

Hypothesis Testing

- T test (paired test, one or two samples, variance equal or not)
- Chi square test
- Type I and Type II Errors

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

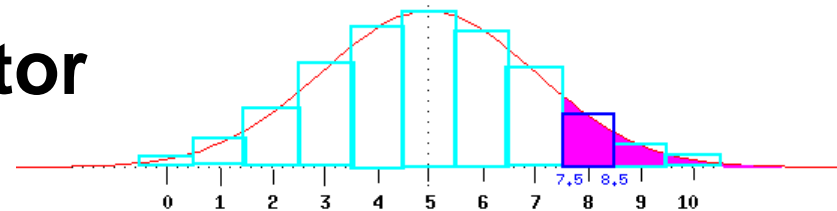
sample estimates:

prop 1 prop 2

0.6190476 0.5172414

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, 20, .25, lower.tail = F)`
- 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.

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

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

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

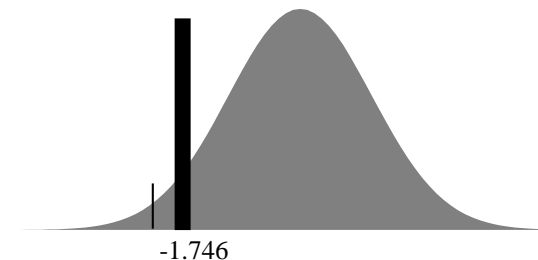
$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{10.71}{10} + \frac{10.71}{8}} = \sqrt{2.41} = 1.55\text{cm}$$

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

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

$$P = 0.0043$$

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)

Compare two indices of diversity

EXAMPLE 8.19

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.

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

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

Hypothesis testing for non-normal distributions

Chi-square test

- A fundamental problem in genetics is determining whether the experimentally determined data fits the results expected from theory
- How can you tell if an observed set of offspring counts is legitimately the result of a given underlying simple ratio?

$$Aa \times Aa$$

- For example, you do a cross and see 290 purple flowers and 110 white flowers in the offspring. This is pretty close to a 3/4 : 1/4 ratio, but how do you formally define "pretty close"? What about 250:150?

Goodness of fit

- Mendel has no way to solve this problem. Shortly after the rediscovery of his work in 1900, Karl Pearson and R. A. Fisher developed the “chi-square” test for this purpose.
- The chi-square test is a “goodness of fit” test: it answers the question of how well do **observed** data fit **expectations**.
- Theory for how the offspring will be distributed (null hypothesis): for a self-pollination heterozygote, the offspring will appear in a ratio of 3/4 dominant to 1/4 recessive.

Formula

- First determine the number of each phenotype that have been observed and how many would be expected given basic genetic theory.
- Then calculate the chi-square statistic using the formula.

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}$$

	exp	obs
purple	300	290
white	100	110

- The “X” is the Greek letter chi; the “ \sum ” is a sigma, it means to sum the following terms for all phenotypes. “obs” is the number of individuals of the given phenotype observed; “exp” is the number of that phenotype expected from the null hypothesis.
- Note that you must use the **number** of individuals, the counts, and NOT proportions, ratios, or frequencies.

Example

As an example, you count offspring, and get 290 purple and 110 white flowers. This is a total of 400 (290 + 110) offspring.

We expect a 3/4 : 1/4 ratio, i.e.

we expect $400 \times 3/4 = 300$ purple, and $400 \times 1/4 = 100$ white.

Purple, obs = 290 and exp = 300.

White, obs = 110 and exp = 100.

Then plugging into the formula:

$$X^2 = \sum \frac{(obs - exp)^2}{exp}$$

$$X^2 = (290 - 300)^2 / 300 + (110 - 100)^2 / 100$$

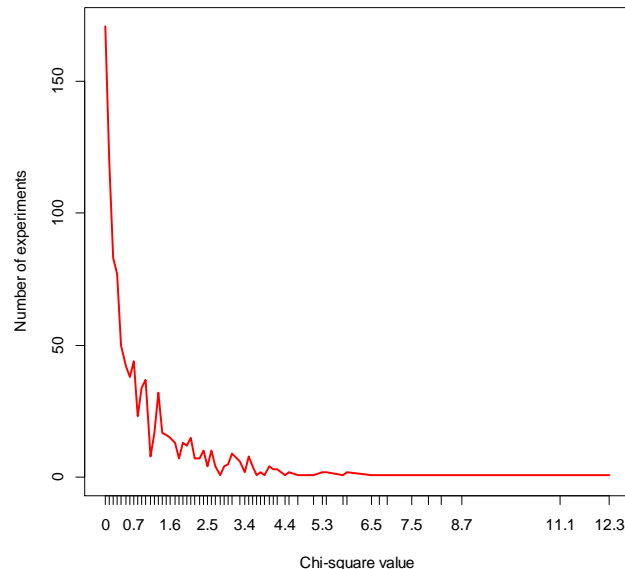
$$= 0.333 + 1.000$$

$$= 1.333.$$

```
1- pchisq(1.333, df = 1) # 0.248
```

Chi-square distribution

- Although the chi-square distribution can be derived through math theory, we can also get it experimentally
- Do the same self-pollination of a Aa heterozygote 1000 times, which should give the 3/4 : 1/4 ratio. For each experiment we calculate the chi-square value, then plot them all on a graph
- The x-axis is the chi-square value calculated from the formula. The y-axis is the number of individual experiments that got that chi-square value.

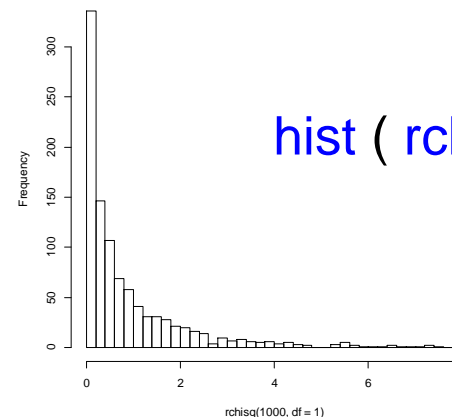


```
plot(table(round(rchisq(1000, df = 1),1)),  
      type = 'l', col = 'red',  
      xlab = 'Chi-square value',  
      ylab = 'Number of experiments')
```

Chi-square distribution

- Most experiments give a small chi-square value (the hump in the graph).
- Note that all the values are greater than 0: that's because we squared the (obs - exp) term
- Sometimes you get really wild results, with obs very different from exp: the long tail on the graph. Really odd things occasionally do happen by chance alone (for instance, you might win the lottery).

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$$



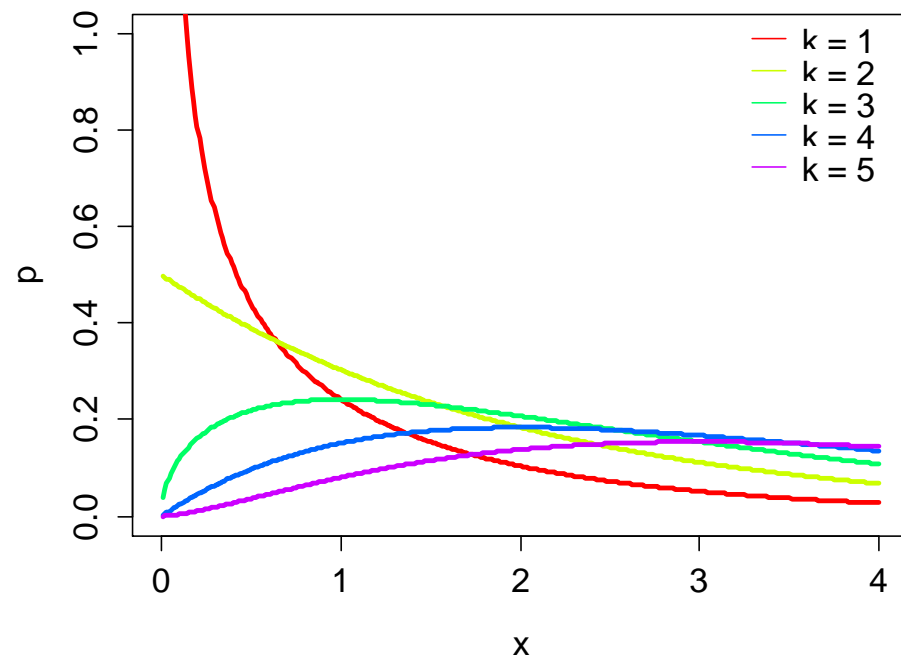
```
hist ( rchisq (1000, df =1),  
      nclass=30 )
```

PDF of Chi-square distribution

$$\frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$$

k: degrees of freedom

$\Gamma(n) = (n-1)!$ (Gamma(n))



```
x <- seq(0.01,4, by=0.01)
```

```
plot(x, dchisq(x,1), type="l", xlab='x',  
      ylab="p", xlim=c(0,4), ylim=c(0,1))
```

```
for(i in 1:5){  
  lines(x, dchisq(x,i), col=rainbow(5)[i], lwd=3)  
  legend(3.2, .11-i/15, paste('k =', i, sep=' '),  
        lty = 1, col = rainbow(5)[i],  
        box.lty=0, cex=1)  
}
```


The critical question

- How do you judge an odd result (e.g., 290:110) is right or wrong?

Most of the time the results fit expectations pretty well, but occasionally very skewed distributions of data occur even though you performed the experiment correctly, based on the correct theory.

- The simple answer is: you can never tell for certain that a given result is “wrong”, that the result you got was completely impossible based on the theory you used. All we can do is to determine whether a given result is likely or unlikely.

The critical question

- Key point: there are two ways of getting a high chi-square value: an unusual result from the correct theory, or a result from another theory.
- Using the example here, how can you tell if your 290:110 offspring ratio really fits a $3/4 : 1/4$ ratio (as expected from selfing a heterozygote) or whether it was the result of a $1/2 : 1/2$ ratio?
- You can't be certain, but you can at least determine whether your result is reasonable.

Reasonable result

- What is a “reasonable” result is subjective.
- For most work, a result is said to not differ significantly from expectations if it could happen at least 1 time in 20.
- We use “fail to reject” instead of “accept”.
- “1 time in 20” can be written as a probability value $p = 0.05$

Degrees of freedom

- A critical factor in using the chi-square test is the “degrees of freedom”, which is essentially the number of independent random variables involved.
- Degrees of freedom is simply the number of classes of offspring minus 1.
- For our example, there are 2 classes of offspring: purple and white. Thus, degrees of freedom (d.f.) = $2 - 1 = 1$.

Critical Chi-square

- Critical values for chi-square are found on tables, sorted by degrees of freedom and probability levels. Or you can have it from statistical software. Be sure to use $p = 0.05$.

`qchisq(0.95, df = 1)`

- If your calculated chi-square value is greater than the critical value from the table, you “reject the null hypothesis”.
- If your chi-square value is less than the critical value, you “fail to reject” the null hypothesis.

Chi-square table

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401

Chi square test

- In our example of 290 purple to 110 white, we calculated a chi-square value of 1.333, with 1 degree of freedom.
- Looking at the table, 1 d.f. is the first row, and $p = 0.05$ is the sixth column. Here we find the critical chi-square value, 3.841.
- Since our calculated chi-square, 1.333, is less than the critical value, 3.841, we “fail to reject” the null hypothesis. Thus, an observed ratio of 290 purple to 110 white is a good fit to a 3/4 to 1/4 ratio.

`chisq.test (c (290,110), p = c (0.75,0.25))`

$P = 0.248$

















Another example: from Mendel

phenotype	observed	expected proportion	expected number
round yellow	315	9/16	312.75
round green	101	3/16	104.25
wrinkled yellow	108	3/16	104.25
wrinkled green	32	1/16	34.75
total	556	1	556





Finding the expected numbers

- You are given the observed numbers, and you determine the expected proportions from a Punnett square.
- To get the expected numbers of offspring, first add up the observed offspring to get the total number of offspring. In this case, $315 + 101 + 108 + 32 = 556$.

Punnett Square of Dihybrid Cross
Gametes from RrYy parent

	RY	Ry	rY	ry
RY	RRYY 	RRYy 	RrYY 	RrYy 
Ry	RRYy 	RRyy 	RrYy 	Rryy 
rY	RrYY 	RrYy 	rrYY 	rrYy 
ry	RrYy 	Rryy 	rrYy 	rryy 

F₁ cross: RrYy × RrYy

-  round yellow
-  round green
-  wrinkled yellow
-  wrinkled green

- Then multiply total offspring by the expected proportion:
 - expected round yellow = $9/16 * 556 = 312.75$
 - expected round green = $3/16 * 556 = 104.25$
 - expected wrinkled yellow = $3/16 * 556 = 104.25$
 - expected wrinkled green = $1/16 * 556 = 34.75$
- Note that these add up to 556, the observed total offspring

Calculating the Chi-square value

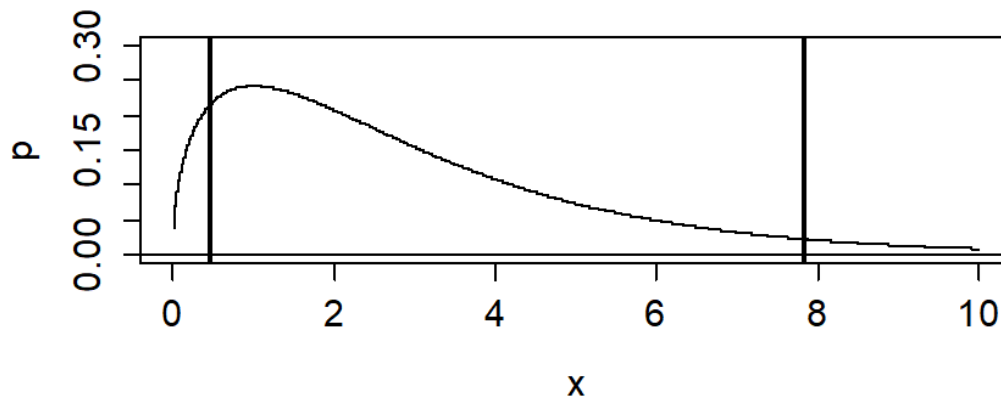
$$X^2 = \sum \frac{(obs - exp)^2}{exp}$$

$$\begin{aligned} X^2 &= (315 - 312.75)^2 / 312.75 \\ &\quad + (101 - 104.25)^2 / 104.25 \\ &\quad + (108 - 104.25)^2 / 104.25 \\ &\quad + (32 - 34.75)^2 / 34.75 \\ &= 0.016 + 0.101 + 0.135 + 0.218 \\ &= 0.470. \end{aligned}$$

```
chisq.test (c(315,101,108,32), p = c(9/16,3/16,3/16,1/16))
```

D.F. and critical value

- Degrees of freedom is 1 less than the number of classes of offspring. Here, $4 - 1 = 3$.
- For 3 d.f. and $p = 0.05$, the critical chi-square value is 7.815.
- Since the observed chi-square (0.470) is less than the critical value, we fail to reject the null hypothesis. We can not reject Mendel's conclusion that the observed results for a $9/16 : 3/16 : 3/16 : 1/16$ ratio so far.
- It should be mentioned that all of Mendel's numbers are unreasonably accurate.



```
x <- seq(0.01,10, by=0.01)
plot(x, dchisq(x,3), type="l", xlab='x',
      ylab="p", xlim=c(0,10), ylim=c(0,.3))
abline(0,0)
abline(v=0.47, lwd=3) # Mendel's result
abline(v=qchisq(.95,3), lwd=3) # 5%
```

Chi square assumptions

The chi square test can only be used on data that has the following characteristics:

- The data must be in the form of frequencies (count)
- The frequency data must have a precise numerical value and must be organised into categories or groups.
- The expected frequency in any one cell of the table must be greater than 5.
- The total number of observations must be greater than 20.

Fisher's exact test

- A statistical significance test used in the analysis of contingency tables
- Be appropriate for small sample sizes (valid for large sample sizes as well)
- One of a class of exact tests, because the significance of the deviation from a null hypothesis (e.g., P-value) can be calculated exactly, rather than relying on an approximation that becomes exact when the sample size grows to infinity

Fisher's exact test

	Men	Women	<i>Row Total</i>
Yoga	2	8	10
Non-yoga	10	4	14
<i>Column Total</i>	12	12	24

- There are 10 of these 24 students are doing Yoga regularly, and that 12 of the 24 are female.
- **Assuming the null hypothesis that men and women are equally likely Yoga**, what is the probability that these 10 Yoga lovers would be so unevenly distributed between the women and the men?
 - ✓ If we choose 10 Yoga lovers at random, what is the probability that 8 or more of them would be among the 12 women, and only 2 or fewer from among the 12 men?

Fisher's exact test

	Men	Women	Row Total
Yoga	<i>a</i>	<i>b</i>	$a + b$
Non-yoga	<i>c</i>	<i>d</i>	$c + d$
Column Total	$a + c$	$b + d$	$a + b + c + d$ (=n)

Fisher showed that the probability of obtaining any such set of values was given by the [hypergeometric distribution](#):

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a! b! c! d! n!}$$

where $()$ is the binomial coefficient and the symbol $!$ indicates the factorial operator.

The formula above gives the exact hypergeometric probability of observing this particular arrangement of the data, assuming the given marginal totals, on the null hypothesis that men and women are equally likely to do Yoga.

R for Fisher's exact test

```
TeaTasting <- matrix(c(8, 2, 3, 7), nrow = 2,
  dimnames = list(Guess = c("Milk", "Tea"),
    Truth = c("Milk", "Tea")))
fisher.test(TeaTasting, alternative = "greater")
```

Guess	Truth	
	Milk	Tea
Milk	8	3
Tea	2	7

```
Fisher's Exact Test for Count Data
data: TeaTasting
p-value = 0.03489
alternative hypothesis: true odds ratio
is greater than 1
95 percent confidence interval:
1.155327 Inf
sample estimates: odds ratio 8.153063
```

```
## A 4 x 4 table Agresti (2002, p. 57) Job Satisfaction
Job <- matrix(c(1,2,1,0, 3,3,6,1, 10,10,14,9, 6,7,12,11), 4, 4,
  dimnames = list(income = c("< 15k", "15-25k", "25-40k", "> 40k"),
    satisfaction = c("VeryD", "LittleD", "ModerateS", "VeryS")))
fisher.test(Job)
fisher.test(Job, simulate.p.value = TRUE, B = 1e5)
```


Test homogeneity of variance – one sample

$$H_0 : \sigma^2 = \sigma_0^2$$

Test the null that the population variance has some specific value. Pick an alpha value. Then:

$$\chi^2_{(N-1)} = \frac{(N-1)s^2}{\sigma_0^2}$$

Plug hypothesized population variance and sample variance into equation along with sample size we used to estimate variance. Compare to chi-square distribution.

```
var.test()  
bartlett.test()  
fligner.test()
```

Example of chi square test of variance (one tailed test)

Test about variance of height of people in inches. Sample 30 people at random and measure the heights.

$$H_0 : \sigma^2 \geq 6.25; H_1 : \sigma^2 < 6.25.$$

$$N = 30; s^2 = 4.55$$

$$\chi^2_{(N-1)} = \frac{(N-1)s^2}{\sigma_0^2}$$

$$\chi^2_{29} = \frac{(29)(4.55)}{6.25} = 21.11$$

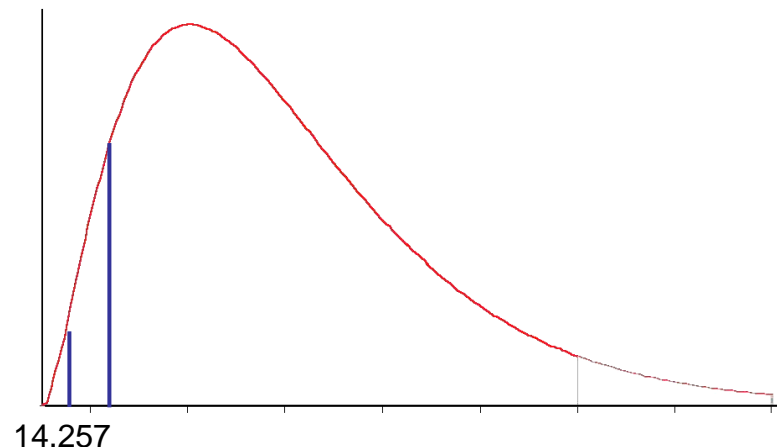
$$p = 0.855$$

1 tailed test on small side.

Set alpha=.01.

For $p=.01$, the value of chi-square is 14.257. Cannot reject null.

`qchisq(0.01, df = 29)`



Example of chi square test of variance (two tailed test)

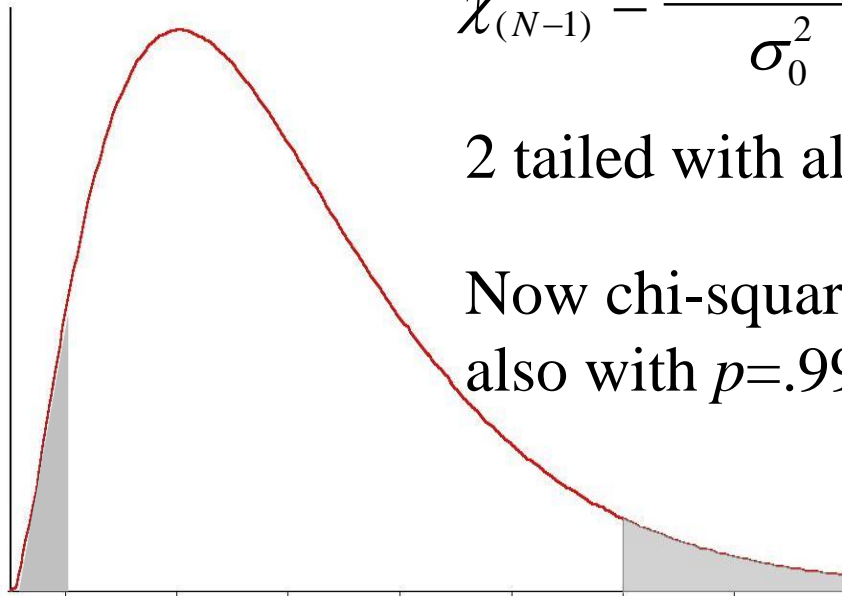
$$H_0 : \sigma^2 = 6.25; H_1 : \sigma^2 \neq 6.25.$$

$$N = 30; s^2 = 4.55$$

$$\chi^2_{(N-1)} = \frac{(N-1)s^2}{\sigma_0^2} \quad \chi^2_{29} = \frac{(29)(4.55)}{6.25} = 21.11$$

2 tailed with alpha=.01.

Now chi-square with $\nu=29$ and $p=.005$ is 13.121 and also with $p=.995$ the result is 52.336. N. S. either way.



`qchisq(0.995, df = 29)`

Confidence intervals for the variance

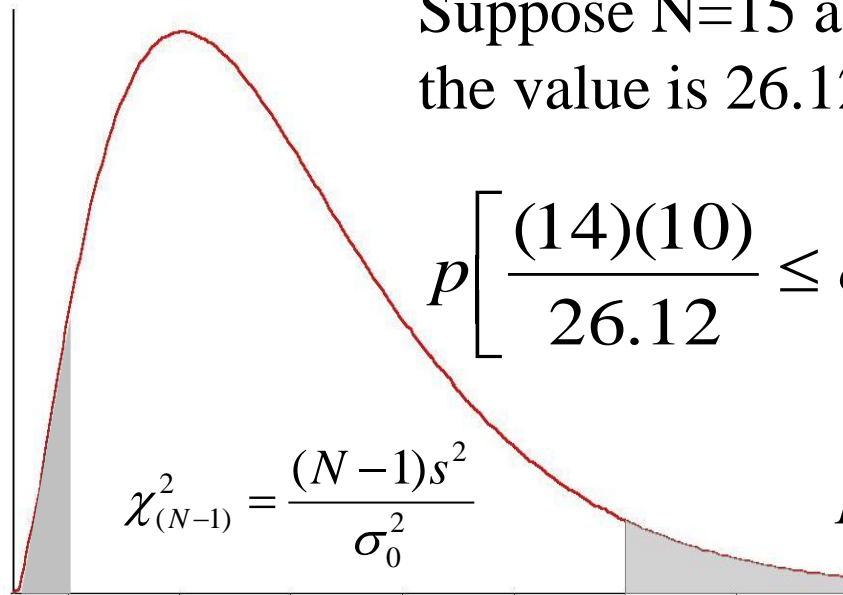
We use s^2 to estimate σ^2 . It can be shown that:

$$p\left[\frac{(N-1)s^2}{\chi^2_{(N-1);.025}} \leq \sigma^2 \leq \frac{(N-1)s^2}{\chi^2_{(N-1);.975}}\right] = .95$$

Suppose $N=15$ and s^2 is 10. Then $df=14$ and for $p=.975$ the value is 26.12. For $p=.025$ the value is 5.63.

$$p\left[\frac{(14)(10)}{26.12} \leq \sigma^2 \leq \frac{(14)(10)}{5.63}\right] = .95$$

$$p[5.36 \leq \sigma^2 \leq 24.87] = .95$$



Test homogeneity of variance – two independent samples

$$F = \frac{s_1^2}{s_2^2}$$

ZAR p138

Test homogeneity of variance – two independent samples

EXAMPLE 8.8 The data are the numbers of moths caught during the night by eleven traps of one style and eight traps of a second style.

The two-tailed variance ratio test for the hypotheses $H_0: \sigma_1^2 = \sigma_2^2$ and $H_A: \sigma_1^2 \neq \sigma_2^2$.

Trap type 1 *Trap type 2*

41 52

34 57

33 62

36 55

40 64

25 57

31 56

37 55

34

30

38

$n_1 = 11$ $n_2 = 8$

$v_1 = 10$ $v_2 = 7$

$SS_1 = 218.73 \text{moths}^2$ $SS_2 = 107.50 \text{moths}^2$

$s_1^2 = 21.87 \text{moths}^2$ $s_2^2 = 15.36 \text{moths}^2$

$$F = \frac{s_1^2}{s_2^2} = \frac{21.87}{15.36} = 1.42$$

$$\alpha = 0.05 \quad F_{0.05(2),10,7} = 4.76$$

Do not reject

H_0

$$s_p^2 = \frac{218.73 \text{moths}^2 + 107.50 \text{moths}^2}{10 + 7} = 19.19 \text{moths}^2$$

You need to know

- How to turn a question into hypotheses
- Failing to reject the null hypothesis DOES NOT mean that the null is true
- Every test has assumptions
 - A statistician can check all the assumptions
 - If the data does not meet the assumptions there are non-parametric versions of the tests

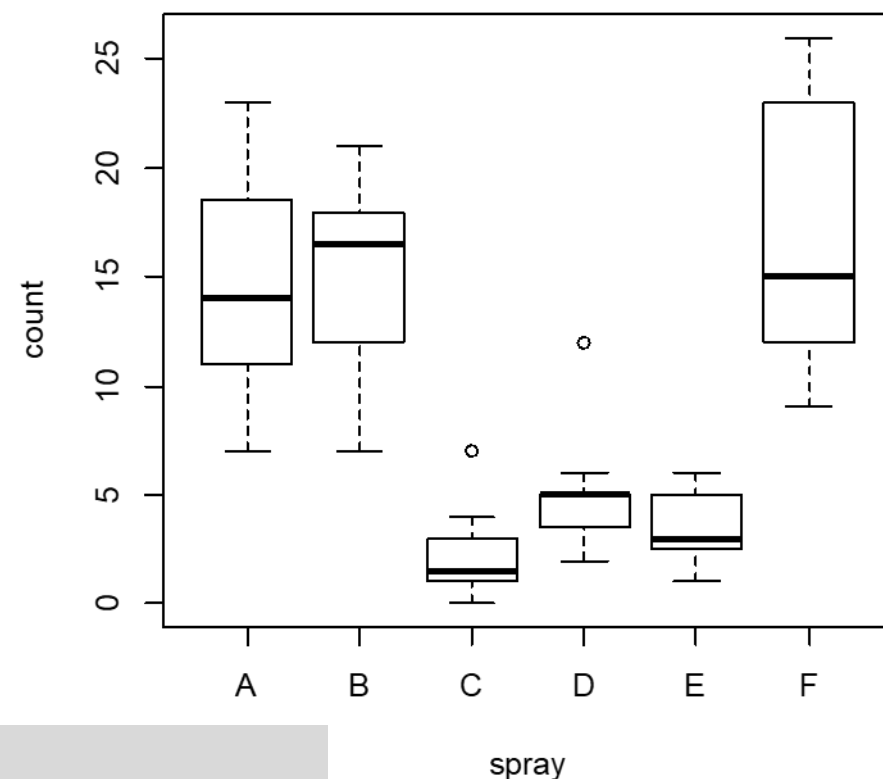
Common mistakes in hypothesis testing

- Lack of independence
- Violation of normality
 - Highly skewed data
- Assume equal variances and the variances are not equal
(Did not do variance test)

ALWAYS graph your data first to assess symmetry and variance

Check variance

- `bartlett.test()`, sensitive to outliers
- `var.test()`, Fisher's F test, sensitive to outliers
- `fligner.test()`, performs a Fligner-Killeen (median) test of the null that the variances in each of the groups (samples) are the same. It is a non-parametric test which uses the ranks of the absolute values.



```
plot(count ~ spray, data = InsectSprays)
```

```
bartlett.test(InsectSprays$count, InsectSprays$spray)
```

```
var.test(InsectSprays$count, InsectSprays$spray)
```

```
fligner.test(InsectSprays$count, InsectSprays$spray)
```

```
fligner.test(count ~ spray, data = InsectSprays)
```

**A standard hypothesis Testing
(the 5 steps): a whole picture?**

No!

Type I and Type II Errors

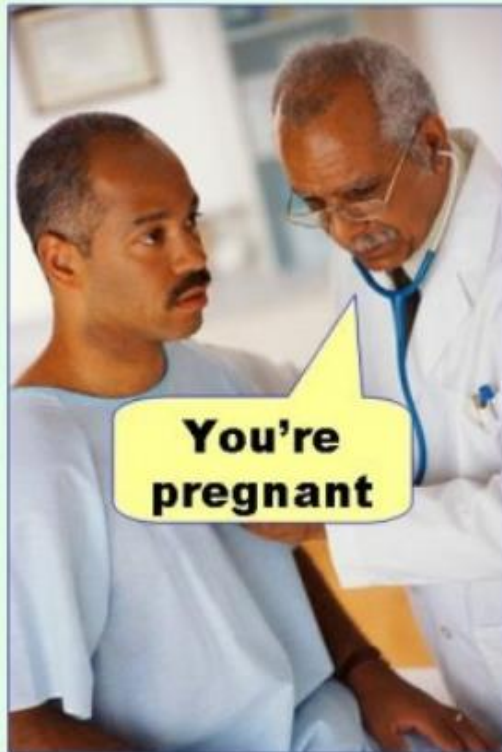
		NULL HYPOTHESIS	
		TRUE	FALSE
D E C I S I O N	Reject the null hypothesis	Type I error α Rejecting a true null hypothesis	CORRECT $1 - \beta$
	Fail to reject the null hypothesis	CORRECT $1 - \alpha$	Type II error β Failing to reject a false null hypothesis

@游识猷 V

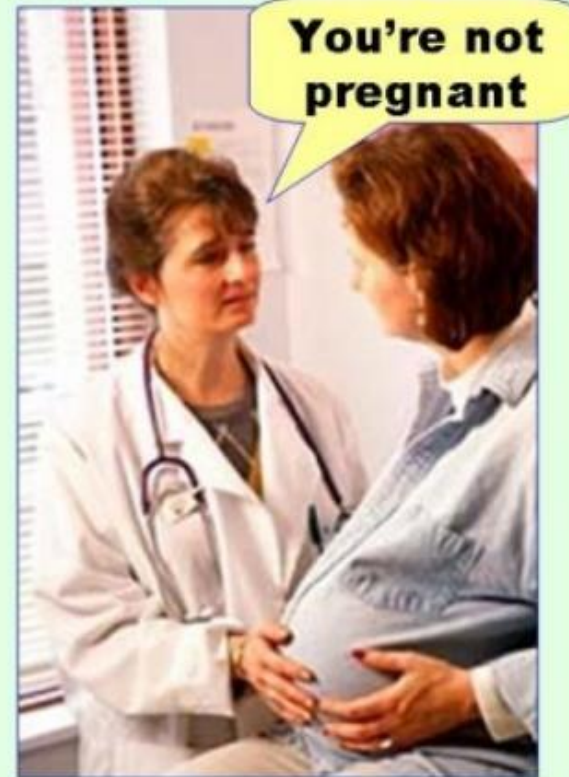
统计中的Type I error 和Type II error 🤪 来自推特aLittleMedic

[收起](#) | [查看大图](#) | [向左旋转](#) | [向右旋转](#)

Type I error
(false positive)



Type II error
(false negative)



@游识猷

Null hypothesis:
everyone is normal
not pregnant



Controlling Type I and Type II Errors

- α , β , and n are interrelated. If one is kept constant, then an increase in one of the remaining two will cause a decrease in the other.
- For any fixed α , an increase in the sample size n will cause a decrease in β .
- For any fixed sample size n , a decrease in α will cause an increase in β .
- Conversely, an increase in α will cause a decrease in β .
- To decrease both α and β , increase the sample size n .

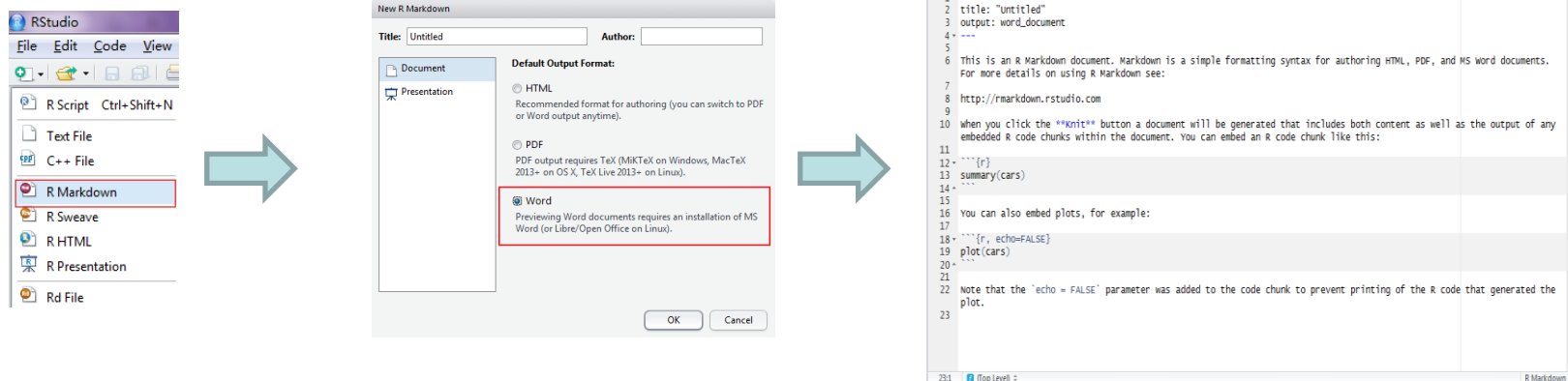
How to use R+Rstudio+knitr for writing assignments

1. Installation

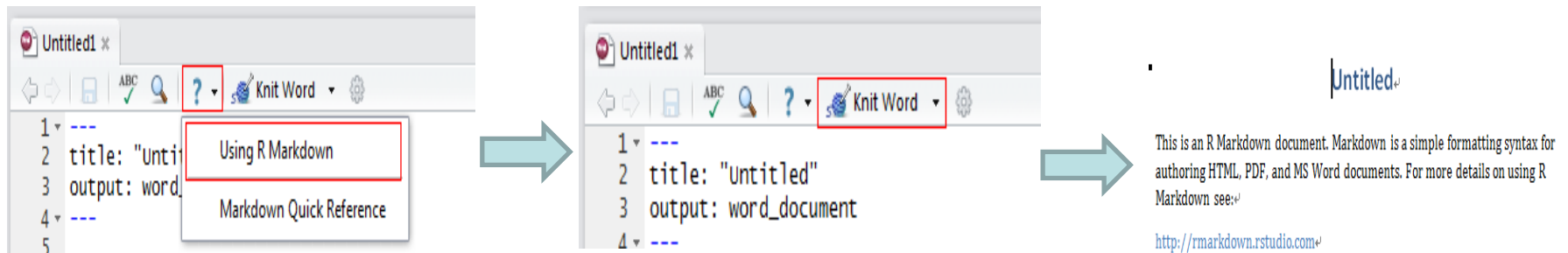
1. Install R <http://ftp.ctex.org/mirrors/CRAN/>
2. Install Rstudio <http://www.rstudio.com/ide/download/preview>.
3. Install package knitr `install.packages("knitr")`
4. Install other packages (following suggestions)

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)
```

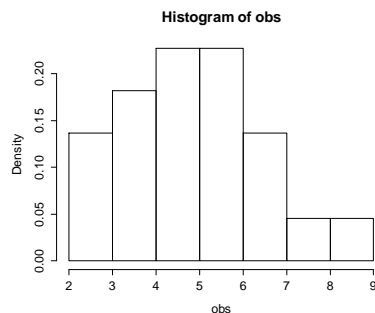


Figure 1. Histogram of rat weight

Assignment

General objectives: learn about hypothesis testing, Chi square test, and R.

For the report, you will provide a **brief introduction** to the data set (a contingency table), 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.