Homework 2

Statistical Inference, Fall 1401

1- (6) Each outcome is equally likely, with probability 1/36

A = {(2,2), (1,3),(3,1)}

B = {(1,5), (5,1), (2,4), (4,2), (3,3)}

C = {((1,2), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,2), (4,2), (5,2),(6,2)}

a. $P(A|C) = \frac{P(A\cap C)}{P(C)} = \frac{\frac{1}{36}}{\frac{11}{36}} = \frac{1}{11}.\ P(A) = \frac{3}{36} \neq P(A|C)$. A and C are not independent

b. $P(B|C) = \frac{P(B\cap C)}{P(C)} = \frac{2}{11}$. $P(B) = \frac{5}{36} \neq P(B|C)$. B and C are not independent

2- (6) $P(F) = 0.55$, $P(C) = 0.08$, $P(FC) = 0.03$

a. $P(F|C) = \frac{0.03}{0.08} = 0.375$

b. $P(C|F) = \frac{0.03}{0.55} \approx 0.055$

3- (9) Part a :

X = the number of people in the sample of 400 adult Richmonders who approve of the President's reaction.

Only two choices, approve or not. Because the sample size is small compared to the population size (all adult Richmonders), it is reasonable to consider the individual responses independent. All have the same probability of success. X is binomial.

Part b:

$P(X \leq 358).\ using\ R\ we\ have:$

```
> pbinom(358, size=400, prob=0.92)
[1] 0.04410268
```

Part c:

$\mu = np = 400 * 0.92 = 368 \qquad \sigma = \sqrt{npq} = \sqrt{400 * 0.92 * 0.08} = 5.426$

Part d:

$P(X \leq 358) = P(Z \leq \frac{358-368}{5.426}) = P(Z \leq -1.843) = 0.0327$

The approximation is not accurate.

4- (6) Let A be the event of playing with an opponent of type i.

$P(A_1) = 0.5, P(A_2) = 0.25, P(A_3) = 0.25$

And B is the event of winnig:

$P(B \mid A_1) = 0.3, P(B \mid A_2) = 0.4, P(B \mid A_3) = 0.5$

by the total probability theorem, the probability of winning is:

$$P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3)$$
$$= 0.5 * 0.3 + 0.25 * 0.4 + 0.25 * 0.5 = 0.375$$

5- (8) Part a:

Poisson distribution is used to describe the number of "events" (such as emails received) happening at some average rate in a fixed interval or volume. Each individual word in a SOP has some very small probability of being word i, and the words are weakly dependent. Here an event means using word i, the average rate is determined by how frequently you use that word overall. It is reasonable to assume that the average rate of occurrence of a particular word is the same for two SOPs by the same author, so we take λ to be the same for X_i and Y_i .

Part b:

Let X be the number of times that "machine learning" is used in the first SOP, and Y be the number of times that it is used in the second SOP. Since X and Y are independent Pois(λ):

$$\begin{aligned}
P(X=0, Y>0) &= P(X=0)\,(1-P(Y=0)) = e^{-\lambda}\left(1-e^{-\lambda}\right) \\
&= e^{-\lambda}\left(1-\left(1-\lambda+\frac{\lambda^2}{2!}-\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}-\ldots\right)\right) \\
&= e^{-\lambda}\left(\lambda-\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}-\frac{\lambda^4}{4!}+\ldots\right).
\end{aligned}$$

6- (8) Part a:

Any given day is rainy with probability p. independent of other days. So the probability that the first day is rainy is p.

Part b:

Same as part a. The probability that days 5 and 8 are rainy is p^2

7- (4) X as a geometric random variable with PMF:

$$P_X(k) = (1-p)^{k-1}p$$

The mean and variance of X are : $E[X] = \frac{1}{p}, Var[X] = \frac{1-p}{p^2}$

8- (6) We use z-table.

$$P(\frac{Z-40}{7}) = 0.95$$

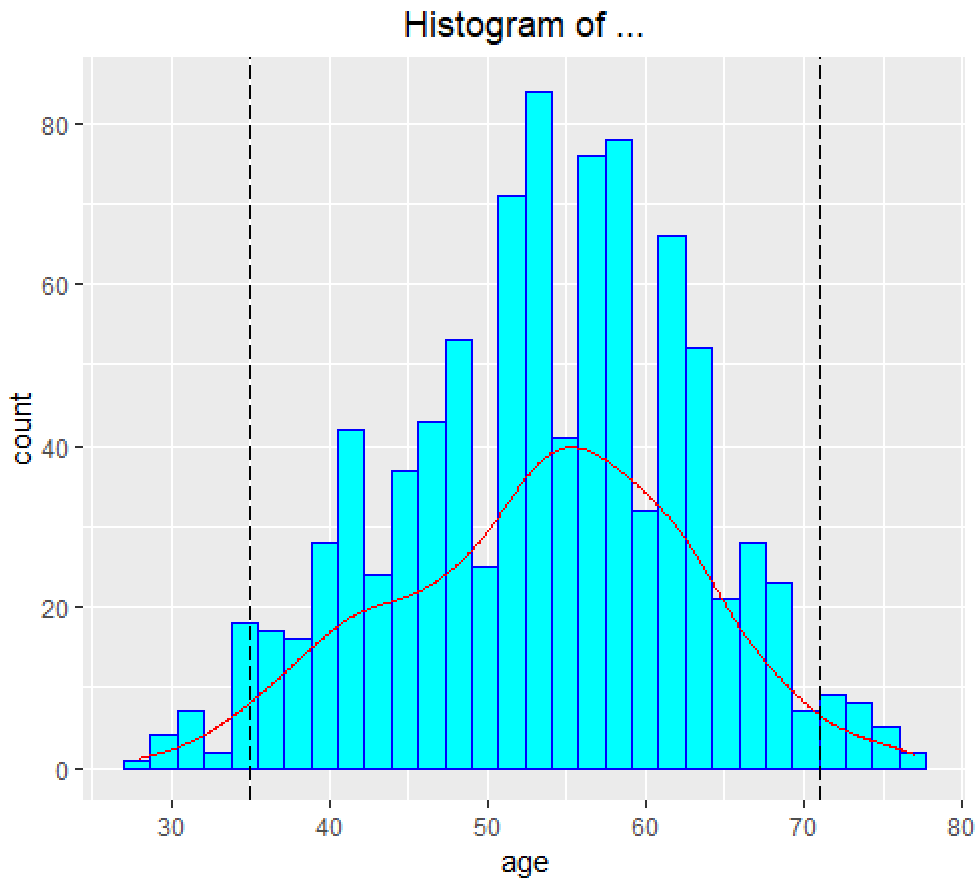
```
> qnorm(0.95)
[1] 1.644854
```

$$Z - 40 = 7 * 1.65$$

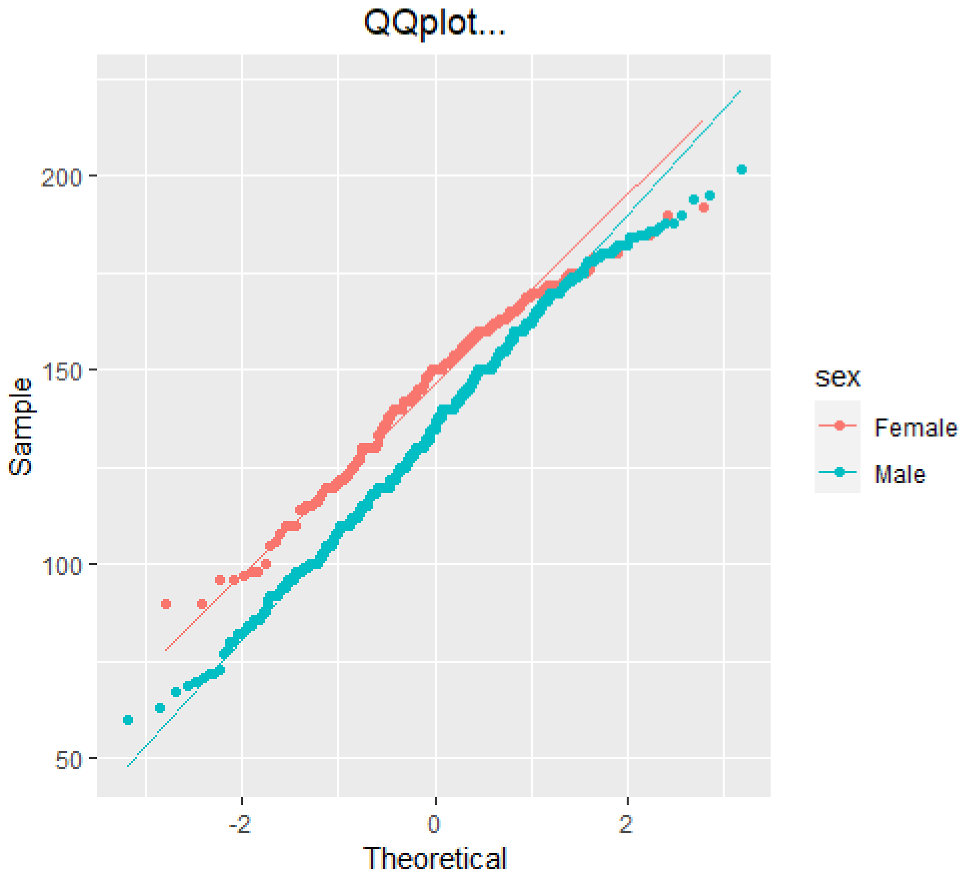
$$Z \approx 51.55$$

So Negar should leave the house at least 51.55 minutes before 1 pm.

9- a. (8)

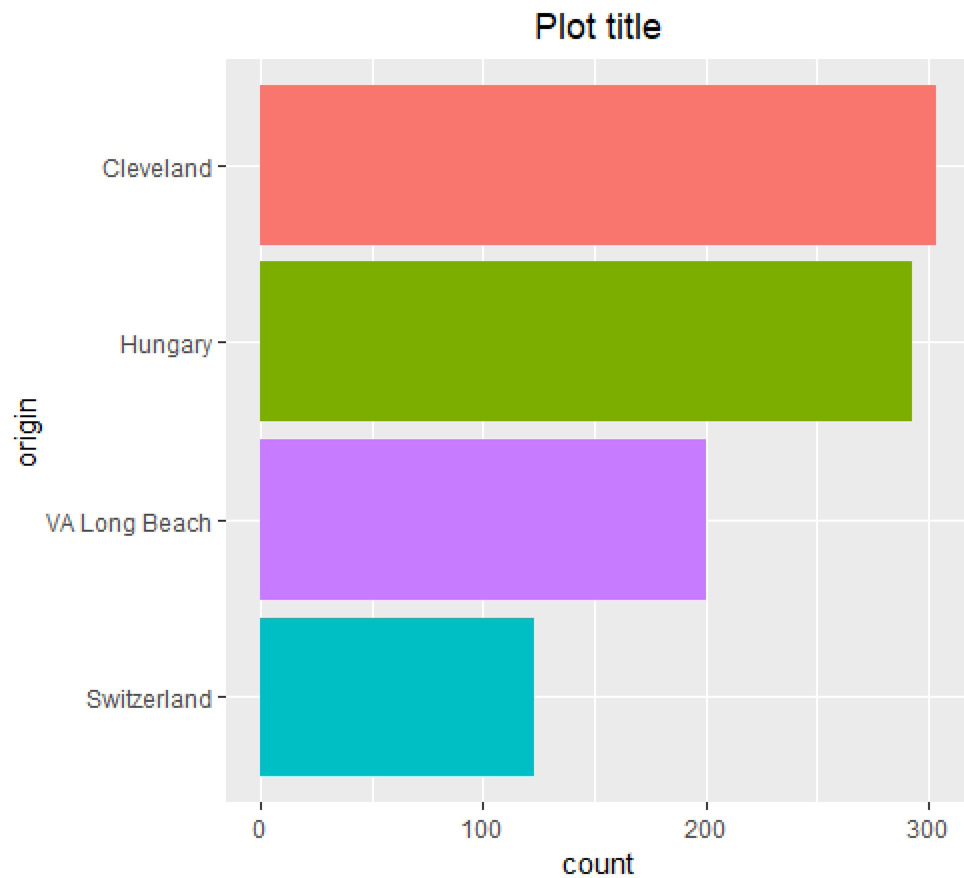


b.(8)



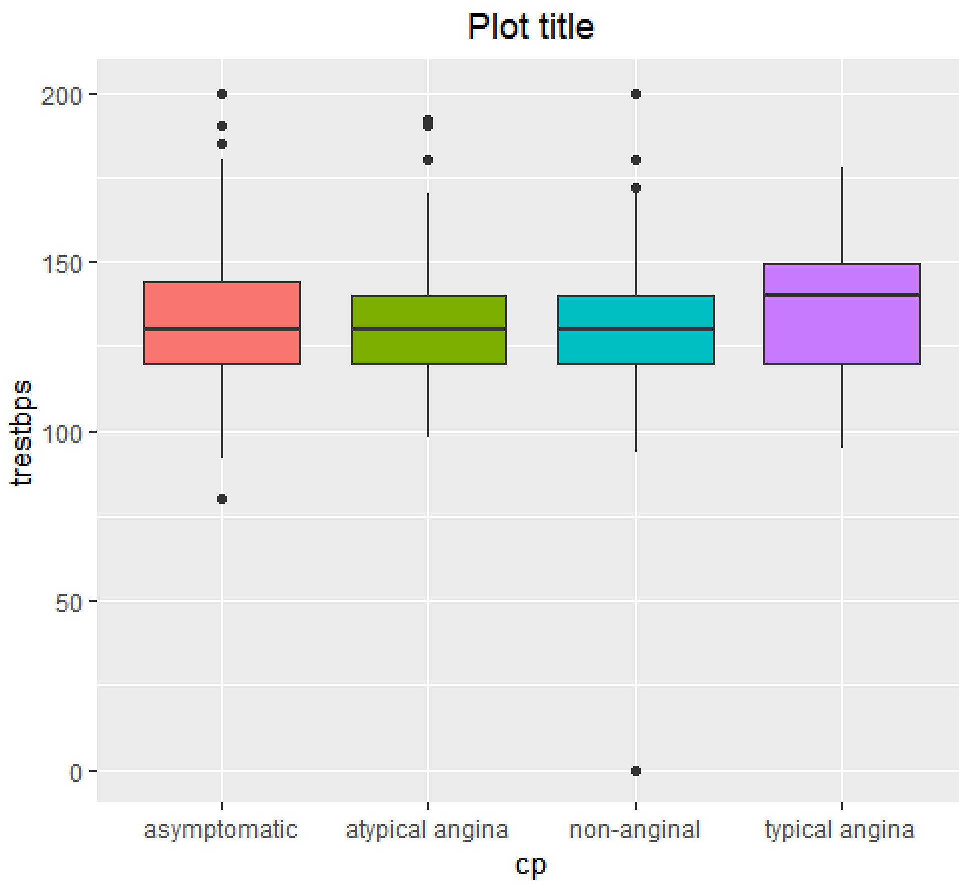
QQplot...
Sample
Theoretical
200
150
100
50
-2
0
2
sex
Female
Male

c. (8)

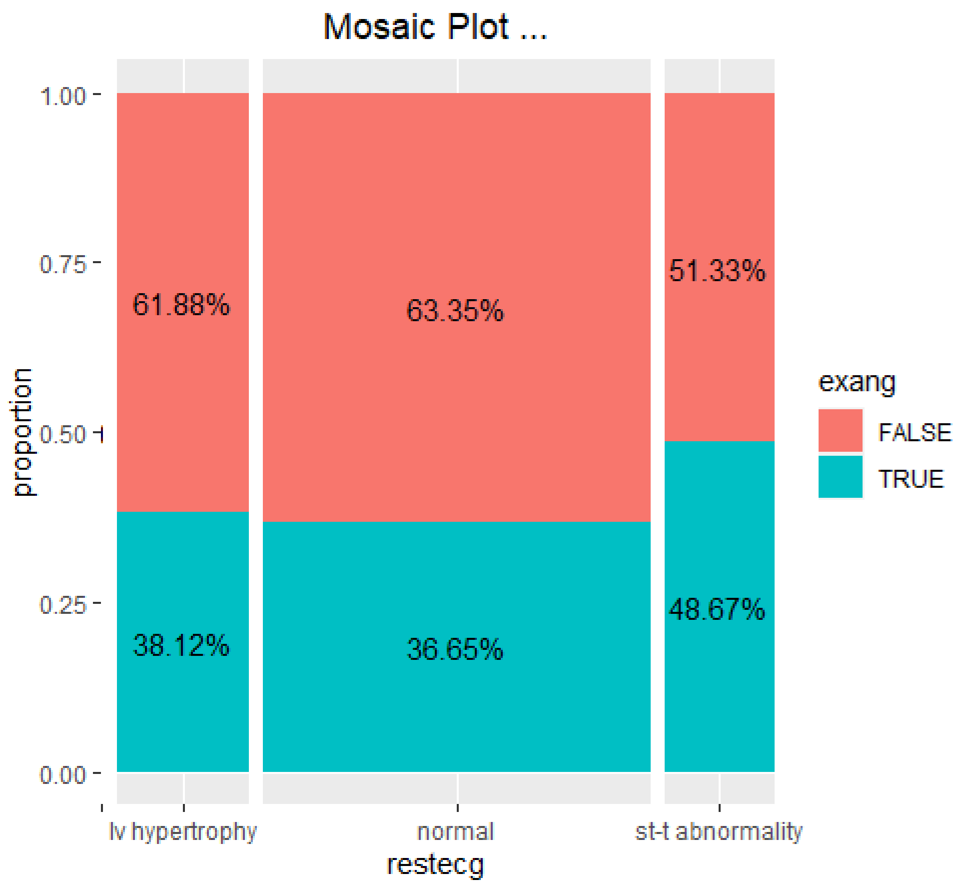


Plot title
Cleveland
Hungary
VA Long Beach
Switzerland
origin
0
100
200
300
count

d. (8)


Plot title
200
150
100
50
0
trestbps
asymptomatic
atypical angina
non-anginal
typical angina
cp

e. (15)



Mosaic Plot ...
proportion
1.00
0.75
0.50
0.25
0.00
61.88%
38.12%
63.35%
36.65%
51.33%
48.67%
lv hypertrophy
normal
st-t abnormality
restecg
exang
FALSE
TRUE