Chapter 1: no key formulas.
Chapter 2: Relative Frequency=freq. of the class$/n$.
Approx. Class Width:
=(largest value-smallest value) /number of classes.
Chapter 3: sample and population means

$$\bar{x} = \sum x_i/n \text{ and } \mu = \sum x_i/N$$

Weighted mean and geometric mean

$$\bar{x} = \sum w_i x_i/w_i \text{ and } \bar{x}_g = [(x_1)(x_2)\ldots(x_n)]^{1/n}.$$

Interquartile Range: IQR $= Q_3 - Q_1$.
Population and sample variance

$$\sigma^2 = \frac{\sum(x_i-\mu)^2}{N} \text{ and } s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}.$$

Population and sample standard deviation

$$\sigma = \sqrt{\sigma^2} \text{ and } s = \sqrt{s^2}.$$

Coefficient of Variation

$$\left(\frac{\text{Standard deviation}}{\text{Mean}} \times 100\right)\%$$

z-Score: $z_i = \frac{x_i-\bar{x}}{s}$.
Population and Sample Covariance

$$\sigma_{xy} = \frac{\sum(x_i-\mu_x)(y_i-\mu_y}{N} \text{ and } s_{xy} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y}}{n-1}$$

Population and Sample Pearson Correlation

$$\rho_{xy} = \sigma_{xy}/(\sigma_x\sigma_y) \text{ and } r_{xy} = s_{xy}/(s_x s_y).$$

Chapter 4: Counting Rule for Combinations

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!}.$$

Counting Rule for Permutations

$$P_n^N = n!\binom{N}{n} = \frac{N!}{(N-n)!}.$$

Probability Rules: $P(A) = 1 - P(A^C)$

Chapter 4 continued:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A).$$

Multiplication Law for Independent Events

$$P(A \cap B) = P(B)P(A).$$

Bayes' Theorem

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

Chapter 5:
Discrete Uniform Probability Mass Function: $f(x) = 1/n$.
Expected Value of a Discrete R. V.: $E(x) = \mu = \sum x f(x)$.
Variance of a Discrete R. V.:

$$Var(x) = \sigma^2 = \sum(x-\mu)^2 f(x).$$

Number of Experimental Outcomes Providing Exactly x Successes in n Trials

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Binomial Probability Mass Function

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{(n-x)}.$$

Expected Value for Binomial Distribution: $E(x) = \mu = np$.
Variance for Binomial Distr.: $Var(x) = \sigma^2 = np(1-p)$.
Poisson Probability Mass Function:

$$P(X = x|\mu) = f(x) = \frac{\mu^x e^{-\mu}}{x!}.$$

Hypergeometric Probability Mass Function and Expected Value:

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \text{ and } E(x) = \mu = \frac{nr}{N}.$$

Chapter 5 continued: Variance for the Hypergeometric Distribution:

$$Var(x) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Chapter 6: Uniform PDF

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Normal PDF The density function is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Converting to the Standard Normal rv:

$$z = \frac{x-\mu}{\sigma}.$$

Exponential PDF and CDF for $x \ge 0$

$$f(x) = \mu^{-1} e^{-x/\mu} \text{ and } P(x \le x_0) = 1 - e^{-x_0/\mu}.$$

Chapter 7: expected value of $\bar{x}$

$$E(\bar{x}) = \mu.$$

Standard Deviation of $\bar{x}$ (Standard Error)

$$\sigma_x = \frac{\sigma}{\sqrt{n}}.$$

Expected Value and Std Dev (Standard Error) of $\bar{p}$

$$E(\bar{p}) = p \text{ and } \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Finite Pop. Correction Factor: $\sqrt{(N-n)/(N-1)}$.
Chapter 8: Interval Estimate of Population Mean, σ known and unknown

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ and } \bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Necessary Sample Size for Interval Estimate of μ

$$n = \frac{(z_{\alpha/2})^2\sigma^2}{E^2}$$

Chapter 8 continued: Interval Estimate of p

$$\bar{p} \pm z_{\alpha/2}\frac{p(1-p)}{\sqrt{n}}$$

Necessary Sample Size for Interval Estimate of p

$$n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2}$$

Chapter 9: Test Statistic for Hypothesis Tests About μ, σ known and unknown

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \text{ and } t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$

Test Stat for Hypothesis About p

$$z = \frac{\bar{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Chapter 10: Point Estimate and Standard Error for Difference in Two Population Means

$$\bar{x}_1 - \bar{x}_2 \text{ and } \sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Interval Estimate and Test Statistic for Difference in Two Means with Known Variances

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \text{ and } z = \frac{\bar{x}_1-\bar{x}_2-D_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Interval Estimate and Test Statistic for Difference in Two Means with Unknown Variances

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \text{ and } t = \frac{\bar{x}_1-\bar{x}_2-D_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

Degrees of Freedom for t, Two Independent Random Samples

$$df = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$$

Chapter 10 continued: Test Statistic (Matched Samples)

$$t = \frac{\bar{d}-\mu_d}{s_d/\sqrt{n}}$$

ANOVA Related:

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \quad s_j^2 = \frac{\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2}{n_j-1} \quad \bar{\bar{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n_T}$$

$$MSTR = \frac{SSTR}{k-1} \quad \text{SSTR} = \sum_{j=1}^{k} n_j(\bar{x}_j-\bar{\bar{x}})^2 \quad MSE = \frac{SSE}{n_T-k}$$

$$SSE = \sum_{j=1}^{k}(n_j-1)s_j^2 \quad F = MSTR/MSE$$

$$\text{SST} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{\bar{x}})^2 \quad \text{SST=SSTR+SSE}$$

Chapter 11: not covered in this course
Chapter 12: $y = \beta_0 + \beta_1 x + \epsilon$

$$E(y) = \beta_0 + \beta_1 x \quad \hat{y} = b_0 + b_1 x \quad b_0 = \bar{y} - b_1\bar{x}$$

$$b_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} \quad SSE = \sum(y_i = \hat{y}_i)^2$$

$$SST = \sum(y_i-\bar{y})^2 \quad SSR = \sum(\hat{y}_i-\bar{y})^2 \quad \text{SST=SSR+SSE}$$

$$r^2 = \frac{SSR}{SST} \quad r_{xy} = (\text{sign of } b_1)\sqrt{r^2} \quad s^2 = MSE = \frac{SSE}{n-2}$$

Standard Error of the Estimate, $s = \sqrt{MSE}$.

$$\sigma_{b_1} = \frac{\sigma}{\sqrt{\sum(x_i-\bar{x})^2}} \quad s_{b_1} = \frac{s}{\sqrt{\sum(x_i-\bar{x})^2}} \quad t = \frac{b_1}{s_{b_1}}$$

For simple regression, $MSR = SSR$ because there is only one independent variable.

$$F = \frac{MSR}{MSE} \quad s_{\hat{y}*} = s\sqrt{\frac{1}{n}+\frac{(x^*-\bar{x})^2}{\sum(x_1-\bar{x})^2}}$$

Confidence Interval for $E(y^*)$: $\hat{y}^*_{\alpha/a}s_{\hat{y}*}$

$$s_{\text{pred}} = s\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{\sum(x_1-\bar{x})^2}}$$

Chapter 12 continued: Prediction Interval for y^*:

$$\hat{y}^* \pm t_{\alpha/a}s_{\text{pred}}$$

Residual for Observation i: $y_i - \hat{y}_i$
Chapter 13:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

$$SST = SSR + SSE \quad R^2 = \frac{SSR}{SST}$$

$$R_a^2 = 1-(1-R^2)\frac{n-1}{n-p-1}$$

$$MSR = \frac{SSR}{p} \quad MSE = \frac{SSE}{n-p-1} \quad F = \frac{MSR}{MSE}$$

$$t = \frac{b_i}{s_{bi}}$$

Other Math Rule Reminders:

$$e^x = \exp(x)$$

$$\ln 1 = 0 \quad \ln e = 1$$

$$x! = (x)(x-1)(x-2)\cdots(2)(1)$$

$$0! = 1 \quad x^0 = 1$$

ECN221 Exam 1A Spring 2015, ASU-COX

Choose the best answer. Do not write letters in the margin or communicate with other students in any way. If you have a question note it on your exam and ask for clarification when your exam is returned. In the meantime choose the best answer. Neither the proctors nor Dr. Cox will answer questions during the exam. Dr. Cox will post a key the day after the exam or in the case of the final exam the day after all finals are given. Grades will be posted on Bb after scores are returned from the testing center.

Please check each question and possible answers thoroughly as questions at the bottom of a page sometimes run onto the next page.

Relax. You studied. You know the material. You can nail it.

1. The sample size
   (a) can be larger than the population size.
   (b) can be larger or smaller than the population size.
   (c) is always smaller than the population size.
   (d) is always equal to the size of the population.

2. Data collected over several time periods are
   (a) time controlled data.
   (b) time series data.
   (c) crossectional data.
   (d) time crossectional data.

3. Statistical inference
   (a) is the same as Data and Statistics.
   (b) refers to the process of drawing inferences about the sample based on the characteristics of the population.
   (c) is the same as a census.
   (d) is the process of drawing inferences about the population based on the information taken from the sample.

4. Product brand is an example of
   (a) categorical data.
   (b) quantitative data.
   (c) either categorical or quantitative data.

(d) time series data.

5. Consider the data set below:

customer information			
customer	$ spent	days since billing	tenure
1	74	3	25
2	989	25	79
⋮	⋮	⋮	⋮
76	4553	15	74

Which variables are quantitative?

(a) $ spent, days since billing, and tenure.

(b) customer and $ spent.

(c) customer, $ spent and tenure.

(d) they are all quantitative.

6. What is true about the data set below:

customer information			
customer	$ spent	days since billing	tenure
1	74	3	25
2	989	25	79
⋮	⋮	⋮	⋮
76	4553	15	74

(a) the data are time series.

(b) the data are time series and cross-sectional.

(c) the data are cross-sectional.

(d) some variables are time-series and some are cross sectional.

7. Which of the following is a quantitative variable.

(a) gender.

(b) education level.

(c) employment status.

(d) interest rate.

8. What is the mode in the following example?

stem	leaf
0	9
1	853
2	8552
3	7775422
4	9964

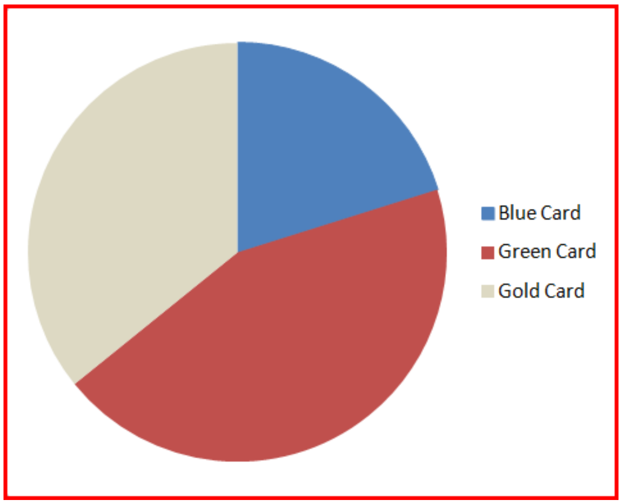
(a) 3.

(b) 7.

(c) 37.

(d) 49.

9. The display below is an example of a



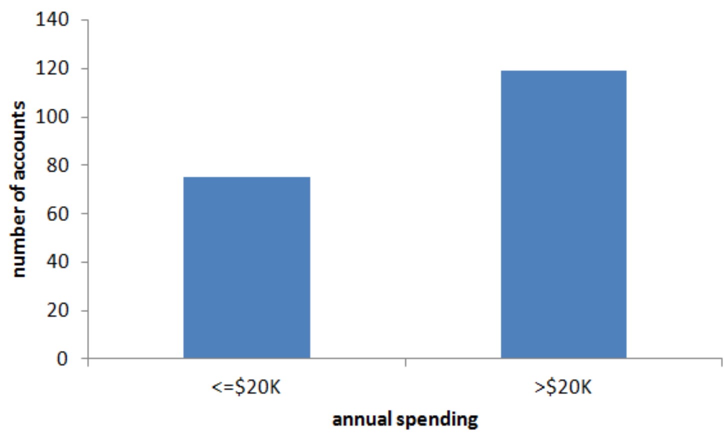
(a) histogram.

(b) bar chart.

(c) scatter plot.

(d) pie chart.

10. The display below has



(a) too many classes.

(b) the right number of classes.

(c) too few classes.

(d) a class range that is too narrow.

11. The relative frequency of platinum cards in the portfolio is:

charge/credit card type		
card	frequency	relative frequency
blue	143	
green	178	
gold	76	
platinum	25	

(a) 1.

(b) 5.9%.

(c) .180.

(d) .059.

12. The relative frequency of blue cards in the portfolio is:

charge/credit card type		
card	frequency	relative frequency
blue	143	
green	178	
gold	76	
platinum	25	

(a) .661.

(b) .421.

(c) 34%.

(d) .339.

13. The *cumulative* frequency of credit card accounts with a balance less than $30,000?

credit card balance		
card	cum. freq.	cum. relative freq.
<= 9,999	143	
10,000-19,999	321	
20,000-29,999	397	
30,000+	422	

(a) 25.

(b) 94%.

(c) 397.

(d) 422.

14. The *cumulative* relative frequency of cards with a balance under $20,000 is:

credit card balance		
card	cum. freq.	cum. relative freq.
<= 9,999	143	
10,000-19,999	321	
20,000-29,999	397	
30,000+	422	

(a) .761.

(b) .421.

(c) .339.

(d) .941.

15. A set of credit card accounts have the following estimated probabilities of default: .067, .743, .042, .022, .031, .194, .001. Find the 30th percentile.

(a) 2.1

(b) .036

(c) .031

(d) .022

16. The variance of a set of cereal prices is 1.38. The standard deviation is

    (a) 1.17
    (b) 1.90
    (c) 1.38
    (d) .69

17. Find the z-score for an observation with a value of 72.2 when the mean is 86.1 and the variance is 22.7.

    (a) -.61
    (b) .61
    (c) -2.92
    (d) 2.92

18. The average of 22, 25, 26, 29 is

    (a) 25.5
    (b) 25
    (c) 26
    (d) 26.5

19. The average is always greater than the standard deviation

    (a) true
    (b) false
    (c) true for ratio data but false for interval data
    (d) false for ratio data but true for interval data

20. Suppose the covariance between two variables is 517 while their individual standard deviations are 18 and 45.2. The correlation coefficient is:

    (a) .403.
    (b) 1.57.
    (c) .635.
    (d) .001.

21. Suppose WP Carey has 14 internships lined up with local companies. Suppose that there are 24 applicant for the internships that meet the GPA requirement; WP Carey will not select an applicant that does not meet the GPA requirement. How many possible combinations of students filling the internships are there?

   (a) 140089.
   (b) 1961256.
   (c) 9806280.
   (d) A really big number not given here.

22. A study by the Institute for Higher Education Policy found the values in the joint probability table below. The underlying data are for former college students that had taken out student loans. The table shows whether the student received a college degree versus whether they are successfully making their student loan payments. What is the probability that a former student completed their degree?

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

   (a) .58.
   (b) .42.
   (c) .26.
   (d) .16.

23. A study by the Institute for Higher Education Policy found the values in the joint probability table below. The underlying data are for former college students that had taken out student loans. The table shows whether the student received a college degree versus whether they are successfully making their student loan payments. What is the probability that a former student completed their degree and is currently satisfactory on their loan payments?

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

   (a) .26.

(b) .42.
(c) .92.
(d) .68.

24. A study by the Institute for Higher Education Policy found the values in the joint probability table below. The underlying data are for former college students that had taken out student loans. The table shows whether the student received a college degree versus whether they are successfully making their student loan payments. What is the probability that a former student did not complete their degree and is currently delinquent on their loan payments?

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

(a) .26.
(b) .74.
(c) .50.
(d) .34.

25. Suppose you observe the following prices for cereals which are given in dollars, 3, 3, 5, 5, 17 where the $17 cereal is a "high end" organic granola. Find the z-score of the organic granola.

(a) z=1.76, and it is an outlier.
(b) z=1.76, which means it is not an outlier.
(c) z=.298, which means it is not an outlier.
(d) z=2.88, which means it is close to being an outlier.

26. A study by the Institute for Higher Education Policy found the values in the joint probability table below. The underlying data are for former college students that had taken out student loans. The table shows whether the student received a college degree versus whether they are successfully making their student loan payments. What is the probability that a former student did not complete their degree given that they are currently delinquent on their loan payments?

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

(a) .34.
(b) .68.
(c) .50.
(d) .59.

27. Consider the *Let's Make a Deal* game we played in class. Instead of three doors suppose that there are four doors, a, b, c and d. What is the probability that you will win if you switch from your original guess after one door is opened? You can apply Bayes' Rule or some other technique to solve this.

(a) 1/4.
(b) 1/2.
(c) 2/3.
(d) 3/8.
(e) 6/8.

28. A study by the Institute for Higher Education Policy found the values in the joint probability table below. The underlying data are for former college students that had taken out student loans. The table shows whether the student received a college degree versus whether they are successfully making their student loan payments. What is the probability that a former student with a loan is delinquent on their loan payments?

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

(a) .16.
(b) .34.
(c) .50.
(d) .42.

29. I have checked that my ID is bubbled in correctly. If it is bubbled in incorrectly I will get this question wrong.

(a) True.
(b) False.

Key

1. c
2. b
3. d
4. a
5. a
6. c
7. d
8. c
9. d
10. c
11. d. To find the answer we need to complete the table. Then we can read the answer out of the table.

charge/credit card type		
card	frequency	relative frequency
blue	143	.339
green	178	.422
gold	76	.180
platinum	25	.059

12. d. To find the answer we need to complete the table. Then we can read the answer out of the table.

charge/credit card type		
card	frequency	relative frequency
blue	143	.339
green	178	.422
gold	76	.180
platinum	25	.059

13. c. The value is found from noting that there are a total of 422 observations and the values are found in the table:

credit card balance		
card	cum. freq.	cum. relative freq.
<= 9,999	143	.339
10,000-19,999	321	.761
20,000-29,999	397	.941
30,000+	422	1

14. a. The value is found from noting that there are a total of 422 observations and the values are found in the table:

credit card balance		
card	cum. freq.	cum. relative freq.
<= 9,999	143	.339
10,000-19,999	321	.761
20,000-29,999	397	.941
30,000+	422	1

15. c. c. $n = 7$ and $p = 30$ so the index point is $i = (7)(30)/100 = 2.1$ which we round up to 3. The third value is .031.

16. a. a. $\sqrt{1.38} = 1.17$.

17. c. c. Find the z-score by using the formula and plugging in,

$$z = \frac{72.2 - 86.1}{\sqrt{22.7}} = \frac{-13.9}{4.76} = -2.92.$$

18. a

19. b

20. c Compute:

$$r = \frac{517}{(18)(45.2)} = .635.$$

21. b. b. Use the formula

$$\binom{24}{14} = \frac{24!}{10!14!} = 1961256.$$

22. b. b. Take the value from the table.

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

23. a. a. Take the value from the table.

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

24. d. d. Take the value from the table.

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

25. b. b. $\bar{x} = 6.6$ and $s = 5.89$ so that

$$z = \frac{17 - 6.6}{5.89} = 1.76.$$

and we would not consider it to be an outlier because $-3 < z < 3$.

26. b. b. Take the values from the table to make the calculations.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.34}{.5} = .68.$$

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

27. d. d. Suppose that you pick a (or any door) and switch to d (or any unrevealed door) after b or c is revealed (or whichever other two doors there are).

$$P(d|\text{picked } a \cap (\text{b or c revealed}))$$
$$= \frac{P(d)P(\text{picked a} \cap (\text{b or c revealed})|d)}{K}$$

where $K = P(a)P(\text{picked a} \cap (\text{b or c rev.})|a) + P(b)P(\text{picked a} \cap (\text{b or c rev.})|b) + P(c)P(\text{picked a} \cap (\text{b or c rev.})|c) + P(d)P(\text{picked a} \cap (\text{b or c rev.})|d)$,

$$= \frac{(1/4)(1/4)}{(1/4)(1/6) + (1/4)(1/8) + (1/4)(1/8) + (1/4)(1/4)} = \frac{3}{8}$$

A simpler way to get the same answer would be to fix the idea that an initial guess would give the prize door with probability 1/4. That means the other doors are right with probability 3/4. Then, after one is revealed there are only two left so picking one of them gives you (1/2)(3/4)=3/8 as the probability you will get the right door.

28. c. c. Take the value from the table.

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

29. a