Quiz #2 Solution

(1) An explanatory variable may explain or influence changes in a/an _________.
(a) quantitative variable, (b) categorical variable, (c) dependent variable, (d) independent variable.

(2) In the normal distribution with mean \( \mu \) and standard deviation \( \sigma \), approximately ________ of the observations fall within \( 2\sigma \) of \( \mu \).
(a) 68%, (b) 90%, (c) 95 %, (d) 99.7%.

(3) You have data for many families on the parents’ income and the years of education their eldest child completes. When you make a scatterplot, the explanatory variable on the \( x \) axis
(a) is families, (b) is years of education, (c) is parents’ income,
(d) can be either income or education.

(4) The points on a scatterplot lie very close to the line whose equation is \( y = 4 + 3x \). The correlation between \( x \) and \( y \) is close to
(a) \(-3\), (b) 1, (c) \(-1\), (d) 0.
(5) To find the regression line, the slope can be found by:
(a) \( b = r \frac{S_x}{S_y} \),
(b) \( b = r \frac{\sqrt{S_x}}{S_y} \),
(c) \( b = r \frac{S_y}{\sqrt{S_x}} \),
(d) \( b = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \).

(6) When coffee prices are high, farmers often clear forest to plant more coffee trees. Here are three years’ data on prices paid to coffee growers in Indonesia and the percent of forest area lost in a national park that lies in a coffee-producing region:

<table>
<thead>
<tr>
<th>price (cents per pound)</th>
<th>29</th>
<th>40</th>
<th>54</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest lost (percent)</td>
<td>0.49</td>
<td>1.59</td>
<td>1.69</td>
</tr>
</tbody>
</table>

The correlation is about
(a) \( r = 0.535 \),
(b) \( r = 0.912 \),
(c) \( r = 0.869 \),
(d) \( r = -0.515 \).

(7) From the observations 1, 6, 5, 7, 3, 8, 10, 3, 11, the third quartile \( Q_3 \) is equal to
(a) 10,
(b) 8,
(c) 8.5,
(d) 9.

(8) The following is the data of the math achievement test scores and the final calculus scores of a class of 4 students.
<table>
<thead>
<tr>
<th>student</th>
<th>Math Achievement test scores</th>
<th>final calculus scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>39</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>43</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>82</td>
</tr>
</tbody>
</table>

The regression line is

(a) \( y = 33.98 + 0.212x \),
(b) \( y = 28.99 + 0.128x \),
(c) \( y = 30.89 + 0.9324x \),
(d) \( y = 39.5817 + 0.7106x \).

**Solution:** Here \( n = 4 \),

\[
\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 664.25,
\]

\[
\sum_{i=1}^{n}(x_i - \bar{x})^2 = 934.75
\]

Then, we can get

\[
b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0.7106
\]

and

\[
a = \bar{y} - b\bar{x} = 69.25 - 0.7106 \times 41.75 = 39.5817
\]
(9) Scores on the Wechsler Adult Intelligence Scale are approximately normal distributed with $\mu = 100$ and $\sigma = 15$. What scores fall in the lowest 60.26% of the distribution?

(a) all $x$ satisfying $x < 96.5$,  
(b) all $x$ satisfying $x < 108.8$,  
(c) all $x$ satisfying $x < 103.9$,  
(d) all $x$ satisfying $x < 80.725$.

**Solution:** Let $X$ be the scores on the Wechsler Adult Intelligence Scale. Then, this question asks you to find $a$ such that

$$P(X \leq a) = 60.26\%$$

or equivalently

$$P\left(\frac{X - 100}{15} \leq \frac{a - 100}{15}\right) = 60.26\%.$$  

Since $\frac{X - 100}{15} \sim Z \sim N(0, 1)$, we have

$$P(Z \leq \frac{a - 100}{15}) = 60.26\%.$$  

From the Table A, we can find that

$$P(Z \leq 0.26) = 60.26\%.$$  

So we have

$$\frac{a - 100}{15} = 0.26.$$  

This gives that $a = 103.9$.

(10) Here are the data on the numbers of degrees earned in 2005-2006, as projected by the National Center for Education Statistics. The table entries are counts of degrees in thousands.
two-way table for degrees earned in 2005-2006

<table>
<thead>
<tr>
<th>Degree</th>
<th>Sex</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>total</td>
<td></td>
</tr>
<tr>
<td>Associate's</td>
<td>431</td>
<td>244</td>
<td>675</td>
<td></td>
</tr>
<tr>
<td>Bachelor's</td>
<td>813</td>
<td>584</td>
<td>1,397</td>
<td></td>
</tr>
<tr>
<td>Master's</td>
<td>298</td>
<td>215</td>
<td>513</td>
<td></td>
</tr>
<tr>
<td>Professional</td>
<td>42</td>
<td>47</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>Doctor’s</td>
<td>21</td>
<td>24</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>1,605</td>
<td>1,114</td>
<td>2,719</td>
<td></td>
</tr>
</tbody>
</table>

The conditional distribution of degree among female is

<table>
<thead>
<tr>
<th>(a) degree</th>
<th>(b) degree</th>
<th>(c) degree</th>
<th>(d) degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Female</td>
<td>Female</td>
<td>Female</td>
</tr>
<tr>
<td>Assoc</td>
<td>0.168</td>
<td>Assoc</td>
<td>0.118</td>
</tr>
<tr>
<td>Bach</td>
<td>0.612</td>
<td>Bach</td>
<td>0.712</td>
</tr>
<tr>
<td>Master’s</td>
<td>0.1212</td>
<td>Master’s</td>
<td>0.10</td>
</tr>
<tr>
<td>Profess</td>
<td>0.03</td>
<td>Profess</td>
<td>0.03</td>
</tr>
<tr>
<td>Doctor’s</td>
<td>0.068</td>
<td>Doctor’s</td>
<td>0.04</td>
</tr>
<tr>
<td>total</td>
<td>100%</td>
<td>total</td>
<td>100%</td>
</tr>
</tbody>
</table>

The conditional distribution of degree among female is
Example: Spell-check software catches "nonword errors" that result in a string of letters that is not a word, as when "the" is typed as "teh". When undergraduates are asked to type a 250-word essay (without spell-checking), the number $X$ of nonword errors has the following distribution:

<table>
<thead>
<tr>
<th>Value of $X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
<td>30%</td>
<td>10%</td>
</tr>
</tbody>
</table>

(a) Is the random variable $X$ discrete or continuous? Why?
(b) Write the event "at least one nonword error" in terms of $X$. What is the probability of this event?
(c) Find the probabilities $P(X \leq 2)$ and $P(X < 2)$.

Solution:
(a) $X$ is discrete because the sample space has only finite number possible outcomes.
(b) "at least one nonword errors" is the event $\{X \geq 1\}$ or $\{X > 0\}$. $P(\{X > 0\}) = 1 - P(\{X = 0\}) = 0.9$.
(c)

$$P(\{X \leq 2\}) = P(\{X = 0\}) + P(\{X = 1\}) + P(\{X = 2\})$$

$$= 0.1 + 0.2 + 0.3 = 0.6$$

$$P(\{X < 2\}) = P(\{X = 0\}) + P(\{X = 1\}) = 0.1 + 0.2 = 0.3$$

Example: A jar contains four coins: a nickel, a dime, a quarter, and a half-dollar. Three coins are randomly or equally likely selected from the jar.
a. List all the possible outcomes in the sample space $S$.
b. What is the probability that the selection will contain the half-dollar?
c. What is the probability that the total amount drawn will equal 0.6 dollar or more?

Solution:
a. Denote:
N: nickel;
D: dime;
Q: quarter;
H: half-dollar.
and $E_1 = (NDQ), E_2 = (NDH), E_3 = (NQH), E_4 = (DQH)$.
Then, $S = \{E_1, E_2, E_3, E_4\}$.
b. 
$$\Pr(\text{choose a half-dollar}) = \Pr(E_2) + \Pr(E_3) + \Pr(E_4) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$
c. All possible outcomes along with their monetary values follow:

$$E_1 = NDQ = 0.4$$
$$E_2 = NDH = 0.65$$
$$E_3 = NQH = 0.80$$
$$E_4 = DQH = 0.85$$

Hence,
$$\Pr(\text{total amount is 0.6 or more}) = \Pr(E_2) + \Pr(E_3) + \Pr(E_4) = \frac{3}{4}.$$
**Example:** A company has five applicants for two positions: two women and three men. Suppose that the five applicants are equally qualified and no preference is given for choosing either gender. Let $X$ equal the number of women chosen to fill the two positions. Find the probability distribution of $X$;

**Solution:** Define short notations as follows:
- $M_1$: Candidate man one;
- $M_2$: Candidate man two;
- $M_3$: Candidate man three;
- $W_1$: Candidate woman one;
- $W_2$: Candidate woman two.

The sample space

$$S = \{ \omega_1 = \{M_1 M_2\}, \omega_2 = \{M_1 M_3\}, \omega_3 = \{M_2 M_3\},$$

$$\omega_4 = \{W_1 M_1\}, \omega_5 = \{W_1 M_2\}, \omega_6 = \{W_1 M_3\}, \omega_7 = \{W_2 M_1\},$$

$$\omega_8 = \{W_2 M_2\}, \omega_9 = \{W_2 M_3\}, \omega_{10} = \{W_1 W_2\} \}$$

$$P(X = 0)$$

$$= P(\{\omega_1, \omega_2, \omega_3\})$$

$$= P(\{\{M_1 M_2\}, \{M_1 M_3\}, \{M_2 M_3\}\})$$

$$= P(\{M_1 M_2\}) + P(\{M_1 M_3\}) + P(\{M_2 M_3\})$$

$$= \frac{3}{10},$$
\[ P(X = 1) = P(\{\omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9\}) \]
\[ = P(\{\{W_1M_1\}, \{W_1M_2\}, \{W_1M_3\}, \{W_2M_1\}, \{W_2M_2\}, \{W_2M_3\}\}) \]
\[ = \frac{6}{10}, \]
\[ P(X = 2) = P(\{\omega_{10} = \{W_1W_2\}\}) = \frac{1}{10}. \]

Therefore,

The distribution table of r.v X

<table>
<thead>
<tr>
<th>Value of X</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>3/10</td>
<td>6/10</td>
<td>1/10</td>
</tr>
</tbody>
</table>

Remark: Probability is based on data about many repetitions of the same random experiment. Don’t be confused about that with the personal probability, which is just a personal judgment of how likely the outcome is.