



# Hypothesis Testing

- What is hypothesis testing
- Standard procedures
- Examples

# What the data refer to?

- If we flip a coin 100 times, and 45 come up heads this could easily occur by chance. There is not sufficient evidence to suggest that the coin is unfair.
- If we flip a coin 100 times, and 25 come up heads this would be an rare event if the coin was fair. The low probability is evidence that the coin may not be fair.

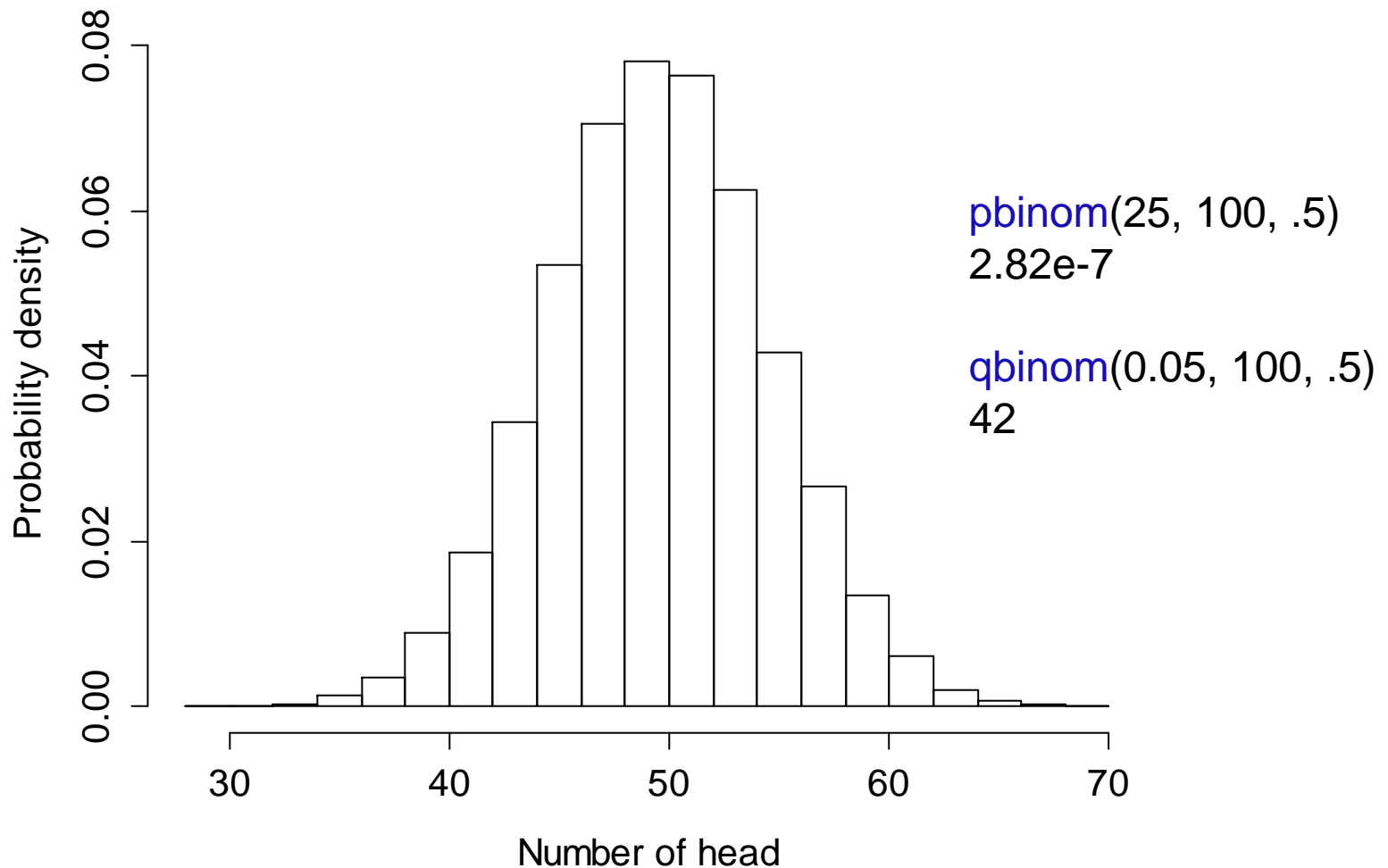
The coin is definitely unfair?

Although rare, 25 heads is still possible by chance from a fair coin.



# Use probability density

```
hist(rbinom(100000, 100, .5), freq = F, main = "",  
     xlab = 'Number of head', ylab = 'Probability density')
```





## A story: Lady tasting tea

- The lady tasting tea is a famous randomized experiment devised by Ronald A. Fisher and reported in his book *The Design of Experiments* (1935).
- The lady (Muriel Bristol) be able to tell whether the tea or the milk was added first to a cup.
- The experiment provided the Lady with 8 randomly ordered cups of tea – 4 prepared by first adding milk, 4 prepared by first adding the tea. She was to select the 4 cups prepared by one method.
- The null hypothesis was that the Lady had no such ability.



# A story: Lady tasting tea

## Tea-Tasting Distribution

Success count	Permutations of selection	Number of permutations
0	oooo	1
1	ooox, ooxo, oxoo, xooo	16
2	ooxx, oxox, oxox, xoxo, xxoo, xoox	36
3	oxxx, xoxx, xxox, xxxo	16
4	xxxx	1
Total		70

If and only if the Lady properly categorized all 8 cups was Fisher willing to reject the null hypothesis – effectively acknowledging the Lady's ability at a 1.4% significance level.

## **Hypothesis testing – confirmatory data analysis**

- A hypothesis is a claim or statement about a property of a population (e.g. the mean or a proportion of the population)
- A hypothesis test (or test of significance) is a standard procedure for testing a claim or statement about a property of a population.

It is extremely important to realize that we are not making definitive conclusions. We are giving probabilistic conclusions. We are either concluding that the results we get are likely due to chance, or unlikely.

(Zar 1999)

# Origins

- Hypothesis testing is the product of Ronald Fisher, Jerzy Neyman, Karl Pearson and Egon Pearson.
- Fisher emphasized rigorous experimental design and methods to extract a result from few samples assuming Gaussian distributions.
- Neyman and E. Pearson emphasized mathematical rigor and methods to obtain more results from many samples and a wider range of distributions (Neyman and Pearson 1933).
- Modern hypothesis testing is a hybrid of the Fisher vs. Neyman/Pearson formulation, methods and terminology developed in the early 20th century.

## Other options rather than frequentist hypothesis testing

- Confidence interval (CI) is a particular kind of interval estimate of a population parameter and is used to indicate the reliability of an estimate
- Effect size represents a standard difference, is not associated with sample size. 
$$d = \frac{\bar{x} - \mu}{s_{pooled}}$$
- The Bayesian approach to hypothesis testing is to base rejection of the hypothesis on the posterior probability
- Other approaches to reaching a decision based on data are available via decision theory and optimal decisions



# t-test using R

*# t-test daily energy intake in kJ for 11 women (Altman, 1991, p. 183)*

```
daily.intake = c(5260, 5470, 5640, 6180, 6390, 6515,
                 6805, 7515, 7515, 8230, 8770)
```

```
mean(daily.intake)
```

```
sd(daily.intake)
```

```
quantile(daily.intake)
```

```
t.test(daily.intake, mu = 7725)
```

*# Nonparametric*

```
wilcox.test(daily.intake, mu = 7725)
```

*# Two samples*

```
x1 = rnorm(300, 0, 1)
```

```
x2 = sample(0:100, 300, rep = T)
```

```
t.test(x1, x2)
```

*# Check normality*

```
plot(x1); hist(x1); qqnorm(x1)
```

```
shapiro.test(x1)
```

## One Sample t-test

data: daily.intake  
t = -2.8208, df = 10, **p-value = 0.01814**  
alternative hypothesis:  
true mean is not equal to 7725

95 percent confidence interval:  
5986.348 7520.925

sample estimates:  
mean of x  
6753.636

## Wilcoxon signed rank test

data: daily.intake  
V = 8, **p-value = 0.0293**  
alternative hypothesis: true location is  
not equal to 7725

# Example

**sample mean v.s. population mean**

Fish in polluted water are larger?



[http://texascoastgeology.com/passes/san\\_bernard\\_10\\_29\\_10%20%20%20004a2.jpg](http://texascoastgeology.com/passes/san_bernard_10_29_10%20%20%20004a2.jpg)

## Research proposition

- Water pollution usually cause eutrophication, resulting in sizes of fish being larger than normal.
- We draw a random sample of 30 fish individuals and calculate their mean length (5.3 cm)
- Based on surveys at upstream, we know the mean length of the fish is 5.1 cm
- By comparing the means, we are asking whether it is reasonable to consider the sample of the fish that is larger than the normal size.

Fish	Length
1	5.76
2	5.21
3	5.44
4	5.46
5	5.45
6	5.07
7	5.14
8	5.64
9	5.10
10	4.62
11	5.08
12	5.45
13	5.21
14	5.03
15	5.40
16	5.52
17	5.81
18	5.54
19	5.35
20	5.36
21	5.69
22	4.57
23	5.83
24	5.18
25	5.01
26	5.45
27	5.31
28	5.34
29	5.13
30	5.34

# Sample and population

Downstream

$$\bar{x} = 5.3$$

$$n = 30$$

Upstream

$$\mu = 5.1$$

$$\sigma = 0.3$$

The sample mean is 0.2 cm larger than the population mean

# What are the possibilities?

- The average fish lengths at downstream area are about the same as the normal size, and this sample happens to show a really large mean.  
**difference = sampling variability**
- The average fish lengths at downstream area are indeed larger than the normal size.  
**difference = real**

# How do we decide which explanation makes more sense?

- The traditional way: hypothesis tests
  - Five steps

# One sample hypothesis test

- **A random sample:** 30 fish with an average of length of 5.3 cm.
- **Population:** The average fish length at upstream is 5.1 cm with a standard deviation of 0.3 cm.

Do fish in polluted water were significantly larger than the population?

# Step one: checking assumptions

Hypothesis testing involves several assumptions that must be met for the results of the test to be valid

For the one sample hypothesis test, we assume:

- random sampling
- the level of measurement is interval-ratio
- the sampling distribution is normal



## Step two: stating hypotheses

- Null hypothesis ( $H_0$ ): a statement of 'no difference'
- Alternative hypothesis ( $H_a$ ): a statement that reflects the research question
- Both are expressed in terms of population parameters

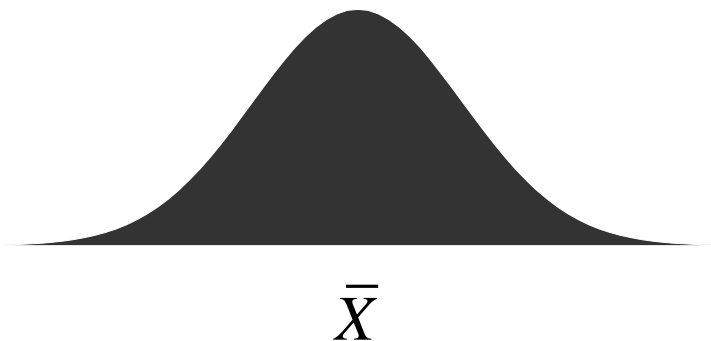
## Step two: stating hypotheses

$$H_0: \mu_{polluted} = 5.1$$

$$H_a: \mu_{polluted} > 5.1$$

## Step three: select the sampling distribution and establish the critical region (1/3)

- Select sampling distribution:



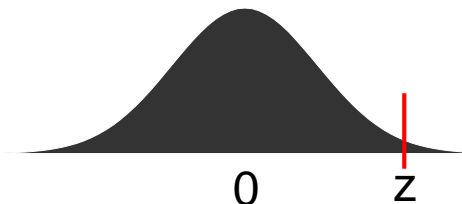
$$\mu = 5.1$$
$$\sigma = 0.3$$

## Step three: selecting sampling distribution and establish the critical region (2/3)

- Sampling distribution  $\bar{X}$   $\rightarrow$  statistic distribution  $Z$ . using  $Z$  statistic and normal distribution when the sample is large
- $Z_{\text{critical}}$ : the score associated with a particular  $\alpha$  level and marking the beginning of the critical region
- Critical region: area under the sampling distribution that includes all unlikely sample results
- P-value: the “chance” of getting the observed sample mean **NOT** further away from the hypothesized population mean?

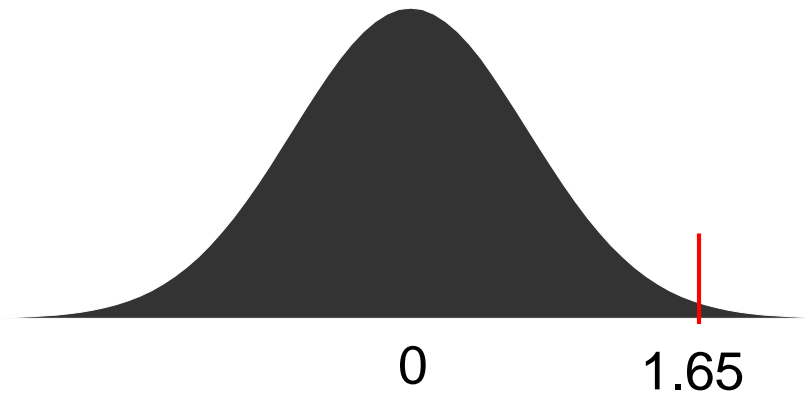
P-value small: reject  $H_0$

P-value large: fail to reject  $H_0$



## Step three: selecting sampling distribution and establish the critical region (3/3)

- Confidence level:  $1-\alpha = 0.95$
- $Z_{\text{critical}} = 1.65$



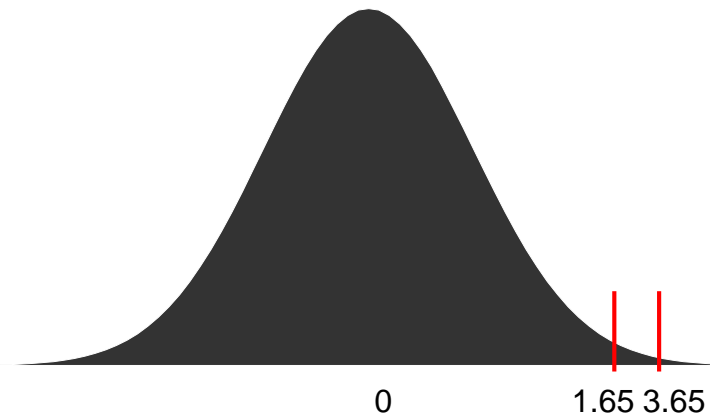
## Step four: compute the test statistic

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}}$$

$$Z = \frac{5.3 - 5.1}{0.3 / \sqrt{30}} = 3.65$$

## Step five: make a decision

- Plot the test statistic on the sampling distribution
- If the test statistic is in the critical region, our decision is: reject the null (“statistically significant”)
- If the test statistic is *not* in the critical region, our decision is: fail to reject the null



## **Step five: make a decision**

- We reject the null hypothesis.
- We are 95% confident that fish in polluted water have significantly larger sizes than normal fish.



# Formal hypothesis testing

1. Making/checking assumptions
2. Convert your claim into a symbolic null and alternative hypothesis
3. Select the sampling distribution and establish the critical region
4. Calculate a test statistic
5. Compare the test statistic to critical values OR a probability, write a conclusion

# Assumptions

# **For the one sample hypothesis test, we assume**

- Random sampling
- The level of measurement is interval-ratio
- The sampling distribution is normal

# When can you assume the shape of the sampling distribution is normal?

- When the population distribution is normal
- When you check sample mean (Based on the Central Limit Theorem)
  - `shapiro.test(x)`  $p > 0.05$
  - `qqnorm(x)` straight line

# Test of normality **shapiro.test**

- Shapiro-Wilk test of normality  
published in 1965 by Samuel Sanford Shapiro and Martin Wilk
- The test may be statistically significant from a normal distribution in any large samples. Thus a [Q–Q plot](#) is required for verification in addition to the test.

```
shapiro.test(rnorm(5000, mean = 5, sd = 100))
```

```
shapiro.test(runif(30, min = 2, max = 4))
```

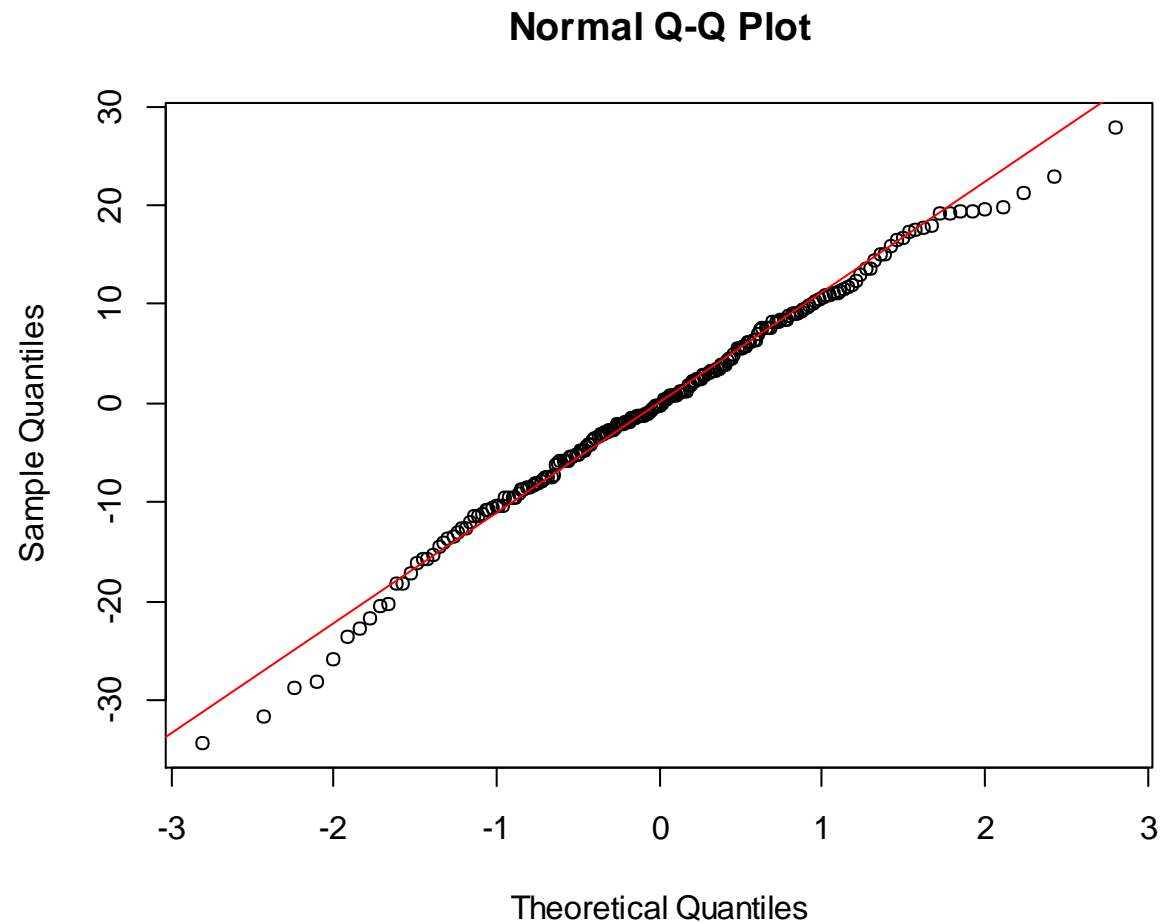
```
Shapiro-wilk normality test
data: rnorm(5000, mean = 5, sd = 100)
W = 0.99934, p-value = 0.06477
```

# Quantile-Quantile Plots

```
y <- rnorm(200, sd = 10)
```

```
qqnorm(y)
```

```
qqline(y, col = 2)
```

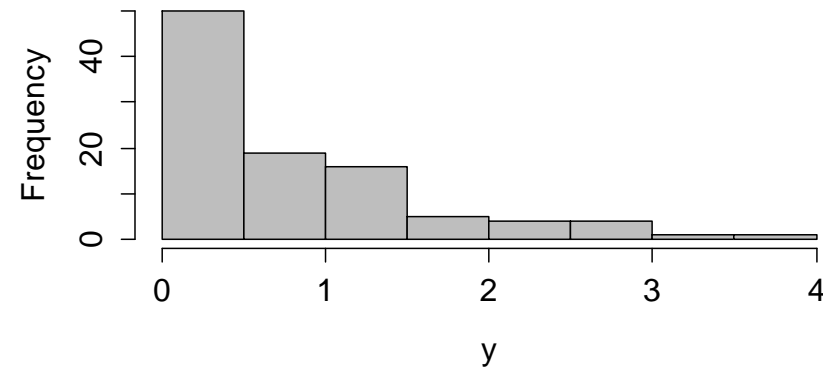


# Central limit theorem

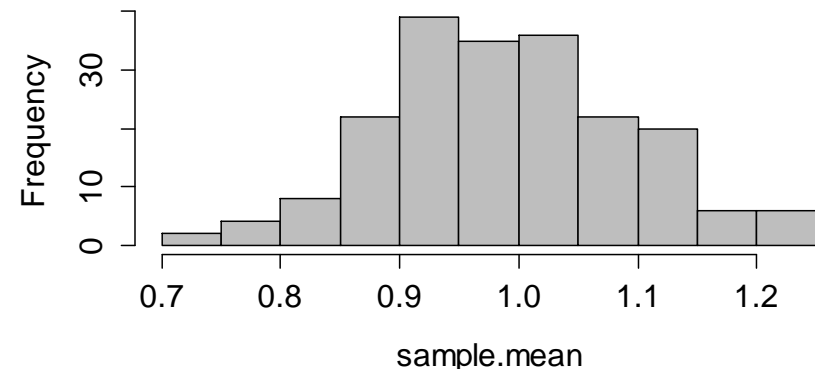
- Sampling distribution of means becomes normal as  $N$  increases, regardless of shape of original distribution.

```
# The central limit theorem  
# Exponential distribution is a skewed distribution  
y <- rexp(100); hist(y, col = 'grey')  
  
# Define a vector with length 200  
sample.mean <- numeric(200)  
  
# The mean of 200 samples  
for(i in 1:200) sample.mean[i] <- mean(rexp(100))  
  
# A normal distribution appears  
hist(sample.mean, col = 'grey')  
shapiro.test(sample.mean)
```

Histogram of y



Histogram of sample.mean



# Hypothesis



# Null hypothesis (denoted $H_0$ )

- A statement that the value of a population parameter (such as proportion or mean) is equal to some claimed value, or has no change.
- The original claim includes equality ( $\leq$ ,  $=$ , or  $\geq$ )

# Null hypotheses

- Some null hypotheses may be:
  - there is no difference between the height of the male and female students
  - there is no difference in the location (distance to downtown) of superstores and small grocers shops

## **Alternative hypothesis (denoted $H_1$ or $H_a$ )**

- A statement that the value of a population parameter differs from the null hypothesis.
- The symbolic form must be a  $>$ ,  $<$  or  $\neq$  statement.

# Understanding hypothesis

- Null hypothesis
  - Results are due to “chance” ( $H_0$ )
- Alternative hypothesis
  - Results are due to a true “effect” ( $H_1$ )
- Assess
  - Assuming  $H_0$  is true, what is the probability or “chance” of obtaining the data we did?
- Decide
  - If the chance is small enough, reject  $H_0$  and infer the “effect” is real.

# Test **statistics**

A value computed from the sample data, used in making the decision whether or not to reject the null hypothesis.

# Test **statistics**

Z value for proportion

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Z value for mean (sigma known)

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

T value for mean (sigma unknown)

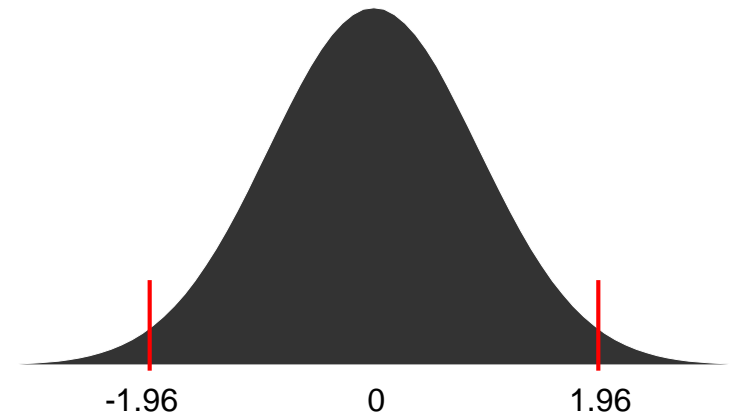
$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

The test statistic indicates how far our sample deviates from the assumed population parameter.

# Z test statistic

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

- Test continuous outcome
- Known variance
- Under  $H_0$   $\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$
- Therefore,



Reject  $H_0$  if  $\left| \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \right| > 1.96$  (gives a 2-sided  $\alpha=0.05$  test)

Reject  $H_0$  if  $\bar{X} > \mu_0 + 1.96 \frac{\sigma}{\sqrt{n}}$  or  $\bar{X} < \mu_0 - 1.96 \frac{\sigma}{\sqrt{n}}$

`pnorm(z)`

`t.test(x, mu=0, var.equal = T)`

# T test statistic

- Test continuous outcome
- Unknown variance
- Under  $H_0$

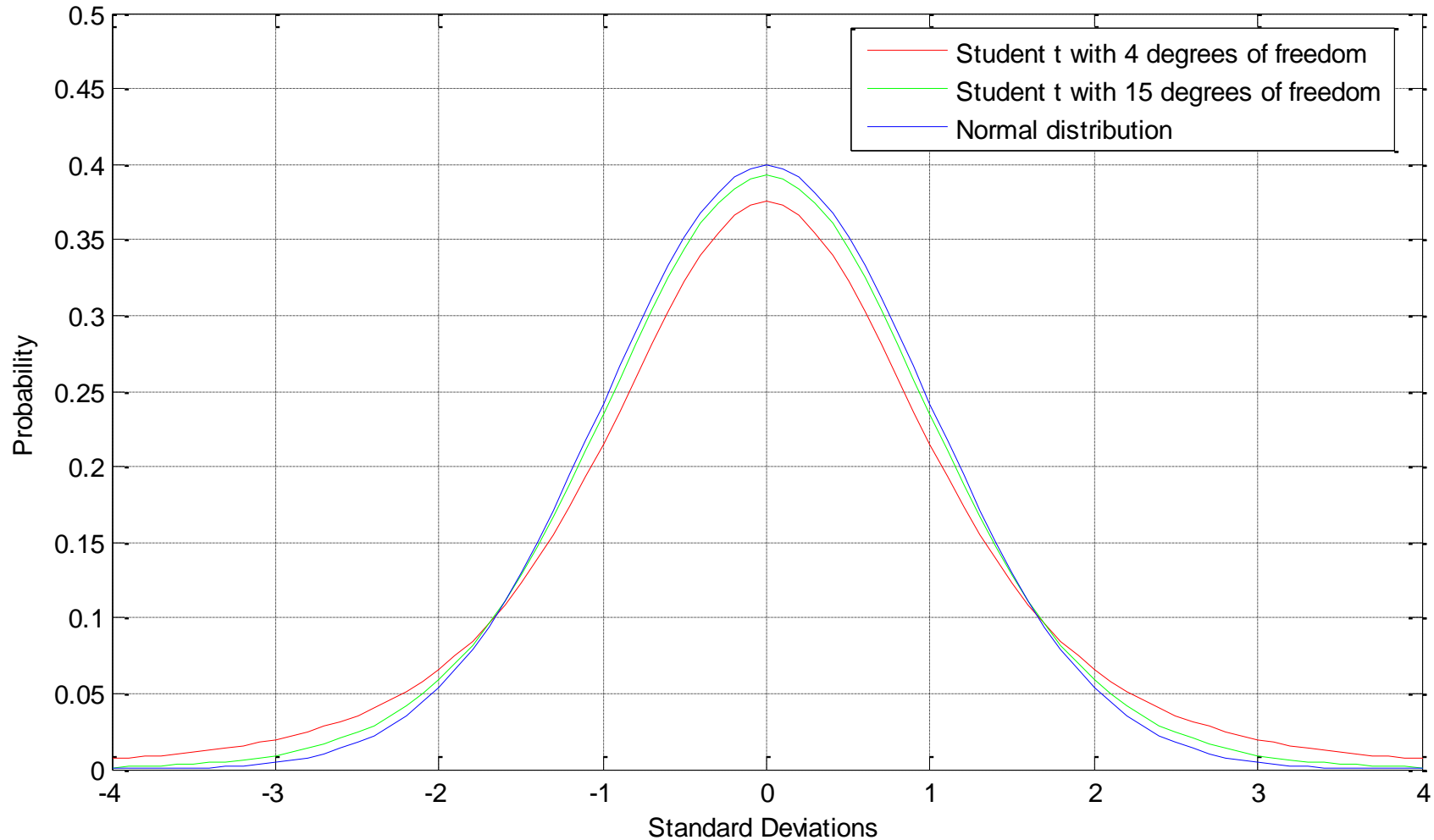
$$\frac{\bar{X} - \mu_0}{s / \sqrt{n}} \sim t_{(n-1)}$$

- Critical values: depend on degree of freedom, from computer or tables in statistics books
- t-distribution approximately normal for degrees of freedom (df) >30

`t.test(x, mu=0, var.equal = F)`



# Comparing the student-t distribution to the normal distribution



# Plot

```
# Plot a normal distribution and t distributions
```

```
x <- seq(-4, 4, len = 1000)
```

```
z <- dnorm(x, 0, 1)
```

```
plot(x, z, type = 'l', col = 'blue') # the normal curve
```

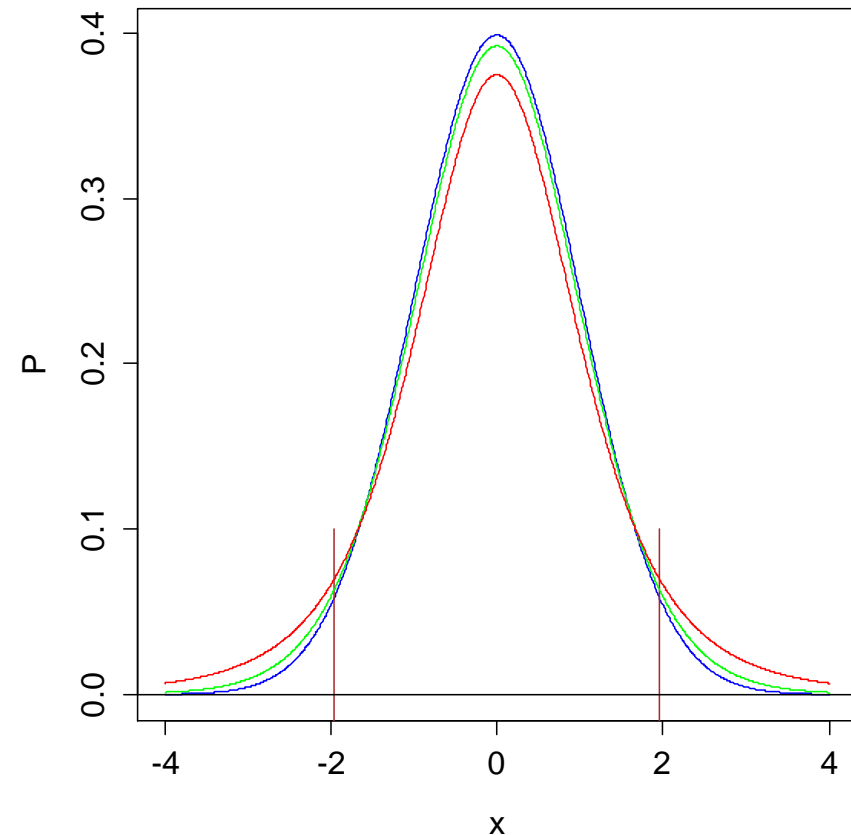
```
segments( 1.96, -0.05, 1.96, 0.1, col = 'brown')
```

```
segments(-1.96, -0.05, -1.96, 0.1, col = 'brown')
```

```
abline(0,0)
```

```
lines(x, dt(x, 15), col = 'green')
```

```
lines(x, dt(x, 4), col = 'red')
```



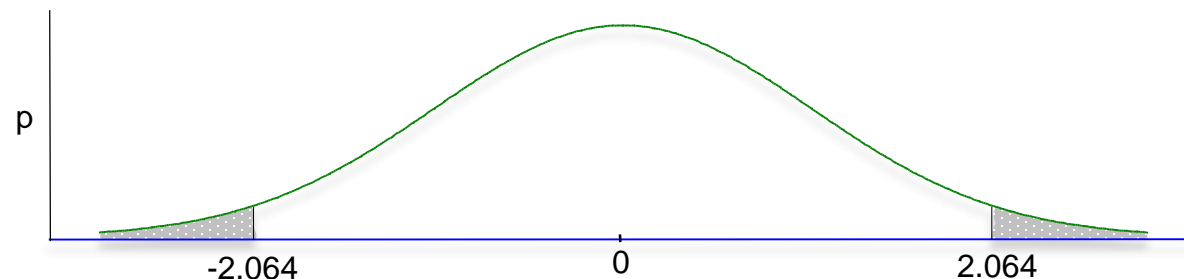
# Critical values (quartiles) of the $t$ distribution

The  $t$  table

DF $\nu$	$P$										
	One tail	0.25	0.20	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
	Two tails	0.50	0.40	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
1		1.000	1.376	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2		0.816	1.061	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3		0.765	0.978	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4		0.741	0.941	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5		0.727	0.920	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6		0.718	0.906	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7		0.711	0.896	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8		0.706	0.889	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9		0.703	0.883	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10		0.700	0.879	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
21		0.686	0.859	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22		0.686	0.858	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23		0.685	0.858	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24		0.685	0.857	1.318	1.711	<b>2.064</b>	2.492	2.797	3.091	3.467	3.745
25		0.684	0.856	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725

$$t_{0.05,24} = 2.064$$

$$P(|t| \geq 2.064) = 0.05$$



# **P value and decision**

# P value

- The probability of getting a value of the test statistic that is at least as extreme as the one obtained for the sample data.
- If the P value is very small (such as less than 0.05), we will reject the null hypothesis.
- Measure of the strength of evidence in the data that the null is not true
- A variable whose value lies between 0 and 1

# Significance level

- Denoted by  $\alpha$
- The probability that the test statistic will fall in the critical region when the null hypothesis is actually true.
- Common choices are 0.05, 0.01, and 0.10

# How to state

## If the original claim contains equality (became $H_0$ )

- Reject  $H_0$ : “There is sufficient evidence to reject the claim that...”
- Fail to Reject  $H_0$ : “There is not sufficient evidence to reject the claim that...”

## If the original claim does not contain equality (was $H_1$ )

- Reject  $H_0$ : “The sample data support the claim that...”
- Fail to Reject  $H_0$ : “There is not sufficient sample evidence to support the claim that...”

# Don't use “Accept”

- Some texts use “accept the null hypothesis”, which is misleading, implying incorrectly that the null has been proven
- We are not proving the null hypothesis (can't **PROVE** equality)
- The phrase ‘fail to reject’ represents the result more correctly.

If the sample evidence is not strong enough to warrant rejection, then the null hypothesis may or may not be true (just as a defendant found NOT GUILTY may or may not be innocent)



# Decisions and conclusions

## P-value method

Reject  $H_0$  if P-value  $\leq \alpha$

Fail to reject  $H_0$  if P-value  $> \alpha$

## Other methods

Give P-value, and leave conclusion to the reader

Look at whether population parameter falls in confidence interval estimate

# **Two-tailed, right-tailed, left-tailed tests**

The tails in a distribution are the extreme regions bounded by critical values.

# Two-tailed Test

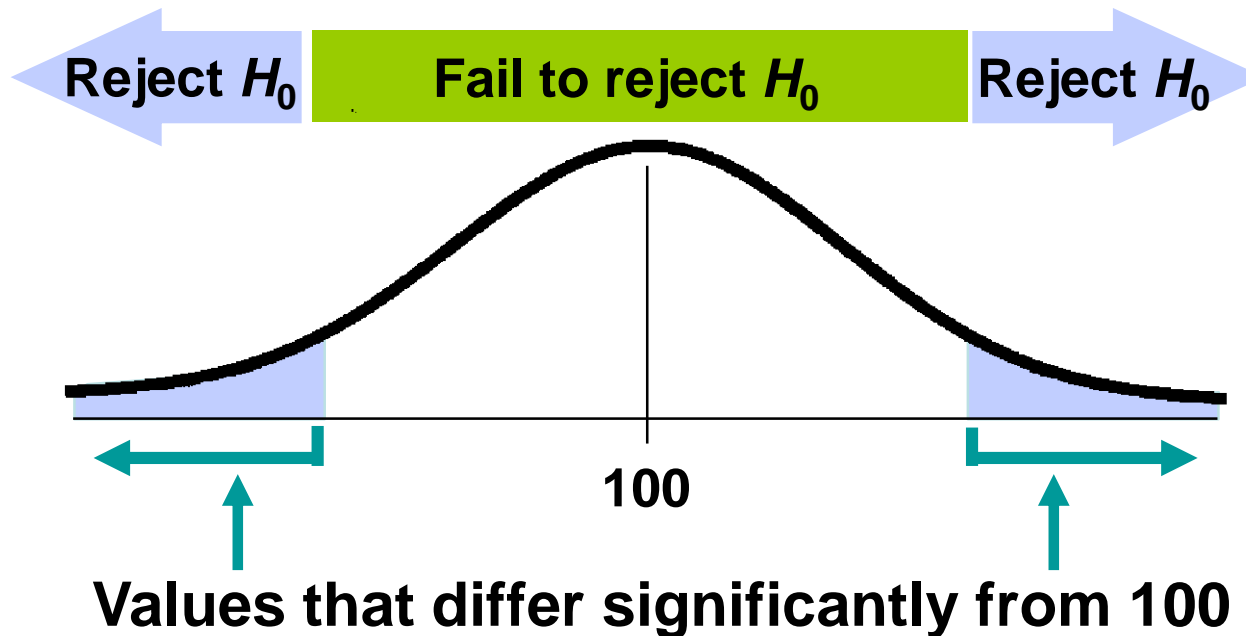
```
t.test(x, mu = 100, alt = "two.sided")
```

$$H_0: \mu = 100$$

$$H_1: \mu \neq 100$$

$\alpha$  is divided equally  
between the two tails of  
the critical region

UNEQUAL means less than or greater than

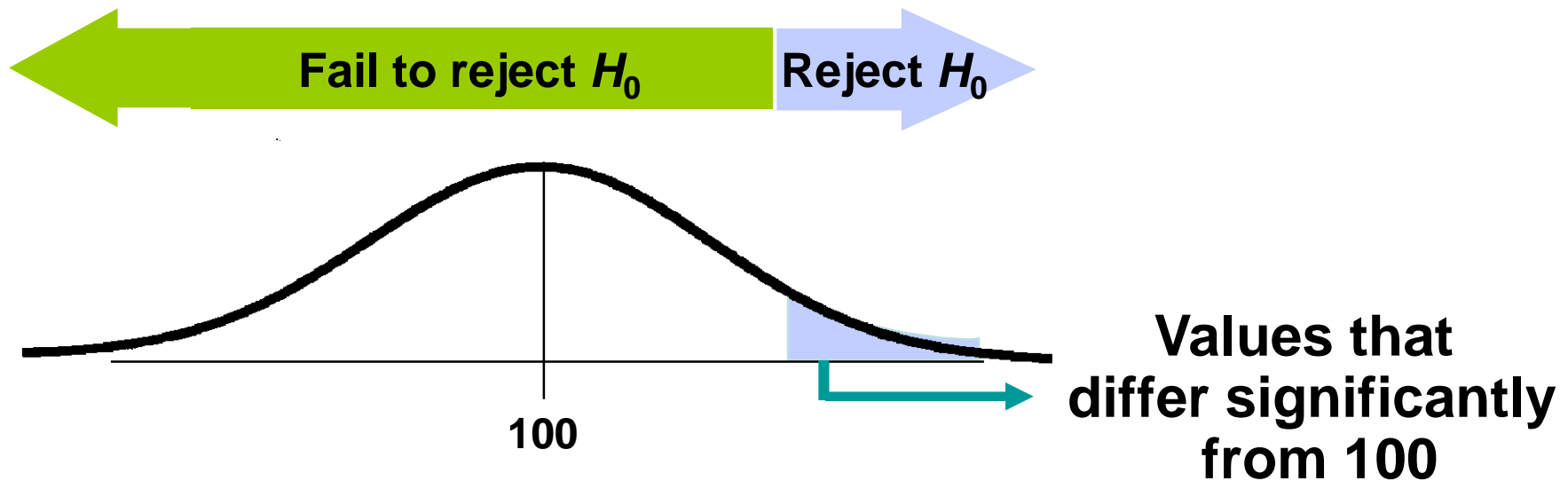


# Right-tailed Test

`t.test(x, mu=100, alt = "greater")`

$$H_0: \mu \leq 100$$

$$H_1: \mu > 100$$



# Left-tailed Test

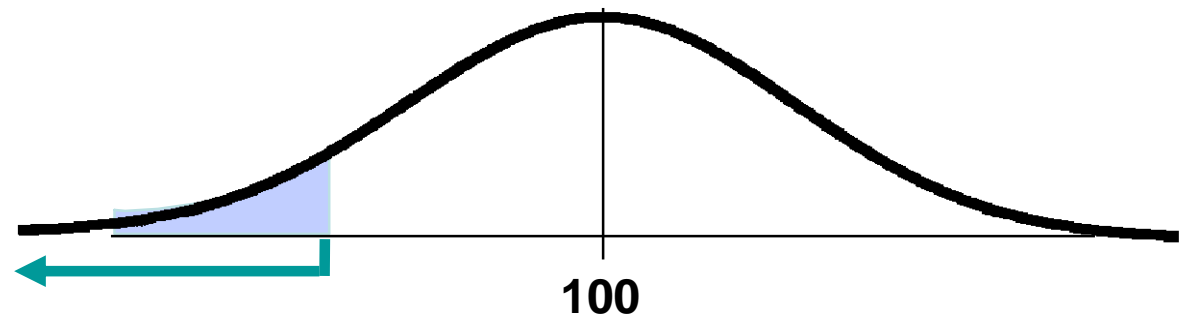
```
t.test(x, mu=0, alt = "less")
```

$$H_0: \mu \geq 100$$

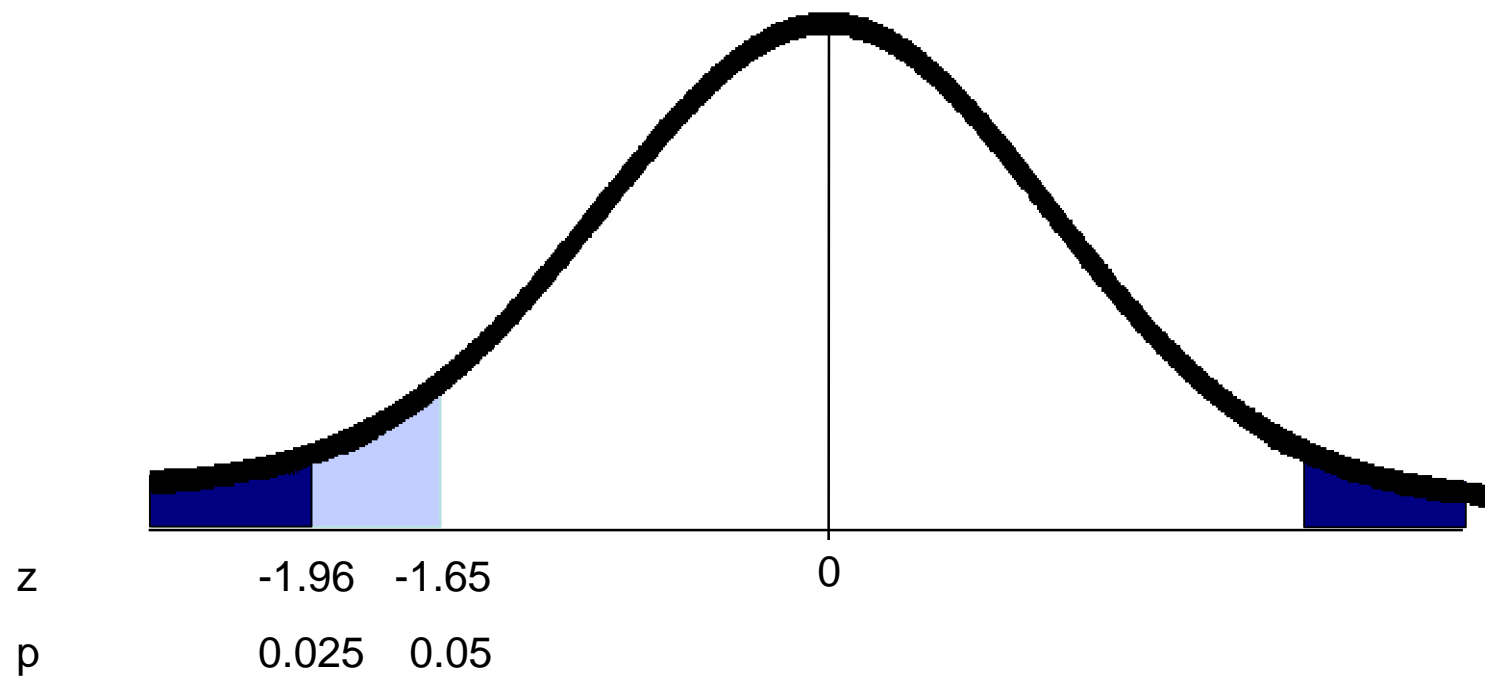
$$H_1: \mu < 100$$



Values that  
differ significantly  
from 100



# Advantage of one-tailed test



# One sample hypothesis test

- Faculty salary equity is a hot political issue: some argue that **earnings of unionized faculty** are considerably **higher** than the **earnings of faculty nationwide**.
- Assume that we know the national mean salary of a full-time faculty member is \$45,000 and that salaries are normally distributed.
- We take a random sample of 23 unionized faculty members and find that they make on average \$47,000 with a standard deviation of \$13,200.
- Do unionized faculty members make significantly more a year than all faculty members in the population? Set  $\alpha = .01$

# Parameters

$$\mu = \$45,000$$

$$\bar{x} = \$47,000$$

$$s = \$13,200$$

$$N = 23$$



# Step one

We assume

- random sampling,
- level of measurement is interval-ratio,
- the sampling distribution is normal

## Step two

$$H_0 : \mu_{unionized} = \$45,000$$

$$H_1 : \mu_{unionized} > \$45,000$$

## Step three

- Sampling distribution = t distribution
- $\alpha = .01$
- $df = N - 1 = 22$
- $t_{\text{critical}} = 2.508$

`qt(0.01, 22) # -2.508`

## Step four

$$t = \frac{\bar{x} - \mu}{s / \sqrt{N}}$$

$$t = \frac{47,000 - 45,000}{13,200 / \sqrt{23}}$$

$$t = .73$$

## Step five

- We fail to reject the null hypothesis.
- We are 99% confident that there is no difference between the income of unionized faculty members and all faculty members in the population.

# One sample hypothesis tests: proportions

- We know that 80% of the population has at least a high school diploma (U.S. Census Bureau 1990).
- Previous studies have shown that the South is economically disadvantaged and that Southerners have less overall education than the rest of the population.
- We randomly sampled 300 Southerners and found that 77% had a high school diploma at least .
- Are Southerners less educated than the population at large? Set  $\alpha = .01$

# Parameters

$$P_u = .80$$

$$P_s = .77$$

$$N = 300$$

# Step one

We assume

- random sampling
- the proportion is interval-ratio
- the sampling distribution is normal



## Step two

$$H_0 : P_s = .80$$

$$H_1 : P_s < .80$$

## Step three

- Sampling distribution = Z distribution
- $\alpha = .01$
- $Z_{\text{critical}} = -2.33$

`qnorm(0.01)`

## Step four

$$Z = \frac{P_s - P_u}{\sqrt{P_u(1 - P_u)/N}}$$

$$Z = \frac{.77 - .80}{\sqrt{.80(1 - .80)/300}}$$

$$Z = -1.30$$

# R script - one sample hypothesis tests: proportions

```
prop.test(x, n, p=null,  
          alternative = c("two.sided", "less", "greater"),  
          conf.level = 0.95, correct = TRUE)
```

```
prop.test(0.77*300, 300, p=0.8)
```

1-sample proportions test with continuity correction

data: 0.77 \* 300 out of 300, null probability 0.8

X-squared = 1.5052, df = 1, p-value = 0.2199

alternative hypothesis: true p is not equal to 0.8

95 percent confidence interval:

0.7173811 0.8155549

sample estimates:

p  
0.77

## Step five

- We fail to reject the null hypothesis.
- We are 99% confident that there is no difference between Southerners and the rest of the U.S. population with respect to the proportion owning a high school diploma.

# Different types of t test

# Paired test: difference between two continuous outcomes

`t.test(x, y, paired = T)`

- Known variance: Z test statistic
- Unknown variance: t test statistic
- $H_0: \mu_d = 0$  vs.  $H_A: \mu_d \neq 0$
- Paired Z-test or paired t-test

$$Z = \frac{\bar{d}}{\sigma / \sqrt{n}} \quad or \quad T = \frac{\bar{d}}{s / \sqrt{n}}$$

$\bar{d}$  is the differences between two observations

# Paired tests

*# input data*

lines <-

"ID x y

1 208 197

2 202 150

3 203 255

4 200 134

5 205 266

6 206 200

7 207 189

8 208 186

9 203 215

10 210 199"

score <- read.table(con <- textConnection(lines), header=TRUE)

close(con)

ID: Individual ID number

x: score after the training course

y: score before the training course

t.test(score\$x, score\$y, paired = T)



# Two samples: common variance, one population

- Known variance: Z test statistic
- Unknown variance: t test statistic
- $H_0: \mu_1 = \mu_2$  vs.  $H_A: \mu_1 \neq \mu_2$
- Assume common variance

$$Z = \frac{\bar{x} - \bar{y}}{\sigma \sqrt{1/n + 1/m}} \quad \text{or} \quad T = \frac{\bar{x} - \bar{y}}{s \sqrt{1/n + 1/m}}$$

`t.test(x, y, var.equal = TRUE)`

# Two samples: not common variance, one/two populations

- Known variance: Z test statistic
- Unknown variance: t test statistic
- $H_0: \mu_1 = \mu_2$  vs.  $H_A: \mu_1 \neq \mu_2$

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_1^2 / n + \sigma_2^2 / m}} \quad \text{or} \quad T = \frac{\bar{x} - \bar{y}}{\sqrt{s_1^2 / n + s_2^2 / m}}$$

`t.test(x, y, var.equal = FALSE)` # Welch's t-test

# A test for equal proportion of one population

- Exact same idea as for mean
- For large samples
  - Use Z test statistic
  - Now set up in terms of proportions, not means

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

`prop.test(12, 20, p=0.5)` # it is a chi square test

# Two population proportions

- For large samples use Z test statistic (sex ratio for two species)

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1(1-P_1)}{n} + \frac{P_2(1-P_2)}{m}}}$$

# Two population proportions

## (chi square test)

```
x = sample(c('A','B'), 50, rep = T)
y = sample(c('F','M'), 50, rep = T)
table(x,y)
```

	F	M
A	13	8
B	15	14

```
prop.test(table(x,y), correct = TRUE)
```

2-sample test for equality of proportions with continuity correction

data: table(x, y)

X-squared = 0.1825, df = 1, p-value = 0.6693

alternative hypothesis: two.sided

95 percent confidence interval:

-0.2153168 0.4189292

sample estimates:

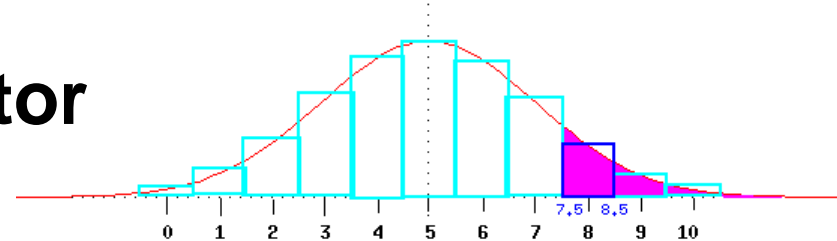
prop 1          prop 2

0.6190476    0.5172414

<http://www.math.uni.edu/~campbell/stat/prob9.html>

# Continuity correction factor

`prop.test(table(x,y), correct = TRUE)`



- The normal distribution can take all real numbers (is continuous) but the binomial distribution can only take integer values (is discrete), a normal approximation to the binomial should identify the binomial event “6” with the normal interval “(5.5, 6.5)” (and similarly for other integer values).
- For example: If  $n=20$  and  $p=.25$ , what is the probability that  $X$  is greater than or equal to 8? `pbinom(8-1, 20, .25, lower.tail = F) = 0.1018`
- The normal approximation without the continuity correction factor yields  $z=(8 - 20 \times .25)/(20 \times .25 \times .75)^{.5} = 1.55$ , hence  $P(X \text{ *greater than or equal to } 8)$  is approximately .0606.
- The continuity correction factor requires us to use 7.5 in order to include 8 since the inequality is weak and we want the region to the right.  $z = (7.5 - 5)/(20 \times .25 \times .75)^{.5} = 1.29$ , hence the area under the normal curve is .0985.
- The exact solution is about 0.1019.
- Hence for small  $n$ , the continuity correction factor gives a much better answer.

# Examples

# Formal hypothesis testing

1. Making/checking assumptions
2. Convert your claim into a symbolic null and alternative hypothesis
3. Select the sampling distribution and establish the critical region
4. Calculate a test statistic
5. Compare the test statistic to critical values OR a probability, write a conclusion



ZAR p123

**A two-sample  $t$  test (two tails)****EXAMPLE 8.1** A two-sample  $t$  test for the two-tailed hypotheses

The data are human blood-clotting times (in minutes) of individuals given one of two drugs.

$H_0: \mu_1 = \mu_2$  and  $H_A: \mu_1 \neq \mu_2$  (which could also be stated as  $H_0: \mu_1 - \mu_2 = 0$  and  $H_A: \mu_1 - \mu_2 \neq 0$ ).

*Given drug B*      *Given drug C*

8.8                      9.9

8.4                      9.0

7.9                      11.1

8.7                      9.6

9.1                      8.7

9.6                      10.4

9.5

$n_1 = 6$                        $n_2 = 7$

$v_1 = 5$                        $v_2 = 6$

$\bar{X}_1 = 8.75 \text{ min}$        $\bar{X}_2 = 9.74 \text{ min}$

$SS_1 = 1.6950 \text{ min}^2$        $SS_2 = 4.0171 \text{ min}^2$

$$s_p^2 = \frac{SS_1 + SS_2}{v_1 + v_2} = \frac{1.6950 + 4.0171}{5 + 6} = 0.5193 \text{ min}^2$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{0.5193}{6} + \frac{0.5193}{7}} = \sqrt{0.0866 + 0.0742} = 0.40 \text{ min}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{8.75 - 9.74}{0.40} = \frac{-0.99}{0.40} = -2.475$$

$$t_{0.05(2), v} = t_{0.05(2), 11} = 2.201 \quad P = 0.03$$

Therefore, reject  $H_0$

**t.test(B, C, var.equal = TRUE)**

ZAR p126

## A two-sample $t$ test (one tail)

### EXAMPLE 8.2 A two-sample $t$ test for the one-tailed hypotheses

The data are heights of plants, each grown with one of two different fertilizers.

$$H_o: \mu_1 \geq \mu_2 \text{ and } H_A: \mu_1 < \mu_2$$

**Present  
fertilizer**

**Newer  
fertilizer**

48.2cm

52.3cm

54.6

57.4

58.3

55.6

47.8

53.2

51.4

61.3

52.0

58.0

55.2

59.8

49.1

54.8

49.9

52.6

 $n_1 = 10$  $n_2 = 8$  $v_1 = 9$  $v_2 = 7$  $\bar{X}_1 = 51.91\text{cm}$  $\bar{X}_2 = 56.55\text{cm}$ 

$$s_p^2 = \frac{102.23 + 69.20}{9 + 7} = \frac{171.43}{16} = 10.71\text{cm}^2$$

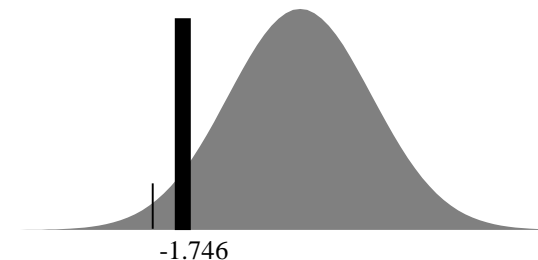
$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{10.71}{10} + \frac{10.71}{8}} = \sqrt{2.41} = 1.55\text{cm}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{51.91 - 56.55}{1.55} = \frac{-4.64}{1.55} = -2.99$$

$$t_{0.05(1),16} = -1.746$$

$$P = 0.0043$$

$H_o$ : is rejected.



$$SS_1 = 102.23 \text{ cm}^2 \quad SS_2 = 69.20 \text{ cm}^2$$

`t.test(present, newer, alt="less", var.equal = T)`



(ZAR 1999)

**EXAMPLE 8.19****Compare two indices of diversity**

$H_0$ : The diversity of plant food items in the diet of Michigan blue jays is the same as the diversity food items in the diet of Louisiana blue jays.

$H_A$ : The diversity of plant food items in the diet of Michigan blue jays is not the same as in the diet of Louisiana blue jays.  $\alpha = 0.05$

**Michigan Blue Jays**

Diet item	$f_i$	$f_i \log f_i$	$f_i \log^2 f_i$
Oak	47	78.5886	131.4078
Corn	35	54.0424	83.4452
Blackberry	7	5.9157	4.9994
Beech	5	3.4949	2.4429
Cherry	3	1.4341	0.6830
Other	2	0.6021	0.1812

$$s_1 = 6 \quad n_1 = \sum f_i \quad \sum f_i \log f_i \quad \sum f_i \log^2 f_i$$

$$= 99 \quad = 144.0751 \quad = 223.1595$$

$$H'_1 = \frac{n_1 \log n_1 - \sum f_i \log f_i}{n_1} = \frac{197.5679 - 144.0751}{99} = 0.5403$$

$$s_{H'_1}^2 = \frac{\sum f_i \log^2 f_i - (\sum f_i \log f_i)^2 / n_1}{n_1^2} = 0.00137602$$

**Louisiana Blue Jays**

Diet item	$f_i$	$f_i \log f_i$	$f_i \log^2 f_i$
Oak	48	80.6996	135.6755
Pine	23	31.3197	42.6489
Grape	11	11.4553	11.9294
Corn	13	14.4813	16.1313
Blueberry	8	7.2247	6.5246
Other	2	0.6021	0.1812

$$s_2 = 6 \quad n_2 = \sum f_i \quad \sum f_i \log f_i \quad \sum f_i \log^2 f_i$$

$$= 105 \quad = 145.7827 \quad = 213.0909$$

$$H'_2 = \frac{n_2 \log n_2 - \sum f_i \log f_i}{n_2} = \frac{212.2249 - 145.7827}{105} = 0.6328$$

$$s_{H'_2}^2 = \frac{\sum f_i \log^2 f_i - (\sum f_i \log f_i)^2 / n_2}{n_2^2} = 0.00096918$$



## Compare two indices of diversity

$$s_{H'_1-H'_2} = \sqrt{s_{H'_1}^2 + s_{H'_2}^2} = \sqrt{0.00137602 + 0.00096918} = 0.0484$$

$$t = \frac{H'_1 - H'_2}{s_{H'_1-H'_2}} = \frac{-0.0925}{0.0484} = -1.911$$

$$v = \frac{\left(s_{H'_1}^2 + s_{H'_2}^2\right)^2}{\frac{\left(s_{H'_1}^2\right)^2}{n_1} + \frac{\left(s_{H'_2}^2\right)^2}{n_2}} = \frac{(0.00137602 + 0.00096918)^2}{\frac{(0.00137602)^2}{99} + \frac{(0.00096918)^2}{105}}$$

$$= \frac{0.000005499963}{0.000000028071} = 196$$

$$t_{0.05(2),196} = 1.972$$

$$P = 0.057$$

Therefore, do not reject  $H_0$ .

## Normality assumption

- We assume normal distributions to figure sampling distributions and thus  $p$  levels.
- Violations of normality have minor implications for testing means, especially as  $N$  gets large.
- Violations of normality are more serious for testing variances.

## Relations among distributions – the children of the normal

- Chi-square is drawn from the normal distribution.  $N(0,1)$  squared ( $df=1$ ) and summed ( $df>1$ ).
- $F$  is the ratio of two chi-squares, each divided by its  $df$ . A chi-square divided by its  $df$  is a variance estimate, that is, a sum of squares divided by degrees of freedom.
- $F = t^2$ . If you square  $t$ , you get an  $F$  with 1  $df$  in the numerator.

$$t_{(v)}^2 = F_{(1,v)}$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

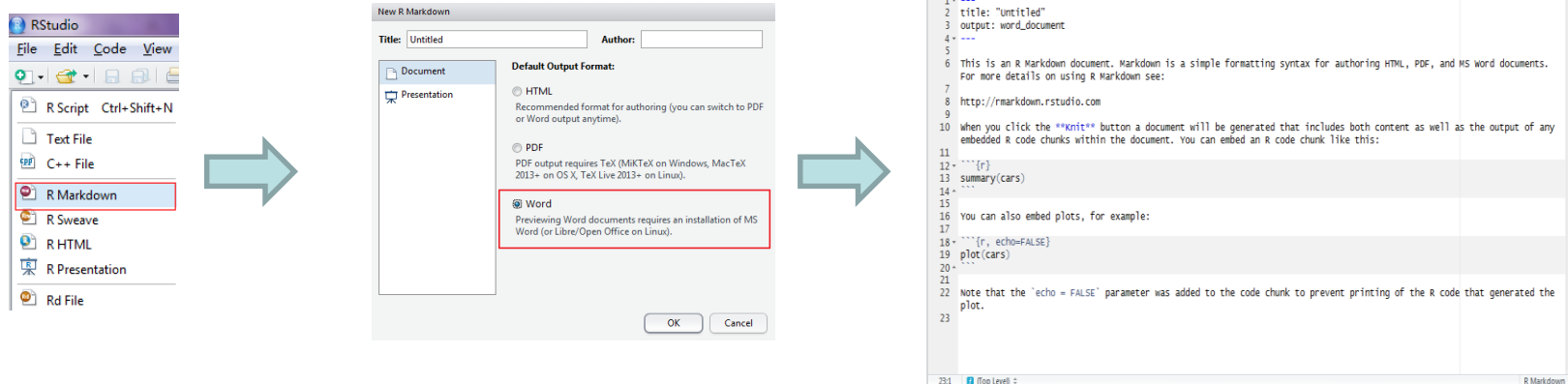
# How to use R+Rstudio+knitr for writing assignments

## 1. Installation

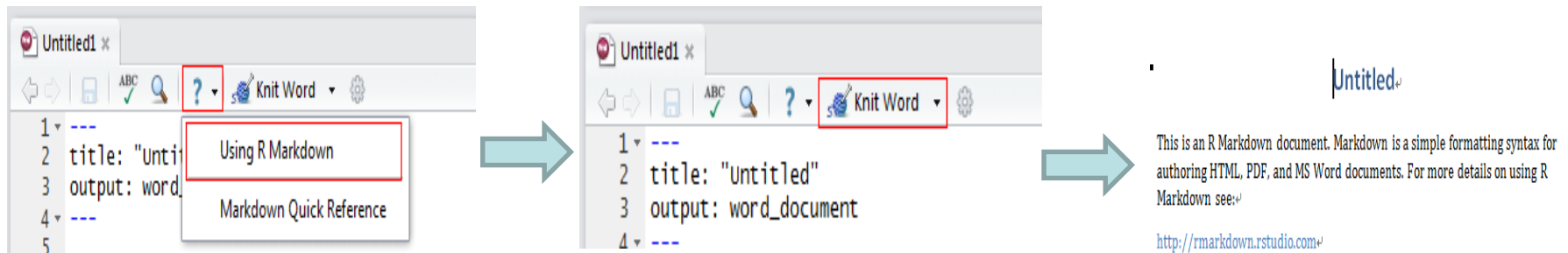
1. Install R <http://ftp.ctex.org/mirrors/CRAN/>
2. Install Rstudio <http://www.rstudio.com/ide/download/preview>.
3. Install package knitr `install.packages("knitr")`
4. Install other packages (following suggestions)

## 2. Usage

1. Open Rstudio, create a new R markdown file



2. Check markdown grammar, export to WORD doc file



# Markdown file.Rmd

```
---
output: word_document
---
# Assignment 2
## **Probability distribution**
```

Xinhai Li  
Group 1, ID: 123456789

```
## Data
```

I measured the weight of 22 rats.

```
```{r}
obs = c(7,5,6,6,7,5,3,4,5,8,2,4,5,6,7,6,4,5,9,3,6,4)
mean = mean(obs)
SD = sd(obs)
```
```

```
```{r}
mean
SD
```
```

```
```{r fig.width=5,fig.height=4}
hist(obs, freq=F)
```
```

**Figure 1. Histogram of rat weight**

```
```{r fig.width=5,fig.height=4}
x = seq(min(obs),max(obs),by=.1)
norm = dnorm(x, mean=mean,sd=SD)
hist(obs, freq=F)
lines(x, norm,type='l',col='red')
abline(v=c(mean-2*SD, mean+2*SD),col='brown')
```
```

**Figure 2. Probability distribution of the weight of rats and the 95% confident interval**

**## Result**

The 95% confident interval of the rat weight is:

```
```{r}
c(qnorm(.025, mean, SD), qnorm(.975, mean, SD))
```
```



# Report example

## Assignment 2

### Probability distribution

Xinhai Li

Group1, ID: 123456789

### Data

I measured the weight of 22 rats.

```
obs = c(7, 5, 6, 6, 7, 5, 3, 4, 5, 8, 2, 4, 5, 6, 7, 6, 4, 5, 9, 3, 6, 4)
mean = mean(obs)
SD = sd(obs)
```

Mean

```
## [1] 5.318
```

SD

```
## [1] 1.701
```

```
hist(obs, freq=F)
```

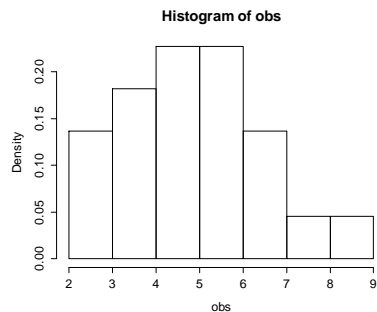


Figure 1. Histogram of rat weight

# Assignment

General objectives: **learn** about hypothesis testing, t-test, and R.

You have a number of **specific** objectives in this assignment. I will provide you with the programs that you will use to test hypotheses and data that **you** develop.

For the report, you will provide a **brief introduction** to the data set, formally state the hypotheses that you are going to test ( $H_0$ 's and  $H_a$ 's), and provide a print out's of the data set, programs and their output. Indicate in your **results and discussion** section what you found, i.e. did you reject your null, and the conclusions that you have drawn from the analysis.

# Tasks

Develop a t test experimental design. Generate your own data and **FORMALIZE** your hypotheses.

Make sure that you go through all of the steps; esp. satisfy assumptions:

- random sampling
- the level of measurement is interval-ratio
- the sampling distribution is normal

# R script for a t test

```
# input data
```

```
lines <-
```

```
"ID x y
```

```
1 208 197
```

```
2 202 150
```

```
3 203 255
```

```
4 200 134
```

```
5 205 266
```

```
6 206 200
```

```
7 207 189
```

```
8 208 186
```

```
9 203 215
```

```
10 210 199"
```

```
weight <- read.table(con <- textConnection(lines), header=TRUE)
```

```
close(con)
```

```
# check data
```

```
weight
```

```
head(weight)
```

```
# t test
```

```
t.test(weight$x, weight$y, var.equal = F) # two samples
```

```
t.test(weight$x, mu = 200) # one sample
```

```
shapiro.test(weight$x) # check normal distribution
```

```
shapiro.test(weight$y)
```

```
hist(weight$y)
```