John Chao
Spring 2009

Econ 422 Problem Set # 2

1 a) Since Y_i, $i = 1, 2, \ldots, n$ are i.i.d. random variables, each distributed $N(10, 4)$,

$$\Pr(9.6 \leq \bar{Y} \leq 10.4) = \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \leq \frac{\bar{Y}-10}{\sqrt{4/n}} \leq \frac{10.4-10}{\sqrt{4/n}}\right)$$

$$= \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \leq Z \leq \frac{10.4-10}{\sqrt{4/n}}\right)$$

where $Z \sim N(0, 1)$

i) when $n = 20$, $\Pr\left(\frac{9.6-10}{\sqrt{4/n}} \leq Z \leq \frac{10.4-10}{\sqrt{4/n}}\right)$

$0.8133 - 0.1867$

$$= \Pr(-0.89 \leq Z \leq 0.89) = 0.63$$

ii) when $n = 100$, $\Pr\left(\frac{9.6-10}{\sqrt{4/n}} \leq Z \leq \frac{10.4-10}{\sqrt{4/n}}\right)$

$0.9772 - 0.0228$

$$= \Pr(-2 \leq Z \leq 2) = 0.954$$

iii) when $n = 1000$, $\Pr\left(\frac{9.6-10}{\sqrt{4/n}} \leq Z \leq \frac{10.4-10}{\sqrt{4/n}}\right)$

$1 - 0$

$$= \Pr(-6.32 \leq Z, \leq 6.32) = 1.00$$

b) $$\Pr(10 - c \leq \bar{Y} \leq 10 + c) = \Pr\left(\frac{-c}{\sqrt{4/n}} \leq \frac{\bar{Y}-10}{\sqrt{4/n}} \leq \frac{c}{\sqrt{4/n}}\right)$$

$$= \Pr\left(\frac{-c}{\sqrt{4/n}} \leq Z \leq \frac{c}{\sqrt{4/n}}\right)$$

As n get large, $\sqrt{4/n}$ gets small and $\frac{c}{\sqrt{4/n}}$ gets large, the probability converges to 1.

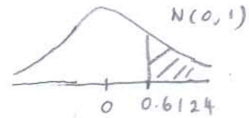
c) $\bar{Y}$ converges in probability to 10 by the argument in (b).

[The sample average $\bar{Y}$ converges in probability to μ_Y if the probability that $\bar{Y}$ is in the range $\mu_Y - c$ to $\mu_Y + c$ becomes arbitrarily close to one <u>as n</u> increases for any constant c.]

2 Since Y_i, $i = 1, 2, \ldots, n$ are i.i.d. Bernoulli random variables with $p = 0.4$, $\mu_Y = p = 0.4$ and $\sigma_Y^2 = p(1-p) = 0.4(0.6)$
$= 0.24$

a i) $$P(\bar{Y} \geq 0.43) = Pr\left(\frac{\bar{Y} - 0.4}{\sqrt{0.24/n}} \geq \frac{0.43 - 0.4}{\sqrt{0.24/n}}\right)$$

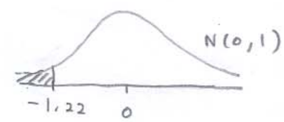
$$= Pr\left(\frac{\bar{Y} - 0.4}{\sqrt{0.24/n}} \geq 0.6124\right)$$



$$= Pr(Z \geq 0.6124) = 0.27$$

ii) $$P(\bar{Y} \leq 0.37) = Pr\left(\frac{\bar{Y} - 0.4}{\sqrt{0.24/n}} \leq \frac{0.37 - 0.4}{\sqrt{0.24/n}}\right)$$

$$= Pr\left(\frac{\bar{Y} - 0.4}{\sqrt{0.24/n}} \leq -1.22\right)$$



$$= Pr(Z \leq -1.22) = 0.11$$

b) We know $Pr(-1.96 \leq Z \leq 1.96) = 0.95$, thus we want

n to satisfy $$\frac{0.41 - 0.4}{\sqrt{0.24/n}} > -1.96 \quad \text{and} \quad \frac{0.39 - 0.4}{\sqrt{0.24/n}} < -1.96$$

Solving these inequalities yield $n \geq 9220$.

3. Denote each voter's preference by Y. $Y = 1$ if the voter prefers the incumbent and $Y = 0$ if the voter prefers the challenger. Y is a Bernoulli random variable with probability $Pr(Y=1) = p$ and $Pr(Y=0) = 1-p$. Y has mean p and variance $p(1-p)$.

a) $$\underbrace{\hat{p}}_{\text{estimate of } p} = \frac{215}{400} = 0.5375$$

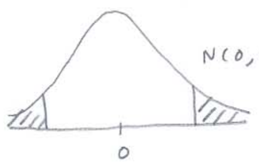
b) $$\widehat{var(\hat{p})} = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.5375(1-0.5375)}{400} = 6.2148 \times 10^{-4}$$

$$SE(\hat{p}) = \sqrt{var(\hat{p})} = \sqrt{6.2148 \times 10^{-4}} = 0.0249$$

c) The t-statistic is $t = \dfrac{\hat{p} - \mu}{SE(\hat{p})} = \dfrac{0.5375 - 0.5}{0.0249} = 1.506$

Because of the large sample size ($n = 400$), we can use the formula for p-value $= 2\underset{\substack{\uparrow \\ \text{cumulative Normal}}}{\Phi}(-\underbrace{|t|}_{\text{absolute value of t-statistic}})$ for the test

$H_0: p = 0.5$ vs $H_1: p \neq 0.5$:

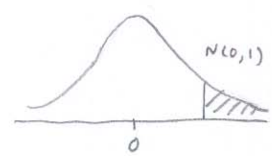


$$\text{p-value} = 2\Phi(-|t|) = 2\Phi(-1.506)$$
$$= 2 \times 0.066$$
$$= 0.132$$

d) Using p-value $= Pr_{H_0}(z > t) = 1 - \Phi(t)$ for the

↑

[area under the standard normal distribution to the right of the calculated t-statistic]



test $H_0: p = 0.5$ vs $H_1: p > 0.5$ is

$$\text{p-value} = 1 - \Phi(t) = 1 - \Phi(1.506)$$
$$= 1 - 0.934$$
$$= 0.066$$

e) Part (c) is a <u>two-sided</u> test and the p-value is the area in the tails of the standard normal distribution outside $\pm$ (calculated t-statistic). Part (d) is a <u>one-sided</u> test and the p-value is the area under the standard normal distribution to the right of the calculated t-statistic.

f) For the test $H_0: p = 0.5$ vs $H_1: p > 0.5$, we cannot reject the null hypothesis at the 5% significance level. The p-value 0.066 is larger than 0.05. Equivalently the calculated t-statistic 1.506 is less than the critical value 1.645 for a one-sided test with a 5% significance level. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

4.

a) 95 % confidence interval for p is

$$\hat{p} \pm \underbrace{1.96}_{\text{critical value}} SE(\hat{p}) = 0.5375 \pm 1.96 \times 0.0249 = (0.4887, 0.5863)$$

b) 99 % confidence interval for p is

$$\hat{p} \pm 2.57\ SE(\hat{p}) = 0.5375 \pm 2.57 \times 0.0249 = (0.4735, 0.6015)$$

c) The interval in (b) is wider because of a larger critical value due to a lower significance level.

d) Since 0.50 lies inside the 95 % confidence interval for p, we cannot reject the null hypothesis at a 5 % significance level.

5. 95 % confidence interval for the population mean test score for high school seniors:

$$1110 \pm 1.96 \underbrace{\left(\frac{123}{\sqrt{1000}}\right)}_{S_{\bar{Y}} = \frac{S_Y}{\sqrt{n}}} \quad \text{or} \quad 1110 \pm 7.62$$