STA305/1004 - L0101 Midterm Test

March 1, 2017, 11:10-12:40

Name:

Student Number:

Instructions: Answer all four questions in 90 minutes.

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

The table below shows the value of each question.

Question	Marks
1	26
2	25
3	25
4	20
Total	96

This test contains 23 pages including this page.

Continue on next page ...

1. Five subjects with advanced colon cancer are enrolled in a two-arm randomized clinical trial. It was decided that three subjects would be randomized to Arm A (standard treatment), and two subjects to Arm B (new treatment). Subjects were assigned to the treatments by using five playing cards: three black and two red. The cards were shuffled ten times then each subject was assigned a card; if the card selected was black then the subject was assigned to arm B; if the card selected was red then the subject was assigned to arm A.

   Subjects' tumour size (cm) was measured before treatment then one month after treatment. One drug will be considered more effective than the other drug if average tumour growth is slower.

   The table below shows the percent change in tumour size for each subject after one month. For example, the first subject's tumour size grew by 1.6%. The investigators are interested to see if the average change in tumour size for the new treatment is different compared to the standard treatment.

Subject	Change in Tumour Size	Arm
1	1.6	B
2	2.7	A
3	2.8	A
4	3.1	A
5	1.5	B

Analysis #1

```
yA <- c(2.7,2.8,3.1)
yB <- c(1.6,1.5,0)
diff <- yA-yB
diff
```

```
[1] 1.1 1.3 3.1
```

```
N <- 32
res <- numeric(32)
LR <- list(c(-1,1))
trtassign <- expand.grid(rep(LR, 5))

for(i in 1:N){
res[i] <- mean(as.numeric(trtassign[i,]*diff))
}

observed <- mean(yA)-mean(yB)
observed
```

```
[1] 1.833333
```

```
tbar <- mean(res)
tbar
```

```
[1] 0
```

```
round(res,2)
```

```
 [1] -1.58 -1.14 -1.06 -0.62 -0.34  0.10  0.18  0.62 -1.14 -0.70 -0.62
[12] -0.18  0.10  0.54  0.62  1.06 -1.06 -0.62 -0.54 -0.10  0.18  0.62
```

Continue on next page ...

```
[23]  0.70  1.14 -0.62 -0.18 -0.10  0.34  0.62  1.06  1.14  1.58
```

```
round(mean(res),2)
```

```
[1] 0
```

```
round(res-tbar,2)
```

```
 [1] -1.58 -1.14 -1.06 -0.62 -0.34  0.10  0.18  0.62 -1.14 -0.70 -0.62
[12] -0.18  0.10  0.54  0.62  1.06 -1.06 -0.62 -0.54 -0.10  0.18  0.62
[23]  0.70  1.14 -0.62 -0.18 -0.10  0.34  0.62  1.06  1.14  1.58
```

```
round(observed-tbar,2)
```

```
[1] 1.83
```

Analysis #2

```
yA <- c(2.7,2.8,3.1)
yB <- c(1.6,1.5)
round(mean(yA)-mean(yB),2)
```

```
[1] 1.32
```

```
y <- c(2.7,2.8,3.1,1.6,1.5)
index <-combn(1:5,3)
res <- numeric(10)
for(i in 1:10)
{
  res[i] <- mean(y[index[,i]])-mean(y[-index[,i]])
}
observed <- mean(yA)-mean(yB)
tbar <- mean(res)
tbar
```

```
[1] -2.218278e-17
```

```
round(res,2)
```

```
 [1]  1.32  0.07 -0.02  0.32  0.23 -1.02  0.40  0.32 -0.93 -0.68
```

```
round(mean(res),2)
```

```
[1] 0
```

```
round(res-tbar,2)
```

```
 [1]  1.32  0.07 -0.02  0.32  0.23 -1.02  0.40  0.32 -0.93 -0.68
```

```
round(observed-tbar,2)
```

```
[1] 1.32
```

Continue on next page ...

Answer the following questions. (26 marks)

(a) State the null and alternative hypotheses for this study. Clearly define the statistical parameters in terms of the experimental context. (5 marks)

(b) What is the probability that a subject will be assigned to standard treatment? What is the probability of a treatment assignment chosen at random? Was the treatment assignment selected at random? Explain. (6 marks)

Continue on next page ...

(c) Describe how the randomization distribution for the difference in means is calculated. What has been assumed in calculating this distribution? Write down all the values of this randomization distribution. (HINT: Look at the analyses done using R) (5 marks)

(d) Calculate the appropriate randomization P-value. Is there a significant difference in effectiveness between the new and standard drugs at the 1% significance level? Explain your reasoning. (HINT: Look at the analyses done using R) (5 marks)

Continue on next page ...

(e) A confused statistician calculated two t-tests: one is appropriate, and the other is inappropriate (see the R output below). If you use the results from the appropriate t-test do you reach the same conclusion as you did using the randomization test (assume the 1% significance level)? What has to be assumed in order for the t-test to be valid. (5 marks)

T-test #1

```
yA <- c(2.7,2.8,3.1)
yB <- c(1.6,1.5)
t.test(yA,yB)
```

```
    Welch Two Sample t-test

data:  yA and yB
t = 10.115, df = 2.5967, p-value = 0.003726
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.8635114 1.7698219
sample estimates:
mean of x mean of y
 2.866667  1.550000
```

T-test #2

```
yA <- c(2.7,2.8,3.1)
yB <- c(1.6,1.5,0)
t.test(yA,yB,paired = TRUE)
```

```
    Paired t-test

data:  yA and yB
t = 2.8828, df = 2, p-value = 0.1022
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9029794  4.5696461
sample estimates:
mean of the differences
               1.833333
```

Continue on next page ...

2. Answer the following questions. (25 marks)

   (a) Two drugs A and B are to be tested on subjects' eyes. The drugs will be randomly assigned to each subject's eyes based on the flip of a fair coin. If the coin toss is heads then a subject will receive drug A in the left eye and drug B in the right eye; if the coin toss is tails then the subject will receive drug A in the right eye and drug B in the left eye. The outcome of interest is intraoccular pressure (IOP). The smallest clinically meaningful difference in IOP between the two drugs is 1.5mmHg. The standard deviation of IOP is 2.5mmHg for each drug, and the standard deviation for the differences of IOP between A and B within each person is 1.5.

   Circle the correct R output below that will allow you to determine how many subjects the researchers should enroll so that the study has 80% power to detect this difference at the 5% significance level (5 marks)?

Power calculation #1

```
power.t.test(n = NULL,power=0.8,delta = 1,sd = 2.5,
             sig.level = 0.05,type = "paired",
             alternative = "two.sided")
```

```
     Paired t test power calculation

              n = 51.00957
          delta = 1
             sd = 2.5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

Power calculation #2

```
power.t.test(n = NULL,power=0.8,delta = 1,sd = 1.5,
             sig.level = 0.05,type = "paired",
             alternative = "two.sided")
```

```
     Paired t test power calculation

              n = 19.66697
          delta = 1
             sd = 1.5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

Continue on next page ...

(b) A clinical trial where an experimental drug is to be compared with the standard treatment for a continuous biomarker measurement is being planned. The biomarker measurements in each group are independent and normally distributed with different means μ_1, μ_2, and common variance σ^2. The power function of the test $H_0 : \theta = 0$ versus $H_1 : \theta > 0$, where $\theta = \mu_1 - \mu_2$ at level α is:

$$1 - \beta = 1 - \Phi\left(z_\alpha - \frac{\theta}{\sigma\sqrt{1/n_1 + 1/n_2}}\right), \ \theta > 0,$$

where, $\Phi(\cdot)$ is the standard normal CDF, and z_α is the $100(1-\alpha)^{th}$ percentile of the standard normal distribution.

Show that if the numbers of patients in the two groups are the same with $n_1 = n_2 = \frac{n}{2}$ then the total sample size required so that the trial has power $1 - \beta$ is:

$$n = \frac{4\sigma^2 \left(z_\alpha + z_\beta\right)^2}{\theta^2}.$$

(5 marks)

Continue on next page ...

(c) Consider again the situation in part (b) on the previous page. The smallest clinically meaningful treatment difference is $\theta = 1$ and the variance is $\sigma^2 = 9$. If we conduct a one-sided Z-test at a significance level of $\alpha = 0.05$, what is the power if the researchers enroll 80 subjects in each arm? The clinical trial statistician calculated the power using R (the value for power is below each R code statement). Which calculation is the correct power for the scenario in this question? (5 marks)

Circle the correct answer

```
# qnorm(x) is the xth percentile of the standard normal distribution
#
# pnorm(x) is the standard normal cumulative distribution function
# evaluated at x
#
# dnorm(x) is the standard normal density function evaluated at x

1-pnorm(qnorm(1-.05)-1/(9*sqrt(1/80+1/80)))
```

```
[1] 0.1730643
```

```
1-pnorm(qnorm(1-.05/2)-1/(3*sqrt(1/80+1/80)))
```

```
[1] 0.5589159
```

```
1-dnorm(qnorm(1-.05)-1/(3*sqrt(1/80+1/80)))
```

```
[1] 0.6416613
```

```
1-pnorm(qnorm(1-.05)-1/(3*sqrt(1/80+1/80)))
```

```
[1] 0.6784366
```

```
1-dnorm(dnorm(.05/2)-1/(3*sqrt(1/80+1/80)))
```

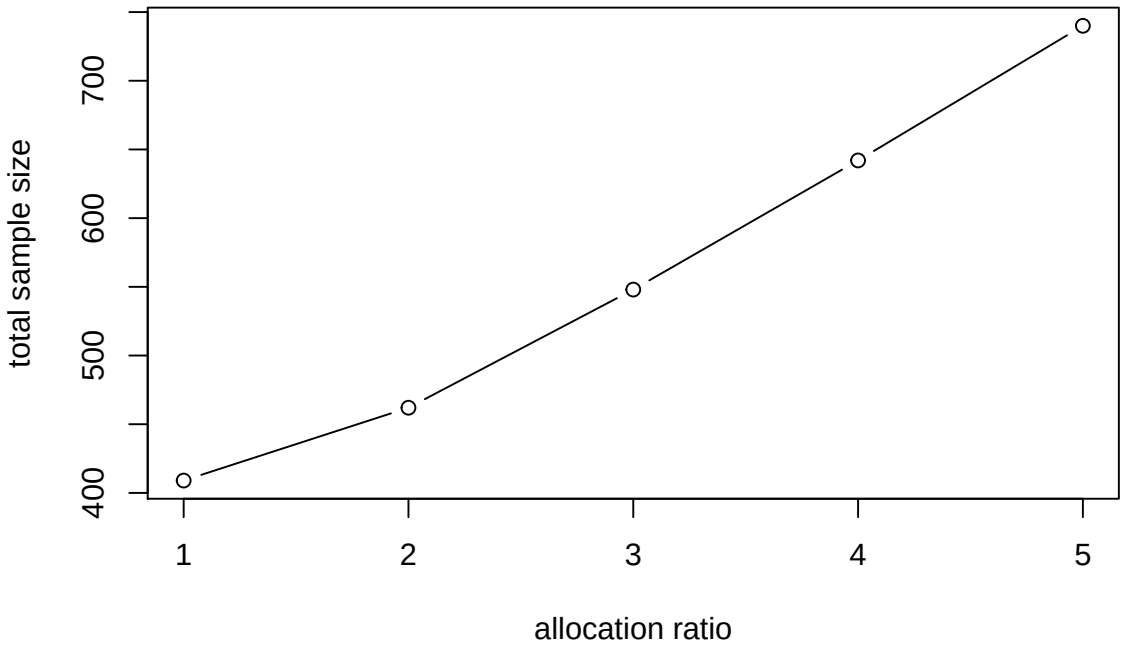
```
[1] 0.9074408
```

```
1-pnorm(dnorm(.05)-1/(3*sqrt(1/80+1/80)))
```

```
[1] 0.9563431
```

Continue on next page ...

(d) A clinical trial to test an experimental drug for prostate cancer is being designed by a group of university researchers. The response rate of the standard chemotherapy for prostate cancer is 22%, and the researchers expect that the experimental drug would increase the response rate of the standard treatment to 35% . The researchers would like to have an allocation ratio between the standard and experimental arms of 2:1. The graph below displays the total sample size required for 90% power at the 0.05 significance level. If the type I error rate is $\alpha = 0.05$ then approximately how many patients should the researchers recruit into the experimental arm so that their study has 90% power? (5 marks)



Continue on next page ...

(e) A study is being designed to compare the success rates of a standard drug versus experimental drug. The researchers estimate that it's feasible to enrol 200 subjects per group. A statistician simulated 30 hypothetical clinical trials using R, where $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 = 0.10$. The simulation assumed that $p_1 = 0.55$ is the success rate in the experimental group and $p_2 = 0.45$ is the success rate in the standard group. The table below shows the 30 simulated studies with their P-values. (5 marks)

Study Number	P-value
1	0.00
2	0.00
3	0.00
4	0.00
5	0.00
6	0.00
7	0.01
8	0.01
9	0.01
10	0.01
11	0.01
12	0.01
13	0.01
14	0.02
15	0.02
16	0.02
17	0.03
18	0.04
19	0.09
20	0.11
21	0.13
22	0.13
23	0.19
24	0.23
25	0.27
26	0.32
27	0.37
28	0.55
29	0.62
30	0.84

Continue on next page ...

Answer the following questions using the results of the simulation.

- Estimate the power of the test at the 1% significance level? Show your work. (3 marks)

- If the success rate in the experimental group were increased to $p_1 = 0.60$, but the success rate in the standard group remained the same (i.e., $p_2 = 0.45$) then would the power increase or decrease if the researchers used a significance level of 1%? Explain. (2 marks)

Continue on next page ...

3. In 1999, Joshua Angrist and Victor Lavy published an unusual study of the effects of class size on academic achievement. They wrote:

> Causal effects of class size on pupil achievement have proved very difficult to measure. Even though the level of educational inputs differs substantially both between and within schools, these differences are often associated with factors such as remedial training or students' socioeconomic background . . . The great twelfth century Rabbinic scholar, Maimonides, interprets the Talmud's discussion of class size as follows: 'Twenty-five children may be put in charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with instruction. If there are more than forty, two teachers must be appointed.' . . . The importance of Maimonides' rule for our purposes is that, since 1969, it has been used to determine the division of enrollment cohorts into classes in Israeli public schools.

Maimonides' rule has the largest impact on a school with about 40 students in a grade cohort. With cohorts of size 40, 80, and 120 students, the steps down in average class size required by Maimonides' rule when an additional student enrolls are, respectively, from 40 to 20.5, from 40 to 27, and from 40 to 30.25.

In Israel (and in many other countries) a child will typically enrol in a school if they live close to the school. Adherence to Maimonides' rule in Israeli public schools is not always followed. In particular, Angrist and Lavy note that the percentage of disadvantaged students in a school "is used by the Ministry of Education to allocate supplementary hours of instruction and other school resources."

The primary goal of the analysis below is to estimate the effect of class size on math scores. Data from 318 grade 5 Israeli public school classes were used to assess this question.

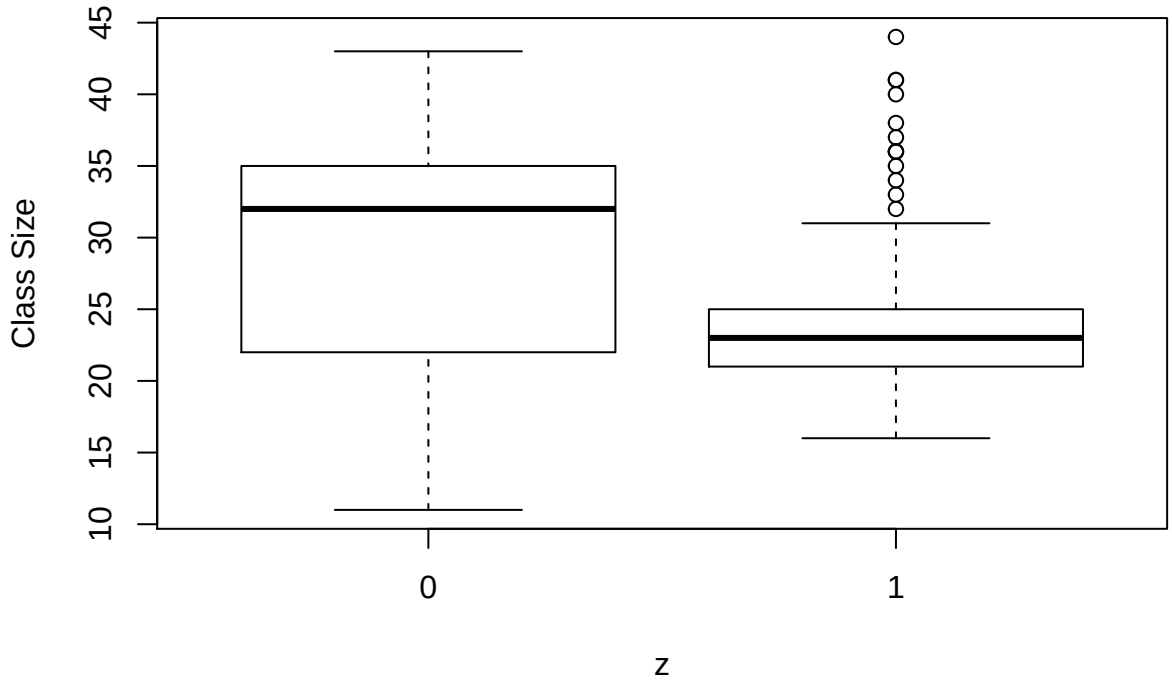
The variable names are defined in the table below

Label	Variable Description
`z`	1= cohort of at least 41 students/2 or more classes, 0= cohort of less than 40 students/1 class.
`pct_disadv`	Percentage of students that are economically disadvantaged
`avgmath`	Average math test score for each class.
`classsize`	Size of classes

Continue on next page ...

Description of the data

```
boxplot(classize~z,data = AngristLavy,xlab="z",ylab="Class Size")
```




```
# distribution of z
table(AngristLavy$z)
```

```
  0   1
 93 225
```

```
# summary of class size when z=1
summary(AngristLavy$classize[AngristLavy$z==1])
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  16.00   21.00   23.00   23.62   25.00   44.00
```

```
# summary of class size when z=0
summary(AngristLavy$classize[AngristLavy$z==0])
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  11.00   22.00   32.00   29.14   35.00   43.00
```

Continue on next page ...

Three different types of data analyses were performed on the data.

Analysis #1

```
# Analysis 1
mod1 <- lm(z~pct_disadv,data=AngristLavy)
summary(mod1)$coefficients
```

```
                Estimate  Std. Error    t value     Pr(>|t|)
(Intercept)  0.7158920892 0.040943590 17.4848392 2.328662e-48
pct_disadv  -0.0003918034 0.001500745 -0.2610727 7.942065e-01
```

```
pscore <- predict(mod1,type = "response")
mod2 <- lm(avgmath~z+pscore,data=AngristLavy)
summary(mod2)$coefficients
```

```
               Estimate Std. Error    t value     Pr(>|t|)
(Intercept) -545.592031  51.489809 -10.596117 1.218817e-22
z              3.737117   1.068721   3.496812 5.385058e-04
pscore       858.098143  72.776868  11.790809 7.952584e-27
```

```
confint(mod2)
```

```
                  2.5 %      97.5 %
(Intercept) -646.899442 -444.284620
z              1.634383    5.839851
pscore       714.907943 1001.288342
```

Analysis #2

```
# Analysis 2
rr <- Match(Y=AngristLavy$avgmath,Tr=AngristLavy$z,X=mod1$fitted)
summary(rr)
```

```
Estimate...  3.1007
AI SE......  1.3593
T-stat.....  2.2811
p.val......  0.022543

Original number of observations..............  318
Original number of treated obs...............  225
Matched number of observations...............  225
Matched number of observations  (unweighted).  610
```

```
# Quantile of the t-distribution on 224 degrees of freedom
qt(p = .975,df = 224 )
```

```
[1] 1.970611
```

Continue on next page ...

Analysis #3

```
# Analysis 3
t.test(avgmath~z,data=AngristLavy)
```

```
    Welch Two Sample t-test

data:  avgmath by z
t = -2.8443, df = 147.82, p-value = 0.005082
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.647189 -1.197136
sample estimates:
mean in group 0 mean in group 1
       61.42195        65.34412
```

Continue on next page ...

Answer the following questions. (25 marks)

(a) What type of design was used in this study? What is the treatment and how was the treatment assigned to the experimental units? Briefly explain each of your answers. (6 marks)

(b) Describe the type of analysis conducted in Analysis #1, Analysis #2, and Analysis #3 to evaluate the treatment effect. (6 marks)

Continue on next page ...

(c) What is the estimated effect and 95% confidence interval of class size on math scores in Analysis #1, Analysis #2, and Analysis #3? Briefly explain. (6 marks)

(d) Is the treatment assignment in this study ignorable? Do smaller class sizes cause higher math scores? Briefly explain. (7 marks)

Continue on next page ...

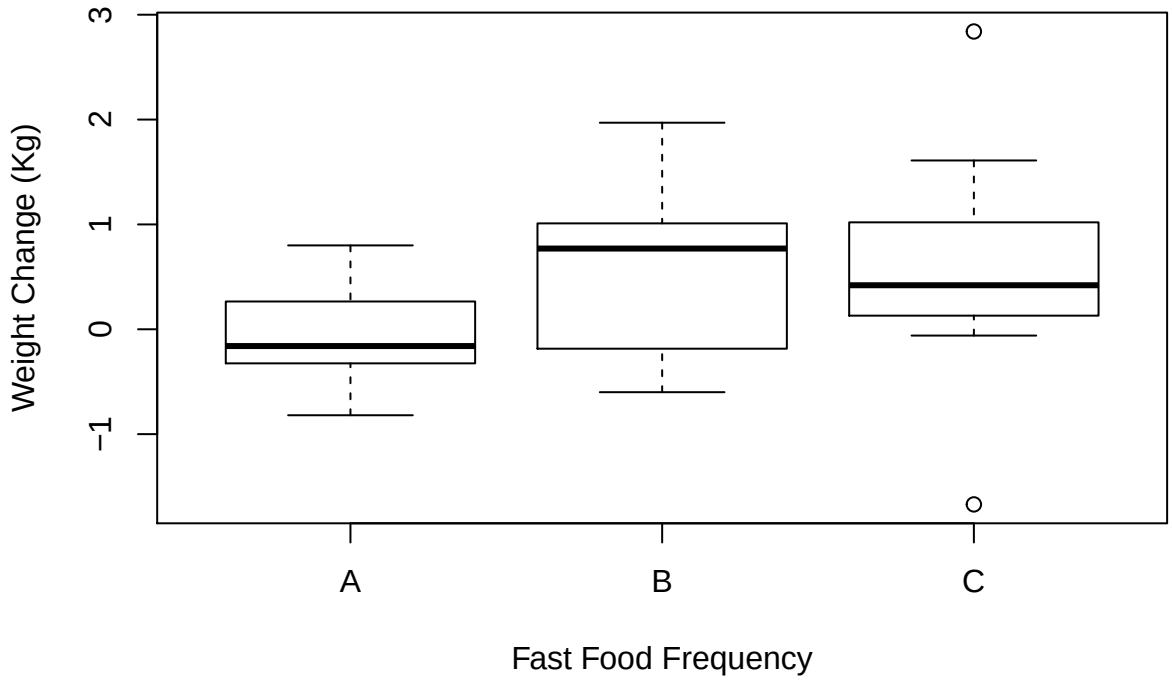
4. A study at UofT recruited twenty-one students to complete a thirty minute survey on diet and eating habits during the academic year. Students were paid $10 to complete the survey. The data below shows their weight gain from September to April classified by the frequency that students ate fast food. In group A students reported never eating fast food; students in group B reported eating fast food twice per month; and students in group C reported eating fast food four times per month.

	A	B	C
	0.52	1.14	-0.06
	-0.82	0.10	0.32
	-0.23	-0.60	2.84
	-0.42	1.97	-1.67
	0.01	-0.47	0.42
	-0.16	0.77	0.43
	0.80	0.88	1.61
Treatment Average	-0.04	0.54	0.56
Treatment SD	0.55	0.92	1.40

Continue on next page ...

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")
```

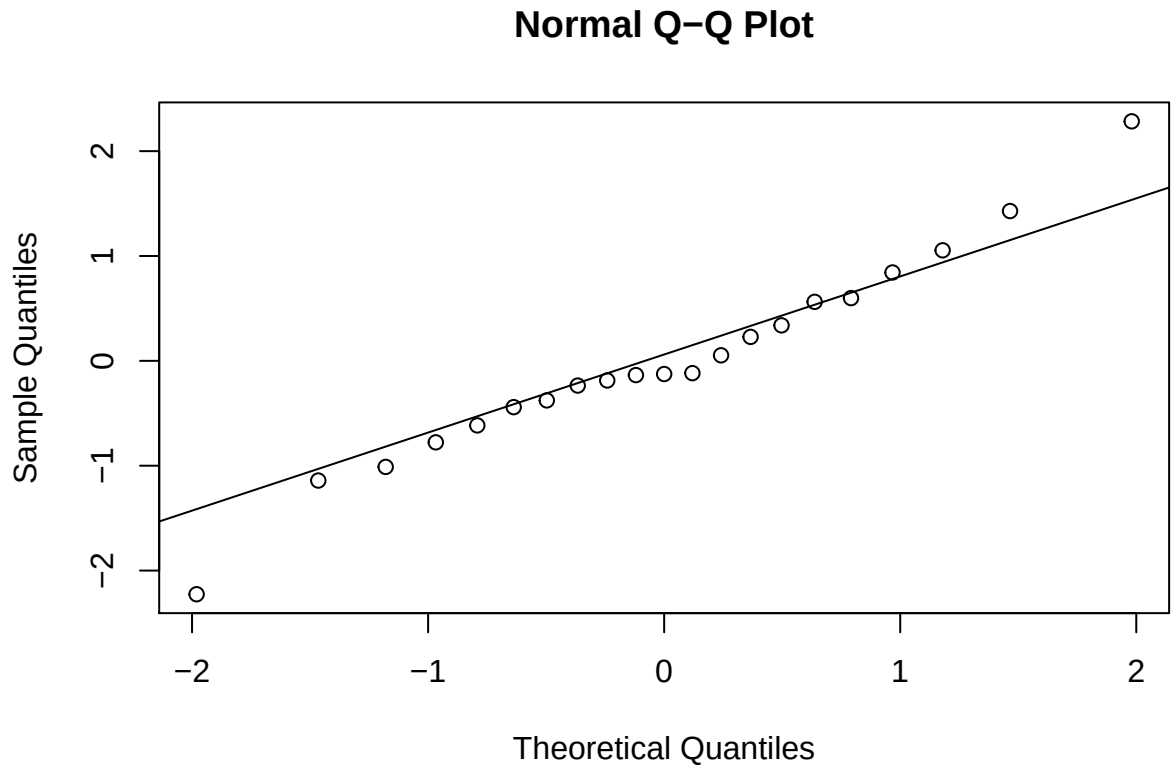



```
aovsurvey <- aov(wtchange~grp,data=surveydat)
summary(aovsurvey)
```

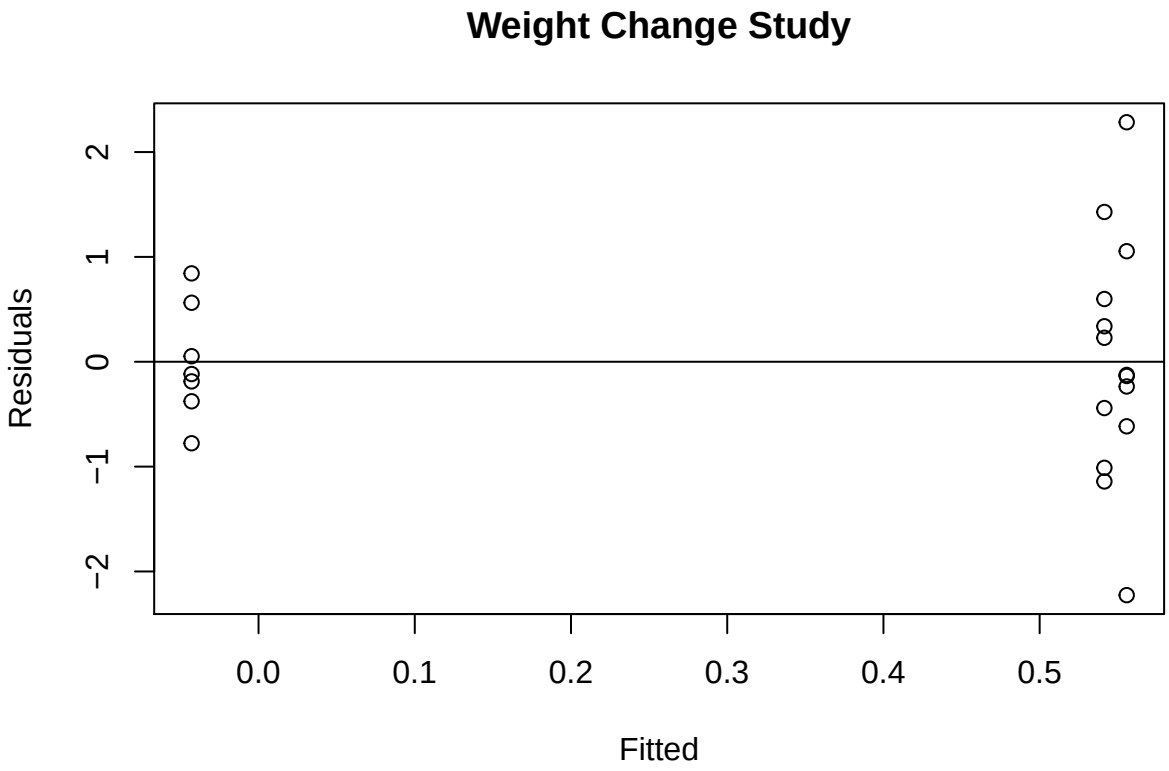
```
             Df Sum Sq Mean Sq F value Pr(>F)
grp           2  1.633  0.8165   0.787   0.47
Residuals    18 18.664  1.0369
```

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

Continue on next page ...




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



Continue on next page ...

Answer the following questions. (20 marks)

(a) Is this study an experiment or observational study? What is the treatment? Briefly explain. (5 marks)

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

Continue on next page ...

(c) What is the ANOVA identity for this data? Is it necessary that the data collected in each treatment group follow a normal distribution for the ANOVA identity to be true? Briefly explain. (5 marks)

(d) 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 twice as large. Is this a valid statement? Briefly explain. (5 marks)

End of Test