

Concepts:

1. In probability theory, what is a distribution?
2. Give an example when the binomial distribution can be applied.
3. Why is the Poisson distribution a good choice for describing counts?
4. What is the interpretation of the joint probability π_{ij} ?
5. What is the interpretation of the probability $P(Y = 1|X = 2)$?
6. What is a contingency table?
7. Which question can be answered with a test for independence?
8. Assume the standardized cell residual for "disease present" and "clinical test positive" equals 5.4. Does this indicate that the clinical test is helping in predicting the presence of the disease?
9. What is the rationale for using Maximum likelihood estimators?
10. What is the difference between Wald and score statistics?
11. If $H_0 : \pi = \pi_0$ is rejected at significance level of 5% using a Wald test, would π_0 be in the 95% Wald confidence interval for π ?
12. Interpret a relative risk of $rr=3$.
13. interpret odds=3.
14. Interpret an odds ratio of 3.
15. What is the rational to test for conditional independence in three way tables?
16. What are the three components of a generalized linear model? Describe the role each component plays in the model.
17. What is measured by Kendall's τ ?
18. What is measured by Spearman's correlation coefficient?
19. Name the random component and link function in the following generalized models:
 - multiple linear regression
 - linear probability model
 - log linear regression
 - Logistic regression
 - ANOVA

20. What is the problem with the linear probability model?

Calculations + Interpretations:

21. Are pesticides present less often in organic foods? The consumers union in the US led a study based in sampling carried out by the US Department of Agriculture and the state of California. The sampling was done as part of regulatory monitoring of foods for pesticides residues. In the table frequencies for "food type" and "pesticide status are displayed".

Food type	Pesticide Status	
	Present	Not Present
Organic	29	98
Conventional	19485	7086

- Use the data to test if the probability for the presence of pesticides is different in the two food groups.
- Find a confidence interval for the relative risk of pesticides in the two groups. Interpret.
- Are the results from (a) and (b) consistent?
- Estimate the odds ratio for presence of pesticides and the two food groups. Interpret.

22. Use the data above to estimate:

- $P(\text{organic AND no pesticide})$
- $P(\text{organic} \mid \text{no pesticide})$
- $P(\text{no pesticide} \mid \text{organic})$

23. The following table summarizes data on religiosity and gender:

Gender	Religiosity			
	Very	Moderately	Slightly	Not
Female	170	340	174	95
Male	98	266	161	123

- Find the expected cell count for the cell Gender=Female/Religiosity=Very
- Find how much this cell would add to Pearson's χ^2 statistic.
- Use $\chi_0^2 = 20.628$ to test if Religiosity and Gender are independent.
- Find the standardized residual for the cell Gender=Female/Religiosity=Very. Interpret.

24. In an Italian study with 100 randomly selected adults researchers related income(in thousands of Euros) with wether a person possesses a travel credit card. The results on the logistic regression are reported in the table below.

Predictor	beta	SE	Z	P
Constant	-3.518	0.71034	-4.95	0.000
income	0.1054	0.02616	4.03	0.000

- (a) State the prediction equation for estimating the odds of owning a travel credit card.
- (b) State the prediction formula for estimating the probability of owning a travel credit card.
- (c) For subjects with income $x = 50$ thousand Euros find the estimated probability of owning a travel credit card.
- (d) Interpret $\hat{\beta}$ or $e^{\hat{\beta}}$, whatever you prefer.
- (e) Find the odds ratio of owning a travel credit card for subjects with income $x = 50$ and subjects with income $x = 65$. Interpret.
- (f) Using the linear probability model the model equation is $\hat{\pi}(x) = -0.159 + 0.0188x$. Use this for $x = 65$. Is the result reasonable?

25. In the data on religiosity and gender using scores female=1, male=2, very=5, moderately=3, slightly=2, not=0.

Kendall's $\tau = -0.108$, and Spearman's correlation coefficient = -0.117. They are both significantly different from 0.

Interpret these results.