Economics 325–003 **Term I 2014**

Introduction to Empirical Economics **Hiro Kasahara**

Final Exam

1. (34 points) State whether each of the following is true or false. [No Explanation Necessary]

   (a) The larger the sample size, the smaller the variance of the sample standard deviation.

   Answer: T

   (b) The central limit theorem is not applied to discrete random variables.

   Answer: F

   (c) The sample standard deviation is an example of an estimator.

   Answer: T

   (d) The p-value of two-sided test is not equal to the p-value of one-sided test.

   Answer: T

   (e) 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 a level of significance of .01.

   Answer: T

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

   Answer: F

   (g) If a null hypothesis is rejected against an alternative at the 5% level, then using the same data, it must be rejected against that alternative at the 10% level.

   Answer: T

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

   Answer: T

2. Let X_1, X_2, X_3, and X_4 be a random sample of observations from a population with mean μ and variance σ^2. Consider the following two point estimators of μ: $\hat{\theta}_1 = 0.25X_1 + 0.25X_2 + 0.40X_3 + 0.10X_4$ and $\hat{\theta}_2 = 0.20X_1 + 0.30X_2 + 0.30X_3 + 0.20X_4$.

   (a) (4 points) Which of the following is true? [No Explanation Necessary]

   (1) $\hat{\theta}_1$ is biased, but $\hat{\theta}_2$ is unbiased estimator of μ. (2) $\hat{\theta}_1$ is unbiased, but $\hat{\theta}_2$ is biased estimator of μ. (3) Both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of μ. (4) Both $\hat{\theta}_1$ and $\hat{\theta}_2$ are biased estimators

of μ.

ANSWER: (3)

(b) (4 points) Which of the following is true? [No Explanation Necessary]

(1) $Var(\hat{\theta}_1) = Var(\hat{\theta}_2)$. (2) $Var(\hat{\theta}_1) > Var(\hat{\theta}_2)$. (3) $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$. (4) Cannot tell the relationship between $Var(\hat{\theta}_1)$ and $Var(\hat{\theta}_2)$.

ANSWER: $Var(\hat{\theta}_1) = 0.295 > Var(\hat{\theta}_2) = 0.26$ **so that (2) is the answer.**

3. (6 points) Assume that the yield per acre for a particular variety of soybeans is $N(\mu, \sigma^2)$. For a random sample of $n = 4$ plots, the yields per acre were 37.4, 48.8, 46.9, and 55.0. Find a 90 percent confidence interval for μ.

ANSWER: $\bar{X} = (1/4)(37.4 + 48.8 + 46.9 + 55) = 47.025$ **and** $s^2 = (1/3)((37.4 - 47.025)^2 + (48.8 - 47.025)^2 + (46.9 - 47.025)^2 + (55 - 47.025)^2) = 53.13$ **and** $s = 7.29$. **From the table for the Student's t distribution when d.f. =3, we have** $t_{0.05} = 2.353$. **Therefore, a 90 percent confidence interval is given by** $\bar{X} \pm 2.353 \times (s/\sqrt{4}) = 47.03 \pm 8.58$ **or** $(38.45, 48.61)$.

Grading Note: if s^2 **is computed by dividing by** 4 **rather than** 3**, give zero point. If** $z_{0.05} = 1.64$ **or** 1.65 **(or any other numbers such as** $t_{0.1} = 1.533$**) is used in place of** $t_{0.05} = 2.353$**, then give zero point. If students cannot find** $t_{0.05}$ **from t-table, give zero point. Any other clear computational mistake (perhaps, typing incorrect numbers in computing** $\bar{X}$ **or** s**), give 3 points.**

4. If a newborn baby has a birth weigh that is less than 2500 grams, we say that the baby has a low birth weight. The proportion of babies with a low birth weight is an indicator of lack of nutrition for the mothers. In Canada, approximately 7 percent of babies have a low birth weight. Let p equal the proportion of babies born in the Sudan who weight less than 2500 grams. In a random sample of $n = 209$ babies in the Sudan, there are 23 babies whose weight is less than 2500 grams.

(a) (6 points) Test the null hypothesis $H_0 : p \leq 0.07$ against the alternative hypothesis $H_1 : p > 0.07$ at a significance level of $\alpha = 0.05$.

ANSWER: $\hat{p} = 0.11$ **and** $\sqrt{\hat{V}ar(\hat{p})} = \sqrt{\hat{p}(1-\hat{p})/209} = \sqrt{0.11 \times 0.89/209} = 0.0216$. **Under** H_0, $\hat{p}$ **is approximately distributed as normal distribution with mean** $p_0 = 0.07$ **and standard deviation** $\sqrt{p_0(1-p_0)/n} = \sqrt{0.07 \times (1-0.07)/209} = 0.0176$. **The null hypothesis is rejected if the test statistic** $\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$ **is greater than** $z_{0.05} = 1.64$ **or** 1.65. **Here,** $\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.11-0.07}{0.0176} = 2.273$. **Therefore,** H_0 **is rejected.**

Grading Note: If the student uses $\sqrt{\hat{p}(1-\hat{p})/209} = \sqrt{0.11 \times 0.89/209} = 0.0216$ **in place**

of $\sqrt{p_0(1-p_0)/n} = \sqrt{0.07 \times (1-0.07)/209} = 0.0176$ **to compute the standard deviation for** $\hat{p}$**, this is OK and give full points because it is not wrong. Any other answer get zero points unless it is clear that a student makes computational mistake (say, typing wrong numbers into calculator), in which case student will get 3 points.**

(b) (6 points) Find the p-value for this test.

ANSWER: In Standard Normal distribution table, $\Pr(z > 2.27) = 0.5 - 0.4884 = 0.0116$ **and** $\Pr(z > 2.28) = 0.5 - 0.4887 = 0.0113$**. Therefore, the p-value is between** 0.0113 **and** 0.0116**.**

Grading Note: both 0.0113 **and** 0.0116 **are acceptable for the answer. Any other numbers will get zero point.**

2.266.

5. Let X, Y, and Z are three Bernoulli random variables that are independent to each other with the probability of success p_x, p_y, and p_z, respectively. We are interested in testing the null hypothesis $H_0 : p_x + p_y \geq p_z$ against the alternative hypothesis $H_1 : p_x + p_y < p_z$. Suppose that we have a random sample of X, Y, and Z with sample size $n = 100$: $\{X_1, ..., X_n\}$, $\{Y_1, ..., Y_n\}$, and $\{Z_1, ..., Z_n\}$. The sample averages (i.e., sample proportions) of X, Y, and Z are given by $\hat{p}_x = (1/n)\sum_{i=1}^n X_i = 0.1$, $\hat{p}_y = (1/n)\sum_{i=1}^n Y_i = 0.3$, and $\hat{p}_z = (1/n)\sum_{i=1}^n Z_i = 0.5$. Denote the true value of p_x, p_y, and p_z by p_x^0, p_y^0, and p_y^0,

(a) (6 points) What is the estimated variance of $\hat{p}_x + \hat{p}_y - \hat{p}_z$?

ANSWER: $Var(\hat{p}_x + \hat{p}_y - \hat{p}_z) = Var(\hat{p}_x) + Var(\hat{p}_y) + Var(\hat{p}_z) = p_x^0(1-p_x^0)/n + p_y^0(1-p_y^0)/n + p_z^0(1-p_z^0)/n$**, where the first equality follows from** $Var(\hat{p}_x + \hat{p}_y - \hat{p}_z) = Var(\hat{p}_x) + Var(\hat{p}_y) + Var(\hat{p}_z) + 2Cov(\hat{p}_x, \hat{p}_y) - 2Cov(\hat{p}_x, \hat{p}_z) - 2Cov(\hat{p}_y, \hat{p}_z)$ **(prove this yourself), where** $Cov(\hat{p}_x, \hat{p}_y) = Cov(\hat{p}_x, \hat{p}_z) = Cov(\hat{p}_y, \hat{p}_z) = 0$ **by random sampling assumption and the assumption that** X**,** Y**, and** Z **are independent to each other. Using the estimator of** $\hat{p}_x$**,** $\hat{p}_y$**, and** $\hat{p}_z$ **in place of** p_x^0**,** p_y^0**, and** p_y^0**, we have** $\hat{Var}(\hat{p}_x + \hat{p}_y - \hat{p}_z) = (1/100) * (0.1 * 0.9 + 0.3 * 0.7 + 0.5 * 0.5) = 0.0055$**. If the answer is given by** $Var(\hat{p}_x + \hat{p}_y - \hat{p}_z) = Var(\hat{p}_x) + Var(\hat{p}_y) + Var(\hat{p}_z) = p_x^0(1-p_x^0)/n + p_y^0(1-p_y^0)/n + p_z^0(1-p_z^0)/n$ **rather than 0.0055, give a full mark of 6 points.**

Grading Note: No point unless students gent 0.0055 unless it is clear that student understand that $Var(\hat{p}_x + \hat{p}_y - \hat{p}_z) = Var(\hat{p}_x) + Var(\hat{p}_y) + Var(\hat{p}_z) = p_x^0(1-p_x^0)/n + p_y^0(1-p_y^0)/n + p_z^0(1-p_z^0)/n$ **and just make mistake in computing 0.0055.**

(b) (6 points) Test the null hypothesis $H_0 : p_x + p_y \geq p_z$ against the alternative hypothesis $H_1 : p_x + p_y < p_z$ at the significance level $\alpha = 0.1$. [Hint: you can use sample proportions $\hat{p}_x$, $\hat{p}_y$, and $\hat{p}_z$ in place of p_x^0, p_y^0, and p_y^0 to compute the standard deviation for $\hat{p}_x + \hat{p}_y - \hat{p}_z$ when you construct a test statistic.]

ANSWER: Under the null hypothesis $H_0 : p_x + p_y - p_z = 0$, the test statistic $z = \frac{\hat{p}_x+\hat{p}_y-\hat{p}_z-0}{\sqrt{Var(\hat{p}_x+\hat{p}_y-\hat{p}_z)}}$ will be approximately distributed as $N(0, 1)$. We reject the null hypothesis if the realized value of $z = \frac{\hat{p}_x+\hat{p}_y-\hat{p}_z-0}{\sqrt{Var(\hat{p}_x+\hat{p}_y-\hat{p}_z)}}$ is smaller than $z_{0.9} = -1.28$. Using 0.0055 as an estimate for $Var(\hat{p}_x + \hat{p}_y - \hat{p}_z)$, we have $\frac{\hat{p}_x+\hat{p}_y-\hat{p}_z-0}{\sqrt{\hat{V}ar(\hat{p}_x+\hat{p}_y-\hat{p}_z)}} = \frac{-0.1}{\sqrt{0.0055}} = -1.35$. Because -1.35 is smaller than -1.28, we reject H_0. Grading Note: If student construct the test statistic $z = \frac{\hat{p}_x+\hat{p}_y-\hat{p}_z-0}{\sqrt{Var(\hat{p}_x+\hat{p}_y-\hat{p}_z)}}$, give 3 points. A full mark requires (i) computing critical value -1.28 and computing the test statistic's value -1.35 and correctly state that H_0 is rejected.

6. 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 25. We are interested in testing the null hypothesis $H_0 : \mu = 10$ against the alternative hypothesis $H_1 : \mu < 10$. Consider hypothesis testing based on the following two different test statistics: (i) sample mean $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$ and (ii) weighted average $\hat{X} = 0.6X_1 + 0.05 \times \sum_{i=2}^{9} X_i$. Suppose that the realized values of $\bar{X}$ and $\hat{X}$ are given by $\bar{X} = 7.5$ and $\hat{X} = 6.5$, respectively.

(a) (6 points) Test the null hypothesis $H_0 : \mu = 10$ against $H_1 : \mu < 10$ *using the test statistic* $\hat{X}$ at the significance level $\alpha = 0.1$. Similarly, test the null hypothesis *using the test statistic* $\bar{X}$.

Answer: $\sqrt{Var(\bar{X})} = \sqrt{25/9} \approx 1.667$ and $\sqrt{Var(\hat{X})} = \sqrt{(0.6)^2 + 8 \times (0.05)^2}\times 5 = \sqrt{0.38}\times 5 \approx 3.082$. The test using $\bar{X}$ will reject H_0 if $\bar{X}$ is less than $10 - 1.28 \times 1.667 = 7.866$ so that, given $\bar{X} = 7.5$, and the null hypothesis is rejected using $\bar{X}$ at $\alpha = 0.1$ because $7.5 < 7.866$. The test using $\hat{X}$ will reject H_1 if $\hat{X}$ is less than $10 - 1.28 \times 3.082 = 6.055$, and the null hypothesis is not rejected using $\hat{X}$ at $\alpha = 0.1$ because $6.5 > 6.055$.

Grading Note: 3 points for each of sub-questions.

(b) (6 points) Compute (i) the power of test using *using the test statistic* $\bar{X}$ when the true value of μ is equal to 8 and (ii) the power of test *using the test statistic* $\hat{X}$ when the true value of μ is equal to 8. Based on the power comparison, which test statistics, $\bar{X}$ or $\hat{X}$, do you recommend using for hypothesis testing?

Answer: the power of test using $\bar{X}$ is the probability of correctly rejecting H_0: $\mathbf{Pr}(\bar{X} <$

$7.866|\mu = 8)$ $=\mathbf{Pr}(\frac{\bar{X}-8}{\sqrt{Var(\bar{X})}} < \frac{7.866-8}{1.667}|\mu = 8)$ $=\mathbf{Pr}(Z < -0.0803) \approx 0.5 - 0.0319 = 0.468$. **The power of test using $\hat{X}$ is the probability of correctly rejecting H_0: $\mathbf{Pr}(\bar{X} < 6.055|\mu = 8)$ $=\mathbf{Pr}(\frac{\hat{X}-8}{\sqrt{Var(\hat{X})}} < \frac{6.055-8}{3.082}|\mu = 8)$ $=\mathbf{Pr}(Z < -0.631) \approx 0.5 - 0.2357 = 0.264$. Test using $\bar{X}$ has a higher power than $\hat{X}$ and, therefore, we recommend using $\bar{X}$ for hypothesis testing. Grading Note: 2 points for each of (i) and (ii). 2 point for recommending $\bar{X}$ with proper reasoning (If there is no reasoning, no point).**

7. (6 points) Suppose X_i for $i = 1, ..., n$ is randomly drawn from normal distribution with mean μ and 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.**

8. (6 points) Let X be a random variable with $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$. Define $Z = \frac{X-\mu}{\sigma}$. Prove that $E[Z] = 0$ and $\text{Var}(Z) = 1$.

   Answer: $E[Z] = E(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma}E[X-\mu] = \frac{1}{\sigma}(E[X]-\mu) = \frac{1}{\sigma} \times 0 = 0$. $Var(Z) = Var(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma^2}Var((X-\mu)) = \frac{1}{\sigma^2}E[((X-\mu) - E(X-\mu))^2] = \frac{1}{\sigma^2}E[(X-\mu)^2] = \frac{1}{\sigma^2}Var(X) = 1$.

9. (6 points) The chi-square distribution with the r degree of freedom, denoted by $\chi^2(r)$, is characterized by the sum of r independent standard normally distributed random variables Z_1^2, Z_2^2, ..., Z_r^2, where $Z_i \sim N(0,1)$ and Z_i and Z_j are independent if $i \neq j$ for $i, j = 1, ..., r$. Namely, $W = Z_1^2 + Z_2^2 + ... + Z_r^2$ has a distribution that is $\chi^2(r)$. Prove that, if W has a $\chi^2(r)$ distribution, then $E[W] = r$.

   Answer: $E[W] = E[Z_1^2 + Z_2^2 + ... + Z_r^2] = E[Z_1^2] + E[Z_2^2] + ... + E[Z_r^2] = 1 + 1 + ... + 1 = r$**, where the third equality follows from $E[Z_i^2] = Var(Z_i) = 1$ because Z_i is a standard normal random variable with mean 0 and variance 1.**