Spring 2019

Lecture: Probability and Statistics - Hypothesis Testing - Week Eight

Evan Agovino
CUNY City College

NYC Tech-in-Residence Corps

Week Eight: Hypothesis Testing

CS 217

Introduction

- Say we want to see if a coin is unfair or not. If we toss it 10 times, how many heads would make us think it was unfair?
    - 9 Heads?
    - 0 Heads?
    - 3 Heads?

Introduction

- Say we want to see if a coin is unfair or not. If we toss it 10 times, how many heads would make us think it was unfair?
  - 9 seems unfair
  - 0 definitely seems unfair
  - 3? Could just be chance
- **Hypothesis Testing is a way of thinking quantitatively about this**

Definitions

- **H_0 - Null Hypothesis** - The default assumption for the model generating the data
- **H_A - Alternative Hypothesis** - The alternate explanation for our data given that we reject the null hypothesis
- **X - Test Statistic** - The result we see in our data
- **Null distribution** - the probability distribution of X assuming H_0
- **Rejection region** - if X is in the rejection region, we reject H_0 in favor of H_A
- **Non-Rejection region** - the **complement** to the rejection region, if X is in this region we do not reject H_0

Definitions

- **H_0 - Null Hypothesis** - The coin that we're flipping is fair.
- **H_A - Alternative Hypothesis** - The coin we're flipping is not fair.
- **X - Test Statistic** - We see _ heads in 10 flips.
- **Null distribution** - The null distribution is a binomial distribution with 10 trials and a probability of success of 50% per trial.
- **Rejection region** - The number of heads in 10 flips that seem suspicious to us.
- **Non-Rejection region** - The number of heads in 10 flips that do not seem suspicious to us.

Definitions

- H_0 - Null Hypothesis - The coin that we're flipping is fair.
- H_A - Alternative Hypothesis - The coin we're flipping is not fair.
- X - Test Statistic - We see _ heads in 10 flips.
- Null distribution - The null distribution is a binomial distribution with 10 trials and a probability of success of 50% per trial.
- **Rejection region** - The number of heads in 10 flips that seem suspicious to us.
- **Non-Rejection region** - The number of heads in 10 flips that do not seem suspicious to us.
- How do we determine these?

What’s a Fair Coin, Anyway?

- We know we can use the PMF function to find the odds of getting a specific number of heads in ten flips.

Head Count	**Odds**	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

What’s a Fair Coin, Anyway?

- We know we can use the PMF function to find the odds of getting a specific number of heads in ten flips.
- From there, it’s up to us to determine what we deem as suspicious
- For example, we can say that 2 or less heads in 10 coin flips or 8 or greater heads in 10 coin flips is ‘suspicious’
- Given the PMFs of the distribution, there is an 11% chance of this occurring if the coin is fair

Head Count	Odds	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

Definitions

- **H_0 - Null Hypothesis** - The coin that we're flipping is fair.
- **H_A - Alternative Hypothesis** - The coin we're flipping is not fair.
- **X - Test Statistic** - We see _ heads in 10 flips.
- **Null distribution** - The null distribution is a binomial distribution with 10 trials and a probability of success of 50% per trial.
- **Rejection region** - Getting 0, 1, 2, 8, 9, or 10 heads in 10 coin flips
- **Non-Rejection region** - Getting 3, 4, 5, 6, or 7 heads in 10 coin flips
- The probability of us rejecting **H_0 given that H_0 is true is 11%**

Definitions

	H_0	H_A
Reject H_0	Type I Error - False Positive	True Positive
Don't Reject H_0	True Negative	Type II Error - False Negative

Definitions

	H_0	H_A
Reject H_0	Type I Error - False Positive	True Positive
Don’t Reject H_0	True Negative	Type II Error - False Negative

This is a little confusing - think of ‘positive’ as rejecting **H_0 and ‘negative’ as failing to reject H_0**

Definitions

- The probability of us **rejecting H_0 given that H_0 is true** is 11%
- Rejecting H_0 given that H_0 is true is a Type I error, or a false positive
- In hypothesis testing, this is also known as the significance level

Hypothesis Test Design

- 1. Pick the null hypothesis H_0
  - Here our null hypothesis is that 'the coin is not biased', or that specifically the probability of landing heads on a given coin flip is 0.5
- 2. Decide if H_A is one-sided or two-sided
  - Do we only care if the coin is biased towards heads (we get 8, 9, 10 coin flips)? Or biased towards heads or tails (we get 0, 1, 2 coin flips)?
- 3. Pick a significance level and determine the rejection region
  - Typical significance levels include 0.1, 0.05, and 0.01.
  - The significance level is equal to the probability of a false positive given the null hypothesis

Hypothesis Test Design

- 1. Pick the null hypothesis $\mathbf{H_0}$
  - **P(H) = 0.5**
- 2. Decide if H_A is one-sided or two-sided
  - Two-sided
- 3. Pick a significance level and determine the rejection region
  - 0.01

Which results are in our rejection region in this case?

Head Count	**Odds**	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

Hypothesis Test Design

- 1. Pick the null hypothesis H_0
    - $P(H) = 0.5$
- 2. Decide if H_A is one-sided or two-sided
    - Two-sided
- 3. Pick a significance level and determine the rejection region
    - 0.01

Which results are in our rejection region in this case?

Head Count	Odds	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

Hypothesis Test Design

- 1. Pick the null hypothesis H_0
  - $P(H) = 0.5$
- 2. Decide if H_A is one-sided or two-sided
  - Two-sided
- 3. Pick a significance level and determine the rejection region
  - 0.05

Which results are in our rejection region in this case?

Head Count	**Odds**	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

Hypothesis Test Design

- 1. Pick the null hypothesis H_0
  - $P(H) = 0.5$
- 2. Decide if H_A is one-sided or two-sided
  - Two-sided
- 3. Pick a significance level and determine the rejection region
  - 0.05

Which results are in our rejection region in this case?

Head Count	Odds	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

Hypothesis Test Design

- 1. Pick the null hypothesis H_0
  - P(H) = 0.5
- 2. Decide if H_A is one-sided or two-sided
  - One-sided, we only care about coins that are biased towards heads
- 3. Pick a significance level and determine the rejection region
  - 0.1

Which results are in our rejection region in this case?

Head Count	**Odds**	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

Hypothesis Test Design

- 1. Pick the null hypothesis H_0
  - $P(H) = 0.5$
- 2. Decide if H_A is one-sided or two-sided
  - One-sided, we only care about coins that are biased towards heads
- 3. Pick a significance level and determine the rejection region
  - 0.1

Which results are in our rejection region in this case?

Head Count	Odds	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

Hypothesis Test Design

- 1. Pick the null hypothesis H_0
  - $P(H) = 0.5$
- 2. Decide if H_A is one-sided or two-sided
  - One-sided, we only care about coins that are biased towards tails
- 3. Pick a significance level and determine the rejection region
  - 0.1

Which results are in our rejection region in this case?

Head Count	**Odds**	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

Hypothesis Test Design

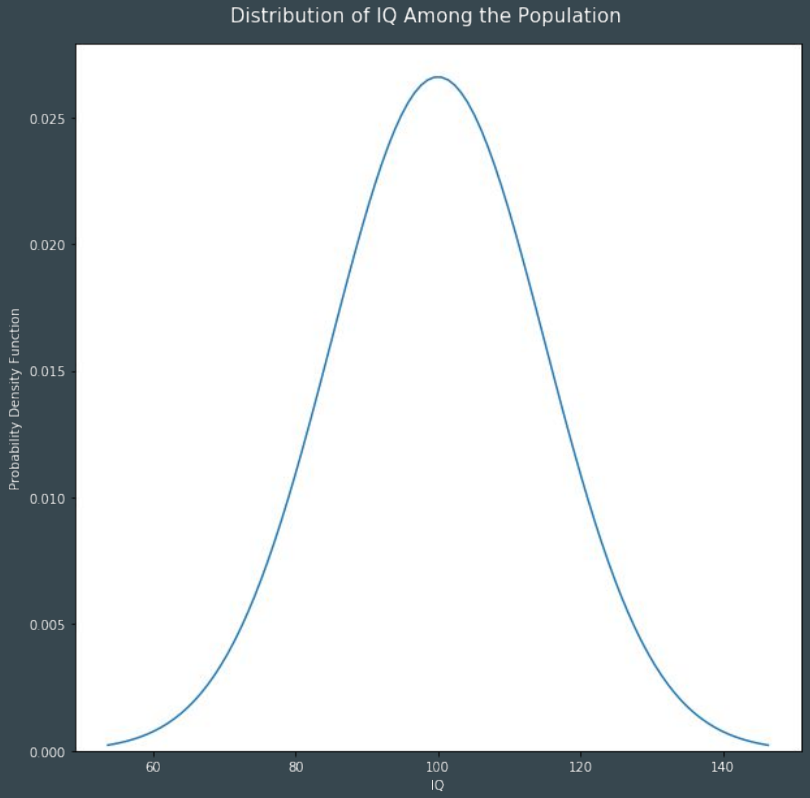
- 1. Pick the null hypothesis H_0
  - $P(H) = 0.5$
- 2. Decide if H_A is one-sided or two-sided
  - One-sided, we only care about coins that are biased towards tails
- 3. Pick a significance level and determine the rejection region
  - 0.1

Which results are in our rejection region in this case?

Head Count	Odds	5	0.246
0	0.001	6	0.205
1	0.01	7	0.117
2	0.044	8	0.044
3	0.117	9	0.01
4	0.205	10	0.001

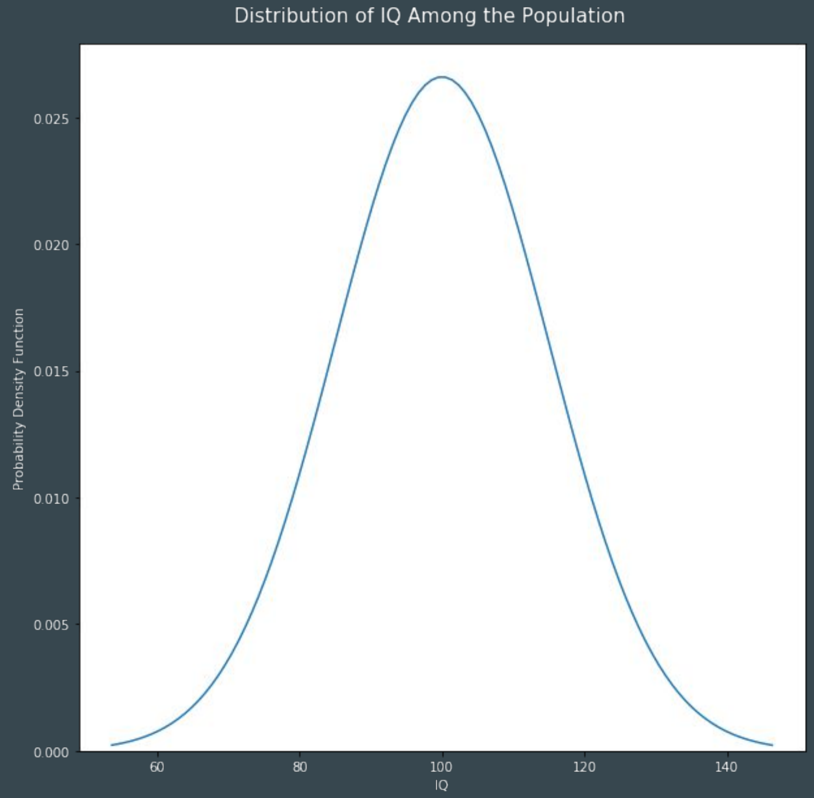
Hypothesis Test Design

- IQ is normally distributed in the population with a mean of 100 and a standard deviation of 15.
- Say we think that CCNY students have an above average intelligence.
- What might our null hypothesis be here?
- What might our alternative hypothesis be?



Hypothesis Test Design

- IQ is normally distributed in the population with a mean of 100 and a standard deviation of 15.
- Say we think that CCNY students have an above average intelligence.
- What might our null hypothesis be here?
  - CCNY students have, on average, an IQ of 100
- What might our alternative hypothesis be?
  - CCNY students have, on average, an IQ greater than 100

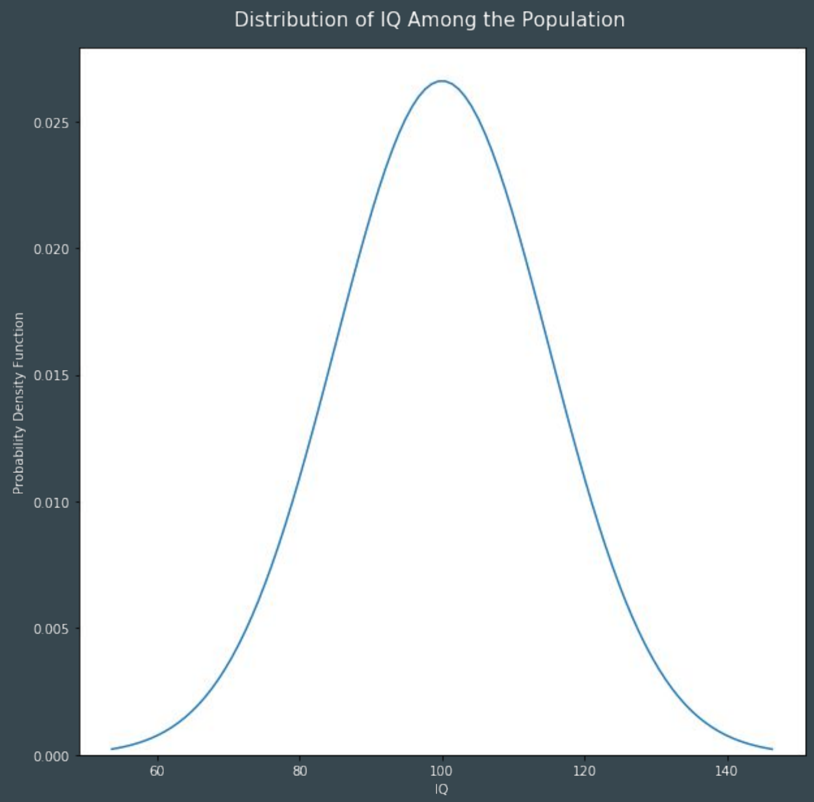


Hypothesis Test Design

- 1. Pick the null hypothesis H_0
  - The mean of IQ of CCNY students is 100
- 2. Decide if H_A is one-sided or two-sided
  - We only care if the mean IQ of CCNY students is greater than 100
  - Thus, our test is one-sided
- 3. Pick a significance level and determine the rejection region
  - Let's use a significance level of 0.05.
  - You can also determine a specific rejection region like we did earlier, but using a significance level is much more common.
- Let's say we ask nine students their IQ and take the mean of that. How do we determine the rejection region?

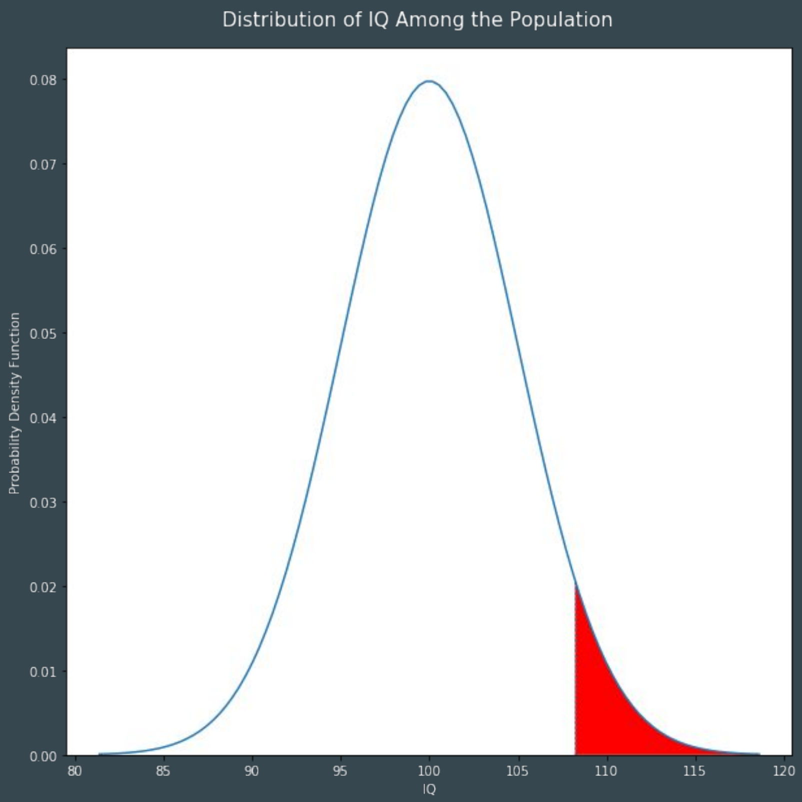
Hypothesis Test Design

- Like we discussed, a few weeks ago, the distribution of the mean IQ for ten people is different from the distribution of IQ for the population.
- The population IQ has a mean of 100 and a standard distribution of 15.
- What does the distribution of the mean IQ for a sample of 9 people have a mean and standard deviation of?



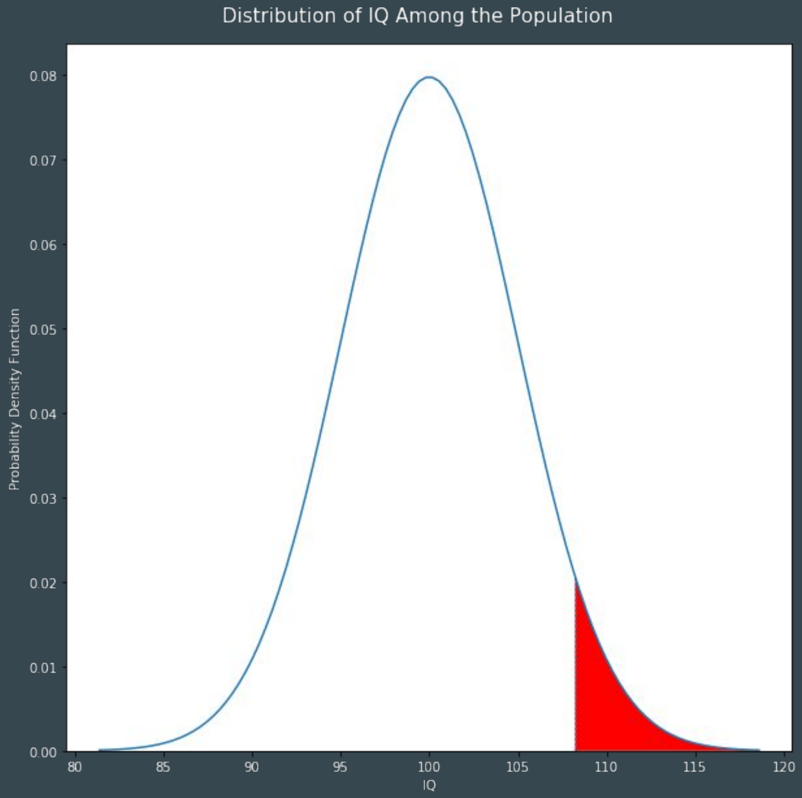
Hypothesis Test Design

- The mean IQ is still 100, but the standard deviation is now 15 divided by the square root of 9, or 5
- Similar to what we did for the discrete distribution, we will **reject** the null hypothesis for any results that come in above the 95th (1 - 0.05) percentile.
- This means anything with a Z-score approximately greater than 1.64
- Specifically that means a mean IQ of 108.22 for the nine people we speak with



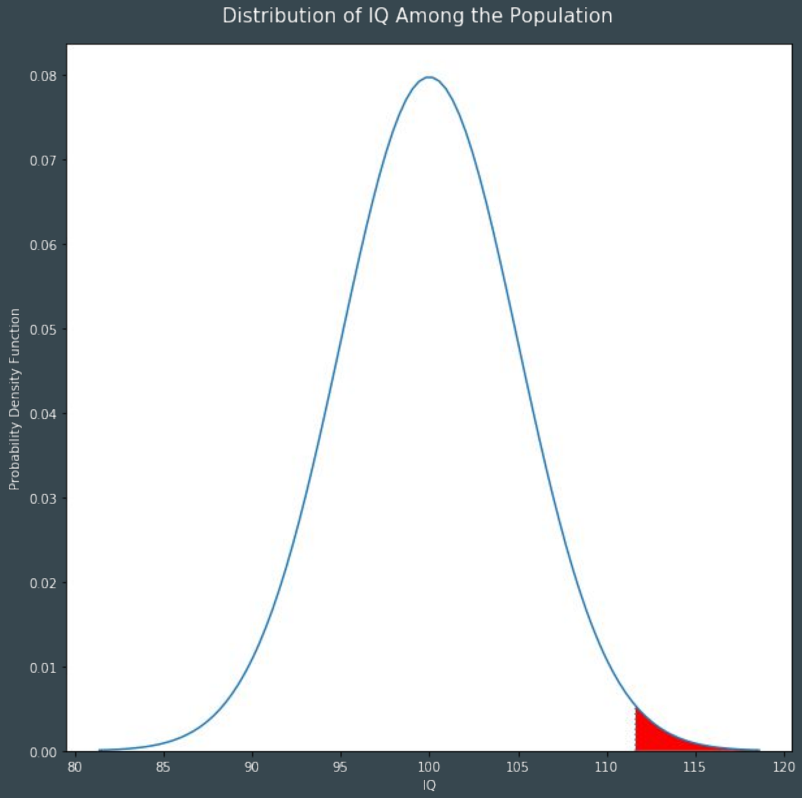
Hypothesis Test Design

- If our nine students have a mean IQ of 110, we can **reject the null hypothesis** that the mean IQ of CCNY students is equal to 100
- If our nine students have a mean IQ of 105, we fail to reject the null hypothesis that the mean IQ of CCNY students is equal to 100
- Remember that our significance level is 0.05. What if we decrease it to 0.01?



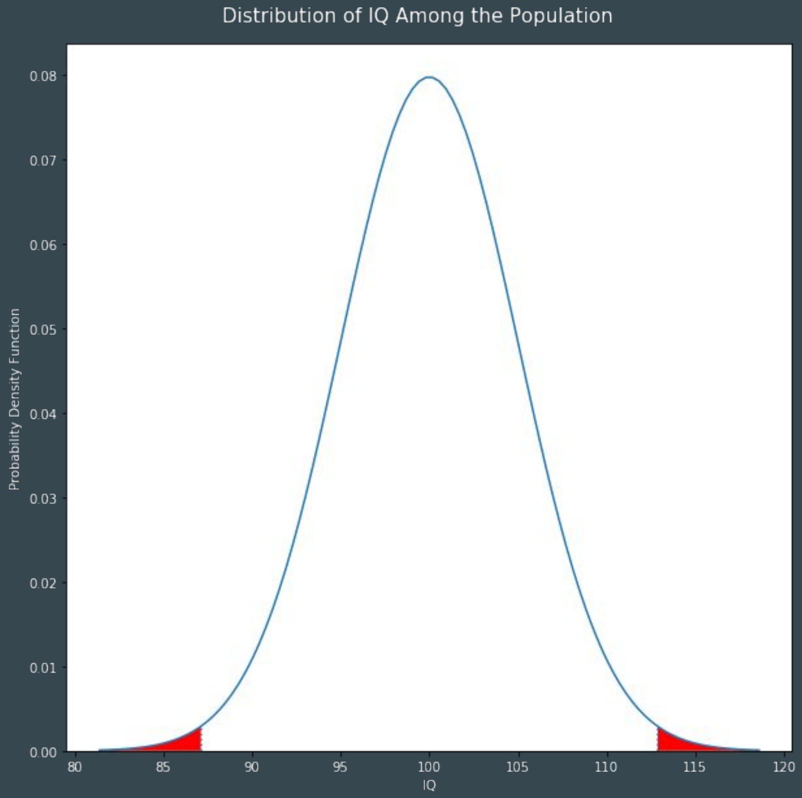
Hypothesis Test Design

- Now we will reject any results that are above the 99th percentile
- This now means anything with a Z-score approximately greater than 2.32
- Specifically that means a mean IQ of 111.63 for the nine people we speak with
- If our nine students have a mean IQ of 110, we now **reject the null hypothesis** that the mean IQ of CCNY students is equal to 100



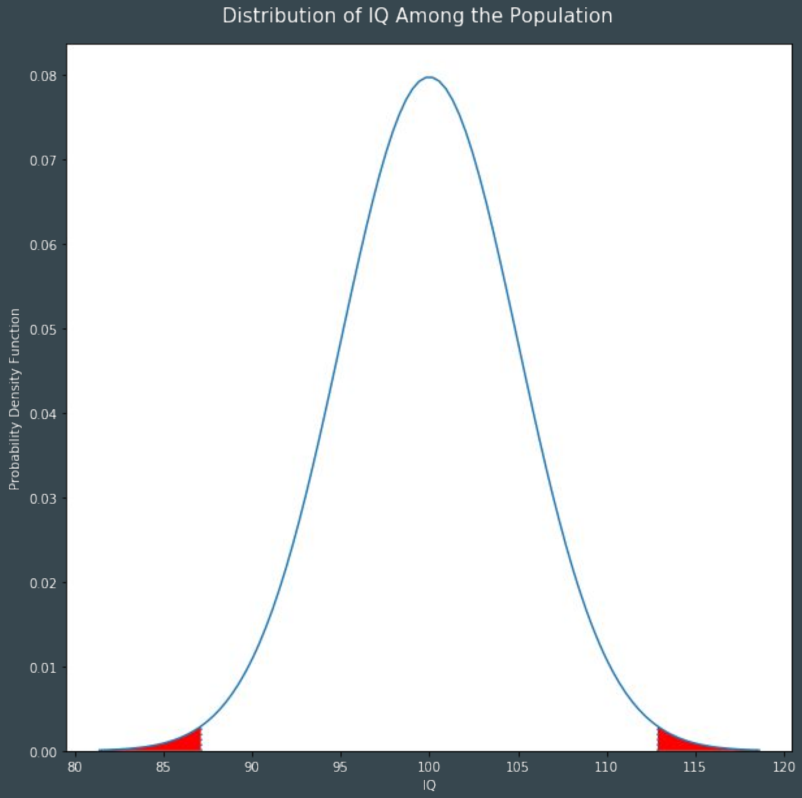
Hypothesis Test Design

- Now let's do a two-sided test rather than a one-sided test
- We are now testing whether the mean IQ of a CCNY is *different* than the mean IQ of the population, rather than greater.
- Now, with a significance level still at 0.01, the rejection region consists of the area before the 0.5th (half of the 1st) percentile and greater than the 99.5th percentile.



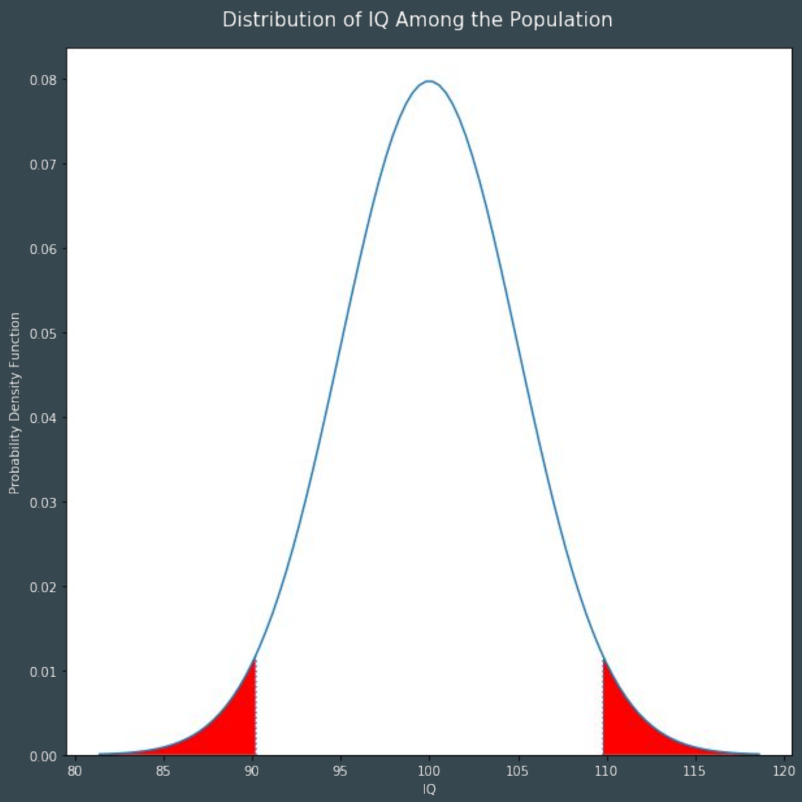
Hypothesis Test Design

- This consists of a Z-score *greater than* 2.57, or *less than* -2.57.
- This is equivalent to an IQ less than 87.12, or greater than 112.87.
- If our nine students have a mean IQ of 110, we will *fail to reject* the null hypothesis that the mean IQ of CCNY students is the same as the mean IQ of the general population.



Hypothesis Test Design

- Now let's change the significance level from 0.01 to 0.05.
- The rejection region will now encompass anywhere between the 2.5th percentile and the 97.5th percentile.
- This consists of a Z-score *greater than* 2.57, or *less than* -2.57.
- This is equivalent to an IQ less than 90.2 or an IQ greater than 109.79
- Now we will *reject* the null hypothesis



Estimation

- Thus far, we have worked with data where we have a known population mean and population variance
- In our IQ example, we know that the population mean is 100 and the population variance is the square of the population standard deviation, or $15^2 = 225$.
- In our coin flip example, we know that the population mean is 5 and the population variance is 1.25
- But what if we don't know the population mean or population variance? We can *estimate* them using our sample data

Estimation

- We previously knew that the nine students at CCNY had an average IQ of 110.
- To the right are the nine student samples that we took.
- The **sample mean** of our nine samples is 110
- What about the sample variance?

1. 112	6. 91
2. 94	7. 142
3. 116	8. 119
4. 140	9. 85
5. 91	

Estimation

- Traditionally the variance entails taking the sum of the square value of all of the values in a dataset subtracted by the mean of the dataset, divided by the length of the dataset.
- Here, the variance is 405.33 and the standard deviation is 20.13

1. 112	6. 91
2. 94	7. 142
3. 116	8. 119
4. 140	9. 85
5. 91	

$$\frac{(112-110)^2+(94-110)^2+(116-110)^2+(140-110)^2+(91-110)^2+(91-110)^2+(142-110)^2+(119-110)^2+(85-110)^2}{10}$$

Estimation

- However, the **sample** variance entails taking the sum of the square value of all of the values in a dataset subtracted by the mean of the dataset, divided by the length of the dataset - 1.
- Here, the sample variance is 456 and the sample standard deviation is 21.35

$$\frac{(112-110)^2+(94-110)^2+(116-110)^2+(140-110)^2+(91-110)^2+(91-110)^2+(142-110)^2+(119-110)^2+(85-110)^2}{9}$$

1. 112	6. 91
2. 94	7. 142
3. 116	8. 119
4. 140	9. 85
5. 91	

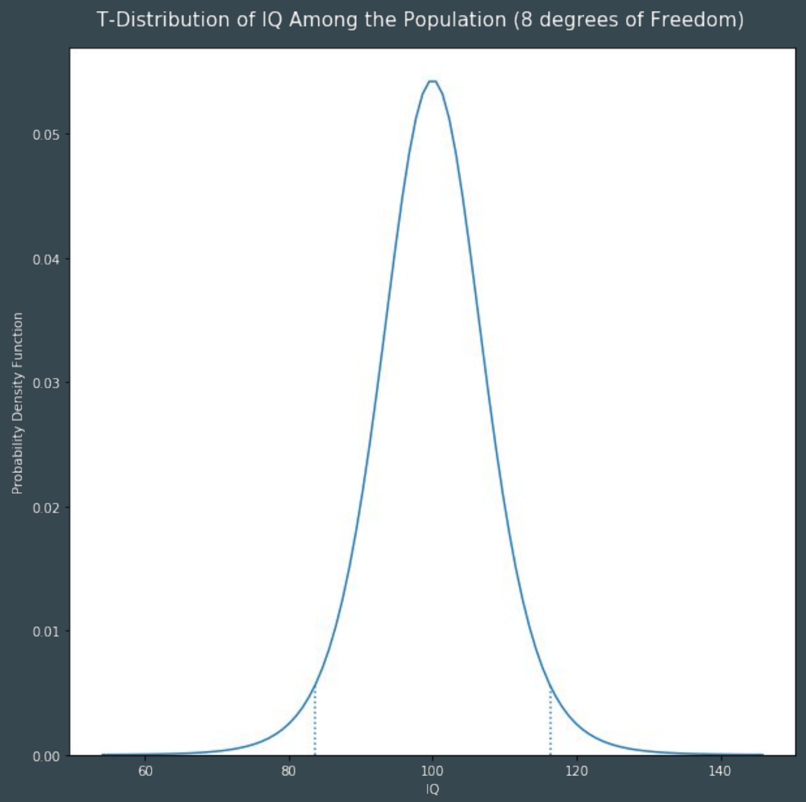
T Distribution

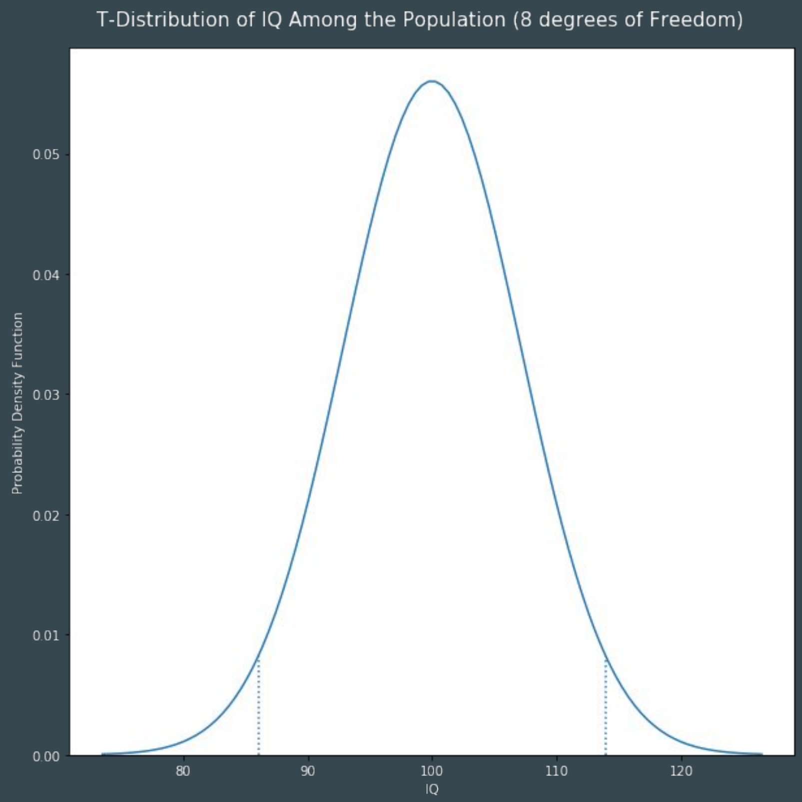
- Now, given our samples, we want to see if we can reject the original null hypothesis that the mean IQ for CCNY students is equal to 100 at a significance level of 0.05
- Originally, we had a normal distribution for the null hypothesis, with a mean of 100 and a standard deviation of 5 (15 divided by the square root of our sample of 9)
- Now, we will use a T-Distribution for the null hypothesis, with a mean of 100 and standard deviation of 7.12 (21.35 divided by the square root of our sample of 9)
- The T-Distribution is a bell-shaped curve, similar to the normal distribution, that has more probability in its tails than the standard normal distribution (due to the uncertainty about the true population variance)
- The T-Distribution changes based on the number of samples its based on. This metric is called *degrees of freedom*.

T Distribution

- The *degrees of freedom* are equal to the number of samples minus 1.
- In our case that will be 8, since there were 9 samples.
- So, to sum it up, we will be building a t-distribution with a mean of 100, a standard deviation of 7.12, and 8 degrees of freedom.
- In the next slide, we'll see a normal distribution with a mean of 100 and a standard deviation of 7.12 side by side with a T-distribution with a mean of 100, standard deviation of 7.12, and 8 degrees of freedom side by side, with dotted lines on each representing the 2.5th and 97.5th percentiles.

T Distribution





T Distribution

- Now we can follow our standard procedure and create rejection regions, given a two-sided test at a significance level of 0.05
- We will reject any values with a Z-score less than -2.30 or greater than 2.30
- Or we will reject any values with an IQ less than 83.58 or greater than 116.41
- This compares to a normal distribution, where we will reject any values with a Z-score less than -1.96 or greater than 1.96
- Or reject any values with an IQ less than 86.04 or greater than 113.95.
- In either case, we *fail to reject* the null hypothesis that the mean IQ of the CCNY student is higher than the mean IQ of the general population.

T Distribution