

Methods in Biostatistics

Luchao Qi

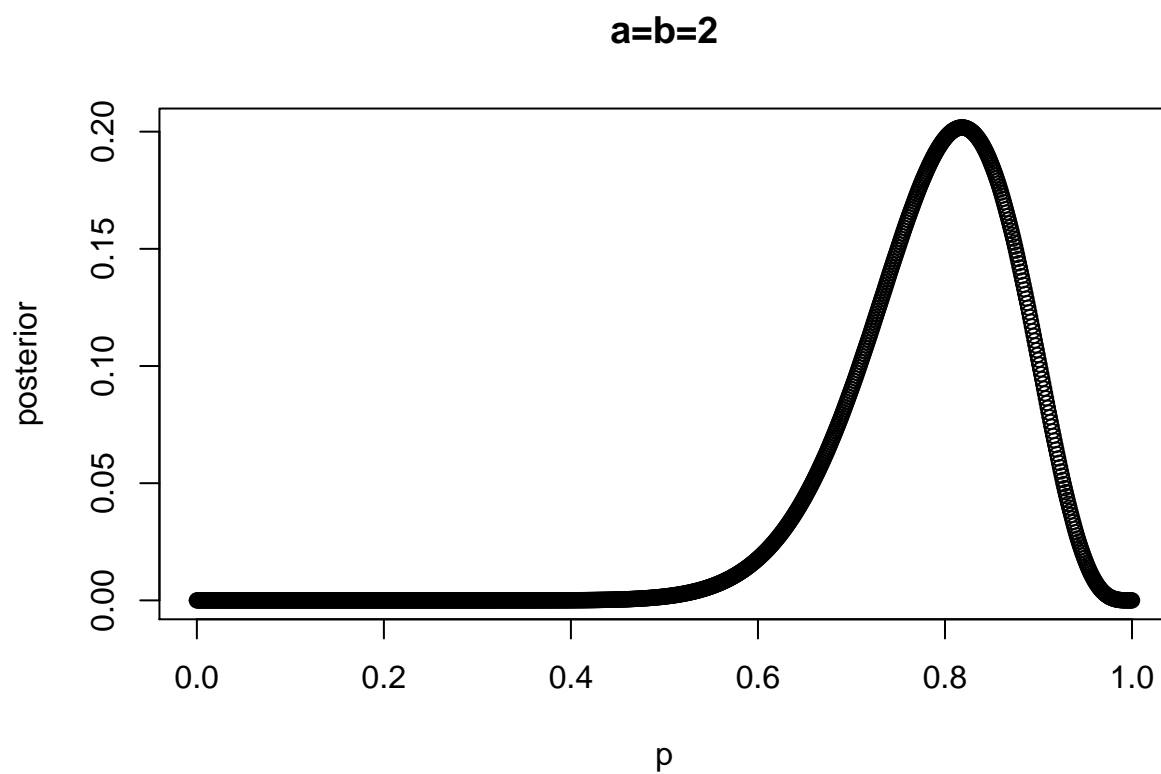
November 23, 2018

HW5

Problem 1

- MLE for p is 0.818, 0.85 and 0.868 with (2,2), (1,1) and (.5,.5), respectively.
The posterior mean is a mixture of the MLE mean estimator and the prior mean
- CI is [0.6314496 0.9378542], [0.6599475 0.9591231] and [0.6773164 0.9698094], respectively. Meaning in interval's lower limit to upper limit was constructed such that in repeated independent experiments, 95% of the intervals obtained would contain p .

```
#1
library(binom)
p=seq(0,1,length=1001)
a=b=2
like=choose(20,17)*p^17*(1-p)^3
prior=gamma(a+b)/(gamma(a)*gamma(b))*p^(a-1)*(1-p)^(b-1)
posterior=prior*like
plot(p,posterior,main="a=b=2")
```



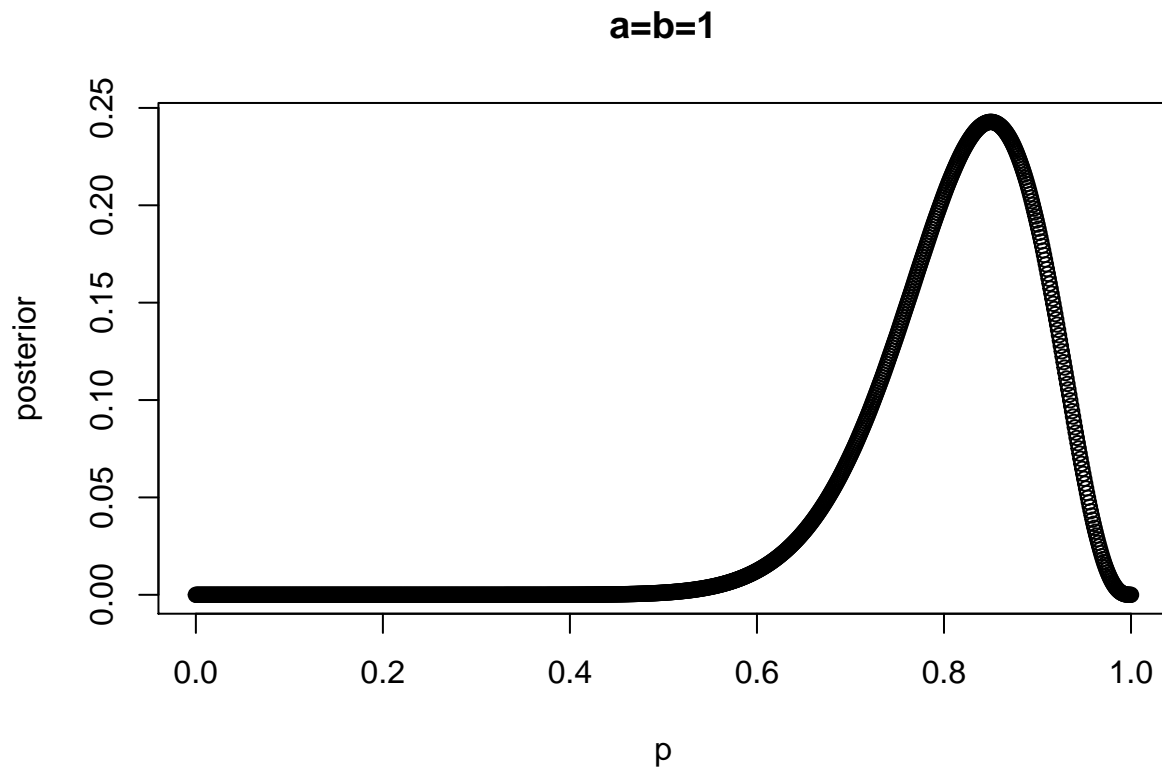
```
p[which.max(posterior)]
```

```
## [1] 0.818
```

```
binom.bayes(17, 20, type = "highest",prior.shape1 = 2,prior.shape2 = 2)
```

```
## method x n shape1 shape2 mean lower upper sig
## 1 bayes 17 20 19 5 0.7916667 0.6314496 0.9378542 0.05
```

```
a=b=1
like=choose(20,17)*p^17*(1-p)^3
prior=gamma(a+b)/(gamma(a)*gamma(b))*p^(a-1)*(1-p)^(b-1)
posterior=like*prior
plot(p,posterior,main="a=b=1")
```



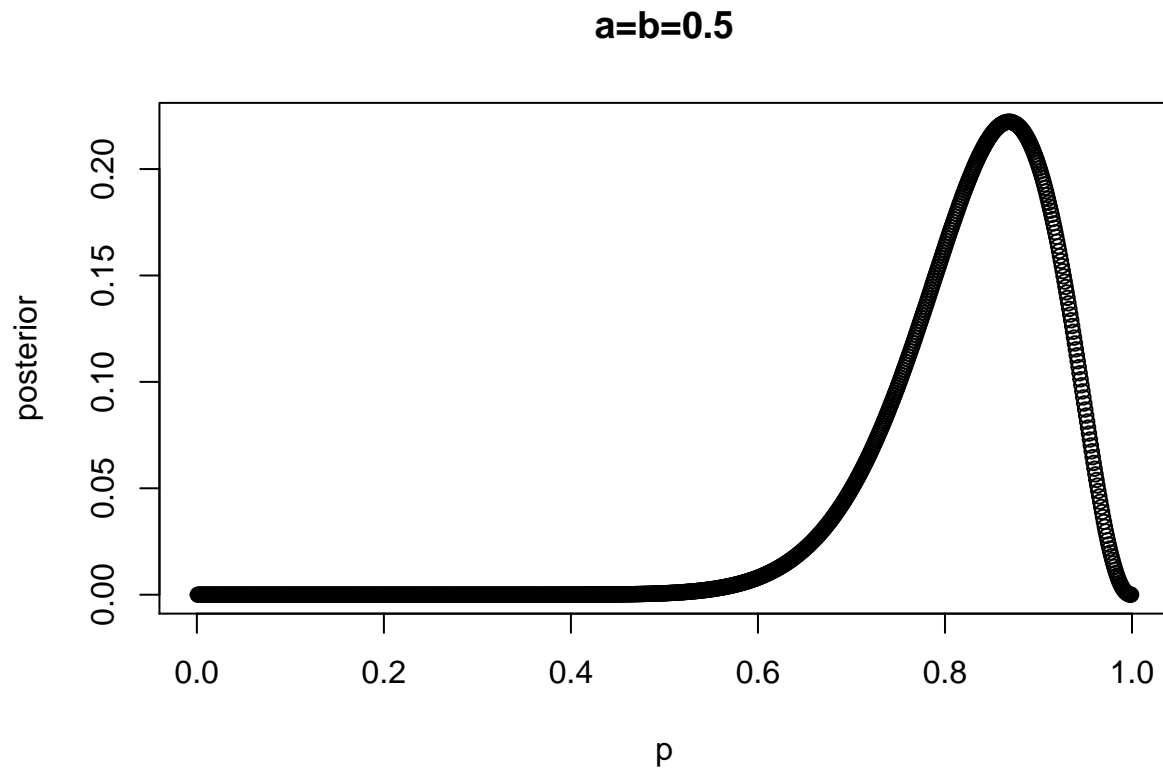
```
p[which.max(posterior)]
```

```
## [1] 0.85
```

```
binom.bayes(17, 20, type = "highest",prior.shape1 = 1,prior.shape2 = 1)
```

```
## method x n shape1 shape2 mean lower upper sig
## 1 bayes 17 20 18 4 0.8181818 0.6599475 0.9591231 0.05
```

```
a=b=0.5
like=choose(20,17)*p^17*(1-p)^3
prior=gamma(a+b)/(gamma(a)*gamma(b))*p^(a-1)*(1-p)^(b-1)
posterior=like*prior
plot(p,posterior,main="a=b=0.5")
```



```
p[which.max(posterior)]
```

```
## [1] 0.868
```

```
binom.bayes(17, 20, type = "highest", prior.shape1 = 0.5, prior.shape2 = 0.5)
```

```
## method x n shape1 shape2 mean lower upper sig
## 1 bayes 17 20 17.5 3.5 0.8333333 0.6773164 0.9698094 0.05
```

Or we could use the code from Lab:

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
library(latex2exp)
```

```
n <- 20
```

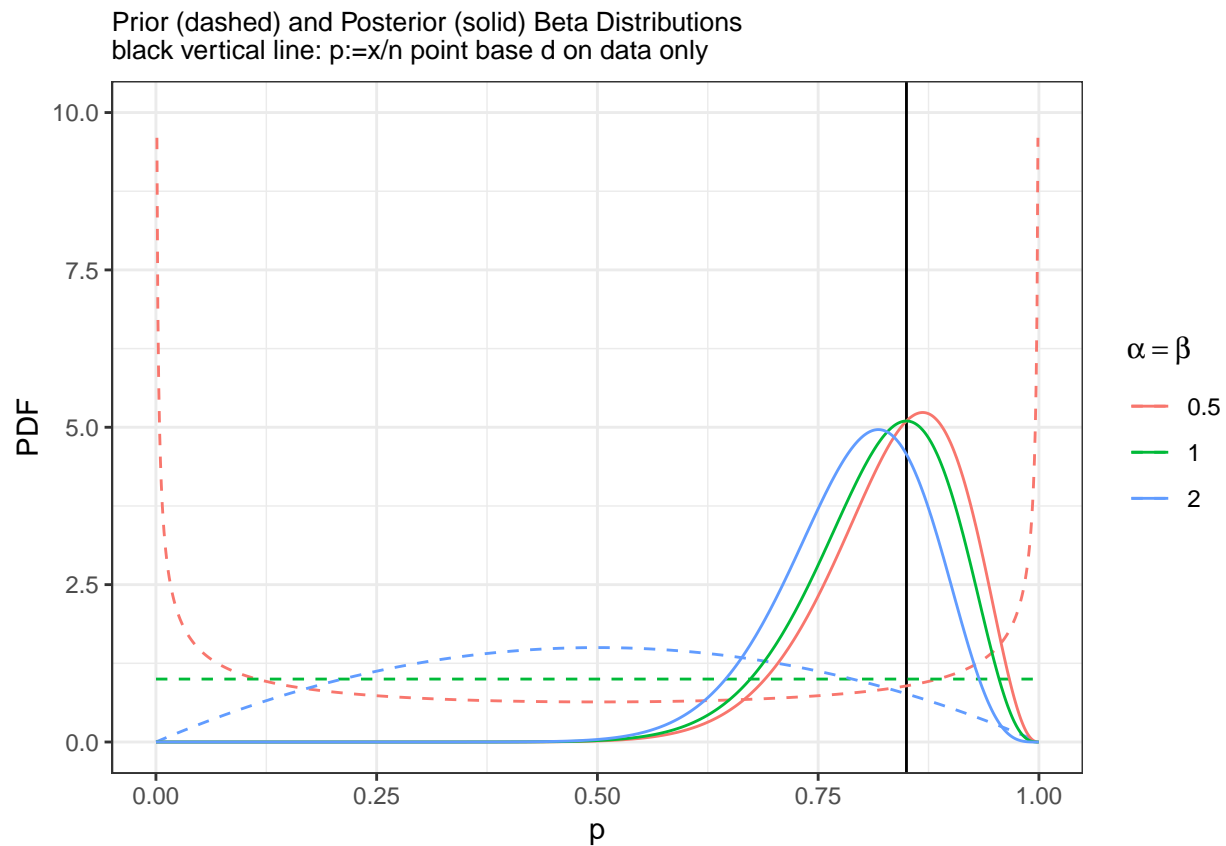
```
x <- 17
```

```
## Grid of p values
```

```
p <- seq(0.0001, 1 - 0.0001, length = 1000)
```

```
## List of prior distribution parameter
beta.params <- list(c(2, 2), c(1, 1), c(0.5, 0.5))
## lapply: apply over each element of list `beta.params` ## here: param - vector from `beta.params` list
priors <- lapply(beta.params, function(param, p)
  dbeta(p, param[1], param[2]), p = p)
## Calculate posteriors for different p
posteriors <- lapply(beta.params, function(param, p)
  dbeta(p, param[1] + x, n - x + param[2]), p = p)
# Translate into data.frame and plot
data.frame(p = rep(p, 3), param = rep(c(2, 1, 0.5), each = 1000), prior = unlist(priors))

## Warning: Removed 2 rows containing missing values (geom_path).
```



Problem 2

Null: true difference in mean equals to 0

Alternative hypothesis: true difference in means is not equal to 0

Assume normality and common variance

p-value = 0.0007046 < 0.05

So we reject null hypothesis

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

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

Problem 3

- a. Null: true difference equals to 0 $H_0 = 0$ Alternative hypothesis: true difference in means is not equal to 0 $H_A \neq 0$ Assume normality and common variance p-value = 0.01295 < 0.05
So we reject null hypothesis

```
x=c(3.22,4.06,3.85,3.5,2.8,3.25,4.2,3.05,2.86,3.5)
y=c(2.95,3.75,4.0,3.42,2.77,3.2,3.9,2.76,2.75,3.32)
dif1=(y-x)
t.test(dif1)
```

```
##
## One Sample t-test
##
## data: dif1
## 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
```

b.

```
x=c(3.22,4.06,3.85,3.5,2.8,3.25,4.2,3.05,2.86,3.5)
y=c(2.95,3.75,4.0,3.42,2.77,3.2,3.9,2.76,2.75,3.32)
dif1=(y-x)
power.t.test(delta = mean(dif1),
              sd = sd(dif1),
              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
```

$n = 10.30858$, so the sample size should be 11.

Problem 4

Assume normality Null: true difference equals to 0

Alternative hypothesis: true difference in means is not equal to 0

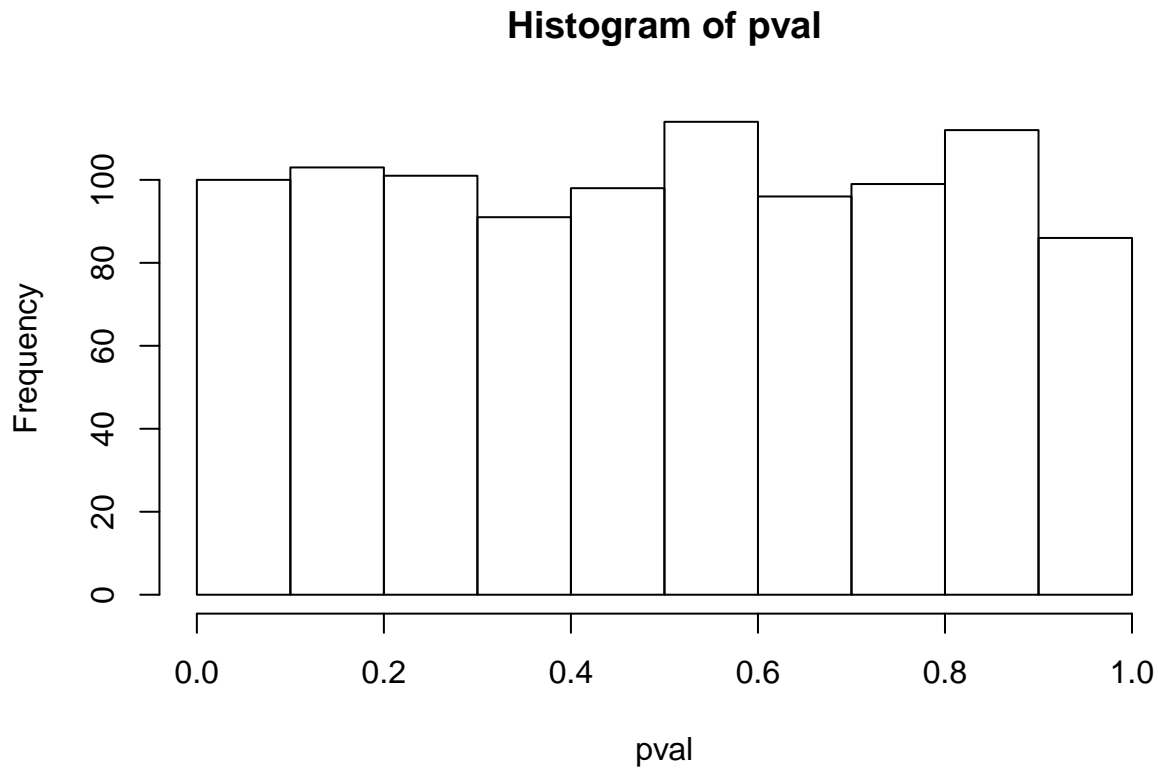
p-value = 0.02188 < 0.05 We reject null. Also, CI > 0, smokers' pulmonary function declines.

```
x=c(2.85,3.32,3.01,2.95,2.78,2.86,2.78,2.9
    ,2.76,3,3.26,2.84,2.5,3.59,3.3)
y=c(2.88,3.4,3.02,2.84,2.75,3.2,2.96,2.74
    ,3.02,3.08,3,3.4,2.59,3.29,3.32)
dif2=(y-x)
t.test(dif1,dif2,var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: dif1 and dif2
## 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
```

Problem 5

```
x=matrix(rnorm(16000,5,1),1000)
y=apply(x,1,mean)
z=apply(x,1,sd)
Y=rep(0,1000)
ts=(y - 5)/(z/sqrt(16))
pval=pt(ts,15)
hist(pval)
```



Problem 6

- a. Null: $\mu = 0$
 Alternative: $\mu \neq 0$

```
11+c(-1,1)*qt(0.975,15)*20/sqrt(16)
```

```
## [1] 0.3427523 21.6572477
```

95%CI for relative change is [0.3427523 21.6572477], which does not contain 0. So we conclude that there is statistically significant change in systolic blood pressure.

b.

```
sp=sqrt((15*20^2+15*28^2)/30)
t=(11-4)/(sp*sqrt(2/16))
2*pt(t,30,lower.tail = FALSE)#p-value
```

```
## [1] 0.4222096
```

p-value = 0.4222096 > 0.05, so we fail to reject null. The change in SBP over the two year period does not differ between oral contraceptive users and controls.

Problem 7

Always agrees with same test in kilograms. Because if we do the test, we could cancel out the constants before mean and standard variance when we do the test.

Problem 8

An increase in sample size generally will make a study more powerful to detect a difference if one truly exists, but will not shield against false rejection of the null hypothesis (in other words, will not protect from rejecting null hypothesis if there is no difference). The significance level $\hat{I}\pm$ is the probability of falsely rejecting null hypothesis, so if a researcher is concerned about falsely rejecting her null hypothesis, she should lower $\hat{I}\pm$.

Problem 9

a.

```
t <- (0.1) / (0.04 / sqrt(16))  
2 * pt(abs(t), df = 16-1, lower.tail = FALSE)
```

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

p-value = 4.996898e-08 < 0.05

Therefore we reject the null hypothesis at significance level $\alpha = 0.05$ for this population and conclude that there is a statistically significant change in grey matter.

b. If n is small, we have:

$$\begin{aligned} \text{power} &= P(TS > t_{1-\alpha, n-1} \mid \mu = \mu_a) \\ &= P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{1-\alpha, n-1} \mid \mu = \mu_a\right) \\ &= P\left(\sqrt{n}(\bar{X} - \mu_0) > t_{1-\alpha, n-1} \cdot S \mid \mu = \mu_a\right) \\ &= P\left(\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} > t_{1-\alpha, n-1} \cdot \frac{S}{\sigma} \mid \mu = \mu_a\right) \\ &= P\left(\frac{\sqrt{n}(\bar{X} - \mu_a + \mu_a - \mu_0)}{\sigma} > t_{1-\alpha, n-1} \cdot \frac{S}{\sigma} \mid \mu = \mu_a\right) \\ &= P\left(\frac{\sqrt{n}(\bar{X} - \mu_a)}{\sigma} + \frac{\sqrt{n}(\mu_a - \mu_0)}{\sigma} > t_{1-\alpha, n-1} \cdot \frac{S}{\sigma} \mid \mu = \mu_a\right) \\ &= P\left(\frac{\sqrt{n}(\bar{X} - \mu_a)}{\sigma} + \frac{\sqrt{n}(\mu_a - \mu_0)}{\sigma} > t_{1-\alpha, n-1} \cdot \sqrt{\frac{S^2}{\sigma^2}} \mid \mu = \mu_a\right) \\ &= P\left(\frac{\sqrt{n}(\bar{X} - \mu_a)}{\sigma} + \frac{\sqrt{n}(\mu_a - \mu_0)}{\sigma} > \frac{t_{1-\alpha, n-1}}{\sqrt{(n-1)}} \cdot \sqrt{\frac{(n-1)S^2}{\sigma^2}} \mid \mu = \mu_a\right) \\ &= P\left(Z + \frac{\sqrt{n}(\mu_a - \mu_0)}{\sigma} > \frac{t_{1-\alpha, n-1}}{\sqrt{(n-1)}} \cdot \sqrt{\chi_{n-1}^2} \mid \mu = \mu_a\right) \end{aligned}$$

If n is large, we have:

$$\begin{aligned}
1 - \beta &= P\left(\frac{\bar{X} - 30}{\sigma/\sqrt{n}} > z_{1-\alpha} \mid \mu = \mu_a\right) \\
&= P\left(\frac{\bar{X} - \mu_a + \mu_a - 30}{\sigma/\sqrt{n}} > z_{1-\alpha} \mid \mu = \mu_a\right) \\
&= P\left(\frac{\bar{X} - \mu_a}{\sigma/\sqrt{n}} > z_{1-\alpha} - \frac{\mu_a - 30}{\sigma/\sqrt{n}} \mid \mu = \mu_a\right) \\
&= P\left(Z > z_{1-\alpha} - \frac{\mu_a - 30}{\sigma/\sqrt{n}} \mid \mu = \mu_a\right)
\end{aligned}$$

we choose T test: plug in $\mu_0 = 0$, $\mu_a = 0.01$, $n = 100$, $\sigma = 0.04$

```

nosim=1000;
n = 100;
sigma = 0.04;
mu0 = 0;
mua <- 0.01
z = rnorm(nosim)
xsq = rchisq(nosim, df = n-1)
t = qt(.95, n-1)
mean(z + sqrt(n) * (mua - mu0) / sigma > t / sqrt(n - 1) * sqrt(xsq))

```

```
## [1] 0.811
```

Or we could use following command

```
power.t.test(n = 100, delta = 0.01/0.04, type = "one.sample", alt = "one.sided")
```

```

##
##      One-sample t test power calculation
##
##              n = 100
##            delta = 0.25
##              sd = 1
##      sig.level = 0.05
##        power = 0.7989855
##  alternative = one.sided

power = 0.7969757

```

Problem 10

If you reject the null hypothesis for $\alpha = 0.05$, this implies 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

```
sp=sqrt((8*1.5^2+8*1.8^2)/16)
4+c(-1,1)*qt(0.975,16)*sp*sqrt(2/9)
```

```
## [1] 2.344301 5.655699
```

```
TS = 4/(sp*sqrt(2/9))
2*pt(TS,16,lower.tail = FALSE)
```

```
## [1] 0.0001025174
```

p-value = 0.0001025174 < 0.05

95% CI interval = [2.344301 5.655699] which does not contain 0. Meaning the change in BMI over the two year period appear to differ between the treated and placebo groups.