Statistics 1040, Section 006, Midterm 2 (200 Points)

Friday, November 11, 2005

Your Name: _____

from: Stat 1040, Test 1, June 26, 2000 (Summer 2000), Question 6
& Stat 1040, Midterm 2, March 26, 2004 (Spring 2004), Question 1

Question 1: Regression **(50 Points)**

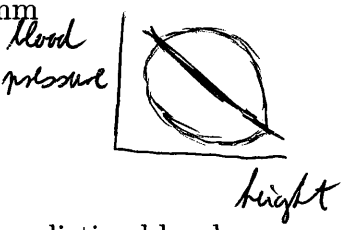
-2 each calculation error

From the subjects (all men) in a health survey, the following data were collected:

x Average height = 68 inches SD = 2.5 inches

y Average blood pressure = 120 mm SD = 15 mm

Correlation = -0.2.



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

The scatter diagram is football–shaped.

Show your work!

1. **(15 Points)** Find the regression equation for predicting blood pressure from height.

$\text{slope} = r \cdot \frac{SD_y}{SD_x} = -0.2 \cdot \frac{15}{2.5} = -1.2$ (6)

$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 120 - (-1.2) \cdot 68 = 120 + 81.6 = 201.6$ (6)

equation: $\boxed{\text{blood pressure} = 201.6 - 1.2 \cdot \text{height}}$ or $\boxed{y = 201.6 - 1.2 \cdot x}$ (3)

2. **(10 Points)** Using your regression equation, estimate the blood pressure of a man who is 73 inches tall.

-2 for old method, correct result
-8 for old method, incorrect result
-5 if result makes no sense

predicted blood pressure of a 73" tall man $= 201.6 - 1.2 \cdot 73$

$= 201.6 - 87.6 = \underline{114 \text{ mm}}$

3. **(15 Points)** Find the r.m.s. error for predicting the blood pressure from the height.

-5 for each major mistake, e.g., SD_x instead of SD_y, $\sqrt{}$ of everything, r instead of r^2, etc.

$\text{r.m.s. error} = \sqrt{1 - r^2} \cdot SD_y$

$= \sqrt{1 - (-0.2)^2} \cdot 15$

$= \sqrt{1 - 0.04} \cdot 15 = \sqrt{0.96} \cdot 15 = 14.69 \text{ mm} \approx \underline{14.7 \text{ mm}}$

4. **(10 Points)** The correlation coefficient tells us that, on average, taller men have **higher / lower** blood pressures than shorter men, and that the relationship between blood pressure and height is quite **strong / weak.** **(Circle the correct word in each pair of choices.)**

(5) (5)

from: FPP, chapter 20, p. 371-372, Question 3 [Solutions → Workbook]

Question 2: Chance Errors in Sampling (**40 Points**)

A group of 50,000 tax forms has an average gross income of \$37,000, with an SD of \$20,000. Furthermore, 20% of the forms have a gross income over \$50,000. A group of 900 forms is chosen at random for audit. To estimate the chance that between 19% and 21% of the forms chosen for audit have gross incomes over \$50,000, a box model is needed.

1. (**5 Points**) Should the number of tickets in the box be **900** or **50,000**?

   Circle your answer. (5)

   Note: 900 is the sample size (i.e., # draws)

2. (**5 Points**) Each ticket in the box shows

   (5) **a zero or a one** or **a gross income**

   Circle your answer.

   Note: 0: gross income less than or equal to \$50,000
   1: gross income over \$50,000

3. (**5 Points**) **True** or **false**: the SD of the box is \$20,000. (5)

   Circle your answer.

   Note: box SD = $\sqrt{0.20 \cdot 0.80} = 0.4$

4. (**5 Points**) **True** or **false**: the number of draws is 900. (5)

   Circle your answer.

5. (**12 Points**) Find the chance (approximately) that between 19% and 21% of the forms chosen for audit have gross incomes over \$50,000. **Show your work!**

   The chance is: 54.67 %

   box: [10,000 × 1, 40,000 × 0] (1)
   # draws = 900 (1)
   box avg = fraction of 1's = $\frac{10,000}{50,000} = 0.2$ (1)
   box SD = $\sqrt{\text{fraction of 1's} \cdot \text{fraction of 0's}} = \sqrt{0.2 \cdot 0.8} = 0.4$ (1)
   EV sum = $900 \cdot 0.2 = 180$
   SE sum = $\sqrt{900} \cdot 0.4 = 30 \cdot 0.4 = 12$ (1)
   [not required]

   $EV_{\%} = 20\%$ (1)
   $SE_{\%} = \frac{12}{900} \cdot 100\% = 1.33\%$ (1)
   s.u.: $\frac{19\% - 20\%}{1.33\%} = -0.75$ (1)
   19% 20% 21%
   $\frac{21\% - 20\%}{1.33\%} = 0.75$ (1)
   area between −0.75 and 0.75: 54.67% (3)

6. (**8 Points**) With the information given, can you find the chance (approximately) that between 9% and 11% of the forms chosen for audit have gross incomes over \$75,000? **Either find the chance, or explain why you need more information.**

   The chance is: / %

   We have no way to calculate this chance! We need to know the percentage of forms that have gross income over \$75,000 in order to find an EV and SE. (8)

   −7 if attempt to calculate a %

from: Stat 1040, Final Test, December 15, 2004 (Fall 2004), Question 2

Question 3: Chance/Probability (**50 Points**)

Six children attend a party. There are 7 party favors: 3 pink favors and 4 blue favors. The children are each given a party favor at random (which they keep — so this is without replacement).

-2 for each calculation error

Answer each of the parts independenly from the other parts. **Show your work!**

-2 if no final %

1. (**10 Points**) What is the chance that the first child gets a pink favor?

   The chance is: 42.86% %

   1st pink: $\frac{3}{7} = 0.4286 = 42.86\%$ (3 ← ④, 7 ← ④, = ②)

2. (**10 Points**) What is the chance that the second child gets a pink favor?

   The chance is: 42.86% %

   2nd pink (and we don't know what the 1st was): $\frac{3}{7} = 0.4286 = 42.86\%$ (3 ← ④, 7 ← ④, = ②)

3. (**10 Points**) If I see that the first 2 children received pink favors, what is the chance that the third child also gets a pink favor?

   The chance is: 20.00% %

   3rd pink, given 1st & 2nd also pink: $\frac{1}{5} = 0.2 = 20.00\%$ (1 ← ④, 5 ← ④, = ②)

4. (**10 Points**) What is the chance that the left–over favor is blue?

   The chance is: 57.14% %

   7th blue (and we don't know what the 1st through 6th were): $\frac{4}{7} = 0.5714 = 57.14\%$ (4 ← ④, 7 ← ④, = ②)

   Note: Think of it this way – we draw one at random, place it aside, and distribute the remaining ones. What is the chance the one drawn is blue? (4 ← ④)

5. (**10 Points**) What is the chance that the first two children get favors of different colors?

   The chance is: 57.14% %

   (1st pink & 2nd blue (given 1st pink)) or (1st blue & 2nd pink (given 1st blue)):

   $\frac{3}{7} \cdot \frac{4}{6} + \frac{4}{7} \cdot \frac{3}{6} = \frac{12}{42} + \frac{12}{42} = \frac{24}{42} = \frac{4}{7} = 0.5714 = 57.14\%$

   (① ② ①) (② ③ ① ② ①)

from: FPP, Chapter 18, p. 328, Question 9 [Solutions → Workbook]

Question 4: Normal Approximation for Probability Histograms (**30 Points**)

One hundred draws are made at random with replacement from a box with ninety–nine tickets marked "0" and one ticket marked "1". True or false? Circle your answer and explain (with box model calculations where necessary).

1. (**15 Points**) The sum will be around 1, give or take 1 or so.

(5) (**True**) / **False**

Explanation:

box: [99 × 0 1 × 1] (1) (1)

draws = 100

box avg = $\frac{1}{100} = 0.01$ (2)

box SD = $\sqrt{\frac{1}{100} \cdot \frac{99}{100}} = 0.0995 \approx 0.1$ (2)

$EV_{sum} = 100 \cdot 0.01 = 1$ (2) ✓

$SE_{sum} = \sqrt{100} \cdot 0.1 = 10 \cdot 0.1 = 1$ (2) ✓

2. (**15 Points**) There is about a 68% chance that the sum will be in the range 0 to 2.

True / (**False**) (10)

Explanation:

We cannot use the normal curve! The box is lopsided and has too many 0's. (5)

Recall, FPP, Chapter 18, p. 320, Figure 6. There we had a box with 9 × 0 and 1 × 1. After 100 draws, the probability histogram for the sum barely followed the normal curve. Now, our box is much, much, more lopsided – 99 100 draws clearly is not enough to justify the use of the normal curve.

Question 5: Sampling (**30 Points**)

Part 1: (20 Points)

For each of the following, decide whether this describes a simple random sample (SRS). **Just circle your answer.**

- (**5 Points**) A student newspaper asked readers to respond to the question "Do you think that there should be more student activities on the weekends?" An overwhelming 95 percent said "yes". The article reporting the results concluded that 95 percent of all students feels this way.

  This is a SRS: **yes** / (**no**) (5)

  Note: this is a voluntary response survey

- (**5 Points**) A researcher selects a sample from a list of all patients at one of five large hospitals in the following manner. A patient is chosen from the first 25 on the list, then every 25th patient from that point forward is selected.

  This is a SRS: **yes** / (**no**) (5)

  Note: this is a systematic sample

- (**5 Points**) Fifteen state parks are to be selected from 1000 state parks in such a way that each has an equal chance of being selected. A random number generator on a computer is used to select 15 integers between 1 and 1000. Based upon those integers, the state parks are selected from a numbered list.

  This is a SRS: (**yes**) / **no**

  (5)

- (**5 Points**) A researcher chooses a random sample of households, then interviews every member of the selected households.

  This is a SRS: **yes** / (**no**) (5)

  Note: this is a cluster sample

Part 2: (10 Points)

A population consists of 100 individuals who have been numbered from 1 to 100 for the purpose of taking a simple random sample of ten individuals. Which of the following sets of ten is most likely to be chosen as the sample?

Just circle the correct answer:

- (a) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- (b) 5, 10, 15, 20, 25, 30, 35, 40, 45, 50
- (c) 3, 17, 24, 39, 41, 47, 66, 73, 87, 96
- (d) These are all equally likely.

(10)

Note: In a SRS, each possible set of 10 (different) individuals has the same chance of being selected. This includes extreme cases such as (a) and (b) on the left.