STAT234: Lecture 6 - Confidence Intervals

Kushal K. Dey

A Romantic Experiment !

Most people are right-handed and even the right eye is dominant for most people. Molecular biologists have suggested that late-stage human embryos tend to turn their heads to the right. German bio-psychologist Onur Güntürkün conjectured that this tendency to turn to the right manifests itself in other ways as well, so he studied kissing couples to see if both people tended to lean to their right more often than to their left (and if so, how strong the tendency is). He and his researchers observed couples from age 13 to 70 in public places such as airports, train stations, beaches, and parks in the United States, Germany, and Turkey. They were careful not to include couples who were holding objects such as luggage that might have affected which direction they turned. In total, 124 kissing pairs were observed with 80 couples leaning right (*Nature*, 2003).

You and your boss !

- **Boss**: Give me the value of p, the actual proportion of people leaning right while kissing !!

- **Boss**: Give me the value of p, the actual proportion of people leaning right while kissing !!
- **You**: I don't know Sir. But seems like it would be around $\hat{p} = \frac{80}{124} = 0.64$. I know this because

$$\hat{p} = \frac{1}{124}\sum_{i=1}^{124} X_i$$

- **Boss**: Give me the value of p, the actual proportion of people leaning right while kissing !!
- **You**: I don't know Sir. But seems like it would be around $\hat{p} = \frac{80}{124} = 0.64$. I know this because

$$\hat{p} = \frac{1}{124} \sum_{i=1}^{124} X_i$$

where

$$X_i \sim Ber(p)$$

- **Boss**: Give me the value of p, the actual proportion of people leaning right while kissing !!
- **You**: I don't know Sir. But seems like it would be around $\hat{p} = \frac{80}{124} = 0.64$. I know this because

$$\hat{p} = \frac{1}{124} \sum_{i=1}^{124} X_i$$

where

$$X_i \sim Ber(p)$$

$$E(X_i) = p \; \forall i \qquad E\left(\frac{1}{124} \sum_{i=1}^{124} X_i\right) = p$$

- **Boss**: Give me the value of p, the actual proportion of people leaning right while kissing !!
- **You**: I don't know Sir. But seems like it would be around $\hat{p} = \frac{80}{124} = 0.64$. I know this because

$$\hat{p} = \frac{1}{124}\sum_{i=1}^{124} X_i$$

where

$$X_i \sim Ber(p)$$

$$E(X_i) = p \; \forall i \qquad E\left(\frac{1}{124}\sum_{i=1}^{124} X_i\right) = p$$

implying

- **Boss**: Give me the value of p, the actual proportion of people leaning right while kissing !!
- **You**: I don't know Sir. But seems like it would be around $\hat{p} = \frac{80}{124} = 0.64$. I know this because

$$\hat{p} = \frac{1}{124} \sum_{i=1}^{124} X_i$$

where

$$X_i \sim Ber(p)$$

$$E(X_i) = p \; \forall i \qquad E\left(\frac{1}{124} \sum_{i=1}^{124} X_i\right) = p$$

implying

- **Boss**: Thats not good enough, can you give me an interval where you are absolutely confident will contain the actual p?

- **Boss**: Thats not good enough, can you give me an interval where you are absolutely confident will contain the actual p?
- **You**: I am 100% confident that the interval $[0, 1]$ will contain the actual p Sir.

- **Boss**: Thats not good enough, can you give me an interval where you are absolutely confident will contain the actual p?
- **You**: I am 100% confident that the interval $[0, 1]$ will contain the actual p Sir.
- **Boss**: Seriously!! Thats all you got from the study?

- **Boss**: Thats not good enough, can you give me an interval where you are absolutely confident will contain the actual p?
- **You**: I am 100% confident that the interval $[0, 1]$ will contain the actual p Sir.
- **Boss**: Seriously!! Thats all you got from the study?
- **You**: No Sir, I am not 100% confident, but I am more or less 95% confident that p lies in the intrval $[0.556, 0.723]$ Sir.

- **Boss**: Thats not good enough, can you give me an interval where you are absolutely confident will contain the actual p?
- **You**: I am 100% confident that the interval $[0, 1]$ will contain the actual p Sir.
- **Boss**: Seriously!! Thats all you got from the study?
- **You**: No Sir, I am not 100% confident, but I am more or less 95% confident that p lies in the intrval $[0.556, 0.723]$ Sir.
- **Boss** : How on earth did you get that interval?

- **Boss**: Thats not good enough, can you give me an interval where you are absolutely confident will contain the actual p?
- **You**: I am 100% confident that the interval $[0, 1]$ will contain the actual p Sir.
- **Boss**: Seriously!! Thats all you got from the study?
- **You**: No Sir, I am not 100% confident, but I am more or less 95% confident that p lies in the intrval $[0.556, 0.723]$ Sir.
- **Boss** : How on earth did you get that interval?

Well, since n is moderately large, I assumed

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Well, since n is moderately large, I assumed

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

That meant

$$\sqrt{n}\frac{\hat{p} - p}{\sqrt{p(1-p)}} \sim N(0, 1)$$

Well, since n is moderately large, I assumed

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

That meant

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} \sim N(0,1)$$

So, I can write this as $Z = \sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}}$. From normal table I know

$$Pr[-1.96 < Z < 1.96] = 0.95$$

Well, since n is moderately large, I assumed

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

That meant

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} \sim N(0,1)$$

So, I can write this as $Z = \sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}}$. From normal table I know

$$Pr[-1.96 < Z < 1.96] = 0.95$$

$$Pr[-1.96 < \sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} < 1.96] = 0.95$$

Well, since n is moderately large, I assumed

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

That meant

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} \sim N(0,1)$$

So, I can write this as $Z = \sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}}$. From normal table I know

$$Pr[-1.96 < Z < 1.96] = 0.95$$

$$Pr[-1.96 < \sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} < 1.96] = 0.95$$

Well, since n is moderately large, I assumed

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

That meant

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} \sim N(0,1)$$

So, I can write this as $Z = \sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}}$. From normal table I know

$$Pr[-1.96 < Z < 1.96] = 0.95$$

$$Pr[-1.96 < \sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} < 1.96] = 0.95$$

- **Boss**: So? How do you calculate the left and right sides? You don't know p ?

- **Boss**: So? How do you calculate the left and right sides? You don't know p ?
- **You**: Well I cheated a bit Sir. and wrote

$$Pr\left[\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

- **Boss**: So? How do you calculate the left and right sides? You don't know p ?
- **You**: Well I cheated a bit Sir. and wrote

$$Pr\left[\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

Then I calculate the right and left side by putting $\hat{p} = .64$ and I get the interval $[0.56, 0.72]$.

- **Boss**: So? How do you calculate the left and right sides? You don't know p ?
- **You**: Well I cheated a bit Sir. and wrote

$$Pr\left[\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

  Then I calculate the right and left side by putting $\hat{p} = .64$ and I get the interval $[0.56, 0.72]$.
- **Boss**: So I get

$$Pr[0.56 < p < 0.72] \approx 0.95$$

- **Boss**: So? How do you calculate the left and right sides? You don't know p ?
- **You**: Well I cheated a bit Sir. and wrote

$$Pr\left[\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

  Then I calculate the right and left side by putting $\hat{p} = .64$ and I get the interval $[0.56, 0.72]$.
- **Boss**: So I get

$$Pr[0.56 < p < 0.72] \approx 0.95$$

- **You**: Well no Sir, that statement is not right!.

- **Boss**: WTH, isn't that what you just did?

- **Boss**: WTH, isn't that what you just did?
- **You**: Well let me explain Sir. I said with 95% chance, the true p will lie in the interval

$$p \in \left[\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

- **Boss**: WTH, isn't that what you just did?
- **You**: Well let me explain Sir. I said with 95% chance, the true p will lie in the interval

$$p \in \left[\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

  and I have one such realization of this random interval from experiment and that is $[0.56, 0.72]$. Thats why I am saying there is a 95% chance this interval could contain p.

- **Boss**: WTH, isn't that what you just did?
- **You**: Well let me explain Sir. I said with 95% chance, the true p will lie in the interval

$$p \in \left[\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

  and I have one such realization of this random interval from experiment and that is $[0.56, 0.72]$. Thats why I am saying there is a 95% chance this interval could contain p.
- **Boss**: Okay whatever! These numbers 0.56 and 0.72 do not look pretty. How about the interval $[0.6, 0.7]$. Can you also say you are 90% or something confident of that?

- **Boss**: WTH, isn't that what you just did?
- **You**: Well let me explain Sir. I said with 95% chance, the true p will lie in the interval

$$p \in \left[\hat{p} - 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

  and I have one such realization of this random interval from experiment and that is $[0.56, 0.72]$. Thats why I am saying there is a 95% chance this interval could contain p.
- **Boss**: Okay whatever! These numbers 0.56 and 0.72 do not look pretty. How about the interval $[0.6, 0.7]$. Can you also say you are 90% or something confident of that?

- **You**: Let me check.

- **You**: Let me check.
  Given that $\hat{p} = 0.64$, and you want to find the confidence for $p \in [0.60, 0.70]$, we can try to look at the probability

$$Pr\left[\hat{p} - 0.04 < p < \hat{p} + 0.06\right]$$

- **You**: Let me check.
  Given that $\hat{p} = 0.64$, and you want to find the confidence for $p \in [0.60, 0.70]$, we can try to look at the probability

$$Pr\left[\hat{p} - 0.04 < p < \hat{p} + 0.06\right]$$

or

$$Pr\left[-0.06 < \hat{p} - p < 0.04\right]$$

- **You**: Let me check.
  Given that $\hat{p} = 0.64$, and you want to find the confidence for $p \in [0.60, 0.70]$, we can try to look at the probability

$$Pr\left[\hat{p} - 0.04 < p < \hat{p} + 0.06\right]$$

or

$$Pr\left[-0.06 < \hat{p} - p < 0.04\right]$$

$$Pr\left[-\sqrt{n}\frac{0.06}{\sqrt{p(1-p)}} < \sqrt{n}\frac{(\hat{p} - p)}{\sqrt{p(1-p)}} < \sqrt{n}\frac{0.04}{\sqrt{p(1-p)}}\right]$$

- **You**: Let me check.
  Given that $\hat{p} = 0.64$, and you want to find the confidence for $p \in [0.60, 0.70]$, we can try to look at the probability

$$Pr\left[\hat{p} - 0.04 < p < \hat{p} + 0.06\right]$$

or

$$Pr\left[-0.06 < \hat{p} - p < 0.04\right]$$

$$Pr\left[-\sqrt{n}\frac{0.06}{\sqrt{p(1-p)}} < \sqrt{n}\frac{(\hat{p} - p)}{\sqrt{p(1-p)}} < \sqrt{n}\frac{0.04}{\sqrt{p(1-p)}}\right]$$

Then I cheat again

$$Pr\left[-\sqrt{n}\frac{0.06}{\sqrt{\hat{p}(1-\hat{p})}} < Z < \sqrt{n}\frac{0.04}{\sqrt{\hat{p}(1-\hat{p})}}\right]$$

$$Pr\left[-1.414214 < Z < 0.942809\right] \approx 0.7484612$$

So, I am 75% confident that the actual p lies in the interval

$$p \in [\hat{p} - 0.04, \hat{p} + 0.06] := [0.60, 0.70]$$

$$Pr\left[-\sqrt{n}\frac{0.06}{\sqrt{\hat{p}(1-\hat{p})}} < Z < \sqrt{n}\frac{0.04}{\sqrt{\hat{p}(1-\hat{p})}}\right]$$

$$Pr\left[-1.414214 < Z < 0.942809\right] \approx 0.7484612$$

So, I am 75% confident that the actual p lies in the interval

$$p \in [\hat{p} - 0.04, \hat{p} + 0.06] := [0.60, 0.70]$$

- **Boss**: What? I just approximate to nearest first decimal, and now you reduced your confidence by 20%? Are you serious?

$$Pr\left[-\sqrt{n}\frac{0.06}{\sqrt{\hat{p}(1-\hat{p})}} < Z < \sqrt{n}\frac{0.04}{\sqrt{\hat{p}(1-\hat{p})}}\right]$$

$$Pr\left[-1.414214 < Z < 0.942809\right] \approx 0.7484612$$

So, I am 75% confident that the actual p lies in the interval

$$p \in [\hat{p} - 0.04, \hat{p} + 0.06] := [0.60, 0.70]$$

- **Boss**: What? I just approximate to nearest first decimal, and now you reduced your confidence by 20%? Are you serious?
- **You**: Not my fault Sir, its the distribution to blame.

$$Pr\left[-\sqrt{n}\frac{0.06}{\sqrt{\hat{p}(1-\hat{p})}} < Z < \sqrt{n}\frac{0.04}{\sqrt{\hat{p}(1-\hat{p})}}\right]$$

$$Pr\left[-1.414214 < Z < 0.942809\right] \approx 0.7484612$$

So, I am 75% confident that the actual p lies in the interval

$$p \in [\hat{p} - 0.04, \hat{p} + 0.06] := [0.60, 0.70]$$

- **Boss**: What? I just approximate to nearest first decimal, and now you reduced your confidence by 20%? Are you serious?
- **You**: Not my fault Sir, its the distribution to blame.

- **Boss**: Okay let me just report the 95% interval then, what was that $(0.56, 0.72)$ right?

- **Boss**: Okay let me just report the 95% interval then, what was that $(0.56, 0.72)$ right?
- **You**: Well kind of.

- **Boss**: Okay let me just report the 95% interval then, what was that $(0.56, 0.72)$ right?
- **You**: Well kind of.
- **Boss**: What do you mean kind of?

- **Boss**: Okay let me just report the 95% interval then, what was that $(0.56, 0.72)$ right?
- **You**: Well kind of.
- **Boss**: What do you mean kind of?
- **You**: Well I cheated in the middle and used replace p by $\hat{p}$ and said essentially

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}} \approx N(0,1)$$

instead of

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} \sim N(0,1)$$

which was the correct statement.

- **Boss**: I can't write *kind of* in my report. Can't you give me a more exact interval?

- **You**: I can say exactly that

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}} \sim t_{n-1}$$

  where t_{n-1} is another distribution (called *Student's t* distribution).
- **Boss**: Where did you come up with that distribution from?
- **You**: Okay let me explain.

How is the $T-$distribution Related to Estimation?

Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ popn.

We know that $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$.

How is the $T-$distribution Related to Estimation?

Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ popn.

We know that $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$.

But what if σ^2 is not known?

How is the $T-$distribution Related to Estimation?

Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ popn.

We know that $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$.

But what if σ^2 is not known? Use $s \approx \sigma$ to estimate it.

Consider

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{s} = Z\frac{\sigma}{s} \quad (\text{where } Z \sim N(0, 1)) \sim t_{(n-1)}$$

How is the $T-$distribution Related to Estimation?

Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ popn.

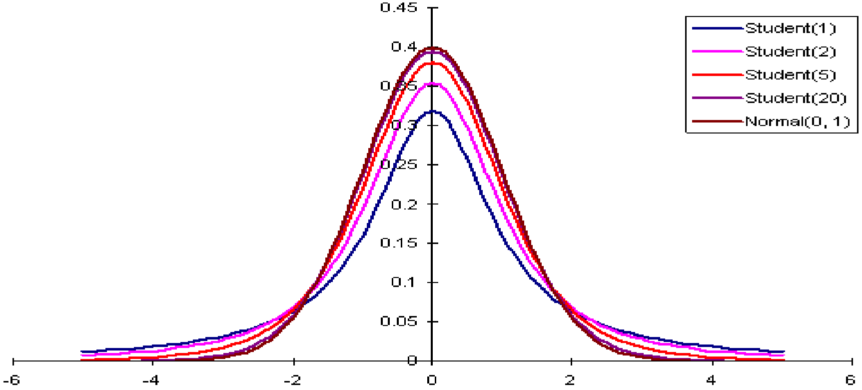
We know that $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$.

But what if σ^2 is not known? Use $s \approx \sigma$ to estimate it.

Consider

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{s} = Z\frac{\sigma}{s} \quad \text{(where } Z \sim N(0,1)\text{)} \sim t_{(n-1)}$$

T distribution look.....



How is the $T-$distribution Related to Estimation?

$T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$

How is the $T-$distribution Related to Estimation?

$T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$

So, even when the popn variance σ^2 is not known, we can still find probabilities for the sample mean $\bar{X}$ for data from a normal popn.

How is the $T-$distribution Related to Estimation?

$$T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$$

So, even when the popn variance σ^2 is not known, we can still find probabilities for the sample mean $\bar{X}$ for data from a normal popn.

We just replace σ by its estimate (s) in the standardization of $\bar{X}$ and then look up probabilities from the $t_{(n-1)}$ distribution instead of the $N(0,1)$ distribution.

How is the $T-$distribution Related to Estimation?

$T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$

So, even when the popn variance σ^2 is not known, we can still find probabilities for the sample mean $\bar{X}$ for data from a normal popn.

We just replace σ by its estimate (s) in the standardization of $\bar{X}$ and then look up probabilities from the $t_{(n-1)}$ distribution instead of the $N(0,1)$ distribution.

The $t_{(n-1)}$ distribution is similar to the $N(0,1)$, but with more spread (to account for the extra uncertainty in the estimate of σ.)

How is the $T-$distribution Related to Estimation?

$T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$

So, even when the popn variance σ^2 is not known, we can still find probabilities for the sample mean $\bar{X}$ for data from a normal popn.

We just replace σ by its estimate (s) in the standardization of $\bar{X}$ and then look up probabilities from the $t_{(n-1)}$ distribution instead of the $N(0, 1)$ distribution.

The $t_{(n-1)}$ distribution is similar to the $N(0, 1)$, but with more spread (to account for the extra uncertainty in the estimate of σ.)

As sample size increases, s gets closer to σ
and the $t_{(n-1)}$ distribution gets closer to $N(0, 1)$.

How is the $T-$distribution Related to Estimation?

$T = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$

So, even when the popn variance σ^2 is not known, we can still find probabilities for the sample mean $\bar{X}$ for data from a normal popn.

We just replace σ by its estimate (s) in the standardization of $\bar{X}$ and then look up probabilities from the $t_{(n-1)}$ distribution instead of the $N(0,1)$ distribution.

The $t_{(n-1)}$ distribution is similar to the $N(0,1)$, but with more spread (to account for the extra uncertainty in the estimate of σ.)

As sample size increases, s gets closer to σ
and the $t_{(n-1)}$ distribution gets closer to $N(0,1)$.

Connection to the problem here

$$s := \sqrt{\hat{p}(1-\hat{p})}$$

$$\bar{X} := \hat{p}$$

Connection to the problem here

$$s := \sqrt{\hat{p}(1-\hat{p})}$$

$$\bar{X} := \hat{p}$$

As a result, we can write

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}} \sim t_{n-1}$$

Connection to the problem here

$$s := \sqrt{\hat{p}(1-\hat{p})}$$

$$\bar{X} := \hat{p}$$

As a result, we can write

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}} \sim t_{n-1}$$

The t-distribution is nice because it is well-documented, meaning there is a table for t_{n-1} of the probabilities just as there is a table for $N(0,1)$ as you amust be familiar with by now.

t distribution

A random variable $T \sim t_{n-1}$ distribution

$$Pr\left[-1.979 < T_{(n-1)=123} < 1.979\right] = 0.95$$

t distribution

A random variable $T \sim t_{n-1}$ distribution

$$Pr\left[-1.979 < T_{(n-1)=123} < 1.979\right] = 0.95$$

We know

$$Pr\left[-1.959 < Z < 1.959\right] = 0.95$$

t distribution

A random variable $T \sim t_{n-1}$ distribution

$$Pr\left[-1.979 < T_{(n-1)=123} < 1.979\right] = 0.95$$

We know

$$Pr\left[-1.959 < Z < 1.959\right] = 0.95$$

$$Pr[\hat{p} - 1.98\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \hat{p} + 1.98\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}] = 0.95$$

t distribution

A random variable $T \sim t_{n-1}$ distribution

$$Pr\left[-1.979 < T_{(n-1)=123} < 1.979\right] = 0.95$$

We know

$$Pr\left[-1.959 < Z < 1.959\right] = 0.95$$

$$Pr[\hat{p} - 1.98\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \hat{p} + 1.98\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}] = 0.95$$

$$95\ \%\ CI\ for\ p\ (t) : [0.554, 0.725]$$

t distribution

A random variable $T \sim t_{n-1}$ distribution

$$Pr\left[-1.979 < T_{(n-1)=123} < 1.979\right] = 0.95$$

We know

$$Pr\left[-1.959 < Z < 1.959\right] = 0.95$$

$$Pr[\hat{p} - 1.98\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < p < \hat{p} + 1.98\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}] = 0.95$$

$$95\ \%\ CI\ for\ p\ (t) : [0.554, 0.725]$$

$$95\ \%\ CI\ for\ p\ (z) : [0.56, 0.72]$$

One sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.

One sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[Z > -1.645\right] \approx 0.95$$

One sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[Z > -1.645\right] \approx 0.95$$

That means if we use the normal approximation

$$Pr\left[\sqrt{n}\frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}} > -1.645\right] \approx 0.95$$

One sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[Z > -1.645\right] \approx 0.95$$

That means if we use the normal approximation

$$Pr\left[\sqrt{n}\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} > -1.645\right] \approx 0.95$$

$$Pr\left[p > \hat{p} - 1.645\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

One sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[Z > -1.645\right] \approx 0.95$$

That means if we use the normal approximation

$$Pr\left[\sqrt{n}\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} > -1.645\right] \approx 0.95$$

$$Pr\left[p > \hat{p} - 1.645\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

One sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[Z > -1.645\right] \approx 0.95$$

That means if we use the normal approximation

$$Pr\left[\sqrt{n}\frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}} > -1.645\right] \approx 0.95$$

$$Pr\left[p > \hat{p} - 1.645\frac{\sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

One-sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.

One-sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[T_{123} > -1.657\right] \approx 0.95$$

One-sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[T_{123} > -1.657\right] \approx 0.95$$

That means if we use the normal approximation

$$Pr\left[\sqrt{n}\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} > -1.657\right] \approx 0.95$$

One-sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[T_{123} > -1.657\right] \approx 0.95$$

That means if we use the normal approximation

$$Pr\left[\sqrt{n}\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} > -1.657\right] \approx 0.95$$

$$Pr\left[p > \hat{p} - 1.657\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

One-sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[T_{123} > -1.657\right] \approx 0.95$$

That means if we use the normal approximation

$$Pr\left[\sqrt{n}\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} > -1.657\right] \approx 0.95$$

$$Pr\left[p > \hat{p} - 1.657\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

One-sided interval

- **Boss**: I want a confidence interval $[c, 1]$ since I know p is very high. So, give me a 95% confidence interval such that $p \geq c$ instead of $a \leq p \leq b$.
- **You**: We know that

$$Pr\left[T_{123} > -1.657\right] \approx 0.95$$

That means if we use the normal approximation

$$Pr\left[\sqrt{n}\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} > -1.657\right] \approx 0.95$$

$$Pr\left[p > \hat{p} - 1.657\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right] \approx 0.95$$

Hypothesis testing

- **Boss**: Okay I am reporting that I am 95% confident that p lies in $[0.5685, 1]$.
- **You**: Sure.
- **Boss**: So we can say now that there is effect of leaning on right with kissing .
- **You**: Yes you can. Thats because the 95% CI does not contain 0.5.
- **Boss**: Ohh, how did you work that out?

Connection: testing and CI

If you are testing

$$H_0 : p = 0.5 \qquad H_1 : p > 0.5$$

Connection: testing and CI

If you are testing

$$H_0 : p = 0.5 \qquad H_1 : p > 0.5$$

then if you want to test and reject the null when p-value < 0.05, then you just check if the null hypothesis p, which is 0.5, lies in the one sided confidence interval $[c, 1]$.

Connection: testing and CI

If you are testing

$$H_0 : p = 0.5 \qquad H_1 : p > 0.5$$

then if you want to test and reject the null when p-value < 0.05, then you just check if the null hypothesis p, which is 0.5, lies in the one sided confidence interval $[c, 1]$.
Does the confidence interval $[0.5685, 1]$ contains 0.5? Ans: NO So, we reject the null.

Connection: testing and CI

If you are testing

$$H_0 : p = 0.5 \qquad H_1 : p > 0.5$$

then if you want to test and reject the null when p-value < 0.05, then you just check if the null hypothesis p, which is 0.5, lies in the one sided confidence interval $[c, 1]$.
Does the confidence interval $[0.5685, 1]$ contains 0.5? Ans: NO So, we reject the null.
If you test

$$H_0 : p = 0.5 \qquad H_1 : p \neq 0.5$$

Connection: testing and CI

If you are testing

$$H_0 : p = 0.5 \qquad H_1 : p > 0.5$$

then if you want to test and reject the null when p-value < 0.05, then you just check if the null hypothesis p, which is 0.5, lies in the one sided confidence interval $[c, 1]$.
Does the confidence interval $[0.5685, 1]$ contains 0.5? Ans: NO So, we reject the null.
If you test

$$H_0 : p = 0.5 \qquad H_1 : p \neq 0.5$$

Then check if the 2-sided 95% interval $[a, b]$ contains 0.5, here we observe that $[0.554, 0.725]$ does not contain 0.5. so, again we reject the null.

Connection: testing and CI

If you are testing

$$H_0 : p = 0.5 \qquad H_1 : p > 0.5$$

then if you want to test and reject the null when p-value < 0.05, then you just check if the null hypothesis p, which is 0.5, lies in the one sided confidence interval $[c, 1]$.
Does the confidence interval $[0.5685, 1]$ contains 0.5? Ans: NO So, we reject the null.
If you test

$$H_0 : p = 0.5 \qquad H_1 : p \neq 0.5$$

Then check if the 2-sided 95% interval $[a, b]$ contains 0.5, here we observe that $[0.554, 0.725]$ does not contain 0.5. so, again we

Connection: testing and CI

If you are testing

$$H_0 : p = 0.5 \qquad H_1 : p > 0.5$$

then if you want to test and reject the null when p-value < 0.05, then you just check if the null hypothesis p, which is 0.5, lies in the one sided confidence interval $[c, 1]$.
Does the confidence interval $[0.5685, 1]$ contains 0.5? Ans: NO So, we reject the null.
If you test

$$H_0 : p = 0.5 \qquad H_1 : p \neq 0.5$$

Then check if the 2-sided 95% interval $[a, b]$ contains 0.5, here we observe that $[0.554, 0.725]$ does not contain 0.5. so, again we

Confidence intervals

- **Boss**: Great! So, I am going to write *I am pretty sure that there is a tendency of people leaning to the right while kissing and I am* 95% *confident that this probability lies in* $[0.5685, 1]$ *if I assume the interval to be one-sided and in* $[0.554, 0.725]$ *if the interval is two-sided*. How does that sound?

Confidence intervals

- **Boss**: Great! So, I am going to write *I am pretty sure that there is a tendency of people leaning to the right while kissing and I am* 95% *confident that this probability lies in* $[0.5685, 1]$ *if I assume the interval to be one-sided and in* $[0.554, 0.725]$ *if the interval is two-sided*. How does that sound?
- **You**: Sounds very good Sir. But you should mention 95% confident under assumption of normality/ t-distribution.

Confidence intervals

- **Boss**: Great! So, I am going to write *I am pretty sure that there is a tendency of people leaning to the right while kissing and I am* 95% *confident that this probability lies in* $[0.5685, 1]$ *if I assume the interval to be one-sided and in* $[0.554, 0.725]$ *if the interval is two-sided*. How does that sound?
- **You**: Sounds very good Sir. But you should mention 95% confident under assumption of normality/ t-distribution.
- **Boss**: What other alternative did we have? Can the conclusion change if we drop the assumption?

Confidence intervals

- **Boss**: Great! So, I am going to write *I am pretty sure that there is a tendency of people leaning to the right while kissing and I am* 95% *confident that this probability lies in* $[0.5685, 1]$ *if I assume the interval to be one-sided and in* $[0.554, 0.725]$ *if the interval is two-sided*. How does that sound?
- **You**: Sounds very good Sir. But you should mention 95% confident under assumption of normality/ t-distribution.
- **Boss**: What other alternative did we have? Can the conclusion change if we drop the assumption?
- **You**: The alternative is called Bootstrap confidence interval and the conclusion may very well change sir, especially for non-normal distributed random variables with small sample size.

Confidence intervals

- **Boss**: Great! So, I am going to write *I am pretty sure that there is a tendency of people leaning to the right while kissing and I am* 95% *confident that this probability lies in* $[0.5685, 1]$ *if I assume the interval to be one-sided and in* $[0.554, 0.725]$ *if the interval is two-sided*. How does that sound?
- **You**: Sounds very good Sir. But you should mention 95% confident under assumption of normality/ t-distribution.
- **Boss**: What other alternative did we have? Can the conclusion change if we drop the assumption?
- **You**: The alternative is called Bootstrap confidence interval and the conclusion may very well change sir, especially for non-normal distributed random variables with small sample size.
- **Boss**: Why don't you calculate this Bootstrap CI and check?

Confidence intervals

- **Boss**: Great! So, I am going to write *I am pretty sure that there is a tendency of people leaning to the right while kissing and I am* 95% *confident that this probability lies in* $[0.5685, 1]$ *if I assume the interval to be one-sided and in* $[0.554, 0.725]$ *if the interval is two-sided*. How does that sound?
- **You**: Sounds very good Sir. But you should mention 95% confident under assumption of normality/ t-distribution.
- **Boss**: What other alternative did we have? Can the conclusion change if we drop the assumption?
- **You**: The alternative is called Bootstrap confidence interval and the conclusion may very well change sir, especially for non-normal distributed random variables with small sample size.
- **Boss**: Why don't you calculate this Bootstrap CI and check?

General idea: Bootstrapping/ Resampling

Let $X_1, X_2, \cdots, X_n$ be data corresponding to n samples coming from a distribution F, where F is not unknown and there is no reason to assume normality.

General idea: Bootstrapping/ Resampling

Let $X_1, X_2, \cdots, X_n$ be data corresponding to n samples coming from a distribution F, where F is not unknown and there is no reason to assume normality.

Suppose we want to find the sampling distribution of $\bar{X}$.

General idea: Bootstrapping/ Resampling

Let $X_1, X_2, \cdots, X_n$ be data corresponding to n samples coming from a distribution F, where F is not unknown and there is no reason to assume normality.

Suppose we want to find the sampling distribution of $\bar{X}$.

We know if n is large, then this sampling distribution is normal, and

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \textit{for} \;\; \textit{large} \;\; n$$

where μ and σ are the population mean and population standard deviation under the population distribution or model F (again may not be normal).

General idea: Bootstrapping/ Resampling

Let $X_1, X_2, \cdots, X_n$ be data corresponding to n samples coming from a distribution F, where F is not unknown and there is no reason to assume normality.

Suppose we want to find the sampling distribution of $\bar{X}$.

We know if n is large, then this sampling distribution is normal, and

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \textit{for} \;\; \textit{large} \;\; n$$

where μ and σ are the population mean and population standard deviation under the population distribution or model F (again may not be normal).

If n is not so large though, normality is not a good assumption for sampling distribution.

Bootstrapping

This is a special technique of resampling.

Bootstrapping

This is a special technique of resampling. This is useful:

- When the theoretical distribution of a statistic of interest is complicated or unknown.

Bootstrapping

This is a special technique of resampling. This is useful:

- When the theoretical distribution of a statistic of interest is complicated or unknown.
- When the sample size is insufficient for straightforward statistical inference.

Bootstrapping

This is a special technique of resampling. This is useful:

- When the theoretical distribution of a statistic of interest is complicated or unknown.
- When the sample size is insufficient for straightforward statistical inference.

Bootstrapping to find CI

You want to find the CI for μ, the population mean.

Bootstrapping to find CI

You want to find the CI for μ, the population mean.

- Take 1000 resamples and evaluate the sample means for them.

Bootstrapping to find CI

You want to find the CI for μ, the population mean.

- Take 1000 resamples and evaluate the sample means for them.
- Those 1000 values of the means give you the *Bootstrap distribution*.

Bootstrapping to find CI

You want to find the CI for μ, the population mean.

- Take 1000 resamples and evaluate the sample means for them.
- Those 1000 values of the means give you the *Bootstrap distribution*.
- Find the 2.5th and the 97.5th quantile from that distribution.

Bootstrapping to find CI

You want to find the CI for μ, the population mean.

- Take 1000 resamples and evaluate the sample means for them.
- Those 1000 values of the means give you the *Bootstrap distribution*.
- Find the 2.5th and the 97.5th quantile from that distribution.
- In other words, sort those values and find 25th and 975th values. Those will give you the CI for the mean.

Bootstrapping to find CI

You want to find the CI for μ, the population mean.

- Take 1000 resamples and evaluate the sample means for them.
- Those 1000 values of the means give you the *Bootstrap distribution*.
- Find the 2.5th and the 97.5th quantile from that distribution.
- In other words, sort those values and find 25th and 975th values. Those will give you the CI for the mean.

You can use this technique for any parameter of interest!

Bootstrapping to find p-value

You want to find the p-value for a test involving test statistic T, say. Suppose you are interested in one-sided alternative (greater than).

Bootstrapping to find p-value

You want to find the p-value for a test involving test statistic T, say. Suppose you are interested in one-sided alternative (greater than).

- For the original sample, evaluate T.

Bootstrapping to find p-value

You want to find the p-value for a test involving test statistic T, say. Suppose you are interested in one-sided alternative (greater than).

- For the original sample, evaluate T.
- Take 1000 resamples and evaluate the test statistics (T) for them.

Bootstrapping to find p-value

You want to find the p-value for a test involving test statistic T, say. Suppose you are interested in one-sided alternative (greater than).

- For the original sample, evaluate T.
- Take 1000 resamples and evaluate the test statistics (T) for them.
- Those 1000 values of the means give you the *Bootstrap distribution* for T.

Bootstrapping to find p-value

You want to find the p-value for a test involving test statistic T, say. Suppose you are interested in one-sided alternative (greater than).

- For the original sample, evaluate T.
- Take 1000 resamples and evaluate the test statistics (T) for them.
- Those 1000 values of the means give you the *Bootstrap distribution* for T.
- Find how many of them are more extreme than T from the original sample.

Bootstrapping to find p-value

You want to find the p-value for a test involving test statistic T, say. Suppose you are interested in one-sided alternative (greater than).

- For the original sample, evaluate T.
- Take 1000 resamples and evaluate the test statistics (T) for them.
- Those 1000 values of the means give you the *Bootstrap distribution* for T.
- Find how many of them are more extreme than T from the original sample.
- The proportion gives you the estimated p-value.

Bootstrapping to find p-value

You want to find the p-value for a test involving test statistic T, say. Suppose you are interested in one-sided alternative (greater than).

- For the original sample, evaluate T.
- Take 1000 resamples and evaluate the test statistics (T) for them.
- Those 1000 values of the means give you the *Bootstrap distribution* for T.
- Find how many of them are more extreme than T from the original sample.
- The proportion gives you the estimated p-value.
- For CI, find the region that contains 95% of the T in the Bootstrap samples (for 2-sided the region between the 2.5 th quantile and 97.5 th quantile of the Bootstrap T samples, for 1-sided the region beyond 5 th quantile)

You can use this technique to test your hypothesis or build

```
obs.samp <- c(rep(1,80), rep(0, 44));
bootstrap_counts <- replicate(1000, sum(sample(obs.samp, replace
summary(bootstrap_counts)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   65.0    77.0    80.0    80.2    84.0    95.0

bootstrap_prop <- bootstrap_counts/124;
## 2-sided

quantile(bootstrap_prop, 0.025)

  2.5%
0.5645

quantile(bootstrap_prop, 0.975)

 97.5%
0.7258

## 1-sided
quantile(bootstrap_prop, 0.05)

    5%
0.5806
```

- **You**: Sir, the Bootstap CIs (both 2-sided and 1-sided) are very close to the normal or t-distribution CIs.

- **You**: Sir, the Bootstap CIs (both 2-sided and 1-sided) are very close to the normal or t-distribution CIs.
- **Boss**: Okay, here is the final report then: *We find strong evidence that there is indeed a tendency to lean to the right while kissing (*95% *2-sided CI under normality:* $[0.554, 0.725]$*, Bootstrap CI:* $[0.564, 0.725]$ *and* 95% *1-sided CI under normality:* $[0.5685, 1]$*, Bootstrap CI:* $[0.5806, 1]$*)*.

- **You**: Sir, the Bootstap CIs (both 2-sided and 1-sided) are very close to the normal or t-distribution CIs.
- **Boss**: Okay, here is the final report then: *We find strong evidence that there is indeed a tendency to lean to the right while kissing (*95% *2-sided CI under normality:* $[0.554, 0.725]$*, Bootstrap CI:* $[0.564, 0.725]$ *and* 95% *1-sided CI under normality:* $[0.5685, 1]$*, Bootstrap CI:* $[0.5806, 1]$*).*
- **You**: That sounds perfect now!

- **You**: Sir, the Bootstap CIs (both 2-sided and 1-sided) are very close to the normal or t-distribution CIs.
- **Boss**: Okay, here is the final report then: *We find strong evidence that there is indeed a tendency to lean to the right while kissing (*95% *2-sided CI under normality:* $[0.554, 0.725]$*, Bootstrap CI:* $[0.564, 0.725]$ *and* 95% *1-sided CI under normality:* $[0.5685, 1]$*, Bootstrap CI:* $[0.5806, 1]$*)*.
- **You**: That sounds perfect now!
- **Boss**: Okay let me submit that then. Thanks.

- **You**: Sir, the Bootstap CIs (both 2-sided and 1-sided) are very close to the normal or t-distribution CIs.
- **Boss**: Okay, here is the final report then: *We find strong evidence that there is indeed a tendency to lean to the right while kissing (*95% *2-sided CI under normality:* $[0.554, 0.725]$*, Bootstrap CI:* $[0.564, 0.725]$ *and* 95% *1-sided CI under normality:* $[0.5685, 1]$*, Bootstrap CI:* $[0.5806, 1]$*)*.
- **You**: That sounds perfect now!
- **Boss**: Okay let me submit that then. Thanks.
- To be continued....

Suppose we want to test

$$H_0 : p = 0.5 \quad H_1 : p = 0.8$$

Suppose we want to test

$$H_0 : p = 0.5 \;\; H_1 : p = 0.8$$

In this case, both the hypothesis are *point* hypothesis.

Suppose we want to test

$$H_0 : p = 0.5 \;\; H_1 : p = 0.8$$

In this case, both the hypothesis are *point* hypothesis.
When we encounter such hypothesis? Case: Another researcher has claimed the probability of leaning to the right while kissing is 80%, you want to test this claim.
How do we test such a hypothesis?

Suppose we want to test

$$H_0 : p = 0.5 \;\; H_1 : p = 0.8$$

In this case, both the hypothesis are *point* hypothesis.
When we encounter such hypothesis? Case: Another researcher has claimed the probability of leaning to the right while kissing is 80%, you want to test this claim.
How do we test such a hypothesis?
We define 2 types of errors.

Suppose we want to test

$$H_0 : p = 0.5 \ \ H_1 : p = 0.8$$

In this case, both the hypothesis are *point* hypothesis.
When we encounter such hypothesis? Case: Another researcher has claimed the probability of leaning to the right while kissing is 80%, you want to test this claim.
How do we test such a hypothesis?
We define 2 types of errors.

	H_0 is true	H_a is true
Reject H_0	Type I error α	Correct decision $1-\beta=\pi$
Not reject H_0	Correct decision $1-\alpha$	Type II error β

- $\alpha = P(\text{Type I Error}) = P(\text{Reject } H_0 | H_0 \text{ is true}) =$ Significance level
- $\beta = P(\text{Type II Error}) = P(\text{Not reject } H_0 | H_a \text{ is true})$

- We also define **power** as

$$\begin{aligned}\pi &= 1-\beta \\ &= P(\text{Reject } H_0 | H_a \text{ is true}) \\ &= \text{Ability to Detect the Alternative Hypothesis}\end{aligned}$$

$$\alpha = Pr(\text{Type I Error}) = Pr(\text{Reject } H_0 | H_0 \text{ is true})$$

$$\alpha = Pr(\text{Reject } H_0 | p = 0.5)$$

Assume we reject H_0 if p-value is < 0.05, this means if true p under null hypothesis H_0 does not lie in the 95% confidence interval, put $\hat{p} = 0.64$, $p = 0.5$,

$$Pr\left[p - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}} < \hat{p} < p + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] = [0.552, 0.728]$$

$p = 0.5$ does not belong in this