

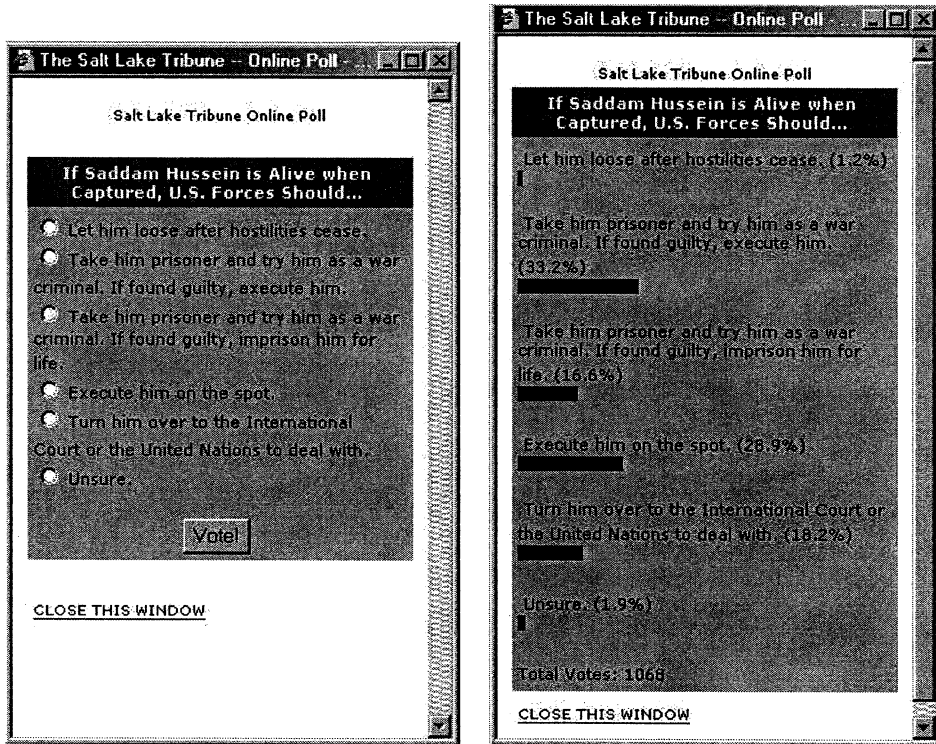
# Statistics 1040, Sections 003 & 004, Midterm 2 (200 Points)

March 28, 2003

Your Name: \_\_\_\_\_

## Question 1: Sampling (30 Points)

The pictures below show the *The Salt Lake Tribune* Online Poll of 03/25/2003:



Conclusion: a correctly conducted survey on the public opinion in the US might reveal that 50% or 60% of the US population would take S.H. prisoner and try him as a war criminal - or just 10%; based on this poll, we just cannot say what the correct percentage is

(30 Points) So, based on this poll, it seems that about 1/3 of the nation would take S.H. prisoner and try him as a war criminal, and a slightly smaller proportion of the nation (28.9%) would execute him on the spot. Do you think the results of this poll objectively represent public opinion in the US? Answer yes or no and give **three arguments** to justify your answer, based on the facts given, but **NOT** on your personal opinion.

- No!
- (12)
- each explanation:
- voluntary answers - only people that know about this web site could access it and had a chance to indicate their opinion
  - there is a selection bias, since older and poorer people with no computer access cannot express their opinion
  - most likely, only people from Salt Lake City or other places in UT are reading the Salt Lake Tribune; these readers are certainly not representative for the public opinion in the US (people in New York, Texas, California, or North Dakota might vote differently)
  - in general, people with a strong opinion are more likely to vote at all
  - people can vote more than once!

based on: FPP, p. 327, Review Exercise 2

**Question 2: The Expected Value and Standard Error (60 Points)**

Three hundred draws will be made at random with replacement from the box

1	3	5	7
---	---	---	---

-2 each calculation error

Show your work!

1. (25 Points) Estimate the chance that the sum of draws will be more than 1,100.

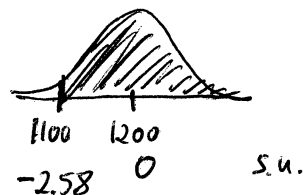
$$\text{box avg} = \frac{1+3+5+7}{4} = \frac{16}{4} = 4 \quad (3)$$

$$\text{box SD} = \sqrt{\frac{(1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2}{4}} = \sqrt{\frac{9+1+1+9}{4}} = \sqrt{\frac{20}{4}} = \sqrt{5} = 2.24 \quad (3)$$

$$EV_{\text{sum}} = 300 \cdot 4 = 1,200 \quad (4)$$

$$SE_{\text{sum}} = \sqrt{300} \cdot 2.24 = 17.32 \cdot 2.24 = 38.8 \quad (4)$$

$$\text{s.u.: } \frac{1,100 - 1,200}{38.8} = \frac{-100}{38.8} = -2.58 \quad (4)$$



area between -2.60 and 2.60: 99.07% (3)

$$\text{area above } -2.60: \frac{99.07\%}{2} + 50\% = 49.54\% + 50\% = \underline{\underline{99.54\%}} \quad (4)$$

2. (35 Points) Estimate the chance that there will be fewer than 80 [5]'s.

new box: 1: [5] (2)  
0: everything else

3	×	0	1	×	1
---	---	---	---	---	---

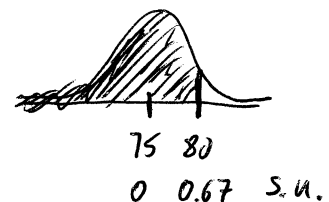
# draws = 300 (2)

$$\text{box avg} = \frac{1}{4} = 0.25 \quad (3)$$

$$\text{box SD} = \sqrt{\frac{1}{4} \cdot \frac{3}{4}} = \sqrt{\frac{3}{16}} = \sqrt{0.1875} = 0.43 \quad (3)$$

$$EV_{\text{sum}} = 300 \cdot 0.25 = 75 \quad (4)$$

$$SE_{\text{sum}} = \sqrt{300} \cdot 0.43 = 17.32 \cdot 0.43 = 7.45 \quad (4)$$



$$\text{s.u.: } \frac{80 - 75}{7.45} = \frac{5}{7.45} = 0.67 \quad (4)$$

area between -0.65 and 0.65: 48.43% (3)

$$\text{area below } 0.65: 50\% + \frac{48.43\%}{2} = 50\% + 24.22\% = \underline{\underline{74.22\%}} \quad (4)$$

based on: Stat 1040, Spring 2002, Final Test, April 30, 2002, Question 6b,c,d

**Question 3: Regression (50 Points)**

-2 each calculation error

For a random sample of 25 car models, the average weight (in pounds) was 3240, with an SD of 530. The average gas mileage (in miles per gallon) was 22.3 with an SD of 4.3. The correlation between weight and gas mileage was -0.79. The scatter diagram was football shaped. Show your work!

1. (15 Points) Find the regression equation for predicting gas mileage from weight.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = -0.79 \cdot \frac{4.3}{530} \approx -0.0064 \quad (6)$$

$$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 22.3 - (-0.0064) \cdot 3240 \approx 43.0 \quad (6)$$

$$\text{equation: } \text{mileage} \approx 43.0 - 0.0064 \cdot \text{weight} \quad (\text{or: } y = 43.0 - 0.0064 \cdot x) \quad (3)$$

-4 if x, y flipped  
-2 if x, y not specified

	avg	SD
x: weight	3240	530
y: mileage	22.3	4.3
		$r = -0.79$

2. (10 Points) Using your regression equation, predict the gas mileage of a car that weights 3200 pounds.

$$\text{mileage for 3200 pounds: } 43.0 - 0.0064 \cdot 3200 = 43.0 - 20.48 = \underline{\underline{22.52}}$$

-2 for old method, correct result  
-8 for old method, incorrect result

3. (10 Points) Find the r.m.s. error for predicting gas mileage from weight.

$$\text{r.m.s. error} = \sqrt{1 - r^2} \cdot SD_y = \sqrt{1 - (-0.79)^2} \cdot 4.3 = \sqrt{1 - 0.6241} \cdot 4.3$$

-4 for each major mistake, e.g.  
SD<sub>x</sub> instead of SD<sub>y</sub>,  $\sqrt{\quad}$  of everything,  
 $r$  instead of  $r^2$ , etc.

$$= \sqrt{0.3759} \cdot 4.3 = 0.61 \cdot 4.3 = \underline{\underline{2.64}}$$

4. (15 Points) Would you be surprised if someone told you that one of these cars weighting 3600 pounds got 24 miles per gallon? Why or why not? Explain your reasoning, using the r.m.s. error.

$$\text{mileage for 3600 pounds: } 43.0 - 0.0064 \cdot 3600 = 43.0 - 23.04 = 19.96 \quad (5)$$

$$\text{s.u.: } \frac{24 - 19.96}{2.64} = \frac{4.04}{2.64} = 1.53 \quad (5) \quad [\text{or any other reasonable explanation}]$$

24 is only 1.53 r.m.s. errors above the predicted value of 19.96;

so this is not unusual at all...

(5)

**Question 4: Probability and Chance (40 Points)**

-2 each calculation error

A bookshelf contains 8 novels, 7 books of poems, 1 dictionary, and 2 copies of Freedman, Pisani, and Purves's "Statistics" textbook. I pick two books at random without replacement. Answer each of the following questions separately. Show your work!

$$8 + 7 + 1 + 2 = 18 \text{ books}$$

1. (5 Points) What is the chance that the first book is a novel or a dictionary?

chance first is novel:  $\frac{8}{18}$  ①

chance first is dictionary:  $\frac{1}{18}$  ①

mutually exclusive

③ correct rule

chance first is novel or dictionary:  $\frac{8}{18} + \frac{1}{18} = \frac{9}{18} = \frac{1}{2} = 0.5 = \underline{\underline{50\%}}$

2. (13 Points) If I want to study statistics, what is the chance that I pick at least one copy of the Stats textbook?

chance first is no stats book:  $\frac{16}{18}$  ③

chance second is no stats book, given that first is no stats book:  $\frac{15}{17}$  ③

③ correct rule

chance both are no stats books:  $\frac{16}{18} \cdot \frac{15}{17} = \frac{240}{306}$

chance at least one stats book:  $1 - \frac{240}{306} = \frac{306}{306} - \frac{240}{306} = \frac{66}{306} = 0.216 = \underline{\underline{21.6\%}}$

3. (10 Points) What is the chance that the first two books are both novels?

chance first is novel:  $\frac{8}{18}$  ①

chance second is novel, given that first is novel:  $\frac{7}{17}$  ④

④ correct rule

chance both are novels:  $\frac{8}{18} \cdot \frac{7}{17} = \frac{56}{306} = 0.183 = \underline{\underline{18.3\%}}$

⑤ correct rule

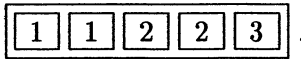
4. (12 Points) What is the chance that I pick one book of poems and a dictionary (in any order)?

chance first is book of poems: $\frac{7}{18}$ ①	chance first is dictionary: $\frac{1}{18}$ ①
chance second is dictionary, given that first is book of poems: $\frac{1}{17}$ ②	chance second is book of poems, given that first is dictionary: $\frac{7}{17}$ ②
chance first is book of poems & second dictionary: $\frac{7}{18} \cdot \frac{1}{17} = \frac{7}{306}$ ② correct rule	chance first is dictionary & second book of poems: $\frac{1}{18} \cdot \frac{7}{17} = \frac{7}{306}$ ② correct rule
chance one book of poems and a dictionary (in any order) = $\frac{7}{306} + \frac{7}{306} = \frac{14}{306} = 0.046 = \underline{\underline{4.6\%}}$	
② correct rule	

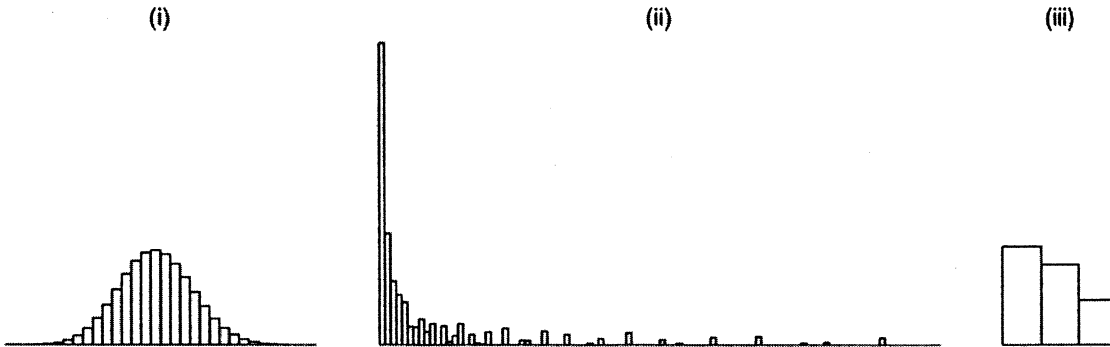
from: FPP, p. 328, Review Exercise 5

**Question 5: Normal Approximation for Probability Histograms (20 Points)**

Twenty-five draws are made at random with replacement from the box



(20 Points) One of the graphs below is an (empirical) histogram for the numbers drawn. One is the probability histogram for the sum. And one is the probability histogram for the product. Which is which? Explain!



- An (empirical) histogram for the numbers drawn is (iii). (4)

Explanation:

According to the probability histogram (see below), we should have 40% 1's, 40% 2's, and 20% 3's. The empirical histogram after 25 draws will somewhat resemble the probability histogram, but not too closely. (3)

- The probability histogram for the sum is (i). (4)

Explanation:

The probability histogram (see below) is not perfectly symmetric, but also not very asymmetric. The probability histogram for the sum will follow the normal curve even after only 25 draws. (3)

- The probability histogram for the product is (ii). (4)

Explanation:

The probability histogram for the product typically does not follow the normal curve. (2)

Probability Histogram for box:

