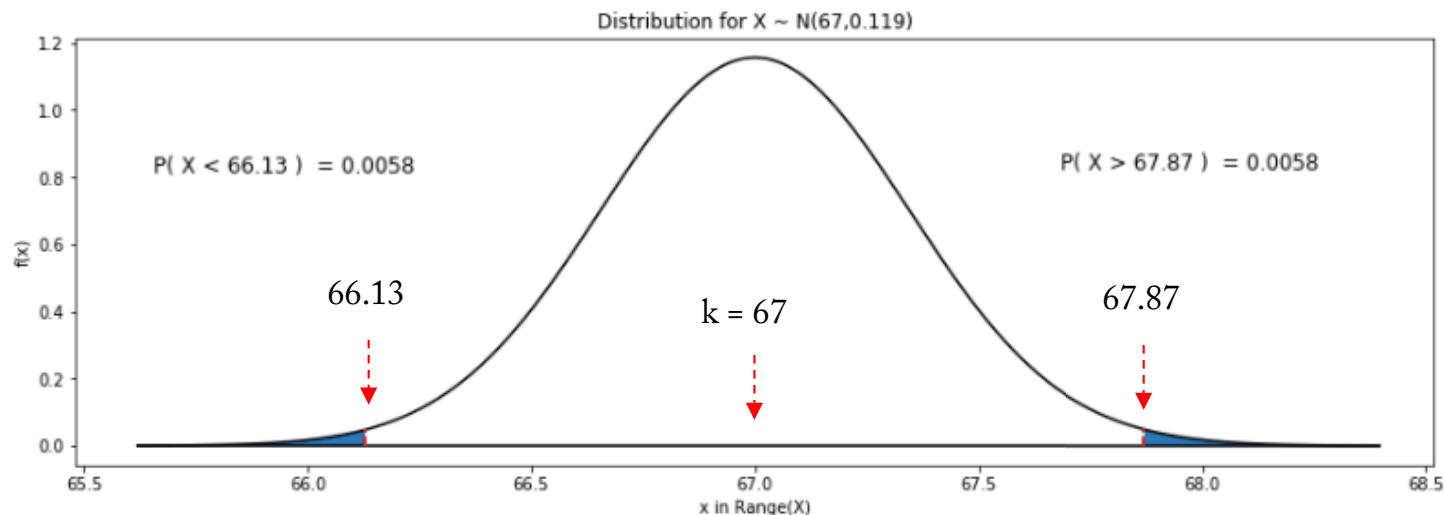
$$2 * P(\bar{X} > \bar{x}) \text{ if } \bar{x} > k$$

# Hypothesis Testing: Two-Tailed Tests

## Hypothesis Two-Sided Test:

**Step Four:** Calculate the **p-value** of the sample mean  $\bar{x}$ , the probability that the random variable  $\bar{X}$  would be farther away from  $k$  (our hypothesis value for the mean) than  $\bar{x}$  is:

$$P(|\bar{X} - k| > |\bar{x} - k|)$$



**Example:** Since  $66.123 < 67$ , we calculate the **p-value** = 0.0117 from the left side:

```
In [31]: 1 2 * norm.cdf(x=66.13, loc=67, scale=0.345)
```

```
Out[31]: 0.01167762737326203
```

# Hypothesis Testing: Two-Tailed Tests

## Hypothesis Two-Sided Test:

**Step Five.** If the p-value  $< \alpha$ , Reject, otherwise Fail to Reject.

Example: Clearly we must **Fail to Reject**, since  $0.0117 > 0.01$  ! We can not reject the hypothesis on the basis of the data!

## Some things to notice:

- (1) If we had set the level of significance at 95%, we would have Rejected! It is important, therefore, to set your parameters **before** doing the test!
- (2) This is precisely the same thing as if we asked “Is 67 inside the 99% confidence interval for our result?” using techniques from last lecture.

# Hypothesis Testing: One-Tailed Tests

When the extreme values are considered in one direction only, you have either an **Upper One-Tailed Test** or a **Lower One-Tailed Test**:

Example of hypothesis for an **Upper One-Tailed Test**:

- I claim Trevor does not have ESP: his chance of guessing the color of a card I hold hidden from him is 0.5 (if he does *much* better I'll reject my hypothesis!)

Example of hypothesis for a **Lower One-Tailed Test**:

- Seagate claims its disk drives last an average of 10,000 hours before failing (if we find the mean is *much* lower we may reject their claim).

# Hypothesis Testing: One-Tailed Tests

**One-Tailed:** When the extreme values are considered in one direction only, you have either an **Upper One-Tailed Test** or a **Lower One-Tailed Test**:

In these cases, you again state a **Null Hypothesis** about the mean of a population **X**:

$H_0 = “\mu_X = k.”$   This is the hypothesis to reject or not.

And you state (or leave implicit) the **Alternative Hypothesis**:

**For Lower:**  $H_1 = “\mu_X < k”$

**For Upper:**  $H_1 = “k < \mu_X”$

You Reject  $H_0$  if your sample mean is very different than  $k$ :

**For Lower:**  $\bar{x} \ll k$

**For Upper:**  $k \ll \bar{x}$

[ The main difference here is that you don't multiply by 2 when calculating the p-value. ]

# Hypothesis Testing: One-Tailed Tests

## Hypothesis Upper One-Tailed Test:

1. State a **Null Hypothesis** which makes a claim about the mean of a population **X**:

$$H_0 = “\mu_X = k.” \quad (\text{and } H_1 = “k < \mu_X”)$$

You will either **Reject** this hypothesis or do nothing (**Fail to Reject**).

2. Determine how willing you are to be wrong, i.e., define the **Level of Significance  $\alpha$**  of the test:  $\alpha$  = probability you are wrong if you Reject  $H_0$  when it is actually correct.

3. Determine a sample size **n**, take a random sample of size **n**, and determine the sample mean  $\bar{x}$ . Establish the standard deviation, either using the (known) population standard deviation or the sample standard deviation (more on this later).

4. Calculate the **p-value** of the mean  $\bar{x}$ , the probability that the random variable **X** would be larger than **k** :  $P( X > \bar{x} )$  The p-value represents the probability of seeing the value  $\bar{x}$  or a value even more unlikely (i.e., larger), if  $H_0$  were true.

4. If the **p-value**  $< \alpha$ , Reject, otherwise Fail to Reject.



# Hypothesis Testing: One-Tailed Tests

## Hypothesis **Lower** One-Tailed Test:

1. State a **Null Hypothesis** which makes a claim about the mean of a population **X**:

$$H_0 = “\mu_X = k.” \quad (\text{and } H_1 = “\mu_X < k”)$$

You will either **Reject** this hypothesis or do nothing (**Fail to Reject**).

2. Determine how willing you are to be wrong, i.e., define the **Level of Significance  $\alpha$**  of the test

$\alpha$  = probability you are wrong if you Reject  $H_0$  when it is actually correct.

3. Determine a sample size  $n$ , take a random sample of size  $n$ , and determine the sample mean  $\bar{x}$ . Establish the standard deviation, either using the (known) population standard deviation or the sample standard deviation (more on this later).

4. Calculate the **p-value** of the mean  $\bar{x}$ , the probability that the random variable  $X$  would be **smaller** than  $\bar{x}$ :  $P(X < \bar{x})$ . The p-value represents the probability of seeing the value  $\bar{x}$  or a value even more unlikely (i.e., even smaller), if  $H_0$  were true.

4. If the **p-value**  $< \alpha$ , Reject, otherwise Fail to Reject.

# Hypothesis Testing: One-Tailed Tests

## Example: Upper One-Tailed Test:

Trevor claims that he has ESP. I disagree. My hypothesis is that Trevor does not have ESP. The question is whether he can guess correctly much **more** than half the time, so this is an upper one-tailed test.

To test, I draw 100 cards from a deck (with replacement) and he guesses the color. The level of significance will be 5%.

$H_0$  = “Trevor's average number of correct cards is 50, because he is randomly guessing.”

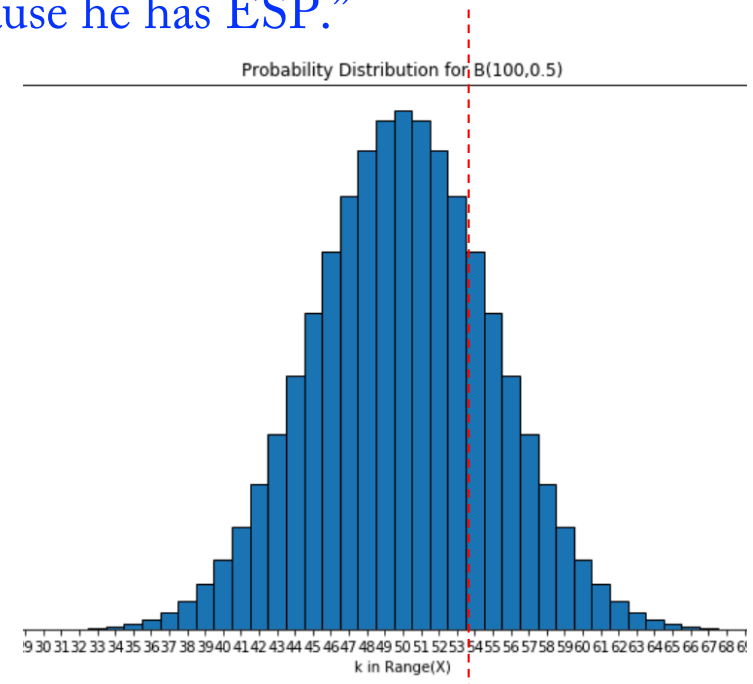
$H_1$  = “Trevor will guess many more than 50 correct, because he has ESP.”

In the experiment, he gets 54 cards correct.

Note that the best model of this experiment is a **Binomial** experiment, not **Normal**. Since this is an upper one-tailed test, the p-value is

$$P(X \geq 54) = \sum_{i=54}^{100} \binom{100}{i} (0.5)^i (0.5)^{100-i} = 0.2431.$$

Since  $0.2431 > 0.05$ , we fail to reject  $H_0$ .

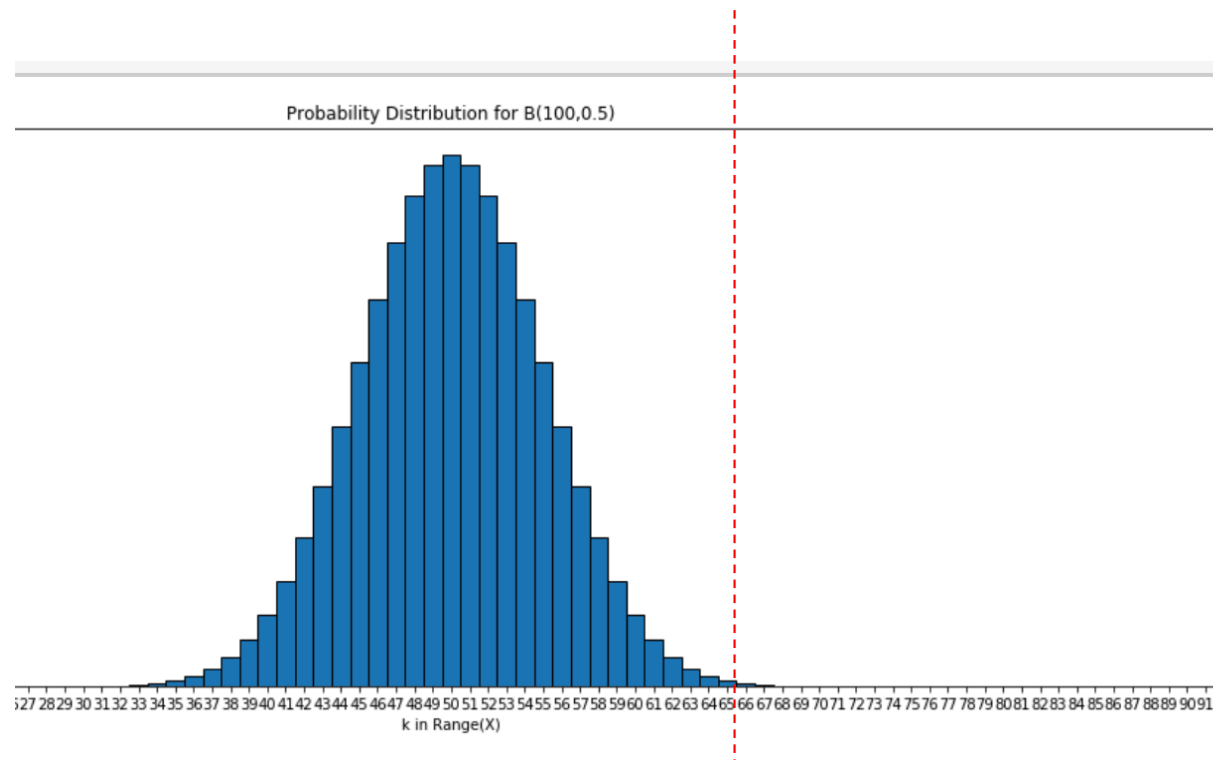


# Hypothesis Testing: One-Tailed Tests

But what if he had guessed 68 of them correctly?

$$P(X \geq 68) = 0.0002044$$

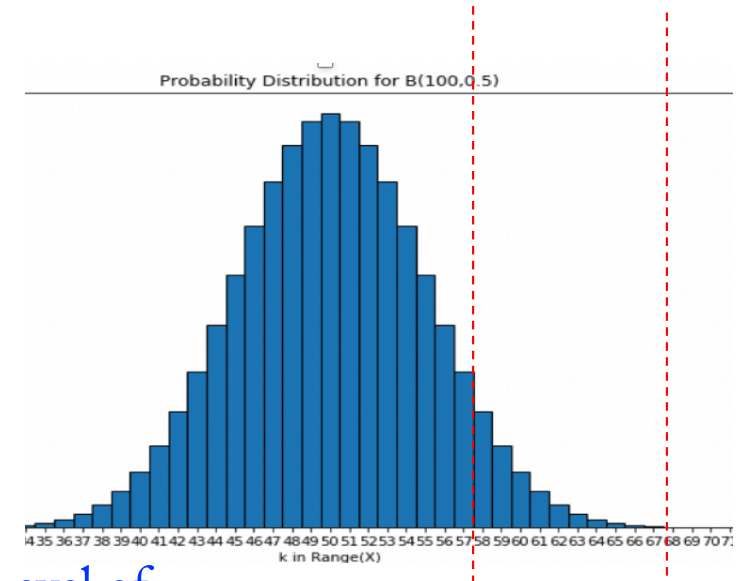
Since  $0.0002 < 0.05$ , we Reject my hypothesis that Trevor does not have ESP, because he did something very, very unlikely!



# Hypothesis Testing: One-Tailed Tests

Here is a table of how probable it is that Trevor guessed  $\geq k$  cards correctly, if in fact he were simply guessing with probability 0.5 of success; these the “p-values” of the outcome of the test:

xbar = 50:	0.460205381306
xbar = 51:	0.382176717201
xbar = 52:	0.308649706795
xbar = 53:	0.242059206804
xbar = 54:	0.184100808663
xbar = 55:	0.135626512037
xbar = 56:	0.0966739522478
xbar = 57:	0.0666053096036
xbar = 58:	0.044313040057
xbar = 59:	0.0284439668205
xbar = 60:	0.0176001001089
xbar = 61:	0.0104893678389
xbar = 62:	0.00601648786268
xbar = 63:	0.00331856025796
xbar = 64:	0.00175882086149
xbar = 65:	0.000894965195743
xbar = 66:	0.000436859918456
xbar = 67:	0.000204388583713
xbar = 68:	9.15716124412e-05
xbar = 69:	3.9250698228e-05
xbar = 70:	1.60800076479e-05



Reject at 5% Level of Significance

Reject at 1% Level of Significant