Ch 7 : One-Sample Test of Hypothesis

Contents

July 11, 2017

1

Preliminaries

[ToC]

1.1 Test of Hypothesis

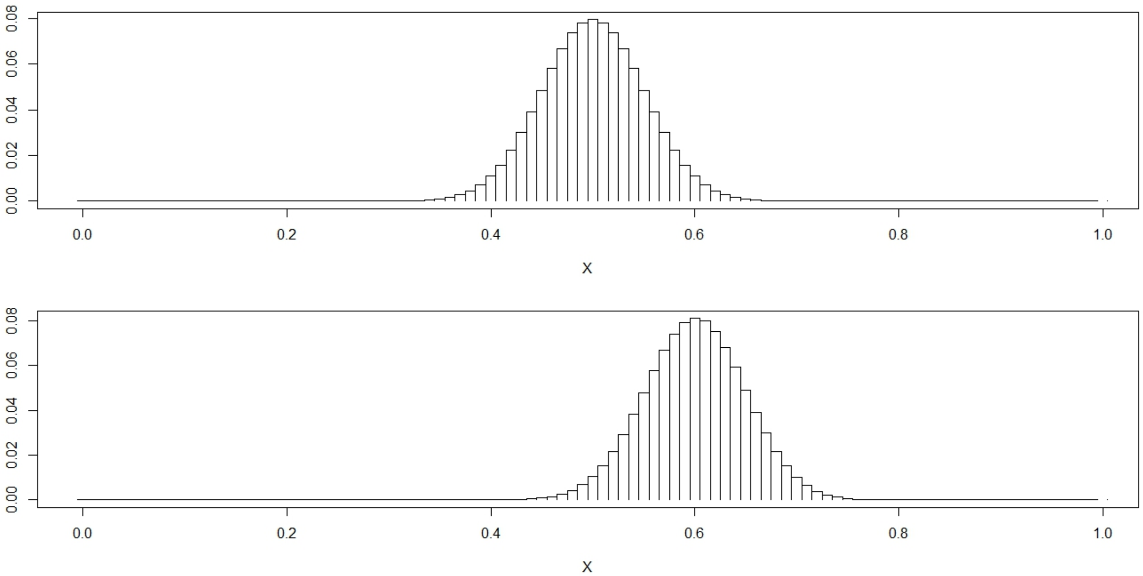
[ToC]

- Suppose we want to test two hypothesis

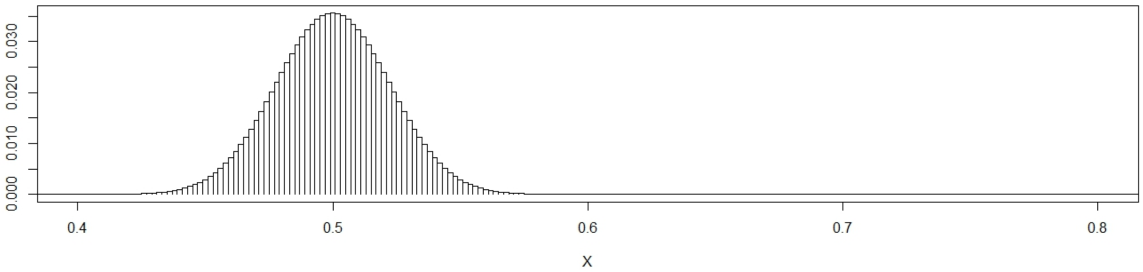
$$
\begin{aligned}
H_0 : &\quad p = .5 \\
H_A : &\quad p = .6
\end{aligned}
$$

- How can we test these hypothesis?

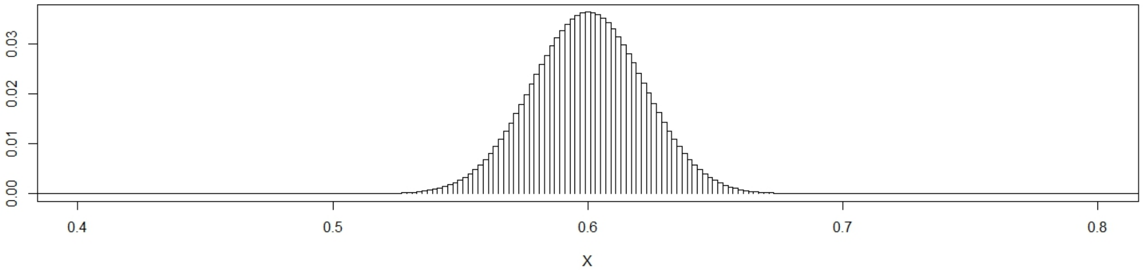
n=100


0.00
0.02
0.04
0.06
0.08
0.0
0.2
0.4
0.6
0.8
1.0
X
0.00
0.02
0.04
0.06
0.08
0.0
0.2
0.4
0.6
0.8
1.0
X

n=500

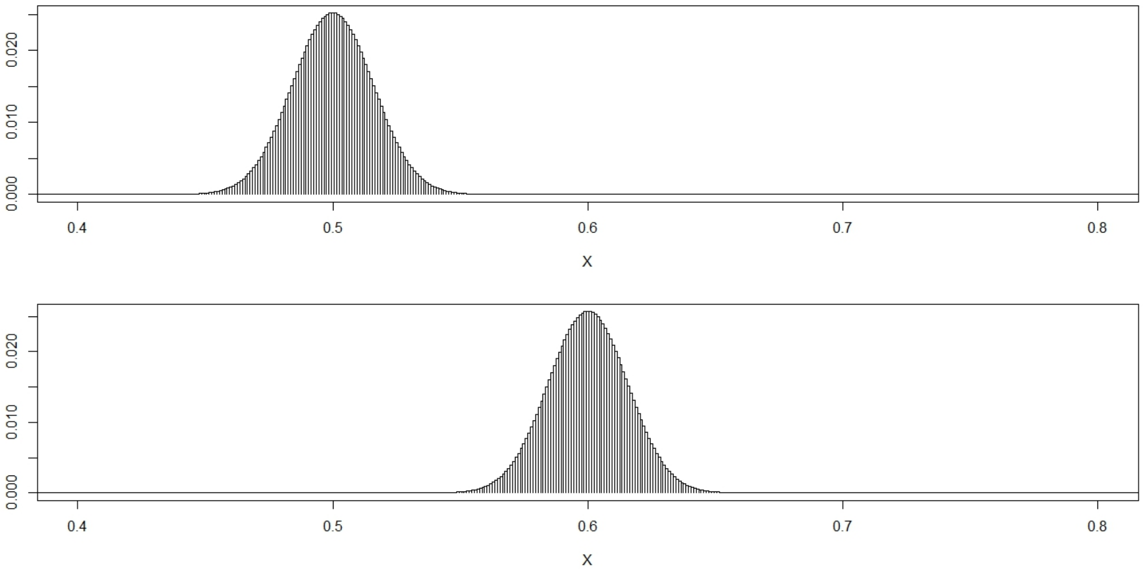


0.000
0.010
0.020
0.030
0.4
0.5
0.6
0.7
0.8
x



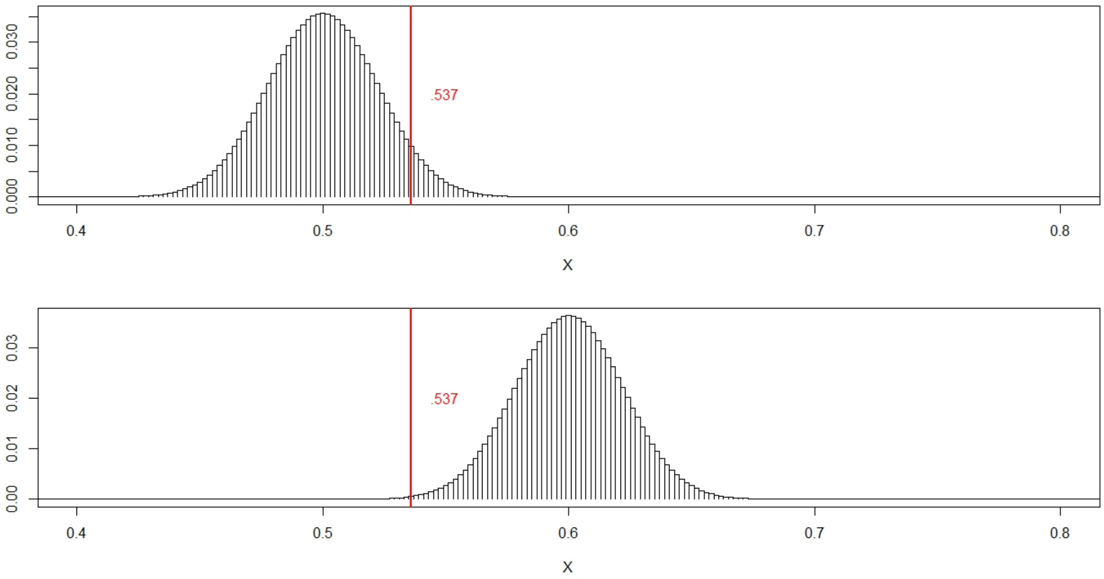
0.00
0.01
0.02
0.03
0.4
0.5
0.6
0.7
0.8
x

n=1000



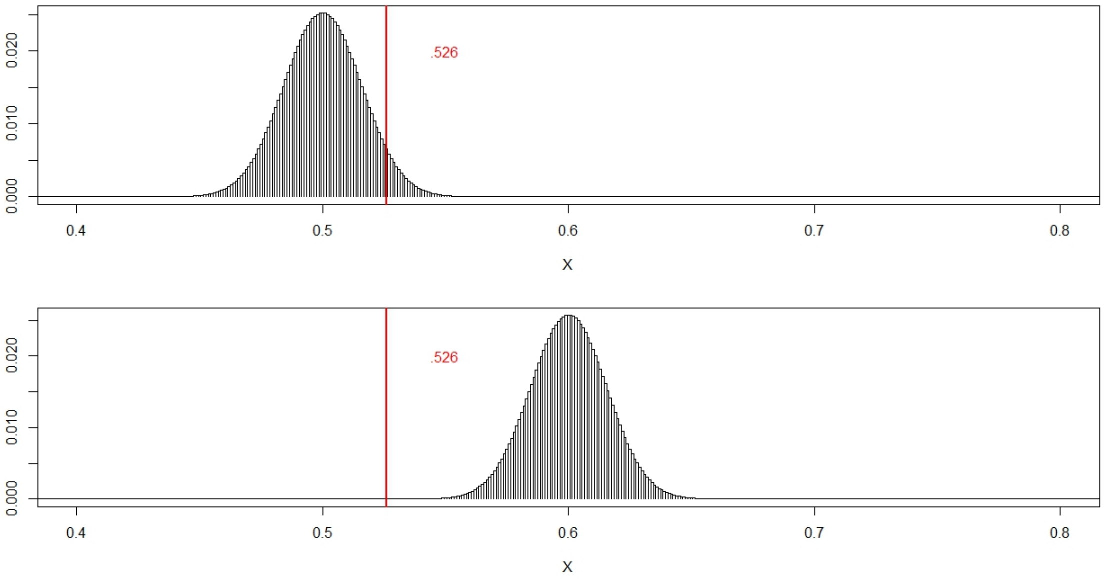
```
x <- 0:100

par(mfrow=c(2,1), mai = c(1, 0.5, 0.2, 0.1))
plot( (x-.5)/100, dbinom(x, 100, .5), type='s', xlab='X', ylab='')
lines((x+.5)/100, dbinom(x, 100, .5), type='h')
plot( (x-.5)/100, dbinom(x, 100, .6), type='s', xlab='X', ylab='')
lines((x+.5)/100, dbinom(x, 100, .6), type='h')

x <- 0:500
plot( (x-.5)/500, dbinom(x, 500, .5), type='s', xlab='X', xlim=c(.4, .8))
lines((x+.5)/500, dbinom(x, 500, .5), type='h')
plot( (x-.5)/500, dbinom(x, 500, .6), type='s', xlab='X', xlim=c(.4, .8))
lines((x+.5)/500, dbinom(x, 500, .6), type='h')

x <- 0:1000
plot( (x-.5)/1000, dbinom(x, 1000, .5), type='s', xlab='X', xlim=c(.4, .8))
lines((x+.5)/1000, dbinom(x, 1000, .5), type='h')
plot( (x-.5)/1000, dbinom(x, 1000, .6), type='s', xlab='X', xlim=c(.4, .8))
lines((x+.5)/1000, dbinom(x, 1000, .6), type='h')
```

n=500



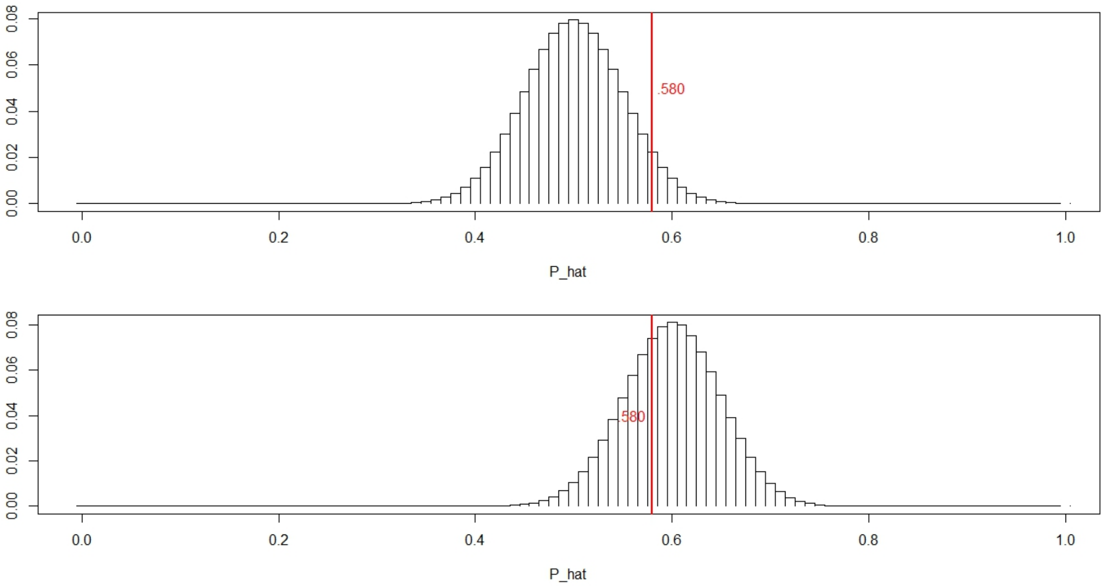
0.000
0.010
0.020
0.030
.537
0.4
0.5
0.6
0.7
0.8
x
0.00
0.01
0.02
0.03
.537
0.4
0.5
0.6
0.7
0.8
x

n=1000



.526
0.000
0.010
0.020
0.4
0.5
0.6
0.7
0.8
X
.526
0.000
0.010
0.020
0.4
0.5
0.6
0.7
0.8
X

n=100



0.00
0.02
0.04
0.06
0.08
.580
0.0
0.2
0.4
0.6
0.8
1.0
P_hat
0.00
0.02
0.04
0.06
0.08
.580
0.0
0.2
0.4
0.6
0.8
1.0
P_hat

```
n=500
 x <- 0:n
 plot( (x-.5)/n, dbinom(x, n, .5), type='s', xlab='X', xlim=c(.4, .8))
 lines((x+.5)/n, dbinom(x, n, .5), type='h')
 abline(v=.537, col="red", lwd=2); text(.55, .02, ".537", col="red")

 plot( (x-.5)/n, dbinom(x, n, .6), type='s', xlab='X', xlim=c(.4, .8))
 lines((x+.5)/n, dbinom(x, n, .6), type='h')
 abline(v=.537, col="red", lwd=2); text(.55, .02, ".537", col="red")

 n=100
 x <- 0:n
 plot( (x-.5)/n, dbinom(x, n, .5), type='s', xlab='P_hat')
 lines((x+.5)/n, dbinom(x, n, .5), type='h')
 abline(v=.580, col="red", lwd=2); text(.6, .05, ".580", col="red")

 plot( (x-.5)/n, dbinom(x, n, .6), type='s', xlab='P_hat')
 lines((x+.5)/n, dbinom(x, n, .6), type='h')
 abline(v=.580, col="red", lwd=2); text(.56, .04, ".580", col="red")


 n=1000
 x <- 0:n
 plot( (x-.5)/n, dbinom(x, n, .5), type='s', xlab='X', xlim=c(.4, .8))
 lines((x+.5)/n, dbinom(x, n, .5), type='h')
 abline(v=.526, col="red", lwd=2); text(.55, .02, ".526", col="red")

 plot( (x-.5)/n, dbinom(x, n, .6), type='s', xlab='X', xlim=c(.4, .8))
 lines((x+.5)/n, dbinom(x, n, .6), type='h')
 abline(v=.526, col="red", lwd=2); text(.55, .02, ".526", col="red")
```

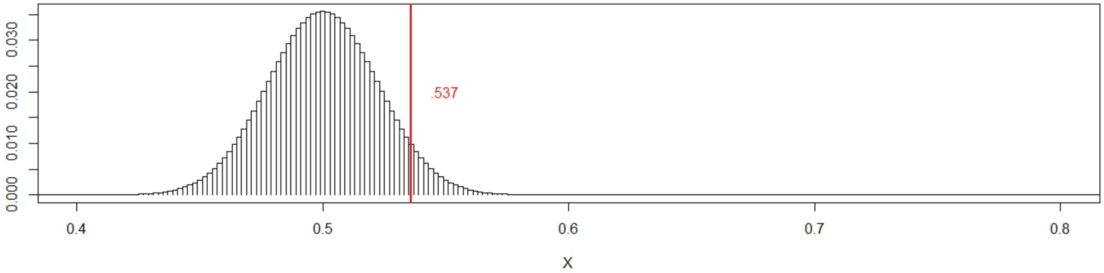
1.2 Test of Hypothesis - upper-tailed alternative

[ToC]

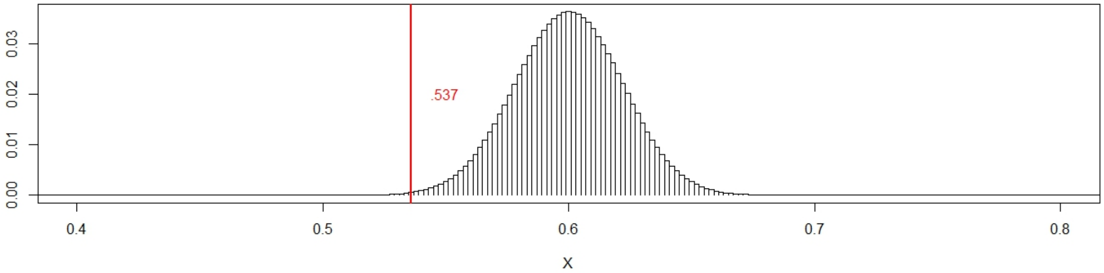
- What if we want to test

$$
\begin{aligned}
H_0 &: \quad p = .5 \\
H_A &: \quad p > .5
\end{aligned}
$$

n=500



0.000
0.010
0.020
0.030
.537
0.4
0.5
0.6
0.7
0.8
x



0.00
0.01
0.02
0.03
.537
0.4
0.5
0.6
0.7
0.8
x

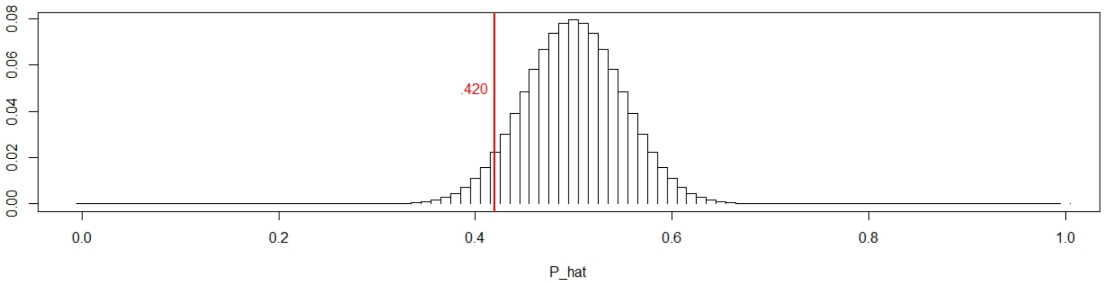
1.3 Test of Hypothesis - lower-tailed alternative

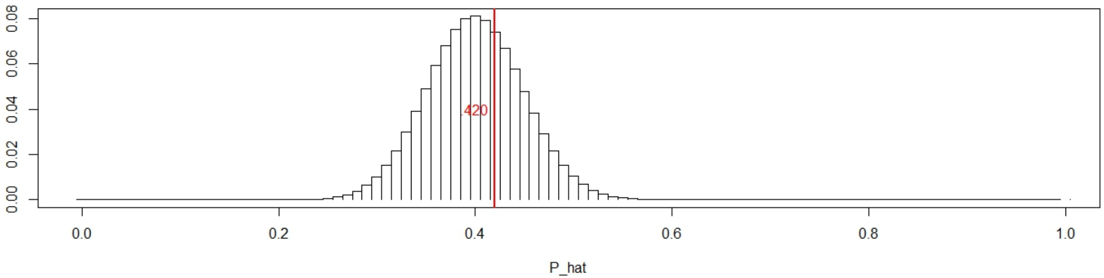
[ToC]

- What if we want to test

$$
\begin{aligned}
H_0: &\quad p = .5 \\
H_A: &\quad p = .4
\end{aligned}
$$

Lower-tailed n=100





```
n=100
x <- 0:n
plot( (x-.5)/n, dbinom(x, n, .5), type='s', xlab='P_hat')
lines((x+.5)/n, dbinom(x, n, .5), type='h')
abline(v=.420, col="red", lwd=2); text(.40, .05, ".420", col="red")

plot( (x-.5)/n, dbinom(x, n, .4), type='s', xlab='P_hat')
lines((x+.5)/n, dbinom(x, n, .4), type='h')
abline(v=.420, col="red", lwd=2); text(.40, .04, ".420", col="red")
```

1.4 Type I and Type II Errors

[ToC]

- With significance level α of our choice,

Prob of ...	don't reject H_0	reject H_0
H_0 is true	$1-\alpha$	(α)
H_A is true	(β)	Power $= 1-\beta$

- $\alpha = P($ type I error $) = P($ false positive$)$
- $\beta = P($ type II error $) = P($ false negative$)$

2

One-Sample Z-test

[ToC]

2.1 One-sample Z-test for μ

To test the null hypothesis of $H_0 : \mu = \mu_0$ against one of the alternatives from below:

$$\begin{aligned} H_A : &\quad \mu > \mu_0 \qquad \text{(Upper-tailed alternative)} \\ H_A : &\quad \mu < \mu_0 \qquad \text{(Lower-tailed alternative)} \\ H_A : &\quad \mu \neq \mu_0 \qquad \text{(Two-tailed alternative)} \end{aligned}$$

We use the test statistic of

$$z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}},$$

and with significance level α,

H_A	rejection region	p-value	β
upper-tailed	$z > z_\alpha$	$1 - \Phi(z)$	$\Phi(z_\alpha - \mu_A)$
lower-tailed	$z < -z_\alpha$	$\Phi(z)$	$1 - \Phi(-z_\alpha - \mu_A)$
Two-tailed	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$	$2\big(1 - \Phi(\lvert z\rvert)\big)$	$\Phi(z_{\frac{\alpha}{2}} - \mu_A) - \Phi(-z_{\frac{\alpha}{2}} - \mu_A)$

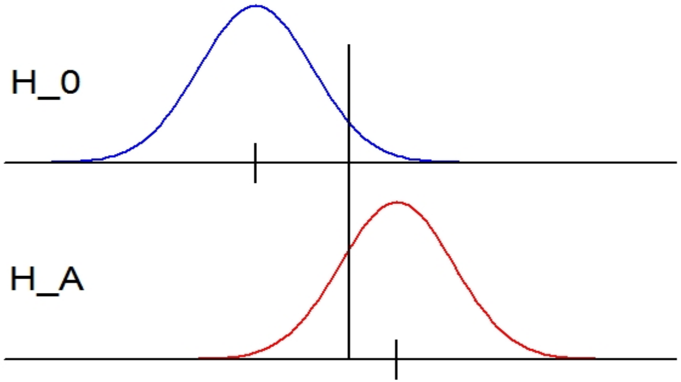
$\mu_A = \frac{\mu - \mu_0}{\sigma/\sqrt{n}}$

Test procedure:

1. Calculate test statistic $z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$.
2. Calculate p-value according to the alternative.
3. Reject H_0 if p-value is LESS than α.

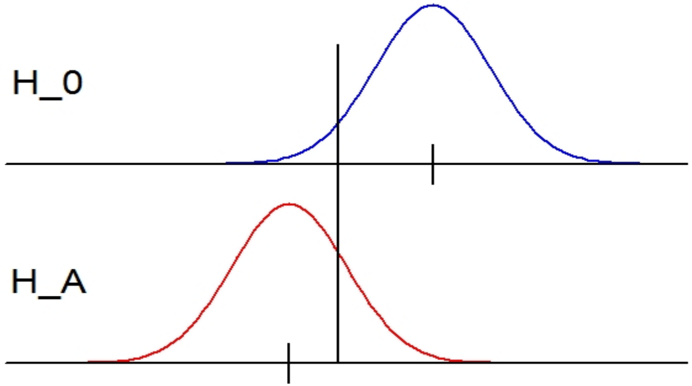
- If you can't reject H_0, then the test is inconclusive.

Behavior of test statistics z (Upper-tailed)



```
x <- seq(-7,7,.1);
plot(x, dnorm(x)+.5, type='l', col="blue", lwd=2, ylim=c(-.2,1), xlab="", ylab="", axes=F)
text(-6.7, .7, paste("H_0"), cex=2); text(-6.7, .2,  paste("H_A"), cex=2)
segments(0,.475,0,.525, lwd=2 ); abline(h=0, lwd=2); abline(h=.5, lwd=2)
text(0, .4, "0")


mu_A <- 2.5; alpha <- .05;
lines(x, dnorm(x, mu_A), col="red", lwd=2)
cv <- qnorm(1-alpha); segments(cv,0, cv, .8, lwd=2 );
segments(mu_A,-.05, mu_A, .05, lwd=2 );
%text(mu_A, -0.05, "(mu - mu_0) / (sigma/ sqrt n)")
```

Behavior of test statistics z (lower-tailed)



Meaning of p-value

- p-value is the probability of getting the observed value of z or 'worse' when H_0 is true.

2.2 Examples: Bus Routs

A metropolitan transit authority wants to determine whether there is any need for changes in the frequency of service over certain bus routes. Wants to know if average miles traveled per person by all residents in the area is 5miles or less. n=120, sample mean=4.66, assume population SD = 1.5

```
mu0=5
Xbar=4.66
Si = 1.5
n=120

testStat = (Xbar-mu0)/Si*sqrt(n); testStat

#--- Lower-tailed setup ---
Pval = pnorm(testStat); Pval

mu=4.5
mu_A = (mu-mu0)/Si*sqrt(n); mu_A
power = pnorm(-1.65, mu_A, 1); power
```

Hydro Turbines

- Hydroelectric miniturbines that generate low-cost clean electric power from small rivers and streams.
- Old model average 25.2 Kw under the lab condition.
- Recently the model design was changed, and that supposed to improve the average output.
- n=10, sample mean 27.1, sigma is known to be 3.2.

```
mu0=25.2
Xbar=27.1
Si = 3.2
n=10

testStat = (Xbar-mu0)/Si*sqrt(n); testStat

#--- Upper-tailed setup ---
Pval = 1-pnorm(testStat); Pval

mu=27.0
mu_A = (mu-mu0)/Si*sqrt(n); mu_A
power = 1-pnorm(1.65, mu_A, 1); power
```

Daily Sales

Average daily sales at small food store are known to be \$452.8. The manager recently implemented some changes in store interior, and want to know if the sales have improved. For last 12 days, the sales averaged \$501.9, and sample SD=\$65. Is the change significant?

```
mu0=452.8
Xbar=501.9
Si = 65
n=10

testStat = (Xbar-mu0)/Si*sqrt(n); testStat

#--- Upper-tailed setup ---
Pval = 1-pnorm(testStat); Pval
```

2.3 When we don't want to reject the null

- We cann't accept H_0 just because we could not reject H_0.
- We must make sure the power is high.

Hydro Turbines

- New model tested: n=10, sample mean 27.1, sigma is known to be 5.2.
- Is this an evidence that μ is higher than 25.2?

```
mu0=25.2;    Xbar=27.1;    Si = 5.2;    n=10

z = (Xbar-mu0) / (Si/sqrt(n)) ; z       # z=1.16

#--- Upper-tail Alternative ---
Pval = 1-pnorm(z); Pval                 # pval=0.1240.
```

- We failed to reject H_0, and get evidence that new model has μ higher than 25.2. Do we hae hope? Should we continule testing?
- State your worst-case acceptable, and check the power.

- Suppose μ have to be > 27.0, otherwise doesn't worth mentioning.
- Suppose $\mu = 27$, and calculate the power
- For upper-tail alt, formula is

$$\text{Power} = 1 - \Phi(z_\alpha - \mu_A) \qquad \text{where} \qquad \mu_A = \frac{\mu - \mu_0}{\sigma/\sqrt{n}}$$

```
mu=27.0
mu_A = (mu-mu0)/Si*sqrt(n); mu_A           # mu_A =1.1
power = 1-pnorm(1.65, mu_A, 1); power      # power=.2893


n=40
mu=27.0
mu_A = (mu-mu0)/Si*sqrt(n); mu_A           # mu_A =2.19
power = 1-pnorm(1.65, mu_A, 1); power      # power=.7051

n=80
mu=27.0
mu_A = (mu-mu0)/Si*sqrt(n); mu_A           # mu_A =3.09
power = 1-pnorm(1.65, mu_A, 1); power      # power=.9259
```

Prescription

- A certain prescription medicine is supposed to contain an average of 247 parts per million(ppm) of a certain chemical.
- n=20, sample mean = 250ppm, sample SD 12ppm.
- Test against two sided alternative. Power and P(type II) when true mean is 253?

```
mu0=25.2;    Xbar=27.1;    Si = 5.2;    n=10

z = (Xbar-mu0) / (Si/sqrt(n)) ; z       # z=1.16

#--- Upper-tail Alternative ---
Pval = 1-pnorm(z); Pval                 # pval=0.1240.
```

Example: Lab Scale

- To assess the accuracy of a laboratory scale, a standard weight that is known to weigh exactly 1 gram is repeatedly weighed a total of 25 times.
- $\bar{X}$ is computed to be 1.0028 grams.
- Suppose the each scale reading is independent of each other, and Normally distributed with unknown mean μ and standard deviation $\sigma = .01g$.

1. 95% CI for $\mu = \bar{X} \pm .00392 = (.99888, 1.00672)$
2. test if $\mu = 1$. report P-value, α. $z = 1.4$ $p - val = .161$
3. power of this test if true mean was 1.001. (.125)
4. get n that will give margin of error $= .001$. (over 1000)
5. What is the probability that μ is within the interval $(\bar{X} - .001, \bar{X} + .002)$?

Example: Tire Life

- The manufacturer of a new fiberglass tire claims that its average life will be at least 40,000 miles. To verify this claim a sample of 12 tires is tested, with their lifetimes (in 1,000s of miles) being as follows:

- $\bar{X} = 37.84$ and $S = 2.56$

```
D <- c(36.1, 40.2, 33.8 , 38.5, 42, 35.8, 37, 41, 36.8, 37.2)
```

Example: pH meter bias

Suppose that an engineer is interested in testing the bias in a pH meter. Data are collected on a neutral substance (pH=7.0). A sample of the measurements were taken with the data as follows:

```
x <- c(7.07, 7.00 , 7.10 , 6.97 , 7.00 , 7.03 , 7.01, 7.01, 6.98, 7.08)
(mean(x) - 7)/ (sd(x)/sqrt(10))

1-pnorm(1.96 - (7.01 - 7)/ (sd(x)/sqrt(10)) )
```

Example: Heat Transfer

- An article in the Journal of Heat Transfer (Trans. ASME, Sec. C, 96. 1974. p. 59) described a new method of measuring the thermal conductivity of Armco iron.
- Using a temperature of 100F and a power input of 550 wtts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-F) were obtained:
- $\bar{X} = 41.92$ and $S = .284$

```
D <- c( 41.60, 41.48, 42.34, 41.95, 41.86, 42.18, 41.72, 42.26, 41.81, 42.04)
```

Example: Mercury in Bass

- An arlicle in the 1993 volume of the TransactionS of the American Fisheries Sociery reports the results of a study to investigate the mercury contamination in largemouth bass. A sample of fish was selected from 53 Florida lakes and mercury concentration in the muscle tissue was measured (ppm). The mercury concentration values are

```
D=c(1.230,  0.490,  0.490,  1.080,  0.590,  0.280, 0.180,  0.100,  0.940,  1.330,  0.190,  1.160,
    0.980,  0.340,  0.340,  0.190,  0.210,  0.400, 0.040,  0.830,  0.050,  0.630,  0.340,  0.750,
    0.040,  0.860,  0.430,  0.044,  0.810,  0.150, 0.560,  0.840,  0.870,  0.490,  0.520,  0.250,
    1.200,  0.710,  0.190,  0.410,  0.500,  0.560, 1.100,  0.650,  0.270,  0.270,  0.500,  0.770,
    0.730,  0.340,  0.170,  0.160,  0.270  )
```

Is this the evidence that $\mu = .4ppm$?

Example: Sprinkler Systems

Manufacturer claims true average system temparature is 130 F. Sample of n=9 was taken with sample average of 131.08 F. Assume true standeard deviation to be 1.5F, test the hypothesis that manufacturer is right against upper-tail alternative.

Example: Pavement

Dynamic Cone Penetrometer (DCP) is used to measure material resistance to penetration (mm/blow) as a cone is driven into pavemant or subgrade.

Suppose for a particular application, it is required that the true average DCP value for pavement is less than 30. The pavement will not be used, unless the evidence is shown by the data. Sample of size 52 was taken:

```
D=c(14.1, 14.5, 15.5, 16.0, 16.0, 16.7, 16.9, 17.1, 17.5, 17.8,
    17.8, 18.1, 18.2, 18.3, 18.3, 19.0, 19.2, 19.4, 20.0, 20.0,
    20.8, 20.8, 21.0, 21.5, 23.5, 27.5, 27.5, 28.0, 28.3, 30.0,
    30.0, 31.6, 31.7, 31.7, 32.5, 33.5, 33.9, 35.0, 35.0, 35.0, 36.7,
    40.0, 40.0, 41.3, 41.7, 47.5, 50.0, 51.0, 51.8, 54.4, 55.0, 57.0)

t.test(D)
```

$\bar{X} = 28.76, S = 12.2647$

One-Sample Z-test for proportion

[ToC]

3.1 Test for a Population Proportion

If we wish to test $H_0 : p = p_0$ against alternatives

$$\begin{aligned} H_A :\quad & p > p_0 \qquad \text{(Upper-tailed alternative)} \\ H_A :\quad & p < p_0 \qquad \text{(Lower-tailed alternative)} \\ H_A :\quad & p \neq p_0 \qquad \text{(Two-tailed alternative)} \end{aligned}$$

We let $\hat{p} = X/n$ and perform one-sample z-test with significance level α of your choice. That is, the test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

H_A	rejection region
upper-tailed	$z > z_\alpha$
lower-tailed	$z < -z_\alpha$
Two-tailed	$z < -z_{\frac{\alpha}{2}}$ or $z > z_{\frac{\alpha}{2}}$

Example: Tire Share

Suppose that the Goodyear Tire company has historically held 42% of the market for automobile tires in US. Recent changes in company operation prompted the firm to test the validity of the assumption that it still controlls 42% of the market. n=550, sample showed 219/550 had Goodyear tires.

Example: Drought and Fertilizer Use

The percentage of farmers using fertilizers in an African country was known to be 35%. The drought and other events of the last few years are believed to have had a potential impact on the proportion of farmars using fertilizers. An international aid program wants to test if it changed. n=150, sample prop = 68/150.

Example 8.1-1: One sample proportion

A study was conducted on the impact characteristics of football helmets used in competitive highschool programs. In the study, a measurement called the Gadd Severity Index (GSI) was obtained on each helmet using a standardized impact test. A ehlmet was deemed to have failed if the GSI was greater than 1200. Of the 81 helmets tested, 19 failed the GSI 1200 criterion.

1. What is the point estimate of the proportion of helmets that fail, and standard error of the estimate?

2. Based on the sample, what is the 90% confidence interval for the true proportion of helmets that would fail the test?

3. Test the null hypothesis that true proportion of helmets that would fail the test is 30% against the lower-tailed alternative.

4. If the test was to be conducted again, how many suspension-type helmets should be tested so that the margin of error does not exceed 0.05 with 95% confidence?

4

One-Sample t-test

[ToC]

4.1 One-sample t-test for μ

4.2 When we have to use S instead of σ

- Let $X_1, \ldots, X_n$ be Random Sample from $\mathrm{N}(\mu, \sigma^2)$ distribution, and assume σ is **unknown**.
- We still have sample distribution,

$$\overline{X} \sim N(\mu, \sigma^2/n),$$

- But since σ is unkown, we must use the sample SD S instead.
- That gives us the test statistic

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(df = n - 1)$$

One Sample t-test

To test the null hypothesis of $H_0 : \mu = \mu_0$ against one of the alternatives from below:

$$\begin{aligned} H_A : \quad & \mu > \mu_0 && \text{(Upper-tailed alternative)} \\ H_A : \quad & \mu < \mu_0 && \text{(Lower-tailed alternative)} \\ H_A : \quad & \mu \neq \mu_0 && \text{(Two-tailed alternative)} \end{aligned}$$

We use the test statistic of $t = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$,

H_A	rejection region	p-value	power
upper-tailed	$t > t_{\alpha,n-1}$	$1 - P(T < t)$	$1 - P(T < t_\alpha - \mu_A)$
lower-tailed	$t < -t_{\alpha,n-1}$	$P(T < t)$	$P(T < -t_\alpha - \mu_A)$
Two-tailed	$t < -t_{\alpha/2,n-1}$ or $t > t_{\alpha/2,n-1}$	$2\big(1 - P(T < \lvert t\rvert)\big)$	$1 - P(T < t_{\frac{\alpha}{2},n-1} - \mu_A)$ $+P(T < -t_{\frac{\alpha}{2},n-1} - \mu_A)$

$$T \sim t(n-1), \qquad \mu_A = \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)$$

4.3 When we can't assume the normalilty

So what can we do when the poppulation distribution does not look normal?

Example: Hockey Players

An article in the Journal of Sports Science (1987, Vol.5, pp. 261-271) presents the results of an investigation of the hemoglobin level of Canadian Olympic ice hockey players. The data reported are as follows (in g/dl)

```
15.3 16.0 14.4
16.1 16.1 14.9
15.7 15.3 14.6
15.7 16.0 15.0
15.7 16.1 14.7
14.8 14.6 15.6
14.5 15.1
```

Central Limit Theorem

If $X_1, \ldots, X_n$ are R.S. from a population with **any distribution** with mean μ and standard deviation σ, then

$$\overline{X} \text{ is approximately distributed as } N(\mu, \sigma^2/n)$$

if n is large enough (> 30). Larger the value of n, better the approximation.

- This means that when n is large, we can use z-test (and z-CI) regardless of the population distribution.

5

Test for Variance

[ToC]

5.1 Test for Variance

Under Case 2,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(\text{df=n-1})$$

To test the null hypothesis of $H_0 : \sigma_1 = \sigma_2$ against alternatives

$$\begin{aligned} H_A :&\quad \sigma > \sigma_0 \qquad \text{(Upper-tailed alternative)} \\ H_A :&\quad \sigma < \sigma_0 \qquad \text{(Lower-tailed alternative)} \\ H_A :&\quad \sigma \neq \sigma_0 \qquad \text{(Two-tailed alternative)}\ , \end{aligned}$$

we use the test statistic

$$X^2 = \frac{(n-1)S^2}{\sigma_0} \sim \chi^2(\text{df=n-1}) \text{ if } H_0 \text{ is true}$$

and perform F-test with rejection regions

H_A	rejection region
upper-tailed	$X^2 > \chi^2_{\alpha,m-1,n-1}$
lower-tailed	$X^2 < \chi^2_{1-\alpha,m-1,n-1}$
Two-tailed	$X^2 < \chi^2_{1-\alpha/2,m-1,n-1}$ or $X^2 > \chi^2_{\alpha/2,m-1,n-1}$