140.652 Problem Set 5 Solutions

Problem 1

A laboratory experiment found that in a random sample of 20 frog eggs having aquaporins, 17 exploded when put into water.

a. Plot and interpret the posteriors for p assuming a beta prior with parameters $(2,2)$, $(1,1)$ and $(.5,.5)$.

Suppose $p \sim Beta(\alpha, \beta)$ and our data $X \sim Binom(n,p)$. Then, the posterior is given by,

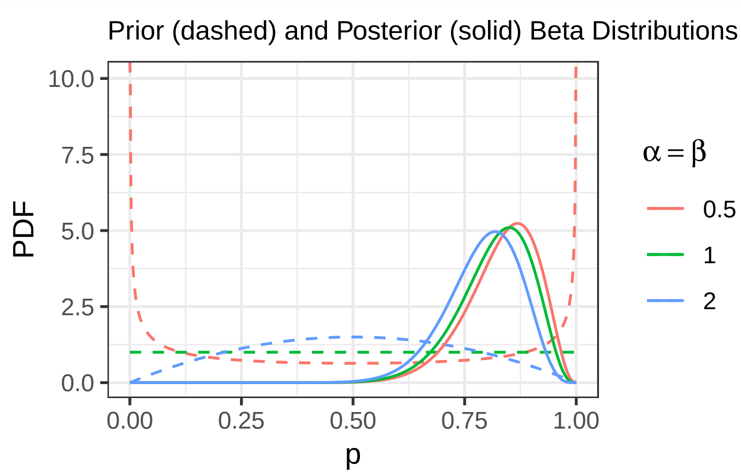
$$\begin{aligned} f(p|\alpha,\beta,X) &= \text{Prior} \times \text{Likelihood} = f(p|\alpha,\beta) \cdot f(X|p) \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \cdot \binom{n}{X} p^X (1-p)^{n-X} \\ &\propto p^{\alpha+X-1}(1-p)^{(n-X)+\beta-1} \end{aligned}$$

We see that the posterior follows a Beta distribution with parameters $\alpha' = \alpha + X$ and $\beta' = (n-X)+\beta$.

```r
n <- 20
x <- 17
p <- seq(0, 1, length = 1000)

# Calculate priors and posterior
beta.params <- list(c(2, 2), c(1, 1), c(0.5, 0.5))
priors <- lapply(beta.params, function(param, p)
  dbeta(p, param[1], param[2]), p = p)
posteriors <- lapply(beta.params, function(param, p)
  dbeta(p, param[1]+ x, n - x + param[2]), p = p)

# Translate into df and plot
data.frame(p = rep(p, 3), param = rep(c(2, 1, 0.5), each = 1000),
           prior = unlist(priors), posterior = unlist(posteriors)) %>%
  ggplot(aes(x = p, color = factor(param))) +
  geom_line(aes(y = prior), linetype = 2) +
  geom_line(aes(y = posterior)) +
  labs(title = "Prior (dashed) and Posterior (solid) Beta Distributions",
       x = "p", y = "PDF", color = TeX('$\\alpha = \\beta$')) +
  theme_bw() + theme(plot.title = element_text(size=10))
```



The plot above shows the prior and posterior distribution. Note that when $\alpha = \beta = 1$, the distribution follows a uniform distribution, and the posterior is thus just the likelihood. We can see that the posterior pools information from both the prior and the likelihood to derive a probability distribution for the value of p.

b. Calculate and interpret the credible interval for each of the beta prior parameter esttings. Note that the R package `binom` may be of use.

For each of the credible intervals below, the probability that p is within the interval is 95%.

```
library(binom)

# 95% credible interval for a Beta(2,2) prior
binom.bayes(x, n, prior.shape1 = 2, prior.shape2 = 2)[c("lower", "upper")]
```

```
      lower     upper
1 0.6314496 0.9378542
```

```
# 95% credible interval for a Beta(1,1) prior
binom.bayes(x, n, prior.shape1 = 1, prior.shape2 = 1)[c("lower", "upper")]
```

```
      lower     upper
1 0.6599475 0.9591231
```

```
# 95% credible interval for a Beta(0.5,0.5) prior
binom.bayes(x, n, prior.shape1 = 0.5, prior.shape2 = 0.5)[c("lower", "upper")]
```

```
      lower     upper
1 0.6773164 0.9698094
```

Problem 2

A study of blood alcohol levels (mg/100 ml) at post mortem examination from traffic accident victims involved taking one blood sample from the leg, A, and another from the heart, B. The results were:

Case	A	B	Case	A	B
1	44	44	11	265	277
2	265	269	12	27	39
3	250	256	13	68	84
4	153	154	14	230	228
5	88	83	15	180	187
6	180	185	16	149	155
7	35	36	17	286	290
8	494	502	18	72	80
9	249	249	19	39	50
10	204	208	20	272	271

Test whether or not the mean blood alcohol level differs between the heart and the leg. Give the appropriate null and alternative hypotheses. Give the relevant P-value. Interpret your results, state your assumptions.

Assume that A and B are normally distributed with equal variance, and let μ_A and μ_B denote the means of this distribution. We want to test,

$$H_0 : \mu_A = \mu_B$$
$$H_A : \mu_A \neq \mu_B$$

Using the `t.test()` function in `R`, we get a p-value of 0.0007. With $\alpha = 0.05$, we reject the null and conclude that the blood alcohol level may be different between samples from the leg and samples from the heart.

```r
A <- c(44,265,250,153, 88,180, 35,494,249,204,
       265, 27, 68,230,180,149,286, 72, 39,272)
B <- c(44,269,256,154, 83,185, 36,502,249,208,
       277, 39, 84,228,187,155,290, 80, 50,271)
diff = A - B

# T-test
t.test(diff, var.equal = TRUE)
```

```
    One Sample t-test

data:  diff
t = -4.0367, df = 19, p-value = 0.0007046
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -7.364725 -2.335275
sample estimates:
mean of x
    -4.85
```

Problem 3

Forced expiratory volume FEV is a standard measure of pulmonary function. We would expect that any reasonable measure of pulmonary function would reflect the fact that a person's pulmonary function declines with age after age 20. Suppose we test this hypothesis by looking at 10 nonsmoking males ages 35-39, heights 68-72 inches and measure their FEV initially and then once again 2 years later. We obtain this data.

Person	Year 0 FEV (L)	Year 2 FEV (L)	Person	Year 0 FEV (L)	Year 2 FEV (L)
1	3.22	2.95	6	3.25	3.20
2	4.06	3.75	7	4.20	3.90
3	3.85	4.00	8	3.05	2.76
4	3.50	3.42	9	2.86	2.75
5	2.80	2.77	10	3.50	3.32

a. Preform and interpret the relevant test. Give the appropriate null and alternative hypotheses. Intepret your results, state your assumptions and give a P-value.

Let D_i denote the difference in FEV between Year 2 and Year 0 for individual i, and assume that $D_i \overset{iid}{\sim} N(\mu_D, \sigma^2)$. Then, we want to test,

$$H_0 : \mu_D = 0$$
$$H_A : \mu_D \neq 0$$

Since n is small, we can use the t-test to test our hypothesis.

```r
fev0 <- c(3.22, 4.06, 3.85, 3.50, 2.80,
          3.25, 4.20, 3.05, 2.86, 3.50)
fev2 <- c(2.95, 3.75, 4.00, 3.42, 2.77,
          3.20, 3.90, 2.76, 2.75, 3.32)
```

```r
diff <- fev2 - fev0
t.test(diff)
```

```
    One Sample t-test

data:  diff
t = -3.0891, df = 9, p-value = 0.01295
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.25465006 -0.03934994
sample estimates:
mean of x
   -0.147
```

For $\alpha = 0.05$, we reject the null and conclude that there may be a change in FEV between Year 2 and Year 0 for individuals in this population. The 95% confidence interval is given by $(-0.25, -0.04)$ which does not contain 0, further verifying that the difference in FEV between the two years is not zero.

b. A large test comparing the two-year decline in non-smokers of a different age. Perform a sample size calculation to detect a change in FEV over two years at least as large as that detected for males age 35-39. Use the data above a for any relevant constants that you might need.

Suppose we want a two-sided test for the change in FEV with 80% power and $\alpha = 0.05$. To determine the sample size necessary, we can use the `power.t.test()` function in the `binom` package. We can estimate the mean and standard deviation of the 2-year FEV difference for non-smokers using the sample mean and sample variance.

```r
power.t.test(delta = mean(diff), sd = sd(diff),
             power = 0.8, sig.level = 0.05,
             alternative = "two.sided", type = "one.sample")
```

```
     One-sample t test power calculation

              n = 10.30858
          delta = 0.147
             sd = 0.1504844
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

Since n is integer valued, we find we need a sample size of 11 individuals to achieve the desired power for an $\alpha = 0.05$ test.

Note: Alternatively, you can treat the data as paired data to get the same results since the FEV at Year 0 is matched with the FEV at Year 2 for a given individual. However, the data is *not* of two unpaired samples.

Problem 4

Another aspect of the preceding study involves looking at the effect of smoking on baseline pulmonary function and on change in pulmonary function over time. We must be careful since FEV depends on many factors, particularly age and height. Suppose we have a comparable group of 15 men in the same age and height group who are smokers and we measure their FEV at year 0. The data are given (For purposes of this exercise assume equal variance where appropriate).

Person	FEV Year 0 (L)	FEV Year 2 (L)	Person	FEV Year 0 (L)	FEV Year 2 (L)
1	2.85	2.88	9	2.76	3.02
2	3.32	3.40	10	3.00	3.08
3	3.01	3.02	11	3.26	3.00
4	2.95	2.84	12	2.84	3.40
5	2.78	2.75	13	2.50	2.59
6	2.86	3.20	14	3.59	3.29
7	2.78	2.96	15	3.30	3.32
8	2.90	2.74			

Test the hypothesis that the change in FEV is equivalent between non-smokers and smokers. State relevant assumptions and interpret your result. Give the relevant P-value.

Let $D_{S,i}$ and $D_{NS,j}$ be the 2-year change in FEV for individuals $i = 1, 2, \cdots 10$ and $j = 1, 2, \cdots 15$. Assume that $D_{S,i} \stackrel{iid}{\sim} N(\mu_S, \sigma^2)$ and $D_{DS,j} \stackrel{iid}{\sim} N(\mu_{NS}, \sigma^2)$. We want to test the hypothesis,

$$H_0 : \mu_S = \mu_{NS}$$
$$H_A : \mu_S \neq \mu_{NS}$$

```
fev1smoker <- c(2.85,3.32,3.01,2.95,2.78,2.86,2.78,2.90,
                2.76,3.00,3.26,2.84,2.50,3.59,3.30)
fev2smoker <- c(2.88,3.40,3.02,2.84,2.75,3.20,2.96,2.74,
                3.02,3.08,3.00,3.40,2.59,3.29,3.32)
diffsmoker <- fev2smoker - fev1smoker

t.test(diff, diffsmoker, var.equal = TRUE)
```

```

    Two Sample t-test

data:  diff and diffsmoker
t = -2.4589, df = 23, p-value = 0.02188
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.36764246 -0.03169087
sample estimates:
  mean of x   mean of y
-0.14700000  0.05266667
```

With $\alpha = 0.05$ and a p-value of 0.0219, we reject the null and conclude that there may be a difference in 2-year FEV changes between smokers and non-smokers.

Problem 5

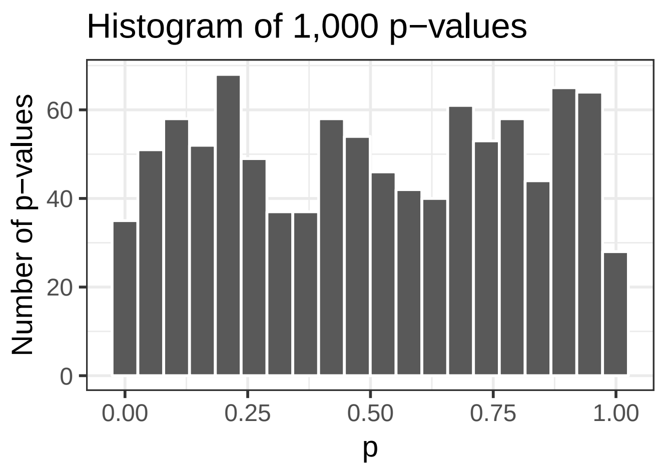
Perform the following simulation. Randomly simulate 1,000 sample means of size 16 from a normal distribution with means 5 and variances 1. Calculate 1,000 test statistics for a test of $H_0 : \mu = 5$ versus $H_a : \mu < 5$. Using these test statistics calculate 1,000 P-values for this test. Plot a histogram of the P-values. Note, this exercise demonstrates the interesting fact that the distribution of P-values is uniform.

```
x <- matrix(rnorm(1000 * 16, mean = 5, sd = 1), nrow = 1000)
means <- apply(x, 1, mean)
```

```r
sds   <- apply(x, 1, sd)

# Create t-statistics
t.stats <- (means - 5)/(sds/sqrt(16))
p.values <- pt(t.stats, df = 16 - 1)

# Plots
ggplot(data.frame(p.values = p.values), aes(x = p.values)) +
  geom_histogram(bins = 20, color = 'white') +
  labs(title = "Histogram of 1,000 p-values",
       x = "p", y = "Number of p-values") +
  theme_bw()
```



Problem 6

Suppose that systolic blood pressures were taken on 16 oral contraceptive users and 16 controls at baseline and again then two years later. The average difference from follow-up SBP to the baseline (followup - baseline) was 11 $mmHg$ for oral contraceptive users and 4 $mmHg$ for controls. The corresponding standard deviations of the differences was 20 $mmHg$ for OC users and 28 $mmHg$ for controls.

a. Calculate and interpret a 95% confidence interval for the **relative** change in systolic blood pressure for oral contraceptive users; assume normality on the log scale.

Let $D_{OC,1}, \cdots, D_{OC,16}$ be the difference in SBP for oral contraceptive users and $D_{C,1}, \cdots, D_{C,16}$ be the difference in SBP for controls. By the central limit theorem, we know that

$$\bar{D}_{OC} \xrightarrow{D} N(\mu_{OC}, \sigma^2_{OC})$$
$$\bar{D}_{C} \xrightarrow{D} N(\mu_{C}, \sigma^2_{C})$$

Since we are interested in a **relative** change, e.g $\frac{\mu_{OC}}{\mu_C}$, we can consider a log transformation. Using the delta method, we know that

$$\sqrt{16} \cdot \log(\bar{D}_{OC}) \xrightarrow{D} N\left(\log(\mu_{OC}), \frac{\sigma^2_{OC}}{\mu^2_{OC}}\right)$$
$$\sqrt{16} \cdot \log(\bar{D}_{C}) \xrightarrow{D} N\left(\log(\mu_{C}), \frac{\sigma^2_{C}}{\mu^2_{C}}\right)$$

which implies that

$$\log(\bar{D}_{OC}) \xrightarrow{D} N\left(\frac{\log(\mu_{OC})}{\sqrt{16}}, \frac{\sigma^2_{OC}}{16 \cdot \mu^2_{OC}}\right)$$
$$\log(\bar{D}_{C}) \xrightarrow{D} N\left(\frac{\log(\mu_{C})}{\sqrt{16}}, \frac{\sigma^2_{C}}{16 \cdot \mu^2_{C}}\right)$$

and assuming independence of the two groups,

$$Var\left(\log\left(\frac{\bar{D}_{OC}}{\bar{D}_C}\right)\right) = Var\left(\log \bar{D}_{OC} - \log \bar{D}_C\right) = Var(\log \bar{D}_{OC}) + Var(\log \bar{D}_C) = \frac{\sigma^2_{OC}}{16 \cdot \mu^2_{OC}} + \frac{\sigma^2_{C}}{16 \cdot \mu^2_{C}}$$

Thus, the 95% confidence interval for the difference in the log-mean SBP changes, assuming normality, is given by,

$$(\log \bar{D}_{OC} - \log \bar{D}_C) \pm Z_{0.975}\sqrt{\frac{\sigma^2_{OC}}{16 \cdot \mu^2_{OC}} + \frac{\sigma^2_{C}}{16 \cdot \mu^2_{C}}}$$

We can estimate $\mu_{OC}, \sigma^2_{OC}, \mu_C, \sigma^2_C$ using sample estimates. Thus, the 95% confidence interval is given by,

```
mu.OC <- 11
mu.C  <- 4
sig.OC <- 20
sig.C  <- 28

exp(log(mu.OC) - log(mu.C) + c(-1, 1) * qnorm(0.975) * sqrt(sig.OC^2/(16*mu.OC^2) + sig.C^2/(16*mu.C^2)))
```

```
[1]  0.07948808 95.14004677
```

The 95% confidence interval for the relative difference is then given by the exponentiated confidence interval. We can interpret the exponentiated interval the interval that will contain the true relative SBP difference between oral contraceptive users and controls 95% of the time in repeated trials. It is important to note that the sample size is small in this case, so asymptotics may not be appropriate.

b. Does the change in SBP over the two year period appear to differ between oral contraceptive users and controls? Perform the relevant hypothesis test and interpret. Give a P-value. Assume normality and a common variance.

Suppose we are interested in testing the following hypothesis:

$$H_0 : \mu_{OC} = \mu_C$$
$$H_A : \mu_{OC} \neq \mu_C$$

We can use a two-sided t-test to test our hypothesis. The p-value is thus given by,

```
sd.pooled <- sqrt((sig.OC^2*(16 - 1) + sig.C^2*(16 - 1))/(16 + 16 - 2))

# Calculate t -statistic
t.stat <- (mu.OC - mu.C)/(sd.pooled*sqrt(1/16 + 1/16))

# Calculate p-value for two-sided test
2* pt(t.stat, df = 16 + 16 - 2, lower.tail = FALSE)
```

```
[1] 0.4222096
```

With $\alpha = 0.05$, we fail to reject the null and conclude that the change in SBP may not differ between OC users and controls. This makes sense as our confidence interval for the relative SBP difference contains 1, and the relative SBP difference equals to 1 when there is no difference in change in SBP between the two groups.

Problem 7

Will a Student's T or Z hypothesis test for a mean with the data recorded in pounds always agree with the same test conducted on the same data recorded in kilograms? (explain)

Yes, the tests will agree because the t-statistic/z-statistic are unitless. To see this, let X_i be iid measurements in pounds and Y_i be iid measurements in kilograms. Then,

1. The mean of X and Y have units of pounds and kilograms, respectively – $\bar{X}$ is the average number of pounds measured and $\bar{Y}$ is the average number of kilograms measured.
2. The standard deviation of X and Y have units of pounds and kilograms.

The test statistic is typically of the form,

$$\frac{\text{Estimator} - \text{True Value}}{\text{Standard Error}}$$

And thus we see that the units cancel out.

For a concrete example, suppose we are interested in performing a Z-test, and let X have true mean and variance μ_X and σ_X^2, respectively. Let Y denote the measurement in kilograms. Then, we know that $Y = 0.45X$, so $\mu_Y = 0.45\mu_X$ and $\sigma_Y^2 = (0.45)^2\sigma_X^2$. Then, the Z statistic for Y is,

$$Z_Y = \frac{\bar{Y} - \mu_Y}{\sigma_Y/\sqrt{n}} = \frac{0.45\bar{X} - 0.45\mu_X}{0.45\sigma_X/\sqrt{n}} = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} = Z_X$$

which is the Z-statistic for X. Similarly, you can show that the t-statistics are equal.

Problem 8

A researcher consulting you is very concerned about falsely rejecting her null hypothesis. As a result the researcher decides to increase the sample size of her study. Would you have anything to say? (explain).

α gives you the probability that you falsely rejects the null hypothesis, so if she is concerned about falsely rejecting her null hypothesis, she should lower alpha. Increasing the sample size of the study generally will make the study more powerful to detect a difference if one truly exists, but will not shield against false rejection.

Problem 9

Researchers studying brain volume found that in a random sample of 16 sixty five year old subjects with Alzheimer's disease, the average loss in grey matter volume as a person aged four years was .1 mm^3 with a standard deviation of .04 mm^3.

a. Calculate and interpret a P-value for the hypothesis that there is no loss in grey matter volumes as people age. Show your work.

Let μ_G be the true average loss in grey matter volume and assume that the average loss in grey matter volume is normally distributed. We want to test the hypothesis,

$$H_0 : \mu_G = 0$$
$$H_A : \mu_G \neq 0$$

Since the sample size is small, we can use a two-sided t-test.

```r
# Calculate t-statistic
t.stat <- 0.1/(0.04/sqrt(16))

# Find p-value for two-sided test
2 * pt(t.stat, df = 16 - 1, lower.tail = FALSE)
```

```
[1] 4.996898e-08
```

With $\alpha = 0.05$, we reject the null and conclude that there may be a loss in grey matter volumes as people age.

b. The researchers would now like to plan a similar study in 100 healthy adults to detect a four year mean loss of .01 mm^3. Motivate a general formula for power calculations in this setting and calculate the power for a test with $\alpha = .05$? Assume that the variation in grey matter loss will be similar to that estimated in the Alzheimer's study.

From your notes (Lec. 16, slide 10), the general formula for power is given by,

$$\begin{aligned}\text{Power} &= P(\text{Probability of Rejecting the Null}|H_A) \\ &= P(T > t_{n-1,1-\alpha}|H_A) \\ &= P\left(Z + \frac{\sqrt{n}\cdot(\mu_a - \mu_0)}{\sigma} > \frac{t_{n-1,1-\alpha}}{\sqrt{n-1}}\sqrt{\chi^2_{n-1}}\right)\end{aligned}$$

where μ_a is a value in the alternative hypothesis and μ_0 is the supposed value of μ in the null hypothesis. Using the `power.t.test()` function in `R`, we get that the power for this setting is,

```r
power.t.test(n= 100, delta = 0.01, sd = 0.04, sig.level = 0.05,
             type = "one.sample", alternative = "two.sided")$power
```

```
[1] 0.6969757
```

Problem 10

A recent Daily Planet article reported on a study of a two week weight loss program. The study reported a 95% confidence interval for weight loss from baseline of [2 lbs, 6 lbs]. (There was no control group, all subjects were on the weight loss program.) The exact sample size was not given, though it was known to be over 200.

a. What can be said of a $\alpha = 5\%$ hypothesis test of whether or not there was any weight change from baseline? Can you determine the result of a $\alpha = 10\%$ test without any additional calculation or information? (explain your answer)

Recall that the definition of a p-value is the probability of seeing a test statistic as or more extreme than the one you obseve, given your null hypothesis.

If you reject the null hypothesis for $\alpha = 0.05$, then the probability of seeing a test statistic as or more extreme than the one you observe is less than 0.05. It follows immediately that this probability is also less than 0.1, leading us to reject the null hypothesis for $\alpha = 0.1$ whenever we reject for $\alpha = 0.05$.

Problem 11

Suppose that 18 obese subjects were randomized, 9 each, to a new diet pill and a placebo. Subjects' body mass indices (BMIs) were measured at a baseline and again after having received the treatment or placebo for four weeks. The average difference from follow-up to the baseline (followup - baseline) was $-3\ kg/m^2$ for the treated group and $1\ kg/m^2$ for the placebo group. The corresponding standard deviations of the differences was 1.5 kg/m^2 for the treatment group and 1.8 kg/m^2 for the placebo group. Does the change in BMI over the two year period appear to differ between the treated and placebo groups? Perform the relevant test and interpret. Give a P-value. Assume normality and a common variance.

Let $B_{D,i}$ and $B_{P,i}$ for $i = 1, \cdots, 9$ be the four-week difference in BMI of the diet pill and control groups, respectively. Suppose $B_{D,i} \overset{iid}{\sim} N(\mu_D, \sigma^2)$ and $B_{P,i} \overset{iid}{\sim} N(\mu_P, \sigma^2)$. We want to test the hypothesis,

$$H_0 : \mu_D = \mu_P$$
$$H_A : \mu_D \neq \mu_P$$

We can test the null-hypothesis using a t-test.

```r
bbar.D <- -3
bbar.P <- 1
sd.D <- 1.5
sd.P <- 1.8

# Calculate pooled variance and t-statistic
pooled.sd <- sqrt((1.5^2*(9-1) + 1.8^2*(9-1))/(9 + 9 - 2))
tstat <- (bbar.P - bbar.D)/(pooled.sd*sqrt(1/9 + 1/9))

# Calculate two-sided p-value
2*pt(tstat, df = 9 + 9 - 2, lower.tail = FALSE)
```

```
[1] 0.0001025174
```

Thus, with $\alpha = 0.5$, we reject the null and conclude that there may be difference in BMI change between the treatement and control groups.