PLSC 502 – Fall 2024
Statistical Inference

October 28, 2024

Where we've been:

- Sampling from a population
- Obtaining an estimate $\hat{\theta}$ of a parameter θ
- Understanding $\hat{\theta}$'s properties

What we'll do today:

- Confidence Intervals
- Hypothesis Testing

The goal: **Quantifying uncertainty about** $\hat{\theta}$.

A range of values for $\hat{\theta}$ (say, $[\hat{\theta}_L, \hat{\theta}_U]$) for which:

- $\Pr(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U)$ is high, and
- $\hat{\theta}_L - \hat{\theta}_U$ is small.

These two goals are fundamentally at odds with each other...

Define:

$$\Pr(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha,$$

"Pivotal method": One where $\hat{\theta}$:

- is a function *only* of the sample data and the population parameter θ, and
- has a sampling distribution that *does not* depend on θ.

This allows us to:

- ...create confidence intervals using simple linear transformations, and
- ..."invert" C.I.s to do hypothesis testing (and vice-versa...).

Constructing C.I.s: The Mean

Recall that:

$$\bar{X} \sim \mathcal{N}(\mu, \sigma^2_{\bar{X}})$$

Because $\mathsf{E}(\bar{X}) = \mu$, we use $\bar{X}$ as the "center" of our C.I.

Suppose $\alpha = 0.05$, so $1 - \alpha = 0.95$. Then

$$\Pr(\bar{X}_L \leq \mu \leq \bar{X}_U) = 0.95$$

Then choose:

$$\Pr(\mu < \bar{X}_L) = \int_{-\infty}^{\bar{X}_L} \phi_{\bar{X}}(u)\,du = 0.025$$

and

$$\Pr(\mu > \bar{X}_H) = \int_{\bar{X}_H}^{\infty} \phi_{\bar{X}}(u)\,du = 0.025.$$

An estimator $\hat{\theta}$ that is:

$$\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2_{\hat{\theta}})$$

implies:

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \sim \mathcal{N}(0, 1)$$

which means that:

$$\begin{aligned} 1 - \alpha &= \Pr\left(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}\right) \\ &= \Pr\left(-z_{\alpha/2}\sigma_{\hat{\theta}} \leq \hat{\theta} - \theta \leq z_{\alpha/2}\sigma_{\hat{\theta}}\right) \\ &= \Pr\left(-\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq -\theta \leq -\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) \\ &= \Pr\left(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \ \theta \ \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) \end{aligned}$$

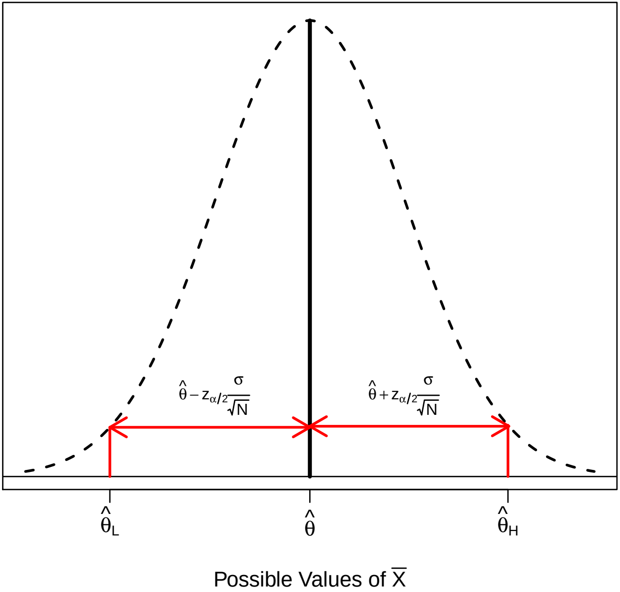
This means that

$$[\hat{\theta}_L, \hat{\theta}_U] = \left[\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}},\ \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right]$$

Steps:

- Select your level of confidence $1 - \alpha$,
- Calculate the sample statistic $\hat{\theta}$,
- Calculate the z-value associated with the $1 - \alpha$ level of confidence,
- Divide that z-value by $\sigma_{\hat{\theta}}$, the standard error of the sampling statistic, and
- Construct the confidence interval according to the above equation.

C.I.s, Illustrated



We have

$$\hat{\theta} = \hat{\pi} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

and

$$\sigma^2_{\hat{\pi}} = \frac{\pi(1-\pi)}{N}$$

so that

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{N}}.$$

We know that:

$$\hat{\pi} \sim \mathcal{N}(\pi, \sigma^2_{\hat{\pi}})$$

Implies:

$$\hat{\pi}_L = \hat{\pi} - z_{\alpha/2} \left[\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N}} \right]$$

and

$$\hat{\pi}_U = \hat{\pi} + z_{\alpha/2} \left[\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N}} \right].$$

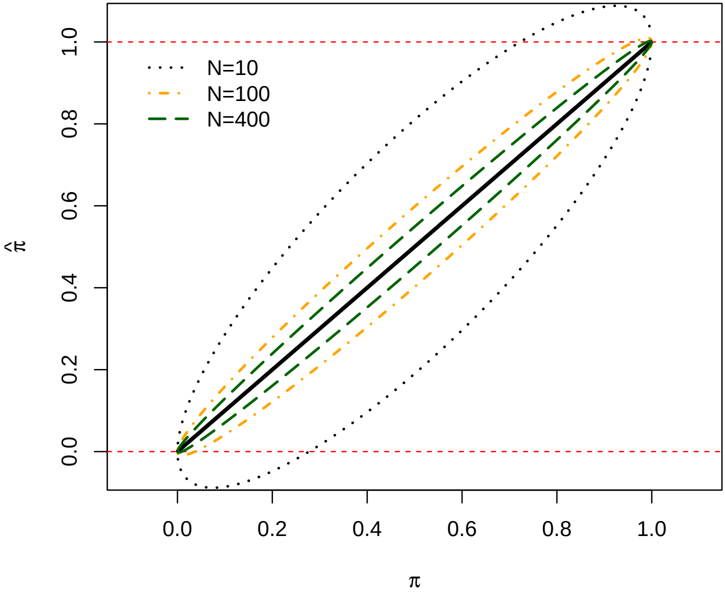
For $N = 20$ and $\hat{\pi} = 0.390$, we have:

$$\begin{aligned}\hat{\pi}_L &= 0.390 - 1.96\left[\sqrt{\frac{0.39(0.61)}{20}}\right] \\ &= 0.390 - 0.214 \\ &= \mathbf{0.176}\end{aligned}$$

and

$$\begin{aligned}\hat{\pi}_U &= 0.390 + 1.96\left[\sqrt{\frac{0.39(0.61)}{20}}\right] \\ &= 0.390 + 0.214 \\ &= \mathbf{0.604}.\end{aligned}$$

C.I.s for Proportions



How Did That Happen?

Some code:

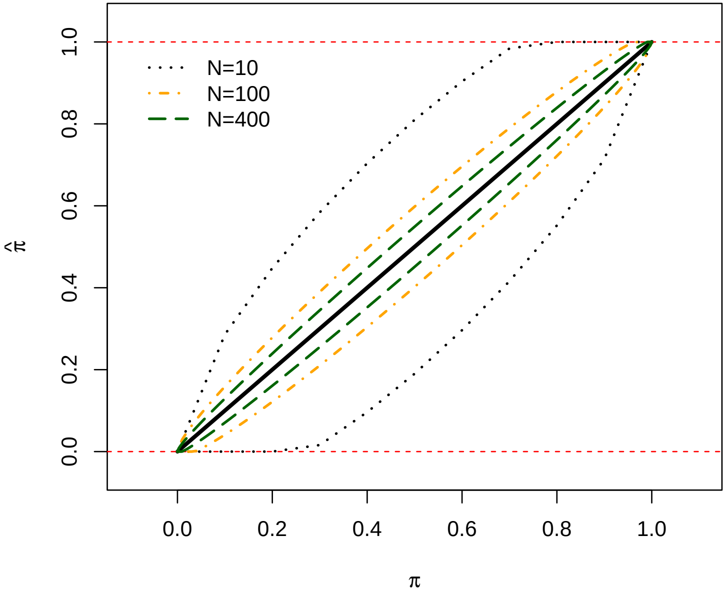
```
> Pi<-seq(0.001,0.999,by=0.001)    # Population value

> ub10 <- Pi + 1.96*(sqrt((Pi*(1-Pi))/(10))) # N=10
> lb10 <- Pi - 1.96*(sqrt((Pi*(1-Pi))/(10)))
> ub100 <- Pi + 1.96*(sqrt((Pi*(1-Pi))/(100))) # N=100
> lb100 <- Pi - 1.96*(sqrt((Pi*(1-Pi))/(100)))
> ub400 <- Pi + 1.96*(sqrt((Pi*(1-Pi))/(400))) # N=400
> lb400 <- Pi - 1.96*(sqrt((Pi*(1-Pi))/(400)))
```

Let's try something different:

```
CI10<-data.frame(BinomCI(c(0:10),10,method="wald"))    # N=10
CI100<-data.frame(BinomCI(c(0:100),100,method="wald")) # N=100
CI400<-data.frame(BinomCI(c(0:400),400,method="wald")) # N=400
```

C.I.s for Proportions Redux



Alternatives

"Wilson" (1927):

$$\hat{\pi}_L = \frac{\hat{\pi} + \frac{z^2_{1-\alpha/2}}{2N} + z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N} + \frac{z^2_{1-\alpha/2}}{4N^2}}}{1 + \frac{z^2_{1-\alpha/2}}{N}}$$

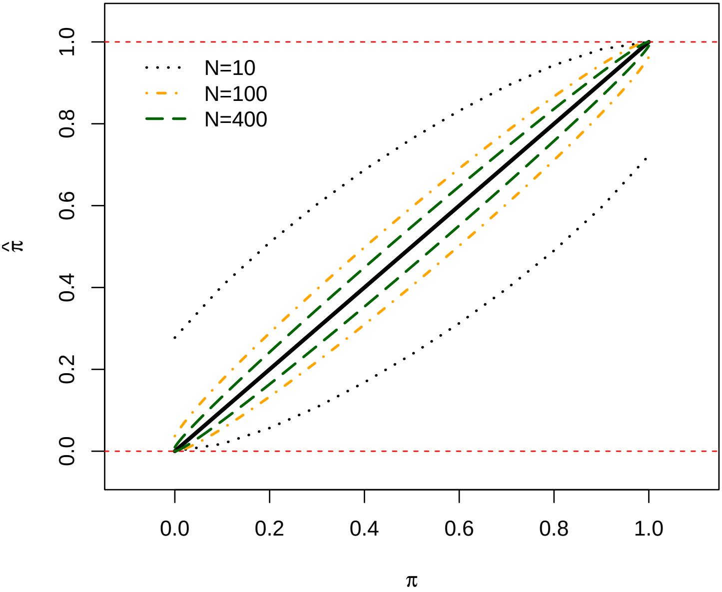
and:

$$\hat{\pi}_U = \frac{\hat{\pi} + \frac{z^2_{\alpha/2}}{2N} + z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N} + \frac{z^2_{\alpha/2}}{4N^2}}}{1 + \frac{z^2_{\alpha/2}}{N}}.$$

Other alternatives:

- "Exact" binomial methods
- Many, many others...

C.I.s for Proportions Re-Redux



Consider:

$$T = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}}$$

As $N \to \infty$, $\hat{\sigma} \to \sigma$ and so $T \to \mathcal{N}(0, 1)$.

However, in small samples,

$$\hat{\sigma}/\sqrt{N} \sim \chi^2_N$$

which means

$$[\bar{X}_L, \bar{X}_U] = \bar{X} \pm t_{\alpha/2} \left(\frac{\hat{\sigma}}{\sqrt{N}} \right).$$

Talking About C.I.s

"$[(1-\alpha)\times 100]\%$ of all confidence intervals constructed from independent simple random samples will contain the population parameter θ, and $(\alpha \times 100)\%$ of them will not."

Never "There is a 95% chance that our confidence interval contains the true population value θ."

Example: SCOTUS Cases, 1946-2020

Data from the Supreme Court Judicial Database:

- All SCOTUS decisions, OT1946-2023 ($N \approx 10800$)
- 50+ variables about each decision
- `Constitutional` $= 1$ if the decision is a constitutional one, 0 otherwise

```
> summary(df$Constitutional)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   0.000   0.000   0.311   1.000   1.000
```

One sample, $N = 20$

Code:

```
> set.seed(7222009)

> DFsample <- with(df, sample(Constitutional,20,replace=F))

> summary(DFsample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0     0.0     0.0     0.4     1.0     1.0

> BinomCI(sum(DFsample),length(DFsample))
     est  lwr.ci  upr.ci
[1,] 0.4 0.21881 0.61342
```

<u>So for this sample</u>:

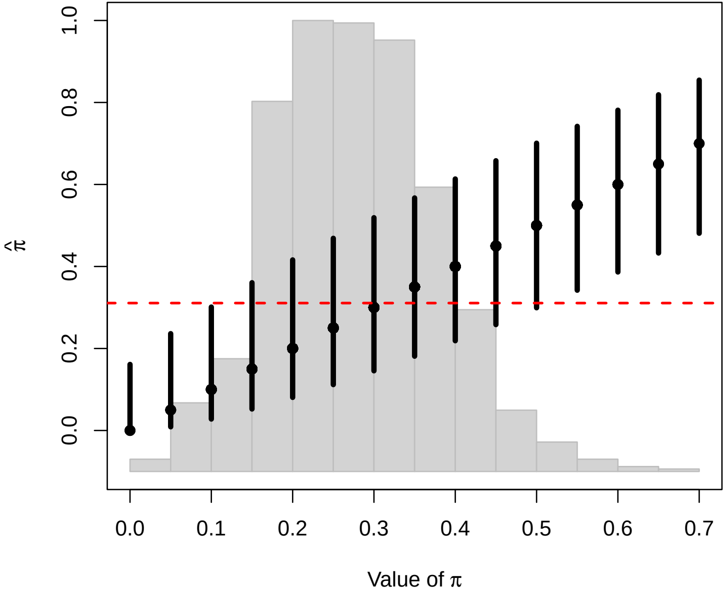
- $\pi = 0.311$
- $\hat{\pi} = 0.4$
- $\widehat{\text{C.I.}} = [0.219, 0.613]$

“Coverage”

How many times (out of 1000) does a C.I. created from a random sample with $N = 20$ contain the population value $\mu = 0.311$?

```
N <- 20
reps <- 1000
PI20 <- numeric(reps)
UB20<-numeric(reps)
LB20<-numeric(reps)
set.seed(7222009)
for (i in 1:reps) {
  foo <- with(df, sample(Constitutional,N,replace=F))
  bar <- data.frame(BinomCI(sum(foo),length(foo)))
  PI20[i] <- bar$est
  LB20[i] <- bar$lwr.ci
  UB20[i] <- bar$upr.ci
}
```

Coverage, $N = 20$

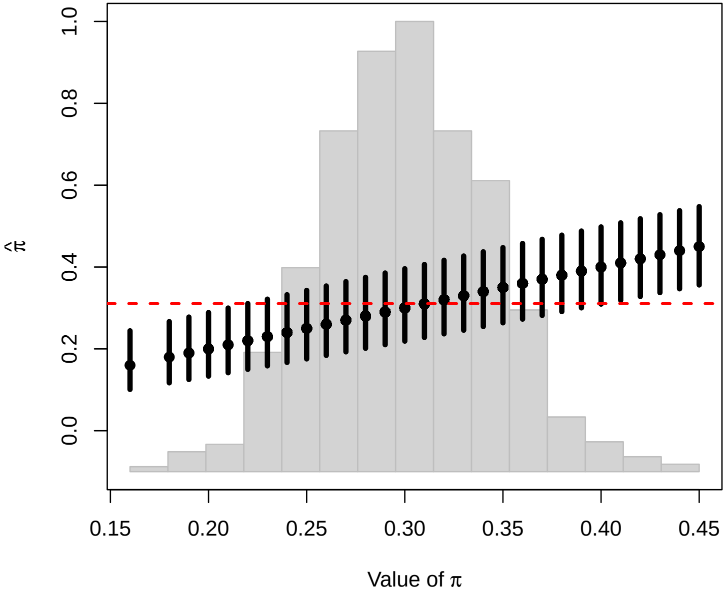


Coverage, $N = 100$

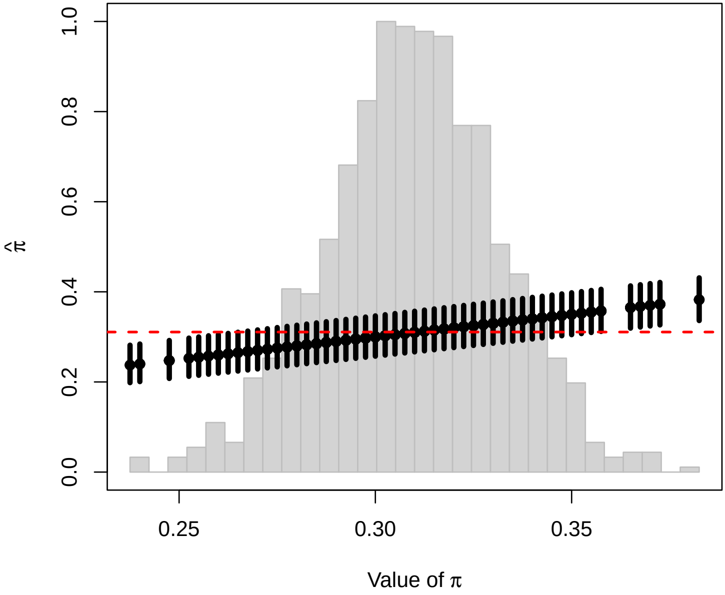
How about for $N = 100$?

```
N <- 100
reps <- 1000
PI100 <- numeric(reps)
UB100<-numeric(reps)
LB100<-numeric(reps)
set.seed(7222009)
for (i in 1:reps) {
  foo <- with(df, sample(Constitutional,N,replace=F))
  bar <- data.frame(BinomCI(sum(foo),length(foo)))
  PI100[i] <- bar$est
  LB100[i] <- bar$lwr.ci
  UB100[i] <- bar$upr.ci
}
```

Coverage, $N = 100$



Coverage, $N = 400$



What proportion of our 95% C.I.s contain the population mean of `Constitutional`?

```
> N = 20:
>
> prop.table(table(ifelse(UB20>popmean & LB20<popmean,1,0)))

    0     1
0.053 0.947


> # N = 100:
>
> prop.table(table(ifelse(UB100>popmean & LB100<popmean,1,0)))

    0     1
0.042 0.958


> # N = 400:
>
> prop.table(table(ifelse(UB400>popmean & LB400<popmean,1,0)))

    0     1
0.045 0.955
```

Hypothesis Testing

Hypothesis Testing: Concepts

Conventional *hypothesis testing* requires:

- A *null hypothesis*, H_0
- an *alternative hypothesis* H_a
- a *test statistic* $\theta = f(\mathbf{X})$
- a *rejection region* in the range of θ.

Example

October 18, 2016 Quinnipiac poll in PA:

- Clinton = 47 percent
- Trump = 41 percent
- $N = 660$ likely voters.

Hypothesis:

$$H_a : \pi > 0.5$$

Corresponding null:

$$H_0 : \pi = 0.5$$

Test statistic: $\hat{\pi} = 0.47$

- **Type I error** = "false positive."
- **Type II error** = "false negative."

	Reality / Population	
Test Statistic / Sample	H_a	H_0
H_a	Correct	Type I error
H_0	Type II Error	Correct

Alphas and Significance

By convention:

$$\Pr(\text{Type I Error}) = \alpha \text{ (``significance level'')}$$

and

$$1 - \alpha = \text{``specificity''}$$

While

$$\Pr(\text{Type II Error}) = \beta$$

and

$$1 - \beta = \text{``sensitivity.''}$$

A New Table

	Reality / Population		
Sample Result	Positive	Negative	Frequency
Positive	True Positive (N_{TP})	Type I error (False Positive) (N_{FP})	$N_P = N_{TP} + N_{FP}$
Negative	Type II Error (False Negative) (N_{FN})	True Negative (N_{TN})	$N_N = N_{TN} + N_{FN}$
Frequency	$N_{(+)} = N_{TP} + N_{FN}$	$N_{(-)} = N_{TN} + N_{FP}$	N

Components...

- *False positive / significance rate* $(\alpha) = N_{FP}/N_{(-)}$,
- *False negative rate* $(\beta) = N_{FN}/N_{(+)}$,
- *False discovery rate* $= N_{FP}/N_P$,
- *False omission rate* $= N_{FN}/N_N$,
- *Accuracy* $= (N_{TP} + N_{FP})/N$

In the Clinton / PA example, we know:

$$\hat{\pi} \sim \mathcal{N}(\pi, \sigma^2_{\hat{\pi}}).$$

and

$$\begin{aligned} \hat{\sigma}^2 &= 0.470(1 - 0.470) \\ &= 0.249 \end{aligned}$$

and

$$\begin{aligned} \hat{\sigma}^2_{\hat{\pi}} &= \frac{0.249}{660} \\ &= 0.00038. \end{aligned}$$

That means:

$$\hat{\pi} \sim \mathcal{N}(0.5, 0.00038).$$

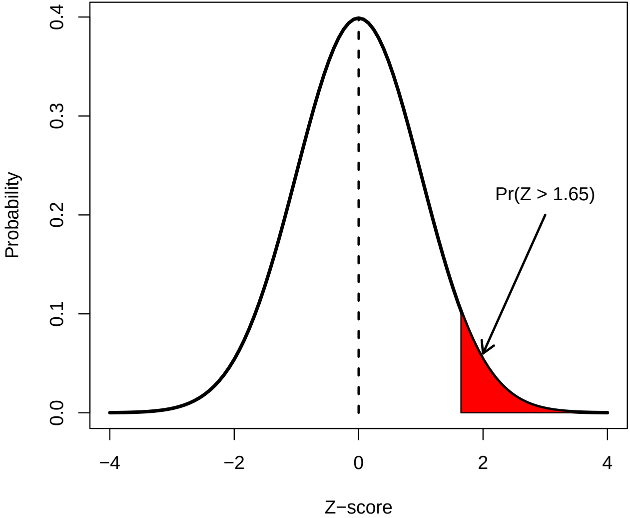
Converting $\hat{\pi}$ to z:

$$\frac{\hat{\pi} - \pi}{\sigma_{\hat{\pi}}} = Z \sim \mathcal{N}(0, 1).$$

Decision rule:

$$\text{Reject } H_0 \text{ if } Z \geq z_\alpha.$$

$$\alpha = 0.05 \rightarrow Z \geq 1.65$$



Example, Continued

Here,

$$Z = \frac{0.470 - 0.50}{0.0195} = -1.54$$

so we fail to reject the null.

Another alternative $H_0 : \pi = 0.40$ yields

$$\begin{aligned} Z &= \frac{0.47 - 0.40}{0.0195} \\ &= 3.59 \end{aligned}$$

meaning we would reject that null.

Tailedness

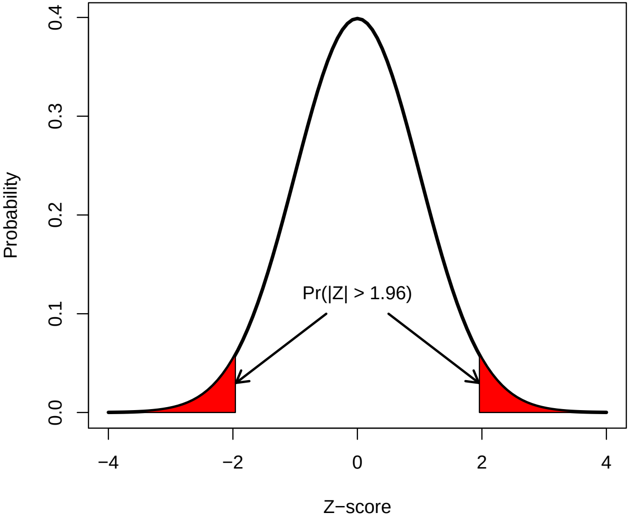
An alternative formulation:

$$H_a : \pi \neq 0.5.$$

Alternative decision rule:

$$\text{Reject } H_0 \text{ if } |Z| \geq z_{\alpha/2}$$

$$\alpha = 0.05 \rightarrow |Z| \geq 1.96$$



P-Values Versus Significance Tests

P-value (or "attained significance level"): *The smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.*

Why we like them:

- **Avoid arbitrary "cutoffs."**
- **Provide more information.**

Significance Tests and Confidence Intervals

C.I.:

$$\text{c.i.}_\alpha = \hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

vs. test:

$$|Z| \equiv \left|\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}\right| \geq z_{\alpha/2}.$$

Significance Tests and Confidence Intervals

"Acceptance region":

$$-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2} = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}},$$

i.e., "Do not reject H_0 at P $= \alpha$ if θ lies within a $(1-\alpha) \times 100$-percent confidence interval around $\hat{\theta}$, and reject H_0 if it does not."

1. P-**values are not "the probability that the null hypothesis is false."**

BAD:

> *"The test statistic allows us to reject the null hypothesis at* $P < 0.01$*, indicating that there is a less than one in 100 chance that the null hypothesis is true."*

GOOD:

> *"The test statistic allows us to reject the null hypothesis at* $P < 0.01$*, which is strong evidence that the observed result is not due to chance."*

2. **One does not, in general, “accept” the null hypothesis.**

BAD:

> *“The P-value for the regression coefficient on Female is 0.56, indicating that there is no relationship between gender and support for immigrants' rights.”*

GOOD:

> *“The P-value for the regression coefficient on Female is 0.56, indicating the data do not support the hypothesized relationship between gender and support for immigrants' rights.”*

3. P-**values are not the long-run frequency of a "statistically significant" test statistic.**

BAD:

> *"The P-value of 0.01 means that 99 out of 100 hypothetical replications would reject the null hypothesis."*

Important Things, IV

4. **Statistical significance does not equate to substantive significance.**

```
> data<-read_csv("https://raw.githubusercontent.com/PrisonRodeo/PLSC502-2024-git/
                  master/Exercises/PLSC502-2024-ExerciseFour.csv") # 2008 Centre Voter File
>
> data$DOB <- with(data, as.Date(DateOfBirth,format = "%d%b%Y"))
> data$Sign <- with(data, Zodiac(DOB))

> popmean <- with(data, prop.table(table(Active)))[2]
> popmean
     1
0.8956

> SC<-data[data$Sign=="Scorpio",]
> with(SC,prop.test(sum(Active,na.rm=TRUE),nrow(SC),p=popmean,
+                    correct=FALSE))

 1-sample proportions test without continuity correction

data:  sum(Active, na.rm = TRUE) out of nrow(SC), null probability popmean
X-squared = 3.2, df = 1, p-value = 0.07
alternative hypothesis: true p is not equal to 0.8956
95 percent confidence interval:
 0.8818 0.8962
sample estimates:
     p
0.8892
```

5. **A statistic can never be "significant in the wrong direction."**

BAD:

> *"Our estimate of the effect of trade liberalization on the probability of a civil war – which we expected to be negative – is in fact positive, and statistically significant at* $P = 0.02$*."*

Important Things, VI and VII

6. **Identical P-values are not “better” or “more reliable” if they are based on a larger sample.**

7. **Failing to reject the null hypothesis in a larger sample is a bigger deal than failing to do so in a small one.**