STA305/1004 Midterm Test - Solutions

March 2, 2016

NB: Some of the solutions shown are brief. More detail may be required for full marks.

Name: _____

Student Number: _____

Instructions:

Time allowed: 90 minutes

Answer all four questions.

Complete all questions in pen. Any questions completed in pencil may not be eligible to be remarked even if there was a marking error.

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

Question	Value	Mark
1	25	
2	20	
3	20	
4	20	
Total	85	

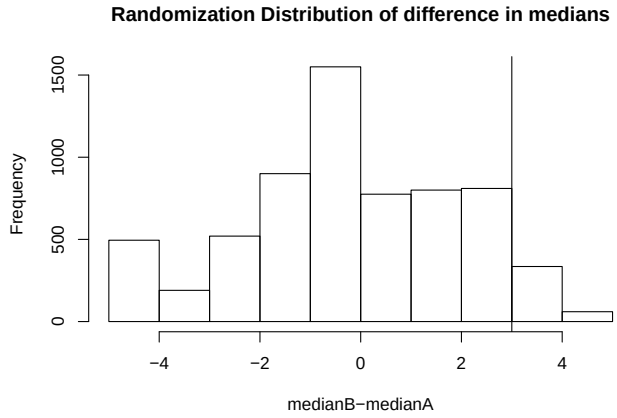
This test should have 14 pages including this page.

1. Fifteen judges were randomly allocated to judge one of two brands of beer, A or B, for taste. Eight judges will be assigned to Brand A and seven judges to brand B. The judges ranked the beer they tasted using a 10-point (Likert) scale with 1 representing "poor taste" and 10 representing "outstanding taste".
   The table below shows the rating from each judge. The number in brackets beside the rating indicates which judge gave the rating. For example, judge 1 gave a rating of 2 to brand A, and judge 9 gave a rating of 3 to brand B.

Brand A	2 (1)	4 (2)	2 (3)	1 (4)	9 (5)	9 (6)	2 (7)	2 (10)
Brand B	8 (8)	3 (9)	5 (11)	3 (12)	7 (13)	7 (14)	4 (15)	

   The randomization distribution and a one-sided p-value for testing $H_0 : \tilde{\mu}_A = \tilde{\mu}_b$ versus $H_1 : \tilde{\mu}_B > \tilde{\mu}_A$, where $\tilde{\mu}_A, \tilde{\mu}_B$ are the median taste rating for brands A and B respectively were calculated. Some of the R code is shown below.

```
yA <- c(2,4,2,1,9,9,2,2)
yB <- c(8,3,5,3,7,7,4)
beer <- c(yA,yB) #pool data
for (i in 1:N)
{
  res[i] <- median(beer[index[,i]])-median(beer[-index[,i]])
}
hist(res,xlab="medianB-medianA", main="Randomization Distribution of difference in medians")
observed <- median(yB)-median(yA) #store observed median difference
abline(v=observed) #add line at observed median diff
```




```
observed # Observed difference in medians
```

```
[1] 3
```

```
sum(res>=observed)/N
```

```
[1] 0.1002331
```

Answer the following questions based on this study. (5 marks each)

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

Experiment. The treatment assignment mechanism is known. In other other words, the mechanism how treatments were assigned to units is known.

(b) Give one example of a possible treatment assignment that is different than the observed treatment assignment. Use the table below to fill in the treatment assigned to each judge. Use A for beer A and B for beer B.

Any treatment assignment besides the observed assignment that assigns 8 judges to A and 7 judges to B.

Judge	Example treatment assignment
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	

(c) What is the propensity score in this study? What is the probability of treatment assignment? Is the treatment assignment ignorable?

The propensity score is the probability of judge assigned, say, brand A $= 8/15$.

The probability of a treatment assignment is $\frac{1}{\binom{15}{7}} = \frac{1}{\binom{15}{8}} = \frac{1}{6435} = 0.0001554002$.

The treatment assignment is ignorable since judges rating should be independent, since beers were randomly assigned, of which beer they were assigned.

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

p-value is 0.1. The p-value > 0.05 therefore there is no evidence of a difference in taste between brands A and B at the 5% level.

(e) Suppose another investigator would like to design a different study where 15 judges rate two beers. But in this new study a randomized paired design will be used instead of the design described above. How would you randomize the treatments to the units? What is the propensity score and probability of a treatment assignment in the paired design?

The beers are the treatments to be assigned to judges which are the units. The experiment could be paired if each judge tasted both beers but in random order. The order of the beers could be randomized by, for example, flipping a coin. For example if a coin toss was heads then the judge tastes A then B, but if the coin toss is tails then judge tastes B then A.

The propensity score is 1/2 and the probability of treatment assignment is $\frac{1}{2^{15}}$, if a fair coin was used for randomization.

2. 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 her experiment is $z = 2.5$ so the p-value=0.012. 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
0.450	-0.13
0.100	1.28
0.050	1.64
0.025	1.96
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) 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. (5 marks)

   There is an error in this question in the normal percentiles table above: $z_{0.45}$ is the 55^{th} percentile. $z_{0.45} = +0.13$ and $z_{0.55} = -0.13$ But, this didn't have an effect on how the question was marked.

   Assume $\mu = 2.5/\sigma\sqrt{20}$.

   Experiment #2 rejects when $\frac{\bar{x}}{\sigma\sqrt{10}} \geq 1.64$ or $\bar{x} \geq 1.64\frac{\sigma}{\sqrt{10}}$. Now standardize

$$P\left(\bar{x} \geq 1.64\frac{\sigma}{\sqrt{10}}\right)$$

   by subtracting μ and dividing by $\sigma/\sqrt{20}$. This gives,

$$\begin{aligned} P\left(\bar{x} \geq 1.64\frac{\sigma}{\sqrt{10}}\right) &= P\left(Z \geq 1.64 - 2.5\sqrt{10/20}\right) \\ &= P\left(Z \geq -0.13\right) \\ &= 1 - P\left(Z < -0.13\right) \\ &= 1 - 0.45 = 0.55. \end{aligned}$$

   This means that the experiment has 55% power.

(b) The psychologist would like a sample size formula for experiment #2 so that she can calculate the sample size required for a level α test that has power $1-\beta$. Show that the sample size formula is

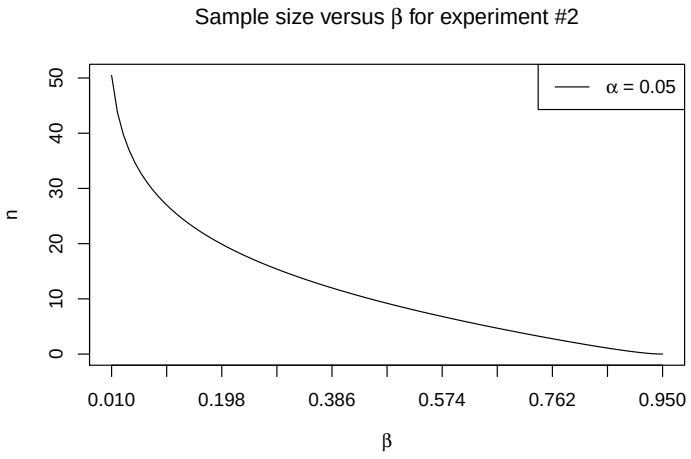
$$n = \left[\frac{\left(z_\alpha - \Phi^{-1}(\beta)\right)\sqrt{20}}{2.5} \right]^2,$$

where $\Phi^{-1}(x)$ is the inverse cumulative distribution function of the $N(0,1)$ distribution, and z_α is the $100(1-\alpha)^{th}$ percentile of the $N(0,1)$ (e.g., $\Phi^{-1}(1-\alpha) = z_\alpha$). (5 marks)

The test rejects when $\bar{x} \geq z_\alpha \frac{\sigma}{\sqrt{10}}$.

$$\begin{aligned}
1-\beta &= P\left(\bar{x} \geq z_\alpha \frac{\sigma}{\sqrt{10}}\right) \\
&= 1 - P\left(Z < z_\alpha - 2.5\sqrt{n/20}\right) \\
&\Rightarrow \Phi^{-1}(\beta) = z_\alpha - 2.5\sqrt{n/20} \\
&\Rightarrow n = \left[\frac{\left(z_\alpha - \Phi^{-1}(\beta)\right)\sqrt{20}}{2.5} \right]^2.
\end{aligned}$$

(c) The sample size formula derived in part (b) was used to create a plot of sample size n versus β.



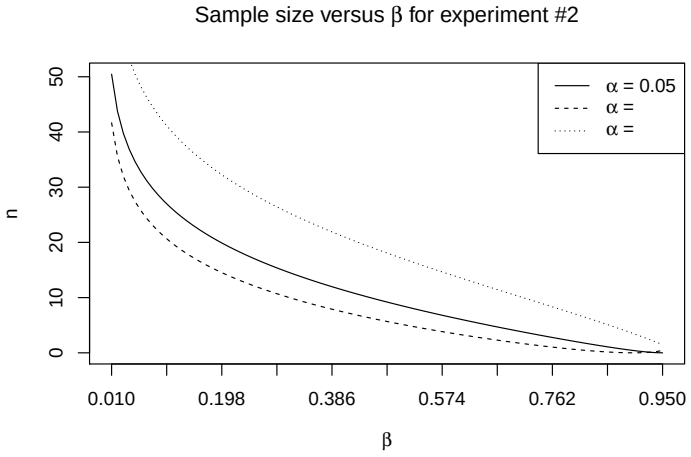
What does the plot tell you about the relationship between sample size and power for experiment #2? Use the plot to estimate how many subjects the psychologist would have to enrol so that experiment #2 will have 80% power at the 5% significance level. If the psychologist wants to be confident of rejecting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

The plot shows that as β decreases n increases which implies that power increases as n increases.

From the plot n is approximately 21.

The psychologist should revise her original sample size of 10 to at least 20 since this would give her approximately 80% power.

(d) Suppose the psychologist decided to change the significance level in experiment #2 from 5% to 1%. The plot of n versus β is now 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. Estimate the sample size required for expereiment #2 to have 80% power at the 1% significance level. Briefly explain how you estimated the sample size. (5 marks)



Using the power curve above $\alpha = 0.05$ curve we have that n is approximately 35. The power curve above is used since as α decreases n should increase for a fixed value of β.

3. What is the effect of smoking on weight gain? Data was used from a voluntary survey to assess this question. Smoking status (cessation/ no cessation) was recorded in 1971, and weight (kg), the outcome of interest, was recorded in 1982. The survey has baseline (1971) information on weight, sex, race, height, education, alcohol use, and intensity of smoking.

   The table below shows the distribution of 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 model based on all 9 covariates; no interactions were included in the propensity score model. Three propensity score methods were used to estimate the average treatment effect (the average difference in 1982 weight between the smoking cessation group and the smoking group): propensity score matching; stratifying on the propensity score; and regression adjustment using the propensity score. Unadjusted means were compared 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 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? Is there evidence of a treatment effect? Did using propensity score methods reduce the bias in estimating the average treatment effect? Briefly explain.

This is an observational study since the assignment mechanism (how subjects are assigned to the smoking or smoking cessation group) is unknown. The are two "treatments" in this study smoking and smoking cessation. In other words the treatment factor smoking has two levels: smoking; smoking cessation. There is evidence (at the 5% level) of a treatment effect since all of the confidence intervals for the average treatment effects exclude 0. All of the propensity score methods reduce the bias since each all three propensity score methods produced average treatment effects that are greater than the unadjusted treatment effects by at least 0.43Kg.

(b) Is the propensity score known or unknown in this study? Briefly describe how the propensity score was calculated for each subject from the logistic regression model.

The propensity score is unknown since the probability of a subject being assigned to the smoking/smoking cessation group is unknown. The propensity score can be estimated from the logistic regression model by calculating the predicted probability of smoking for each subject.

(c) 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?

The potential outcomes in this study are 1982 weight if the subject was a smoker and non-smoker, $Y(1), Y(0)$, respectively. The treatment assignment, T, is smoking status. The treatment assignment is strongly ignorable if conditional on the propensity score smoking status is independent of the unobserved potential outcomes. This can be written symbolically as:

$$Y(0), Y(1) \perp T | e(x).$$

It seems plausible if the 9 covariates are all the important confounding covariates and no important covariates are missing. Conversely, it might not be plausible if at least one important covariate is missing.

(d) The study authors claim that the observed weight difference between the smoking and smoking cessation groups must be caused by smoking cessation. A critic claims that the study did not measure an important covariate, and that the difference in this unobserved covariate is the real reason that weight differs in the smoking and smoking cessation groups. Is it plausible that if the unobserved covariate was measured and accounted for in the propensity score analysis then there would not be a significant ($\alpha = 0.05$) treatment effect? In other words, does the critic have a valid point? Briefly explain.

Yes, the critic has a valid point. Propensity score methods balance observed not unobserved covariates. If an important covariate is not balanced then it may lead to a difference between the groups that is not due to treatment but due to the unobserved covariate.

4. In order to determine the effect of diet on blood coagulation 18 animals were randomized to three diets: A, B, C. The table below gives coagulation times for blood samples drawn from 18 animals receiving three different diets A, B, and C.

	A	B	C
	60	65	71
	63	66	66
	59	67	68
	63	63	68
	62	64	67
	59	71	68
Treatment Average	61	66	68
Grand Average	64	64	64
Difference	-3	2	4

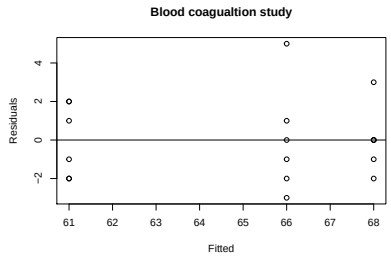
NB: There is a typo in the table above. The grand average should be 65 and the difference for each treatments A, B, C is then 61-65=-4, 66-65=1, 68-65=3.

The data was analyzed using R. Some of the output is shown below.

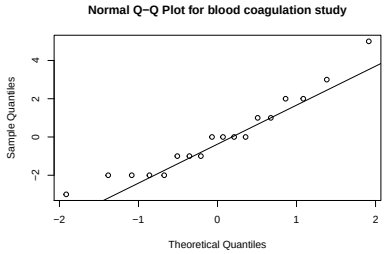
```
aov.diets <- aov(y~diets,data=dietdat)
summary(aov.diets)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)
diets          156
Residuals       72
---
```

```
plot(aov.diets$fitted.values,aov.diets$residuals,ylab="Residuals",
     xlab="Fitted",main="Blood coagulation study")
abline(h=0)
```




```
qqnorm(aov.diets$residuals, main="Normal Q-Q Plot for blood coagulation study")
qqline(aov.diets$residuals)
```



Answer the following questions based on the blood coagulation study.

(a) Formulate null and alternative hypotheses to compare the mean coagulation times between the three diets. (5 marks)

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad \text{versus} \quad H_1 : \mu_i \neq \mu_j,$$

$i, j = 1, 2, 3$, for at least one pair (i, j), where $i \neq j$. μ_1, μ_2, μ_3 are the treatment means for diets A, B, C respectively.

(b) Calculate MS_{treat}, MS_E, and the observed F statistic. What statistical assumptions are required for these calculations? (5 marks)

	Df	Sum Sq	Mean Sq	F value
diets	2	156	78.0	16.25
Residuals	15	72	4.8	

$MS_{Treat} = 78, MS_E = 4.8, F = 16.25$ require no statistical assumptions.

(c) Is there statistical evidence of a difference between the three diets at the 1% level? What statistical assumptions are required so that

$$\frac{MS_{Treat}}{MS_E}$$

follows an F distribution? Are these assumptions satisfied in the blood coagulation study? Briefly explain. (10 marks)

NB: The 1% critical value of the appropriate F distribution is 6.34. In other words, $P\left(F_{df_1,df_2} > 6.34\right) = 0.01$, where df_1, df_2 are the numerator and denominator degrees of freedom of the appropriate F distribution for this study.

The p-value is $P\left(F_{2,25} > 16.25\right) < 0.01$ since $16.25 > 6.34$. Therefore there is evidence of a difference between the diets at the 1% level.

The following three assumptions are required so that $\frac{MS_{Treat}}{MS_E} \sim F_{2,15}$.

1. Additive model.

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}.$$

The parameters τ_i are interpreted as the treatment effect of the i^{th} mean. That is, if μ_i is the mean of i^{th} group and μ is the overall mean then $\tau_i = \mu_i - \mu$.

This model seems plausible for this experiment since each observation y_{ij}, the jth observation under the ith treatment, is a single measurement obtained under treatment i with mean μ_i plus random error ϵ_{ij}.

2. If $\epsilon_{ij} \sim N(0, \sigma^2)$ then MS_{Treat} and MS_E are independent. Under the null hypothesis that $\sum_{i=1}^{a} \tau_i^2 = 0$ the ratio $F = \frac{MS_{Treat}}{MS_E}$ is the ratio of two independent estimates of σ^2. Therefore, $\frac{MS_{Treat}}{MS_E} \sim F_{a-1,N-a}$.

The constant variance assumption can be assessed by checking if the residual plot is a random scatter of points around 0. The plot exhibits this random scatter for this data.

The QQ plot indicates that the residuals are normally distributed since the points don't deviate systematically from the straight line.