Economics 325–003 **Term I 2015**

Introduction to Empirical Economics **Hiro Kasahara**

Final Exam

1. State whether each of the following is true or false. No explanation necessary.

   (a) (5 points) If the null hypothesis is rejected based on sample evidence using the test at the significance level $\alpha = 0.1$, the research has absolutely proven that the null hypothesis is false without any doubts.

   (b) (5 points) If the p-value of a two sided test for the mean of a population is .005, then the null hypothesis will be rejected at the 1% significance level.

   (c) (5 points) If a null hypothesis is rejected at the 5% significance level, then using the same data, the null hypothesis will be rejected at the 10% significance level.

   (d) (5 points) Let $\{X_1, X_2, ..., X_n\}$ be n observations, each of which is randomly drawn from a distribution with mean μ and variance σ^2. Let $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2}$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Then, the distribution of a statistic $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ is always given by t-distribution with the degree of freedom $n-1$.

   (e) (5 points) Given a realized sample, the confidence interval contains the population parameter with probability either one or zero.

   (f) (5 points) As the sample size increases to infinity, the variance of the sample mean approaches zero.

2. (6 points) Let b be a constant. Prove that $E[(X-b)^2] = E(X^2) - 2bE(X) + b^2$. What is the value of b that gives the minimum value of $E[(X-b)^2]$?

3. (6 points) A company produces electrical devices operated by a thermostatic control. According to the engineering specifications, the variance of the temperature at which these controls actually operate should not exceed 4.0 degrees Fahrenheit. We assume that the temperature is normally distributed. For a random sample of 25 of these controls, the sample variance of operating temperatures was $s^2 = 2.36$ degrees Fahrenheit. Compute the 95 percent confidence interval for the population variance σ^2.

4. Let $\{X_1, X_2, X_3, X_4\}$ be $n = 4$ observations, each of which is randomly drawn from normal distribution with mean μ and variance σ^2. The value of μ is not known while σ^2 is known and equal to 100. We are interested in testing the null hypothesis $H_0 : \mu \geq 10$ against $H_1 : \mu < 10$. Consider the following two

different test statistics: (i) $\bar{X} = (1/4)(X_1+X_2+X_3+X_4)$ and (ii) $\hat{X} = 0.1X_1+0.1X_2+0.1X_3+0.7X_4$. Suppose that the realized values of $\bar{X}$ and $\hat{X}$ are given by $\bar{X} = 2.0$ and $\hat{X} = 1.0$, respectively.

(a) (6 points) Compute the p-value for testing the null hypothesis $H_0 : \mu \geq 10$ against $H_1 : \mu < 10$ *using the test statistic* $\bar{X}$ and test the null hypothesis at the significance level $\alpha = 0.1$.

(b) (6 points) Compute the mean and the variance of a statistic $\hat{X}$ when $\mu = 10$ and $\sigma^2 = 100$.

(c) (6 points) Test $H_0 : \mu \geq 10$ against $H_1 : \mu < 10$ *using the test statistic* $\hat{X}$ at the significance level $\alpha = 0.1$.

(d) (8 points) Compute (i) the power of test using *using the test statistic* $\bar{X}$ at the significance level $\alpha = 0.1$ when the true value of μ is equal to 5 and (ii) the power of test *using the test statistic* $\hat{X}$ at the significance level $\alpha = 0.1$ when the true value of μ is equal to 5. Based on the power comparison, which test statistics, $\bar{X}$ or $\hat{X}$, do you recommend using for hypothesis testing?

5. (6 points) Show that the expected value of the sample variance is equal to σ^2 $\underline{\text{when } n = 2}$, i.e.,

$$E[s^2] = \sigma^2,$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = (X_1-\bar{X})^2+(X_2-\bar{X})^2$ with $\bar{X} = (X_1+X_2)/2$ and $\sigma^2 = E[(X_i-\mu)^2]$.

6. (6 points) Suppose X_i for $i = 1, ..., n$ is randomly drawn from a distribution with variance σ^2. Let $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$ be the sample average. Prove that $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

7. Table IAverage Amount of Tobacco Smoked Daily Over the 10 yearstable.1 reports the number of non-smokers and the numbers of smokers with the daily average of 1-14 cigarettes, 15-24 cigarettes, and more than 25 cigarettes among 1357 patients with lung-cancer and the number of smokers among 1357 patients with other diseases. Suppose that Doll and Hill randomly sampled 1357 patients with lung-cancer from a population of patients with lung-cancer and 1357 patients with other diseases from a population of patients with other diseases.

   Denote the proportions of smokers with more than 25 cigarettes per day for patients with lung-cancer and for patients with other diseases by p_x and p_y. Our concern is whether heavy smoking is associated with lung cancer in population so that we are interested in the population difference $p_x - p_y$.

   (a) (6 points) What is the estimator of the difference between two population proportions of smokers, $p_x - p_y$?

Table I: Average Amount of Tobacco Smoked Daily Over the 10 years

Disease Group	No. of Non-Smokers	No. of Smokers with the Daily Average of 1-14 Cigs.	15-24 Cigs.	25+ Cigs.
Men:				
1357 lung-cancer	7	55	964	331
1357 other diseases	61	129	1001	166

Notes: Computed from Table V of Doll and Hill (1952).

(b) (6 points) Compute the 95 percent confidence interval of the difference between two population proportions of smokers.

(c) (8 points) Test the null hypothesis that $H_0 : p_x \leq p_y$ against $H_1 : p_x > p_y$ at the 5 percent significance level.