Economics 325–003 **Term I 2016**

Introduction to Empirical Economics **Hiro Kasahara**

Final Exam

You have 2 hours and 30 minutes. When you use your calculator, try to keep the number up to four digits below decimal point. For example, if you will get 6.164414003 in calculator, pelase use 6.1644 for your subsequent calculation. When you refer to the standard normal distribution table, report the number that is closest or report two numbers. Good luck!

1. (20 points) State whether each of the following is true or false. No explanation necessary.

   (a) The confidence interval for population parameter θ with confidence level $1-\alpha$ is always given by $\hat{\theta} \pm z_{\alpha/2}\sqrt{Var(\hat{\theta})}$, where $\hat{\theta}$ is an unbiased point estimator of θ and $z_{\alpha/2}$ is the critical value such that $P(Z > z_\alpha) = \alpha$ given $Z \sim N(0,1)$.

   (b) 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 $n-1$.

   (c) If the null hypothesis is not rejected based on sample evidence, the researcher has proven beyond any doubt that the null hypothesis is true.

   (d) The central limit theorem cannot be applied to discrete random variables.

2. (10 points) Define $W = (X - E(X))/\sqrt{Var(X)}$ and $Z = (Y - E(Y))/\sqrt{Var(Y)}$. (You don't necessarily need to use the summation operator but, if you like, you can use the summation operator.)

   (a) (5 points) Prove that $E[W] = 0$.

   (b) (5 points) Prove that $Cov(W, Z) = Corr(X, Y)$.

3. (6 points) Let X_1 and X_2 be a pair of random variables. Show that the covariance between the random variables $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$ is 0 if and only if X_1 and X_2 have the same variance.

4. (10 points) Suppose that X_1, X_2, X_3 are randomly sampled from a population with mean μ and variance σ^2. Consider the following two point estimators of μ:

$$\hat{\mu}_1 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3,$$
$$\hat{\mu}_2 = \frac{1}{4}X_1 + \frac{1}{4}X_2 + \frac{1}{2}X_3,$$

(a) (5 points) Prove that both estimators are unbiased.

(b) (5 points) Prove that $\hat{\mu}_1$ is more efficient than $\hat{\mu}_2$.

5. (22 points) Let X_1, X_2, ..., X_n be $n = 9$ 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 $H_0 : \mu = 5$ against the alternative hypothesis $H_1 : \mu < 5$. Consider the following two different test statistics (i) $\tilde{X} = 0.6X_1 + \sum_{i=2}^{9} 0.05X_i = 0.6X_1 + 0.05X_2 + 0.05X_3 + ... + 0.05X_9$ and (ii) $\hat{X} = (1/4)(X_1 + X_2 + X_3 + X_4)$. Suppose that the realized values of $\tilde{X}$ and $\hat{X}$ are given by $\tilde{X} = -2.1$ and $\hat{X} = -2.0$, respectively.

   (a) (5 points) Compute the variance of $\tilde{X}$.

   (b) (5 points) Test the null hypothesis $H_0 : \mu = 5$ agains the alternative hypothesis $H_1 : \mu < 5$ using the test statistic $\tilde{X}$ at the significant level $\alpha = 0.1$.

   (c) (6 points) Compute the power of test using the test statistic $\tilde{X}$ at the significance level $\alpha = 0.1$ when the value of μ is equal to -3.

   (d) (6 points) Compute the power of test using the test statistic $\hat{X}$ at $\alpha = 0.1$, and discuss which of tests, the test based on $\tilde{X}$ or the test based on $\hat{X}$, is more powerful and which of tests you recommend.

6. (12 points) Table I reports the number of smokers among 400 patients with lung-cancer and the number of smokers among 400 patients with other diseases, which Prof. Kasahara randomly sampled from a population of patients with lung-cancer and from a population of patients with other diseases, respectively. Denote the population proportions of smokers for patients with lung-cancer by p_x and for patients with other diseases by p_y. We test the null hypothesis of $H_0 : p_x \leq p_y$ against $H_1 : p_x > p_y$.

   (a) (6 points) Using the data reported in Table I, test the null hypothesis of $H_0 : p_x = p_y$ against $H_1 : p_x > p_y$ at the significance level $\alpha = 0.05$.

   (b) (6 points) Compute the p-value of the test.

7. (10 points) The survey asks randomly sampled eligible voters in the U.S. whether he or she would vote for Trump. An individual i's voting preference is recorded as $X_i = 1$ if she/he would vote for Trump and as $X_i = 0$ otherwise. We also define $Y_i = 1$ if she/he would not vote for Trump and $Y_i = 0$ otherwise. Note that $Y_i = 1 - X_i$.

Table I: Tobacco Smoked Daily Over the 10 years (Hypothetical Example)

Disease Group	No. of Non-Smokers	No. of Smokers
Men:		
400 lung-cancer patients	30	370
400 patients with other diseases	50	350

Let p_x and p_y represent a fraction of Trump supporters and Trump non-supporters in population so that $p_x = E[X_i]$ and $p_y = E[Y_i]$. We are interested in estimating the population difference $p_x - p_y$.

Suppose that we randomly sample $n = 400$ voters and we construct two data sets, $\{X_1, X_2, ..., X_n\}$ and $\{Y_1, Y_2, ..., Y_n\}$, based on the same sample of $n = 400$ voters. Let $\hat{p}_x = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{p}_y = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

(a) (5 points) Derive the variance of the difference between $\hat{p}_x$ and $\hat{p}_y$ in terms of p_x and p_y and n.

(b) (5 points) Suppose that $\hat{p}_x = 0.52$ and $\hat{p}_y = 0.48$. Construct the 95 percent confidence interval for $p_x - p_y$, assuming that n is large enough to apply the Central Limit Theorem.

8. (10 points) Suppose that we have 2 data sets, each of which is randomly sampled. For each data set, we construct a 90 percent confidence interval for the population parameter θ. For $i = 1, 2$, define a Bernoulli random variable X_i, where $X_i = 1$ if the confidence interval constructed from the i-th data set contains the population parameter θ and $X_i = 0$ if the confidence does not contain the population parameter θ. Let $\Pr(X_i = 1) = p$ for $i = 1, 2$. Define $\bar{X} = \frac{1}{2}\sum_{i=1}^{2} X_i$.

   (a) (5 points) What is the probability mass function of $\bar{X} = \frac{1}{2}\sum_{i=1}^{2} X_i$ if $p = 0.9$? [Hint: What are the possible values $\bar{X}$ can take? What is the probability that each value of $\bar{X}$ happens?]

   (b) (5 points) Suppose that the realized value of $\bar{X}$ is equal to 0 because neither of two confidence intervals contains the population parameter θ.[1] Test the null hypothesis of $H_0 : p \geq 0.9$ against the alternative hypothesis of $H_1 : p < 0.9$ at the significance level $\alpha = 0.05$ by (i) constructing the rejection region and (ii) examining if the realized value of $\bar{X} = 0$ is in the rejection region or not.

[1] Here, we assume that we know the true value of θ so that we know whether each of confidence interval contains the true parameter value or not.