Financial Time Series I and Methods of Statistical Prediction

Homework 1: Review on Basics

1. (a) The null hypothesis might be "This person is innocent." It is better to release a guilty person than to convict an innocent one. So we would rather to make Type II error than Type I error that we need to minimize Type I error.

   (b) The null hypothesis might be "This drug will care an illness," and one agrees that spending time and money on a useless drug. So we need to use a more strictly judgement when we think that it does a go job but it actually does not. So we need to minimize Type II error.

2. Recall that Type I error is the probability of "reject H_0 where H_0 holds" and Type II error is the probability of "not reject H_0 where H_a holds". And the null hypothesis H_0 is that " The individual came from Sample #1 ". We now find the probabilities of Type I and Type II error for four different strategies.

   (a) P [Type I error] = P [not Caucasian | H_0]= $\frac{10}{25} = 0.4$

   P [Type II error] = P [Caucasian | H_a] = $\frac{3}{20} = 0.15$

   (b) P [Type I error] = P [neither Caucasian nor Oriental| H_0]= $\frac{4}{25} = 0.16$

   P [Type II error] = P [Caucasian or Oriental | H_a] = $\frac{6}{20} = 0.30$

   (c) P [Type I error] = P [Native American | H_0]= $\frac{1}{25} = 0.04$

   P [Type II error] = P [not Native American | H_a] = $\frac{16}{20} = 0.8$

   (d) P [Type I error] = P [African American | H_0]= $\frac{3}{25} = 0.12$

   P [Type II error] = P [not African American | H_a] = $\frac{10}{20} = 0.5$

3. (a) The histogram can be obtained by "R". Its command is as follows.

   >data<-c(5,4,4,5,4,3,3,4,3,2,3,3,3,2,2,1,3,2,1,3)

   >hist(data)

   (b) Histogram gives the frequency on producing various number of defective items. We can know how many times the event that find n defective items happen during 20 days.

   (c) The scatter plot gives daily change of the number of defective items.

4. (a) We put this problem in the framework of hypothesis testing by setting them as "$H_0 : p \leq 0.1$ v.s $H_a : p > 0.1$". Let $\hat{p}$ be the estimator of the probability where

p is the unknown percentage of graduate students looking for house on campus. And we know $\hat{p} = \frac{62}{481}$

by C.L.T

$$\frac{(\hat{p} - 0.1)}{\sqrt{\frac{0.1 \times 0.9}{481}}} = 2.111 > Z_{0.05} = 1.64$$

I have enough evidence to reject H_0 under significance level $\alpha = 0.05$. That is to say, the percentage of graduate students who are looking for housing on campus is larger then 10%.

(b) We use the worst case to analyze. Assume that the answer of these 19 graduate students who did not respond to the survey will be that they do not look for housing on campus. The test will change to

$$\frac{(\hat{p} - 0.1)}{\sqrt{\frac{0.1 \times 0.9}{500}}} = 1.789 < Z_{0.05} = 1.64$$

where $\hat{p} = \frac{62}{500}$. Then the final recommendation of my decision will not be changed.

5. The hypothesis of this problem can be set as follows

H_0 : For those hikers who are novice or experienced, the direction chosen by the hiker will be independent of whether he is novice or experienced.

H_a : For those hikers who are novice or experienced, the direction chosen by the hiker will be dependent of whether he is novice or experienced.

$$\begin{aligned}
\sum \frac{(o_i - e_i)^2}{e_i} &= \frac{(20 - \frac{120\times 30}{200})^2}{\frac{120\times 30}{200}} + \frac{(50 - \frac{120\times 80}{200})^2}{\frac{120\times 80}{200}} + \frac{(50 - \frac{120\times 90}{200})^2}{\frac{120\times 90}{200}} \\
&\quad + \frac{(10 - \frac{80\times 30}{200})^2}{\frac{80\times 30}{200}} + \frac{(30 - \frac{80\times 80}{200})^2}{\frac{80\times 80}{200}} + \frac{(40 - \frac{80\times 90}{200})^2}{\frac{80\times 90}{200}} \\
&= 1.5046 < X^2_{(2-1)(3-1),0.05} = 5.99
\end{aligned}$$

I do not have enough evidence to reject H_0 under significance level $\alpha = 0.05$. So these data do not provide convincing evidence of an association between the level of hiking expertise and the direction the hiker would head if lost.

6. According to the problem, the diameter of the pearl found in an oysters can be described by a random variable X which is N (8,4).

(a) Use C.L.T.

$$P[7 \le X \le 9] = P[\frac{7-8}{2} \le \frac{X-8}{2} \le \frac{9-8}{2}] = p[-0.5 \le Z \le 0.5] = 0.383$$

(b) Let the probability of an oyster meets the request of the customer is denoted by p. Note that $p = 0.384$ .Therefore the number of the oysters he needs to open can be described by a random variable Y which is Negative-Bin(r,p).

$$E(Y) = \frac{r}{p} = 2\frac{2}{0.383} = 5.2219$$

So we can choose n=6.

7. (a) Method 1 : We can do this hypothesis testing : $H_0 : p \leq 0.5$ vs $H_a : p > 0.5$. The estimation of the data is $\hat{p} = 0.563$.

$$\frac{0.563 - 0.5}{\frac{0.5\times0.5}{20\times100}} = 5.635 > Z_{0,05}$$

Method 2 : In this problem, need to test $H_0 : \mu \leq 50$ vs $H_a : \mu > 50$. The standard error of the data is $s^2 = \frac{\sum_{i=1}^{20}(y_i - \bar{y})^2}{19}$

$$\frac{56.3 - 50}{s} = 1.333 > t_{19,0.05} = 1.729$$

So, we do not have enough evidence to reject H_0 under significance level $\alpha = 0.05$. This means that the training is not useful to people.

(b) No, the test in (a) can not provide evidence that this training is effective in improving a person's ability.

Method1 : In this problem, we can try to use sign test. First, set up the hypothesis : $H_o : p_1 = p_2$vs $H_a : p_1 \neq p_2$. Then, define $r = " + "$ if $X > Y$, $r = " - "$ if $X < Y$, and $r = 0$ if $X = Y$. X is the number of before training and Y is the number of after training. There are 11 $" + "$ and 8 $" - "$. So, define T is the number of $" - "$.

$$P(T \geq 11|n = 19, p = 0.5) > 0.05)$$

Method 2 : We should use paired t-test in this problem. First, set up the hypothesis.

H_0 : the mean of the difference of the paired data is smaller or equal to 0

H_a : the mean of the difference of the paired data is larger to 0.

Let the difference be d. Then the standard error is $s^2 = \frac{\sum_{i=1}^{2} 0(d_i - \bar{d})^2}{19}$.

$$\frac{0.3 - 0}{S} = 0.1962 < t_{19,0.05} = 1.729$$

But if you do not think that the effect between individuals will be small, you can use two sample t-test.

So, we do not have evidence to reject H_0 under significance level $\alpha = 0.05$.

(c) Plot scatter plot and discuss whether you should consider individual effects.