DSC 462: Assignment 4

Shouman Das

November 14, 2017

Q1

(a) Standard deviation of the sample mean $\sigma_{\bar{X}_{12}} = \frac{1.2}{\sqrt{12}}$

```
1.2/sqrt(12)
```

```
## [1] 0.3464102
```

(b) We have, $\bar{X}_{12} \sim N(76.5, 1.2^2/12)$. The desired probability is then

$$\begin{aligned} P(76.5 - 4 < \bar{X}_{12} < 76.5 + 4) &= P(72.5 < \bar{X}_{12} < 80.5) \\ &= P(-4/0.3464102 < Z < 4/0.3464102) \quad \approx 1 \end{aligned}$$

Q2

(a) Here the true variance of population is unknown. So we use $t_{n-1,\frac{\alpha}{2}}$ for $z_{\frac{\alpha}{2}}$.

```
x = c(23.72, 18.95, 9.79, 12.04, 9.54)
t.test(x, conf.level = .9)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 5.2823, df = 4, p-value = 0.006161
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##   8.831717 20.784283
## sample estimates:
## mean of x
##    14.808
```

So 90% confidence interval is $(8.831717, 20.784283)$. We can also check this by hand calculation with $t_{4,.05} = 2.132$ and $\bar{X}_n = 14.808, S_n = 39.29347$,

$$(14.808 - 2.132\sqrt{\frac{39.29347}{5}}, 14.808 + 2.132\sqrt{\frac{39.29347}{5}})$$

$$= (8.831717, 20.784283)$$

(b)

```
xbar = mean(x)       # sample mean
mu0 = 24             # hypothetical value
s = sqrt(var(x))     # sample standard deviation
n = 5                # sample size
Tobs = (xbar - mu0)/(s/sqrt(n))
Tobs                 # test statistic
```

```
## [1] -3.27895
```

```
alpha = 0.05
p_val = 2*pt(Tobs, df = n-1)
p_val
```

```
## [1] 0.03053087
```

For significance level $\alpha = .05$ our P-value ≤ 0.05, we have strong evidence to reject H_0 in favor of H_a. But for $\alpha = 0.01$, we don't have enough evidence to reject H_0.

Q3

```
### set up labels with mathematical typesetting
ex0 = expression(paste("Power Curve with sample size ",
italic(n)," and ",alpha," = 0.05",sep=''))
ex1 = expression(italic(mu))
ex2 = expression(Power(mu))

### grid for horizontal axis
mu = seq(150, 175, by = 0.1)

### set up graphics window, draw empty plot (type='n')
par(mar=c(4,5,2,2), oma=c(4,4,4,4), cex=1,cex.axis=1.2, cex.lab=1.2, cex.main=1.2)
plot(range(mu), c(0,1), xlab=ex1, ylab=ex2, type='n')
title(ex0)

### grid
for (x in seq(150,175,by=2.5)) {abline(v=x,col='gray')}
for (y in seq(0,1,by=0.05)) {abline(h=y,col='gray')}
abline(h=0.05,col='blue')
ex4 = bquote(alpha == 0.05)
text(170, 0.05, ex4)

for(n in c(5,10,15,20,25,30)){
    # power curve for n
    alpha = 0.05
    z.alpha = qnorm(1-alpha)
    y = 1 - pnorm((150 + z.alpha*10/sqrt(n) -mu)/(10/sqrt(n)))
    lines(mu, y, type = 'l')
```

```
    # label n=5,30 plot
    if(n==5){
        ex3 = bquote(italic(n) == .(n))
        text(155, y[mu==155], ex3)
    }else if(n==30){
        ex3 = bquote(italic(n) == .(n))
        text(155, y[mu==155], ex3)
        }
}
```

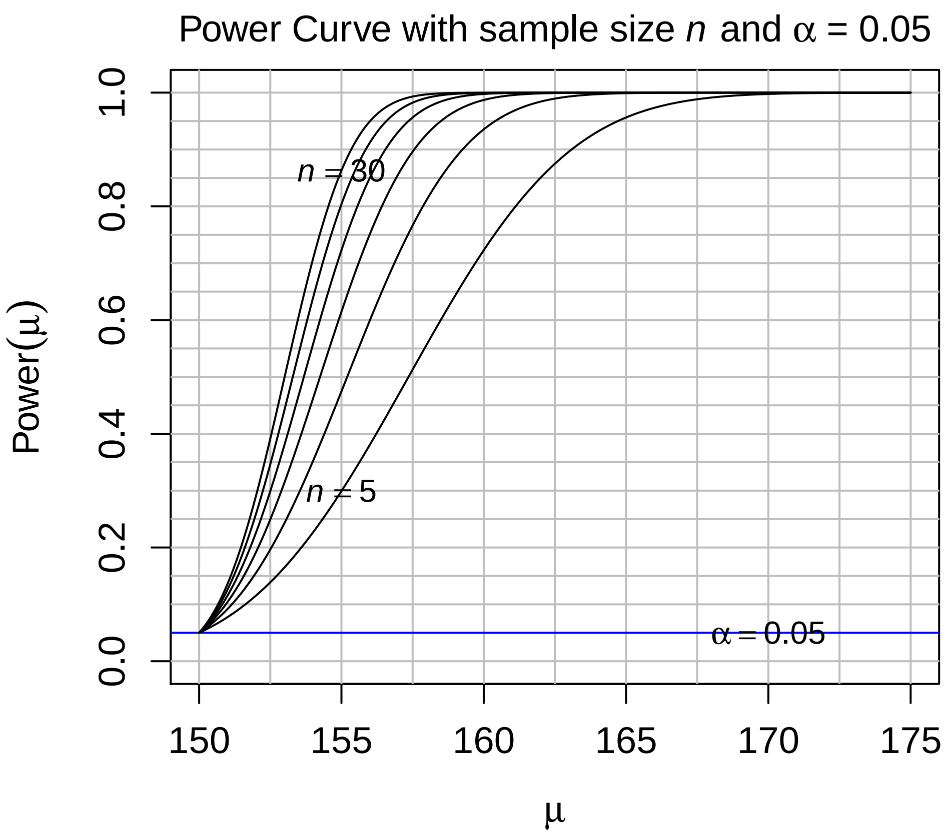


Figure 1: Power curve

(b)

	μ					
n	155	156	157	158	159	160
5	0.2992	0.3809	0.4683	0.5573	0.6434	0.7228
10	0.4746	0.5997	0.7152	0.8119	0.8852	0.9354
15	0.6147	0.7514	0.8568	0.9270	0.9672	0.9871
20	0.7228	0.8505	0.9313	0.9734	0.9913	0.9977
25	0.8038	0.9123	0.9682	0.9907	0.9978	0.9996
30	0.8630	0.9497	0.9857	0.9969	0.9995	0.9999

From the table, we can find the minimum sample size for 80% power.

	155	156	157	158	159	160
Min size for .8 power	25	20	15	10	10	10

Q4

(a)i) Equal variance:

```
alpha = 0.02
xbar1 = 96.3
xbar2 = 121.5
S1 = 16.5
S2 = 12.5
n1 = 15
n2 = 8
Sp = sqrt(((n1-1)*S1^2+(n2-1)*S2^2)/(n1+n2-2))
t.crit = qt(1-alpha/2, df = n1+n2-2)
conf.int = c(xbar2-xbar1-t.crit*Sp*sqrt(1/n1+1/n2), xbar2-xbar1+t.crit*Sp*sqrt(1/n1+1/n2))
conf.int
```

```
## [1]  8.354293 42.045707
```

ii) Unequal variance: Welch's procedure

```
nu.w = floor((S1^2/n1+S2^2/n2)^2/((S1^2/n1)^2/(n1-1)+(S2^2/n2)^2/(n2-1)))
t.crit = qt(1-alpha/2, df = nu.w)
conf.int = c(xbar2-xbar1-t.crit*sqrt(S1^2/n1+S2^2/n2), xbar2-xbar1+t.crit*sqrt(S1^2/n1+S2^2/n2))
conf.int
```

```
## [1]  9.532204 40.867796
```

(b) Hypothesis test for $H_o : \mu_2 - \mu_1 = 0$ against $H_a : \mu_2 - \mu_1 > 0$.

```
# Equal variance
T.obs = (xbar2-xbar1)/(Sp*sqrt(1/n1+1/n2))
P.value = 1 - pt(T.obs, df = n1+n2-2)
P.value
```

```
## [1] 0.0005673169
```

```
# Unequal variance
T.obs = (xbar2-xbar1)/(sqrt(S1^2/n1+S2^2/n2))
P.value = 1 - pt(T.obs, df = nu.w)
P.value
```

```
## [1] 0.0003322196
```

In both cases, we can reject H_o in favour of H_a.

Q5

(a) t-test

```r
#paired t-test
pipe = c(6.64, 7.89, 1.83, 0.42, 0.85, 0.29, 0.57, 0.63, 0.32, 0.37, 0.00, 0.11, 4.86, 1.80, 0.23, 0.58
brush = c(9.73, 8.21, 2.17, 0.75, 1.61, 0.75, 0.83, 0.56, 0.76, 0.32, 0.48, 0.52, 5.38, 2.33, 0.91, 0.7
t.test(pipe, brush, paired = TRUE, conf.level = 0.99)
```

```
##
##  Paired t-test
##
## data:  pipe and brush
## t = -3.0496, df = 15, p-value = 0.00811
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -1.07038573 -0.01836427
## sample estimates:
## mean of the differences
##               -0.544375
```

Since the p-value is less than $\alpha = 0.001$ we can reject the null hypothesis and say that brush is more effective than pipe.

(b)

```r
t.test(pipe, brush, var.equal = TRUE, conf.level = .99, paired = FALSE)
```

```
##
##  Two Sample t-test
##
## data:  pipe and brush
## t = -0.56979, df = 30, p-value = 0.5731
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -3.171716  2.082966
## sample estimates:
## mean of x mean of y
##  1.711875  2.256250
```

```r
t.test(pipe, brush, var.equal = FALSE, conf.level = .99, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  pipe and brush
## t = -0.56979, df = 29.251, p-value = 0.5732
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -3.176247  2.087497
## sample estimates:
## mean of x mean of y
##  1.711875  2.256250
```

This we can see that 0 is in the confidence interval. So we cannot reject H_o.

Q5

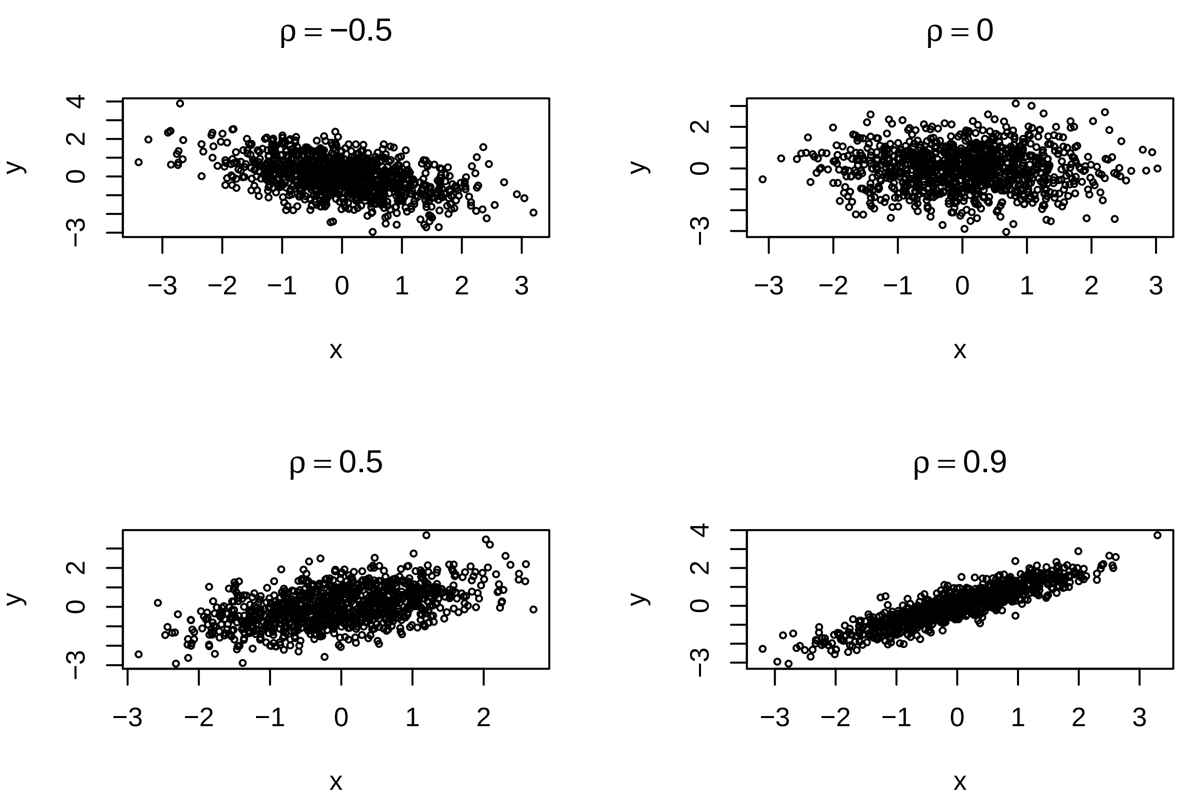
(a) Note that $Var(Y) = \beta_1^2 \; Var(X) + \beta_2^2 \; Var(\epsilon) = \beta_1^2 + \beta_2^2$. So $\sigma_Y = \sqrt{\beta_1^2 + \beta_2^2}$. Also, $cov(X, Y) = cov(X, \beta_1 \; X) = \beta_1$. So,

$$corr(X, Y) = \frac{\beta_1}{\sqrt{\beta_1^2 + \beta_2^2}}$$

which doesn't depend on β_0.

(b) We can choose $\beta_0 = 0, \beta_1 = \rho, \beta_2 = \sqrt{1 - \rho^2}$

```r
set.seed(1234)
par(mfrow=c(2,2))
mat = matrix(nrow = 4, ncol = 4)
colnames(mat) <- c("rho", "Var(X)", "Var(Y)", "Correlation")
df  = as.data.frame(mat)
i = 1
for(rho in c(-0.5, 0, 0.5, 0.9)){
    x = rnorm(1000)
    epsilon = rnorm(1000)
    beta1 = rho
    beta2 = sqrt(1-rho^2)
    y = beta1*x+beta2*epsilon
    ex5 = bquote(rho == .(rho))
    plot(x,y, main = ex5, cex = 0.5)
    df$rho[i] = rho
    df$`Var(X)`[i] = var(x)
    df$`Var(Y)`[i] = var(y)
    df$Correlation[i] = cor(x,y)
    i=i+1
}
```




```r
df # Summary of sample variance and correlation of X,Y
```

```
##    rho    Var(X)    Var(Y)  Correlation
## 1 -0.5 0.9946825 0.9227325 -0.469032926
## 2  0.0 1.0248294 0.9875567 -0.008370712
## 3  0.5 0.9447794 0.9389075  0.494799937
## 4  0.9 0.9983888 0.9689276  0.896860116
```