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.

   ANSWER: F – type I error probability is not equal to zero.

   (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.

   ANSWER: T

   (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.

   ANSWER: T

   (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 $n-1$.

   ANSWER: F — because t statistic does not have t-distribution when X_i is not normally distribution.

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

   ANSWER: T

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

   ANSWER: T

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 $E[(X-b)^2]$?

ANSWER: Because $(X-b)^2 = X^2 - 2bX + b^2$, we have

$$E[(X-b)^2] = E[X^2 - 2bX + b^2] = E[X^2] - 2bE(X) + b^2.$$

Noting that $E[X^2] - 2bE(X) + b^2$ is a quadratic convex function of b, we may find the minimum by differentiating $E[(X-b)^2]$ with respect to b and set $\frac{\partial}{\partial b}E[(X-b)^2] = 0$, i.e.,

$$\frac{\partial}{\partial b}E[(X-b)^2] = -2E(X) + 2b = 0,$$

and, therefore, setting the value of b equal to

$$b = E(X)$$

minimizes $E[(X-b)^2]$. Grading: 3 points for the proof of $E[(X-b)^2] = E(X^2) - 2bE(X) + b^2$ and 3 points for $b = E(X)$, where the latter 3 points will be given as long as students got the right answer.

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.

   ANSWER: the distribution of $\frac{(n-1)s^2}{\sigma^2}$ is given by chi-square distribution with $(n-1)$ degrees of freedom. Let $\chi^2 n - 1$ be a random variable distributed by chi-square distribution with $(n-1)$ degrees of freedom and let $\bar{\chi}^2_{n-1,\alpha}$ be the upper critical value such that $\Pr(\chi^2_{n-1} > \bar{\chi}_{n-1,\alpha}) = \alpha$. Then, $\Pr(\bar{\chi}^2_{n-1,1-\alpha/2} \leq \frac{(n-1)s^2}{\sigma^2} \leq \bar{\chi}^2_{n-1,\alpha/2}) = \alpha$ and, therefore, $\Pr\left(\frac{(n-1)s^2}{\bar{\chi}^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\bar{\chi}^2_{n-1,1-\alpha/2}}\right)$. Now, when $\alpha/2 = 0.025$, chi-square table gives $\bar{\chi}^2_{24,\alpha/2} = 39.364$ and $\bar{\chi}^2_{24,1-\alpha/2} = 12.401$ so that the lower bound of the 95 percent CI is $\frac{(n-1)s^2}{\bar{\chi}^2_{n-1,\alpha/2}} = \frac{24\times 2.36}{39.364} = 1.439$ and the upper bound is $\frac{(n-1)s^2}{\bar{\chi}^2_{n-1,1-\alpha/2}} = \frac{24\times 2.36}{12.401} = 4.567$. Therefore, the 95 percent CI is $[1.439, 4.567]$.

   Grading: give 3 points if the logic is right but students make computational mistakes. No partial point unless it is clear from the answer shows that students understand how to construct the confidence interval for sample variance.

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 $\alpha = 0.1$.

ANSWER: The null distribution of $\bar{X}$ when $\mu = 10$ is given by $N(10, \sigma^2/n)$ with $\sigma^2/n = 25$ so that $(\bar{X} - 10)/5$ is a standard normal random variable. $\Pr(Z < (2 - 10)/5)$ $=\Pr(Z < -1.6) = 1 - .9452 = 0.0548$. Thus, the p-value is 0.0548. Therefore, we reject 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 $\sigma^2 = 100$.

ANSWER: $E[\hat{X}] = \mu$ and $Var(\hat{X}) = [(0.1)^2 + (0.1)^2 + (0.1)^2 + (0.7)^2] \times 100 = 0.52 \times 100 = 52$.

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

ANSWER: The null distribution of $\hat{X}$ when $\mu = 10$ is given by $N(10, 52)$ so that $(\hat{X} - 10)/\sqrt{52}$ is a standard normal random variable. The p-value is given by $\Pr(Z < (1 - 10)/\sqrt{52})$ $=\Pr(Z < -9/7.2111)$ $=\Pr(Z < -1.25) = 0.1056$. Thus, we do not reject the null hypothesis at $\alpha = 0.1$.

(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?

ANSWER: The distribution of $\bar{X}$ when $\mu = 5$ is given by $N(5, \sigma^2/n)$ with $\sigma^2/n = 25$ so that $(\bar{X} - 5)/5$ is a standard normal random variable. The critical value of the test at $\alpha = 0.1$ based on $\bar{X}$ is given by $10 - 1.28 \times 5 = 3.6$ so that the power (=the probability of correctly reject H_0 when $\mu = 5$) is $\Pr((\bar{X} - 5)/5 < (3.6 - 5)/5) = \Pr(Z < (3.6 - 5)/5) = \Pr(Z < -0.28) = 1 - 0.6103 = 0.3897$.

The distribution of $\hat{X}$ when $\mu = 5$ is given by $N(5, 52)$ so that $(\bar{X} - 5)/7.2111$ is a standard normal random variable. The critical value of the test at $\alpha = 0.1$ based on $\hat{X}$ is given by $10 - 1.28 \times 7.2111 = 0.7698$ so that the power (=the probability of correctly reject H_0 when $\mu = 5$) is $\Pr((\hat{X} - 5)/7.2111 < (0.7698 - 5)/7.2111) = \Pr(Z < (0.7698 - 5)/7.2111) = \Pr(Z < -0.59) = 1 - 0.7224 = 0.2776$.

Thus, the power is higher for the test based on $\bar{X}$ than for the test based on $\hat{X}$, and we recommend using the test based on $\bar{X}$.

5. (6 points) Show that the expected value of the sample variance is equal to σ^2 <u>when $n = 2$</u>, 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 $\sigma^2 = E[(X_i-\mu)^2]$.

ANSWER: When $n = 2$, $\bar{X} = \frac{X_1+X_2}{2}$ and $(X_1-\bar{X})^2 = (X_2-\bar{X})^2 = \frac{(X_1-X_2)^2}{4}$ so that $s^2 = \frac{1}{2-1}\left(\frac{(X_1-X_2)^2}{4}+\frac{(X_1-X_2)^2}{4}\right) = \frac{(X_1-X_2)^2}{2}$. Because $(X_1-X_2)^2 = ((X_1-E[X])-(X_2-E[X]))^2 = (X_1-E[X])^2+(X_1-E[X])^2+2(X_1-E[X])(X_1-E[X])$,

$$\begin{aligned} E[s^2] &= \frac{1}{2}E\left((X_1-X_2)^2\right) \\ &= \frac{1}{2}\left\{E(X_1-E[X])^2+E(X_2-E[X])^2+2E(X_1-E[X])(X_2-E[X])\right\} \\ &= \frac{1}{2}\left\{\sigma^2+\sigma^2+0\right\} = \sigma^2. \end{aligned}$$

Grading: No partial points unless there is a good reason to give partial points.

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 $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

   ANSWER: $Var(\bar{X}) = Var((1/n)(X_1+X_2+...+X_n)) = (1/n)^2 Var(X_1+X_2+...+X_n) = (1/n)^2(Var(X_1)+Var(X_2)+...+Var(X_n)) = (1/n)^2(n\times\sigma^2) = \frac{\sigma^2}{n}$**, where the third equality follows because $Cov(X_i, X_j) = 0$ if $i \neq j$ by random sampling.** Grading: No partial points unless there is a good reason to give partial points.

7. Table I 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, $p_x - p_y$?

   ANSWER: $\hat{p}_x - \hat{p}_y = \frac{331}{1357} - \frac{166}{1357} = 0.2439 - 0.1223 = 0.1216$.

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

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

	No. of Non- Smokers	No. of Smokers with the Daily Average of		
Disease Group		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).

ANSWER: by the Central Limit Theorem (CLT), the distribution of $\hat{p}_x - \hat{p}_y$ is approximated by the normal distribution with mean $p_x - p_y$ and variance $\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{n}$. The standard deviation can be estimated as $\sqrt{\hat{p}_x(1-\hat{p}_x) + \hat{p}_y(1-\hat{p}_y)\}/n} = \sqrt{\{0.2439(1-0.2439) + 0.1223(1-0.1223)\}/1357} = 0.01466$. Therefore, the 95 percent confidence interval can be computed as by $\hat{p}_x - \hat{p}_y \pm 1.96 \times \sqrt{\{\hat{p}_x(1-\hat{p}_x) + \hat{p}_y(1-\hat{p}_y)\}/n} = 0.1216 \pm 1.96 \times 0.01466$, which is $[0.0929, 0.1503]$.

(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.

ANSWER: We first derive the distribution of $\hat{p}_x - \hat{p}_y$ when the null hypothesis is true. Let $p_x = p_y = p_0$ under the null hypothesis. By the CLT, under the null hypothesis, the distribution of $\hat{p}_x - \hat{p}_y$ is approximated by the normal distribution with mean 0 and variance $p_0(1-p_0)/n + p_0(1-p_0)/n = 2p_0(1-p_0)/n$. We estimate $\hat{p}_0 = \frac{331+166}{2\times 1357} = 0.1831$ so that the standard deviation of $\hat{p}_x - \hat{p}_y$ under H_0 is estimated by $\sqrt{2 \times 0.1831(1-0.1831)/1357} = 0.01485$. Because $Z = \frac{\hat{p}_x - \hat{p}_y}{0.01485}$ follows standard normal distribution, $\Pr(\frac{\hat{p}_x - \hat{p}_y}{0.01485} > 1.645) = 0.05$. Therefore, we reject H_0 if $\frac{\hat{p}_x - \hat{p}_y}{0.01485} > 1.645$. Because $\frac{\hat{p}_x - \hat{p}_y}{0.01485} = 0.1216/0.01485 = 8.189 > 1.645$, we reject H_0.