STA305/1004 Midterm Test - Solutions

February 25, 2015

Name:

Student Number:

Instructions

Time allowed: 90 minutes

This test should contain 20 pages. Complete all parts of the 4 questions.

The value of each question is shown in the table below.

Question	Mark	Total
1		30
2		20
3		25
4		25
Total Marks		100

Complete all questions in pen. Any questions completed in pencil *may* not be eligible to be remarked even if there was a marking error. Show all of your work. Feel free to use the back side of the test if necessary.

Aids allowed: You are allowed to bring in an 8.5'x11' sheet with hand writing on both sides and a calculator.

1. Suppose that two drugs A and B are to be tested on subjects' eyes. The drugs will be randomly assigned to each subject based on the flip of a fair coin. If the coin toss is heads then a subject will receive drug A in both eyes; if the coin toss is tails then a subject will recieve drug B in both eyes. The outcome of interest is intraoccular pressure (IOP). The smallest clinically meaningful difference in IOP between the two drugs is 1mmHg. The standard deviation of IOP is 3mmHg for each drug, and the standard deviation for the differences of IOP between A and B within each person is 1.

(a) Use the R output below to determine how many subjects the researchers should enrol so that the study has 90% power to detect a mean difference of 1 mmHg at the 5% significance level (5 marks)? Explain your reasoning.

This is a two-sample design not a paired design. Therefore, the total sample size is $2 \times 190 = 380$.

```
power.t.test(n = NULL,power=0.9,delta = 1,sd = 3,
             sig.level = 0.05,type = "two.sample",
             alternative = "two.sided")
```

```
##
##      Two-sample t test power calculation
##
##               n = 190.0991
##           delta = 1
##              sd = 3
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
power.t.test(n = NULL,power=0.9,delta = 1,sd = 1,
             sig.level = 0.05,type = "paired",
             alternative = "two.sided")
```

```
##
##      Paired t test power calculation
##
##               n = 12.58547
##           delta = 1
##              sd = 1
##       sig.level = 0.05
##           power = 0.9
```

```
##     alternative = two.sided
##
## NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

(b) A clinical trial where an experimental drug is to be compared with the standard treatment in terms of a continuous biomarker measurement is being planned. The biomarker measurements are independent and normally distributed with different means μ_1, μ_2, and the same variance σ^2. The sample size formula for calculating, n, the **total sample size** for a test of $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$, where $\theta = \mu_1 - \mu_2$ at level α with power $1 - \beta$ is:

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

where z_α is the $100(1-\alpha)^{th}$ percentile of the standard normal distrbution.

Show that the power function of this test is:

$$1 - \beta = 1 - \Phi \left(z_{\alpha/2} - \frac{\sqrt{n}|\theta|}{2\sigma} \right),$$

where $|\theta| = \sqrt{\theta^2}$, for $-\infty < \theta < \infty$ (5 marks).

$$\begin{aligned}
n &= \frac{4\sigma^2}{\theta^2} \left(z_{\alpha/2} + z_{\beta} \right)^2 \\
\Rightarrow \sqrt{\frac{\theta^2 n}{4\sigma^2}} &= z_{\alpha/2} + z_{\beta} \\
\Rightarrow \frac{|\theta|\sqrt{n}}{2\sigma} - z_{\alpha/2} &= z_{\beta} \\
\Rightarrow \Phi \left(\frac{|\theta|\sqrt{n}}{2\sigma} - z_{\alpha/2} \right) &= \Phi \left(z_{\beta} \right) = 1 - \beta \\
\Rightarrow 1 - \Phi \left(z_{\alpha/2} - \frac{|\theta|\sqrt{n}}{2\sigma} \right) &= 1 - \beta, since, \Phi(x) = 1 - \Phi(-x)
\end{aligned}$$

(c) Consider again the situation in part 1. (b). The smallest clinically meaningful treatment difference is $\theta = 1$ and the variance is $\sigma^2 = 9$. If the means will be compared using a two-sided Z-test at a significance level of $\alpha = 0.05$ then the power function in part (b) can be used.

The power of the study with 80 subjects in each arm was calculated by the clinical trial statistician using R (the value for power is below each R code statement). Which calculation gives the correct power for the scenario in this question? (5 marks)

Circle the correct answer

The correct answer is the second set of code, since $1 - \Phi\left(z_{.05/2} - \frac{\sqrt{160\times1}}{2\times3}\right) = 0.5589159$.

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

1-pnorm(qnorm(1-.05/2)-sqrt(160)*1/(2*9))
```

```
## [1] 0.1043342
```

```
1-pnorm(qnorm(1-.05/2)-sqrt(160)*1/(2*3))
```

```
## [1] 0.5589159
```

```
1-pnorm(qnorm(.05)-sqrt(160)*1/(2*3))
```

```
## [1] 0.9999126
```

```
1-pnorm(qnorm(1-.05)-sqrt(160)*1/(2*3))
```

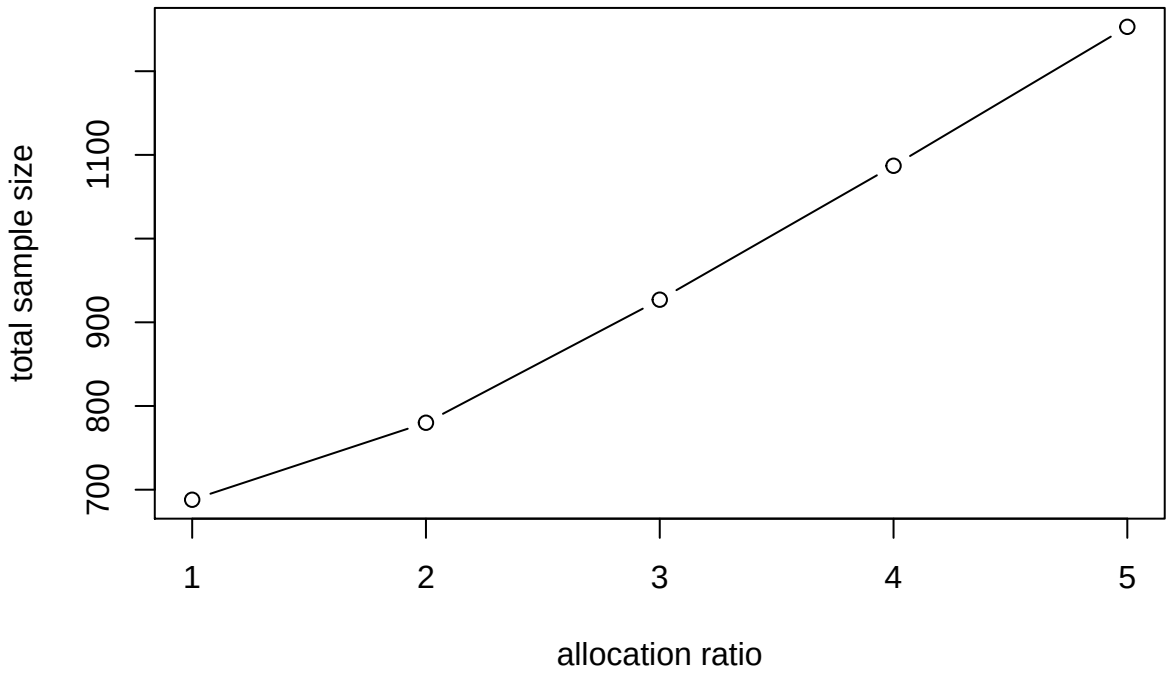
```
## [1] 0.6784366
```

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

```
## [1] 0.438749
```

(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 by 50% . The researchers would like to have an allocation ratio between the experimental and standard arms of 2:1 (i.e., for every subject in the standard arm there will be two subjects in the experimental arm).

The graph below displays the total sample size required for 90% power at the 0.025 significance level. If the type I error rate is $\alpha = 0.025$ then approximately how many patients should the researchers recruit into the experimental arm so that their study has 90% power? (5 marks)



The total sample size is $n = n_{exp} + n_{standard}$ *and* $n_{exp} = 2n_{standard}$*. The graph shows that* $n = 800$*. So,* $800 = n_{exp} + \frac{n_{exp}}{2} = \frac{3}{2}n_{exp} \Rightarrow n_{exp} = 2/3 \times 800 = 533$

(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 476 subjects per group. The hypothesized success rates in the two groups are: $p_1 = 0.4$ in the experimental group; and $p_2 = 0.33$ in the standard group. A statistician simulated 25 hypothetical clinical trials using R, where $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 = 0.07$. The table below shows the 25 simulated studies with their P-values.

Answer the following questions using the R output on the next page:

- What is the power of the test at the 5% significance level? Show your work. (2 marks)

  20 out of the 25 p-values are significant at the 5% level. Therefore the power is 80%.

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

  15 out of the 25 p-values are significant at the 1% level. Therefore the power is 60%.

- If the success rate in the experimental group were decreased to $p_1 = 0.38$, but the success rate in the standard group remained the same (i.e., $p_2 = 0.33$) then would the power increase or decrease if the researchers used a significance level of 5%? Explain. (2 marks)

  The power would decrease since the alternative value is closer to the null hypothesis value.

```r
#power simulation
N <- 25
#simulate N counts from Bin(476,0.33)
x1_sim <- rbinom(n = N,size =476,prob = 0.33 )
#simulate N counts from Bin(476,0.4)
x2_sim <- rbinom(n = N,size =476,prob = 0.4 )
pvalres <- numeric(N) #store p-values from simulation
for (i in 1:N)
{
  # calculate p-value for comparing two proportions
  pvalres [i] <- prop.test(x=c(x1_sim[i],x2_sim[i]),n=c(476,476),
  alternative = "two.sided",correct = F)$p.value
}
```

Study Number	P-value
1	0.00
2	0.00
3	0.00
4	0.00
5	0.00
6	0.00
7	0.00
8	0.00
9	0.01
10	0.01
11	0.01
12	0.01
13	0.01
14	0.01
15	0.01
16	0.02
17	0.02
18	0.02
19	0.03
20	0.03
21	0.11
22	0.20
23	0.21
24	0.25
25	0.79

(f) Twenty one student athletes, were randomly grouped into three teams A, B, and. C. The number of successful attempts (y) to toss a basketball through a hoop within a fixed time period were recorded. Is there a significant difference at the 1% significance level between at least two of the teams in the mean number of successful shots? The $F_{2,18}$ 1% critical value is 6.01 (i.e., $P(F_{2,18} > 6.01) = 0.01$). Justify your answer. (5 marks)

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
trt        2     98  49.000       9
Residuals 18     98   5.444
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

There is a significance difference since $P(F_{2,18} > 9) < 0.01$. Therefore we reject $H_0 : \mu_1 = \mu_2 = \mu_3$ in favour of $H_1 : \mu_i \neq \mu_j$.

2. A study to determine the amount of wear in the soles of boys shoes was conducted. The shoe soles were made of two different synthetic materials: a standard material A and a cheaper material B. Each boy wore a special pair of shoes, the sole of one shoe made with material A and the sole of the other with B. The decision as to whether the left or right sole was made with A or B was determined by the flip of a fair coin. The data from the study is shown in the table below (matA is the amount of wear from the shoe using material A, sideA is the foot (L - left, R -right) material A was used on. matB and sideB are defined in a similar manner)

boy	matA	sideA	matB	sideB	diff
1	13.2	L	14.0	R	0.8
2	8.2	L	8.8	R	0.6
3	10.9	R	11.2	L	0.3
4	14.3	L	14.2	R	-0.1
5	10.7	R	11.8	L	1.1
6	6.6	L	6.4	R	-0.2
7	9.5	L	9.8	R	0.3
8	10.8	L	11.3	R	0.5
9	8.8	R	9.3	L	0.5

```
#Two-sample t-test equal variance
t.test(shoedat$matA,shoedat$matB,paired=FALSE,var.equal=TRUE)
```

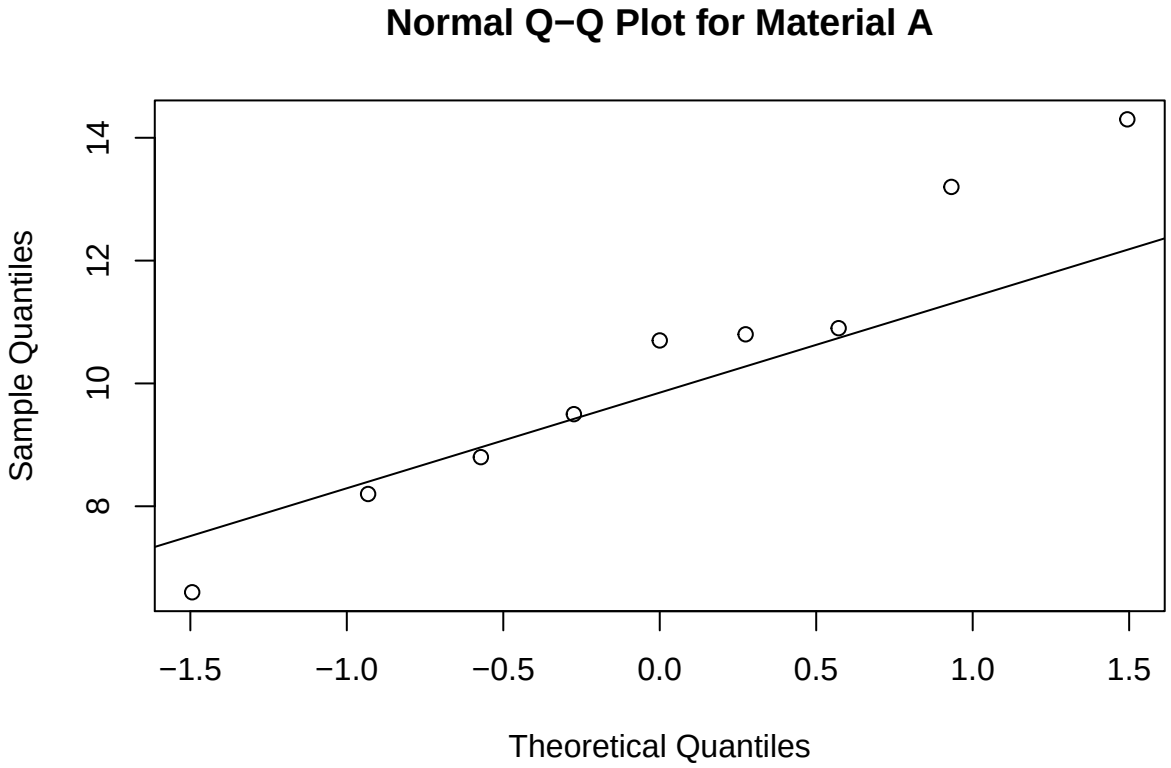
```
##
##  Two Sample t-test
##
## data:  shoedat$matA and shoedat$matB
## t = -0.3657, df = 16, p-value = 0.7194
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.869596  2.025152
## sample estimates:
## mean of x mean of y
##  10.33333  10.75556
```

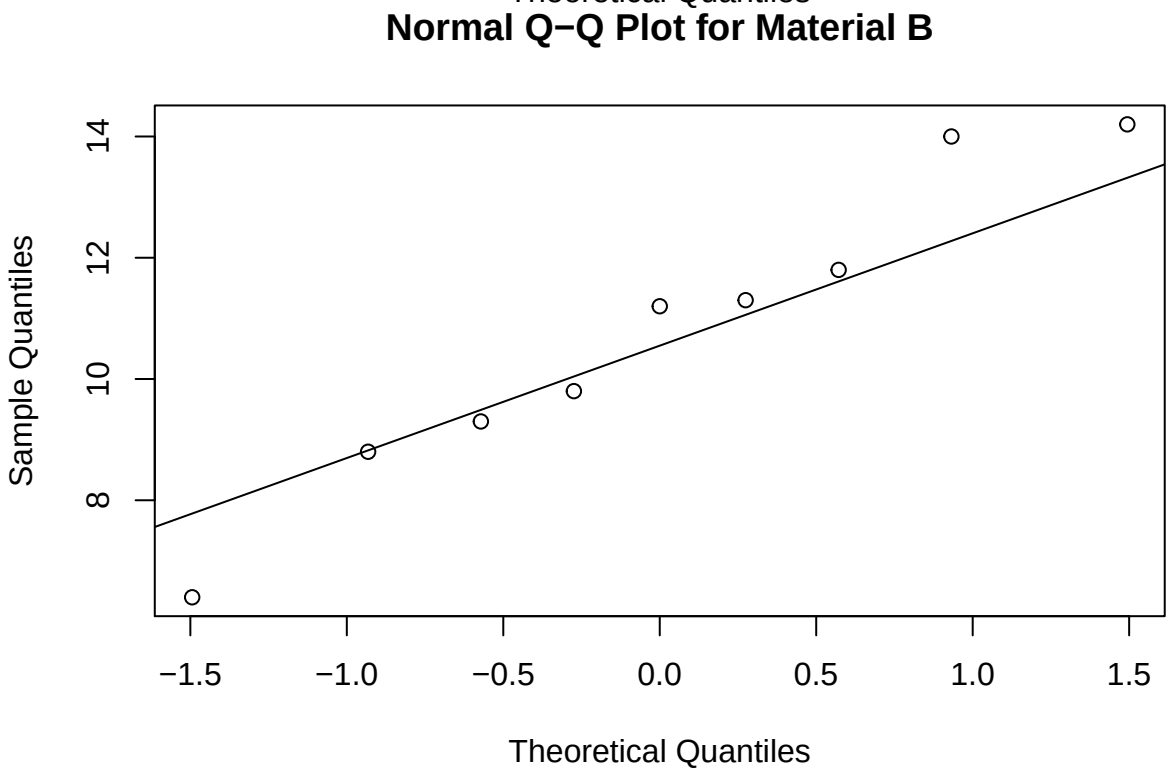
```r
#Two-sample t-test unequal variance
t.test(shoedat$matA,shoedat$matB,paired=FALSE,var.equal=FALSE)
```

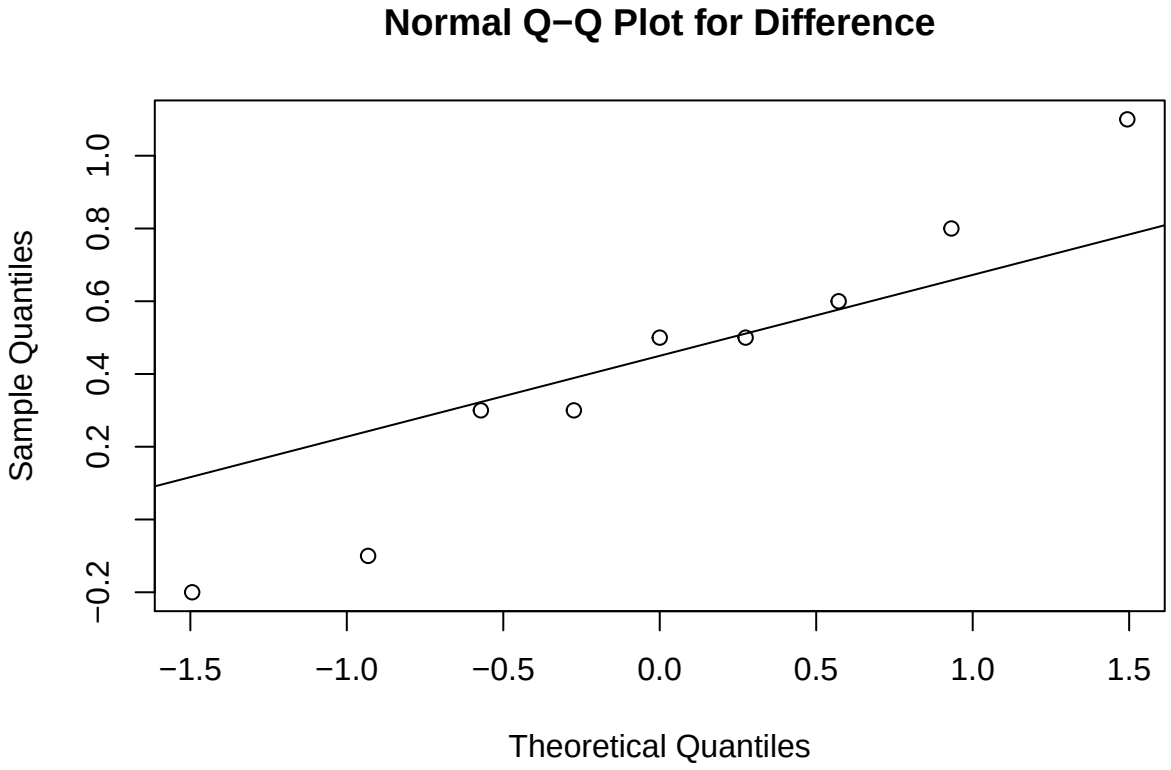
```
##
##  Welch Two Sample t-test
##
## data:  shoedat$matA and shoedat$matB
## t = -0.3657, df = 15.977, p-value = 0.7194
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.869883  2.025438
## sample estimates:
## mean of x mean of y
##  10.33333  10.75556
```

```r
#Paired t-test
t.test(shoedat$matA,shoedat$matB,paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  shoedat$matA and shoedat$matB
## t = -3.1001, df = 8, p-value = 0.01466
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7362910 -0.1081534
## sample estimates:
## mean of the differences
##              -0.4222222
```


Normal Q−Q Plot for Material A
Sample Quantiles
8
10
12
14
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Theoretical Quantiles


Normal Q−Q Plot for Material B
Sample Quantiles
8
10
12
14
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Theoretical Quantiles


Normal Q−Q Plot for Difference
Sample Quantiles
−0.2
0.2
0.4
0.6
0.8
1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Theoretical Quantiles

(a) What is the name of the study design? Justify your answer (5 marks)

Randomized paired design. Material was randomized within each boy.

(b) Is this study an experiment or observational study? Justify your answer. (5 marks)

An experiment since the probability of treatment allocation is randomized.

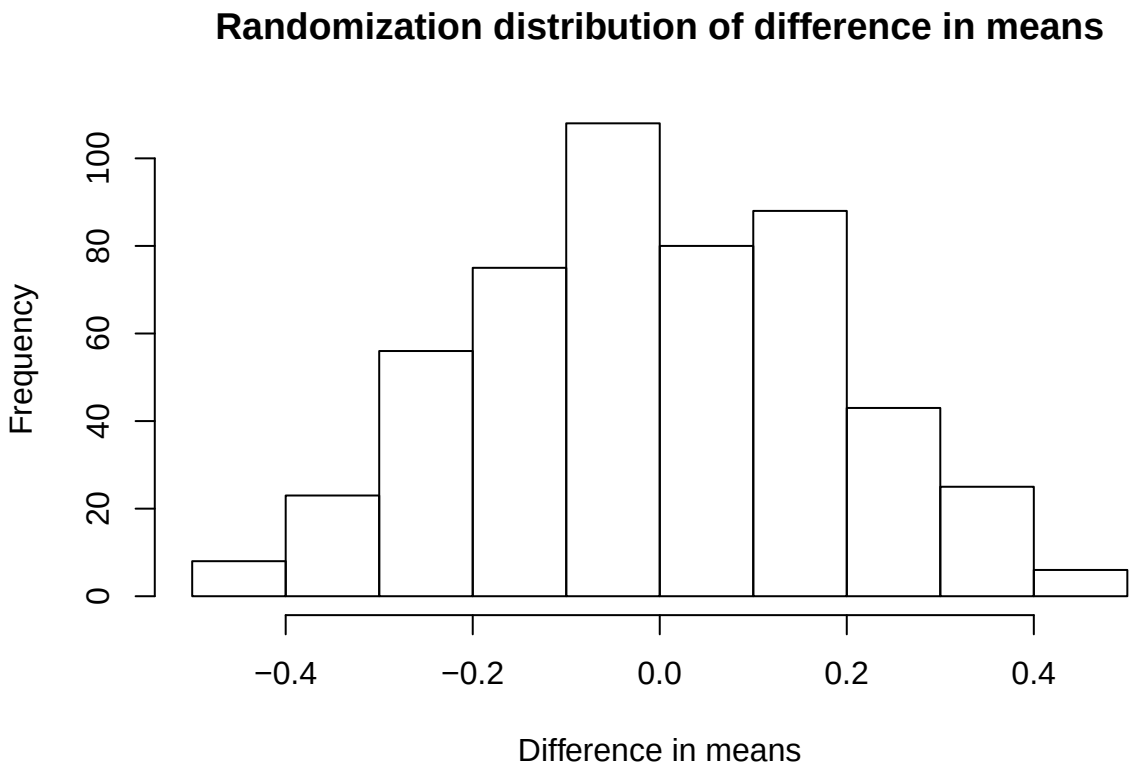
(c) Is there a statistically significant difference between the mean wear of materials A and B at the 5% significance level? Justify your answer. (5 marks)

Yes, since the p-value is less than 0.05.

(d) Are the assumptions behind the appropriate significance test for these data justified? (5 marks)

Yes. The Q-Q plot for the differences does not deviate too much from the straight line.

(3) The randomization distribution for the difference in means between materials A and B for the boy's shoe study (from the previous question) was generated. A plot of the randomization distribution is below:



The mean observed difference between material A and material B was -0.42. The number of mean differences in the randomization distribution that are greater than or equal to the observed difference is 506.

(a) How many outcomes make up the randomization distribution? (5 mark)

$2^9 = 512$.

(b) Describe how the randomization distribution was generated. (5 mark)

Generate all 512 treatment allocations. This corresponds to calculating the average for each of the 512 arrangements of signs in:

$$\bar{d} = \frac{\pm 0.8 \pm 0.6 \pm 0.3 \pm 0.1 \pm 1.1 \pm 0.2 \pm 0.3 \pm 0.5 \pm 0.5}{9}$$

(c) Give an example of how you would calculate two different values in this distribution using the table containing the data in question 2. (5 marks)

Two examples are:

$$\bar{d}_1 = \frac{0.8 + 0.6 + 0.3 + 0.1 + 1.1 + 0.2 + 0.3 + 0.5 + 0.5}{9}$$
$$\bar{d}_2 = \frac{0.8 + 0.6 + 0.3 + 0.1 - 1.1 - 0.2 - 0.3 - 0.5 - 0.5}{9}$$

(d) Is there evidence at the 5% significance level that material A wears less than material B? Calculate the randomization p-value for the test $H_0 : \mu_A - \mu_B = 0$ versus $H_0 : \mu_A - \mu_B < 0$, where μ_A and μ_B are the mean wear of materials A and B respectively to assess this question. What do you conclude? (5 marks)

The p-value is $\frac{\{\#\bar{d}_i \leq -0.42\}}{512} = \frac{512-506+1}{512} = \frac{7}{512} = 0.01$. *Therefore, there is evidence at the 5% level that material A wears less than material B.*

(e) In another similar study boys wore both left and right shoes with the same material (i.e., either both shoes were made of material A or both were made of material B). The decision whether a boy recieved a pair of shoes with A or B was determined by how hard each boy wore his shoes in the past, age, and level of activity. Is the propensity score known or unknown? If it is known then state the propensity score otherwise state how you would estimate the propensity score. (5 marks)

The propensity score is unknown. The propensity score can be estimated using the predicted probabilities from a logistic regression model of the form:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

where p is the probability of wearing shoes with material A or B and x_1, x_2, x_3 *are the covariates corresponding to past hardness, age, and level of activity.*

4. (*Adapted from Hill and Gelman, Chapter 9*) The table below describes a hypothetical study on 2400 persons. Each row of the table specifies a category of person, as defined by his or her pre-treatment covariate x, treatment indicator T , and potential outcomes $Y(0)$, $Y(1)$. (For simplicity, we assume unrealistically that all the people in this experiment fit into these eight categories.)

Category	persons in category	x	T	$Y(0)$	$Y(1)$
1	300	0	0	4	6
2	300	1	0	4	6
3	500	0	1	4	6
4	500	1	1	4	6
5	200	0	0	10	2
6	200	1	0	10	2
7	200	0	1	10	2
8	200	1	1	10	2

In making the table we are assuming omniscience (knowing everything), so that we know both $Y(0)$ and $Y(1)$ for all observations. But the (nonomniscient - not knowing everything) investigator would only observe x, T, and $Y(T)$ for each unit. (For example, a person in category 1 would have $x = 0$,$T = 0$,$Y(0) = 4$, and a person in category 3 would have $x = 0$,$T = 1$,$Y(1) = 6$.)

(a) Give an example of a context for this study. Define $x, T, Y(0), Y(1)$. (5 marks)

A study to estimate the effect of smoking cessation on weight loss. In this case, let Y be weight loss (continuous), T be smoking cessation (0 or 1), and x be sex (male or female).

(b) *What is the average treatment effect in this population of 2400 persons? (5 marks)*

The average treatment effect is $\bar{Y}(0) - \bar{Y}(1)$.

$$\bar{Y}(0) = \frac{4 \times (600 + 1000) + 10 \times 800}{2400} = 6$$
$$\bar{Y}(1) = \frac{6 \times (600 + 1000) + 2 \times 800}{2400} = 4.7$$
$$\Rightarrow \bar{Y}(0) - \bar{Y}(1) = 6 - 4.7 = 1.3.$$

(c) Is it plausible to believe that these data came from a randomized experiment? Defend your answer. (5 marks)

In a randomized experiment we would expect that the covariate x to have a similar distribution in each treatment group (i.e., balanced).

$$P(x = 0|T = 0) = \frac{300 + 200}{300 + 300 + 200 + 200} = 0.5$$
$$P(x = 0|T = 1) = \frac{500 + 200}{500 + 500 + 200 + 200} = 0.5$$

The covariate x is balanced between the two treatment groups. Therefore, it's plausible that these data came from a randomized experiment.

(d) Another population quantity is the mean of Y for those who received the treatment minus the mean of Y for those who did not. What is the relation between this quantity and the average treatment effect? (5 marks)

This quantity is the observed treatment effect. It's calculated based on subjects' outcome under the treatment that was received. Subjects' potential responses under treatments that they did not receive are missing when this quantity is used. This quantity is calculated as:

$$\bar{Y}^{obs}(0) = \frac{4 \times (300 + 300) + 10 \times (200 + 200)}{300 + 300 + 200 + 200} = 6.4$$
$$\bar{Y}^{obs}(1) = \frac{6 \times (500 + 500) + 2 \times (200 + 200)}{500 + 500 + 200 + 200} = 4.9$$
$$\Rightarrow \bar{Y}^{obs}(0) - \bar{Y}^{obs}(1) = 6.4 - 4.9 = 1.5.$$

(e) *For these data, is it plausible to believe that treatment assignment is ignorable given the covariate x? Defend your answer. (5 marks)*

If treatment assignment is ignorable then treatment assignment is independent of the potential responses conditional on the covariate. So, for example, if treatment assignment is ignorable then $P(T = 0|Y(0) = 4, Y(1) = 6, x = 0) = P(T = 0|x = 0)$. These probabilities can be calculated from the data and we already calculated $P(T = 0|x = 0) = 0.42$ in part (c).

$$P(T = 0|Y(0) = 4, Y(1) = 6, x = 0) = \frac{300}{300 + 500} = 0.375.$$

Therefore, treatment assignment depends on the potential responses and is non-ignorable.