1. (15%) The Bureau of Weather has documented the maximum temperatures for July in a specific city between 1920 and 2020, with a mean of 33.2 degrees and a standard deviation of 5.1. Meteorologists randomly select 36 years from these records, averaging the maximum temperatures for July in these samples and recording the sample mean. They repeat this process 1,000 times to create a distribution of sample means.

   (1) What would you anticipate the shape of this distribution to be—symmetric, right-skewed, or left-skewed? Explain your reasoning. (5%)
   (2) Determine the variability of this distribution and specify the appropriate term used to describe this value. (5%)
   (3) If these meteorologists decide to select random samples of 9 years, recording the sample mean of the maximum temperature, and repeat this process 1,000 times to form a new distribution of sample means, how would the variability of this new distribution compare to the variability of the original distribution? (5%)

2. (15%) Social scientists surveyed 196 randomly selected undergraduates at a college regarding the number of welcome parties they attended during their freshman year. The histogram (**Figure 1**) displays the distribution of this sampled data, with an average count of 3.2 and a standard deviation of 2.0.

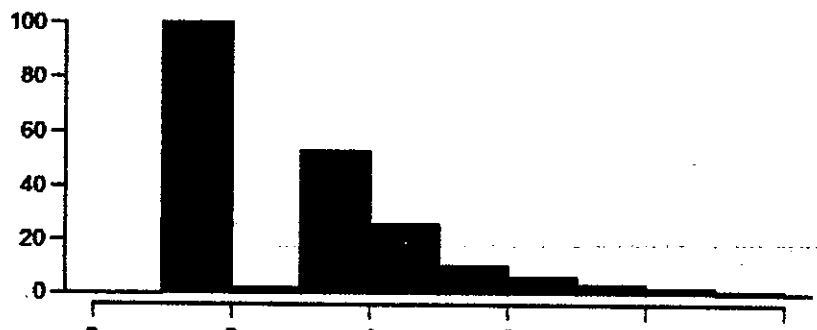

Figure 1

   Estimate the average number of welcome parties attended by the college student using a 90% confidence interval and provide an interpretation of this interval within the context. (10%)
   Additionally, describe the possible biases and assumptions required for statistical inference in this case before performing calculations and drawing conclusions. (5%)

3. (40%) The Department of Agriculture compiled data in **Table 1** from a 23-year study on crop cold damages. The table presents temperatures in Fahrenheit（華氏溫度）and the reported number of damaged farms (in hundreds) due to low temperatures.

Table 1

Year ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Temperature $[^\circ F]$	53	57	58	63	66	67	67	67	68
Damaged farms	5	3	6	2	0	0	0	0	0
Year ID	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Temperature $[^\circ F]$	69	70	70	70	70	72	73	75	75
Damaged farms	0	1	0	1	0	0	0	0	1
Year ID	**19**	**20**	**21**	**22**	**23**				
Temperature $[^\circ F]$	76	76	78	79	81				
Damaged farms	0	0	0	0	0				

見背面

(1) Officials are discussing the observed relationship between temperatures and damaged farms. Please suggest potential probability distributions and the type of regression that would be appropriate for modeling this relationship and describe your reasons (10%).

(2) Researchers created a new variable "Damage" coded as "1" for damaged farms > 0 and "0" for no damaged farms. Please suggest the suitable *likelihood* function for regression that could effectively model the relationship between "Temperature" and "Damage" and describe your reasons. Additionally, explain the specific *link* function that could be used to establish a linear relationship between these two variables (15%).

(3) **Table 2** displays regression results from the aforementioned question (2). Describe the relationship between "Temperature" and "Damage" based on the Table, and provide interpretations of how decreases in temperature might impact "Damage" (15%).

Table 2: Regression results

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	11.6630	3.2963	3.54	0.0004
Temperature	-0.2162	0.0532	-4.07	0.0000

4. (30%)

(1) The measurements of two variables across six samples are displayed in Matrix A. Determine the sample covariance matrix based on this data. (15%)

$$A = \begin{bmatrix} 3 & -3 & 1 & 7 & -4 & -4 \\ 7 & -7 & -1 & 8 & -6 & -1 \end{bmatrix}$$

The sample covariance matrix is $S = AA^T/(n-1)$.

(2) Assume there is a new axis with unit vector of $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$. Calculate the variances on the new axis. (15%)

試題隨卷繳回