

Nonparametric tests - EXERCISES

- 2 → 1) known discrete
- 9a → 2) known cont → discrete
- 9b, 7b, 5b } unknown discrete → $\hat{\theta}$, df
- 4 → 4) unknown cont → $\hat{\theta}$, discrete, df.

1)

	door 1	door 2	door 3
$X = \# \text{ cust.}$	$O_1 = 83$	$O_2 = 62$	$O_3 = 56$

$k = 3$ groups

known H_0 distribution (discrete)

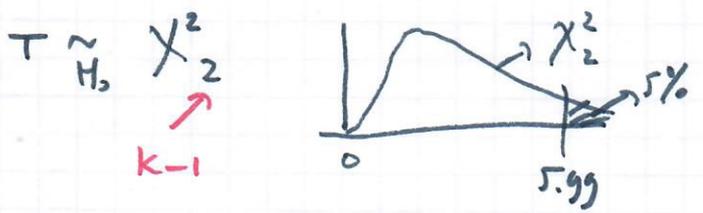
$H_0:$

$P_1 = 1/3$	$P_2 = 1/3$	$P_3 = 1/3$
-------------	-------------	-------------

$n = 201:$

$E_1 = 67$	$E_2 = 67$	$E_3 = 67$
------------	------------	------------

$$T_{\text{obs}} = \sum_{i=1}^{k=3} \frac{(O_i - E_i)^2}{E_i} = \frac{(83 - 67)^2}{67} + \frac{(62 - 67)^2}{67} + \frac{(56 - 67)^2}{67} = 6$$



As $T_{\text{obs}} = 6 > 5.99$, we reject H_0 .

$\text{obs} \leftarrow c(83, 62, 56)$
 $\text{prob-H}_0 \leftarrow c(1/3, 1/3, 1/3)$
 $\text{chisq.test}(x = \text{obs}, p = \text{prob-H}_0)$
 $(\text{qchisq}(0.95, df = 2))$ # critical value
 $1 - \text{pchisq}(6, df = 2)$ # p-value

Known discrete H_0 distribution, $E_i < 5$

2) $H_0: X \sim \text{Poisson}(\lambda = 0.2)$

	0	1	2	≥ 3
Obs	800	175	21	4

under H_0 :

$$P(X=0) = e^{-0.2} \frac{0.2^0}{0!} \approx 0.81873 \rightarrow E_1 = 818.73$$

$$P(X=1) \approx 0.16375 \rightarrow E_2 = 163.75$$

$$P(X=2) = 0.01637$$

$$P(X \geq 3) = 0.00115 \Rightarrow E_4 = 1.15 < 5 !$$

join groups 3 & 4:

	0	1	≥ 2
	800	175	25

$$P(X=0) = 0.81873$$

$$P(X=1) = 0.16375$$

$$P(X \geq 2) = 1 - P(X=0) - P(X=1) = 0.01752$$

$$T_{obs} = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} = \frac{(800 - 818.73)^2}{818.73} + \frac{(175 - 163.75)^2}{163.75} + \frac{(25 - 17.52)^2}{17.52}$$

$$= \cancel{4.39} 4.39$$



As $T_{obs} = 4.39 \neq 5.99$, we do not reject H_0 .

```

(R) p1 <- dpois(0:2, lambda=0.2)
p2 <- 1 - sum(p1)
chisq.test(x=c(800,175,21,4), p=c(p1,p2)) warning!

p1 <- dpois(0:1, lambda=0.2)
p2 <- 1 - sum(p1)
chisq.test(x=c(800,175,25), p=c(p1,p2))
    
```

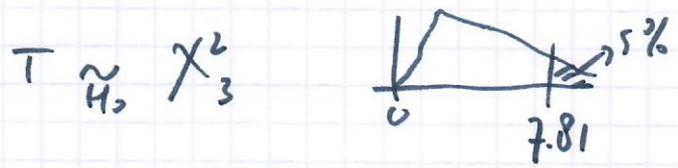
known discrete H_0 distribution

3) H_0 :

^① 9-12	^② 12-14	^③ 14-17	^④ 17-19	
$p_1 = 1/4$	$p_2 = 1/4$	$p_3 = 1/4$	$p_4 = 1/4$	
$E_1 = 200$	$E_2 = 200$	$E_3 = 200$	$E_4 = 200$	all ≥ 5
$O_1 = 150$	$O_2 = 220$	$O_3 = 235$	$O_4 = 195$	

$$T_{obs} = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} = \frac{(150-200)^2}{200} + \frac{(220-200)^2}{200} + \frac{(235-200)^2}{200} + \frac{(195-200)^2}{200}$$

$$= 20.75$$



As $T_{obs} = 20.75 > 7.81$, we reject H_0

(R) chisq. test ($x = c(150, 220, 235, 195)$, $p = rep(1/4, 4)$)

4) $X \sim \exp(\lambda=?)$ unknown continuous distribution

$Y = X$ discretized into intervals $\overset{(1)}{[0,3)}, \overset{(2)}{[3,6)}, \overset{(3)}{[6,9)}, \overset{(4)}{[9,12)}, \overset{(5)}{[12,15)}, \overset{(6)}{[15,\infty)}$
 $o_i = [205 \mid 135 \mid 80 \mid 34 \mid 26 \mid 20]$

total observed waiting time $= \sum_{i=1}^{500} X_i = 2500 \Rightarrow \bar{x} = 5$
 $\Rightarrow \hat{\lambda}_{me} = \frac{1}{\bar{x}} = \frac{1}{5} = 0.2$

\Rightarrow test $H_0: Y \sim$ discretized $\exp(\lambda = \frac{1}{5})$
 using $T \underset{H_0}{\sim} \chi^2_{k-1} \rightarrow$ estimated λ

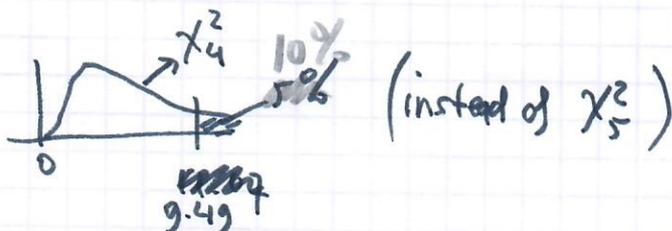
~~calculations~~

$P_1 = P(0 < X < 3) = P(X < 3) = P(\exp(\lambda = \frac{1}{5}) < 3) = 0.451188 \quad E_1 = 225.594$
 $P_2 = P(3 < X < 6) = P(X < 6) - P(X < 3) = 0.247617 \quad E_2 = 123.805$
 $P_3 = P(6 < X < 9) = P(X < 9) - P(X < 6) = 0.135895 \quad E_3 = 67.948$
 $P_4 = P(9 < X < 12) = 0.07458 \quad E_4 = 37.290$
 $P_5 = P(12 < X < 15) = 0.04093 \quad E_5 = 20.465$
 $P_6 = P(X \geq 15) = 1 - P_1 - P_2 - P_3 - P_4 - P_5 = 0.049787 \quad E_6 = 24.8935$

$$T_{obs} = \sum_{i=1}^6 \frac{(o_i - e_i)^2}{E_i} = \frac{(205 - 225.594)^2}{225.594} + \frac{(135 - 123.805)^2}{123.805} + \frac{(80 - 67.948)^2}{67.948} + \frac{(34 - 37.290)^2}{37.290} + \frac{(26 - 20.465)^2}{20.465} + \frac{(20 - 24.8935)^2}{24.8935}$$

$$= 7.775$$

$T \underset{H_0}{\sim} \chi^2_5$



As $T_{obs} = 7.78 \neq 9.49$, we do not reject H_0 .

(R)

$$p1 \leftarrow \text{pexp}(3*(1:5), 0.2) - \text{pexp}(3*(0:4), 0.2)$$

$$p2 \leftarrow 1 - \text{sum}(p1)$$

$$\text{chisq.test}(x = c(205, 135, 80, 34, 26, 20), p = c(p1, p2))$$

critical value corrected for estimating lambda:

$$qchisq(0.95, df = 4)$$

p-value corrected for estimating lambda:

$$1 - \text{pchisq}(7.7785, df = 4)$$

independence test

	Port	BenRice	Sponting
$X = \text{age} \leq 35$	$O_{11} = 75$	$O_{12} = 75$	$O_{13} = 50$
$X = \text{age} > 35$	$O_{21} = 75$	$O_{22} = 125$	$O_{23} = 100$

$r = 2$

$C = 3$

$n = 500$

Are X and Y independent ($\alpha = 5\%$)?

expected $f_{X,Y}$ if $X \perp Y$:

0.12	0.16	0.12	$P(X=1) = \frac{200}{500} = 0.4$
0.18	0.24	0.18	$P(X=2) = 0.6$

$P(Y=1) = \frac{150}{500} = 0.3$
 $P(Y=2) = \frac{200}{500} = 0.4$
 $P(Y=3) = 0.3$

expected obs E_{ij} if $X \perp Y$:

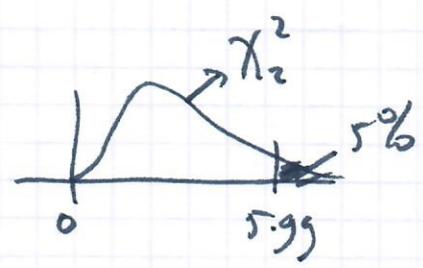
60	80	60
90	120	90

$$T_{obs} = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \frac{(75-60)^2}{60} + \frac{(75-80)^2}{80} + \frac{(50-60)^2}{60} + \frac{(75-90)^2}{90} + \frac{(125-120)^2}{120} + \frac{(100-90)^2}{90}$$

$$= 9.55$$

$$T \sim \chi^2 = (r-1)(c-1)$$



As $T_{obs} = 9.55 > 5.99$ we reject H_0

(R)

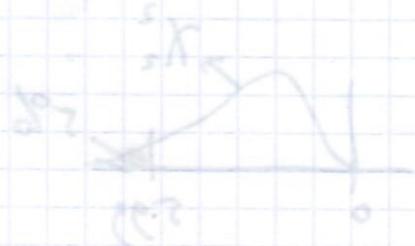
table <- matrix(c(75, 75, 75, 125, 50, 100), nrow=2)

chisq.test(x=table, correct=FALSE)

51.0	41.0	92.0
61.0	45.0	106.0
112.0	86.0	198.0
0.0	0.0	0.0

00	80	80
00	150	150

$$T_{adj} = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \frac{(75-92)^2}{92} + \frac{(75-41)^2}{41} + \frac{(75-45)^2}{45} + \frac{(125-50)^2}{50} + \frac{(50-100)^2}{100} = 15.27$$



As $T_{adj} = 15.27 > 2.0$ we reject H_0

6) independence test

chisq.test(x = ^{matrix} c(44, 36, 37, 33, 23, 27), nrow = 2, correct = FALSE)

gives

$$T_{obs} = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 1.0302 \quad p\text{-value} = 0.5974$$

(critical value = $qchisq(0.95, df=2) = 5.99$)

We do not reject H_0 .

~~discrete~~ (known and unknown discrete H_0 distribution & $E_i < 5$)

7)

0	1	2	≥ 3
$O_1 = 63$	$O_2 = 25$	$O_3 = 8$	$O_4 = 2$

$n = 100$ gP
(not: 100)

a) $H_0: X \sim \text{Poisson}(\lambda = 0.4)$

$P_1 = \frac{0.4^0 e^{-0.4}}{0!} = 0.6703 \Rightarrow E_1 = 98 \times 0.6703 = 65.6894$
 $P_2 = \frac{0.4^1 e^{-0.4}}{1!} = 0.2681 \Rightarrow E_2 = 26.2738$
 $P_3 = \frac{0.4^2 e^{-0.4}}{2!} = 0.0536 \Rightarrow E_3 = 5.2528$
 $P_4 = 1 - P_1 - P_2 - P_3 = 0.008 \Rightarrow E_4 = 0.7840$

0	1	≥ 2
63	25	10

$T = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} = 2.7737$
 $P_3 = \frac{0.4^3 e^{-0.4}}{3!} = 0.0616$
 $E_3 = 6.0368$
 $P\text{-value} = 0.2455$
 $\text{critical value } \chi^2_2 = 5.99$

We do not reject H_0

(R)

$p_1 \leftarrow \text{round}(\text{dpois}(0:2, 0.4), 4)$

$p_2 \leftarrow 1 - \text{sum}(p_1)$

$\text{prob-H}_0 \leftarrow c(p_1, p_2)$

$gP * \text{prob-H}_0$

$\text{chisq.test}(c(63, 25, 8, 2), p = \text{prob-H}_0)$
gives warning

$p_1 \leftarrow \text{round}(\text{dpois}(0:1, 0.4), 4)$

$p_2 \leftarrow 1 - \text{sum}(p_1)$

$\text{chisq.test}(c(63, 25, 10), p = c(p_1, p_2))$

b) $H_0: X \sim \text{Poisson} (\lambda = ?)$

$$o_1 = 63 \mid o_2 = 25 \mid o_3 = 8 \mid o_4 = 2$$

$$\hat{\lambda}_{MLE} = \bar{x} \approx \frac{63 \times 0 + 25 \times 1 + 8 \times 2 + 2 \times 3}{98} = 0.4796$$

$$\Rightarrow p_1 = 0.6190$$

$$\Rightarrow E_1 = 98 \times p_1 = 60.6620 \geq 5$$

$$p_2 = 0.2565$$

$$\Rightarrow E_2 = 25.0562 \geq 5$$

$$p_3 = 0.0712$$

$$\Rightarrow E_3 = 6.9776 \geq 5$$

$$p_4 = 1 - p_1 - p_2 - p_3 = 0.0129 \Rightarrow E_4 = 1.2642 < 5 !$$

$$o_1 = 15 \mid o_2 = 25 \mid o_3 = 10$$

$$p_1 = 0.6150 \Rightarrow E_1 = 60.6620$$

$$p_2 = 0.2565 \Rightarrow E_2 = 25.0562$$

$$p_3 = 0.0841 \Rightarrow E_3 = 8.2418 \geq 5$$

$$T = \sum_{i=1}^3 \frac{(o_i - E_i)^2}{E_i} = 1.0418$$

$$p\text{-value} = 0.594$$

we do not reject H_0

$$\text{crit. value } \chi^2_{2, 0.95} = 5.99$$

$$\chi^2_{1, 0.95} = 6.63$$

(R) $p1 \leftarrow \text{round}(\text{dpois}(0:1, \text{lambda} = \text{lambda_hat}), 4)$

$p2 \leftarrow 1 - \text{sum}(p1)$

$\text{chisq.test}(x = c(63, 25, 10), p = (p1, p2))$

3) H_0 - distribution discrete and known

$H_0: X \sim \text{Poisson}(\lambda = 1.5)$

①	②	③	④	⑤	⑥
23	22	18	15	1	1

$P_1 = P(X=0) = 0.2231 \rightarrow E_1 = 80 \times p_1 = 17.848 \quad O_1 =$

$P_2 = P(X=1) = 0.3397 \rightarrow E_2 = 26.776$

$P_3 = P(X=2) = 0.2510 \rightarrow E_3 = 20.080$

$P_4 = P(X=3) = 0.1255 \rightarrow E_4 = 10.040$

$P_5 = P(X=4) = 0.0471 \rightarrow E_5 = 3.768$

$P_6 = P(X \geq 5) = 0.0186 \rightarrow E_6 = 1.488$

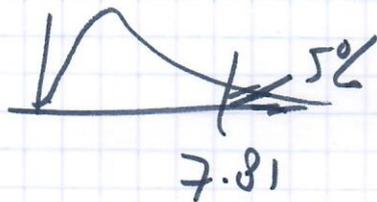
$O_4 = 17$
 $P_4 = 0.1912$
 $E_4 = 15.296$

22	22	18	17
----	----	----	----

$$T_{obs} = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i}$$

$$= 2.7444$$

$T \underset{H_0}{\sim} \chi^2_3$



As $T_{obs} = 2.74 \neq 7.81$, we do not reject H_0

② $p1 \leftarrow \text{round}(\text{dpois}(0:2, 1.5), 4)$

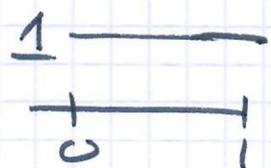
$p2 \leftarrow 1 - \text{sum}(p1)$

$\text{chisq.test}(c(23, 22, 18, 17), p = c(p1, p2))$

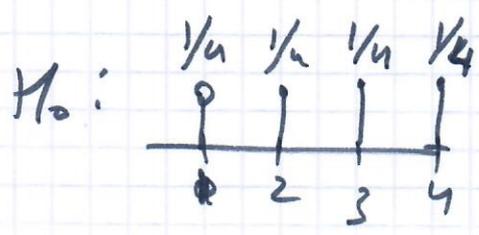
$qchisq(0.95, 3)$

- g) a) known continuous distribution
 b) independence test

a) $X \sim \text{Unif}(0,1)$



$Y \sim \text{discretized } X \text{ in intervals } (0, 0.25)$
 $(0.25, 0.5)$
 $(0.5, 0.75)$
 $(0.75, 1)$



\Rightarrow

$E_1 = 25$	$E_2 = 25$	$E_3 = 25$	$E_4 = 25$	$n = 100$
$O_1 = 20$	$O_2 = 35$	$O_3 = 35$	$O_4 = 10$	

$T_{obs} = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} = 18 \neq 7.81$ $p\text{-value} = 0.0004358$
 we reject H_0

(R) chisq. test ($x = c(20, 35, 35, 10)$, $p = rep(1/4, 4)$)

b)

	< 0.5	$0.25 - 0.5$		
1st	20	35	35	10
2nd	15	25	40	20

$T_{obs} = \sum_{i=1}^2 \sum_{j=1}^4 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 6.0476$

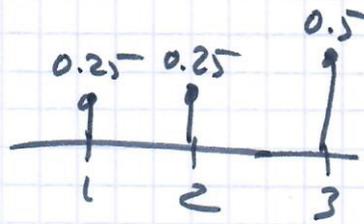
$T \underset{H_0}{\sim} \chi^2_3$ crit. value = 7.81

we do not reject H_0

(R) chisq. test ($x = \text{matrix}(c(20, 15, 35, 35, 40, 10, 20), \text{nrow} = 2)$, $\text{correct} = \text{FALSE}$)

10) known discrete distribution + independence

a) H_0 $\left. \begin{matrix} p_1 = p_2 \\ p_3 = 2p_2 = 2p_1 \end{matrix} \right\} \Rightarrow$



$n=350$: $E_1 = \frac{350}{4} = 87.5$ $O_1 = 150$
 $E_2 = 87.5$ $O_2 = 110$
 $E_3 = 175$ $O_3 = 90$

$T_{obs} = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i} = 51.714 > 5.99$ reject H_0

(R) chisq.test(x=c(150,110,90),p=(0.25,0.25,0.5))

b) $H_0: X \perp Y$

$T_{obs} = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 1.8939 \neq 5.99$
do not reject H_0

$T_{obs} \sim \chi^2_2$

(R) chisq.test(x=matrix(c(150,70,110,50,90,30),nrow=2),
correct=FALSE)