Homework Assignment 4

Due Date: Friday October 7, 2022 at 5PM

Total Points: 140

Please email your answer (compiled pdf file from R markdown) and R code to Yen-Yi Ho (hoyen@stat.sc.edu).

1 Effect of asymmetrical distribution in one-sample test

(a) Plot the probability density of a log-normal distribution ($\mu = 0, \sigma = 1$). Note that a positive random variable X is log-normally distributed if the logarithm of X is normally distributed: (5 points)

$$ln(X) \sim N(\mu, \sigma^2).$$

(b) Simulate observations from log-normal ($\mu = 0, \sigma = 1$) distribution with n=3, 5, 10, 30. With $H_0 : \mu = 0$ (population mean=$e^{\mu+\frac{\sigma^2}{2}} = e^{0.5}$ or population median= $e^{\mu} = e^{0}$) versus $H_a : \mu \neq 0$, calculate type I error rate using one-sample t-test, signed test, signed rank test. Filled the type I error rate in the table below. (15 points)

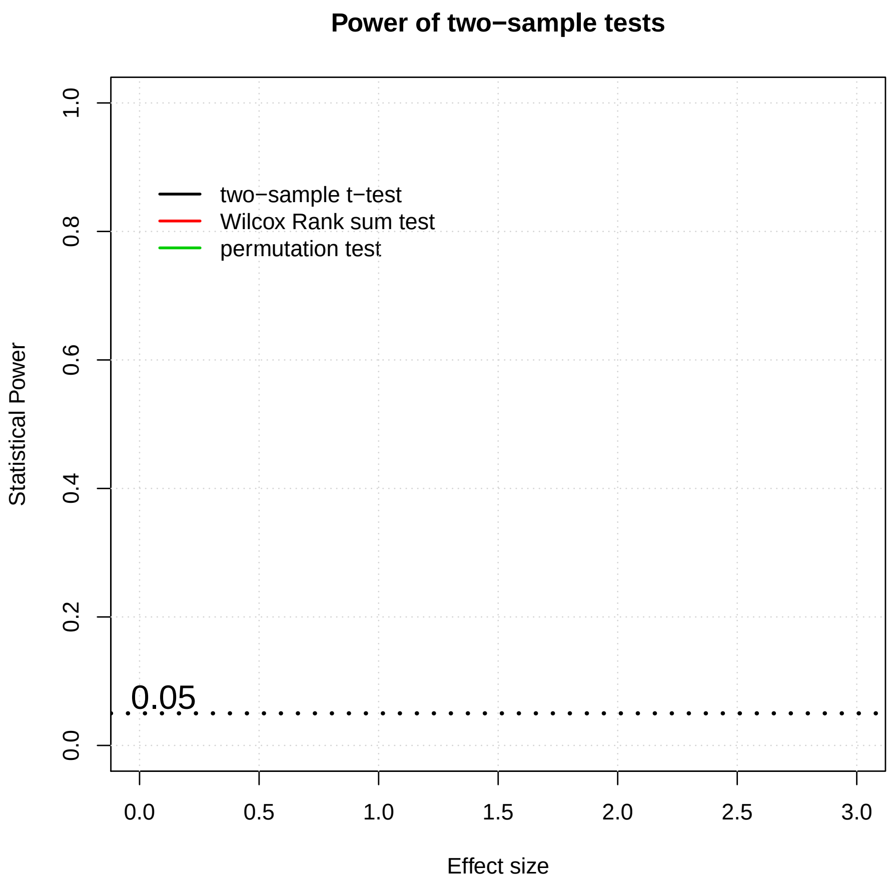
	n=3	n=5	n=10	n=30
One sample t-test				
Signed test				
Signed-rank test				

Recall that type I error rate can be calculated as:

$$\text{Type I error rate} = \frac{\text{\# of times test results are significant}}{\text{\# of simulation iterations}}.$$

2 Large sample efficiency

(a) Simulate data from Normal distribution for two groups (n=30 per group) with $\sigma = 1$ and mean difference ($\mu_1 - \mu_2 = 0, 0.5, 1, 2, 3$). Calculate statistical power using two-sample t-test, Mann-Whitney-Wilcoxon (rank sum test), and permutation test. Add power curves



for each of the three methods in the figure below. (20 points)

(b) based on (a), comment on the statistical power of the three tests. (5 points)

3 Effect of unequal variances in two-sample test

Assume a model $Y_{ij} = \mu_{ij} + \epsilon_{ij}, Var(\epsilon_{ij}) = \sigma_i^2, i = 1,2$ and $j = 1, 2, ...n$. The usual t-statistic used in forming a confidence interval for $\mu_1 - \mu_2$ is

$$T = \frac{\overline{Y_1} - \overline{Y_2} - (\mu_1 - \mu_2)}{S(n_1^{-1} + n_2^{-1})^{1/2}}$$

where

$$S^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n - 2}, \qquad n = n_1 + n_2.$$

If $\sigma_1^2 = \sigma_2^2$ and ϵ_{ij} is normally distributed, the $T \sim t_{n-2} \approx N(0,1)$ for large n. The confidence interval (CI) at level α for $\mu_1 - \mu_2$ in this cane is the $\overline{Y_1} - \overline{Y_2} \pm t_{1-\alpha/2} S(n_1^{-1} + n_2^{-1})^{1/2}$

where $t_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the t-distribution on $n-2$ degrees of freedom, or $\overline{Y_1} - \overline{Y_2} \pm \Phi^{-1}1 - \alpha/2S(n_1^{-1} + n_2^{-1})^{1/2}$ for large n.

However, assume that $\sigma_1^2 \neq \sigma_2^2$. Then heuristically $S^2 \approx \frac{1}{n}(n_1\sigma_1^2 + n_2\sigma_2^2)$ for large n, and T is approximately normally distributed with mean 0 and variance υ.

(a) Show that $\upsilon = Var(T) \approx \frac{\frac{\sigma_1^2}{\sigma_2^2}+\frac{n_1}{n_2}}{\frac{n_1\sigma_1^2}{n_2\sigma_2^2}+1}$ (10 points)

(b) Investigate the error rate of the confidence interval, i.e. calculate $P(\mu_1 - \mu_2 \notin \text{CI}|\mu_1 = \mu_2)$ at $\alpha = 0.05$, using (i) the above approximation to Normal of the test statistics T, (10 points) and (ii) a simulation study. Discuss in detail your simulation set up (5 points).

Present 2 tables of error rates using (i) normal approximation and (ii) simulation where the columns and rows are specified by: (30 points)

$$\sigma_1^2/\sigma_2^2 = \frac{1}{10}, \frac{1}{2}, 1, 5, 10$$
$$n_1/n_2 = \frac{1}{10}, \frac{1}{2}, 1, 5, 10.$$

(c) Based on results from (a) and (b), comment on the effect of unequal variances in relation to sample size using t-test. (10 points)
(d) Repeat (b), present another table of error rates using permutation test. (20 points)
Based on (b) & (d), comment on the effect of unequal variances in relation to sample size using t-test and permutation test (10 points).