Name: ID:

Homework for 3/10

1. [§9-36] The National Center for Health Statistics (1970) gives the following data on distribution of suicides in the United States by month in 1970. Is there any evidence that the suicide rate varies seasonally, or are the data consistent with the hypothesis that the rate is constant? (Hint: Under the latter hypothesis, model the number of suicides in each month as a multinomial random variable with the appropriate probabilities and conduct a goodness-of-fit test. Look at the signs of the deviations, $O_i - E_i$, and see if there is a pattern.)

Month	# of suicides
Jan.	1867
Feb.	1789
Mar.	1944
Apr.	2094
May	2097
June	1981
July	1887
Aug.	2024
Sept.	1928
Oct.	2032
Nov.	1978
Dec.	1859

Let p_i denote the probability that a suicide occurs in the ith month (January is the 1st month). Thus the null hypothesis is that the rate is constant, that is, $H_0 : p_1 = p_2 = \cdots = p_{12}$. Since we must have $p_1 + \cdots p_{12} = 1$, the null hypothesis is actually

$$H_0 : p_1 = p_2 = \cdots = p_{12} = \frac{1}{12}.$$

And the alternative hypothesis is

$$H_0 : p_1 + p_2 + \cdots + p_{12} = 1.$$

The sample size, that is the total number of observations, is

$$1867 + 1789 + \cdots + 1859 = 23480.$$

Under the null hypothesis, the expected counts for each month will be

$$23480 \cdot \frac{1}{12} = 1956.7.$$

Thus the Pearson's χ^2 test statistic is

$$X^2 = \frac{(1867 - 1956.7)^2}{1956.7} + \frac{(1789 - 1956.7)^2}{1956.7} + \cdots + \frac{(1859 - 1956.7)^2}{1956.7}$$
$$= 51.79.$$

Under H_0, there is no free variables. Thus $\dim \omega_0 = 0$. Under H_A, there are 11 free variables. Thus $\dim \Omega = 11$. Therefore, the asymptotic distribution of the Pearson χ^2 statistic is χ^2 with degrees of freedom $\dim \Omega - \dim \omega_0$, that is,

$$X^2 \sim \chi^2_{11}.$$

Consequently, the p-value of $X^2 = 51.79$ is approximately 2.98×10^{-7}. Since the p-value is very small, we will reject H_0, that is, there is strong evidence that the suicide rate is not constant throughout the 12 months.

We can also use the χ^2 table to find a rejection region. Since the rejection region is of the form $\{X^2 > c\}$, and $\chi^2_{0.995,11} = 26.76$, we will reject H_0 at significance level $\alpha = 0.005$.

2. [§9-42] Nylon bars were tested for brittleness (Bennett and Franklin 1954). Each of 280 bars was molded under similar conditions and was tested in five places. Assuming that each bar has uniform composition, the number of breaks on a given bar should be binomially distributed with five trials and an unknown probability p of failure. If the bars are all of the same uniform strength, p should be the same for all of them; if they are of different strengths, p should vary from bar to bar. Thus, the null hypothesis is that the p's are all equal. The following table summarizes the outcome of the experiment:

Breaks/Bar	Frequency
0	157
1	69
2	35
3	17
4	1
5	1

a. Under the given assumption, the data in the table consist of 280 observations of independent binomial random variables. Find the mle of p.

b. Pooling the last three cells, test the agreement of the observed frequency distribution with the binomial distribution using Pearson's chi-square test.

c. ~~Apply the test procedure derived in the previous problem.~~

a. For the kth bar, we apply tests at 5 different places. For each test, the probability that the bar breaks at that place is p_k. Therefore, the number of breaks, X_k, follows a binomial distribution with parameter $n = 5$ and p_k, that is $X_k \sim Bin(5, p_k)$. In particular, we have

$$q_{k,i} = \binom{5}{i} p_k^i (1 - p_k)^{5-i}, \qquad 0 \leq i \leq 5,$$

where $q_{k,i} = \mathbb{P}(X_k = i)$ is the probability that the kth bar breaks at i places.

The null hypothesis is that the p_k's are all equal for $1 \leq k \leq 280$. Consequently, we have for each $0 \leq i \leq 5$, $q_{k,i}$'s are all equal for $1 \leq k \leq 280$. Thus we can use θ_i to denote the common value, that is, θ_i is the probability that a randomly selected bar will break at i places. Then our null hypothesis is

$$H_0 : \theta_i = \theta_i(p) = \binom{5}{i} p^i (1 - p)^{5-i},\ 0 \leq i \leq 5.$$

The alternative hypothesis is simply

$$H_A : \theta_0 + \theta_1 + \cdots + \theta_5 = 1.$$

Let N_i denote the number of bars that have i breaks in the sample of size $n = 280$. Thus for this particular sample,

$$N_0 = 157,\ N_1 = 69,\ N_2 = 35,\ N_3 = 17,\ N_4 = 1,\ N_5 = 1.$$

Under H_0, the likelihood function of p (joint pmf of $N_0, \ldots, N_5$) is

$$\text{lik}(p) = \frac{280!}{N_0!N_1!N_2!N_3!N_4!N_5!}\theta_0^{N_0}\theta_1^{N_1}\theta_2^{N_2}\theta_3^{N_3}\theta_4^{N_4}\theta_5^{N_5}.$$

The log-likelihood function is

$$\begin{aligned} l(p) = \log \text{lik}(p) &= \log 280! - \sum_{i=0}^{5} \log N_i! + \sum_{i=0}^{5} N_i \log \theta_i \\ &= \log 280! - \sum_{i=0}^{5} \log N_i! \\ &\qquad + \sum_{i=0}^{5} N_i \left[\log \binom{5}{i} + i \log p + (5-i)\log(1-p) \right]. \end{aligned}$$

It follows that

$$\begin{aligned} l'(p) &= \sum_{i=0}^{5} N_i \left[\frac{i}{p} - \frac{5-i}{1-p} \right], \quad \text{and} \\ l''(p) &= \sum_{i=0}^{5} N_i \left[-\frac{i}{p^2} - \frac{5-i}{(1-p)^2} \right] < 0. \end{aligned}$$

Thus $\max_{0 \le p \le 1} l(p)$ is achieved at $\hat{p}$ with $l'(\hat{p}) = 0$. Solving $l'(p) = 0$, we have

$$\begin{aligned} \sum_{i=0}^{5} N_i \left[\frac{i}{p} - \frac{5-i}{1-p} \right] = 0 &\Leftrightarrow \sum_{i=0}^{5} N_i \left[i(1-p) - (5-i)p \right] = 0 \\ &\Leftrightarrow \sum_{i=0}^{5} N_i \left[i - 5p \right] = 0 \\ &\Leftrightarrow \sum_{i=0}^{5} iN_i = 5 \left(\sum_{i=0}^{5} N_i \right) p \\ &\Leftrightarrow p = \frac{\sum_{i=0}^{5} iN_i}{5 \left(\sum_{i=0}^{5} N_i \right)}. \end{aligned}$$

Also notice that

$$p = \frac{\sum_{i=0}^{5} iN_i}{5\left(\sum_{i=0}^{5} N_i\right)} = \frac{\sum_{k=1}^{280} X_k}{280} \cdot \frac{1}{5} = \frac{1}{5}\overline{X},$$

where $\overline{X}$ is the sample mean of $X_1, \ldots, X_{280}$, which is the sample average number of breaks of these 280 bars.
Therefore, the mle for p under H_0 is

$$\hat{p} = \frac{\sum_{i=0}^{5} iN_i}{5\left(\sum_{i=0}^{5} N_i\right)} = \frac{\sum_{k=1}^{280} X_k}{280} \cdot \frac{1}{5} = \frac{1}{5}\overline{X},$$

For this particular sample, we have

$$\hat{p} = \frac{0 \cdot 157 + 1 \cdot 69 + 2 \cdot 35 + 3 \cdot 17 + 4 \cdot 1 + 5 \cdot 1}{280} \cdot \frac{1}{5} = 0.711.$$

b. Under H_0, we have $\theta_i(\hat{p}) = \binom{5}{i}\hat{p}^i(1-\hat{p})^{5-i}$. Thus for this sample, we have

$$\theta_0(\hat{p}) = 0.465,\ \theta_1(\hat{p}) = 0.385,\ \theta_2(\hat{p}) = 0.128,$$
$$\theta_3(\hat{p}) = 0.0211,\ \theta_4(\hat{p}) = 0.00175,\ \theta_5(\hat{p}) = 5.80 \times 10^{-5}$$

Accordingly, the expected counts are $E_i = 280 \cdot \theta_i(\hat{p})$:

$$E_0 = 130.1,\ E_1 = 107.8,\ E_2 = 35.7,\ E_3 = 5.9,\ E_4 = 0.5,\ E_5 = 0.$$

Since the last two counts are less than 5, we merge them with the observation $i = 3$ in order to form a single observation so that the expected count is larger than 5. As a result, the updated expected counts are:

$$E_0 = 130.1,\ E_1 = 107.8,\ E_2 = 35.7,\ E_3 = 6.4(= 5.9 + 0.5 + 0).$$

Since the observed counts are $O_i = N_i$, we have (with merge)

$$O_0 = 157,\ O_1 = 69,\ O_2 = 35,\ O_3 = 19(= 17 + 1 + 1).$$

Therefore, Pearson's χ^2 test statistic is

$$\begin{aligned} X^2 &= \sum_{i=0}^{5} \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(157 - 130.1)^2}{130.1} + \frac{(69 - 107.8)^2}{107.8} + \frac{(35 - 35.7)^2}{35.7} + \frac{(19 - 6.4)^2}{6.4} \\ &= 44.347. \end{aligned}$$

Since the number of free variables under H_0 is 1 (it is p) and the number of free variables under H_A is 5 $(= 6-1)$, X^2 is approximately χ^2_4.
Thus the p-value of this test statistic is 5.4×10^{-9}. Since the p-value is very small, we reject H_0. In other words, there is strong evidence that the bars are not of the same uniform strength.

If we use the table, we will have $\chi^2_{0.995,4} = 4.604$. Thus we will reject H_0 at $\alpha = 0.005$.

Name: ID:

Homework for 3/14

1. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution $N(\mu_X, \sigma^2)$, and $Y_1, \ldots, Y_m$ be a random sample from a normal distribution $N(\mu_Y, \sigma^2)$, with μ_X, μ_Y and σ unknown. We assume the X's and Y's are independent. Let $\overline{X}$ and s_X denote the sample mean and sample standard deviation of the $X_1, \ldots, X_n$, respectively. (Similar meaning for $\overline{Y}$ and s_Y.) Define the pooled sample variance as

$$s_p = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}.$$

We want to test the hypotheses $H_0 : \mu_X - \mu_Y = \Delta$ v.s. $H_A : \mu_X - \mu_Y > \Delta$. We use the test statistic

$$T = \frac{(\overline{X} - \overline{Y}) - \Delta}{s_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}},$$

and rejection region $T > c$.

a. Show that the significance level of this test is $1 - F_{n+m-2}(c)$, where F_{n+m-2} is the cdf of the t-distribution with degrees of freedom $n+m-2$, that is, $F_{n+m-2}(t) = \mathbb{P}(U \le t)$ with $U \sim t_{n+m-2}$.

b. Show that if the test is required to have significance level α, then $c = t_{1-\alpha, n+m-2}$. Here $t_{1-\alpha,n+m-2}$ satisfies $\mathbb{P}(U \le t_{1-\alpha,n+m-2}) = 1 - \alpha$.

c. Show that if the sample gives $T = t$, then the p-value in this case is $1 - F_{n+m-2}(t)$.

d. Assume σ is known. Show that if in fact $\mu_X - \mu_Y = \delta$ with $\delta > \Delta$, then the power of the test is $1 - \Phi\left(c + \dfrac{\Delta - \delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right)$. If the significance level is α, then the power is $1 - \Phi\left(z_{1-\alpha} + \dfrac{\Delta - \delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right)$.

a. Since the significance level is the probability of making a type I error, we have

$$\alpha = \mathbb{P}(\text{reject } H_0 \mid H_0) = \mathbb{P}(T > c | H_0) = \mathbb{P}\left(\frac{(\overline{X} - \overline{Y}) - \Delta}{s_p\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} > c \,\middle|\, H_0\right)$$

$$= 1 - F_{n+m-2}(c),$$

since

$$\frac{(\overline{X}-\overline{Y})-\Delta}{s_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim t_{n+m-2} \text{ under } H_0.$$

b. From the above derivation, we have

$$\alpha = 1 - F_{n+m-2}\left(c\right), \quad \text{or} \quad F_{n+m-2}\left(c\right) = 1-\alpha.$$

By the definition of $t_{1-\alpha,n+m-2}$, we have

$$c = t_{1-\alpha,n+m-2}.$$

c. Since the p-value is the probability of making type I error when we reject H_0 based on the sample, we have (by following a similar argument as in (a))

$$\begin{aligned} p &= \mathbb{P}(\text{reject } H_0 \text{ based on the sample } |H_0) = \mathbb{P}(T > t|H_0) \\ &= \mathbb{P}\left(\frac{(\overline{X}-\overline{Y})-\Delta}{s_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} > t \middle| H_0\right) \\ &= 1 - F_{n+m-2}(t), \end{aligned}$$

since

$$\frac{(\overline{X}-\overline{Y})-\Delta}{s_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim t_{n+m-2} \text{ under } H_0.$$

d. Note that here we assume σ is known. Correspondingly, we will change our test statistic to

$$T = \frac{(\overline{X}-\overline{Y})-\Delta}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}},$$

Therefore, we have

$$\frac{(\overline{X}-\overline{Y})-\Delta}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim N(0,1) \text{ under } H_0,$$

instead of

$$\frac{(\overline{X}-\overline{Y})-\Delta}{s_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim t_{n+m-2} \text{ under } H_0.$$

A similar argument as in (a) shows that

$$\alpha = \mathbb{P}(\text{reject } H_0 \mid H_0) = \mathbb{P}(Z > c \mid H_0) = \mathbb{P}\left(\frac{(\overline{X}-\overline{Y})-\Delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} > c \,\middle|\, H_0\right)$$

$$= 1 - \Phi\left(c\right),$$

since

$$\frac{(\overline{X}-\overline{Y})-\Delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim N(0,1) \text{ under } H_0,$$

Thus $c = z_{1-\alpha}$. The probability of making type II error is

$$\begin{aligned}
\beta &= \mathbb{P}(\text{accept } H_0 \mid H_A) = \mathbb{P}(Z \le c \mid H_A) \\
&= \mathbb{P}\left(\frac{(\overline{X}-\overline{Y})-\Delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \le c \,\middle|\, H_A\right) \\
&= \mathbb{P}\left(\frac{(\overline{X}-\overline{Y})-\delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \le c + \frac{\Delta-\delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \,\middle|\, H_A\right) \\
&= \mathbb{P}\left(Z \le c + \frac{\Delta-\delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right) \qquad Z \sim n(0,1) \\
&= \Phi\left(c + \frac{\Delta-\delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right).
\end{aligned}$$

Furthermore, the power of the test is

$$\text{power} = 1-\beta = 1-\Phi\left(c + \frac{\Delta-\delta}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right).$$

When the significance level is α, we derived from part (b) that

$$c = z_{1-\alpha}.$$

Thus, by combining the above two equations, the power of this test is

$$1 - \Phi\left(z_{1-\alpha} + \frac{\Delta - \delta}{\sigma\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}\right).$$

2. Low-back pain (LBP) is a serious health problem in many industrial settings. The article "Isodynamic Evaluation of Trunk Muscles and Low-Back Pain Among Workers in a Steel Factory" (*Ergonomics*, 1995: 2107-2117) reported the accompanying summary data on lateral range of motion (degrees) for a sample of workers without a history of LBP and another sample with a history of this malady.

Condition	Sample Size	Sample Mean	Sample SD
No LBP	28	91.5	5.5
LBP	31	88.3	7.8

We assume that the lateral range of motion under both conditions are normally distributed and have a common (unknown) standard deviation σ.

a. Calculate a 90% confidence interval for the difference between population mean extent of lateral motion for the two conditions.

b. Does the data suggest that population mean lateral motion differs for the two conditions?

Let X denote the sample for "no LBP" and Y for "LBP". Thus the sample size, sample mean, and sample standard deviation for X and Y are

$$\begin{aligned} \bar{x} &= 91.5, \quad s_X = 5.5, \quad n = 28, \\ \bar{y} &= 88.3, \quad s_Y = 7.8, \quad m = 31, \end{aligned}$$

respectively. Let μ_X and μ_Y denote the population mean extent of lateral motion for the two conditions, respectively.

(a) A $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is given by

$$\left((\bar{x}-\bar{y}) - t_{1-\frac{\alpha}{2},n+m-2} \cdot s,\ (\bar{x}-\bar{y}) - t_{1-\frac{\alpha}{2},n+m-2} \cdot s\right),$$

where s is the estimated standard error (deviation) of $\overline{X} - \overline{Y}$:

$$s = s_p\sqrt{\frac{1}{n} + \frac{1}{m}},$$

and s_p is the pooled sample standard deviation:

$$s_p = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}.$$

Since $\alpha = 0.1 (= 1 - 0.9)$, we have $t_{1-\frac{\alpha}{2},n+m-2} = t_{0.95,57} = 1.672$ (if we use TABLE 4 of the textbook, we can use $t_{0.95,60} = 1.671$ to approximate). Moreover,

$$s_p = \sqrt{\frac{(28-1)\cdot 5.5^2 + (31-1)\cdot 7.8^2}{28+31-2}} = 6.808,$$

and

$$s = 6.808 \cdot \sqrt{\frac{1}{28} + \frac{1}{31}} = 1.775.$$

Therefore, a 90% confidence interval for $\mu_X - \mu_Y$ is

$$\begin{aligned}&((91.5 - 88.3) - 1.672 \cdot 1.775,\ (91.5 - 88.3) + 1.672 \cdot 1.775)\\ &= (0.232, 6.168).\end{aligned}$$

If we use $t_{0.95,57} \approx t_{0.95,60} = 1.671$, then the C.I. is $(0.234, 6.166)$.

b. The hypotheses are

$$H_0 : \mu_X = \mu_Y \qquad H_A : \mu_X \neq \mu_Y.$$

The test statistic is

$$T = \frac{\overline{X} - \overline{Y}}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

and the rejection region is of the form $\{|T| > c\}$. Since for this sample

$$T = \frac{91.5 - 88.3}{1.775} = 1.803$$

and $T \sim t_{57}$, the p-value is

$$p = 2[1 - F_{57}(|1.803|)] = 0.077.$$

Thus if our significance level is $\alpha = 5\%$, we will not reject H_0, that is, we will conclude the population mean lateral motion does not differ for the two conditions.
However, if we use $\alpha = 10\%$, then we will reject H_0, that is, we will conclude that population mean lateral motion differs for the two conditions.

If we use TABLE 4, we will use $df = 60$ to approximated $df = 57$. In this case, we have $t_{0.95,60} = 1.671 < 1.803 < 2.000 = t_{0.975,60}$. Thus the p-value satisfies

$$(2 \cdot (1 - 0.975) =)\ 0.05 < p < 0.1\ (= 2 \cdot (1 - 0.95)),$$

which means we reject H_0 at $\alpha = 0.1$ but not at $\alpha = 0.05$.

Or equivalently, we can use rejection region $\{T > 1.671 = t_{0.95,60}\}$ corresponding to $\alpha = 0.1$, and $\{T > 2.000 = t_{0.975,60}\}$ corresponding to $\alpha = 0.05$ in order to reach the same conclusion.

3. The article "The Effects of a Low-Fat, Plant-Based Dietary Intervention on Body Weight, Metabolism, and Insulin Sensitivity in Postmenopausal Women" (*Amer. J. of Med.*, 2005: 991-997) reported on the results of an experiment in which half of the individuals in a group of 64 postmenopausal overweight women were randomly assigned to a particular vegan diet, and the other half received a diet based on National Cholesterol Education Program guidelines. The sample mean decrease in body weight for those on the vegan diet was 5.8 kg, and the sample SD was 3.2, whereas for those on the control diet, the sample mean weight loss and standard deviation were 3.8 and 2.8, respectively. Does it appear the true average weight loss for the vegan diet exceeds that for the control diet by more than 1 kg? Assume the weight loss under both diet are normally distributed with an unknown variance σ^2. Carry out an appropriate test of hypotheses at significance level .05 based on calculating a p-value.

Let X denote the sample for the group with the vegan diet and Y for the group with control diet. Thus the sample size, sample mean, and sample standard deviation for X and Y are

$$\begin{aligned} \bar{x} &= 5.8, \quad s_X = 3.2, \quad n = 32, \\ \bar{y} &= 3.8, \quad s_Y = 2.8, \quad m = 32, \end{aligned}$$

respectively. Let μ_X and μ_Y denote the population average weight loss for the two diets, respectively. They hypotheses are

$$H_0 : \mu_X - \mu_Y = 1 \qquad H_A : \mu_X - \mu_Y > 1.$$

The test statistic is

$$T = \frac{(\overline{X} - \overline{Y}) - 1}{s_p \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}},$$

and the rejection region is of the form $\{T > c\}$. For this sample, we have

$$s_p = \sqrt{\frac{(32-1)\cdot 3.2^2 + (32-1)\cdot 2.8^2}{32+32-2}} = 3.007,$$

and

$$T = \frac{(5.8 - 3.8) - 1}{3.007 \cdot \sqrt{\dfrac{1}{32} + \dfrac{1}{32}}} = 1.330.$$

Since $T \sim t_{62}$, the p-value is

$$p = 1 - F_{62}(2.660) = 0.094.$$

Therefore, we will not reject H_0 at significance level $\alpha = 0.05$, that is, there is no significant evidence that the true average weight loss for the vegan diet exceeds that for the control diet by more than 1 kg.

We can also find the rejection region according to $\alpha = 0.05$: $\{T > t_{0.95,62} = 1.670\}$. The same conclusion will be made since $1.330 < 1.670$.

If we use TABLE 4 of the textbook, we will use $df = 60$ to approximate $df = 62$. In this case, we will have $t_{0.0.90,60} = 1.296 < 1.330 < 1.671 = t_{0.95,60}$ which implies that $\alpha = 0.05 < p < 0.10$, and $t_{0.95,60} = 1.671 > 1.330$.

4. [§11-15] Suppose that n measurements are to be taken under a treatment condition and another n measurements are to be taken independently under a control condition. It is thought that the standard deviation of a single observation is about 10 under both conditions. How large should n be so that a 95% confidence interval for $\mu_X - \mu_Y$ has a width of 2? Use the normal distribution rather than the t distribution, since we assume $\sigma = 10$.

When σ is known, we use the fact

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right).$$

Correspondingly, a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is given by

$$\left((\overline{X} - \overline{Y}) - z_{1-\frac{\alpha}{2}} \cdot \sigma\sqrt{\frac{1}{n} + \frac{1}{m}},\ (\overline{X} - \overline{Y}) + z_{1-\frac{\alpha}{2}} \cdot \sigma\sqrt{\frac{1}{n} + \frac{1}{m}}\right).$$

Thus the width of the confidence interval is

$$w = 2z_{1-\frac{\alpha}{2}} \cdot \sigma\sqrt{\frac{1}{n} + \frac{1}{m}}.$$

In this case, we have $m = n$. Therefore

$$w = 2z_{1-\frac{\alpha}{2}} \cdot \sigma\sqrt{\frac{2}{n}}.$$

In order to reach a desired width of confidence intervals with given confidence level, we need sample size to satisfy

$$n = \left(\frac{2\sqrt{2} \cdot z_{1-\frac{\alpha}{2}} \cdot \sigma}{w}\right)^2.$$

Since $\alpha = 0.05 (= 1 - 0.95)$, we have $z_{0.975} = 1.96$. In this case, $\sigma = 10$ and $w = 2$. Thus,

$$n = \left(\frac{2\sqrt{2} \cdot 1.96 \cdot 10}{2}\right)^2 = 768.32 \approx 769.$$

Note that sample size n must be a positive integer. And we typically round up when the calculated value is not an integer. The reason is that we want to guarantee that the width is no more than the given value and the confidence level is no less than the given value.

Appendix

1' Assume the same setting as in Problem 1. In this case we want to test the hypotheses $H_0 : \mu_X - \mu_Y = \Delta$ v.s. $H_A : \mu_X - \mu_Y < \Delta$. We use the test statistic

$$T = \frac{(\overline{X} - \overline{Y}) - \Delta}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

and rejection region $T < c$.

a. The significance level of this test is $F_{n+m-2}(c)$.

b. If the test is required to have significance level α, then $c = t_{\alpha,n+m-2}$.

c. If the sample gives $T = t$, then the p-value in this case is $F_{n+m-2}(t)$.

d. Assume σ is known. If in fact $\mu_X - \mu_Y = \delta$ with $\delta < \Delta$, then the power of the test is $\Phi\left(c + \frac{\Delta - \delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right)$. If the significance level is α, then the power is $\Phi\left(z_\alpha + \frac{\Delta - \delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right)$.

1" Assume the same setting as in Problem 1. In this case we want to test the hypotheses $H_0 : \mu_X - \mu_Y = \Delta$ v.s. $H_A : \mu_X - \mu_Y \neq \Delta$. We use the test statistic

$$T = \frac{(\overline{X} - \overline{Y}) - \Delta}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

and rejection region $|T| > c$.

a. The significance level of this test is $2(1 - F_{n+m-2}(c))$.

b. If the test is required to have significance level α, then $c = t_{1-\frac{\alpha}{2},n+m-2}$.

c. If the sample gives $T = t$, then the p-value in this case is $2(1 - F_{n+m-2}(|t|))$.

d. Assume σ is known. If in fact $\mu_X - \mu_Y = \delta$ with $\delta \neq \Delta$, then the power of the test is $1 - \Phi\left(c + \frac{\Delta - \delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right) + \Phi\left(-c + \frac{\Delta - \delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right)$. If the significance level is α, then the power is

$$1 - \Phi\left(z_{1-\frac{\alpha}{2}} + \frac{\Delta - \delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right) + \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{\Delta - \delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right).$$