Two Sample Problems

```
*************two.tex ****************
```

It is often of interest to test a hypothesis that the means of two populations are the same. Imagine two populations $\mathcal{P}_1$ and $\mathcal{P}_2$ of items we are interested in buying. Assume that we can observe some quality feature (length of life, durability, etc). Let μ_1 and μ_2 represent the means of these observable quality features in the populations $\mathcal{P}_1$ and $\mathcal{P}_2$. We are interested in testing $H_0 : \mu_1 = \mu_2$ against one-sided or two sided alternative. For example, items from $\mathcal{P}_1$ are substantially less expensive than those from $\mathcal{P}_2$. Accepting H_0 would mean we are buying P_1 items. Or, we rejected H_0 in favor of $H_1 : \mu_1 < \mu_2$. If the items from $\mathcal{P}_1$ and $\mathcal{P}_2$ are equally priced, then the rational decision would be to buy $\mathcal{P}_2$ items.

The above setup is equivalent to testing that the difference between the two means is zero. Without much additional effort one generalize the testing problem. For example, one may be interested in the hypothesis that $\mu_1 - \mu_2 = c$, where c is any number.

Depending on the sample sizes, relations between populations, as well as the information we have beforehand, there are several methods for dealing with testing of equality of two means.

These methods are given in the subsequent subsections.

6.1 Known population variances

Mushrooms. Making spore prints is an enormous help in identifying genera and species of mushrooms. To make a spore print, mushroom fans take a fresh, mature cap and lay it on a clean piece of glass. Left overnight or possibly longer the cap should give you a good print. Family of Amanitas is one that has the most poisonous (Amanita Phaloides, Amanita verna, Amanita virosa, Amanita Pantherina, etc) and the most delicious species (Amanita Cesarea, Amanita Rubescens) A. Pantherina = 7 microns A. Rubescens = 5.5 microns.

1. Dr. Mendel injected two groups of rats with two different drugs to determine how the drug affects the speed with which the rats run a maze. The 45 rats treated with drug A needed an average of 17 minutes to run the maze. The standard deviation was 2.3 minutes. The 53 rats treated with drug B needed an average of 19 minutes to run the maze. The standard deviation was 3.6. Is there a significant difference between the effects of the two drugs on the average time it takes the rats to run the maze? Use 10% level of significance.

6.2 Unknown population variances: Small samples

Aerobic Capacity. The peak oxygen intake per unit of body weight, called *the aerobic capacity* of an individual performing a strenuous activity is a measure of work capacity. For comparative study, measurements of aerobic capacities are recorded [1] for a group of 20 Peruvian Highland natives and for a group of 10 U.S. lowlanders acclimatized as adults in high altitudes.

	Peruvian Natives	U.S. Subjects Acclimatized
Sample mean	46.3	38.5
Sample st. deviation	5.0	5.8

Test the hypothesis that the population mean aerobic capacities are the same against one sided alternative. Take $\alpha = 0.05$.

[1] Frisancho, A.R., *Science,* Vol 187, (1975), 317.

```
----------------------------------------------
             Two sample t-test
 Testing H_0: mu1-mu2 = 0  v.s. H_1: mu1-mu2 > 0 .
----------------------------------------------
 :-( Reject H_0.
 p-value= 0 is smaller than alpha= 0.05 .
 t-statistic= 3.821 .
 n1= 20  n2= 10  pooled s= 5.27
 The  1 -sided rejection region is determined by
 0.95  quantile of t distribution with  28  degrees of freedom: 1.699 .
```

6.2.1 Problems

Growth Hormone. An investigation was undertaken to determine how the administration of a growth hormone affects the weight gain of pregnant rats. Weight gains during gestation are recorded for 6 control rats and for 6 rats receiving the growth hormone. The summary of the results[2] is given in the table below.

	Control rats	Hormone rats
Mean	41.8	60.8
Standard deviation	7.6	16.4

(i) State the assumptions about the populations and test to determine if the mean weight gain is significantly higher for the rats receiving the hormone than for the rats in the control group.

(ii) Do the data indicate that you should be concerned about the possible violation of any assumptions? If so, which one?

3. **Eating Disorders.** An example involving heterogeneous variances can be found in an extensive study of eating disorders in adolescents by Gross (1985). Among other things, Gross examined subjects who had a disorder known as bulimia. "Simple bulimia" is a psychological eating disorder involving uncontrollable eating (often called binge eating), coupled with the knowledge that the eating is abnormal and an associated state of dysphoria (feeling bad). In many cases, but not all, binge eating is followed by intentional vomiting or the use of laxatives. When this behavior is present, the disorder is labeled "bulimia with purging." As one of many variables, Gross investigated whether there was a weight difference between people classified in the two categories of bulimia. Although Gross' actual data are not available, the data given below were generated to have the same means and variances as she reported for her subjects. Fictional data have been provided because they are necessary for the application of O'Brien's test for homogeneity of variance. The dependent variable shown on the left of table is the mean percentage deviation of an individual's actual weight from the close to normal - that is, the mean percentage deviation is near zero. If we ignored the unequal variances and simply pooled them, we would obtain $t = 1.87$, a nonsignificant result at $\alpha = .05$.

[2]Sara at al., *Science* 186 (1974), 446

	Original Data (X_{ij})		Transformed Data (r_{ij})	
	Simple	Purging	Simple	Purging
	24.01	10.23	385.87	127.18
	14.50	-6.20	98.63	28.86
	-5.00	-6.13	92.96	28.03
	7.71	-1.88	7.61	-0.19
	35.25	1.83	966.13	6.15
	-22.18	-10.79	738.28	102.59
	-5.13	4.87	95.57	32.85
	-13.27	16.56	327.46	316.50
	9.11	-15.82	18.59	234.04
	2.54	1.04	2.08	2.39
	...	...	...	...
Mean	4.61	-0.83	219.04	79.21
Variance	219.04	79.21	65432.73	8144.20
N	49	32	49	32

Our first step in dealing with these data involves testing for heterogeneity of variance. This is done using the values on the right of table, which have been obtained with O'Brien's transformation. In the above table notice that the means of the transformed values (r_{ij}) are equal to the variances of the original values (X_{ij}), reflecting the fact that the t test we are about to apply on the means of the *transformed* values is actually comparing the variances of the original values. From the means and variances given in the table, we can compute a t test of the null hypothesis that the data were sampled from populations with equal variances.

4. **Streakers.** In the early 1970s, students started a phenomenon called *streaking.* Within a two=week period following the first streaking sighted on campus, a standard psychological test was given to a group of 19 males who were admitted streakers and to a control group of 19 males who were non streakers. Stoner and Watman (*Psychology* Vol. 11, No 4 (1975), 14-16.) reported the following numbers regarding the scores on a test designed to determine extroversion:

Streaker	Non Streaker
$\bar{X} = 15.26$	$\bar{Y} = 13.90$
$s_1 = 2.26$	$s_2 = 4.11$

(a) Construct 95% confidence interval for the difference in population means. Does there appear to be a difference between the two groups?

(b) It may be true that those who admit to streaking differ from those who do not admit to streaking. In light of this possibility, what criticism can be made for the conclusions in the part (a).

5. **Brain tissue.** Specimens of brain tissue are collected by performing autopsies on 9 schizophrenic patients and 9 control patients of comparable ages. A certain enzyme activity is measured for each subject in terms of the amount of substance formed per gram of tissue per hour. The following means and standard deviations are calculated from the data./footnote Wyatt et al., *Science, Vol. 187 (1975), 369.*

	Control subjects	Schizophrenic subjects
Mean	39.8	35.5
St. deviation	8.16	6.93

(a) Test to determine if the mean activity is significantly lower for the schizophrenic subjects than for the control subjects. Use $\alpha = 0.05$.

(b) Construct 99% confidence interval for the mean difference in enzyme activity between the two groups.

6.3 Difference between two population proportions

6.4 Comparing variances in two populations

6.5 Dependent samples: Paired Comparisons

6.5.1 Exercises

Feminism and Authoritarianism. A study[3] compared peoples attitudes toward feminism with their degree of authoritarianism. Two independent samples were used, one consisting of 30 subjects who were rated high in authoritarianism, and a second sample of 31 subjects who were rated low. Each subject was given an 18-item test designed to reveal attitudes on feminism, with scores reported on a scale from 18 to 90 (High scores indicated pro-feminism). Summary statistics from the study are as follows:

Authoritarianism	n	$\bar{X}$	s
High	30	67.7	11.8
Low	31	52.4	13.0

Assume that variances in the 'High' and 'Low' subpopulations are the same.

(a) State H_0. What type of test is appropriate and why?

(b) Perform the test against the two sided alternative. Use $\alpha = 0.05$. (c) Which one-sided alternative will be appropriate in this problem. You may find this piece of Splus output useful:

```
n1= 30  n2= 31  pooled s= 12.425
```

Solution: (a) H_0 says that there is no significant difference between the means in each of two populations.

(b) $s_p = \sqrt{\frac{29 \cdot 11.8^2 + 30 \cdot 13^2}{59}} = 12.425$.

$t = \frac{67.7-52.4}{12.425\sqrt{1/30+1/31}} = \frac{15.3}{3.182} = 4.808$.

```
> 2*(1-pt(4.808, 59))
[1] 1.090644e-05

> qt(0.025, 59)
[1] -2.000995
```

Reject H_0. (p-value ¡ α or t is in rejection region $(-\infty, -2) \cup (2, \infty)$)

(c) $H_1 : \mu_1 > \mu_2$ (The mean of the 'high' subpopulation is greater than the mean of the 'low' subpopulation.)

P.**Teaching by imitation** Howel, D (1994) reports the following results from an experiment. For 6 month psychologist worked with a group of 15 severely retarded individuals in an attempt to teach them self-care skills trough imitation. For a second 6 month period the psychologist used psychically guided practice with the same individuals. For each 6 month session the ratings on the required assistance level (high=bad) for each person are recorded. The data are summarized in the following table.

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Imitation	14	11	19	8	4	9	12	5	14	17	18	0	2	8	6
Guidance	10	13	15	5	3	6	7	9	16	10	13	1	2	3	6

Aggression in children. Albert Bandura has conducted a number of studies on aggression in children. In one study (Bandura, Ross, and Ross, 1963), one group of children were shown a film showing violence. Another group was not shown the film. Afterward, both groups were allowed to play with Bobo dolls in a playroom, and the number of violent contacts were counted. The following data are obtained:

[3] Sarup, G. (1976). Gender, authoritarianism, and attitude towards feminism. *Soc.Behav.Personality* **4** *57-64.*

Subject	Before	After
1	137	130
2	201	180
3	167	150
4	150	153
5	173	162

	n	$\bar{X}$	s
Beginning			
Treatment	14	165.09	71.20
Control	10	159.00	67.45
End			
Treatment	14	123.63	74.09
Control	10	162.17	67.01

Film Group	20	65	41	80	52	35	15	75	60	50	33
No Film Group	5	20	0	0	10	8	30	13	0	25	

(i) Test the hypothesis that the Film Group has significantly higher number of violent contacts. Take $\alpha = 0.10$. Assume the unknown variances are equal. [Useful numbers: $\bar{X} = 47.82, s_X^2 = 452.16, \bar{Y} = 11.1, s_Y^2 = 116.77$.]

(ii) What would you change in the design of the experiment so that the problem becomes paired data problem.

5. In the past, many bodily functions were thought to be beyond conscious control. However, recent experimentation suggests that it may be possible for a person to control certain body functions if that person is trained in a program of *biofeedback* exercises. An experiment is conducted to show that blood pressure levels can be consciously reduced in people trained in this program. The blood pressure measurements (in millimeters of mercury) listed in the table represent readings before and after the biofeedback training of five subjects.

(a) If we want to test whether the mean blood pressure decreases after the training, what are the appropriate null and alternative hypotheses?

(b) Perform the test in (a) with $\alpha = 0.05$.

(c) What assumptions are needed to assure validity of results.

[$D_i = \{7, 21, 17, -3, 11\}, \bar{D} = 10.6, s_D = 9.32, t = 10.6/(9.32/sqrt5) = 2.54, t_{4,0.95} = 2.131847$.]

6. Helping smokers kick the habit is big business in today's no-smoking environment. One of the more commonly used treatments according to an article in the *Journal of Imagination, Cognition and Personality* (Spanos et al., 1992/93) is Spiegel's three point message:

- • For your body, smoking is poison.
- • You need your body to live.
- • You owe your body this respect and protection.

To determine the effectiveness of this treatment, the authors conducted a study consisting of a sample of 52 smokers placed in two groups, a Spiegel treatment group or a Control group (no treatment). Each participant was asked to record the number of cigarettes he or she smoked each week. The results for the study are shown below for the beginning period and the end-of-experiment period.

Test the hypothesis that the difference in means between treatment and control groups **at the end of experiment** is significant.

Use one sided alternative and $\alpha = 0.05$. Assume that population variances are the same ($\sigma_1^2 = \sigma_2^2$), though unknown.

Interpret results.

7. Of 40 recently hired marksmen for the *Sherwood Rascals* company, half were assigned to a special one-day orientation course (held by Robin Hood himself), and half received no orientation. After 3 months, a special committee was conducting "on-the-job" evaluations and they reported the following results:

Received Orientation	No Orientation
$n_1 = 40$	$n_2 = 40$
$\bar{X}_1 = 84.1$	$\bar{X}_2 = 81.4$
$s_1 = 3.6$	$s_2 = 4.1$

Do the data indicate that the marksmen receiving orientation performed better than those who did not? Take $\alpha = 0.05$.

8. Little John had revealed the results of a secret shooting match between Robin Hood and the Sheriff of Nottingham.

	Robin	Sheriff
Number of Shoots	$n_1 = 16$	$n_2 = 20$
Average number of points	$\bar{X}_1 = 7.5$	$\bar{X}_2 = 6.9$
Sample standard deviation	$s_1 = 2.9$	$s_2 = 3.1$

Using the data above try to prove that Robin is better archer. Use $\alpha = 0.05$.

9. Decision makers of *Sherwood Rascals* company have a rough time. They have to choose between two suppliers of arrows: *Arrows Unlimited* and *Sharp Wily*.

To make intelligent and statistically sound choice 4 randomly chosen archers shoot at the target with 10 arrows from each supplier. The number of arrows that hit the target is given.

Archer	*Arrows Unlimited*	*Sharp Wily*
1	7	5
2	8	7
3	5	5
4	9	7

Test the hypothesis that two producers produce arrows of the same precision. Choose $\alpha = 0.05$. [HINT: USE PAIRED t-TEST. $\bar{d} = 1.25$ AND $s_d = 0.957$.]

10. Ten individuals participated in a a study on the effectiveness of two sedatives, A and B. Each individual was given A on some nights and B on other nights. The average number of hours he slept after taking the first sedative is compared with the normal amount of sleep; a similar comparison is made with the second drug. Table below gives the increase in sleep due to each sedative for each individual. (A negative value indicates a decrease in sleep.)

Patient	Drug A	Drug B
1	1.9	0.7
2	0.8	1.6
3	1.1	-0.2
4	0.1	-1.2
5	-0.1	-0.1
6	4.4	3.4
7	5.5	3.7
8	1.6	0.8
9	4.6	0.0
10	3.4	2.0

(a) Compute the mean increase for drug A and the mean increase for drug B.
(b) For each individual, compute the difference (increase for drug A minus increase for drug B).
(c) Compute the mean of these differences.
(d) Verify that the mean of the differences is equal to the difference between the means.
(e) Test the hypothesis ...

11. Two machines are used for filling plastic bottles with a net volume of 12.0 ounces. The filling processes can be assumed normal, with standard deviations $\sigma_1 = 0.015$ and $\sigma_2 = 0.018$. The quality control department suspects that both machines fill to the same net volume, whether or not this volume is 12.0 ounces. A random sample is taken from the output of each machine.

Machine 1:	12.03	12.04	12.05	12.05	12.02	12.01	11.96	11.98	12.02	11.99
Machine 2:	12.02	11.97	11.96	12.01	11.99	12.03	12.04	12.02	12.01	12.00

Do you think that the quality control department is correct?

Student (W. S. Gosset) (1908). "The probable error of a mean." *Biometrika*, 6, 1-25.

?? In the study *"Interrelationships Between Stress, Dietary Intake, and Plasma Ascorbic Acid During Pregnancy"* conducted at the Virginia Polytechnic Institute and State University in May 1983, the plasma ascorbic acid levels of pregnant women were compared for smokers versus non-smokers. Thirty-two women in the last three months of pregnancy, free of major health disorders, and ranging in age from 15 to 32 years were selected for the study. Prior to the collection of 20 ml of blood, the participants were told to avoid breakfast, forego their vitamin supplements, and avoid foods high in ascorbic acid content. From the blood samples, the following plasma ascorbic acid values of each subject were determined in milligrams per 100 milliliters:

Plasma Ascorbic Acid Values		
Non-smokers		Smokers
0.97	1.16	0.48
0.72	0.86	0.71
1.00	0.85	0.98
0.81	0.58	0.68
0.62	0.57	1.18
1.32	0.64	1.36
1.24	0.98	0.78
0.99	1.09	1.64
0.90	0.92	
0.74	0.78	
0.88	1.24	
0.94	1.18	

Economic fuel. An industrial plant wants to determine which of two types of fuel (gas or electric) will produce more useful energy at the lower cost. One measure of economical energy production, called the *plant investment per quad,* is calculated by taking the amount of money (in dollars) invested in the particular utility by the plant and dividing by the delivered amount of energy (in quadrillion British thermal units). The smaller this ratio, the less an industrial plant pays for its delivered energy.

Random samples of 11 plants using electrical utilities and 16 plants using gas utilities were taken, and the plant investment per quad was calculated for each. The data produced the results shown in the table.

	Electric	Gas
Sample size	$n_1 = 11$	$n_2 = 16$
Mean Investment/Quad(Billions)	$\bar{x_1} = 44.5$	$\bar{x_2} = 34.5$
Sample Variance	$s_1^2 = 76.4$	$s_2^2 = 63.8$

Do the data provide sufficient evidence at the $\alpha = 0.05$ level to indicate a difference in the average investment per quad between the plants using gas and those using electrical utilities?

Fatigue. According to the article *"Practice and Fatigue Effects on the Programming of a Coincident Timing Response,"* published in the *Journal of Human Movement Studies* in 1976, practice under fatigued conditions distorts mechanisms which govern performance. An experiment was conducted using 15 college males who were trained to make a continuous horizontal right-to-left arm movement from a micro-switch to a barrier, knocking over the barrier coincident with the arrival of a clock sweephand to the 6 o'clock position. The absolute value of the difference between the time, in milliseconds, that it took to knock over the barrier and the time for the sweephand to reach the 6 o'clock position (500 msec) was recorded. Each participant performed the task five times under pre-fatigue and post-fatigue conditions, and the sums of the absolute differences for the five performances were recorded as follows:

Subject	Absolute Time differences (msec)	
	Pre-fatigue	Post-fatigue
1	158	91
2	92	59
3	65	215
4	98	226
5	33	223
6	89	91
7	148	92
8	58	177
9	142	134
10	117	116
11	74	153
12	66	219
13	109	143
14	57	164
15	85	100

An increase in the mean absolute time differences when the task is performed under post-fatigue conditions would support the claim that practice under fatigued conditions distorts mechanisms that govern performance. Assuming the populations to be normally distributed, test this claim.

New Mexico wells. The accompanying data are calcium carbonate ($CaCO_3$) readings (parts per million cubic centimeters) for ten wells in the Atrisco well field (one of the water sources for Albuquerque, New Mexico) for 1961 and 1966.

Well No.	YEAR	
	1961	1966
1	185	256
2	92	58
3	112	190
4	82	98
5	108	142
6	117	142
7	62	138
8	64	166
9	92	64
10	76	130

There was a concern that the $CaCO_3$ levels in the water supply were rising during that period. Is this concern substantiated by the data? Test at 10% significance level. You will find the following Splus calculations useful.

```
> y1961_c(185, 92, 112, 82, 108, 117,62, 64, 92, 76)
> y1966_c(256, 58, 190, 98, 142,142, 138, 166, 64, 130)
> diff_y1961-y1966
> diff
 [1]  -71   34  -78  -16  -34  -25  -76 -102   28  -54
> mean(diff)
[1] -39.4
> var(diff)
[1] 2074.933
> sqrt(var(diff))
[1] 45.55144
```

Solution: This is paired t-test. The alternative is $H_1 : \mu + 1 - \mu_2 = d < 0$. $t = \frac{-39.4}{45.55/\sqrt{10}} = -2.7353$.

$t_{9,.9} = 1.383$

Rejection Region is $(-\infty, -1.383)$. H_0 is rejected.

Methods of reading.

In a psychological experiment a random sample of 20 students is randomly divided into two groups: *phonetic group* and *memorization group* with 10 students in each group. At the end of instruction, we measure all 20 students' reading times on a standard passage. The data are shown in the table below.

Phonetic (X)	5.8	5.1	6.6	4.7	5.6	5.9	5.7	4.3	4.5	5.0
Memorization (Y)	5.9	6.1	5.1	4.7	4.6	6.4	6.7	5.1	5.0	4.6

[$\bar{X} = 5.32, \bar{Y} = 5.42, s_X = 0.72, s_Y = 0.78.$]
Test the hypothesis that the two types of instruction are different. Use $\alpha = 10\%$. Assume $\sigma_X = \sigma_Y$.
[Sol: $\bar{X} = 5.32, \bar{Y} = 5.42, s_X = 0.72, s_Y = 0.78, s_p = 0.753, t = -0.297, t_{18,0.05} = 1.734$, Do not reject H_0.]

Energy. Two relatively new energy-saving concepts in home building are solar-powered homes and earth-sheltered homes. An individual is drawing up plans for a new home and wants to compare expected annual heating costs for the two types of innovation. Independent random samples of solar powered homes (which receive 50% of their energy from the sun) and earth-sheltered homes yielded the accompanying summary data on annual heating costs.

Solar-powered	Earth-sheltered
$n_1 = 120$	$n_2 = 60$
$\bar{X} = \$285$	$\bar{Y} = \$280$
$s_X = \$35$	$s_Y = \$30$

Is there evidence ($\alpha = 5\%$) that the annual costs of heating earth-sheltered homes is significantly less than the annual costs of heating solar-powered homes. [Hint. You can use z cut-points. $\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}} = 5.02$.]

Milking Cows. A feeding test is conducted on a herd of 25 milking cows to compare two diets, one of de-watered alfalfa and the other of field-wilted alphalpha. A sample of 12 cows randomly selected from the herd are fed de-watered alfalfa; the remaining 13 cows are fed field- wilted alfalfa. From observations made over a three-week period, the average daily milk production is recorded for each cow.

Field-wilted alphalpha (X)	44, 44, 56, 46, 47, 38, 58, 53, 49, 35, 46, 30, 41
De-watered alphalpha (Y)	35, 47, 55, 29, 40, 39, 32, 41, 42, 57, 51, 39

Researchers are interested in comparing the mean daily milk yields per cow between two diets. As a matter of fact, researchers suspect that the field-wilted alphalpha diet gives significantly larger mean.

Assume $\alpha = 0.05$ and perform the appropriate test. State clearly your decision. Assume that measurements come from the normal populations with the same (but unknown) variances.
[You may find the following info useful: $\bar{X} = 45.15, \bar{Y} = 42.25, s_X = 8, s_Y = 8.74.$]
$s_p = 8.361252, s_p^2 = 69.91053, t = 0.8664011, t_{23,0.05} = 1.71.$

Left-handed grippers. Measurements of the left- and right-hand gripping strengths of 10 left-handed writers are recorded.

Person	1	2	3	4	5	6	7	8	9	10
Left hand (X)	140	90	125	130	95	121	85	97	131	110
Right hand (Y)	138	87	110	132	96	120	86	90	129	100

- Do the data provide strong evidence that people who write with left hand have a greater gripping strength in the left hand then they do in the right hand? Use $\alpha = 0.05$.
- Would you change your opinion on significance if α were 0.1?

[You may find the following info useful: $\bar{d} = \bar{X} - \bar{Y} = 3.6, s_d = 5.46.$]
$t = 1.978; t_{9,0.05} = 1.833; t_{9,0.1} = 1.383$ **Durham and Raleigh.** A local investigation is conducted to determine

the mean age of welfare recipients between cities Durham and Raleigh, NC. Random samples of 75 and 100 welfare recipients are selected from the cities and the following computations are made:

	Durham	Raleigh
Sample Size	75	100
Sample Mean	39	43
Sample Standard Deviation	6.8	7.5

Do the data provide strong evidence that the mean ages of welfare recipients are different in Durham and Raleigh? Test at $\alpha = 0.02$.
$t = \frac{\bar{X}-\bar{Y}}{\sqrt{s_X^2/n_1+s_Y^2/n_2}} = -3.684.$

Marijuana. Investigators have studied the effects of marijuana on human physiology. One common belief held by laypersons is that marijuana affects pupil size. Weil et al.[4] studied number of subjects. Each was administered a high dose of marijuana by smoking a potent marijuana cigarette. The subjects ware all males, 21 to 26 years of age, all of whom smoked tobacco cigarettes regularly but have never tried marijuana. In this study, pupil size was measured with a millimeter rule under constant illumination with eyes focused on an object at a constant distance. Pupil size was measured before and after smoking marijuana. The part of data are given below.

Individual	1	2	3	4	5	6
Before marijuana	6	5	3	3	5	3
After marijuana	6	7	9	5	9	9

1. Describe the hypotheses of interest for testing. (Hint. The alternative should be one sided)
2. What is the error of II kind in the terms of the problem?
3. Perform the test at 5% significance level.
4. You assumed data come from normal populations. Why then you can not use z cut-points.

Solution.

```
> b_c(6,5,3,3,5,3)
> a_c(6,7,9,5,9,9)
> Ttest(a-b, alt=">")

----------------------------------------------
                  t-test
 Testing H_0: mu= 0  v.s. H_1: mu > 0 .
----------------------------------------------
 :-( Reject H_0.
 p-value= 0.01 is smaller than alpha= 0.05 .
 t-statistic= 3.371 .
 The rejection region cut-point is (+/-) 2.015 .
```

IQ test pairing In a study, children were first given an IQ test. The two lowest-scoring children were randomly assigned, one to a "noun-first" task, the other to a "noun-last" task. The two next-lowest IQ children were similarly assigned, one to "noun-first" task, the other to a "noun-last" task, and so on until all children were assigned. The data (scores on a word-recall task) are shown here, listed in order from lowest to highest IQ score

Noun-first	12	21	12	16	20	39	26	29	30	35	38	34
Noun-last	10	12	23	14	16	8	16	22	32	13	32	35

1. Are these two samples (Noun-first, Noun-last) independent?
2. Test the hypothesis that the population mean difference is 0 assuming the two sided alternative. Take $\alpha = 10\%$. *The following info may be useful: the difference sample mean is 6.583 and the difference sample standard deviation is 11.041.*

Duke Wear Pricing Practices.[5] Ever since the Duke Blue Devils won back-to-back National Basketball championship, the demand for Duke sweatshirts has skyrocketed not only at Duke, but across the nation as well. However after three years of buying their swearshirts on campus, many students have found that their friends at other schools often purchase twice as many Duke shirts from department stores far from Duke. This has led many students to complain that they are being unfairly overcharged because Duke sweatshirts are apparently priced higher on campus than they are off campus and elsewhere in the United States.

[4] Weil, A. T., Zinberg, N. E., and Nelson, J. (1968). Clinical and psychological effects of marijuana in man. *Science, 1968, No 162, 1234-1242.*

[5] From STA110 student projects

One particularly disgruntled group of students in their STA 110 project wanted to test the hypothesis that higher retail prices are being charged for sweatshirts in Duke stores than are charged off campus. They obtain random samples of 72 retail sweatshirt sales on campus and 55 such retail sales from stores off campus over the same time period and for the same style of sweatshirts. The following data were obtained:

Duke Sales	Off-Campus Sales
$X_1 = \$49.35$	$X_2 = \$43.05$
$s_1 = \$6.70$	$s_2 = \$7.98$

(a) Do these data provide sufficient evidence to support the students' claim that the mean sales price of Duke sweatshirts is higher at Duke than it is off campus? State the null and the alternative hypothesis and perform the test at $\alpha = 0.05$.

(b) Since samples are large, you can use z approximation for the exact t test in (a). Calculate approximate p-value for the test in (a).

Stairs for Stats. For their STA110e project Gretchen and Montaye[6] decided to measure heights of individual stairs on West and East campuses and then compare the means. They hypothesized that there might be a difference in heights due to different styles of architecture, Gothic on West and Georgian on East.

Gothic architecture evolved during the 12th century in Europe, primarily France, and was popular there until the 15th century. *High Gothic* was perfected in the 13th century and it was named such for its higher ceilings, vaults and form. Gothic architecture has long been admired for its ornateness, high-reaching towers and spires; Gretchen and Montaye believed that Gothic steps on West were taller in height than those on Georgian East campus.

Georgian architecture was primarily in vogue during the 16th and 17th centuries; it is known for its rounded arches, red brick, simple lines and smooth, flowing form.

Campus	Data Source and Number of Stairs	Mean	St. Deviation
West	Allan 20, Perkins 25, West Union 15	17.53	2.74
East	Lilly 5, East Union 6, Baldwin 11, Brown 5, Alspaugh 5, Pegram 5, Giles 5, Wilson 5, Carr 3, Jarvis 2, East Duke 8	14.99	0.58

Without assuming equality of underlying (unknown) variances test the hypothesis that the mean heights of stairs are the same. Consider the one sided alternative. Take $\alpha = 0.05$.

- (i) State your decision.
- (ii) What assumption(s) you have made?
- (iii) Is the p-value smaller than 0.01? (Do not calculate p-value.)

[6] Gretchen Anderson and Montaye Sigmon: Stairs for Stats, Sta110E Project, Fall 1995.