1. [11 marks] The data below show the 1 year returns (as of September 30, 2006) for randomly selected Canadian and U.S. equity mutual funds. It has been suggested that Canadian equity funds have outperformed U.S. equity funds in the past year. The Minitab output gives some appropriate and inappropriate analyses of the data.

CdnEq1	3.46	5.84	-0.01	16.97	14.6	4.39	10.26	13.06	4.85	11.33	5.19	3.87
USEq2	13.04	0.45	4.96	6.09	4.68	-0.32	0.29	3.52	9.87	4.23	2.81	-1.44

(a) [5 marks] Test whether Canadian funds have outperformed U.S. funds. Use the .05 level of significance. Show how the test statistic is calculated.

-H0: mu1=mu2, H1: mu1 > mu2;
-pooled stdev is sqrt [(5.25^2 + 4.25^2)/2] = 4.77
-t-stat is (7.82-4.02)/sqrt[4.77^2 * (1/12 + 1/12)] = 1.95
-p-value is .032 (half of the two-sided p-value) < .05, or critical value is 1.717 (22df)
-decide to reject the null H and conclude Cdn funds have outperformed US funds

accept a test not assuming equal variance for full marks. Here the critical value is 1.721 (21 df) and the p-value is the same.
-deduct 2 marks for paired t-test here if everything else is correct.

(b) [3 marks] Explain how you selected the specific test above and explain whether the assumptions of the test are warranted.

-data are from 2 random samples as there is no evidence of matched data;
-since these are small samples, we must assume the data come from normally distributed populations (.5 mark) from the first two boxplots, there is a little bit of skewness but no outliers; therefore, it is reasonable to assume these data come from normal distributions (.5 mark)
-the sample variances are close in value; it is reasonable to assume equal variances, or accept statement that it is not necessary to assume equal variance.
If paired test in (a), look for comment about boxplot of distribution of differences showing a normal distribution for max 2 marks out of 3.

(c) [3 marks] Given that INGDirect offered a 3% rate of return on savings accounts during the past year, test whether Canadian equity funds have performed better than funds placed in INGDirect.

```
Test of mu = 3 vs > 3
                                                 95%
                                               Lower
Variable   N     Mean    StDev  SE Mean        Bound     T      P
CdnEq1    12  7.81750  5.24635  1.51449      5.09765  3.18  0.004
```

-one mark for hypotheses
-one mark for calculation of t-statistic and 1.796 (11 df),
-one mark for deciding to reject the null H and concluding that Cdn equity funds have performed better than 3%.

2. [6 marks] The data below compares the rates of return on equity last year and this year for a sample of Canadian small businesses. The Minitab output gives some output that may or may not be helpful.

RofR1	13.966	11.1395	8.3104	10.0634	18.3224	7.8564	10.783	5.3086	12.5726	1.1173	7.3576	6.3016	5.4648	15.8778	18.5251
RofR2	15.2405	9.6101	10.6785	11.5555	11.9431	6.2496	4.6547	6.5608	12.6578	1.3674	8.1469	7.2978	4.1749	15.7108	15.9875
Diff2	-1.27452	1.52944	-2.36808	-1.49208	6.3794	1.60678	6.12832	-1.25215	-0.08517	-0.25008	-0.7893	-0.99617	1.28993	0.16691	2.53763

(a) [4 marks] Test whether there is sufficient evidence to conclude that the rates of return are different between the two years. Show how the test statistic is computed.

- H0: mu1 = mu2 or average change = 0, H1: mu1 >< mu2 or average change >< 0
-t-statistic based on average difference is .742/.675 = 1.1
-p-value is 2 * .145 = **.29** not < .05 or critical value is **2.145** (14 df).
note that output is for a 1-sided paired t-test.
-decide not to reject null H, conclude rates of return have not changed.

-deduct 2 marks for performing 2-sample test, but otherwise full marks if everything else is correct: output shows that the t-statistic is .43 (critical values are 2.052 (27df), not assuming equal variance, or 2.048 (28 df), assuming equal variance), and p-value is .671.

(b) [2 marks] Explain whether the assumptions of the test above are warranted and justify your selection of the test completed.

-data are paired since we have the same sample of firms for both years
-assume the differences must come from a normal distribution (.5 mark) ; the third boxplot (of differences) has two outliers, suggesting that the assumption above is not warranted. (.5 mark)

If the 2-sample test is completed in (a), then deduct 1 mark for thinking the data are independent.
Then give .5 mark for assuming each sample comes from a normal distribution and .5 mark for saying the first two boxplots indicate the data come from normal distributions since they are relatively symmetric and have no outliers.

3. Some psychologists believe that people think better when they are properly rested and that performance on an exam depends on the time of day. To investigate this, some data have been collected on students' performance on exams scheduled in the morning (start time before noon), in the afternoon (start time between noon and 4pm), and in the evening (start time after 4pm). The summary of these data (number of people at each performance level and time) is presented below.

Grade	Morning	Afternoon	Evening	Total
A	50	65	54	169
B	69	87	74	230
C	33	59	48	140
D	25	31	31	87
F	14	22	18	54
Total	191	264	225	680

a) [2 marks] Based on this data, find the 95% confidence interval for the overall failure rate.

54/680±1.96*sqrt((54/680*(1-54/680)/680))=0.0794±0.0203=(0.0591,0.0997)

1 point for formula, 1 point for answer

b) [2 marks] If you were to estimate the overall failure rate with a 95% confidence level and a 1% margin of error, how large a sample size would be required?

n=1.96^2*0.5*0.5/0.01^2=9604 or use pilot sample
1.96^2*(54/680)*(1-54/680)/0.01^2=2808 or 2809

1 point for formula, 1 point for answer (any approach is OK)

c) [4 marks] Do you think the failure rate is different for morning and afternoon exams? Test your hypothesis at 5% significance level.

H0: $\pi_1=\pi_2$ HA: $\pi_1\neq\pi_2$ (1 Point)

p_1=14/191=0.0733 p_2=22/264=0.0833
p_{pooled}=(14+22)/(191+264)=0.0791 (1 point for the use of p_{pooled})

z=(0.0733-0.0833)/sqrt(0.0791*(1-0.0791)*(1/191+1/264))=-0.39 (1 point)

Since |-0.39|<1.96, do not reject H0 and may assume that the failing rate is the same for morning and afternoon exams (1 point)

d) [3 marks] Of 20 students who took the course for the second time, 2 failed again. Can you be 95% sure that the probability of failing the same course for the second time is less than 20%?

H0: $\pi\geq0.2$ HA: $\pi<0.2$ (0.5 marks)
We have small sample size, therefore, we have to use Binomial distribution
(Give no credits for the remaining parts for students who decided to do normal approx.)
p-value = Prob(X≤2) (1 point)
Prob(X≤2)=P(0)+P(1)+P(2)
=0.8^20+20*0.2*0.8^19+19*20/2*0.2^2*0.8^18=0.2061 (1 point)
Since 0.1545>0.05, we cannot reject H0 and cannot claim that the probability to fail the same course for the second time is less than 20% (0.5 marks)

```
Binomial with n = 20 and p = 0.2

x  P( X <= x )
2     0.206085
```

e) [3 marks] Suppose you want to determine if the probability of getting a C is the same for all three times. State the appropriate test and hypotheses, rearrange the data into a form required to perform this test, and briefly explain how you would set up the next set of calculations. Please do not carry out any actual computations.

Use Contingency analysis test of hypothesis that "getting C" is independent of the time of the exam (1 point)

Need to rearrange data:

Grade	Morning	Afternoon	Evening	Total
C	33	59	48	140
Not C	158	205	177	540
Total	191	264	225	680

(1 point)

After that you need to find expected frequencies for each of the cells based on the assumption that rows and columns are independent (i.e., use formula e_{ij}=(# in row i)*(# in column j)/(# total) and compute chi-square score. (1 point)

f) [4 marks] Test whether there is or is not an equal distribution of As, Bs, and Cs among students who got at least a C on the final exam. Start by stating the appropriate test and use the .05 level of significance.

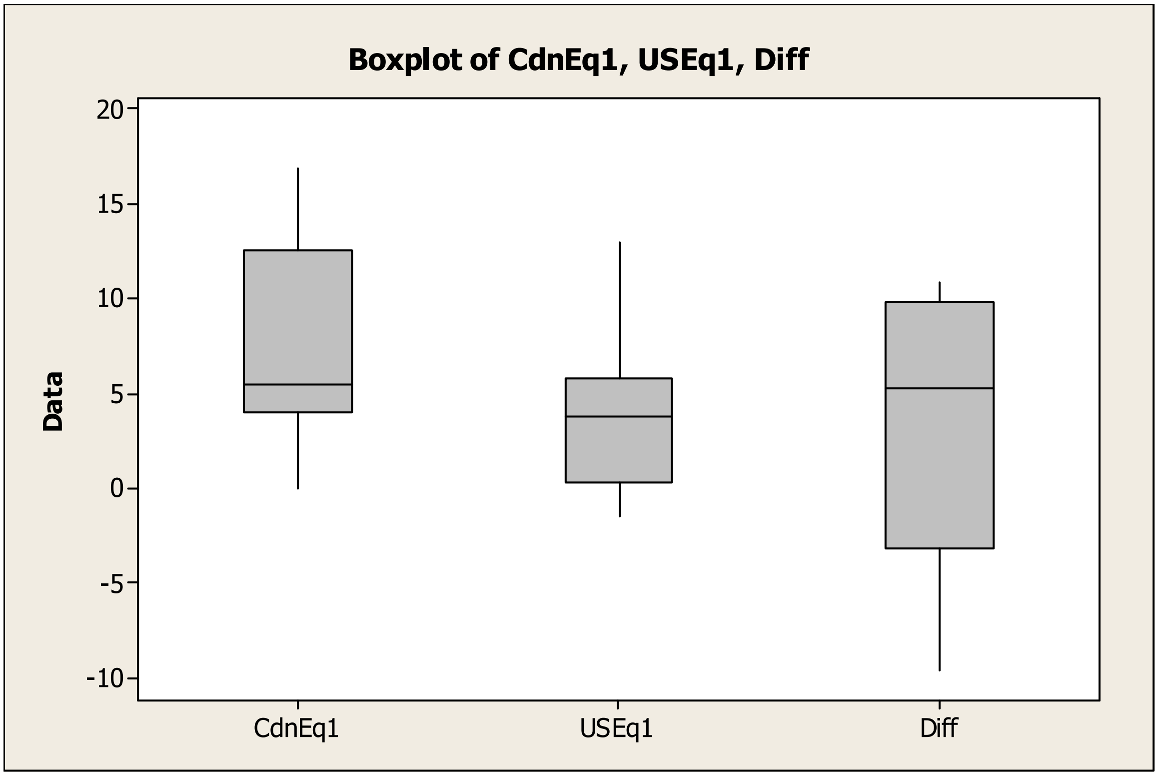
Goodness of fit test (.5 mark)
H0: $\pi_A=\pi_B=\pi_C=1/3$ (0.5 point)

	Observed	probability	expected
A	169	1/3	179.67
B	230	1/3	179.67
C	140	1/3	179.67
Total	539		

(1 marks: 0.5 marks for observed, 0.5 marks for expected)

Chi-sq=(169-179.67)^2/179.67+(230-179.67)^2/179.67+(140-179.67)^2/179.67=23.49
(0.5 point)
d.f.=3-1=2 (0.5 marks)
Chi-sq(5%, 2d.f.)=5.99 (0.5 marks)
Since 23.39>5.99, we reject H0, so, we can conclude that students that got at least C on the final exam do not have approximately the same chance to get A B or C (0.5 marks)

1. CANADIAN AND U.S. FUND RETURNS




```
         N  Mean  StDev  SE Mean
CdnEq1  12  7.82   5.25      1.5
USEq1   12  4.02   4.25      1.2
```

Two-Sample T-Test and CI: CdnEq1, USEq1

```
Difference = mu (CdnEq1) - mu (USEq1)
Estimate for difference:  3.80250
95% CI for difference:  ( ______ , ______)
T-Test of difference = 0 (vs not =): T-Value = ______  P-Value = ______  DF = 21
```

Two-Sample T-Test and CI: CdnEq1, USEq1

```
Difference = mu (CdnEq1) - mu (USEq1)
Estimate for difference:  3.80250
95% CI for difference:  ( ______ , ______ )
T-Test of difference = 0 (vs not =): T-Value = _______  P-Value = ______  DF = 22
Both use Pooled StDev = ______
```

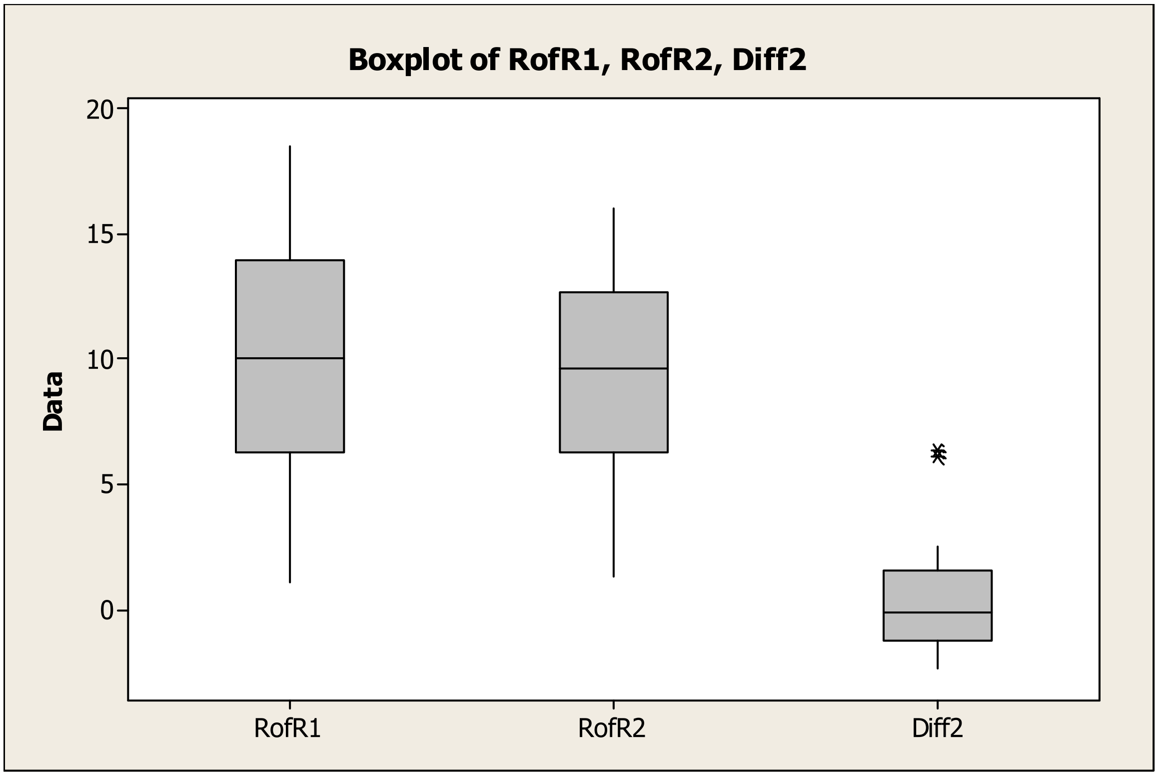
Paired T-Test and CI: CdnEq1, USEq1

```
Paired T for CdnEq1 - USEq1

             N     Mean    StDev  SE Mean
CdnEq1      12  7.81750  5.24635  1.51449
USEq1       12  4.01500  4.24532  1.22552
Difference  12  3.80250  6.81489  1.96729

95% lower bound for mean difference: _______
T-Test of mean difference = 0 (vs > 0): T-Value = ________  P-Value = ______
```

2. RETURN ON EQUITY



Two-Sample T-Test and CI: RofR1, RofR2

```
Two-sample T for RofR1 vs RofR2

        N   Mean  StDev  SE Mean
RofR1  15  10.20   5.00      1.3
RofR2  15   9.46   4.46      1.2

Difference = mu (RofR1) - mu (RofR2)
Estimate for difference:  0.742057
95% CI for difference:  ( _____ , _____)
T-Test of difference = 0 (vs not =): T-Value = ____  P-Value = _____  DF = 27
```

Paired T-Test and CI: RofR1, RofR2

```
Paired T for RofR1 - RofR2

             N      Mean     StDev   SE Mean
RofR1       15   10.1978    4.9983    1.2905
RofR2       15    9.4557    4.4642    1.1527
Difference  15  0.742057  2.615299  0.675267

95% lower bound for mean difference: _______
T-Test of mean difference = 0 (vs > 0): T-Value = ____  P-Value = _____
```