Problem 1: This study aimed to evaluate the diagnostic value of the whole-blood interferon γ (INF-) assay for the diagnosis of latent tuberculosis (TB) infection. Table 1 shows the INF- γ assay and tuberculin skin test (TST) data for the group with low risk of infection (n=99).

Table 1

		TST		
		Positive	Negative	Total
IFN- γ	Positive	4	0	4
	Negative	46	49	95
	Total	50	49	99

Please compute the Kappa statistics. (20 points)

Problem 2: This study aimed to determine the utility of APOE gene ε4 allele in the diagnosis of Alzheimer's disease (AD). Table 2 shows the data. The sample included 2188 patients with dementia.

Table 2

***APOE* Genotype**	**Pathological Diagnosis**	
	Alzheimer's disease	Other causes of dementia
≧1 *e4* alleles	1142	133
No *e4* alleles	628	185
Total	1770	418

Please compute the sensitivity and specificity of the presence of an *APOEe4* allele, with pathologically confirmed Alzheimer's disease used as the standard. (20 points)

Problem 3: Table 3.presents a two by two contingency table.

Table 3

Group	Outcome		
	Presence	Absence	Total
Nonexposed	Y_1	$n_1 - Y_1$	n_1
Exposed	Y_2	$n_2 - Y_2$	n_2
Total	$Y_1 + Y_2$	$n_1 + n_2 - Y_1 - Y_2$	$n_1 + n_2$

(a) Assume that $Y_1 \sim Binomial(n_1, P_1)$ and $Y_2 \sim Binomial(n_2, P_2)$ such that

$$\Pr(Y_1 = y_1) = \binom{n_1}{y_1} P_1^{y_1}(1-P_1)^{n_1-y_1}, y_1 = 0,\ldots,n_1;$$

$$\Pr(Y_2 = y_2) = \binom{n_2}{y_2} P_2^{y_2}(1-P_2)^{n_2-y_2}, y_2 = 0,\ldots,n_2$$

In addition, the Y_1 and Y_2 are statistically independent. Please show that when $P_1 = P_2$

$$\Pr(Y_1 = y_1 \mid Y_1 + Y_2 = t) = \frac{\binom{n_1}{y_1} \times \binom{n_2}{y_2}}{\binom{n_1+n_2}{t}}$$

where $t = y_1 + y_2$. (15 points)

We want to test whether the exposure is associated with the presence of the outcome. The null hypothesis is that $H_0 : P_1 = P_2$ and the alternative hypothesis is either one-tailed or two-tailed. The Fisher's exact test uses the conditional distribution derived in (a) for computing p-value.

(b) Table 4 shows a data set taken from an experimental study. Please compute the probability of $\Pr(Y_1 \le 1 \mid Y_1 + Y_2 = 5)$ under the null hypothesis (i.e., p-value). (15 points)

Table 4

Group	Cured		
	Yes	No	Total
Control	$Y_1 = 1$	$n_1 - Y_1 = 9$	$n_1 = 10$
Intervention	$Y_2 = 4$	$n_2 - Y_2 = 6$	$n_2 = 10$
Total	$Y_1 + Y_2 = 5$	$n_1 + n_2 - Y_1 - Y_2 = 15$	$n_1 + n_2 = 10$

Problem 4: Let X represent the outcome variable in a two parallel group study and it is a continuous random variable. Let $X_1,\ldots,X_m$ represent a sample from the first group (e.g., the intervention group) and $X_{m+1},\ldots,X_{m+n}$ represent a sample from the second group (e.g., the control group). Assume that $X_1,\ldots,X_m$ follow a probability distribution $F_1(x)$ and $X_{m+1},\ldots,X_{m+n}$ follow a probability distribution $F_2(x)$. We want to test whether $F_1(x) = F_2(x)$. Let $R_1,\ldots,R_m$ denote the ranks of the first group, and $R_{m+1},\ldots,R_{m+n}$ denote the ranks of the second group when $X_1,\ldots,X_m, X_{m+1},\ldots,X_{m+n}$ are ordered by their values. Under the null hypothesis of $H_0: F_1(x) = F_2(x)$, please show that the Wilcoxon rank sum statistics for the first group given as

$S_1 = R_1 + \ldots + R_m$

has mean and variance, respectively, as follows:

$$E(S_1) = \frac{m \times (m+n+1)}{2} \quad \text{(15 points)}$$

and

$$Var(S_1) = \frac{mn(m+n+1)}{12}. \text{ (15 points)}$$

Hint: S_1 is the sample total of ranks obtained from a population of ranks consisting of {1,…, m+n} and the sample size is m.