Stat 31-1 Second Midterm Exam

Thursday March 29, 2001

Name: _____

Instructions:

1. Answer all questions and **show all work** on the exam in the space provided (you may use the backs of pages if necessary). You will **not** receive credit if you do not justify your answers.
2. All multiple choice questions have only one correct answer.
3. Please draw a **circle** or **box** around your final answers.
4. Unless otherwise specified, give exact answers.
5. Remember to sign the Honor Pledge.

I have neither given nor received any unauthorized help on this exam and I have conducted myself within the guidelines of the University Honor Code.

Pledge: _____

1. (24 points) The original simple form of the Connecticut state lottery awarded the following prizes for each 100,000 tickets sold. The winners were chosen by drawing tickets at random.

1	\$5000 prize
18	\$200 prizes
120	\$25 prizes
270	\$20 prizes

(a) Let X be the amount of winning for one ticket sold, make a probability table showing the distribution of X.
Answer

X	\$0	\$20	\$25	\$200	\$5000
prob.	0.99591	0.0027	0.0012	0.00018	0.00001

(b) Find $P(X \leq \$200)$.
Answer

$$P(X \leq \$200) = 1 - P(X = \$5000) = 0.99999$$

(c) Find $P(X = \$20 | X \leq \$200)$.
Answer

$$P(X = \$20 | X \leq \$200) = \frac{P(X = \$20 \text{ and } X \leq \$200)}{P(X \leq \$200)} = \frac{P(X = \$20)}{P(X \leq \$200)} = \frac{0.0027}{0.99999} \simeq 0.0027$$

(d) What is the expected amount won μ_X?
Answer

$$\mu_X = \sum x_i * p_i$$

So,

$$\mu_X = (0*0.99591) + (20*0.0027) + (25*0.0012) + (200*0.00018) + (5000*0.00001) = 0.17.$$

The expected amount won is 17 cents.

(e) What is the standard deviation of the amount won?
Answer

$$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 * p_i}$$

So,

$$\sigma_X = 16.09$$

(f) If you hold two tickets, what is the expected total amount you will win?
Answer
Let μ_X be the expected amount won for the first ticket, and μ_Y the expected amount won for the second ticket. Then $\mu_X = \mu_Y = .17$, and the expected total amount you will win is $\mu_X + \mu_Y = 34$ cents.

Continued...

2. (12 points) A household is called prosperous if its income exceeds \$75,000, and called educated if the householder completed college. 20% of all households are prosperous, 30% are educated, and 19% are prosperous and educated. If a household is chosen at random

   (a) What is the probability that it either is educated, or is prosperous?
   Answer
   Let A and B be the following events:
   - *A: the household is prosperous.*
   - *B: the household is educated.*

   Then,

   $$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.20 + 0.30 - 0.19 = 0.31$$

   (b) What is the probability that it is educated given that it is prosperous?

   Answer

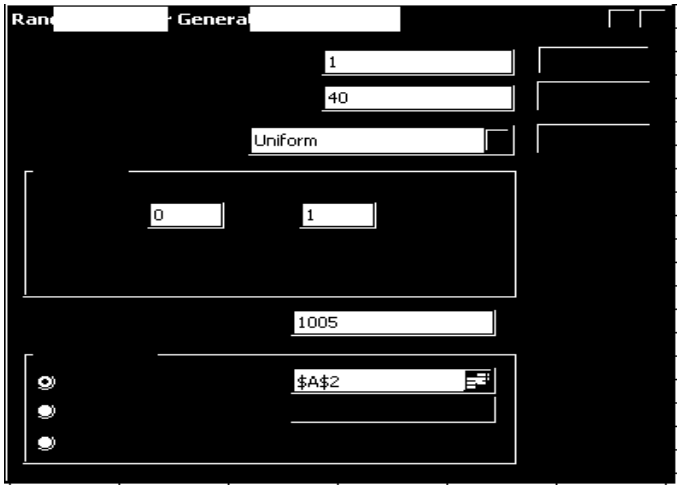
   $$P(B|A) = \frac{P(A \text{and} B)}{P(A)} = \frac{.19}{.20} = 0.95$$

   (c) Is the event that it is educated independent of the event that it is prosperous? Why or why not?
   Answer

   $$\begin{aligned} P(B|A) &= 0.95 \\ P(B) &= 0.30 \\ \text{So,} & \\ P(B|A) &\neq P(B) \end{aligned}$$

   Then the two events A and B are not independent.

Continued...



3. (a) (8 points) Fill out the menu to generate 40 observations from a uniform distribution, starting at the cell A2 in your spreadsheet.

   (b) What is the the probability that any generated number is greater than 0.5 and smaller than 1.2?
   Answer: A generated number from a uniform distribution can take values from 0 to 1.

$$P(0.5 \leq X \leq 1.2) = P(0.5 \leq X \leq 1) = 0.5$$

Continued...

4. (12 points) A High school guidance counsellor wants to study the later earnings of the Class of 1990. The High school had 800 graduates in the class of 1990; among them, 400 obtained college degrees and the other 400 did not. In the year 2000, the counsellor randomly selects 20 individuals from among those who obtained college degrees, and another 20 individuals among those who didn't obtain college degrees. Earnings data is collected on these 40 selected individuals.

   (a) This study is

      1. a designed experiment, in which the "treatment group" receives a college degree, and the "control" group " gets no college degree.
      2. based on "anecdotal evidence", and hence is not reliable.
      3. an observational study, in which the 40 selected individuals are a Simple Random Sample.
      4. an observational study, in which the 40 selected individuals are a Stratified Random Sample.

      Answer 4

   (b) The counsellor selects the individuals by using a Random Digits Table, beginning here:
   98120 23651 35938 59284 10773 94395 39582 10307
   The list of those numbers of the Class 1990 who have a college degrees is alphabetized and labelled in the usual appropriate way. Which individual from this list is the first to be selected?

      1. The 9^{th} individual in the list.
      2. The 120^{th} individual in the list.
      3. The 202^{nd} individual on the list.
      4. The 236^{th} individual on the list.

      Answer 3

   (c) The counsellor finds that the average earnings of the 20 individuals who didn't obtain a college degrees is \$23,000. The value \$23,000 is

      1. a parameter.
      2. $\hat{p}$ for the sample.
      3. a statistic.
      4. $\hat{p}$ for the population.

      Answer 3

Continued...

5. (12 points) Marketers use 2 types of price comparisons when promoting a product

   - "was \$25, now \$20" (comparison to previous price at the same store)
   - "is \$5 cheaper than anywhere else" (comparison to competitor's price).

   The effectiveness of these promotions can depend on the type of comparison (comparison to previous price at the same store vs competitor's price) or situation of the advertisement(advertisements read at home vs. in store ad). Marketers also believe that there might be a gender bias in the response to the promotion. To answer this question, you are asked to design an experiment on 52 undergraduate students (28 males and 24 females). You would use flyers promoting the same product from the same store. You could distribute the flyers in mailboxes or at the store. Each flyer is either making a price comparison to previous price at the same store, or to a competitor's price. The effectiveness of the type of advertisement could be measured by the answers of the subjects to questions on attitude toward the product, and intention to purchase it.

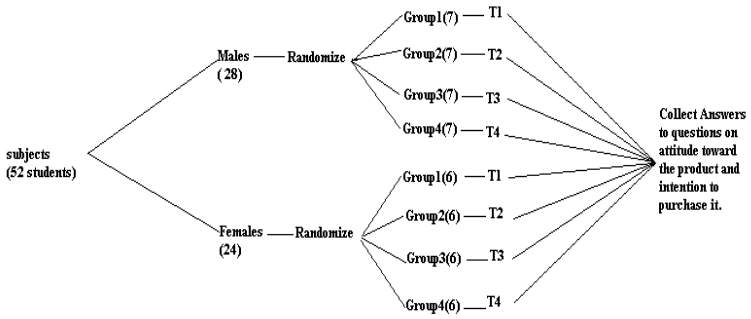
   (a) What are the blocks in this experiments?
   Answer The two blocks are the block of males and the block of females

   (b) Identify the experimental units or subjects, the factors (and levels), the treatments, and the response variables.
   Answer

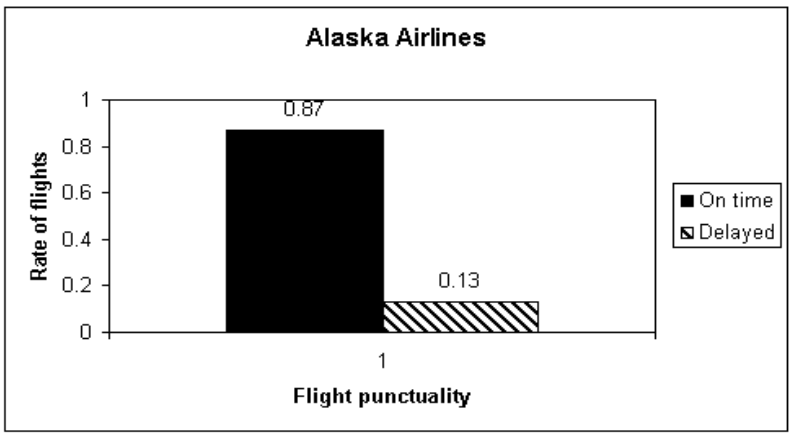
   - *Subjects: 52 undergraduate students.*
   - *Factors:*
     - *Factor 1: Type of comparison (2 levels: comparison to previous price, and comparison to competitor's price)*
     - *Factor 2: Situation of the advertisement (2 levels: read at home, and read in store)*
   - *treatments: there are 4 treatments.*
     - T_1*: flyer making a price comparison to previous price at the same store distributed at home.*
     - T_2*: flyer making a price comparison to previous price at the same store distributed at the store.*
     - T_3*: flyer making a price comparison to competitor's price at distributed at home.*
     - T_4*: flyer making a price comparison to competitor's price at distributed at the store.*
   - *Response: Answers of subjects to questions on attitude toward the product, and intention to purchase it.*

   (c) Outline the design of this experiment.

Continued...



6. (8 points) Two competing airlines "America West" and "Alaska Airline" serve the same five US airports. They reported the number of flights on time or delayed for the last year. As is shown in the two following bar plots, overall "America West" did better than "Alaska Airline".



(a) Which information about the two bar plots is **not true**:

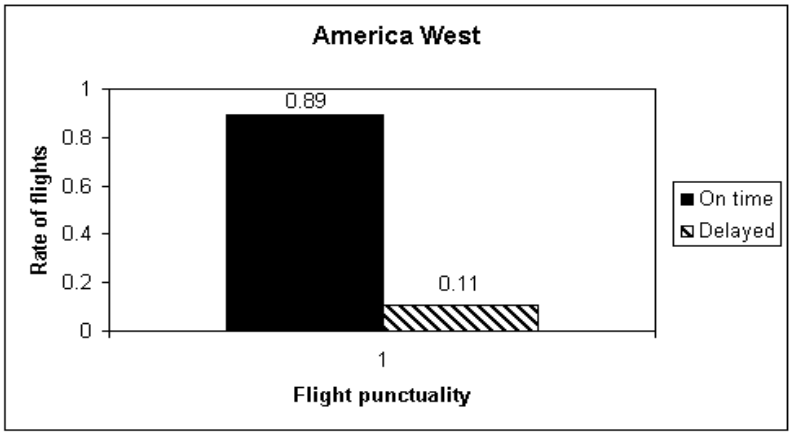
1. The two bar plot could easily be produced from a two way table of the two categories Airline (takes the value "America West" or "Alaska Airline") and flight punctuality (takes the values "Delayed" or "On time").
2. The two bar plots represent the number of flights on time or delayed after aggregation of the numbers for all five cities.
3. The two bar plot represent the conditional distribution of flight punctuality (on time or delayed) by airline.
4. The two bar plots represent the marginal distribution of flight punctuality (on time or delayed).

Answer 4

(b) An illustration of Simpson's Paradox for this example would be that:

1. "America West" did better than "Alaska Airlines" at three different airports, but "Alaska airlines" did better in the two other airports.

Continued...

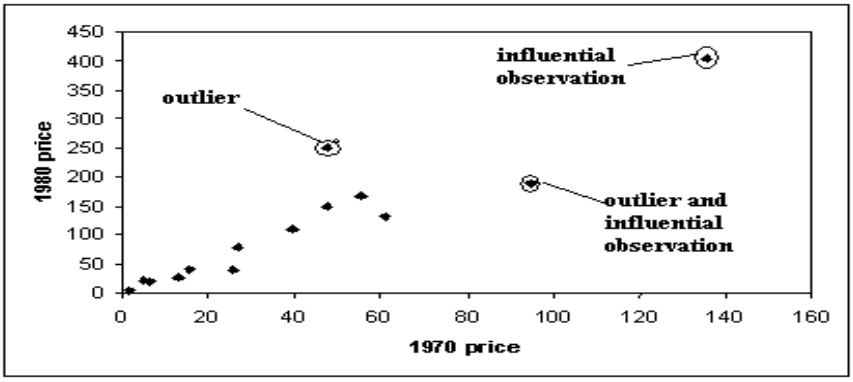


2. "America West" did better than "Alaska Airlines" in two different airports, but "Alaska Airlines" did better in the three other airports.
3. "America West" did worse than "Alaska Airlines" at all five airports.
4. "America West" did better than "Alaska Airlines" at all five airports.

Answer 3

Continued...

7. (24 points) The price of seafood varies with species and time. The following scatterplot represents the prices in cents per pound in 1970 and 1980 for several species.



(a) Describe the relationship of the prices in 1970 and prices in 1980.
Answer the relationship is linear, positive and fairly strong. There are a couple of outliers

(b) Flag any possible outliers in the scatterplot. Are there any influential observations?
Answer cf Graphic above

(c) A regression line is fit to the data, what are the values for slope, intercept and correlation?

1. 0.4 , 0.2 , 0.80.
2. 2.74, 4.44, 0.92.
3. 2.51, 6.27 , 2.11.
4. 1.79, 5.12 , -0.33.

Answer 2

(d) How much of the variation in the prices in 1980 is explained by the prices in 1970?

1. About 86%.
2. About 30%.
3. About 92%.
4. About 69%.

Answer 1. Note: the percent of variation explained by the regression is given by r^2, in this case $r^2 \simeq 86\%$.

(e) If the price of a fish was \$1 per pound in 1970, what would be the price of this fish in 1980 (Use your answer in (c) to answer this question)? Are you confident with this estimation? Why or why not?
Answer

$$y = 4.44 + 2.74 * (100) = 278.44 \text{ cents.}$$

Yes, I would be confident with this estimation because the linear regression is a good fit for this data ($r^2 \approx 86\%$, and there aren't many outliers), and the observation \$1 is in the range of the data.

(f) If we standardize both prices, what would be the new value of

- The slope of the regression of 1980 standardized price on 1970 standardized price.
- The correlation of 1980 standardized price and 1970 standardized price.

Answer: The correlation would stay the same. The new slope will be equal to the correlation.

End of exam.