STA305/1004 Winter 2016 - Homework #2 - Answers

Due Date: February 23, 2016

Due date: Electronic submission on UofT Learning Portal course page by Tuesday, February 23, 2016 at 22:00. NB: e-mail submissions will NOT be accepted.

- STA305 and STA1004 students are *not* required to complete the bonus questions 4. (a). If the bonus question is completed then it will be worth extra marks. You will not be penalized for trying the bonus question (i.e., you cannot lose marks)
- STA1004 students are required and STA305 students are *not* required to complete 4 (b).

If you work with other students on this assignment then:

- indicate the names of the students on your solutions;
- your solutions must be written up independently (i.e., your solutions should not be the same as another students solutions).

1. Suppose that you have run an experiment on 20 subjects, and have 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 $z = 2.1$ so the p-value=0.03. In order to confirm the results the researcher 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.

(a) 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. You may assume that the true mean in experiment #2 is the mean obtained in the experiment #1.

In experiment #1 $z = \frac{\bar{x}}{\frac{\sigma}{\sqrt{20}}} = 2.1 \Rightarrow \bar{x} = 2.1\frac{\sigma}{\sqrt{n}}$. So, in experiment #2 we will assume that (the true) $\mu = 2.1\frac{\sigma}{\sqrt{20}}$.

In experiment #2 we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu > 0$ if

$$\frac{\bar{x}}{\frac{\sigma}{\sqrt{10}}} \geq 1.645 \Rightarrow \bar{x} \geq 1.645\frac{\sigma}{\sqrt{10}}.$$

So, the probability that the test in experimet #2 is rejected when $\mu = 2.1\frac{\sigma}{\sqrt{20}}$ is

$$\begin{aligned}
P\left(\bar{x} \geq 1.645\frac{\sigma}{\sqrt{10}}\right) &= P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{10}} \geq \frac{1.645\frac{\sigma}{\sqrt{10}} - \mu}{\sigma/\sqrt{10}}\right) \\
&= P\left(Z \geq \frac{1.645\frac{\sigma}{\sqrt{10}} - 2.1\frac{\sigma}{\sqrt{20}}}{\sigma/\sqrt{10}}\right) \\
&= P\left(Z \geq 1.645 - 2.1\sqrt{10/20}\right) \\
&= P\left(Z \geq 1.645 - 2.1\sqrt{10/20}\right) \\
&= P\left(Z \geq 0.16\right)
\end{aligned}$$

```r
1-pnorm(0.16)
```

```
## [1] 0.4364405
```

Therefore, the probability that the results of experiment #2 will be significant (power of the study) is $P(Z \geq 0.16) = 0.4364405$.

(b) The researcher strongly believes in his theory, and is a bit depressed about the probability that he obtained in part (a); he feels it's too low for him to proceed to experiment #2. He wants the probability in part (a) to be higher. If he would like the probability to be $1-\beta$, where β is the probability of a type II error, show that the sample size formula is,

$$n_2 = \left[\frac{(1.645 + z_\beta)\sqrt{20}}{2.1}\right]^2,$$

where z_β is the $100 * (1-\beta)^{th}$ percentile of the $N(0,1)$.

In experiment #2 $H_0 : \mu = 0$ versus $H_1 : \mu = \mu_1$, where $\mu_1 = 2.1\frac{\sigma}{\sqrt{20}}$.

The test rejects if and only if

$$\frac{\bar{x}}{\sigma/\sqrt{n_2}} \geq 1.645.$$

Therefore, under H_1,

$$\begin{aligned}
1-\beta = P(\bar{x} \geq 1.645\sigma/\sqrt{n_2}) &= P\left(\frac{\bar{x}-\mu_1}{\sigma/\sqrt{n_2}} \geq \frac{1.645\sigma/\sqrt{n_2} - 2.1\sigma/\sqrt{20}}{\sigma/\sqrt{n_2}}\right) \\
&= P\left(Z \geq 1.645 - 2.1\frac{\sqrt{n_2}}{\sqrt{20}}\right) \\
&\Rightarrow 1 - \Phi\left(1.645 - 2.1\frac{\sqrt{n_2}}{\sqrt{20}}\right) = 1-\beta \\
&\Rightarrow \Phi\left(1.645 - 2.1\frac{\sqrt{n_2}}{\sqrt{20}}\right) = \beta \\
&\Rightarrow 1.645 - 2.1\frac{\sqrt{n_2}}{\sqrt{20}} = \Phi^{-1}(\beta) \\
&\Rightarrow n_2 = \left[\frac{\left(1.645 - \Phi^{-1}(\beta)\right)\sqrt{20}}{2.1}\right]^2.
\end{aligned}$$

Let z_β be the $100(1-\beta)$ percentile of the $N(0,1)$.

$$\Phi^{-1}(\beta) = \begin{cases} z_\beta & \text{if } \beta > 0.5 \\ -z_\beta & \text{if } \beta < 0.5 \end{cases}$$

`qnorm()` calculates $\Phi^{-1}(\beta)$.

For example, if $\beta = 0.2$ then $z_{0.2}$ is

```
qnorm(.2)
```

```
## [1] -0.8416212
```

If $\beta = 1 - .2 = 0.8$ then $z_{0.8}$ is

```
qnorm(0.8)
```

```
## [1] 0.8416212
```

In practice we are only intrested in sample sizes for $\beta < 0.5$. Therefore, $\Phi^{-1}(\beta) = -z_\beta$ so

$$n_2 = \left[\frac{(1.645 + z_\beta)\sqrt{20}}{2.1}\right]^2.$$

(c) Use the formula in part (b) to calculate the sample sizes required for $1 - \beta$ equal to 0.7, 0.8, and 0.9. (**Hand in your R code and output**)

```
(((1.645-qnorm(.3))*sqrt(20))/(2.1))^2 # 70% power
```

```
## [1] 21.34376
```

```
(((1.645-qnorm(.2))*sqrt(20))/(2.1))^2 # 80% power
```

```
## [1] 28.04211
```

```
(((1.645-qnorm(.1))*sqrt(20))/(2.1))^2 # 90% power
```

```
## [1] 38.8422
```

NB: Use `-qnorm(p)`, since `p=0.3,0.2,0.1`.

(d) If the researcher stuck to his original plan do you think he is likely to confirm his theory? Briefly explain.

It's very unlikely since the power of the test when $n = 10$ is only 44%. The calculations in part (c) show that at least 28 subjects are required for at least 80% power.

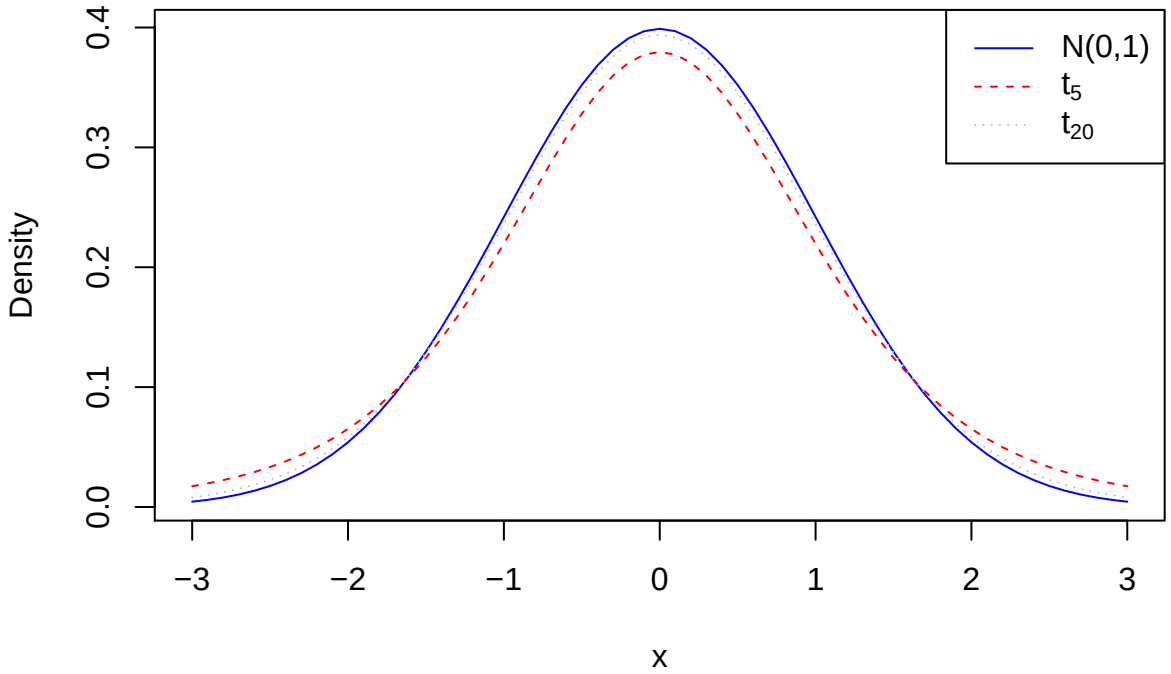
2. A group of medical researchers are interested in conducting a randomized clinical trial to determine whether acetazolamide reduces mechanical ventilation duration in critically ill patients with COPD and metabolic alkalosis. The treatments are Acetazolamide vs placebo, initiated within 48 hours of ICU (intensive care unit) admission and continued during the ICU stay for a maximum of 28 days. The primary outcome is the duration (in hours) of invasive mechanical ventilation via endotracheal intubation or tracheotomy.

Another randomized study found that the duration of invasive mechanical ventilation follows a non-normal symmetric distribution. The statistician on the team decided that the most appropriate distribution for duration of invasive mechanical ventilation via endotracheal intubation or tracheotomy is a location-scale t_m distribution - the t-distribution on m degrees of freedom. Let y_i be the duration of the i^{th} patient then

$$y_i = \mu + \sigma T,$$

where $T \sim t_m$. The mean of the distribution is $E(y_i) = \mu$. The variance of the distribution is $Var(y_i) = \sigma^2 \left(\frac{m}{m-2}\right)$, where m is the degrees of freedom. Through various plots and tests the statistician decided that the appropriate value of m (the degrees of freedom) is 5.

The density functions of the t_5, t_{20} and $N(0,1)$ distributions are shown in the plot below.



A random sample from a location-scale t_5 distribution with mean μ and scale σ can be generated in R using the command `mu+sigma*rt(n=,df=5)` where `n` is the number of observations and `df` is the degrees of freedom.

The statistician on the research team needs to calculate the power of the study to compare duration of invasive mechanical ventilation in the Acetazolamide and placebo groups at the 5% significance level. The team hypothesized that the placebo arm will have a mean duration of 27 hours and the Acetazolamide arm will have a mean duration of 24 hours ($\sigma = 10$ in both arms).

(a) The statistician decides to ignore that duration follows a location-scale t_5 distribution, but instead assumes that the duration is normally distributed in each arm. The researchers would like to have twice as many patients in the Acetazolamide arm. The plan is to enroll 300 subjects in the Acetazolamide arm. Calculate the power of the two-sample t-test using the 5% significance level. (**Hand in your R code and output**)

The researchers want to test if acetazolamide *reduces* duration of mechanical ventilation compared to a palcebo. Most clinical trials would be designed to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ instead of $H_1 : \mu_1 > \mu_2$, where μ_1, μ_2 are the mean duration times in the placebo and acetazolamide arms. However, both answers are acceptable since strictly speaking the alternative should be $H_1 : \mu_1 > \mu_2$, but it's difficult to interpret a one-sided test if the study produces evidence that $\mu_1 < \mu_2$.

$H_1 : \mu_1 \neq \mu_2$.

```
#use the pwr library
library(pwr)
pwr.t2n.test(n1 = 300,n2 = 150,d = (27-24)/10)
```

```
##
##      t test power calculation
##
##              n1 = 300
##              n2 = 150
##               d = 0.3
##       sig.level = 0.05
##           power = 0.8493385
##     alternative = two.sided
```

```
# use the power function given in class
pow.t <- function(theta){
  alpha <-0.05
  nA <- 300
  nB <- 150
  t.crit <-qt(1-alpha/2,nA+nB-2)
  t.gamma <- theta/(sqrt(1/nA+1/nB))
  t.power <- 1-pt(t.crit,nA+nB-2,ncp=t.gamma)+pt(-t.crit,nA+nB-2,ncp=t.gamma)
  return(t.power)
}

pow.t((27-24)/10)
```

```
## [1] 0.8493385
```

$H_1 : \mu_1 > \mu_2$.

```
#use the pwr library
library(pwr)
pwr.t2n.test(n1 = 300,n2 = 150,d = (27-24)/10,alternative = "less")
```

```
##
##      t test power calculation
##
##              n1 = 300
##              n2 = 150
##               d = 0.3
##       sig.level = 0.05
##           power = 0.9115901
##     alternative = less
```

```
# use the power function given in class
pow.t <- function(theta){
  alpha <-0.05
  nA <- 300
  nB <- 150
```

```
  t.crit <-qt(1-alpha,nA+nB-2) # replace alpha/2 with alpha.
  t.gamma <- theta/(sqrt(1/nA+1/nB))
  t.power <- 1-pt(t.crit,nA+nB-2,ncp=t.gamma)+pt(-t.crit,nA+nB-2,ncp=t.gamma)
  return(t.power)
}

pow.t((27-24)/10)
```

```
## [1] 0.9115918
```

(b) Calculate the power for the scenario in part (a) except now assume that duration in each arm follows a location-scale t_5 distribution. This means that in the Acetazolamide arm $y_i = 24 + 10T_i$ and $y_i = 27 + 10T_i$ in the placebo arm, where y_i is the duration of the i^{th} patient and $T \sim t_5$.

A random sample of observations in each arm can be simulated using `rt()`. These can be further used to simulate the power of the test under the assumption that the data in each arm follows a location-scale t_5. Ideally the number of simulations should be 250,000, but other large numbers are acceptable such as 10,000 used below.

```
degf <- 5
scale <- 10
mu1 <- 24
mu2 <- 27
pvals <- replicate(10000,t.test(mu1+scale*rt(300,degf),mu2+scale*rt(150,degf),var.equal = F)$p.value)
sum(pvals<=0.05)/10000
```

```
[1] 0.6421
```

The power is approximately 65%.

(c) Calculate the power using the scenario in part (b), but now assume that the duration in each arm follows a location-scale t_{20} distribution.

```
degf <- 20
scale <- 10
mu1 <- 24
mu2 <- 27
pvals <- replicate(10000,t.test(mu1+scale*rt(300,degf),mu2+scale*rt(150,degf),var.equal = F)$p.value)
sum(pvals<=0.05)/10000
```

```
[1] 0.8107
```

The power is approximately 80%.

(c) Explain why the power that you obtained in parts (b) and (c) is different than the power you obtained in part (a)?

The assumption of normality was violated in parts (b) and (c). The assumption in part (b) was more severly violated than in part (c). This is why the power in (c) is closer to the power in part (a).

3. A statistics professor is able to choose the best type of lecture style, traditional class or inverted class, for each student planning to enroll in a certain statistics course. The professor is able to perfectly predict students' course grades in both lecture styles. She chooses the lecture style for each student by choosing the lecture style where a student's final course grade will be the highest; if there is no difference then the statistics professor flips a coin. Let

$$T = \begin{cases} 1 & \text{if student is in inverted class} \\ 0 & \text{if student is in traditional class} \end{cases},$$

and $Y(0)$ represent the potential outcome after completing the traditional course, and, $Y(1)$ represent the potential outcome after completing the inverted class.

The data for seven students is shown in the following table.

Student	$Y(0)$	$Y(1)$
1	95	79
2	76	77
3	61	54
4	53	47
5	71	69
6	81	82
7	65	70
average	**72**	**68**

(a) What data would we actually observe under the statistics professor's assignment mechanism? Create a table with three columns: T; $Y(0)$; $Y(1)$, and calculate the observed averages.

This would be the observed data. A ? indicates that the data would not be observed.

Student	$Y(0)$	$Y(1)$	T
1	95	?	0
2	?	77	1
3	61	?	0
4	53	?	0
5	71	?	0
6	?	82	1
7	?	70	1

$\bar{Y}(0) = 70$, $\bar{Y}(1) = 76.33$.

(b) Is the assignment mechanism non-ignorable? Briefly explain.

The assignment mechanism is non-ignorable since it depends on the potential outcomes.

(c) Is it possible to draw the correct causal conclusion about the effectiveness of the inverted classroom method versus the traditional classroom method based on the observed data? Briefly explain.

No. The observed data indicates that inverted is better than traditional. But, the true values indicate that traditional is better than inverted.

(d) Suppose that the professor had randomly assigned three students to the traditional class and four students to the inverted class instead of using her awesome predictive powers. The observed data from her experiment is in the table below.

Student	Y^{obs}	T^{obs}
1	79	1
2	76	0
3	54	1
4	53	0
5	69	1
6	82	1
7	65	0

T^{obs} represents the class that a student was assigned, and Y^{obs} represents students' final grade. This is the same data as in the table above except that some of the potential outcomes are missing since the students were randomly assigned to one of the lecture styles.

Conduct a randomization test to investigate if grades in the traditional class are greater than grades in the inverted class. What have you assumed about the *observed* treatment assignment? Is this assumption valid? Does the randomization test lead to the correct causal inference? Briefly explain why it does or does not lead to the correct causal inference. (**Hand in your R code and output**)

```
y1 <- c(79,54,69,82)
y0 <- c(76,53,65)
y <- c(y1,y0) #pool data
N <- choose(7,4)
res <- numeric(N) # store the results
#install.packages("combinat") # if package not installed then remove comment
library(combinat)
index <-combn(1:7,4)
for (i in 1:N)
{
  res[i] <- mean(y[index[,i]])-mean(y[-index[,i]])
}

observed <- mean(y1)-mean(y0)
tbar <- mean(res)
pval <- sum(abs(res-tbar)>=abs(observed-tbar))/N
round(pval,2)
```

```
[1] 0.46
```

The randomization test H_0 is $Y^{obs} = Y(1) = Y(0)$. The randomization test does not provide evidence to reject this hypothesis. But, it does not lead to the correct causal inference since the randomization test is inconclusive (p-value=0.46). Moreover, the observed difference is $\bar{Y}(1) - \bar{Y}(0) = 6.3$. This implies that student's in the inverted lecture have an average final grade that is 6 points higher than student's in the traditional lecture section, although the difference is not statistically significant and is likely due to chance. Indeed, it must be due to chance since the truth is that students in the traditional section are 4 points higher than students in the inverted lecture section.

4. Consider a study comparing two treatments labelled 1 and 0. The propensity score is the conditional

probability that a unit with vector $\mathbf{x}$ of observed covariates will be assigned treatment 1

$$e(\mathbf{x}) = P\left(Z = 1|\mathbf{x}\right).$$

(a) (Bonus Question for all students - STA305 and STA1004) Show that $\mathbf{x}$ and Z are conditionally independent given $e = e(\mathbf{x})$,

$$P(\mathbf{x}, Z|e) = P(\mathbf{x}|e)P(Z|e).$$

This is called the balancing property of the propensity score. The balancing property says that treated ($Z = 1$) and control ($Z = 0$) subjects with the same propensity score $e(x)$ have the same distribution of the observed covariates x.

HINTS:

- First show that $P(\mathbf{x}, Z|e) = P(\mathbf{x}|e)P(Z|x, e)$.
- Recall the facts about conditional expectation: if X, Y are random variables, then $E(X) = E(E(X|Y))$.

We first show that $P(\mathbf{x}, Z|e) = P(\mathbf{x}|e)P(Z|x, e)$.

$$\begin{aligned} P(\mathbf{x}, Z|e) = \frac{P(\mathbf{x}, Z, e)}{P(e)} &= \frac{P(\mathbf{x}, Z, e)}{P(x, e)} \cdot \frac{P(\mathbf{x}, e)}{P(e)} \\ &= P(\mathbf{x}|e)P(Z|x, e) \end{aligned}$$

So, it suffices to show that $P(Z|\mathbf{x}, e) = P(Z|e)$. But, e is a function of $\mathbf{x}$ so $P(Z|\mathbf{x}, e) = P(Z|\mathbf{x})$ so it suffices to show $P(Z|\mathbf{x}) = P(Z|e)$.

By definition $e(\mathbf{x}) = P(Z|\mathbf{x})$. But,

$$\begin{aligned} P(Z|e) &= E(Z|e), \quad \text{since } Z \text{ is an indicator random variable} \\ &= E(E(Z|e)|e) \\ &= E(e|e) = e. \end{aligned}$$

(b) (STA1004 Additional Question/ STA305 Bonus) Explain why randomization is a much more powerful tool for balancing covariates than matching on an estimate of the propensity score.

Randomization balances both observed and unobserved covariates. Matching on the propesnity score balances only observed covariates.