

AMS 5 – Statistics (Quiz 2)
24 August 2006

Due at the beginning of class on Tuesday 29 August 2006

The Instructor

Dimitris Fouskakis

Student Name _____

1. One hundred draws are made at random with replacement from the box -1 -2 0 1 3. Find the expected value and standard error for
- a) the sum of the draws. (6 pts)
 - b) the average of the draws. (6 pts)
 - c) the percentage of times there was a positive number. (6 pts)
 - d) the sum of the positive numbers. (6 pts)

Answer: **a)** The average of the box is 0.2 and the SD = 1.72. Therefore, the expected value for the sum is $100 \times 0.2 = 20$ and the SE = $\sqrt{100} \times 1.72 = 17.2$. **b)** The expected value of the average is 0.2 with a SE of $17.2/100 = 0.172$. **c)** The new box model is 0 0 0 1 1. The expected value therefore is $2/5 = 40\%$ with a SE = $\sqrt{100} \times \sqrt{3/5 \times 2/5} = 4.9$, i.e., 4.9%. **d)** The new box model is 0 0 0 1 3. The average of the box is 0.8, and the SD is about 1.2. So the sum will be around $100 \times 0.8 = 80$ with a SE = $\sqrt{100} \times 1.2 = 12$.

2. One hundred draws are made at random with replacement from the box 1 2 3 4 5 6.
- a) If the sum of draws is 321, what is their average? (6 pts)
 - b) If the average of the draws is 3.78, what is the sum? (6 pts)
 - c) Estimate the chance that the average of the draws is between 3 and 4. (6 pts)

Answer: **a)** $321/100 = 3.21$. **b)** $3.78 \times 100 = 378$. **c)** The average of the draws will be between 3 and 4 when the sum is between 300 and 400. The expected value of the sum is 350, and the SE is $\sqrt{100} \times 1.7 = 17$, so the chance is about 99.7% (± 3 SEs).

3. A simple random sample of 3,500 people age 18 or over is taken in a large town to estimate the percentage of people who read newspaper. It turns out that 2,487 people in the sample are newspaper readers. Find a 95% confidence interval for the percentage of the newspaper readers in town. (7 pts)

Answer: The population percentage is estimated as $\frac{2,487}{3,500} \times 100\% \approx 71\%$. The

$SE = \sqrt{3,500} \times \sqrt{0.71 \times 0.29} = 27$, i.e. 0.77%. A 95% confidence interval for the percentage of the newspaper readers in town is therefore $(71\% \pm 2 \times 0.77\%) = (69.5\%, 72.5\%)$.

4. A multiple-choice quiz has 50 questions. Each question has 3 possible answers, one of which is correct. Two points are given for each correct answer, but a point is taken off for a wrong answer. The passing score is 50. If a student answers all the questions at random, what is the chance of passing? (7 pts)

Answer: If a student answers at random, the score is like the sum of 50 draws from the box $\underline{2 \ -1 \ -1}$. The average of the box is 0, and the SD is 1.41. The

expected value for the sum of the box is $50 \times 0 = 0$; the SE is $\sqrt{50} \times 1.41 \approx 10$. So the score will be around 0, give or take 10 or so. A cutoff of 50 is 5 SEs, the chance is zilch.

5. In the printing of a newspaper, frequent mistakes in the printing process often require that many tons of newsprint must be thrown away. A large newspaper company is concerned that's its printing operations waste too much newsprint. To judge the severity of their problem and to help decide if management action is warranted, the company has collected data on the weight of daily newsprint waste for each of the last 50 days of operations. The observed sample mean is 273.1 tons/day and the observed sample standard deviation is 64.2 tons/day.

- Estimate the chance that the average weight of the daily newsprint waste is above 300 tons. (6 pts)
- Compute a 95% confidence interval for the average weight of the daily newsprint waste. (6 pts)
- Say whether the following statement is true or false and explain. In order to get answers for the previous two questions you assumed that the weight of the daily newsprint waste follows the normal curve. (6 pts)

Answer: **a)** This is a trap question. The average weight of the daily newsprint waste is an unknown quantity that has a fixed value (not a variable). Therefore it will be either above 300 tons or below. What you can do if you still want to quantify your uncertainty about the required chance, you can check the null hypothesis that the average is at least 300 tons, versus the alternative that is less than 300 tons, and based on the P-value to see if you have enough evidence to reject the null. **b)** An estimate for the average weight of the daily newsprint waste

is 273.1 tons with a $SE = \frac{64.2}{\sqrt{50}} = 9.08$. Therefore, the required confidence interval is $(273.1 \pm 2 \times 9.08) = (254.94, 291.26)$. **c)** False, you only need the normality of the average.

6. Many companies are experimenting with “flex-time”, allowing employees to choose their schedules within broad limits set by management. Among other things, flex-time is supposed to reduce absenteeism. One firm knows that in the past few years, employees have averaged 6.3 days off work (apart from vacations). This year, the firm introduces flex-time. Management chooses a simple random sample of 100 employees to follow in detail, and at the end of the year, these employees average 5.5 days off from work, and the SD is 2.9 days.

- a) Does this mean that the flex-time reduced absenteeism? (6 pts)
- b) Repeat the previous question for a sample average of 5.9 days and an SD of 2.9 days. (6 pts)

Answer: a) The sample is like 100 draws made at random from a box which has one ticket from each employee, showing the number of days that the employee was absent. Null hypothesis: the average of the box is 6.3. Alternative hypothesis: the average of the box is less than 6.3 days. The SD of the box is estimated from the sample as 2.9 days, so the SE for the average is $\frac{2.9}{\sqrt{100}} = 0.29$ days. The z-score therefore is $z = \frac{5.5 - 6.3}{0.29} \approx -2.8$, so P-value = 0.3%. This is very strong evidence against the null. **b)** Now, $z = \frac{5.9 - 6.3}{0.29} \approx -1.4$, so P-value = 8%. Therefore, there is not enough evidence to reject the null hypothesis.

7. A car company says that the mean gas mileage for its luxury sedan is at least 21 miles per gallon. You believe that the claim is incorrect and find that a random sample of 5 cars has a mean gas mileage of 19 miles per gallon with a standard deviation of 4 miles per gallon. Is there enough evidence in the data to reject the company’s claim? (7 pts)

Answer: Null hypothesis: gas mileage is 21 miles per gallon or more. Alternative hypothesis: gas mileage is less than 21 miles per gallon. The sample size is very small and therefore we should use the t-test. The $SD^+ = \sqrt{5/4} \times 4 = 4.47$. The t-score therefore is $t = \frac{19 - 21}{4.47/\sqrt{5}} = -1.0$. For a t with 4 degrees of freedom the table has the value of 1.53 corresponding to a right hand tail of 10%. This implies that -1 will have a left hand tail of more than 10%. Therefore, there is not enough evidence to reject the null hypothesis.

8. In a study conducted to compare surgical and drug treatments for chronic lower back pain 80 out of 100 patients treated surgically experienced relief from pain, while 50 out of 100 patients receiving a nonsteroidal anti-

inflammatory agent experienced relief. Did the surgical treatment increase the percentage of patients experienced relief? (7 pts)

Answer: The parameter to be estimated is the difference between the proportion of patients experiencing relief after surgery (P_S) and the proportion of experiencing relief after drug treatment (P_D). The estimate for the difference between the two population proportions is the difference between the two sample proportions i.e. $0.8 - 0.5 = 0.3$. The SE associated with this estimate is

$$\sqrt{SE_S^2 + SE_D^2} = \sqrt{\frac{0.8 \times 0.2}{100} + \frac{0.5 \times 0.5}{100}} = 0.064.$$

Null hypothesis: the difference between the percentages of patients experienced relief after surgery and drug treatment is 0. Alternative hypothesis: the difference between the percentages of patients experienced relief after surgery and drug treatment is positive.

The z-score therefore is $z = \frac{0.3 - 0.0}{0.064} = 4.68$ and so the P-value is much smaller than 1%. Therefore, there is very strong evidence against the null hypothesis.