

# **STA260: PROBABILITY AND STATISTICS II**

## **SPRING 2021**

### **TUTORIAL 4 (TUT9101)**

**MAY 14, 2021**

**YUEHUAN HE**



**UNIVERSITY OF  
TORONTO**

**7.95** The *coefficient of variation* (CV) for a sample of values  $Y_1, Y_2, \dots, Y_n$  is defined by

$$CV = S/\bar{Y}.$$

This quantity, which gives the standard deviation as a proportion of the mean, is sometimes informative. For example, the value  $S = 10$  has little meaning unless we can compare it to something else. If  $S$  is observed to be 10 and  $\bar{Y}$  is observed to be 1000, the amount of variation is small relative to the size of the mean. However, if  $S$  is observed to be 10 and  $\bar{Y}$  is observed to be 5, the variation is quite large relative to the size of the mean. If we were studying the precision (variation in repeated measurements) of a measuring instrument, the first case ( $CV = 10/1000$ ) might provide acceptable precision, but the second case ( $CV = 2$ ) would be unacceptable.

Let  $Y_1, Y_2, \dots, Y_{10}$  denote a random sample of size 10 from a normal distribution with mean 0 and variance  $\sigma^2$ . Use the following steps to find the number  $c$  such that

$n=10$

$$P\left(-c \leq \frac{S}{\bar{Y}} \leq c\right) = .95.$$

prob. 7.2.30

- a Use the result of Exercise 7.33 to find the distribution of  $(10)\bar{Y}^2/S^2$ .
- b Use the result of Exercise 7.29 to find the distribution of  $S^2/[(10)\bar{Y}^2]$ .
- c Use the answer to (b) to find the constant  $c$ .

prob 7.2.29

a.  $T \sim t(n)$  (prob 7.2.30)

$$U = T^2 \sim F_{1,n}$$

$$\frac{\bar{Y}}{\frac{S}{\sqrt{n}}} \sim t_9$$

$$\therefore \left( \frac{\bar{Y}}{\frac{S}{\sqrt{n}}} \right)^2 \sim F_{1,9}$$

**7.95** The *coefficient of variation* (CV) for a sample of values  $Y_1, Y_2, \dots, Y_n$  is defined by

$$\text{CV} = S/\bar{Y}.$$

This quantity, which gives the standard deviation as a proportion of the mean, is sometimes informative. For example, the value  $S = 10$  has little meaning unless we can compare it to something else. If  $S$  is observed to be 10 and  $\bar{Y}$  is observed to be 1000, the amount of variation is small relative to the size of the mean. However, if  $S$  is observed to be 10 and  $\bar{Y}$  is observed to be 5, the variation is quite large relative to the size of the mean. If we were studying the precision (variation in repeated measurements) of a measuring instrument, the first case ( $\text{CV} = 10/1000$ ) might provide acceptable precision, but the second case ( $\text{CV} = 2$ ) would be unacceptable.

Let  $Y_1, Y_2, \dots, Y_{10}$  denote a random sample of size 10 from a normal distribution with mean 0 and variance  $\sigma^2$ . Use the following steps to find the number  $c$  such that

$$P\left(-c \leq \frac{S}{\bar{Y}} \leq c\right) = .95.$$

- a Use the result of Exercise 7.33 to find the distribution of  $(10)\bar{Y}^2/S^2$ .
- b Use the result of Exercise 7.29 to find the distribution of  $S^2/[(10)\bar{Y}^2]$ .
- c Use the answer to (b) to find the constant  $c$ .

b. (prob 7.2.29)

$Y \sim F_{n_1, n_2}$

$U = \frac{1}{Y} \sim F_{n_2, n_1}$

$$\therefore \frac{S^2}{(10)\bar{Y}^2} \sim F_{9,1}$$

**7.95** The *coefficient of variation* (CV) for a sample of values  $Y_1, Y_2, \dots, Y_n$  is defined by

$$CV = S/\bar{Y}.$$

This quantity, which gives the standard deviation as a proportion of the mean, is sometimes informative. For example, the value  $S = 10$  has little meaning unless we can compare it to something else. If  $S$  is observed to be 10 and  $\bar{Y}$  is observed to be 1000, the amount of variation is small relative to the size of the mean. However, if  $S$  is observed to be 10 and  $\bar{Y}$  is observed to be 5, the variation is quite large relative to the size of the mean. If we were studying the precision (variation in repeated measurements) of a measuring instrument, the first case ( $CV = 10/1000$ ) might provide acceptable precision, but the second case ( $CV = 2$ ) would be unacceptable.

Let  $Y_1, Y_2, \dots, Y_{10}$  denote a random sample of size 10 from a normal distribution with mean 0 and variance  $\sigma^2$ . Use the following steps to find the number  $c$  such that

$$P\left(-c \leq \frac{S}{\bar{Y}} \leq c\right) = .95.$$

- a Use the result of Exercise 7.33 to find the distribution of  $(10)\bar{Y}^2/S^2$ .
- b Use the result of Exercise 7.29 to find the distribution of  $S^2/[(10)\bar{Y}^2]$ .
- c Use the answer to (b) to find the constant  $c$ .

c.  $F_{0.05} = 240.5$

for  $F_{9,1}$

$$P\left(-c \leq \frac{S}{\bar{Y}} \leq c\right)$$

$$= P\left(\frac{S^2}{\bar{Y}^2} \leq c^2\right)$$

$$= P\left(\frac{S^2}{10\bar{Y}^2} \leq \frac{c^2}{10}\right) = 0.95$$

$$\Rightarrow \frac{c^2}{10} = 240.5 \quad c = 49.04$$

- 7.96 Suppose that  $Y_1, Y_2, \dots, Y_{40}$  denote a random sample of measurements on the proportion of impurities in iron ore samples. Let each variable  $Y_i$  have a probability density function given by

$$f(y) = \begin{cases} 3y^2, & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

beta distribution  $\alpha = 3, \beta = 1$

The ore is to be rejected by the potential buyer if  $\bar{Y}$  exceeds .7. Find  $P(\bar{Y} > .7)$  for the sample of size 40.

$$\mu = \frac{3}{4}, \quad \sigma^2 = \frac{3}{80} \quad n = 40$$

$$P(\bar{Y} > 0.7) \approx P\left(\frac{\bar{Y} - \mu}{\sqrt{\frac{\sigma^2}{n}}} > \frac{0.7 - 0.75}{\sqrt{\frac{3}{4}}}\right)$$

or directly

$$E(Y) = \int_0^1 y f(y) dy = \frac{3}{4}$$

$$= P(Z > -1.63)$$

$$V(Y) = E(Y^2) - (E(Y))^2 =$$

$$= 0.948$$

$$= \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$$

- \*7.97 Let  $X_1, X_2, \dots, X_n$  be independent  $\chi^2$ -distributed random variables, each with 1 df. Define  $Y$  as

$$Y = \sum_{i=1}^n X_i. \quad \Rightarrow \quad Y = n \cdot \bar{X}$$

It follows from Exercise 6.59 that  $Y$  has a  $\chi^2$  distribution with  $n$  df.

- a Use the preceding representation of  $Y$  as the sum of the  $X$ 's to show that  $Z = (Y - n)/\sqrt{2n}$  has an asymptotic standard normal distribution.
- b A machine in a heavy-equipment factory produces steel rods of length  $Y$ , where  $Y$  is a normally distributed random variable with mean 6 inches and variance .2. The cost  $C$  of repairing a rod that is not exactly 6 inches in length is proportional to the square of the error and is given, in dollars, by  $C = 4(Y - \mu)^2$ . If 50 rods with independent lengths are produced in a given day, approximate the probability that the total cost for repairs for that day exceeds \$48.

$$\text{a. } \frac{Y-n}{\sqrt{2n}} = \frac{n\bar{X}-n}{\sqrt{2n}} = \frac{\bar{X}-1}{\sqrt{\frac{2}{n}}} \xrightarrow[\text{C.L.T.}]{d} N(0,1) \quad \because \bar{X} \sim \chi^2(1). \quad \therefore \mu = E(\bar{X}) = 1 \\ \sigma^2 = 2$$

- \*7.97 Let  $X_1, X_2, \dots, X_n$  be independent  $\chi^2$ -distributed random variables, each with 1 df. Define  $Y$  as

$$Y = \sum_{i=1}^n X_i.$$

It follows from Exercise 6.59 that  $Y$  has a  $\chi^2$  distribution with  $n$  df.

- a Use the preceding representation of  $Y$  as the sum of the  $X$ 's to show that  $Z = (Y - n)/\sqrt{2n}$  has an asymptotic standard normal distribution.
- b A machine in a heavy-equipment factory produces steel rods of length  $Y$ , where  $Y$  is a normally distributed random variable with mean 6 inches and variance .2. The cost  $C$  of repairing a rod that is not exactly 6 inches in length is proportional to the square of the error and is given, in dollars, by  $C = 4(Y - \mu)^2$ . If 50 rods with independent lengths are produced in a given day, approximate the probability that the total cost for repairs for that day exceeds \$48.

b. total cost  $C_{\text{Total}} = \sum_i C_i = \sum_i 4(Y_i - \mu)^2$

$$C_{\text{Total}} = \left( \sum_i \frac{(Y_i - \mu)^2}{\sigma^2} \right) \cdot (4\sigma^2)$$

$$\begin{aligned} P(C_{\text{Total}} > 48) &= P \left( \sum_i \frac{(Y_i - \mu)^2}{\sigma^2} > \frac{48}{4\sigma^2} \right) = P \left( \frac{\left( \sum_i \frac{(Y_i - \mu)^2}{\sigma^2} - n \right)}{\sqrt{2n}} > \frac{60n}{\sqrt{2n}} \right) \\ &= P(Z > \frac{10}{\sqrt{100}}) = P(Z > 1) = 0.1587. \end{aligned}$$

$$Y \sim N(\mu, \sigma^2)$$

$$\mu = 6, \sigma^2 = 0.2$$

$$\frac{Y_i - \mu}{\sigma} \sim N(0, 1)$$

$$\frac{(Y_i - \mu)^2}{\sigma^2} \sim \chi^2(1)$$

$$\frac{\left( \sum_i \frac{(Y_i - \mu)^2}{\sigma^2} - n \right)}{\sqrt{2n}} \sim \frac{60n}{\sqrt{2n}}$$



- 8.2 a If  $\hat{\theta}$  is an unbiased estimator for  $\theta$ , what is  $B(\hat{\theta})$ ?  
b If  $B(\hat{\theta}) = 5$ , what is  $E(\hat{\theta})$ ?

a  $\because \text{unbiased} \quad \therefore B(\hat{\theta}) = 0$

b  $B(\hat{\theta}) = E(\hat{\theta}) - \theta = 5$

$\therefore E(\hat{\theta}) = \theta + 5.$

- 8.3 Suppose that  $\hat{\theta}$  is an estimator for a parameter  $\theta$  and  $E(\hat{\theta}) = a\theta + b$  for some nonzero constants  $a$  and  $b$ .

prob 8.3.6

- a In terms of  $a$ ,  $b$ , and  $\theta$ , what is  $B(\hat{\theta})$ ?  $[(a-1)\theta + b]$

- b Find a function of  $\hat{\theta}$ —say,  $\hat{\theta}^*$ —that is an unbiased estimator for  $\theta$ .

$$\hat{\theta}^* = \frac{\hat{\theta} - b}{a}$$

- 8.5 Refer to Exercises 8.1 and consider the unbiased estimator  $\hat{\theta}^*$  that you proposed in Exercise 8.3.

- a Express  $MSE(\hat{\theta}^*)$  as a function of  $V(\hat{\theta})$ .

- b Give an example of a value of  $a$  for which  $MSE(\hat{\theta}^*) < MSE(\hat{\theta})$ .

- c Give an example of values for  $a$  and  $b$  for which  $MSE(\hat{\theta}^*) > MSE(\hat{\theta})$ .

$$a. MSE(\hat{\theta}^*) = V(\hat{\theta}^*) + B(\hat{\theta}^*)^2 = V(\hat{\theta}^*) = V\left(\frac{\hat{\theta} - b}{a}\right) = \frac{1}{a^2} V(\hat{\theta}).$$

$\downarrow$   
0

$$b. MSE(\hat{\theta}) = V(\hat{\theta}) + (a\theta + b)^2$$

e.g.  $a=2$   $b=0$ ,

c. solve for  $MSE(\hat{\theta}) < MSE(\hat{\theta}^*)$

8.8

Suppose that  $Y_1, Y_2, Y_3$  denote a random sample from an exponential distribution with density function

Problem 3

$$f(y) = \begin{cases} \left(\frac{1}{\theta}\right)e^{-y/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases} \Rightarrow E(Y_i) = \theta$$

$$V(Y_i) = \theta^2$$

Consider the following five estimators of  $\theta$ :

$$\hat{\theta}_1 = Y_1, \quad \hat{\theta}_2 = \frac{Y_1 + Y_2}{2}, \quad \hat{\theta}_3 = \frac{Y_1 + 2Y_2}{3}, \quad \hat{\theta}_4 = \min(Y_1, Y_2, Y_3), \quad \hat{\theta}_5 = \bar{Y}.$$

- a Which of these estimators are unbiased?
- b Among the unbiased estimators, which has the smallest variance?

a.  $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_5 \quad E(\hat{\theta}_i) = \theta, \quad \text{for } \hat{\theta}_4. \quad E(\hat{\theta}_4) = \frac{\theta}{3}.$

b.  $V(\hat{\theta}_1) = \theta^2, \quad V(\hat{\theta}_2) = \frac{\theta^2}{2}, \quad V(\hat{\theta}_3) = \frac{5}{9}\theta^2, \quad \underline{V(\hat{\theta}_5) = \frac{\theta}{3}} \quad \checkmark$

8.9

Suppose that  $Y_1, Y_2, \dots, Y_n$  constitute a random sample from a population with probability density function

$$f(y) = \begin{cases} \left(\frac{1}{\theta+1}\right) e^{-y/(\theta+1)}, & y > 0, \theta > -1, \\ 0, & \text{elsewhere.} \end{cases}$$

exponential with mean  
 $(\theta+1)$

Suggest a suitable statistic to use as an unbiased estimator for  $\theta$ . [Hint: Consider  $\bar{Y}$ .]

- 8.9** Suppose that  $Y_1, Y_2, \dots, Y_n$  constitute a random sample from a population with probability density function

$$f(y) = \begin{cases} \left(\frac{1}{\theta+1}\right)e^{-y/(\theta+1)}, & y > 0, \theta > -1, \\ 0, & \text{elsewhere.} \end{cases}$$

Suggest a suitable statistic to use as an unbiased estimator for  $\theta$ . [Hint: Consider  $\bar{Y}$ .]

$$E(\bar{Y}) = \frac{1}{n} \cdot \sum E(Y_i) = \theta + 1$$

$$\therefore E(\bar{Y} - 1) = \theta \quad , \quad \hat{\theta} = \bar{Y} - 1$$

8.12

The reading on a voltage meter connected to a test circuit is uniformly distributed over the interval  $(\theta, \theta + 1)$ , where  $\theta$  is the true but unknown voltage of the circuit. Suppose that  $Y_1, Y_2, \dots, Y_n$  denote a random sample of such readings.

Problem 5

a Show that  $\bar{Y}$  is a biased estimator of  $\theta$  and compute the bias.

$$Y_i \sim U(\theta, \theta + 1)$$

b Find a function of  $\bar{Y}$  that is an unbiased estimator of  $\theta$ .

$$E(Y_i) = \theta + 0.5$$

c Find  $MSE(\bar{Y})$  when  $\bar{Y}$  is used as an estimator of  $\theta$ .

$$V(Y_i) = \frac{1}{12}$$

a  $E(\bar{Y}) = \theta + 0.5$        $B(\bar{Y}) = 0.5$       biased

b.  $\hat{\theta} = \bar{Y} - 0.5 \rightarrow E(\hat{\theta}) = \theta$

c.  $MSE(\bar{Y}) = V(\bar{Y}) + B(\bar{Y})^2$

$$= \frac{1}{n^2} \cdot \left(n \frac{1}{12}\right) + \frac{1}{4} = \frac{1}{12n} + \frac{1}{4}.$$

8.13

We have seen that if  $Y$  has a binomial distribution with parameters  $n$  and  $p$ , then  $Y/n$  is an unbiased estimator of  $p$ . To estimate the variance of  $Y$ , we generally use  $n(Y/n)(1 - Y/n)$ .

Problem 6

a Show that the suggested estimator is a biased estimator of  $V(Y)$ .

$$E(Y) = np$$

b Modify  $n(Y/n)(1 - Y/n)$  slightly to form an unbiased estimator of  $V(Y)$ .

$$V(Y) = npq$$

$$a. E\left(n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)\right) = E\left(Y\left(1-\frac{Y}{n}\right)\right) = E\left(Y - \frac{Y^2}{n}\right)$$

$$q = (1-p).$$

$$= E(Y) - E\left(\frac{Y^2}{n}\right).$$

$$= np - \frac{1}{n}(npq + n^2p^2).$$

$$= np - pq + np^2.$$

biased

$$= np(1-p) - pq = (n-1)pq \neq npq$$

8.13

We have seen that if  $Y$  has a binomial distribution with parameters  $n$  and  $p$ , then  $Y/n$  is an unbiased estimator of  $p$ . To estimate the variance of  $Y$ , we generally use  $n(Y/n)(1 - Y/n)$ .

Problem 6

a Show that the suggested estimator is a biased estimator of  $V(Y)$ .

b Modify  $n(Y/n)(1 - Y/n)$  slightly to form an unbiased estimator of  $V(Y)$ .

$$\text{b } E(n(Y/n)(1 - Y/n)) = (n-1)Pq$$

$$E\left(\frac{n}{n-1} \cdot n(Y/n)(1 - Y/n)\right) = \frac{n}{n-1} \cdot (n-1)Pq = nPq$$

unbiased



**8.14** Let  $Y_1, Y_2, \dots, Y_n$  denote a random sample of size  $n$  from a population whose density is given by

$$f(y) = \begin{cases} \alpha y^{\alpha-1}/\theta^\alpha, & 0 \leq y \leq \theta, \\ 0, & \text{elsewhere,} \end{cases} \quad Y_{(n)} \sim \text{power family with n.d. } \Theta$$

where  $\alpha > 0$  is a known, fixed value, but  $\theta$  is unknown. (This is the power family distribution introduced in Exercise 6.17.) Consider the estimator  $\hat{\theta} = \max(Y_1, Y_2, \dots, Y_n)$ .

- a Show that  $\hat{\theta}$  is a biased estimator for  $\theta$ .
- b Find a multiple of  $\hat{\theta}$  that is an unbiased estimator of  $\theta$ .
- c Derive  $\text{MSE}(\hat{\theta})$ .

$$E(Y) = \frac{\alpha}{\alpha+1} \theta$$

$$E(Y^2) = \frac{\alpha}{\alpha+2} \theta^2$$

8.14 Let  $Y_1, Y_2, \dots, Y_n$  denote a random sample of size  $n$  from a population whose density is given by

$$f(y) = \begin{cases} \alpha y^{\alpha-1}/\theta^\alpha, & 0 \leq y \leq \theta, \\ 0, & \text{elsewhere,} \end{cases}$$

where  $\alpha > 0$  is a known, fixed value, but  $\theta$  is unknown. (This is the power family distribution introduced in Exercise 6.17.) Consider the estimator  $\hat{\theta} = \max(Y_1, Y_2, \dots, Y_n)$ .

- a Show that  $\hat{\theta}$  is a biased estimator for  $\theta$ .
- b Find a multiple of  $\hat{\theta}$  that is an unbiased estimator of  $\theta$ .
- c Derive  $\text{MSE}(\hat{\theta})$ .

a  $E(\hat{\theta}) = E(Y_{(n)}) = \frac{n\alpha}{n\alpha+1} \theta \neq \theta$

b.  $\hat{\theta}^* = \frac{n\alpha+1}{n\alpha} \hat{\theta} \quad E(\hat{\theta}^*) = \frac{n\alpha+1}{n\alpha} E(\hat{\theta}) = \frac{n\alpha+1}{n\alpha} \cdot \frac{n\alpha}{n\alpha+1} \theta = \theta.$

c.  $\text{MSE}(\hat{\theta}) = E[(Y_{(n)} - \theta)^2] = E(Y_{(n)}^2) - 2\theta E(Y_{(n)}) + \theta^2$   
 $= \frac{n\alpha}{n\alpha+2} \theta^2 - \frac{2n\alpha}{n\alpha+1} \theta^2 + \theta^2 = \frac{2}{(n\alpha+1)(n\alpha+2)} \theta^2$

8.15

Let  $Y_1, Y_2, \dots, Y_n$  denote a random sample of size  $n$  from a population whose density is given by

Problem 8

$$f(y) = \begin{cases} 3\beta^3 y^{-4}, & \beta \leq y, \\ 0, & \text{elsewhere,} \end{cases}$$

where  $\beta > 0$  is unknown. (This is one of the Pareto distributions introduced in Exercise 6.18.)

Consider the estimator  $\hat{\beta} = \min(Y_1, Y_2, \dots, Y_n)$ .

- a Derive the bias of the estimator  $\hat{\beta}$ .
- b Derive  $\text{MSE}(\hat{\beta})$ .

$$E(Y_{(1)}) = \frac{3n}{3n-1} \beta$$

$$E(Y_{(1)}^2) = \frac{3n}{3n-2} \beta^2$$

**8.15** Let  $Y_1, Y_2, \dots, Y_n$  denote a random sample of size  $n$  from a population whose density is given by Problem 8

$$f(y) = \begin{cases} 3\beta^3 y^{-4}, & \beta \leq y, \\ 0, & \text{elsewhere,} \end{cases}$$

$$E(Y) = \frac{2}{3}\beta \quad E(Y^2) = 3\beta^2$$

where  $\beta > 0$  is unknown. (This is one of the Pareto distributions introduced in Exercise 6.18.)

Consider the estimator  $\hat{\beta} = \min(Y_1, Y_2, \dots, Y_n)$ .

$$Y_{(1)} : g(y) = 3n\beta^{3n}y^{-(3n+1)} \quad y \geq \beta.$$

a Derive the bias of the estimator  $\hat{\beta}$ .

b Derive  $MSE(\hat{\beta})$ .

$$\therefore E(Y_{(1)}) = \frac{3n}{3n-1} \beta$$

$$E(Y_{(1)}^2) = \frac{3n}{3n-2} \beta^2$$

$$a. B(\hat{\beta}) = E(\hat{\beta}) - \beta$$

$$= \frac{3n}{3n-1} \beta - \beta = \frac{1}{3n-1} \beta$$

$$b. MSE(\hat{\beta}) = E(Y_{(1)}^2) - E(Y_{(1)})^2 + \left(\frac{1}{3n-1} \beta\right)^2 = \frac{3n}{3n-2} \beta^2 - \left(\frac{3n}{3n-1}\right)^2 \beta^2 + \left(\frac{1}{3n-1}\right)^2 \beta^2$$

$$= \frac{2}{(3n-1)(3n-2)} \beta^2.$$

- 8.27** A random sample of 985 “likely voters”—those who are judged to be likely to vote in an upcoming election—were polled during a phone-athon conducted by the Republican Party. Of those contacted, 592 indicated that they intended to vote for the Republican running in the election.

- a According to this study, the estimate for  $p$ , the proportion of all “likely voters” who will vote for the Republican candidate, is  $\hat{p} = .601$ . Find a bound for the error of estimation.
- b If the “likely voters” are representative of those who will actually vote, do you think that the Republican candidate will be elected? Why? How confident are you in your decision?

$$\text{a. } 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.031$$

↑  
standard deviation

$$\text{b. } \hat{p} - 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.570 > 0.5$$

**8.29** Sometimes surveys provide interesting information about issues that did not seem to be the focus of survey initially. Results from two CNN/USA Today/Gallup polls, one conducted in March 2003 and one in November 2003, were recently presented online.<sup>4</sup> Both polls involved samples of 1001 adults, aged 18 years and older. In the March sample, 45% of those sampled claimed to be fans of professional baseball whereas 51% of those polled in November claimed to be fans.

- a Give a point estimate for the difference in the proportions of Americans who claim to be baseball fans in March (at the beginning of the season) and November (after the World Series). Provide a bound for the error of estimation.
- b Is there sufficient evidence to conclude that fan support is greater at the end of the season? Explain.

a.  $\hat{\theta}$  estimator for  $P_2 - P_1$        $\hat{\theta} = \hat{P}_2 - \hat{P}_1 = 6\%$

$$2 \cdot \text{std}(\hat{\theta}) = 2 \cdot \sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}} = 2 \sqrt{\frac{0.45 \times 0.55}{1001} + \frac{0.51 \times 0.49}{1001}} \\ = 0.045$$

**8.29** Sometimes surveys provide interesting information about issues that did not seem to be the focus of survey initially. Results from two CNN/USA Today/Gallup polls, one conducted in March 2003 and one in November 2003, were recently presented online.<sup>4</sup> Both polls involved samples of 1001 adults, aged 18 years and older. In the March sample, 45% of those sampled claimed to be fans of professional baseball whereas 51% of those polled in November claimed to be fans.

- a Give a point estimate for the difference in the proportions of Americans who claim to be baseball fans in March (at the beginning of the season) and November (after the World Series). Provide a bound for the error of estimation.
- b Is there sufficient evidence to conclude that fan support is greater at the end of the season? Explain.

b.  $0.06 - 0.045 = 0.015 > 0$

$(\hat{\theta}) \quad (2 \text{ std } \hat{\theta})$

# THANK YOU FOR ATTENDING TODAY'S TUTORIAL

Feedbacks: [yuehuan.he@utoronto.ca](mailto:yuehuan.he@utoronto.ca)