Problem 1

1. Order the 6, 830 p-values from the smallest to the largest. Plot the ordered p-values as the y-axis and their corresponding ranks as the x-axis. If all the 6, 830 null hypotheses are true, the p-values should be uniformly distributed. Are the observed p-values likely from a uniform distribution?

\*\*Solution:\*\* Yes, it is uniform distribution.

2. What will be the p-value cutoﬀ if you would like to control the Family Wise Error Rate under 0.05 using the Bonferroni correction? How many genes will be identiﬁed as differentially

expressed using this cutoﬀ?

\*\*Solution:\*\* The cutoff will be 0.05/6830=7.32\*10^-6. None of the genes will be idetnified as differentially.

3. What will be the p-value cutoﬀ if you would like to control the False Discovery Rate under 0.05

using the Benjamini-Hochberg procedure? How many genes will be identiﬁed as diﬀerentially

expressed using this cutoﬀ?

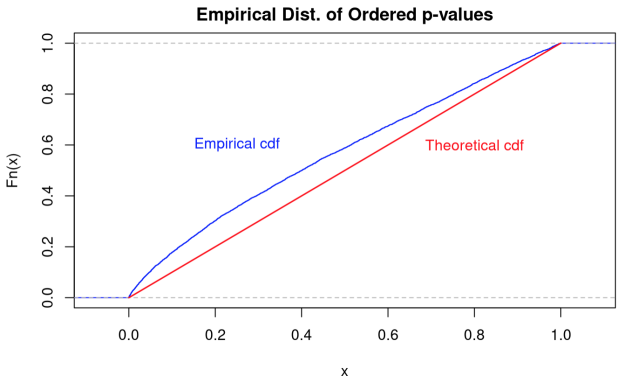
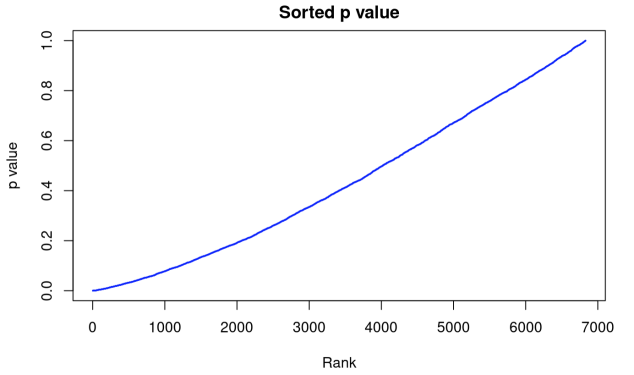
\*\*Solution:\*\* The cutoff is 0. Because the null hypothesis is rejected by BH procedure.

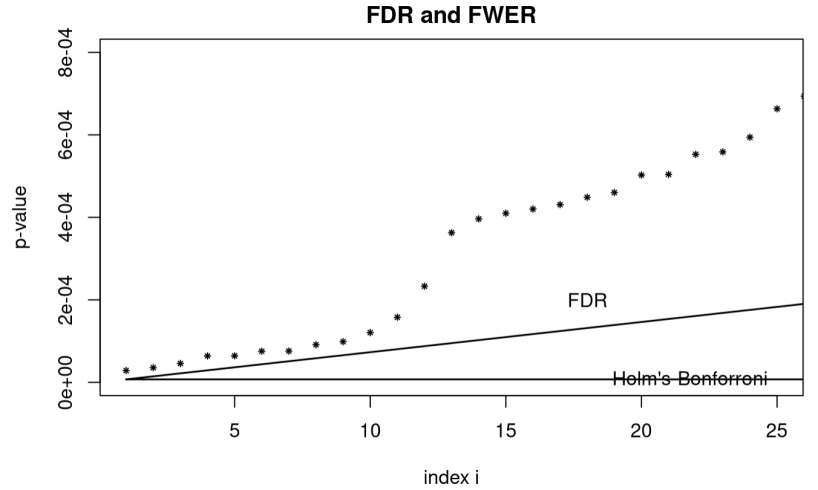
4. What will be the p-value cutoﬀ if you would like to control the Per-Comparison Error Rate

(i.e., signiﬁcance level) under 0.05? How many genes will be identiﬁed as diﬀerentially expressed

using this cutoﬀ?

\*\*Solution:\*\* The cutoff is 0.05. And there are 707 genes will be identified as differentially expressed.





Problem 2

Use the NCI60 data, give an example to each of the following data analysis tasks (please refer to the Leek and Peng, Science (2015) articl). Note: this is an open question;diﬀerent people may give diﬀ erence examples to the same task.

\*\*Solutions:\*\*

1. Descriptive study

In this dataset, there are 6830 columns and 64 rows.

\*\*Solutions:\*\*

2. Exploratory study

Just calculate the genes' mean

\*\*Solutions:\*\*

3. Inferential study

Using different gene to get a SLM and see if there is any linear relationship.

\*\*Solutions:\*\*

4. Predictive study

Use different model or plot to distinguish each gene.

\*\*Solutions:\*\*

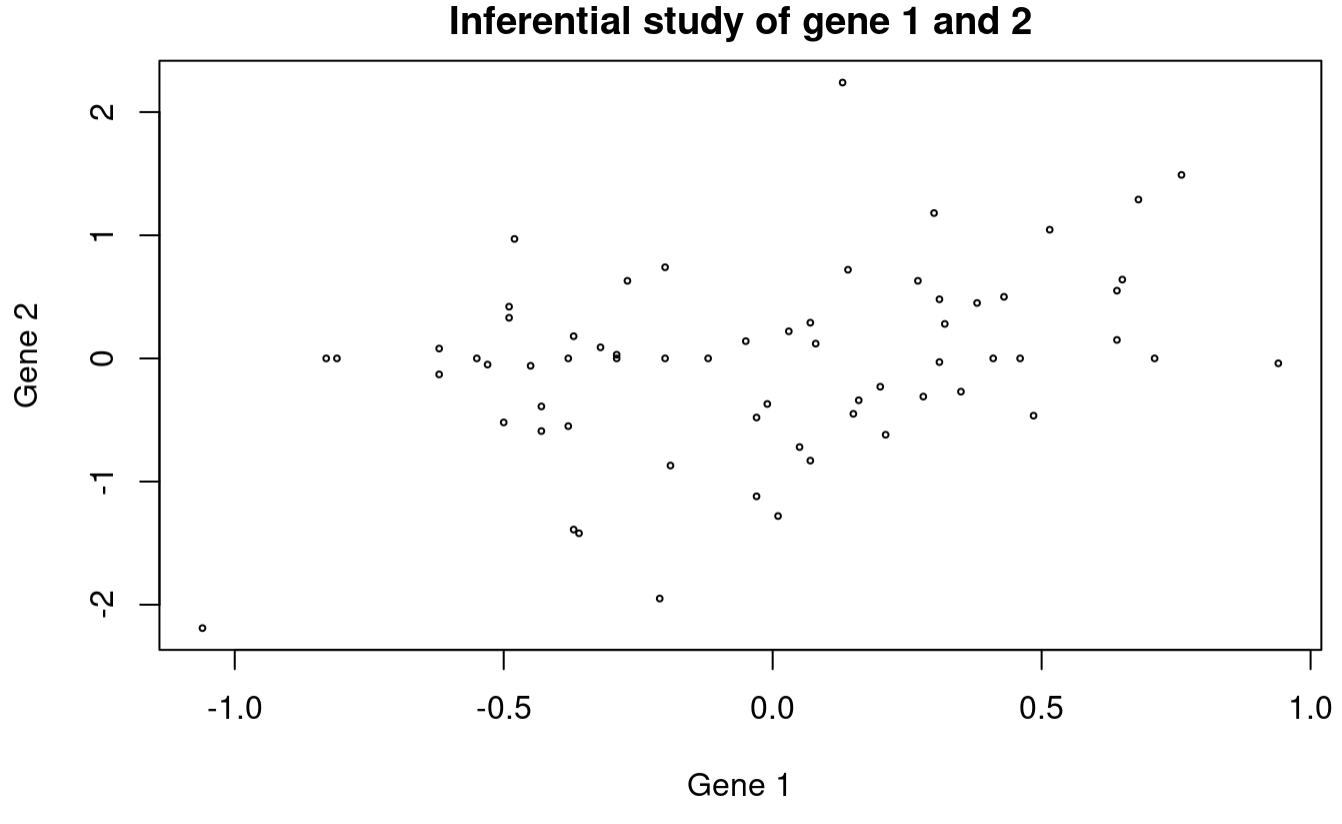
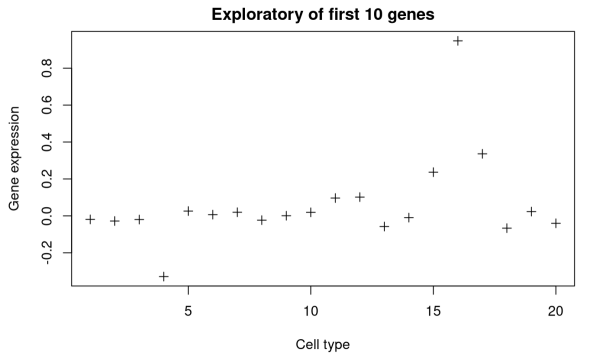
5. Causal study

Use different model or plot to see different between to genes.

\*\*Solutions:\*\*

6. Mechanistic study

Use model or plot to see a genes experssion among other type.

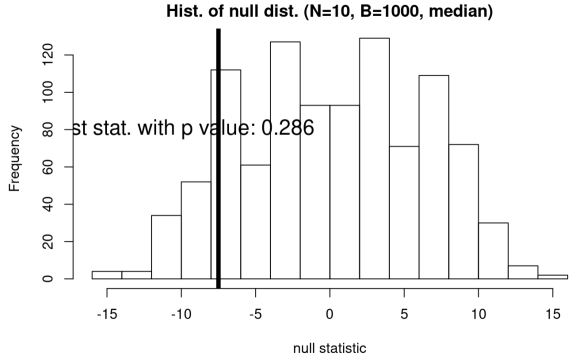
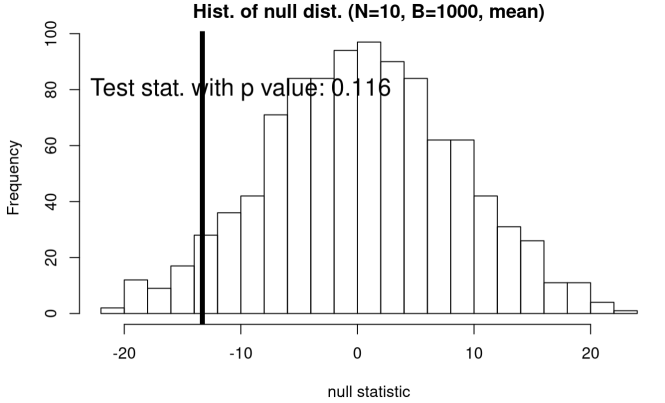


Problem 3

\*\*Solution:\*\*

For 3.1 p-value is 0.116

For 3.2 p-value is 0.286



Problem 4

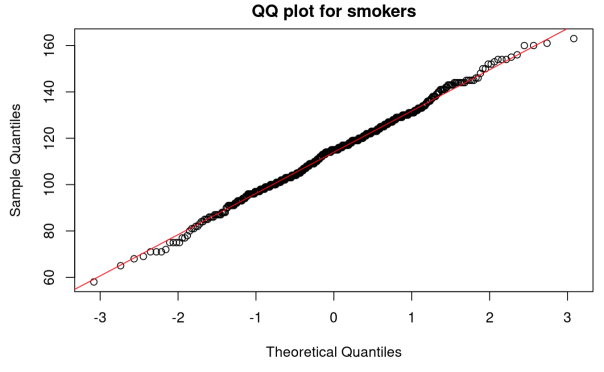
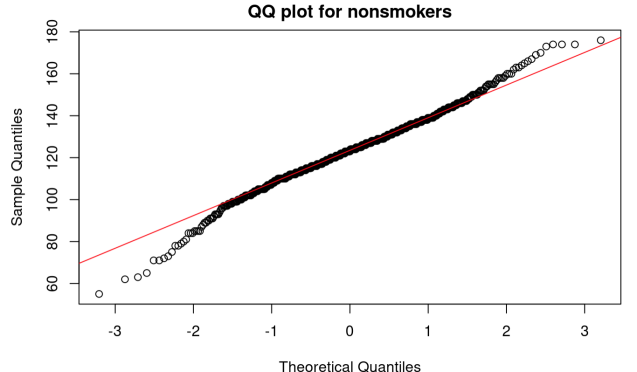
\*\*Soluiton 4.1:\*\*

For smokers, the data is normal distributed.

For non-smokers, the data is skewed.

\*\*Soluiton 4.2:\*\*

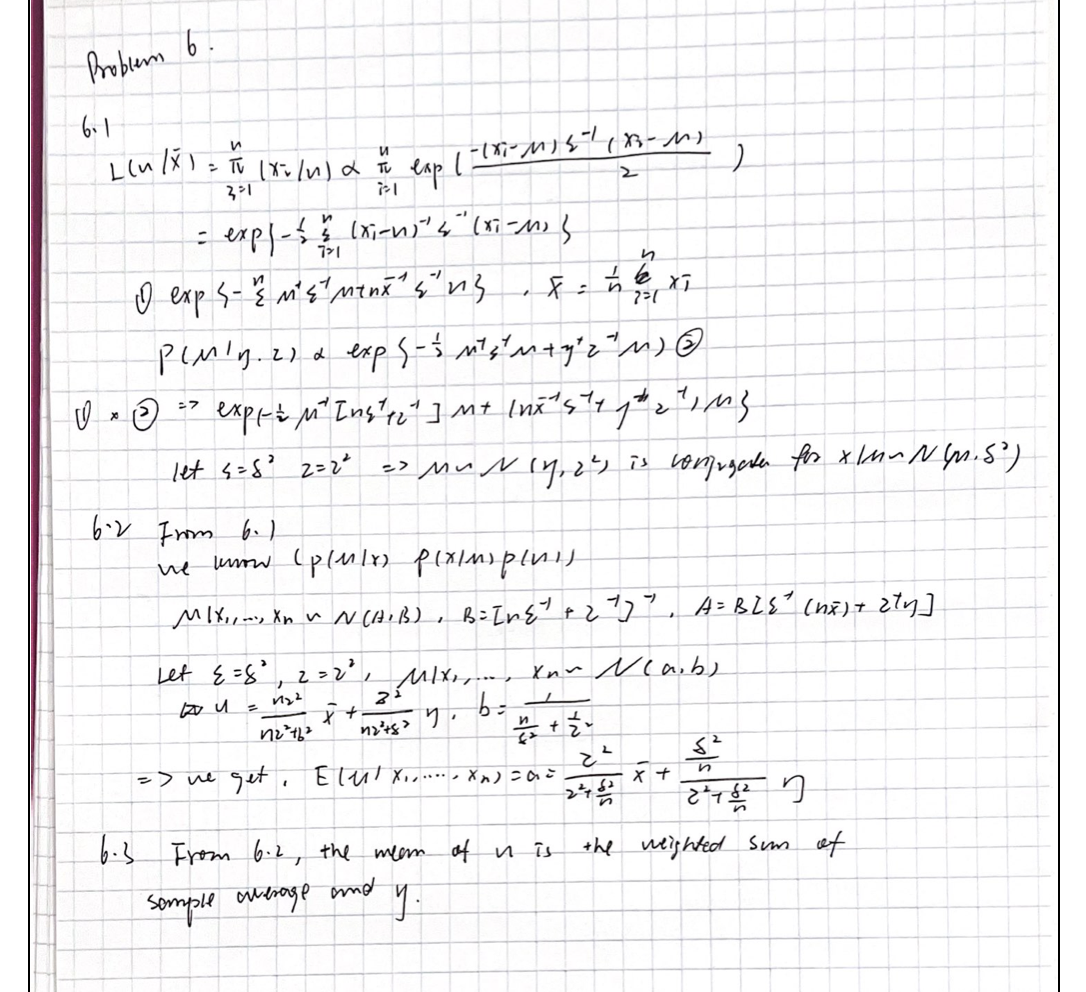
The p-value is 0.056, which shows that there is not significant and we do not reject the null hypothesis.



Problem 5

P = (0.99/3900)/((0.99/3900)+(0.01\*3899/3900))=2.46%

Problem 6



Problem 7

\*\*Soluiton:\*\*

The distribution of simple mean will concertrate on a simple point.

