Homework Problems

- Note. Always show your work in your homework solutions to receive full points.

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) 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?

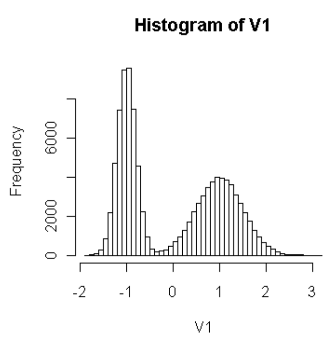
   (a) Reporting the sample mean.

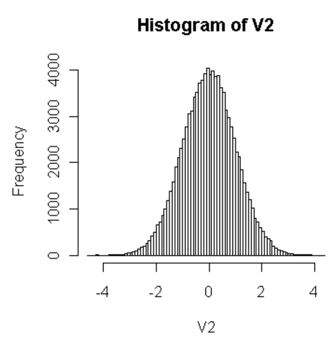
   (b) Reporting the sample variance.

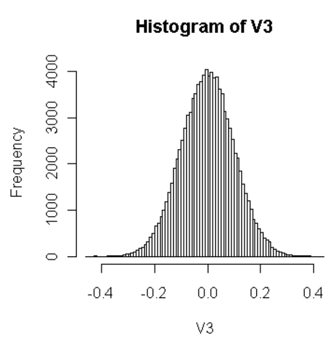
   (c) Reporting the sample median.

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

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?

6. (5 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.

7. (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 weights (in kilograms) of new born babies in Taiwan in 2023.

8. (5 pts) Chebyshev's theorem (sample version) states that for $k > 0$, the proportion of sample observations that are in the interval $[\bar{X} - kS, \bar{X} + kS]$ 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: for $a > 0$, let m_a be the number of X_is that are outside the interval $[\bar{X} - a, \bar{X} + a]$. Construct a lower bound for the sample variance S^2 based on m_a and a. Choose a suitable a to derive Chebyshev's theorem.

9. (5 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 (隨機選取二個物件, 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 $P(A|B)$.

10. (5 pts) Suppose that during the past 49 years, Flower Airlines experienced no air accident in 36 years. The probability that Flower Airlines experiences no air accident in one year is estimated to be 36/49. Which concept of probability (classical or empirical probability) is used to make the estimate?

11. (5 pts) For a disease screening test, the sensitivity of the test is the conditional probability that the test classifies a person as having the disease given that the person has the disease, and the specificity for the test is the conditional probability that the test classifies a person as not having the disease given that the person does not have the disease.

    Suppose that a screening test is available for detecting COVID-19, and the sensitivity and specificity for the test are 0.7 and 0.9 respectively. Suppose that in a small town, 20% of the residents have COVID-19. Suppose that a resident of the town is randomly selected to take the screening test. Find the conditional probability that the resident has COVID-19 given that the resident is detected to have COVID-19 based on the screening test ($P(\text{true positive}|\text{test positive})$).

12. (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 20 hours.

    (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.

13. (6 pts) Suppose that X is a discrete random variable with PMF p_X, where

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

    (a) Find $P(1.5 < X < 4.5)$.

    (b) Find $E(X)$.

14. (3 pts) Consider the events A and B in Problem 9. Determine whether A and B are independent and justify your answer.

15. (6 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. For $i = 1, 2$, let

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th selected item is defective;} \\ 0 & \text{if the } i\text{-th selected item is not defective.} \end{cases}$$

    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 (probability mass function)?

16. (3 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 $P((X,Y) = (x,y)) \neq P(X = x)P(Y = y)$.

17. (6 pts) Suppose that A and B are two events and $P(B) > 0$ and $P(B^c) > 0$. Show that if $P(A|B) = P(A|B^c)$, then $P(A) = P(A|B)$.

    Note. The result in Problem 17 implies that if $P(A|B) = P(A|B^c)$, then A and B are independent. Indeed, it can be shown that if $P(A|B) \neq P(A|B^c)$, then A and B are not independent.

18. (12 pts) In the handout "Mean, variance and standard deviation", some properties of expectation and variance are listed, which are summarized in the following fact.

    Fact 1 Suppose that X and Y are random variables with finite expectations. Then the following properties hold.

(P1) $E(X+Y) = E(X) + E(Y)$.

(P2) For a constant k, $E(kX) = kE(X)$.

(P3) For a constant k, $E(k) = k$.

(P4) If X and Y are independent, then $E(XY) = E(X)E(Y)$.

Suppose that X and Y are random variables and k is a constant. Apply Fact 1 to derive the following results in (a)–(c) assuming that $E(X)$, $E(X^2)$, $E(Y)$ and $E(Y^2)$ are finite.

(a) $Var(X) = E(X^2) - [E(X)]^2$.

(b) $Var(kX) = k^2 Var(X)$.

(c) If X and Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$.

Note. From now on, we will assume that all expectations/variances involved in the homework problems are finite unless otherwise stated.

19. (6 pts) Show that $Var(X+a) = Var(X)$ for every constant a. You may use Fact 1 in Problem 18.

20. (6 pts) Suppose that $(X_1, \ldots, X_n)$ is a random sample. Let $\mu = E(X_1)$ and $\sigma = \sqrt{Var(X_1)}$. Let $\bar{X} = \sum_{i=1}^{n} X_i/n$ be the sample mean. Show that

$$E(\bar{X}) = \mu$$

and

$$Var(\bar{X}) = \frac{\sigma^2}{n}.$$

You may apply Fact 1 and the properties in (a)(b)(c) in Problem 18 to solve this problem.

21. (18 pts) Consider the X and Y in Problem 16.

(a) (3 pts) Find the PMF of Y.

(b) (3 pts) Find $E(Y)$.

(c) (3 pts) Find $Var(Y)$.

(d) (3 pts) Find the PMF of XY.

(e) (3 pts) Find $E(XY)$ using the PMF of XY.

(f) (3 pts) Find $E(XY)$ using the result that for two discrete random variables X and Y and a function g such that $g(X,Y)$ is defined, we have

$$E(g(X,Y)) = \sum_{(x,y)\in S} g(x,y) P((X,Y) = (x,y)),$$

where $S = \{(x,y) : P((X,Y) = (x,y)) > 0\}$.

22. (6 pts) Suppose that X is a random variable with mean μ and standard deviation σ. Chebyshev's theorem (distribution version) states that for $k > 0$,

$$P(X \in [\mu - k\sigma, \mu + k\sigma]) \geq 1 - \frac{1}{k^2}.$$

Prove Chebyshev's theorem (distribution version).

Hint: define a random variable Z by

$$Z = \begin{cases} (k\sigma)^2 & \text{if } |X - \mu| > k\sigma; \\ 0 & \text{if } |X - \mu| \leq k\sigma, \end{cases}$$

then apply the fact that

$$X \le Y \Rightarrow E(X) \le E(Y)$$

to deduce that $E(Z) \le E(X-\mu)^2$.

23. (6 pts) Suppose that X is a random variable with mean 800 and standard deviation 0.01. Find an interval $[a,b]$ such that $P(X \in [a,b]) \ge 0.95$.

24. (6 pts) Suppose that $X_1, \ldots, X_n$ are IID random variables with $E(X_1) = \mu$ and $Var(X_1) = 1$.

    (a) (3 pts) Apply Chebyshev's theorem to find a lower bound for $P(|\bar{X} - \mu| \le n^{-1/4})$.

    (b) (3 pts) Find a constant N (as small as possible) such that the lower bound found in Part (a) is at least 0.99 for $n \ge N$.

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 distribution of each observation in the sample.

26. (5 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?

27. (5 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.

28. (5 pts) Suppose $X \sim H(1000, 5, 2)$. Find $E(X)$ using the PMF of X.

29. (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 $X \ge 1$ using the proposed distribution. You may use the following R output for probability evaluation.

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

30. (7 pts) Suppose that $X \sim Poisson(\mu)$ for some $\mu > 0$.

    (a) (5 pts) Show that $E(X(X-1)) = \mu^2$.

    (b) (2 pts) Show that $Var(X) = \mu$ using the result in Part (a) and the result that $E(X) = \mu$.

31. (6 pts) Consider the X_1 and X_2 in Problem 15.

    (a) (3 pts) Find $P((X_1, X_2) = (x_1, x_2))$ for all x_1, $x_2 \in \{0, 1\}$.

    (b) (3 pts) 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 $X_1 + X_2 \sim H(N, S, n)$.

32. (9 pts) Suppose that X has CDF F_X, where

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0; \\ 0.5x + 0.5 & \text{if } 0 \le x < 1; \\ 1 & \text{if } x \ge 1. \end{cases}$$

    (a) (2 pts) Find $P(X = 0.3)$.

    (b) (2 pts) Find $P(0 < X \le 0.3)$.

    (c) (2 pts) Find $P(0 < X < 0.3)$.

    (d) (3 pts) Find $P(0 \le X \le 0.3)$.

    Note that X is not a discrete random variable since

$$\sum_{x : P(X=x) > 0} P(X = x) < 1.$$

33. (5 pts) Suppose that a random variable X has PDF f, where

$$f(x) = \begin{cases} \frac{8\sqrt{x(1-x)}}{\pi} & \text{if } 0 \le x \le 1; \\ 0 & \text{otherwise.} \end{cases}$$

    Running the following R commands

```
 f <- function(x){ 8*sqrt(x*(1-x))/pi }
 h <- function(x){ x*f(x) }
integrate(h, 0, 1)$value
```

    gives an output that is approximately 0.5, and running the following R commands

```
 f <- function(x){ 8*sqrt(x*(1-x))/pi }
 h <- function(x){ (x^2)*f(x) }
integrate(h, 0, 1)$value
```

    gives an output that is approximately 0.3125.

    Find $Var(X)$ based on the given R outputs. Note that the R command "`pi`" gives the value of π and the R command "`sqrt`(x)" gives $\sqrt{x}$.

34. (5 pts) Show that if $Y \sim U(a, b)$, then $E(Y) = (a + b)/2$ and $Var(Y) = (b - a)^2/12$. You may use the following results:

    - If $X \sim U(0, 1)$, then $E(X) = 1/2$ and $Var(X) = 1/12$.

- If $Y \sim U(a,b)$, then $(Y-a)/(b-a) \sim U(0,1)$.

35. (5 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) The graph of the CDF of X has no jump at any point.

    (d) If $f(x) = 0$ for $x \le 0$, then $P(X \ge 0) = 1$.

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

    Note. The table "Normal probabilities" at

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

    can be used for finding normal probabilities for problems hereafter.

36. (25 pts) Suppose that $X \sim N(0,1)$. Find the following probabilities.

    (a) (2 pts) $P(0 < X < 1.51)$.

    (b) (2 pts) $P(0 < X < 0.5)$.

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

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

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

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

    (g) (3 pts) $P(-0.5 < X < 1.51)$.

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

    (i) (3 pts) $P(-1.51 < X < -0.5)$.

37. (14 pts) Suppose that X is a random variable with mean μ and variance σ^2, where $\sigma > 0$. Let $Y = (X - \mu)/\sigma$.

    (a) (4 pts) What are the mean and variance of Y?

    (b) (3 pts) Give two constants a and b such that $P(a \le Y \le b) \ge 0.75$ using Chebyshev's Theorem.

    (c) (3 pts) Suppose that the distribution of Y is a normal distribution. For the a and b found in Part (b), what is $P(a \le Y \le b)$?

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

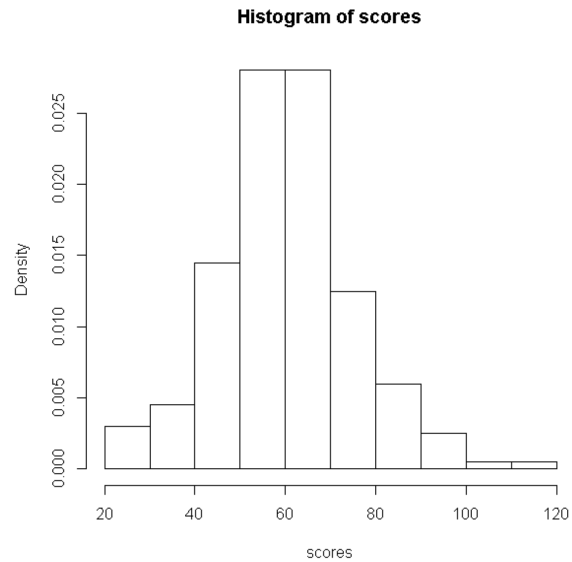
38. (6 pts) Suppose that $X \sim N(1, 3^2)$.

    (a) (3 pts) Find $P(2.5 < X < 5.53)$.

    (b) (3 pts) 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 $P(X \le 22)$ using a normal probability with continuity correction.

40. (6 pts) Suppose that 7000 students took an English examination and we are given a sample of 200 scores which were randomly selected from the 7000 scores one at a time with replacement. 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 scores). 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, or the uniform distribution $U(0, 100)$? Justify your answer.

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 bus waiting time (in minutes) for the next arriving passenger at the bus stop.

    (a) (3 pts) 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. $N(10, 20^2)$.
        iii. $Exp(1/20)$ (the exponential distribution with mean 20).
        iv. $U(5, 30)$.
        v. $U(0, 15)$.

    (b) (3 pts) Base on your choice for the distribution of T, compute $P(0 < T < 10)$.

Note. We use $Exp(\lambda)$ to denote the exponential distribution with mean $1/\lambda$ for $\lambda > 0$ for problems hereafter.

42. (6 pts) Suppose that X is a random variable and the distribution of X is $Exp(\lambda)$,where $\lambda > 0$. What is the distribution of $2X$? Justify your answer. You may use the following fact:

Fact 2 For $a > 0$,

$$X \sim Exp(a) \Leftrightarrow aX \sim Exp(1).$$

43. (20 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) (5 pts) Find the expected waiting time for Ms. Yu to catch the first fish the next time she goes fishing.

   (b) (5 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.

   (c) (5 pts) Find the probability that the next time Ms. Yu goes fishing, she will not catch any fish in the first 30 minutes.

   (d) (5 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.

   You may use the following R output for probability evaluation.

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

   Note. Starting from Problem 44, use the table "Quantiles for t distributions" on the course web page to find the quantiles of $N(0, 1)$ and t distributions.

44. (10 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 a normal distribution with standard deviation of NT$4,000 per month.

   (a) (5 pts) Suppose that we would like to obtain a random sample of size n of rents for one-bedroom apartments near NCCU, and then construct a 95% confidence interval for the mean of the rent distribution based on the sample. Suppose that the maximum allowable margin of error of the 95% confidence interval is 1500 (NTD per month). How large the sample size n needs to be?

   (b) (5 pts) 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 based on the sample.

45. (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 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.

46. (10 pts) Suppose that a US state governor candidate would like to conduct a survey to construct an approximate 90% confidence interval for p: the proportion of the candidate's supporters in the state. For the survey to be conducted, n eligible voters in the state will be randomly selected to indicate whether they support the candidate or not.

    (a) (5 pts) Suppose that the maximum allowable margin of error of the 90% confidence interval is 10%. Find the minimum n required.

    (b) (5 pts) Suppose that the survey has been conducted with $n = 1000$. According to the survey result, among the 1000 selected voters, 100 voters indicated that they would support the candidate. Find an approximate 90% confidence interval for p based on the survey result.

47. (10 pts) Suppose that $(X_1, \ldots, X_n)$ is a random sample from $Poisson(\mu)$, and we are interested in estimating μ based on $(X_1, \ldots, X_n)$. Let $\bar{X}$ be the sample mean $\sum_{i=1}^{n} X_i/n$.

    (a) (5 pts) Apply C.L.T. (Central Limit Theorem) to find an approximate distribution for $\sqrt{n}(\bar{X} - \mu)$ when n is large. You may use the results given in Problem 30 to solve this problem.

    (b) (5 pts) Propose an approximate 95% C.I. for μ when n is large. Justify your answer.

48. (5 pts) For $\lambda > 0$, let F_λ be the CDF of $Exp(\lambda)$, then

$$F_\lambda(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0, \end{cases}$$

    Prove Fact 2 in Problem 42 using the CDF of $Exp(\lambda)$ and the CDF of $Exp(1)$.

49. (5 pts) Let

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

    Then f is a PDF of $Exp(1)$. Running the R commands

```
f <- function(x){ exp(-x) }
g <- function(x){ x*f(x) }
integrate(g,0, Inf)$value
```

    gives the output

```
[1] 1
```

    and running the R commands

```
f <- function(x){ exp(-x) }
g <- function(x){ (x^2)*f(x) }
integrate(g,0, Inf)$value
```

    gives the output

```
[1] 2
```

Find the mean and variance for $Exp(1)$ based on the above R outputs and then apply Fact 2 in Problem 42 to deduce that

$$X \sim Exp(\lambda) \Rightarrow \begin{cases} E(X) = 1/\lambda; \\ Var(X) = 1/(\lambda^2). \end{cases}$$

50. (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 μ (mmHg) on average. Suppose that after using a conventional drug, a typical patient's systolic blood pressure is lowered by μ_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.

51. (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 μ_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. Show that the size of the test is $P(N(0, 1) < C)$.

52. (5 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 μ_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 $P(t(n - 1) < C)$.

53. (9 pts) Consider the rent distribution in Problem 44. Suppose that we have a random sample of 30 rents from the rent distibution, and the sample mean is NT$11,000 per month.

    (a) (3 pts) 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) (3 pts) 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) (3 pts) Can we conclude that the mean of the rent distribution is different from NT$14,000 per month at the 0.01 significance level?

54. (9 pts) Consider the rent distribution in Problem 45. 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.

    (a) (3 pts) 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) (3 pts) 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) (3 pts) Can we conclude that the mean of the rent distribution is different from NT$14,000 per month at the 0.01 significance level?

55. (3 pts) Suppose that the p-value for a test is 0.023. Determine whether we can reject the null hypothesis based on the test at each of the following levels: 0.01, 0.025, 0.05.

56. (6 pts) Suppose that a test rejects H_0 at level a if a test statistic $Z_n \leq -z_a$. Show that the p-value for the test is $P(N(0,1) <$ observed value for $Z_n)$.

57. (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

$$H_0 : \mu \leq 0 \text{ v.s. } H_1 : \mu > 0 \tag{1}$$

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 (1). When $\mu = 0.001$, which test has larger Type II error probability, $T_{0.05}$ or $T_{0.01}$? Justify your answer.

58. (7 pts) Consider the rent distribution in Problem 44. Suppose that we have a random sample of 30 rents from the rent distribution, and the sample mean is NT$11,000 per month.

(a) (4 pts) 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?

(b) (3 pts) Is there any evidence that the mean of the rent distribution is more than NT$10,000 per month?

59. (12 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. You may use the approximate Z test based on the approximation of binomial probabilities using normal probabilities.

(a) (4 pts) Can we conclude that the defective rate is more than 0.01 at the 0.05 significance level?

(b) (4 pts) Can we conclude that the defective rate is less than 0.01 at the 0.05 significance level?

(c) (4 pts) Can we conclude that the defective rate is different from 0.006 at the 0.01 significance level?

60. (5 pts) Suppose that C, n, p_1 and p_2 are constants, where n is a positive integer and $0 \leq p_1 < p_2 \leq 1$. Show that

$$P(Bin(n, p_1) \leq C) \geq P(Bin(n, p_2) \leq C).$$

Hint: let X_1, …, X_n, Y_1, …, Y_n be independent random variables such that $X_i \sim Bin(1, p_1)$ and

$$Y_i \sim Bin\left(1, \frac{p_2 - p_1}{1 - p_1}\right)$$

for $i = 1, \ldots, n$. For $i = 1, \ldots, n$, define

$$Z_i = \begin{cases} 1 \text{ if } X_i = 1; \\ Y_i \text{ if } X_i = 0, \end{cases}$$

then $Z_i \geq X_i$, so $n\bar{Z} \geq n\bar{X}$. Show that $Z_i \sim Bin(1, p_2)$, then

$$P(Bin(n, p_1) \leq C) \geq P(Bin(n, p_2) \leq C)$$

follows from the fact that

$$\{n\bar{Z} \leq C\} \subset \{n\bar{X} \leq C\}$$

and the fact that

$$n\bar{Z} \sim Bin(n, p_2) \text{ and } n\bar{X} \sim Bin(n, p_2).$$

61. (6 pts) Consider the rent distribution in Problem 45. Suppose that we have a random sample of n 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 μ is the mean of the rent distribution.

   (a) (4 pts) Suppose that $n = 30$. According the table "Quantiles of t distributions", what can be said about the range of the p-value of the t test?

   (b) (2 pts) Suppose that $n = 36$. The output after running the R command

   ```
   1-pt(1.5, df=c(35,36))
   ```

   is

   ```
   [1] 0.07129092 0.07116710
   ```

   Find the p-value of the t test. Is there a strong evidence for $\mu > 10000$ based on the p-value?

62. (4 pts) For this problem, you may write down your answers directly without justification. The table "Quantiles for χ^2 distributions" is available on the course web site.

   (a) (2 pts) For a number x in the interval $(k_{0.02,15}, k_{0.01,15})$, what can be said about the range of $P(\chi^2(15) > x)$?

   (b) (2 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 $X + Y$?

Note. For problems hereafter, you may use the table "Quantiles for χ^2 distributions" on the course web site to solve the problems.

63. (10 pts) Let S be the sample standard deviation of a random sample of size 11 from $N(\mu, \sigma^2)$. Note that running the R command

```
1/qchisq(c(0.02, 0.97, 0.98, 0.99),df=11)
```

gives

```
[1] 0.27710910 0.04685688 0.04421269 0.04044494
```

and running the R command

```
1/qchisq(c(0.02, 0.97, 0.98, 0.99),df=10)
```

gives

```
[1] 0.32689872 0.05019599 0.04725727 0.04308627
```

(a) (4 pts) Find a 96% C.I. for σ^2.

(b) (2 pts) Suppose that we observe $S = 10$. Compute the observed 96% C.I. for σ^2 using the C.I. from Part (a).

(c) (4 pts) Propose a level 0.04 test for testing

$$H_0 : \sigma = 12 \text{ v.s. } H_1 : \sigma \neq 12$$

based on S. Suppose that we observed $S = 10$. Can we conclude that $\sigma \neq 12$ at level 0.04 based on the proposed test?

64. (8 pts) Suppose that we are given the sample in Problem 45 and the assumptions in Problem 45 hold.

(a) (4 pts) Can we conclude that the standard deviation for the rent distribution is greater than NT$3,200 at level 0.1?

(b) (4 pts) For each $a \in \{0.01, 0.02, 0.05\}$, can we conclude that the standard deviation for the rent distribution is greater than NT$3,200 at level a?

65. (8 pts) Suppose that $(X_1, \cdots, X_n)$ is a random sample from $N(\mu, \sigma^2)$. Let S be the sample standard deviation.

(a) (4 pts) Propose a test of size a based on S for the testing problem

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

You need to verify that the proposed test is of size a.

(b) (4 pts) Suppose that the testing problem is changed to

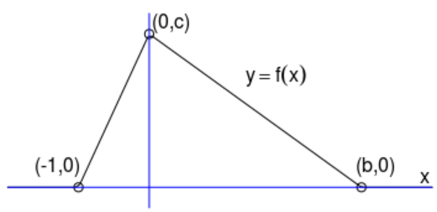
$$H_0 : \sigma \geq \sigma_0 \text{ v.s. } H_1 : \sigma < \sigma_0. \qquad (2)$$

Is the proposed test in Part (a) still of size a for the new testing problem in (2)?

66. (8 pts) Suppose that $b > 1$ and $c > 0$ are constants and $\mathcal{D}$ is a distribution with PDF f, where

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

The graph of f is shown below.



(a) (2 pts) Is it possible that the median of $\mathcal{D}$ is 0? Justify your answer.

(b) (2 pts) Find b when $c = 0.1$.

(c) (2 pts) Find the median of $\mathcal{D}$ when $c = 0.1$.

(d) (2 pts) It can be shown that

$$\int_{-\infty}^{\infty} x f(x) dx = \frac{c(b^2 - 1)}{6}. \tag{3}$$

Use the result in (3) to determine whether $\mathcal{D}$ is positively skewed when $c = 0.1$. The definition of a positively skewed distrbution is given in Definition 1 in the handout "Chi-squared distributions and testing for population variance".