Statistics I, Fall 2012
Homework 12

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

Note. For each problem, define the notations you use, if any.

1. (15 points; 5 points each) Your food factory produces one product that contains one ingredient. A study shows that if the amount of ingredient is larger than 5 mg, eating the product will harm one's health. Suppose you want to test whether your products in average contain more than 5 mg of that ingredient. If so, you should stop using the current production system in producing the product and invest a lot in improving your factory.
   (a) Write down the research hypothesis.
   (b) If your first priority is to save money, what should be the statistical hypothesis?
   (c) If your first priority is to protect your consumers, what should be the statistical hypothesis?
2. (15 points; 5 points each) As a government officer, you are asked to study whether teenagers spend more on fast food than adults. If the answer is true, you should suggest the government to incentivize more non-fast-food restaurant specifically targeting at teenagers. Let μ_1 and μ_2 be the average amount of money spent on fast food per teenager and adult, respectively.
   (a) Write down the statistical hypothesis.
   (b) Explain how did you construct your null and alternative hypotheses.
   (c) Suppose the significance level $\alpha = 0.05$ and the p-value of your survey is 0.07, write down your conclusion formally.
3. (15 points; 5 points each) You have been selling a product at $50 per unit for one year. It seems to you that the demand is higher than the supply, so you are considering whether you should increase the unit price. In order to understand how much a consumer is willing to pay for this product, you randomly sampled 1000 consumers and asked this question. If the average willingness-to-pay is around $50, you will not change the price. You will increase your price if and only if the average willingness-to-pay is higher than $60 (not $50!).
   (a) Write down the statistical hypothesis.
   (b) Let the significance level be α, write down an equation to link the significance level and the error probability that you want to control in the hypothesis testing.
   (c) Suppose $\alpha = 0.05$ and the p-value of your survey is 0.04, write down your conclusion and the managerial implication formally.
4. (10 points; 5 points each) In a hypothesis testing regarding the population mean, the population is normal and the population variance is known to be 1600. The sample size is 16. The population mean is assumed to be 70.
   (a) Suppose we want to do a two-tailed test. Write down the statistical hypothesis. If the observed sample mean is 90, what is the p-value?
   (b) Suppose we want to do a one-tailed test and we want to know whether the population mean is higher than 70. Write down the statistical hypothesis. If the observed sample mean is 60, what is the p-value?
5. (20 points; 2 points each) Answer these true/false questions without providing any explanation.
   (a) The significance level can be calculated based on a p-value.

(b) The p-value can be calculated based on a significance level.

(c) The critical values can be calculated based on a significance level.

(d) The critical values can be calculated based on a p-value.

(e) For a one-tailed test, once the p-value is lower than the significance level, we can reject H_0 and conclude that, with the specified significance level, H_a is true.

(f) For a one-tailed test, once the p-value is higher than the significance level, we can accept H_0 and conclude that, with the specified significance level, H_0 is true.

(g) If H_0 is true, H_a cannot be true.

(h) When we reject H_0, the significance level α is the probability for H_0 to be true.

(i) The p-value is the probability of rejecting H_0 when H_0 is true.

(j) For a one-tailed test regarding the population mean, suppose the p-value is smaller than α. In this case, observing a smaller p-value implies that the the true population mean deviates more from the assumed value.

6. (20 points; 5 points each) For one study regarding the population mean μ, Alice and Bob observe the same sample mean $\bar{x} = 90$. However, they adopt two different statistical hypotheses

$$\begin{array}{l} H_0 \colon \mu \geq 100 \\ H_a \colon \mu < 100 \end{array} \quad \text{and} \quad \begin{array}{l} H_0 \colon \mu \leq 100 \\ H_a \colon \mu > 100. \end{array}$$

Alice uses the first one while Bob thinks the second one is right. Suppose based on the significance level is $\alpha = 0.05$, the rejection region found by Alice is $(-\infty, 95)$.

(a) Write down Alice's conclusion formally.

(b) What should be Bob's rejection region?

(c) Write down Bob's conclusion formally.

(d) Are their conclusions equivalent? If yes, explain why. If no, what are the differences?

7. (5 points) Consider a one-tailed test regarding the population mean μ.[1] Suppose we write the statistical hypothesis as

$$\begin{aligned} H_0 &\colon \mu = \mu_0 \\ H_a &\colon \mu < \mu_0, \end{aligned}$$

where μ_0 is our original belief of μ. To construct a rejection rule, we calculate a distance d such that

$$\Pr(\overline{X} < \mu - d | \mu = \mu_0) = \alpha, \tag{1}$$

where α is the significance level. By doing so, the error probability is controlled to be α. Alternatively, many statisticians prefer to say that the distance is the *minimum* distance that can limit the error probability $\Pr(\overline{X} < \mu - d | \mu = \mu_0)$ to be below α:

$$d' = \min \left\{ x \middle| \Pr(\overline{X} < \mu - x | \mu = \mu_0) \leq \alpha \right\}. \tag{2}$$

Prove or explain why (1) and (2) are equivalent, i.e., $d = d'$.[2]

[1] The idea illustrated in this problem also applies to two-tailed tests.
[2] As long as your explanations are good enough, you do not need to write a mathematical proof.