Homework Problems

- Note. Always show your work in your homework solutions to receive full points unless it is stated otherwise in the problem.

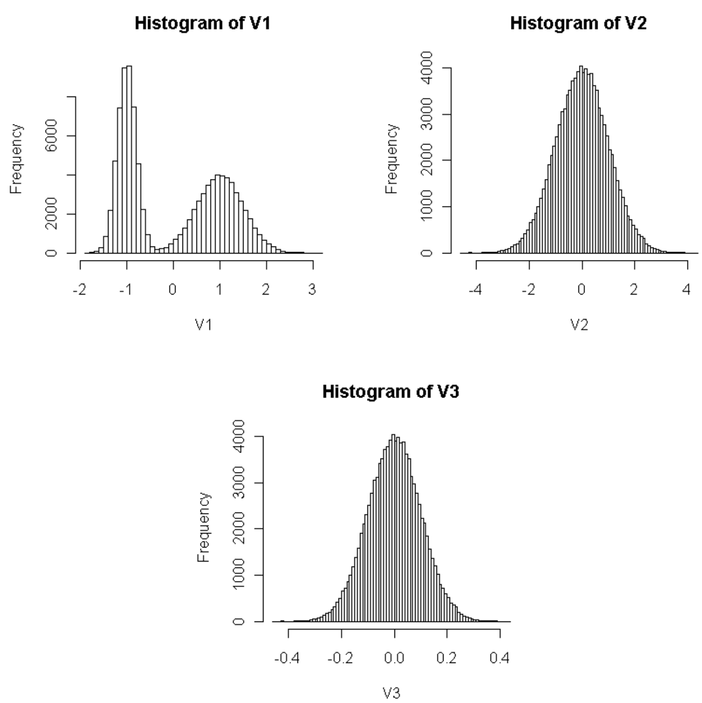
1. (5 pts) Suppose that we have a sample of scores with sample mean 80 and sample standard deviation 1. At least what percentage of the scores are between 76 and 84?

2. (5 pts) Suppose that we have a sample of scores with sample mean 80 and sample standard deviation 1. Find a range that covers at least 70% of the scores using Chebyshev's theorem.

3. (5 pts) Chebyshev's theorem states that for $k > 0$, the proportion of sample observations that are in the range $[\bar{X} - kS, \bar{X} + kS]$ is at least $(1 - 1/k^2)$, where $\bar{X}$ is the sample mean and $S$ is the sample standard deviation. Prove Chebyshev's theorem.

   Hint: construct a lower bound for the sample variance $S^2$ by considering only the terms involving the $X_i$s that are outside the range $\bar{X} \pm kS$.

4. (5 pts) Suppose that we have observations for three continuous variables V1, V2 and V3 respectively, and the histograms for the three samples are given below. Based on the histograms, which of the three samples can be best represented by zero as a typical value? (根據直方圖, 三組樣本中哪一組最適合用0作爲樣本的代表值?) Why?




5. (5 pts) Suppose that we have a sample of scores with sample mean 80 and sample standard deviation 1. At least what percentage of the scores are between 76 and 90?

   Hint: try to find $k$ such that the interval $[80 - k, 80 + k]$ is contained in $[76, 90]$.

6. (3 pts) For a sample of size 1000 with minimum 0, maximum 1 and sample standard deviation 0.3, determine the number of classes for drawing a histogram using Scott's rule. You may use 3.5 as an approximate value for $(24\sqrt{\pi})^{1/3}$ in your calculation.

7. (5 pts) Suppose that a survey of of 8 singers has been conducted and a group of 6000 people were asked to choose their favorate singers among the 8 signers in the survey. The survey response for each participant is recorded as $i$ if the $i$-th singer is the participant's favorate singer, where $i \in \{1, 2, \ldots, 8\}$. Which of the following ways for summarizing the survey data are meaningful?

   (a) Reporting the sample mean.

   (b) Reporting the sample variance.

   (c) Reporting the sample median.

   (d) Reporting the mode.

   (e) Reporting the frequencies of 1's, 2's, ..., and 8's in the sample.

8. (5 pts) What is the level of measurement (nominal, ordinal, interval or ratio) for the data in each of the following cases?

   (a) The number of cups of coffee sold at Starbucks each Sunday during 2008.

   (b) The most challenging courses for the first year students at NCCU (National Chengchi University) during 2006, such as statistics, calculus, probability, etc.

   (c) The monthly average temperatures at Taipei over the past 21 years.

   (d) The rankings of all universities in Taiwan in 2009 according to the QS World University Rankings.

   (e) The waiting times for service for the customers at a post office yesterday.

9. (3 pts) Suppose that during the past 49 years, Flower Airlines experienced no air accident in 36 years. Estimate the probability that Flower Airlines experiences no air accident in one year. Which concept of probability (classical or empirical probability) is used to make the estimate? You may write down your answer without providing justification.

10. (9 pts) Suppose that we have a box of 1000 items, and 5 of them are defective. Suppose that we randomly select two items one at a time without replacement (選取二個物件時, 一次取一個, 取出不放回). Let $A$ be the event that the first item is defective and $B$ be the event that the second item is defective. Find $P(A)$, $P(B)$ and $P(A|B)$.

11. (3 pts) Suppose that a screening test is available for detecting COVID-19. We say that a person is test positive if the person is detected to have COVID-19 based on the screening test, and a person is test negative if the person is not test positive. The sensitivity for the test is defined as the probability that a person is test positive given that the person has COVID-19, and the specificity for the test is defined as the probability that a person is test negative given that the person does not have COVID-19. Suppose that the sensitivity and specificity for the test are 0.7 and 0.9 respectively. Suppose that in a city, 20% of the residents have COVID-19,

and residents of the city are randomly selected to take the screening test. Find the conditional probability that a resident has COVID-19 given that the resident is test positive.

12. (6 pts) Suppose that for the test in Problem 11, a modified testing procedure is proposed. When a person takes the modified test, the person has to take the original test three times. If the person is detected to have COVID-19 in any of the three trials, then the person is detected to have COVID-19 based on the modified test. Find the sensitivity and specificity of the modified test and determine whether the modified test has better sensitivity and specificity than the original test.

13. (6 pts) Determine whether $X$ is a discrete random variable in each of the following cases. You may write down your answers without justification.

    (a) $X$ is the total number of cars sold in Taiwan in the next month.

    (b) $X$ is the average weight of all the babies born in Taiwan next year.

    (c) $X$ is the amount of time you need to wait for service when you visit the post office the next time.

14. (6 pts) Consider the events $A$ and $B$ in Problem 10. Let

$$X_1 = \begin{cases} 1 & \text{if } A \text{ occurs;} \\ 0 & \text{if } A \text{ does not occur,} \end{cases}$$

and let

$$X_2 = \begin{cases} 1 & \text{if } B \text{ occurs;} \\ 0 & \text{if } B \text{ does not occur.} \end{cases}$$

Answer the following questions and justify your answers.

    (a) Are $X_1$ and $X_2$ independent?

    (b) Do $X_1$ and $X_2$ have the same PMF?

15. (12 pts) Suppose that $X$ is a discrete random variable with PMF $p_X$, where

$$p_X(x) = \begin{cases} 0.2 & \text{if } x = 0; \\ 0.5 & \text{if } x = 2; \\ 0.3 & \text{if } x = 4; \\ 0 & \text{otherwise.} \end{cases}$$

Find $P(1.5 < X < 4.5)$, $E(X)$ and $Var(X)$.

16. (6 pts) Suppose that $X$ and $Y$ are two random variables with the same PMF, $P(X = 1) = 1 - P(X = 0)$, and $0 < P(X = 1) < 1$. Show that if $P(Y = 1|X = 0) \neq P(Y = 1|X = 1)$, then $X$ and $Y$ are not independent.

17. (6 pts) Suppose that $X \sim Bin(n, p)$. Verify that $E(X) = np$ and $Var(X) = np(1 - p)$ using the PMF of $X$. You may assume that $n = 2$.

18. (6 pts) For a constant $k$ and a random variable $X$, show that $Var(kX) = k^2 Var(X)$ and $Var(X + k) = Var(X)$.

19. (4 pts) Suppose that $X$ is a random variable with mean 800 and standard deviation 0.01. Find an interval $[a, b]$ such that $P(X \in [a, b]) \geq 0.95$.

20. (6 pts) Suppose that $X_1, \ldots, X_n$ are IID random variables with $E(X_1) = \mu$ and $Var(X_1) = 1$. Find an interval $[a_n, b_n]$ such that $P(\bar{X} \in [a_n, b_n]) \geq 0.95$ and the interval length $(b_n - a_n)$ decreases as $n$ increases. Find the smallest $n$ so that $(b_n - a_n) < 0.01$.

21. (4 pts) Ms. Li is the manager of the human resource department of a company. Based on her experience, she estimates that the probability that a new employee will quit in one year is 0.025. The company hired 40 people last month. What is the probability that exactly 2 of the 40 new employees will quit in one year?

22. (4 pts) Suppose that a class of 15 students are given 4 movie tickets. To be fair, the students would like to choose 4 ticket winners among themselves randomly. Suppose that 10 of the 15 students are males and the rest are females. Find the probability that exactly 3 of 4 ticket winners are males.

23. (4 pts) Suppose $X \sim H(1000, 5, 2)$. Find $E(X)$ using the PMF of $X$.

24. (6 pts) Consider the $X_1$ and $X_2$ in Problem 14.
    (a) Find $P((X_1, X_2) = (x_1, x_2))$ for $(x_1, x_2)$ such that $x_1,\ x_2 \in \{0, 1\}$.
    (b) Find the PMF of $X_1 + X_2$ using the results in Part (a). Is the distribution of $X_1 + X_2$ a hypergeometric distribution? If so, identify the $(N, S, n)$ so that $X_1 + X_2 \sim H(N, S, n)$.

25. (6 pts) Suppose that $X_1, \ldots, X_n$ are IID random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$. Let $\bar{X} = (X_1 + \cdots + X_n)/n$. Show that

$$E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = (n-1)\sigma^2.$$

    - Hint: express $\sum_{i=1}^{n}(X_i - \bar{X})^2$ using $\sum_{i=1}^{n} X_i^2$ and $(\bar{X})^2$, and then compute the expectations of the two terms using the fact that $E(Y^2) = Var(Y) + (E(Y))^2$ and Equations (6) and (7) in the handout "Mean, variance and standard deviation".
    - Remark. This result shows that the expectation of the sample variance of a random sample is the variance of the population distribution.

26. (5 pts) Suppose that $X$ and $Y$ are two random variables and the following table gives the possible values for $(X, Y)$ and the corresponding probabilities.

| $P((X,Y) = (x,y))$ | $y = 0$ | $y = 1$ | $y = 2$ |
|---|---|---|---|
| $x = 0$ | 0.4 | 0.1 | 0 |
| $x = 1$ | 0.1 | 0.3 | 0.1 |

Show that $X$ and $Y$ are not independent by finding a pair $(x, y)$ such that $P((X, Y) = (x, y)) \neq P(X = x)P(Y = y)$.

27. (5 pts) Consider the $X$ and $Y$ in Problem 26. Find $E(XY)$ using the PMF of $XY$.

28. (5 pts) Suppose that Ms. Yu goes fishing every weekend, and she catches 3 fishes in 2 hours on average. Suppose that she plans to spend 2 hours on fishing this weekend. Let $X$ be the number of fishes that she will

catch this weekend. Choose a distribution from binomial distributions, hypergeometric distributions, and Poisson distributions as the distribution of $X$ and give some explanation for your choice. Find the probability that $X \geq 1$ using the distribution that you choose. You may use the following result: the output after running the R command

```
exp(c(-1,-2,-3))
```

is

```
[1] 0.36787944 0.13533528 0.04978707
```

29. (4 pts) Suppose a random variable $X$ has PDF $f$. Determine whether each of the following statements is true. You may write down the answers directly without justification.

    (a) If $f(x) = f(-x)$ for all $x \in (-\infty, \infty)$, then $P(X > 0) = 0.5$.

    (b) If $f(x_1) > f(x_2)$ for some $x_1,\ x_2 \in (-\infty, \infty)$, then $P(X = x_1) > P(X = x_2)$.

    (c) If $f(x) = 0$ for $x < 0$, then $P(X \geq 0) = 1$.

    (d) $\int_{-\infty}^{\infty} f(x)dx = 1$.

30. (6 pts) Suppose that $X$ is a random variable with PDF $f$, where

$$f(x) = \begin{cases} 0 & \text{if } x \leq -1; \\ 2[-x(1+x)]^{0.5}/\pi & \text{if } -1 < x \leq 0; \\ 6[x(1-x)]^{0.5}/\pi & \text{if } 0 < x \leq 1; \\ 0 & \text{if } x > 1. \end{cases}$$

The output after running the following R commands

```
h <- function(x){ 2*((-x)*(1+x))^(0.5)/pi }
integrate(h,-1,0)$value -0.25

h <- function(x){ x*2*((-x)*(1+x))^(0.5)/pi }
integrate(h,-1,0)$value
h <- function(x){ x*6*(x*(1-x))^(0.5)/pi }
integrate(h,0,1)$value

h <- function(x){ (x^2)*2*((-x)*(1+x))^(0.5)/pi }
integrate(h,-1,0)$value
h <- function(x){ (x^2)*6*(x*(1-x))^(0.5)/pi }
integrate(h,0,1)$value
```

is

```
[1] 9.528381e-08
[1] -0.125
[1] 0.375
[1] 0.07812505
[1] 0.2343751
```

and running the R command "`pi`" gives the value of $\pi$. Find $P(-1 < X < 0)$, $E(X)$ and $Var(X)$.

31. (5 pts) Suppose that $X \sim Poisson(\mu)$ for some $\mu > 0$. Show that $E(X(X-1)) = \mu^2$ using the PMF of $X$.

32. (5 pts) Suppoes that $a < b$. Let $F$ be the CDF of $U(a, b)$. That is, $F$ is the CDF of a random variable whose distribution is $U(a, b)$. Show that

$$F(x) = \begin{cases} 0 & \text{if } x \leq a; \\ (x-a)/(b-a) & \text{if } a < x < b; \\ 1 & \text{if } x \geq b. \end{cases}$$

Remark. From the result in Problem 32, we can show that for $X \sim U(a, b)$, the CDF of $(X - a)/(b - a)$ is the CDF of $U(0, 1)$. Since two random variables with the same CDF are identically distributed, we have

$$X \sim U(a, b) \Rightarrow \frac{X-a}{b-a} \sim U(0, 1).$$

33. (5 pts) Suppose that $X \sim N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$ are constants. Show that

$$E(X) = \mu \text{ and } Var(X) = \sigma^2 \tag{1}$$

using the fact that

$$\frac{X-\mu}{\sigma} \sim N(0, 1)$$

and the following results:

(a) The mean and variance for a random variable whose distribution is $N(0, 1)$ are 0 and 1 respectively.

(b) For any constants $a$ and $b$ and a random variable $Y$,

$$E(a + bY) = a + bE(Y) \text{ and } Var(a + bY) = b^2 Var(Y).$$

Note. From now on, to find the probability that a normal random variable is in some interval, you may use the table "Normal probabilities" at

`https://stat.walkup.tw/teaching/statistics/tables/normal.pdf`

34. (15 pts) Suppose that $X$ is a random variable with mean $\mu$ and variance $\sigma^2$, where $\sigma > 0$. Let $Y = (X - \mu)/\sigma$.

(a) (4 pts) Show that the mean and variance of $Y$ are 0 and 1 respectively.

(b) (2 pts) Give two constants $a$ and $b$ such that $P(a \leq Y \leq b) \geq 0.75$ using Chebyshev's Theorem.

(c) (4 pts) Suppose that the distribution of $Y$ is a normal distribution. For the $a$ and $b$ found in Part (b), what is $P(a \leq Y \leq b)$?

(d) (5 pts) Suppose that the distribution of $Y$ is a uniform distribution. For the $a$ and $b$ found in Part (b), what is $P(a \leq Y \leq b)$?

35. (15 pts) Suppose that $X \sim N(0, 1)$. Find the following probabilities.

(a) (3 pts) $P(0.5 < X < 1.51)$.

(b) (3 pts) $P(X > 1.51)$.

(c) (3 pts) $P(X > -0.5)$.

(d) (3 pts) $P(X < 0.5)$.

(e) (3 pts) $P(X < -1.51)$.

36. (5 pts) Suppose that $X \sim N(1, 3^2)$. Find $P(2.5 < X < 5.53)$.

37. (5 pts) Let

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0; \\ e^{-x} & \text{if } x > 0, \end{cases}$$

and let $X$ be a random variable with PDF $f$. Suppose that $\lambda$ is a positive constant. Let $Y = \lambda X$. Find a PDF of $Y$ using Fact 1 in the handout "The probability density function (PDF) of a continuous random variable".

38. (6 pts) Suppose that $X \sim N(1, 3^2)$. Let $Y = 2 - 2X$. Find $P(Y < 0)$.

39. (6 pts) Suppose that a factory produces pens with defective rate 0.02. Let $X$ be the number of defective pens in the next 1000 pens produced by the factory. Approximate $P(X \leq 22)$ using a normal probability with continuity correction.

40. (6 pts) The output after running the R command `pbinom(24, size=50, prob=0.5)` is

```
[1] 0.4438624
```

Compute the normal probability approximation for the above binomial probability with and without continuity correction. For the approximation without continuity correction, use

$$P(Bin(n,p) \leq k) \approx P(N(np, np(1-p)) \leq k),$$

where $k \geq 0$ is an integer. Is it better to use the approximation with continuity correction?

41. (6 pts) Suppose that Route 1819 buses departure from Taoyuan International Airport to Taipei Main Station every 15 minutes. Suppose that a Route 1819 bus just left the bus stop at Taoyuan International Airport, taking away all the passengers who were waiting for the bus. Let $T$ be the waiting time (in minutes) for the next arriving passenger at the bus stop to wait for the bus.

(a) Choose one of the following distributions as the distribution of $T$ (choose the most suitable one). Justify your answer.

i. $Bin(15, 0.5)$.

ii. $N(7.5, 1)$.

iii. $Exp(2)$ (the exponential distribution with mean 1/2).

iv. $U(5, 30)$.

v. $U(0, 15)$.

(b) Base on your choice for the distribution of $T$, compute $P(0 < T < 10)$.

42. (12 pts) Suppose that Ms. Yu goes fishing every weekend, and she catches 4 fishes in 2 hours on average. Let $N(t)$ be the number of fishes she catches in the first $t$ hours for $t \geq 0$. Suppose that $\{N(t) : t \geq 0\}$ is a Poisson process.

(a) (3 pts) Find the expected waiting time for Ms. Yu to catch the first fish the next time she goes fishing. Write down your answer directly.

(b) (3 pts) Find the probability that the next time Ms. Yu goes fishing, she will not catch any fish in the first 30 minutes.

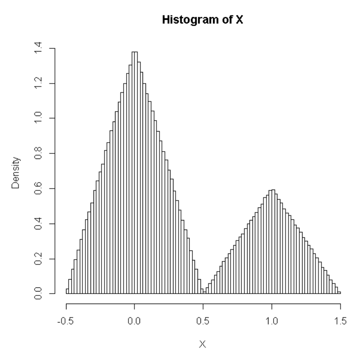
(c) (3 pts) Find the probability that the next time Ms. Yu goes fishing, she will not catch any fish in the first 30 minutes yet will catch at least one fish in the first hour.

(d) (3 pts) Find the probability that the next time Ms. Yu goes fishing, she will catch at least one fish in the last 30 minutes in the first hour.

You may use the following R output for probability evaluation.

```
> exp(c(-1,-2,-3))
[1] 0.36787944 0.13533528 0.04978707
```

43. (5 pts) Suppose that $X = (X_1, \ldots, X_n)$ is a random sample of size $n = 10^6$ from a distribution $\mathcal{D}$ with a PDF and the normalized histogram based on $X$ is given below:




For $\mu \in (-\infty, \infty)$ and $\sigma > 0$, let $\phi_{\mu,\sigma}$ be the function defined by

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}, \qquad x \in (-\infty, \infty).$$

Choose a reasonable PDF for the distribution $\mathcal{D}$ from the following $f$s. You do not need to juistify your answer.

(a) For $x \in (-\infty, \infty)$,

$$f(x) = \begin{cases} 0 & \text{if } x \le 0; \\ e^{-x} & \text{if } x > 0. \end{cases}$$

(b) For $x \in (-\infty, \infty)$,

$$f(x) = 0.5\phi_{0,1}(x) + 0.5\phi_{1,1}(x).$$

(c) For $x \in (-\infty, \infty)$,

$$f(x) = \begin{cases} 0 & \text{if } x \le -0.5; \\ 1.4 - 2.8|x| & \text{if } x \in (-0.5, 0.5]; \\ 0.6 - 1.2|x-1| & \text{if } x \in (0.5, 1.5]; \\ 0 & \text{if } x > 1.5. \end{cases}$$

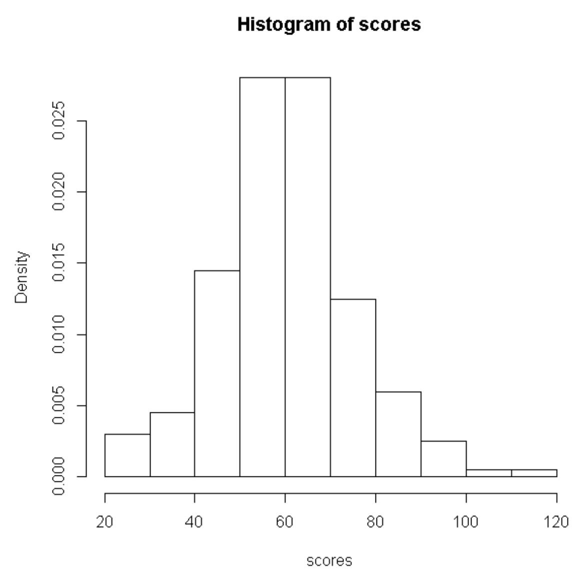
(d) For $x \in (-\infty, \infty)$,

$$f(x) = \begin{cases} 0 & \text{if } x \leq -0.5; \\ 0.6 & \text{if } x \in (-0.5, 0.5]; \\ 0.4 & \text{if } x \in (0.5, 1.5]; \\ 0 & \text{if } x > 1.5. \end{cases}$$

(e) For $x \in (-\infty, \infty)$,

$$f(x) = \begin{cases} 0 & \text{if } x \leq -1; \\ 1 - |x| & \text{if } x \in (-1, 1]; \\ 0 & \text{if } x > 1. \end{cases}$$

44. (5 pts) Suppose that 7000 students took an English examination and we are given a random sample of 200 scores. Below is a normalized histogram of the 200 sample scores. The sample mean and sample standard deviation are 58.1 and 14.9 respectively.

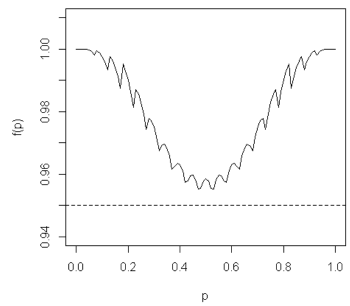



Suppose that we would like to choose a distribution to approximate the score distribution (the common distribution of the 200 IID data). Which of the following distribution is most suitable for approximating the score distribution, the normal distribution $N(60, 15^2)$, the exponential distribution with mean $1/4$, the uniform distribution $U(0, 100)$, the $t(200)$ distribution, or the normal distribution $N(80, 1)$? Justify your answer.

45. (5 pts) Suppose that $(X_1, \ldots, X_n)$ is a random sample from $Bin(1, p)$, where $p \in (0, 1)$ and $n = 20$. Let $\bar{X}$ be the sample mean and let $f$ be the function defined by

$$\begin{aligned} f(p) &= P(-\sqrt{n} < n\bar{X} - np \leq \sqrt{n}) \\ &= P(np - \sqrt{n} < Bin(n, p) \leq np + \sqrt{n}) \end{aligned}$$

for $p \in (0, 1)$. The graph of $f$ is given below:




From the graph of $f$, it is clear that

$$f(p) > 0.95 \text{ for } p \in (0, 1). \tag{2}$$

Construct a 95% C.I. for $p$ based on $\bar{X}$ using (2).

Note. Starting from Problem 46, use the table "Quantiles for $t$ distributions" on the course web page to find the quantiles of $N(0, 1)$ and $t$ distributions.

46. (5 pts) Suppose that we are interested in the rents for one-bedroom apartments near NCCU. Suppose that for a randomly selected rent, the rent distribution is approximately a normal distribution with standard deviation of NT$4,000 per month. Suppose that we have a random sample of 30 rents for one-bedroom apartments near NCCU, and the sample mean is NT$11,000 per month. Find a 95% confidence interval for the mean of the rent distribution.

47. (5 pts) Suppose that we are interested in the rents for one-bedroom apartments near NCCU. Suppose that for a randomly selected rent, the rent distribution is approximately a normal distribution. Suppose that we have a random sample of 30 rents for one-bedroom apartments near NCCU, and the sample mean and sample standard deviation are NT$11,000 per month and NT$4,000 per month respectively. Find a 90% confidence interval for the mean of the rent distribution.

48. (5 pts) Suppose that $(X_1, \ldots, X_n)$ is a random sample from $Poisson(\mu)$, and we are interested in estimating $\mu$ based on $(X_1, \ldots, X_n)$. Let $\bar{X}$ be the sample mean $\sum_{i=1}^{n} X_i/n$. Apply C.L.T. (Central Limit Theorem) to find an approximate distribution for $\sqrt{n}(\bar{X} - \mu)$ when $n$ is large and propose an approximate 95% C.I. for $\mu$ when $n$ is large. Justify your answer.

49. (5 pts) A new drug has been developed to lower systolic blood pressure. Suppose that after using the new drug, a typical patient's systolic blood pressure is lowered by $\mu$ (mmHg) on average. Suppose that after using a conventional drug, a typical patient's systolic blood pressure is lowered by $\mu_0$ (mmHg) on average. We would like to know whether the new drug is more effective than the conventional drug. Suppose that we want to control the probability of falsely claiming the new drug is more effective than the conventional drug. Which statement should be the null hypothesis, $\mu > \mu_0$ or $\mu \leq \mu_0$? You may write down the final answer only.

50. (5 pts) Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\sigma > 0$ is known. Consider the testing problem

$$H_0 : \mu \geq \mu_0 \text{ versus } H_1 : \mu < \mu_0,$$

where $\mu_0$ is given. Consider the test that rejects $H_0 : \mu \geq \mu_0$ when $\sqrt{n}(\bar{X} - \mu_0)/\sigma < C$, where $\bar{X}$ is the sample mean and $C$ is a constant.

$$\max_{\mu \geq \mu_0} P(\sqrt{n}(\bar{X} - \mu_0)/\sigma < C) = P(N(0,1) < C). \tag{3}$$

Note. (3) implies that the size of the test is $P(N(0,1) < C)$.

51. (9 pts) Suppose that the assumptions in Problem 46 hold.

    (a) Can we conclude that the mean of the rent distribution is more than NT$10,000 per month at the 0.01 significance level?

    (b) Can we conclude that the mean of the rent distribution is less than NT$18,000 per month at the 0.05 significance level?

    (c) Can we conclude that the mean of the rent distribution is different from NT$14,000 per month at the 0.01 significance level?

52. (9 pts) Suppose that the assumptions in Problem 47 hold.

    (a) Can we conclude that the mean of the rent distribution is more than NT$10,000 per month at the 0.01 significance level?

    (b) Can we conclude that the mean of the rent distribution is less than NT$18,000 per month at the 0.05 significance level?

    (c) Can we conclude that the mean of the rent distribution is different from NT$14,000 per month at the 0.01 significance level?

53. (4 pts) Suppose that the $p$-value for a test is 0.023. Determine whether we can reject the null hypothesis at each of the following levels: 0.01, 0.05.

54. (6 pts) Suppose that $X \sim Bin(15, p)$ is observed and consider the problem of testing

$$H_0 : p \geq 0.5 \text{ v.s. } H_1 : p < 0.5$$

based on $X$. For $a \in (0, 1)$, let $C_a$ be the largest integer such that $P(Bin(15, 0.5) < C_a) \leq a$. Consider the test that rejects $H_0$ at level $a$ if $X < C_a$. Suppose that we observe $X = 2$ and running the R command

```
pbinom(1:3, 15, 0.5)
```

gives $P(Bin(15, 0.5) \leq x)$ for $x = 1, 2, 3$:

```
[1] 0.0004882812 0.0036926270 0.0175781250
```

What is the smallest $a$ at which the test can reject $H_0$ at level $a$ based on the observed $X$?

- Note. The smallest $a$ at which the test can reject $H_0$ at level $a$ based on the observation is the $p$-value.

55. (6 pts) Suppose that for $a \in (0, 1)$, a test rejects $H_0$ at level $a$ if and only if a test statistic $|Z_n| > z_{a/2}$. Show that the $p$-value for the test is

$$2P(N(0,1) > \text{ observed value for } |Z_n|).$$

56. (6 pts) Consider the rent distribution in Problem 46. Suppose that we have a random sample of 30 rents from the rent distribution, and the sample mean is NT$11,000 per month. For each $a$ in $\{0.02, 0.04, 0.06, 0.08, 0.1\}$, can we conclude that the mean of the rent distribution is more than NT$10,000 per month at level $a$? Is there any evidence that the mean of the rent distribution is more than NT$10,000 per month?

57. (6 pts) Consider the rent distribution in Problem 47. Suppose that we have a random sample of 30 rents from the rent distribution, and the sample mean and sample standard deviation are NT$11,000 per month and NT$4,000 per month respectively. Suppose that we apply the one sample $t$ test for testing $H_0 : \mu \leq 10000$ versus $H_1 : \mu > 10000$, where $\mu$ is the mean of the rent distribution. According the table "Quantiles of $t$ distributions", what can be said about the range of the $p$-value?

58. (6 pts) A mobile phone manufacturer would like to learn about the defective rate for one of their new products. Suppose that a sample of 1000 is taken and 5 out of the 1000 items are defective. Use the approximate test based on C.L.T. (central limit theorem) with known standard deviation.

    (a) Can we conclude that the defective rate is more than 0.01 at the 0.05 significance level?

    (b) Can we conclude that the defective rate is less than 0.01 at the 0.05 significance level?

    (c) Can we conclude that the defective rate is different from 0.006 at the 0.01 significance level?

59. (4 pts) Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\sigma > 0$ is unknown. Consider the testing problem

$$H_0 : \mu \geq \mu_0 \text{ versus } H_1 : \mu < \mu_0,$$

where $\mu_0$ is given. Let $\bar{X}$ and $S$ be the sample mean and the sample standard deviation respectively. Consider the test that rejects $H_0 : \mu \geq \mu_0$ when

$$\frac{\sqrt{n}(\bar{X} - \mu_0)}{S} < C,$$

where $C$ is a constant. Show that the size of the test is $P(t(n-1) < C)$.

60. (6 pts) Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, 2^2)$, where $n = 100$. Consider the testing problem

$$H_0 : \mu \leq 0 \text{ v.s. } H_1 : \mu > 0 \tag{4}$$

For $a \in (0, 1)$, let $T_a$ be the test that rejects $H_0$ if

$$\sqrt{100}\bar{X}/2 > z_a.$$

Note that both $T_{0.05}$ and $T_{0.01}$ are level 0.05 tests for the testing problem in (4). When $\mu = 0.001$, which test has larger Type II error probability, $T_{0.05}$ or $T_{0.01}$?

61. (6 pts) In Problem 47, can we conclude that the standard deviation for the rent distribution is greater than NT$3,200 at level $a$ for each $a \in \{0.01, 0.02, 0.05, 0.1\}$?

62. (6 pts) Suppose that $(X_1, \cdots, X_n)$ is a random sample from $N(\mu, \sigma^2)$. Let $S$ be the sample standard deviation. Consider the testing problem

$$H_0 : \sigma \geq \sigma_0 \text{ v.s. } H_1 : \sigma < \sigma_0$$

and the test that rejects $H_0$ at level $a$ if and only if

$$\frac{(n-1)S^2}{\sigma_0^2} < k_{1-a,n-1},$$

where $k_{1-a,n-1}$ is the $a$ quantile of $\chi^2(n-1)$ such that $P(\chi^2(n-1) > k_{1-a,n-1}) = 1 - a$.

(a) Show that the test is of size $a$.

(b) Show that the $p$-value of the test is

$$P\left(\chi^2(n-1) < \text{ observed } \frac{(n-1)S^2}{\sigma_0^2}\right).$$

63. (6 pts) Suppose that $c$ is a positive constant and $\mathcal{D}$ is a distribution with PDF $f$, where

$$f(x) = \begin{cases} 0 & \text{if } x \leq -1; \\ c + cx & \text{if } -1 < x \leq 0; \\ c - cx/19 & \text{if } 0 < x \leq 19; \\ 0 & \text{if } x > 19. \end{cases}$$

(a) Show that $c = 0.1$ using the fact that $\int_{-\infty}^{\infty} f(x)dx = 1$. Note that you may use the following results:

- From the linearity of integration, we have that for a constant $k$ and a function $f_1$ defined on an interval $(a, b)$,

$$\int_a^b k f_1(x)dx = k \int_a^b f_1(x)dx.$$

- The output after running the following R commands

```
g <- function(x){ 1+x }
h <- function(x){ 1-x/19}
integrate(g,-1,0)$value
integrate(h,0,19)$value
```

is

```
[1] 0.5
[1] 9.5
```

(b) Find the median of $\mathcal{D}$ when $c = 0.1$.

(c) Running the R commands

```
f <- function(x){ x*0.1*(1+x) }
g <- function(x){ x*0.1*(1-x/19) }
integrate(f,-1,0)$value + 1/60
integrate(g,0,19)$value - 6 - 1/60
```

gives the output

```
[1] -3.469447e-18
[1] 8.291978e-16
```

Determine whether $\mathcal{D}$ is positively skewed when $c = 0.1$.

64. (3 pts) Suppose that $X$ and $Y$ are independent random variables, $X \sim \chi^2(n_1)$, and $Y \sim \chi^2(n_2)$, where $n_1$ and $n_2$ are positive integers. What is the distribution of $X + Y$? You may write down your answer without justification.

65. (5 pts) Suppose that in the latest 100 times when a pitcher A faced a batter B, A striked out B 60 times. Let $p$ be the probability that A strikes out B. Can we conclude that $p > 0.5$ at level 0.01 based on the $Z$ test for population proportion?

66. (5 pts) Suppose that $a$ is a number in the interval $(t_{0.02,15}, t_{0.01,15})$, is it true that $P(t(15) > a) < 0.05$? Justify your answer.