MATH334/STAT434 **Spring 2021**

Homework 10

Problems 1-11: Each problem is worth 20 points. Problem 12 is optional.

1. Let $X_1, ..., X_{25}$ be a sample from a normal distribution with mean μ and having a variance of 100.
(a) Consider testing μ with the sample mean $\overline{X}$ as the test statistics, find the rejection region for a test at level $\alpha = .10$ of $H_0 : \mu = 0$ versus $H_a : \mu = 1.5$. What is the power of the test? Repeat for $\alpha = .01$.
(b) Determine the power functions for the two tests in part (a) for testing $H_0 : \mu = 0$ versus $H_a : \mu > 0$. Graph the power functions and overlay the two curves on the same plot; make sure you add legends to the plot. Discuss your finding and comment
(c) Repeat part (b) but for sample size of 50 instead of 25. Overlay all four power function curves ($\alpha = .05, n = 25$ and $50; \alpha = .10, n = 25$ and 50) on the same plot; make sure you add legends to the plot. Discuss your finding and comment

2. Read Example 8.1.1 on page 470 before you answer the following questions. Let X be a discrete random variable with four possible values a_1, a_2, a_3, a_4. Suppose the distribution of X follows one of the following two possible distributions (described in H_0 and H_1):

X	H_0	H_a
a_1	.2	.1
a_2	.3	.4
a_3	.3	.1
a_4	.2	.4

(a) Compare the likelihood ratio, Λ, for each possible value X and order the a_i according to Λ.

(b) What is the likelihood ratio test of H_0 versus H_a at level $\alpha = .2$? What is the test at level $\alpha = .5$?
(c) Describe briefly how to answer (a) and (b) when a sample of size 2 is given.

3. Read Definition of UMP on page 479 and Example 8.2.4 on page 481 before you answer the following questions. Let $X_1, ..., X_n$ be a sample from a Poisson distribution with mean λ.
(a) Determine the Uniformly Most Powerful (UMP) test for testing $H_0 : \lambda = 1$ versus $H_a : \lambda > 1$. Find the rejection region when $n = 10$ at level $\alpha = 0.05$.
(b) Compute the power for the test in (a) when $\lambda = 1.5$. Briefly describe the meaning of the resulting value.
(c) Plot the power function of the test in (a).
(d) Use simulation to verify your interpretation of the power in (b). Use at least 2000 simulation runs.

4. You may use the derived UMP test from the previous problem, Problem 3. Let $X_1, ..., X_n$ be a sample from a Poisson distribution with mean λ.
(a) Determine the exact power function of the UMP test for testing the null hypothesis

$H_0 : \lambda = 2$ versus the alternative hypothesis $H_a : \lambda > 2$ when $n = 25$ and at level $\alpha = 0.05$. Compute the power for the test when $\lambda = 3$
(b) Determine the approximate power function (Hint: use Central Limit Theorem) of the UMP test for testing the null hypothesis $H_0 : \lambda = 2$ versus the alternative hypothesis $H_a : \lambda > 2$ when $n = 25$ and at level $\alpha = 0.05$. Compute the power for the test when $\lambda = 3$. Is the value close to the exact power obtained in (a)?
(c) Obtain a graph of the exact power function you obtained in (a) and overlay the approximate power function you obtained in (b); make sure you add legends to the plot.

5. Let $\{X_1, ..., X_n\}$ be a random sample from an exponential distribution with mean $1/\lambda$.
(a) Determine the UMP test for testing $H_0 : \lambda = 1$ versus $H_a : \lambda > 1$. Find the rejection region when $n = 10$ at level $\alpha = 0.05$. Determine the probability of making Type-II error when $\lambda = 2$.
(b) Plot the power function of the test in (a).

6. Suppose that $X_1, X_2, ..., X_n$ are i.i.d. normal random variables with common mean $\mu = 0$ and common variance σ^2.
(a) Determine the UMP test for testing $H_0 : \sigma^2 = c$ versus $H_a : \sigma^2 > c$, where $c > 0$ is specified.
(b) When $n = 25$, find the test in part (a) for testing $H_0 : \sigma^2 = 4$ versus $H_a : \sigma^2 > 4$ at level $\alpha = 0.05$. Given the following 25 observed values, determine the p-value of the test and state your conclusion.
2.86 0.39 -1.98 -0.47 6.67 -1.21 3.56 -0.19 0.63 1.54 1.18 -0.01 -0.36 3.40 -1.84 2.28 0.81 2.58 1.74 -2.16 3.36 1.32 4.31 -2.14 4.41
(c) Use Monte Carlo simulation to approximate the p-value for the test in (b); use at least 2000 simulation runs. Compare this approximate p-value with the theoretical p-value you obtained in (b).
(d) Obtain the histogram of the simulated values of the test statistic in (c) and then overlay the theoretical pdf curve.

7. You may use the derived UMP test result from the previous problem. Suppose that $X_1, X_2, ..., X_n$ are i.i.d. normal random variables with common mean $\mu = 0$ and common variance σ^2.
(a) Determine the exact power function of the UMP test for testing $H_0 : \sigma^2 = 4$ versus $H_a : \sigma^2 > 4$ when n = 36 and at level $\alpha = 0.05$. Repeat it when the sample size n = 10.
(b) Obtain an overlayed graph of the two power functions; make sure you add legends to the plot.

8. Read Example 8.3.1 on pages 488-491. Use Monte-Carlo simulation to first generate 2000 X-samples $(X_1, ..., X_9)$ (i.e., n = 9) then independently generate 2000 Y-samples $(Y_1, ..., Y_{16})$ (i.e., m = 16); use $\theta_1 = \theta_2 = 0$ and $\theta_3 = 1$. Then use them to form the 2000 (X-sample, Y-sample)'s.
(a) Check numerically that the simulated values of the transformed likelihood ratio $(\Lambda)^{2/(m+n)}$, where Λ is defined in the middle of page 490, are identical to $\frac{n+m-2}{(n+m-2)+T^2}$ with T defined on page 491, expression (8.3.4).
(b) Obtain the histogram of the simulated values of T and then overlay the appropriate theoretical pdf curve.

(c) For testing $H_0 : \theta_1 = \theta_2$ versus $H_0 : \theta_1 \neq \theta_2$, determine the critical region when $n = 9, m = 16$ and $\alpha = 0.05$.
(d) Use the simulated values of T in (b) to obtain the simulated critical region with $\alpha = 0.05$. Is the region close to the theoretical one you obtained in (c)?

9. (One Sample t-test and non-central t-distribution) Read Example 8.3.2 on page 492.
(a) For testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$, use the R code given in the example to determine the power function of the one-sample t-test. Plot the power function when $n = 10$ and $\alpha = 0.05$; use $\sigma^2 = 1$.
(b) Simulate the values of the t-statistic given in the example under the null hypothesis $\mu = 0$. Obtain the histogram and then overlay the appropriate theoretical pdf curve.

10. Problem 8.2.2 on page 486.

11. Problem 8.2.9 on page 486.

12. (Optional Extra-Credit Problem) Submit the work separately. Write a Summary Report to include all the work, results and R code (include necessary remarks to explain your R code lines). If possible use R Markdown and submit both the knitted HTML file and the .RMD file). Read Section 8.3 on Likelihood Ratio Tests on page 487 before you start the project.

The amount of extra credit will depend on the quality of the work. The Report is due by 5:00 pm EST, Friday, 04/23/2021; submit the Report to the Submission portal on coursesite.

Part I.
A simple linear regression model is defined as

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 0, ..., n$$

where $Y_i, i = 1, ..., n$, are the response values;
$x_i, i = 1, ..., n$ are the predictor values (non-random constants) ;
β_0 is the intercept;
β_1 is the slope;
$\epsilon_i, i = 1, ..., n$ are i.i.d. $N(0, \sigma^2)$ error variables.
Suppose that the intercept β_0, the slope β_1, and σ^2 are all unknown. **Derive** the likelihood ratio test for testing $H_0 : \beta_1 = c$ versus $H_a : \beta_1 \neq c$, where the constant c is specified.
(a) Explain in detail how do we obtain the exact critical region when n = 9.
(b) Given the following 9 observed (x_i, Y_i) pairs, determine the p-value of the likelihood-ratio test. Consider two different cases: $c = 0$ and $c = 1.5$.
-4 -7.28
-3 -3.10
-2 -0.15
-1 -1.83
0 -1.80
1 1.82
2 2.56

3	4.17
4	7.98

Part II. Write a computer program in R to simulate both (a) and (b) in Part I.

(A) Use n = 9 and the same x_i values as given in (b) above. Carry out the simulation with at least 2,000 runs and with $\beta_0 = 1, \beta_1 = 2$ and $\sigma = 1$ (so you are able to simulate Y_i values). You are NOT allowed to use any of the linear model or regression packages or functions in R for your work. Obtain the simulated critical value (use $\alpha = 0.05$) and the simulated p-value and then compare with the theoretical values.

(B) Repeat your work in (A) by using linear model or regression functions in R (for example (lm()). Check if your results are consistent with the results from (A).

(C) Consider a different estimator of the slope β_1, $\widetilde{\beta}_1 = \text{median}\ \{\frac{Y_i - Y_j}{x_i - x_j}, x_i \neq x_j, i < j = 1, ..., n\}$. Use n = 9 and the same (x_i values as given in (b) above. Carry out the simulation with at least 2,000 runs and with $\beta_0 = 1, \beta_1 = 2$ and $\sigma = 1$ (so you are able to simulate Y_i values). Obtain the simulated critical value (use $\alpha = 0.05$) and the simulated p-value. You are NOT allowed to use any of the linear model or regression packages or functions in R for your work.

(D) Explore the simulated power functions of the two tests (the likelihood ratio test and the test based on $\widetilde{\beta}_1$. Which one is more powerful? Why?

(E) Simulated the sampling distribution of the p-value for the likelihood ration test. Consider two cases: $\beta_1 = c = 2$ and $\beta_1 = 2$ but $\sigma = 1.5$.

(F) Consider two estimators of β_1: (1) the mle, $\widehat{\beta}$, used in the likelihood ratio test and (2) $\widetilde{\beta}_1$. Obtain simulated MSEs for these two estimators. Which estimator has smaller MSE? Carry out the simulation with at least 2,000 runs and with $\beta_0 = 1, \beta_1 = 2$ and $\sigma = 1$ (so you are able to simulate Y_i values).