MAST90105, Assignment 4, 2019

Tingqian Wang

2019/6/7

Q1

(a)
The confidence interval required is [-4.0568,-0.3265] - see below.
In this case, variance of two samples are unkown but equal. Assign smaple A to X and B to Y.

```
X=c(3.9,4.7,7.1,6.9,4.3,6.3)
Y=c(7.3,6.9,7.6,9.1)
mX=mean(X)
mY=mean(Y)
mX
```

```
## [1] 5.533333
```

```
mY
```

```
## [1] 7.725
```

```
sX=sd(X)
sY=sd(Y)
sX
```

```
## [1] 1.399524
```

```
sY
```

```
## [1] 0.9604686
```

The pooled estimate of the standard deviation is

```
Sp=sqrt(((6-1)*sX^2+(4-1)*sY^2)/(4+6-2))
Sp
```

```
## [1] 1.253038
```

Hence, a 95% confidence interval for $\delta = \mu_A - \mu_B$ is[-4.056834, -0.3264988], known $t_{0.025}(8)$ = 2.306 - see the R output below.

```
CI_up=mX-mY+2.306*Sp*sqrt(1/6+1/4)
CI_lw=mX-mY-2.306*Sp*sqrt(1/6+1/4)
CI_up
```

```
## [1] -0.3264988
```

```
CI_lw
```

```
## [1] -4.056834
```

or using t.test

```
t.test(X,Y,conf.level = 0.95, var.equal = T)$conf.int
```

```
## [1] -4.0568378 -0.3264955
## attr(,"conf.level")
## [1] 0.95
```

(b)

We can reject H_0 at the 5% (or any lower) significance level - see R output below. The confidence interval of 95% for $\delta = \mu_A$ - μ_B is [-4.0568378, -0.3264955], excluding δ=0. And the p-value of this test is 0.02667.

```
t.test(X,Y,conf.level=0.95,alternative = "two.sided",var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  X and Y
## t = -2.7097, df = 8, p-value = 0.02667
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.0568378 -0.3264955
## sample estimates:
## mean of x mean of y
##  5.533333  7.725000
```

(c)

In the same case but assumed that the variances are 2 and 1, respectively, for the two samples. See the R output below, the 95% confidence interval for δ is [-3.9147468, -0.4685866].

```
t.test(X,Y,conf.level=0.95,alternative = "two.sided",var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  X and Y
## t = -2.9364, df = 7.9484, p-value = 0.01895
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.9147468 -0.4685866
## sample estimates:
## mean of x mean of y
##  5.533333  7.725000
```

(d)
Compare the confidence interval obtain in (a) and (c), the confidence interval in (c) is narower than the one in (a). The pooled estimate of the common variance in (a) is 1.253038 as a denominator of the pivot, however, in (c), the denominator is

$$\sqrt{S_X^2/n + S_Y^2/m} = 0.957427$$

. The smaller the denomiantor is, the narrower the width of the confidence interval is. And as the variances are known, the confidence interval shringks and more precise wiht higher accuracy.

(e)
Set $Y_1 < Y_2 < Y_3 < Y_4 < Y_5 < Y_6$ as ordered statistics for iid rv's $X_1, \ldots X_6$
The median concentration of carbon monoxide in City A $m = \pi_{0.5}$ to be between Y_1 and Y_6 must have at least one $X_i < m$ but not five $X_i < m$. So let W be the number of $X's < m$, then $W \sim$ Binom$(6, 0.5)$ and

$$P(Y_1 < m < Y_6) = P(0 < W < 5) = \sum_{k=1}^{5} \binom{6}{k} (\frac{1}{2})^k (\frac{1}{2})^{5-k} = 0.96875$$

which is approximately 97% - see R output below.

```
P_confint=dbinom(1,6,0.5)+dbinom(2,6,0.5)+dbinom(3,6,0.5)+dbinom(4,6,0.5)+dbinom(5,6,
0.5)
P_confint
```

```
## [1] 0.96875
```

Q2

Given X follows exponential distribution with parameter λ and sample size of $n = 10$, and$Y = \sum_{i=1}^{10} X_i$, there is $\lambda Y \sim$ Gamma$(10, 1)$.
$H_0 : \lambda = 1$ is rejected in favor of $H_1 : \lambda = 2$ if $Y < 6$ is observed.

(a)
Probability of tpye 1 error is P(reject $H_0 \mid H_0$ is true), which is equals to
$P(Y < 6|\lambda = 1) = P(Y < 6|Y \sim$ Gamma$(10, 1)) = 0.08392402$ - see the R output below.

```
pgamma(6,10,1,lower.tail = TRUE)
```

```
## [1] 0.08392402
```

(b)
Probability of tpye 2 error is P(can't reject $H_0 \mid H_1$ is true), which is equals to
$P(Y \geq 6|\lambda = 2) = P(2Y \geq 12|2Y \sim \text{Gamma}(10, 1)) = 0.2423922$ - see the R output below.

```
pgamma(12,10,1,lower.tail = FALSE)
```

```
## [1] 0.2423922
```

(c)
if $y = 5$ is the observed value of Y, p-value of this test is $P(Y > 5|\lambda = 1)$. So
$P(Y < 5|Y \sim \text{Gamma}(10, 1)) = 0.03182806$. This means that H_0 could have been rejected at significance level $\alpha = 0.03182806$, which is much stronger than rejecting it at $\alpha = 0.05$.

```
pgamma(5,10,1,lower.tail = TRUE)
```

```
## [1] 0.03182806
```

(d)
As 95% confidence interval, $\alpha = 0.05$.
That is $P(\text{Gamma}_{0.025}(10, 1) \leq \lambda Y \leq \text{Gamma}_{0.975}(10, 1))$
And given $y = 5$,

$$\frac{\text{Gamma}_{0.025}(10, 1)}{Y} \leq \lambda \leq \frac{\text{Gamma}_{0.975}(10, 1)}{Y}$$

```
gq=qgamma(c(0.025,0.975),scale = 1,shape = 10)
## given y=5
c(gq[1]/5,gq[2]/5)
```

```
## [1] 0.9590777 3.4169607
```

So the interval of λ at significance level of 0.05 is $0.9590777 \leq \lambda \leq 3.4169607$.

Q3

(a)
For $Y \sim \text{Poisson}(\lambda)$,

$$f(y) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Then the likelihood function is

$$L(\lambda) = \frac{\lambda^{\sum_{i=1}^{100} x_i} e^{-100\lambda}}{\prod_{i=1}^{100} x_i!}$$

Let $l = ln(L(\lambda))$, then

$$l = \sum_{i=1}^{100} X_i \ln \lambda + \ln e^{-100\lambda} - \ln(\prod_{i=1}^{100} x_i!)$$

Then

$$\frac{\partial l}{\partial \lambda} = \frac{\sum_{i=1}^{100} X_i}{\lambda} - 100$$

Let $\frac{\partial l}{\partial \lambda} = 0$,get

$$\hat{\lambda} = \frac{119}{100} = 1.19$$

(b)
H_0:the number of faults for a data-transmission line follows a Poisson distribution.
H_1:the number of faults for a data-transmission line doesn't follows a Poisson distribution.
And the significance level is $\alpha = 0.05$.

```
# Build the matirx of the obseved sample.
# Note "x=6" repsents that "x>5" here.
x <- matrix(c(0,1,2,3,4,5,6), ncol = 7)
f <- matrix(c(38,30,16,9,5,2,0), ncol = 7)
t <- rbind(x,f, col.names = NULL)
colnames(t) <- c("x=0","x=1","x=2","x=3","x=4","x=5","x>5")
rownames(t) <- c("x","Frequency")
t <- as.table(t)
t
```

```
##            x=0 x=1 x=2 x=3 x=4 x=5 x>5
## x            0   1   2   3   4   5   6
## Frequency   38  30  16   9   5   2   0
```

```
p0=dpois(x[1],1.19)
p1=dpois(x[2],1.19)
p2=dpois(x[3],1.19)
p3=dpois(x[4],1.19)
p4=dpois(x[5],1.19)
p5=dpois(x[6],1.19)
p6=1-p0-p1-p2-p3-p4-p5
p=c(p0,p1,p2,p3,p4,p5,p6)
# probability of $x_i$
p
```

```
## [1] 0.304221264 0.362023304 0.215403866 0.085443534 0.025419451 0.006049829
## [7] 0.001438752
```

```
e=rep(0,7)
for (k in 1:7){
  e[k]=p[k]*100
}
# get the expected value of frequency.
e
```

```
## [1] 30.4221264 36.2023304 21.5403866  8.5443534  2.5419451  0.6049829
## [7]  0.1438752
```

Note that the last 3 expected frequencies are less than 5. So combine them with the 4th group as one group. Therefore the last observed frequencies and expected frequencies is update to 16 and 11.83516 respectively.

```
## Rebuilt the matrix
Of <- matrix(c(38,30,16,16), ncol = 4)
Ef <- matrix(c(30.4221264,36.2023304,21.5403866,11.83516), ncol = 4)
T <- rbind(Of,Ef, col.names = NULL)
colnames(T) <- c("x=0","x=1","x=2","x>=3")
rownames(T) <- c("Observed frequency","Expected frequency")
T <- as.table(T)
T
```

```
##                        x=0      x=1      x=2     x>=3
## Observed frequency 38.00000 30.00000 16.00000 16.00000
## Expected frequency 30.42213 36.20233 21.54039 11.83516
```

```
V=rep(0,4)
for(n in 1:4){
  V[n]=(Of[n]-Ef[n])^2/Ef[n]
}
V
```

```
## [1] 1.887579 1.062608 1.425039 1.465624
```

The test statistic $\chi^2 = \sum_{i=1}^{4} V_i$

```
chisqrt=sum(V)
chisqrt
```

```
## [1] 5.84085
```

And the degree of freedom is $n - 1 - 1 = 4 - 2 = 2$.

```
# Calculate the p-value
1-pchisq(chisqrt,2)
```

```
## [1] 0.05391077
```

Since the p-value is approximately 0.05, and $0.05391077 > \alpha = 0.05$, the decision is fail to reject H_0. Actually, the difference between p-value of this test and the significance level is quite small. So even though being fail to reject H_0 means that the number of faults for a data-transmission line follows a Poisson distribution, there should be more smaple to be test.