Math 217 – Module 5 Review

Module 5

Two-way Table

- A table that summarizes the relationship between two **categorical** variables: the explanatory variable and the response variable.
- The table cells can include either counts or **conditional percentages**.
- Use conditional percentages when comparing the distributions of the response variable.

For example, we investigated the relationship between body image and gender. We compared males to females.

	Underweight	About Right	Overweight	Total
Female	33	60	49	142
Male	21	53	20	94
Total	54	113	69	236

The explanatory variable is gender. The response variable is body image. The total number of participants in the survey is 236. The proportion of females that view themselves as "about right" is 60/142 = 0.423. The proportion of males that view themselves as "about right" is 53/94 = .564. Theses are examples of conditional percentages that can be used to compare how males and females view their own bodies.

Probability

- Probability is the likelihood or chance of an event occurring. For example, using the above data, the probability that a randomly selected adult thinks he/she is underweight is 54/236 = .229.
- The notation for this example is P(underweight) = .229

Three kinds of probability related to two-way tables:

- A **marginal probability** is the probability of a categorical variable taking on a particular value **without regard to the other categorical variable**. A marginal probability is calculated by dividing a row or column total by the table total (lower right cell). P(underweight) = .229 is an example of marginal probability.
- A **conditional probability** is the probability of a categorical variable taking on a particular value **given the condition that the other categorical variable has some particular value**. For example, the probability that a female thinks she is "underweight" is 33/142 = .232. Here the condition is that the adult is female, so we use the total number of females as the denominator. The notation is P(underweight|female) = .232.
- A **joint probability** is the probability that the **two categorical variables each take on a specific value**. For example, the probability that a randomly selected adult is both male and thinks he is overweight is 20/236 = .085. In calculating this probability we divide the count in one inner cell of the table by the overall total count. The notation is P(male and overweight) = .085.

Building Hypothetical two-way tables: used to compute conditional probabilities not given in the original information, such as the probability of a positive drug test for someone who does not use drugs. Usually you will be given a marginal probability **P(A)** and two conditional probabilites **P(B|A)** and **P(B|not A).**

- Start with a grand total of 10000.
- Determine the number for **A** and complement of A (**not A**) by taking 10000 times **P(A)** and using subtraction for **not A.**
- Determine the numbers for **B|A** and **B|not A** by multiplying the totals from the prior step times the appropriate conditional probability.
- Rest of the table can be completed by subtraction.

Module 6 Review

- The **probability** of an event is a measure of the likelihood that the event occurs. Probabilities are always between 0 and 1. The closer the probability is to 0, the less likely the event is to occur. The closer the probability is to 1, the more likely the event is to occur.

- Some common **probability rules**:
  - **Complement Rule**, P(not A) = 1 – P(A)
  - **Addition Rule**, P(A or B) = P(A) + P(B) - P(A and B)
  - If events are **mutually exclusive**, P(A or B) = P(A) + P(B), since P(A and B) = 0
  - When the knowledge of the occurrence of one event A does not affect the probability of another event B, we say the events are independent. (That is P(A|B) = P(A). If A and B are **independent events**, P(A and B) = P(A) · P(B)

- When we have a quantitative variable with outcomes that occur as a result of some random process (*e.g.,* rolling a die, choosing a person at random, etc.), we call it a **random variable.** There are two types of random variables:
  - **Discrete** random variables have numerical values that can be listed and often can be counted. We find probabilities using areas in a probability histogram.
  - We can find the **Expected Value** of a discrete random variable by using the formula $\mu = \sum x \cdot p(x)$
  - **Continuous** random variables can take any value in an interval and are often measurements. We use a density curve to assign probabilities to intervals of X-values. We use **the area under the density curve to find probabilities.**

- We use a **normal distribution curve**to model the probability distribution for many variables, such as weight, shoe sizes, foot lengths, and other human physical characteristics. For a normal curve, the **Empirical Rule for Normal Curves** tells us that 68% of the observations fall within 1 standard deviation of the mean, 95% within 2 standard deviations of the mean, and 99.7% within 3 standard deviations of the mean.

- To compare X-values from different distributions, we standardize the values by finding a Z-score**.**
  - Z-score measures how far X is from the mean in standard deviations. In other words, the Z-score is the number of standard deviations X is from the mean of the distribution. For example, Z = 1 means the X-value is one standard deviation above the mean.

$$Z = \frac{x - \mu}{\sigma}$$

- If we convert the X-values into Z-scores, the distribution of Z-scores is also a normal curve. This curve is called the **Standard Normal distribution.** We use TI calculator to find probabilities and percentiles for any normal distribution.

  - To find probabilities using X, use calculator:
    - $P(a \le X \le b) = NORMCDF(a, b, \mu, \sigma)$
    - $P(X \le b) = NORMCDF(-10 \wedge 99, b, \mu, \sigma)$
    - $P(X > a) = NORMCDF(a, 10 \wedge 99, \mu, \sigma)$
    - To Find the pth percentile, Use Calculator $X_p = INVNORM(p, \mu, \sigma)$
  - To find probabilities using Z-scores, first find $Z = \frac{x - \mu}{\sigma}$, then use calculator:
    - $P(a \le Z \le b) = NORMCDF(a, b)$
    - $P(Z \le b) = NORMCDF(-10 \wedge 99, b)$
    - $P(Z > a) = NORMCDF(a, 10 \wedge 99)$
    - To Find the pth percentile, Use Calculator $Z = INVNORM(p)$, then use formula $X = \mu + Z\sigma$

Math 217 NAME ____________________________________
(First) (Last)

Exam 3 (Modules 5 and 6) Practice Problems

1. A sample of American adults was asked the question: "Suppose you could only have one child. Would you prefer that it be a boy or a girl?" The results summarized by sex are given in the following table. (Data simulated from Newport, 2007).

Sex of Respondent	Preferred Sex of Child			**Total**
	Boy (B)	Girl (G)	No Preference (NP)	
Women (W)	174	196	140	
Men (M)	237	111	142	
Total				

a. Fill in the missing totals from the table.

b. What is the probability that a randomly chosen adult from this sample is a woman? Use probability notation. Is this a marginal, joint, or conditional probability?

c. What is the probability that a randomly chosen adult from this sample prefers a boy? Use probability notation. Is this a marginal, joint, or conditional probability?

d. What is the probability that a randomly chosen adult from this sample is a man and prefers a girl? Use probability notation. Is this a marginal, joint, or conditional probability?

e. What is the probability that a randomly chosen woman from this sample has no preference? Use probability notation. Is this a marginal, joint, or conditional probability?

f. What is the probability that a randomly chosen man from this sample prefers a girl?

g. What is the probability that a randomly chosen woman from this sample prefers a girl?

h. Who is more likely to prefer a girl in this study, men or women?

2. A Gallup poll taken in the first half of 2007 asked the question “Did the United States make a mistake in sending troops to Iraq?” It was found that 52 percent of Americans adults under 50 years old said yes, while 64 percent of Americans 50 years old or over said yes. Fifty percent of those polled were 50 years old or over (Newport, 2007).

Age Group	Did the US make a mistake in sending troops to Iraq?		**Total**
	Yes	No	
Under 50			
50 and older			
Total			10,000

a. Assuming a hypothetical population of 10,000 people fill in the following table.
b. What percent of the adults are under 50 and said yes?

c. What percent of the adults are 50 or over and said yes?

d. What percentage of the adults said yes?

e. What percentage of those who said yes are under 50?

3. In a study of hospital nurses from a Washington state hospital it was found that 92% of the nurses were female, that 74% had received their nursing degree in Washington state and 70% were both female and had received their nursing degree in Washington state. (Smith, 2007). If a nurse is chosen randomly from the hospital, find the probability of the following events.

a. Getting a nurse who did not receive their nursing degree in Washington state.

b. Getting a nurse who is either female or received their nursing degree in Washington state.

4. 10% of drivers will get in an accident in the next year. 60% of drivers have imported cars. Assume these are independent events

a. Find the probability a driver gets in an accident in the next year **and** is driving an imported car.

b. Find the probability a driver gets in an accident in the next year **or** is driving an imported car.

5. Thelesbian and bisexual student organization at a women's college is holding a 50-50 raffle as a fundraiser. They will sell 100 raffle tickets at $1 apiece, keep $50 for the organization and return $50 to the winner of the raffle. Suppose you buy one ticket. Remember that if you lose, your "winnings" are then negative.

    a. What is the probability distribution of your winnings?

    b. What is the expected value of your winnings?

6. When an automobile is brought into a car dealer, the number of hours required by a certain mechanic to determine the problem and repair it has the following probability distribution:

Hours worked, x	1	2	3	4	5	6	7	8
Probability	0.1	0.2	0.25	0.15	0.1	0.05	0.05	0.1

    a. What is the probability that the mechanic will work 5 or more hours on a car? Write the corresponding probability statement.

    b. What is the probability that the mechanic will work less than 4 hours on a car? Write the corresponding probability statement.

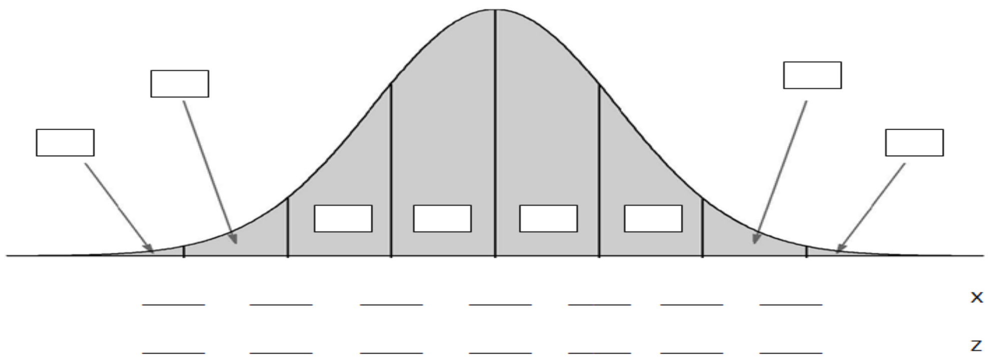
    c. What is the probability that the mechanic will work 2 to 5 hours, inclusive, on a car? Write the corresponding probability statement.

    d. What is the mean number of hours that the mechanic will work on a car?

7. Apples have a mean weight of 9 ounces and a standard deviation of 2 ounces. Oranges have a mean weight of 12 ounces and a standard deviation of 3 ounces. What is more unusual, an apple that weighs 13 ounces or an orange that weighs 17 ounces? Justify your answer with Z-score calculations.

8. At a local college, GPAs are distributed normally with a mean of 2.8 and a standard deviation of 0.4.

   a. Create a normal distribution graph. Demonstrate the Empirical Rule by labeling areas based on number of standard deviations that data values are from the mean. Include the X and the Z values.



   b. Use your graph to find the probability that a randomly selected student has a GPA
      i. Greater than 3.2.

      ii. Between 2.0 and 3.2.

      iii. Less than 2.8.

   c. Those students whose GPAs are in the top 2.5% make the Dean's List. Find the Z-score of the minimum GPA required to make the Dean's List.

   d. What is the minimum GPA required to make the Dean's List?

9. For salaried workers working for an advertising firm, the yearly salary is normally distributed with a mean of $42,100 and a standard deviation of $4,800. What proportion of these workers earn less than $34,000 per year?

10. Each American uses an average of 3 gallons of water flushing toilets each day (Carey, A. &Kereselidze, G., 2007). Suppose this is normally distributed with a standard deviation of 1 gallon.

   a. What percent of Americans use less than 0.5 gallon of water flushing toilets each day?

   b. Above how many gallons of water places an American in the heaviest 15% of users of water in flushing toilets each day?

Partial Answers

1. b. P(W)=0.51 Marginal c. P(B)=0.411 Marginal d. P(M and G) = 0.111 Joint
   e. P(NP|W) = 0.2745 Conditional f. P(G|M) = 0.2265 g. P(G|W) = 0.3843
   h. Women

2. b. P(Under 50 AND Yes) = 0.26 c. P(50 or over AND Yes) = 0.32 d. P(Yes) = 0.58
   e. P(Under 50 | Yes) = 0.4483

3. a. 0.26 b. 0.96

4. a. 0.06 b. 0.64

5. a.

x	\$49	-\$1
P(x)	0.01	0.99

   b. -\$0.50

6. a. $P(X \geq 5) = 0.3$ b. $P(X < 4) = 0.55$ c. $P(2 \leq X \leq 5) = 0.7$ d. 3.8 hours

7. Apple Z-score = 2.50 Orange Z-score = 1.67, so Apple is more unusual.

8. bi. $P(X > 3.2) = 0.16$ bii. $P(2.0 < X < 3.2) = 0.815$ c. Z = 2 d. Minimum GPA = 3.6

9. $P(X < 34{,}000) = 0.0458$

10. a. $P(X < 0.5) = 0.0062$ = 0.62% of the Americans use less than 0.5 gallon of water per day in toilets
    b. Above 4.04 gallons; Use the invNorm feature to find the 85th percentile.