STA305/1004 L0101 & L0201- Midterm Term Review

February 24-25, 2020

Aids allowed: One 8.5'x11' sheet with writing on both sides, and a non-programmable calculator.

Test Instructions: Answer all five questions in the space provided within 90 minutes. Work written on the back of pages will not be graded.

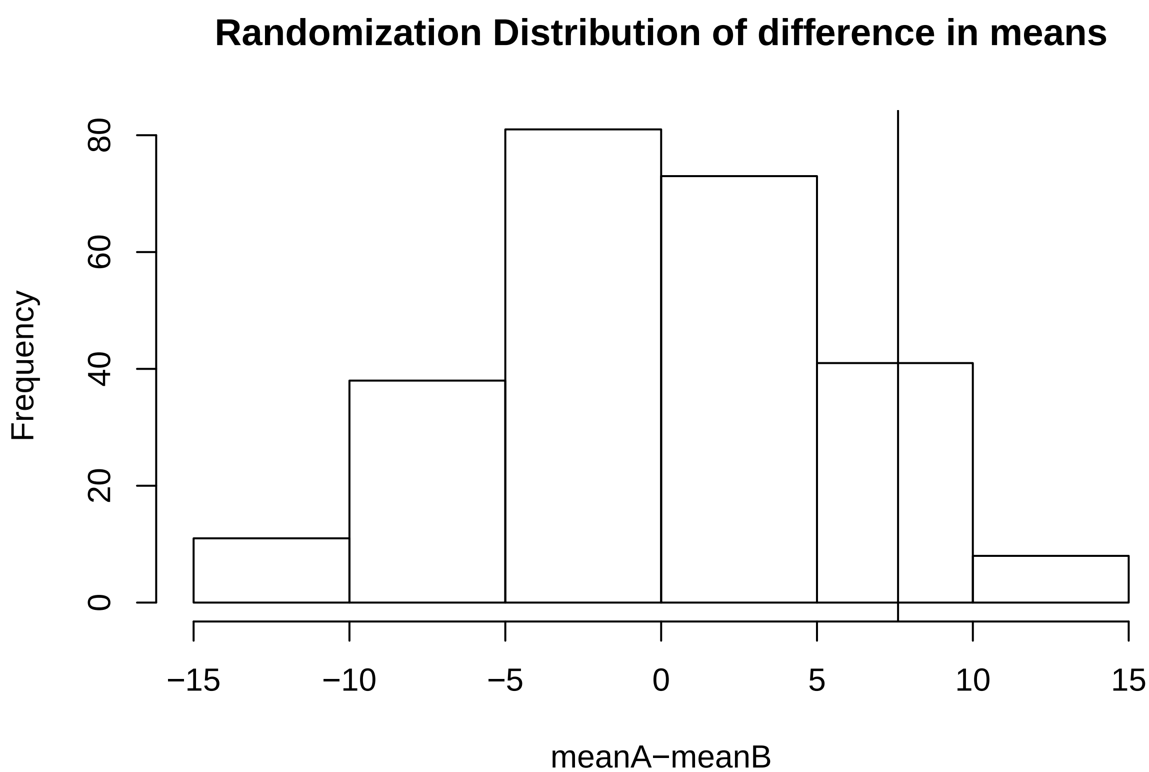
Week	Topic
Week 1	Background topics and distribution theory
Week 2	Comparing two treatments: Completely Randomized Design
Week 3	Comparing two treatments: Randomized Paired Design
Week 4	Power, Sample size calculations and related factors
Week 5a	Causal Inference
Week 5b, 6a	Estimating and Using the Propensity Score
Week 6b	Introduction to One-way Analysis of Variance
Total Marks	70

1. (30 marks) (Adapted from Box, Hunter and Hunter) Ten hens were randomly allocated to two different diets, A or B. Five were assigned to Diet A and five to Diet B. After one year the number of eggs produced are given in the table below. The hens are identified by the number in parentheses.

Diet A	166 (1)	174 (2)	150 (3)	166 (4)	165 (5)
Diet B	158 (6)	159 (7)	142 (8)	163 (9)	161 (10)

The randomization distribution and related output for testing $H_0 : \mu_A = \mu_b$ versus $H_1 : \mu_B > \mu_A$, where μ_A, μ_B are the mean number of eggs for diets A and B respectively, were calculated. Some of the R code is shown below.

```r
yA <- c(166,174,150,166,165)
yB <- c(158,159,142,163,161)
eggs <- c(yA,yB) #pool data
for (i in 1:N)
{
  res[i] <- mean(eggs[index[,i]])-mean(eggs[-index[,i]])
}
hist(res,xlab="meanA-meanB", main="Randomization Distribution of difference in means")
observed <- mean(yA)-mean(yB) #store observed mean difference
abline(v=observed) #add line at observed mean diff
```




```r
observed # Observed difference in means
```

```
[1] 7.6
```

```r
sum(res>=observed)
```

```
[1] 30
```

Answer the following questions based on this study.

(a) (5 marks) Is this study an experiment or observational study? Briefly explain your reasoning.

(b) (5 marks) Give two examples of possible treatment assignment for this study under the null hypothesis that are not the observed treatment assignment. Use the table below to fill in the treatment assigned to each hen. Use A for diet A and B for diet B.

Hen	Treatment assignment #1	Treatment assignment #2
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

(c) (5 marks) What is the propensity score in this study? What is the probability of treatment assignment? Is the treatment assignment ignorable? Briefly explain.

(d) (5 marks) What is the p-value of the randomization test? Is there evidence at the 5% significance level that diet A is better than diet B? Briefly explain your reasoning.

(e) (10 marks) Suppose another investigator would used a different set of 10 hens to compare the 2 diets, but would like to use a randomized paired design instead of the design described above. In one or two sentences describe how this study could have been designed as a randomized paired design. What are the treatments and experimental units? How would you randomize the treatments to the units? What is the propensity score and probability of treatment assignment in your paired design?

2. (20 marks) (Adapted from Box, Hunter and Hunter) Paint used for marking lanes on highways must be very durable. Double yellow lines are used to mark the centre of a road to separate lanes of traffic travelling in opposite directions. In one trial, yellow paint from two suppliers, labelled A and B, was randomly assigned to the two centre lines on six different highway sites, denoted 1, 2, 3, 4, 5, 6. After a considerable length of time, the average wear for the samples at the six sites were as follows:

Site	Paint A	Paint B
1	59	69
2	65	83
3	64	74
4	52	61
5	71	78
6	64	69

```r
paintA <- c(59,65,64,52,71,64)
paintB <- c(69,83,74,61,78,69)
```

```r
#Two-sample t-test; equal variance
t.test (paintA, paintB, paired=FALSE, var.equal=TRUE)
```

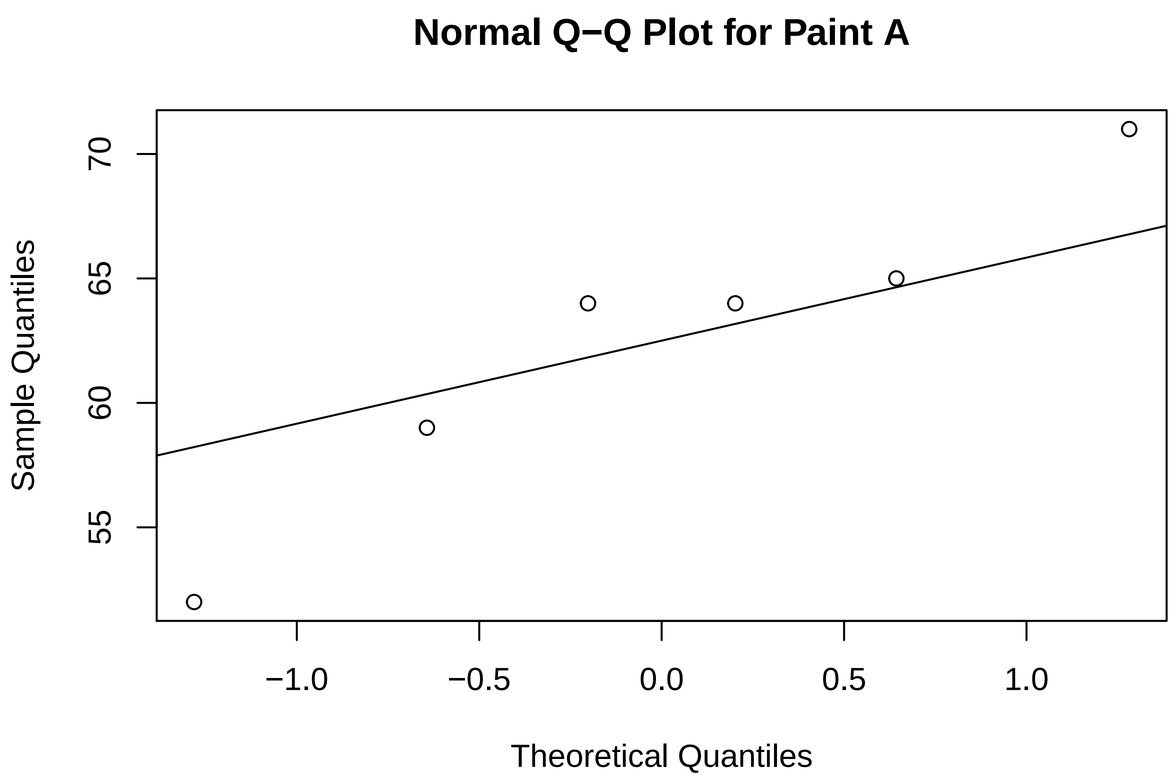
```
##
##  Two Sample t-test
##
## data:  paintA and paintB
## t = -2.3971, df = 10, p-value = 0.0375
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.9735320  -0.6931347
## sample estimates:
## mean of x mean of y
##  62.50000  72.33333
```

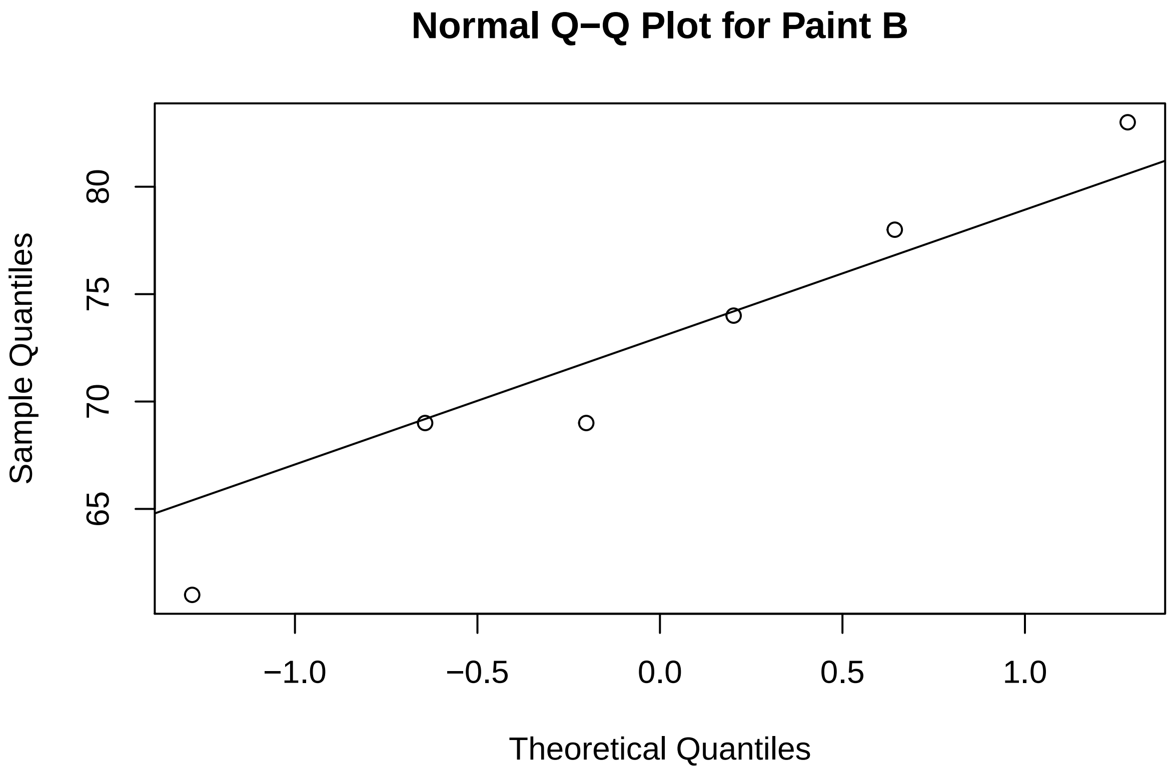
```r
#Two-sample t-test; unequal variance
t.test (paintA, paintB, paired=FALSE, var.equal=FALSE)
```

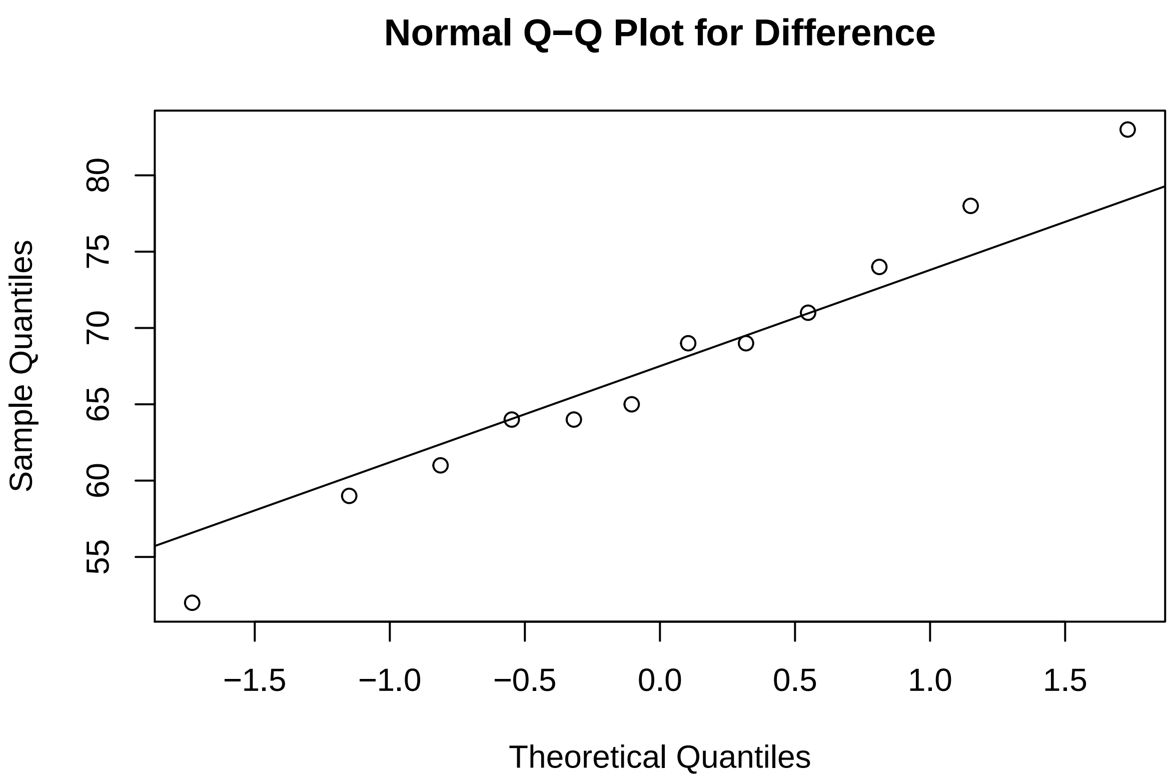
```
##
##  Welch Two Sample t-test
##
## data:  paintA and paintB
## t = -2.3971, df = 9.6661, p-value = 0.0383
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -19.0165158  -0.6501509
## sample estimates:
## mean of x mean of y
##  62.50000  72.33333
```

```r
#Paired t-test; equal variance
t.test (paintA, paintB, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  paintA and paintB
## t = -5.4176, df = 5, p-value = 0.002901
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.499095  -5.167572
## sample estimates:
## mean of the differences
##               -9.833333
```




Normal Q−Q Plot for Paint B
Sample Quantiles
65
70
75
80
−1.0
−0.5
0.0
0.5
1.0
Theoretical Quantiles


Normal Q−Q Plot for Difference
Sample Quantiles
55
60
65
70
75
80
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Theoretical Quantiles

(a) (5 marks) What is the name of the study design? Justify your answer.
(b) (5 marks) Is this study an experiment or an observational study? Justify your answer.
(c) (5 marks) Is there a statistically significant difference between the mean wear of A and B at the 1% significance level? Justify your answer.
(d) (5 marks) Would it have been more appropriate to conduct a randomization test for this data? Justify your answer.

3. (25 marks) A psychologist studying body language conducted an experiment on 20 subjects, and obtained a significant result from a two-sided z-test ($H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$). Let's call this experiment #1. The observed value of the z statistic from your experiment is $z = 2.2$ so the p-value=0.028. In order to confirm the results the psychologist is planning to run the same experiment on an additional 10 subjects (i.e., the same experiment will be done on 10 different subjects). Let's call this experiment #2.

   The following percentiles from the $N(0, 1)$ distribution might be required to carry out some of the calculations in the questions below.

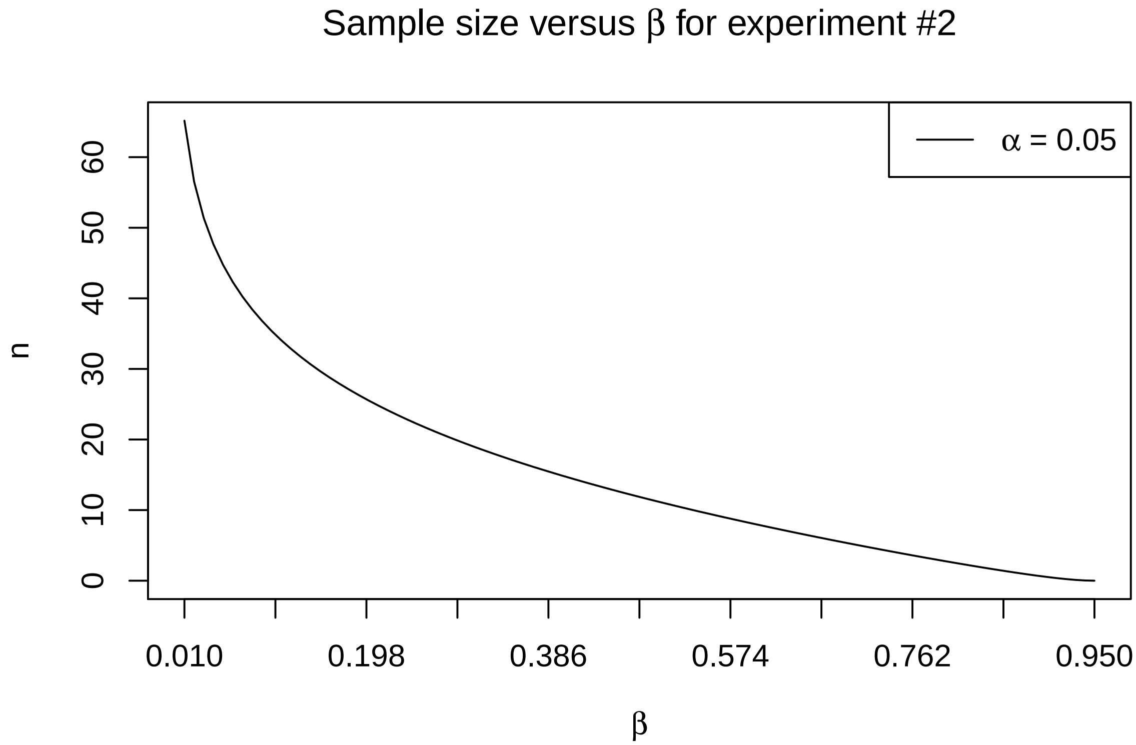
α	z_α
0.468	0.08
0.100	1.28
0.050	1.64
0.025	1.96
0.020	2.05
0.010	2.33

   z_α is the $100(1 - \alpha)^{th}$ percentile of the $N(0, 1)$. For example, the 90^{th} percentile is $z_{0.10} = 1.28$.

   (a) (5 marks) Assume that the true mean in experiment #2 is the sample mean obtained in the experiment #1. What is the probability that the results of experiment #2 will be significant at the 5% level by a one-tailed z-test ($H_1 : \mu > 0$)? Provide a brief interpretation of this probability.

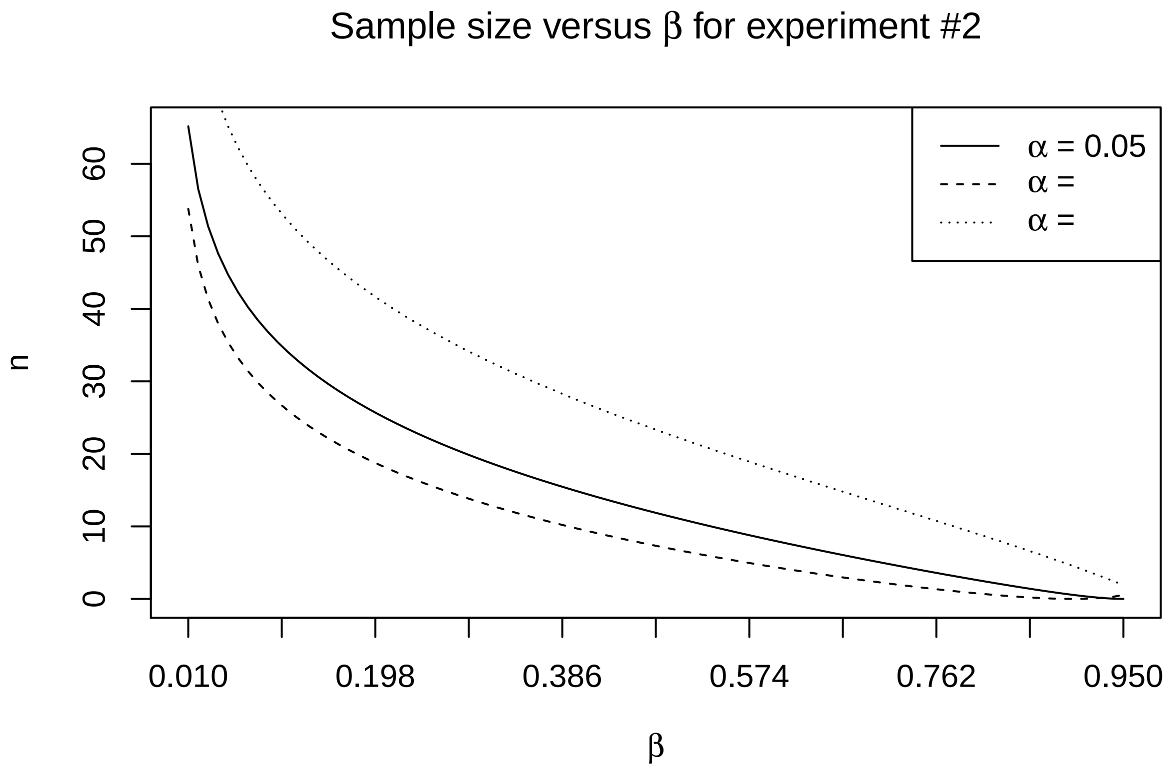
   (b) (10 marks) The psychologist strongly believes in her theory, and would like a sample size formula for experiment #2 so she can calculate the sample size given α - type I error rate and β - type II error rate. Derive such a sample size formula for the psychologist as a function of α, β and $\Phi(\cdot)$ the cumulative distribution function of the $N(0, 1)$.

(c) (5 marks) The power function derived in part (b) was used to create a plot of sample size n versus β, the probability of a type II error.



What does the plot tell you about the relationship between sample size and power for experiment #2? Use the plot to estimate how many subjects the psychologist would have to enrol so that experiment #2 will have 80% power at the 5% significance level. Should she revise her original design and enrol more than 10 subjects in experiment #2, if she want to be more confident of rejecting H_0 when in fact $\mu > 0$?

(d) (5 marks) Suppose the psychologist decided to change the significance level in experiment #2 from 5% to 10%. The plot of n versus β is shown for three values of α, the type I error rate, but the statistician that created the graph forgot to label two of the three curves in the plot. Should she use the curve above or below the curve where $\alpha = 0.05$ to estimate the sample size? Estimate the sample size required for experiment #2 to have 80% power at the 10% significance level. Briefly explain.



4. (25 marks) What is the effect of smoking on weight gain? Data was used from The National Health and Nutrition Examination Survey Data I Epidemiological Follow-up Study (NHESFS) survey to assess this question. The NHESFS has information on sex, race, weight, height, education, alcohol use, and intensity of smoking at baseline (1971-75) and follow-up (1982) visits. The survey was designed to investigate the relationships between clinical, nutritional, and behavioural factors. A cohort of persons 25-74 who completed a medical exam in 1971-75 followed by a series of follow-up studies.

   The table below shows the distribution of the baseline covariates in the two groups. 403 people are in the smoking group (No cessation T=0) and 1163 people are in the smoking cessation group (T=1).

	Cessation (T=1)	No cessation (T=0)
age, years (mean)	46.2	42.8
men, %	54.6	46.6
white, %	91.1	85.4
university, %	15.4	9.9
weight, kg (mean)	72.4	70.3
Cigarettes/day (mean)	18.6	21.2
year smoking (mean)	26.0	24.1
little/no exercise, %	40.7	37.9
inactive daily life, %	11.2	8.9

   The propensity score was estimated using a logistic regression based on all 9 covariates; no interactions were included in the propensity score model. Three propensity score methods were used to estimate the treatment effect (the average difference in weight gain between the smoking cessation group and the smoking group): propensity score matching, stratifying on the propensity score, regression adjustment using the propensity score, and no adjustment using the two-sample t-test. The propensity score methods successfully balanced all the covariates (i.e., the absolute standardized differences are less than 10%). The table below shows the average treatment effect for each method with the 95% confidence interval. The propensity score matching method was able to match the 403 subjects in the smoking group to 403 subjects in the smoking cessation group.

Method	Average Treatment Effect	95% Confidence Interval
Matched	2.93	1.8 - 4.0
Stratified	3.26	1.7 - 3.4
Regression	3.40	2.5 - 4.3
Unadjusted	2.54	1.7 - 3.4

   Answer the following questions. (5 marks each)

   (a) Is this an experiment or observational study? What is the treatment in this study? Briefly explain.

   (b) Is the propensity score known or unknown in this study? Briefly describe how the propensity score was calculated for each subject. Give a formula showing how the propensity score is calculated and explain the terms in the formula.

   (c) Briefly compare and contrast the methods of propensity score matching and stratification by propensity score.

   (d) In the context of this study briefly explain what it means for the treatment assignment to be strongly ignorable. Does this seem like a plausible assumption for this study?

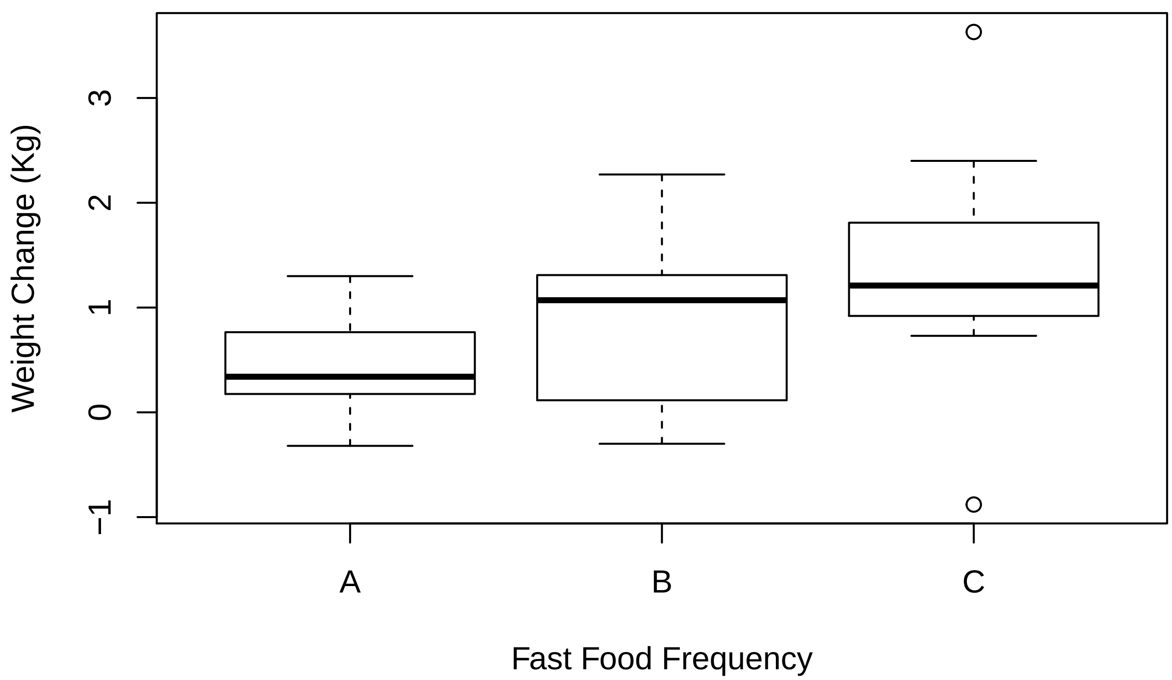
   (e) For this study, is the treatment assignment unconfounded? Briefly explain.

5. (30 marks) A study at UofT recruited twenty-one students to complete a thirty minute survey on their diet and eating habits at the end of an academic year. Students were paid $10 to complete the survey and answer a few questions. The data below shows their weight gain from September to April classified by the frequency that students ate fast food. In group A students reported eating fast food once per month; the students in group B reported eating fast food twice per month; and the students in group C reported eating fast food four times per month.

	A	B	C
	1.02	1.44	0.73
	-0.32	0.40	1.11
	0.27	-0.30	3.63
	0.08	2.27	-0.88
	0.51	-0.17	1.21
	0.34	1.07	1.22
	1.30	1.18	2.40
Treatment Average	0.46	0.84	1.35
Treatment SD	0.55	0.92	1.40

The researchers analyzed the data using R.

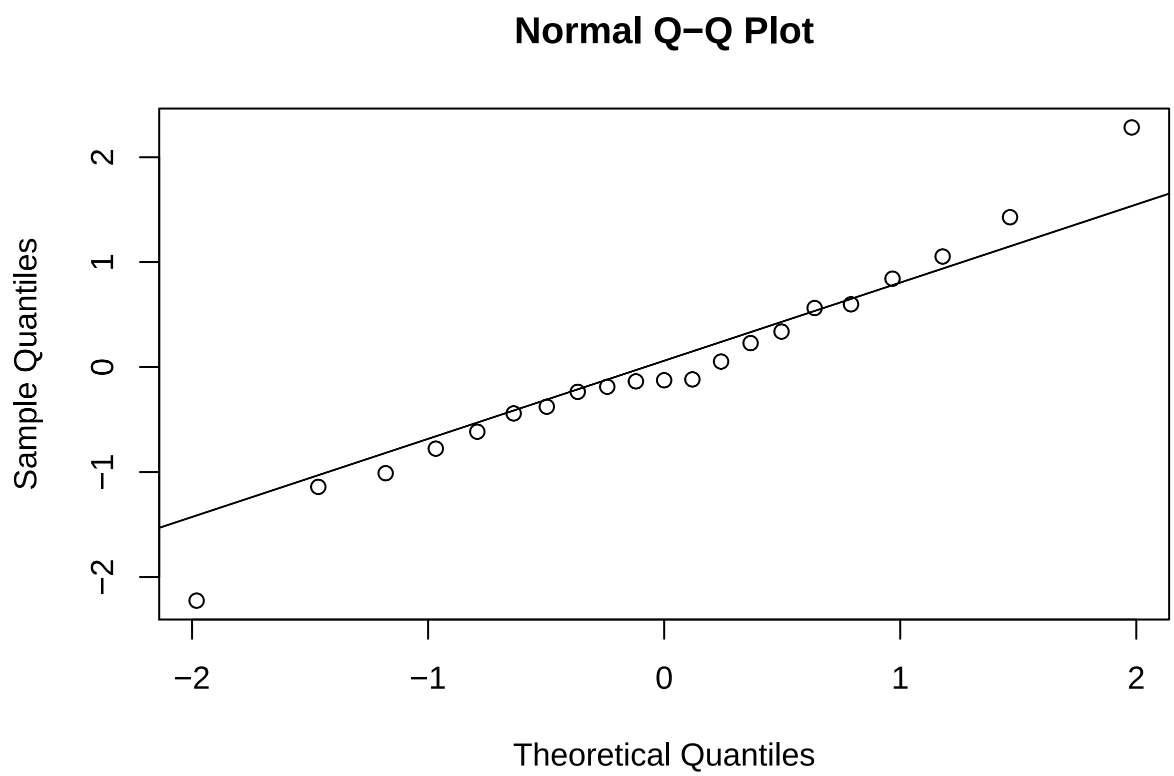
```
surveydat <- read.csv("surveydat.csv")
boxplot(wtchange~grp,data = surveydat,ylab="Weight Change (Kg)", xlab="Fast Food Frequency")
```



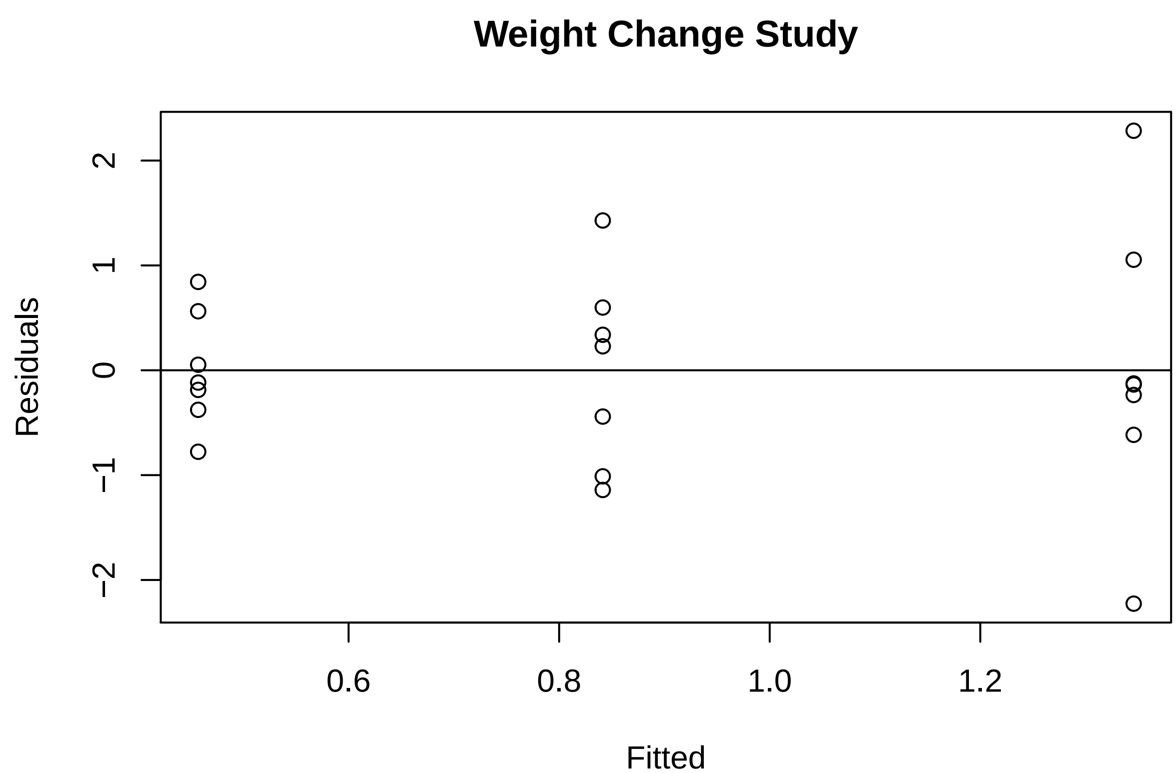

```
aovsurvey <- aov(wtchange~grp,data=surveydat)
summary(aovsurvey)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## grp           2   2.78   1.390   1.341  0.287
## Residuals    18  18.66   1.037
```

```
qqnorm(aovsurvey$residuals);qqline(aovsurvey$residuals)
```




```r
plot(aovsurvey$fitted.values, aovsurvey$residuals,ylab="Residuals",
     xlab="Fitted",main="Weight Change Study")
abline(h=0)
```



(a) (5 marks) Is this study an experiment or observational study? What are the treatments?

(b) (5 marks) Would it have been feasible for the researcher to randomized students to the treatments? What randomization scheme (assigning the subjects to the treatments) could the researcher use to accomplish the randomization?

(c) (5 marks) What are the null and alternative hypotheses that the researchers are testing in the data analysis? Is there evidence to reject the null hypothesis? Explain.

(d) (10 marks) What are the statistical assumptions behind the data analysis? Which tools can be used to check the assumptions? Are the assumptions satisfied? Explain.

(e) (5 marks) The researcher is convinced that the results of the study would have provided strong evidence that eating fast food four times per month causes students to gain weight, if the sample size in each group was larger. Is this a valid statement? Explain.