CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 19:

- Review: Confidence Intervals
- Hypothesis Testing
- Two-Tailed Tests
- One-Tailed Tests, Upper and Lower
- [If Time] Introduction to the Exponential Distribution

Sampling When the Population Parameters are Unknown

This is the usual case, and there are various solutions (we are only discussing the first two):

(1) (**Special Distributions**) 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 use the "sample standard deviation," with n-1 in denominator instead of n.

(3) (Small samples from normal population), when sampling with n <= 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.)

(4) (**Small samples not from normal**). There are various special "non-parametric" methods for small samples not known to be normal. (Not covered in CS 237.)

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):

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

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

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

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

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

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)
```

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%."

Hypothesis Testing

Hypothesis Testing is a probabalistic version of a Refutation by Counter Example of a mathematical hypothesis, or a Proof by Contradiction.

Example of Refutation by Counter-Example:

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.

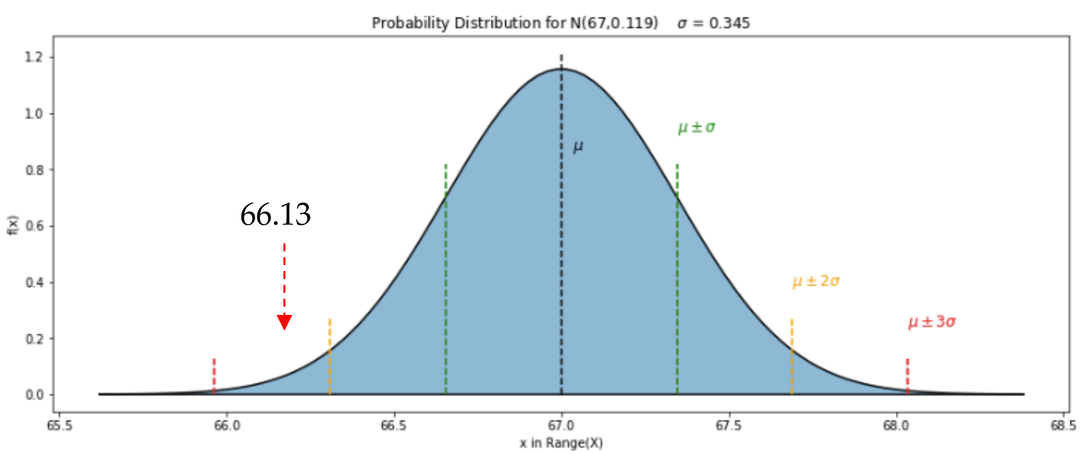
Nope! Because then $\exists$ k such that $n = 2k$ and so $n^2 = (2k)^2 = 4(k)^2$ and hence n^2 is divisible by 2 and even. Therefore, the hypothesis is false, and the theorem (the inverse of the hypothesis) must be true. Q.E.D.

Hypothesis Testing

When we refute a hypothesis probabalistically, instead of showing that is is impossible, we show that the hypothesis is extremely unlikely given the result of our sampling experiment. Here's an example:

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

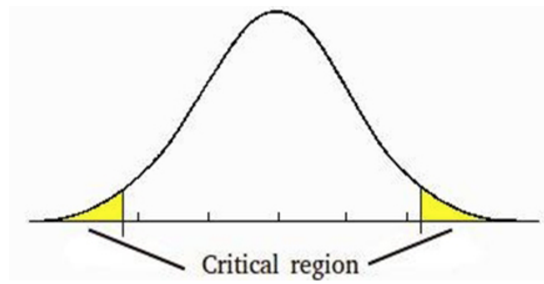
Now we do our experiment, with n = 100, and we find a sample mean of 66.13 inches and a sample standard deviation of 3.45 inches, and so our hypothesis implies that this sample mean should have the following distribution:



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-Sided ("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**:

$\mathbf{H}_0$ = "$\mu_X = k$." This is the hypothesis to reject or not.

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

$\mathbf{H}_1$ = "$\mu_X < k \text{ or } k < \mu_X$" or, more simply, $\mathbf{H}_1$ = "$\mu_X \neq k$."

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-Sided ("Two-Tailed") Tests

Hypothesis Two-Sided Test:

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

$\mathbf{H_0}$ = "$\mu_X = k$." (and $\mathbf{H_1}$ = "$\mu_X \neq k$.")

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** $\boldsymbol{\alpha}$ of the test:

α = probability you are wrong if you Reject $\mathbf{H_0}$ when it is actually correct.

Example:

1. $\mathbf{H_0}$**:** 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:

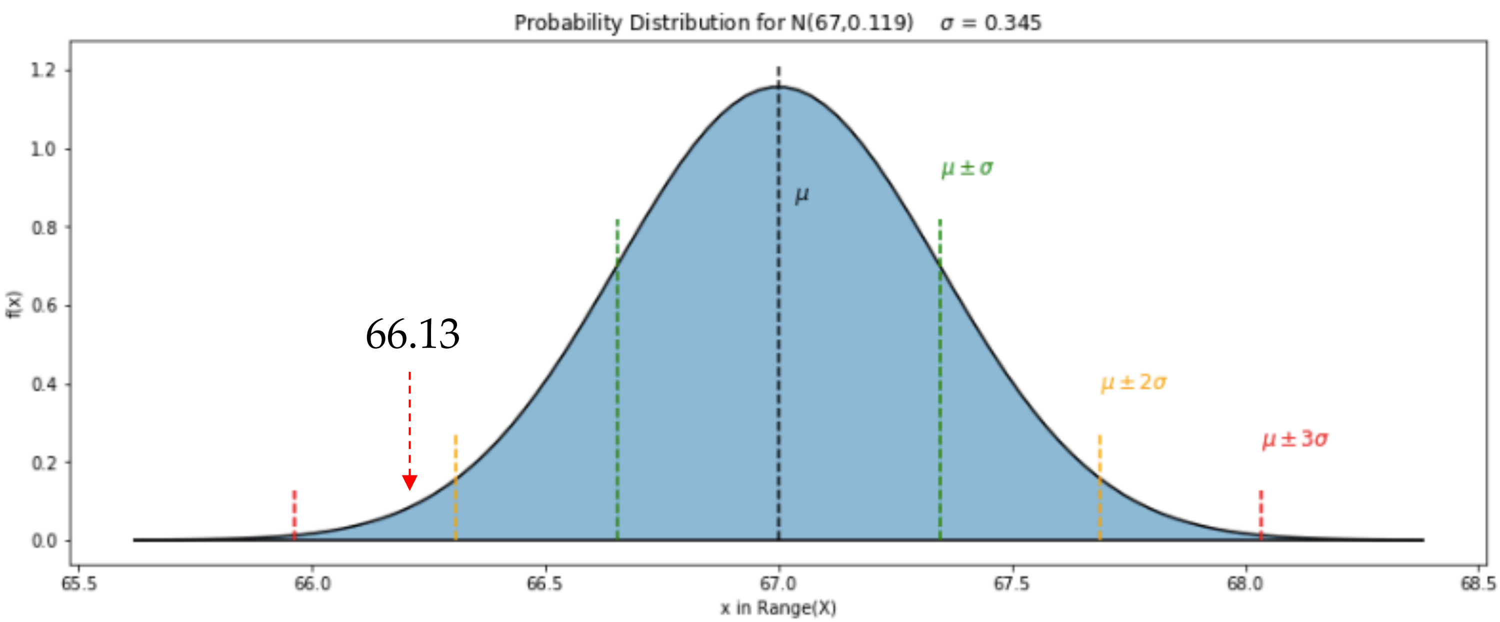
3. We perform the sampling experiment for n = 100, and find: $\bar{x}$ = 66.13 and s = 3.45.

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

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

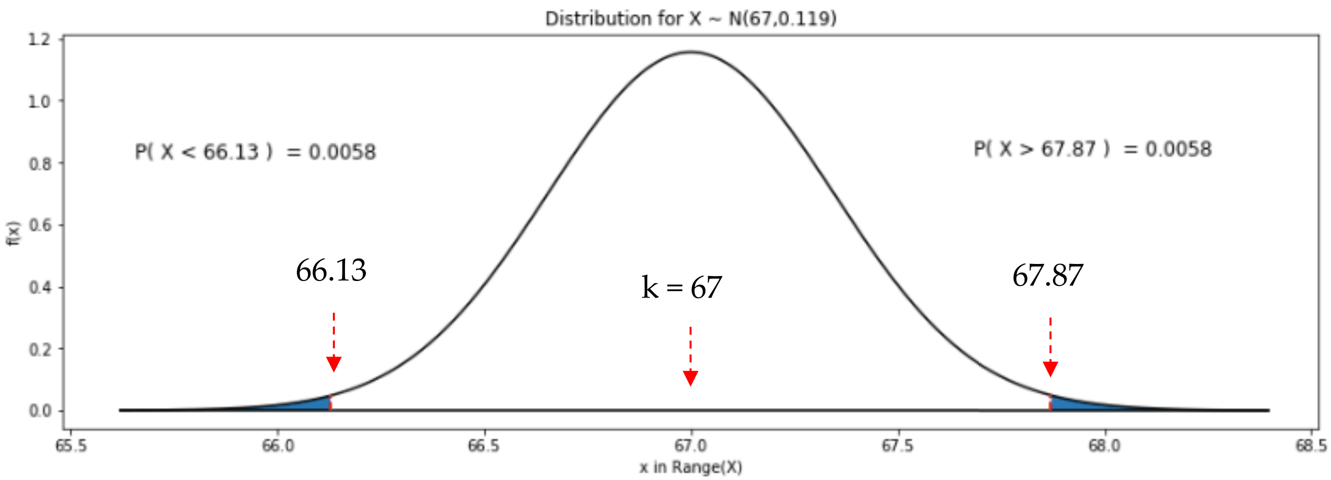


The question is, of course, how likely our actual value of 66.13 is under this assumption!

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:

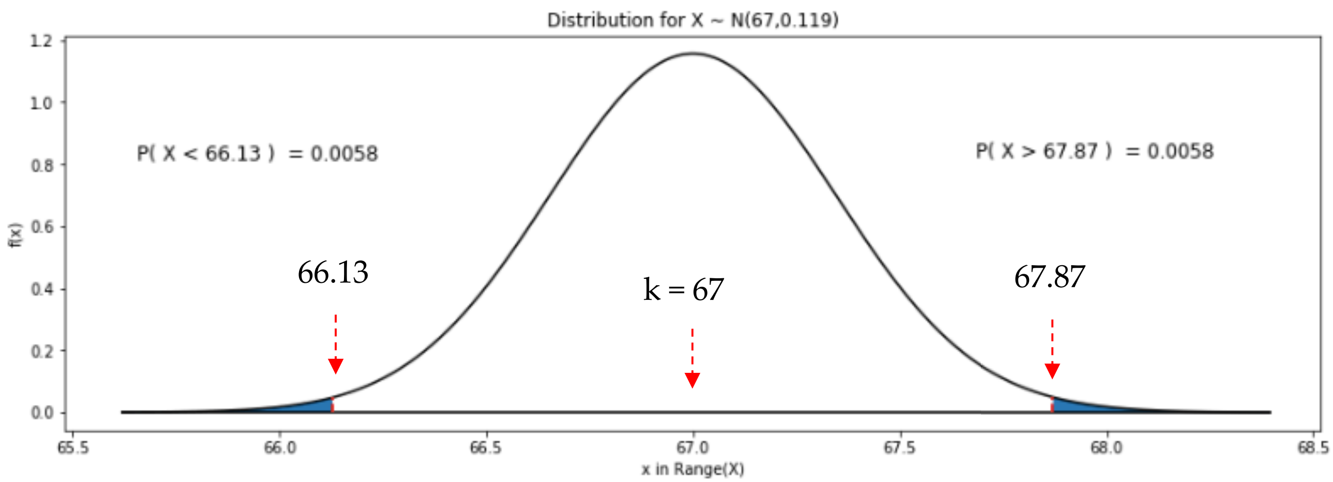
$$2 * P(\bar{X} < \bar{x}) \text{ if } \bar{x} < k \qquad 2 * P(\bar{X} > \bar{x}) \text{ if } \bar{x} > k$$

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]:  2 * norm.cdf(x=66.13,loc=67,scale=0.345)
Out[31]:  0.011677627373262034
```

Hypothesis Testing: Two-Tailed Tests

Hypothesis Two-Sided Test:

Step Five. If the **p-value < α** , 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 Richard 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**:

$\mathbf{H_0} = $ "$\mu_X = k.$" ← This is the hypothesis to reject or not.

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

For Lower: $\mathbf{H_1} = $ "$\mu_X < k$" **For Upper:** $\mathbf{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**:

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

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 *α*** of the test: $\alpha =$ probability you are wrong if you Reject $\mathbf{H_0}$ 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 $\bar{x}$ e 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.

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**:

$\mathbf{H_0}$ = "$\mu_X = k$."   (and $\mathbf{H_1}$ = "$\boldsymbol{\mu_X < k}$")

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** $\boldsymbol{\alpha}$ of the test

α = probability you are wrong if you Reject $\mathbf{H_0}$ 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.

Hypothesis Testing: One-Tailed Tests

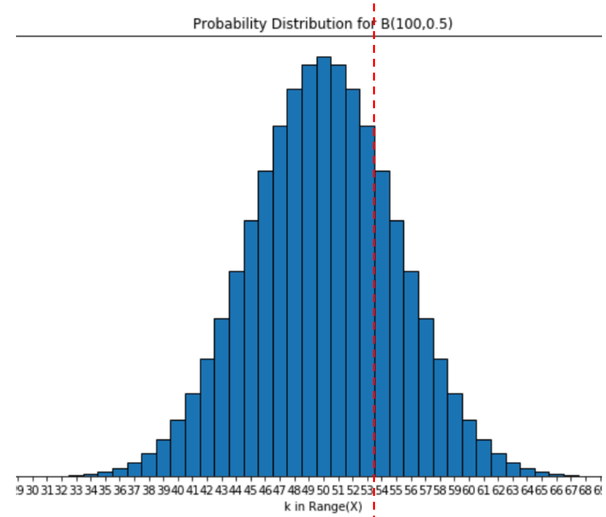
Example: Upper One-Tailed Test:

Richard claims that he has ESP. I disagree. My hypothesis is that Richard 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 = "Richard's average number of correct cards is 50, because he is randomly guessing."
H_1 = "Richard 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

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

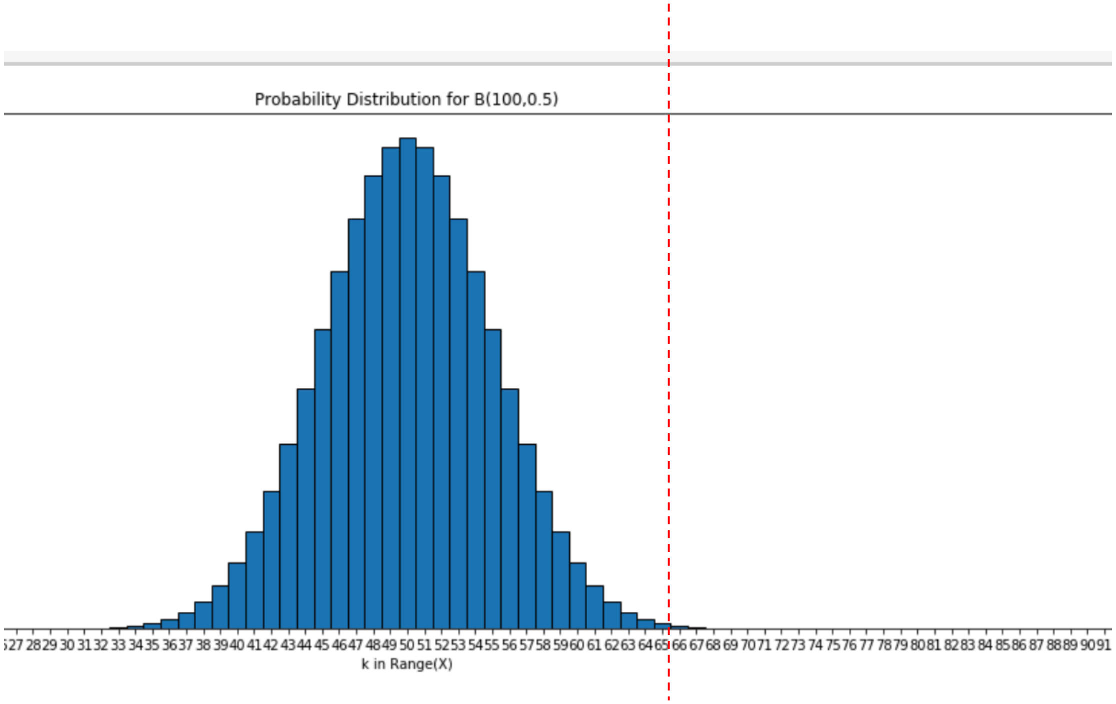
Since $0.2431 > 0.05$, we fail to reject H_0.

Hypothesis Testing: One-Tailed Tests

But what if he had guessed 68 of them correctly?

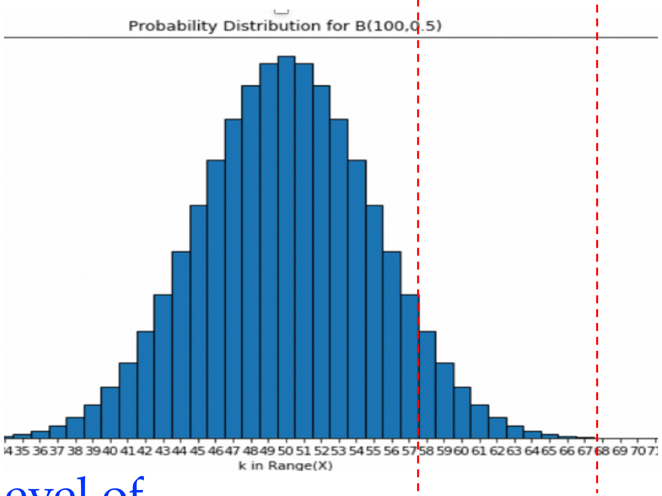
$P(X>=68) = 0.0002044$

Since 0.0002 < 0.05, we Reject my hypothesis that Richard 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 Richard guessed ≥ 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:  0.18410080866 3
xbar = 55:  0.135626512037
xbar = 56:  0.0966739522478
xbar = 57:  0.0666053096036
xbar = 58:  0.04431304005 7
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