

**Answer any and only four of the next five multiple-part questions**

I. (25 pts) The below article from *Autoblog* (July 1, 2009) claims that for 36 of 50 states car ownership is closely related to the political leaning of the state.

- A. With a 50-50 chance of any state having car ownership that matches the state's political leaning, what is the expected number of matches for the 50 states? **25 matches**

Let X be matches:  $E(X) = .5 * 50 = 25$

- B. With a 50-50 chance of any state having car ownership that matches the state's political leaning, what is the probability of observing 36 out of 50 states matching car ownership and political leaning? **0.00083**

$0.000833 = \text{BINOMDIST}(36, 50, 0.5, 0)$

- C. With a 50-50 chance of any state having car ownership that matches the state's political leaning, what is the probability of observing at least 36 out of 50 states matching car ownership and political leaning? **0.0013**

$0.998699 = \text{BINOMDIST}(35, 50, 0.5, 1)$

$0.001301 = 1 - A1$

- D. What did you have to assume to obtain the probabilities in parts C and B and expected value in part A?

**From state to state the probability of a match remains fixed at 0.5; that is, state matches are independent.**

- E. Based on your probability calculations and expected value is it reasonable to conclude that the *Autoblog* 36 of 50 states car ownership and political leaning occurred by chance? Explain.

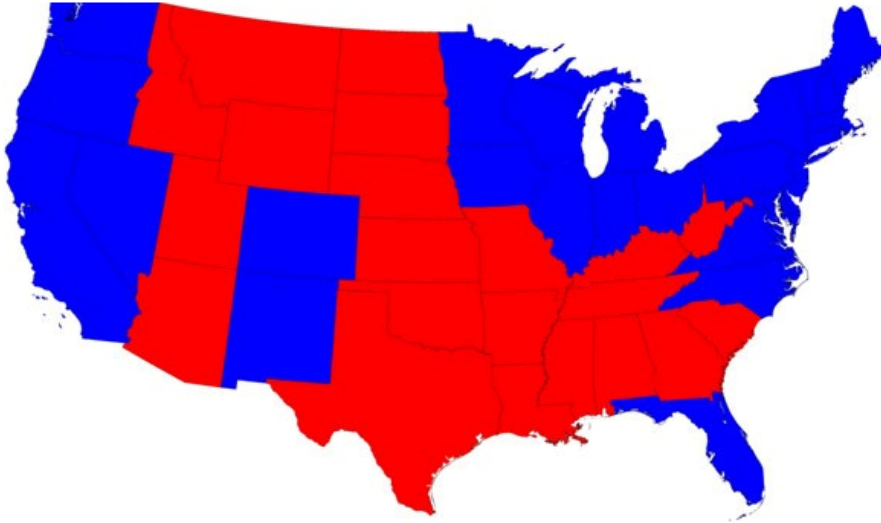
**NO, the likelihood of getting 36 or more matches, if 25 is expected, is too small to believe these observed results occurred by chance.**



Filed under: **Car Buying, Government/Legal**

## **Voter Drive: Study says political ideologies correlate with what cars we buy**

by **Chris Shunk** on Jul 1st 2009 at 12:57PM



A new study by RL Polk & Co. shows that the kind of car you drive is tied to the state you live in, and whether that state is traditionally Republican or Democrat-leaning. The study shows that 36 of the 50 United States tie closely to political lines.

Red states (Republican) typically buy domestic, while blue states (Democrat) typically skew import.

II. (25 pts) Students in S370 were confronted with calculating the expected dollar payoff and standard deviation for the following game of chance:

Payout	Probability of Payout
\$1.00	0.2
\$2.25	0.3
\$9.50	0.5

Student One said that the expected value and standard deviation were \$4.25 and \$4.59, based on the following EXCEL spreadsheet calculations:

	A	B	C	D
1	X	Prob(X)	X	
2	\$1.00	0.2		
3	\$2.25	0.3	Mean	4.25
4	\$9.50	0.5	Standard Error	2.649686
5			Median	2.25
6			Mode	#N/A
7			Standard Deviation	4.58939
8			Sample Variance	21.0625
9			Kurtosis	#DIV/0!
10			Skewness	1.588621
11			Range	8.5
12			Minimum	1
13			Maximum	9.5
14			Sum	12.75
15			Count	3

**Descriptive Statistics**

Input  
 Input Range:

Grouped By:  Columns  Rows

Labels in first row

Output options  
 Output Range:

New Worksheet Ply:

New Workbook

Summary statistics

Confidence Level for Mean:  %

Kth Largest:

Kth Smallest:

OK Cancel Help

Student Two said that the expected value and standard deviation were \$5.63 and \$4.11, based on the following EXCEL spreadsheet calculations:

	A	B	C	D
17	X	Prob(X)	Column1	
18	\$1.00	0.2		
19	\$2.25	0.3	Mean	5.625
20	\$9.50	0.5	Standard Error	1.299706
21			Median	5.875
22	\$1.00		Mode	9.5
23	\$1.00		Standard Deviation	4.110032
24	\$2.25		Sample Variance	16.89236
25	\$2.25		Kurtosis	-2.47938
26	\$2.25		Skewness	-0.04454
27	\$9.50		Range	8.5
28	\$9.50		Minimum	1
29	\$9.50		Maximum	9.5
30	\$9.50		Sum	56.25
31	\$9.50		Count	10

**Descriptive Statistics**

Input  
 Input Range:

Grouped By:  Columns  Rows

Labels in first row

Output options  
 Output Range:

New Worksheet Ply:

New Workbook

Summary statistics

Confidence Level for Mean:  %

Kth Largest:

Kth Smallest:

OK Cancel Help

A. Which, if either, neither or both, is or are correct? For each student state what is correct or wrong and why.

Student One ignored the probability weights completely and incorrectly used the descriptive statistics in EXCEL on just the population values.

Student Two recognized that the probability weights needed to be employed but did so as if these were sample data appropriate for EXCEL, which they are not because the weights are not relative frequencies but probabilities, imply population data.

- B. For this game of chance, at what price for playing the game would both the game provider and contestants expect to break even? What does this break even value tell you about repeated play of the game?
- C. What value could be used as a measure of risk for this game?

The break even price is the expected value (\$5.625), which says on average (over an infinite number of plays) neither the game provider nor the contestants will gain or lose. The standard deviation (\$3.90) could be employed as a measure of risk.

X	Prob(X)	XP(X)	dev <sup>2</sup> *P(X)
\$1.00	0.2	0.2	4.278125
\$2.25	0.3	0.675	3.417188
\$9.50	0.5	4.75	7.507813
		5.625	15.20313
			3.899118

III. (25pts) William Holstein, Managing Editor of *CEO Magazine*, reported on the *CNBC* that “20 percent of chief executive officers last less than one year in the job and the median tenure of a CEO is 4 years.”

- A. If time in the job (tenure) as a CEO is normally distributed, what is the 1) mean, 2) range, and 3) standard deviation of this distribution?

The mean is 4 years, i.e.,  $X \sim N(4, \sigma)$ . The range is infinity. To find out how many standard deviations away is the left-tail area of 20%, use  $\text{NORMSINV}(0.20) = -0.84$ . Therefore, the three-year difference between the four year (the median and mean) and the first year should correspond to 0.84 standard deviations. Therefore, the standard deviation is equal to  $3/0.84 = 3.57$  years.

- B. If time in the job (tenure) as a CEO is normally distributed, how likely would it be to find a CEO who lasts more than 8 years in the job? 13.1%  
 $1 - \text{NORMDIST}(8, 4, 3.57, 1) = 0.1313$ .

- C. Why is it reasonable or unreasonable to assume that CEO tenure is normally distributed? What are the implications of this for your calculations in part C?

Taking what *CNBC* reported as representative, it is not a good idea to assume that CEO tenure is normally distributed. First, because some CEO's are known to be in the job a real long time (20 to 30) years, we would suspect the distribution of time in the job to be right skewed. Second, no CEO could have a negative time in the job; yet, notice that the truncation point of 0 year is only 1.12 S.D. below the mean, and the probability of negative time getting  $P(\text{time} < 0) = P(Z < -1.12) = 0.13$ . If time were normal, then 13% would have to have negative time, but this is impossible. If the population of CEO tenure is not normal then for a sample of 4, the distribution of the sample mean could not be normal.

IV. (25 points) In “Congratulations, Grads – You’re Bankrupt” (*Business Week*) reporter Charles Haddad provided information showing that 95% of all college students had credit cards with some form of debt in 2000; 25% of those owed more than \$3,000, and 10% owed more than \$7,000. He also wrote that for those who have at least some credit card debt “since 1968, the average (mean) student’s debt from plastic (credit cards) jumped 46%, to \$2,748.”

- A. Construct a relative frequency distribution for the implied distribution of student credit card debt, including those who have zero credit card debt and the three classes on credit card debt that can be inferred from the information in the *BW* article.

### Distribution of Credit Card Debt for All Students

Class	Rel. Freq.
0	0.0500
0 to 3000	0.7125
3000 to 7000	0.1425
7000 +	0.0950
	1.000

- B. What is the mean credit card debt for all college students . . . ?  
 C. If all observations are assumed to take place at the midpoint of each class, what would be the midpoint of the class of more than \$7,000?

	Mid Pt	Some	0 or some	All students	Cell mean	Overall Mean
0			0.05	0.050	\$0.00	
0 to 3000	\$ 1,500.00	0.75		0.713		\$2,610.60
3000 to 7000	\$ 5,000.00	0.15	0.95	0.143	\$2,748.00	
7000 +	\$ 8,730.00	0.1		0.095		
				1.000		

- D. Can a standard deviation for student credit card debt be calculated? Why or Why not?

Without making arbitrary assumptions about the distribution of observations within each class a standard deviation cannot be calculated. Calculation of the standard deviation requires information on the occurrence of each individual value. For a sample standard deviation we also need to know the sample size.

V. (25pts) The following three investments (or games of chance) are available to you at a cost of \$100 per attempt. The net return for each option is distributed independently of the others and is normal. Option C is formed by placing \$60 in A and \$40 in B.

Option A: net return is  $a$       Option B: net return is  $b$       Option C: net return is  $c$

$$E(a) = \$12$$

$$\sqrt{E[a - E(a)]^2} = \$6$$

$$E(b) = \$13$$

$$\sqrt{E[b - E(b)]^2} = \$5$$

$$E(c) = \text{????}$$

$$\sqrt{E[c - E(c)]^2} = \text{????}$$

- A. The coefficient of variation for Option A is 0.5 (show work or give reason)

$$= \sqrt{E[a - E(a)]^2} / E(a) = \$6 / \$12 = 0.5\#$$

	A	B	C
E(x)	12	13	12.4

St. Dev.	6	5	4.118252
fraction	0.6	0.4	
Coef. Var.	0.5	0.384615	0.332117
P(x<0)	0.02275	0.004661	0.001302

B. The covariance of options A and B is 0 (show work or give reason)  
Because the net return for option A and B are distributed independently.

C. The expected net return for Option C is \$12.4 (show work or give reason)  
=  $0.6 \times \$12 + 0.4 \times \$13 = \$12.4$

D. The standard deviation for Option C is \$4.118 (show work or give reason)  
=  $\sqrt{0.6^2 \cdot \text{var}(a) + 0.4^2 \cdot \text{var}(b)} = \sqrt{0.36 \times 36\$sq + 0.16 \times 25\$sq} = \sqrt{16.96\$sq} = \$4.118$

E. The probability that Option A yields a negative net return is 0.025  
(show work or give reason)

Negative net return are those values that 2 SD below the mean, and because the expected net return is normally distributed, therefore the corresponding probability is approximate 0.025 =  $(1-0.95)/2$ , or precisely, = NORMDIST(0,12,6,1) = 0.02275

F. The probability that Option C yields a negative net return is 0.0013.  
(show work or give reason)

Similar to question 5, probability is = NORMDIST(0,12.4,4.118,1) = 0.0013

G. If you have the ability to play one and only one option, then which option

i. should yield you the most money in the “long term,”(infinite repeated play)? and why?  
Option B yields highest return in the long term, because it has the highest expected net return.

ii. has the highest potential to yield you the most money in one and only one attempt? and why?  
Pick a large value of net return and we compare the 3 options:

$$P(a > \$24) = 1 - \text{NORMDIST}(24,12,6,1) = \underline{0.02275}$$

$$P(b > \$24) = 1 - \text{NORMDIST}(24,13,5,1) = \underline{0.0139}$$

$$P(c > \$24) = 1 - \text{NORMDIST}(24,12.4,4.118,1) = \underline{0.0024}$$

Option A has the highest potential to yield net return more than \$24 in one and only one attempt because it has the higher probabilities for the bigger payouts.