

Name Key
 (Please ALSO write your name on the back of the last page.)

Ag 400 - Quiz 1
 Fall 2008

- A. In Mudville high school, there are 30 student clubs or organizations. Information on the number of members in these clubs is summarized below:

Number of Members in the Club	Midpoints \bar{X}	Number of Clubs	$\bar{X}F$
5 - 9 students	7	9	63
10 - 14 students	12	10	120
15 - 19 students	17	6	102
20 - 24 students	22	3	66
25 - 29 students	27	2	54
		30	405

- (6) 1. Calculate the mean number of students per club or indicate why the mean cannot be calculated.

$$\bar{X} = \frac{405}{30} = 13.5$$

- (6) 2. Calculate the Median for the above distribution.

$$\text{Md position} = \frac{30+1}{2} = 15.5 \quad \text{Md class } 9.5 \text{ to } 14.5$$

$$\text{Md} = 9.5 + \left(\frac{15.5-9}{10}\right)5 = 9.5 + 3 = 12.5$$

- (18) 3. Answer the following by indicating whether each of the statements is true (T) or false (F) in terms of the above data. If any part of a statement is untrue, it should be marked false (F). Add comments if you wish to clarify your answers.

- T a. In the above data, the variable, number of members in the club, is measured by an interval scale that is presented as grouped data with equal intervals.

- T b. The mode for the above data is 12.
- F c. The range for the above data is 8.
- T d. The distribution of number of members in a club is skewed to the right.
- T e. The standard deviation for the above data is the square root of the variance.
- T f. The figures calculated in (1) and (2) above are descriptive statistics that measure the central tendency of the distribution of number of members in the clubs.
- F g. The standard deviation and variance cannot be calculated from the above data.
- F h. The statistical unit in the above data is a student.
- F i. The independent variable here is number of clubs. The dependent variable is number of members in the club.

- B. The number of spelling errors (out of 15 words) made by a sample of 10 third grade students is recorded below:

Student Number	X Number of Spelling Errors Made	X^2
1	0	0
2	1	1
3	3	9
4	3	9
5	6	36
6	6	36
7	1	1
8	3	9
9	2	4
10	<u>5</u>	<u>25</u>
	30	130

- (5) 1. Calculate the mean number of words misspelled per student or indicate why the mean cannot be calculated.

$$\bar{X} = \frac{30}{10} = 3.0$$

- (5) 2. Calculate the standard deviation for the distribution of number of spelling errors.

$$S = \sqrt{\frac{130 - \frac{(30)^2}{10}}{10}} = \sqrt{4} = 2.0$$

(10) 3. Answer the following questions in terms of the above data by indicating whether each statement is true (T) or false (F). If any part of a statement is untrue, it should be marked false (F). Add comments if you wish to clarify your answers.

F a. The statistical unit here is a spelling error.

T b. The above distribution is presented as ungrouped data.

F c. The median number of spelling errors made per student is 6.0.

F d. The range for the above distribution is 9.

T e. The mode for this distribution is 3.

- (20) C. Suppose that you are interested in ascertaining whether a person's political affiliation is related to how he/she votes on Issue A. To that end, you survey a random sample of Mudville voters and compile the following information:

Political Affiliation	Vote on Issue A	
	Yes	No
	-----number of voters-----	
Republican	18 60%	12
Democrat	25 50%	25
Independent	32 40%	48
	75	85

Indicate whether each of the following statements is true (T) or false (F) in terms of the above data:

- T 1. The statistical unit is a voter.
- T 2. The dependent variable, "Vote on Issue A," is measured by a two category nominal scale.
- F 3. In these data, Republicans were less likely to vote "yes" than were Democrats or Independents.
- F 4. "Independents" were the most likely of the three political affiliation categories to vote "yes;" they were also the most likely to vote "no."
- F 5. The above data are presented in a contingency table where "political affiliation" is the Dependent Variable and "vote on Issue A" is the Independent Variable.
- F 6. The mean score for the distribution of number of voters is 26.67.
- F 7. The mode is 48 for the above data.
- F 8. The ratio of Republicans to Independents in the sample .6:1 or .6.
- T 9. Sixty percent of the Republicans in this sample voted "yes" but only 40% of the Independents voted "yes."
- T 10. In the total sample, Mudville residents were more likely to vote "no" than "yes" but this was true only for Independents, not for Republicans or for Democrats.

- (30) D. A survey of a random sample of Penn State graduate students carried out last year obtained (among other variables) information on each student's age (in years). The attached output shows the distribution of age (in years) for this sample and the distribution of a second variable derived from age and called AGECE. AGECE is defined in terms of the following categories:

Code	Age in Years
1	< 25 years
2	25 - 29 years
3	30 - 34 years
4	35 years and older

Indicate whether each of the following statements is true (T) or false (F) in terms of these data. If any part of a statement is untrue, it should be marked false (F). Feel free to add comments as you wish to clarify your responses.

- T 1. The statistical unit for this analysis is a graduate student.
- T 2. The variable AGE is measured by an interval scale.
- F 3. AGECE presents AGE as grouped data with equal intervals. If only the AGECE data were available you could not calculate the mean and standard deviation for AGE.
- T 4. The attached output was obtained using the task procedure called FREQUENCIES.
- T 5. There are 357 rows and a least two columns in the data file used for obtaining this output.
- F 6. The figures reported in the box labeled "statistics" at the top of Page 1 are actually parameters that describe the central tendency and dispersion of the distribution of AGE and AGECE in the population.
- F 7. AGECE could have been derived using RECODE (into a different variable) and specifying the following categories:

(22 THRU 24 = 1)
 (25 THRU 29 = 2)
 (30 THRU 34 = 3)
 (35 THRU 999 = 4)

- T 8. The distribution of AGE is skewed to the right.
- T 9. The range in values for AGE in these data equals 34; the variance is 39.6.
- T 10. The mode for AGE in this data set is 22.
- T 11. In this sample 12.9% of the 357 graduate students who participated in the survey did not provide information on their ages, and 12.3% reported that they were just 22 years old.
- T 12. Of those graduate students in the sample who reported their ages, 74% indicated that they were 30 years of age or younger.
- T 13. The median age of graduate students who participated in the survey and reported their ages equals 26 years.
- T 14. The distribution of AGE on Page 1 of the output is in the form that we have called semi-grouped data.
- F 15. To calculate the mean age for the sample of graduate students in this data set, you could use the COMPUTE command in SPSS.

Statistics

		AGE	AGEC
N	Valid	311	311
	Missing	46	46
Mean		27.95	2.06
Std. Deviation		6.294	1.036

Frequency Table

AGE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	22	44	12.3	14.1	14.1
	23	41	11.5	13.2	27.3
	24	32	9.0	10.3	37.6
	25	25	7.0	8.0	45.7
	26	26	7.3	8.4	54.0
	27	17	4.8	5.5	59.5
	28	23	6.4	7.4	66.9
	29	9	2.5	2.9	69.8
	30	13	3.6	4.2	74.0
	31	15	4.2	4.8	78.8
	32	14	3.9	4.5	83.3
	33	6	1.7	1.9	85.2
	34	5	1.4	1.6	86.8
	35	3	.8	1.0	87.8
	36	7	2.0	2.3	90.0
	37	5	1.4	1.6	91.6
	38	4	1.1	1.3	92.9
	39	1	.3	.3	93.2
	40	4	1.1	1.3	94.5
	41	1	.3	.3	94.9
	42	2	.6	.6	95.5
	43	3	.8	1.0	96.5
44	3	.8	1.0	97.4	
45	1	.3	.3	97.7	
46	2	.6	.6	98.4	
47	1	.3	.3	98.7	
51	1	.3	.3	99.0	
54	2	.6	.6	99.7	
56	1	.3	.3	100.0	
Total	311	87.1	100.0		
Missing	999	46	12.9		
Total	357	100.0			

AGEC

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 less than 25 yrs	117	32.8	37.6	37.6
	2 25-29 years	100	28.0	32.2	69.8
	3 30-34 years	53	14.8	17.0	86.8
	4 35 years and older	41	11.5	13.2	100.0
	Total	311	87.1	100.0	
Missing	System	46	12.9		
Total		357	100.0		