Mathematical Statistics
Homework 4

Nolan R. H. Gagnon

Wackerly 10.126

Suppose that $X_1, X_2, \ldots, X_{n_1}$, $Y_1, Y_2, \ldots, Y_{n_2}$, and $W_1, W_2, \ldots, W_{n_3}$ are independent random samples from normal distributions with respective unknown means μ_1, μ_2, and μ_3 and common variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_2$. Suppose that we want to estimate a linear function of the means: $\theta = a_1\mu_1 + a_2\mu_2 + a_3\mu_3$. Because the maximum-likelihood estimator (MLE) of a function of parameters is the function of the MLEs of the parameters, the MLE of θ is $\hat{\theta} = a_1\bar{X} + a_2\bar{Y} + a_3\bar{W}$.

(a) What is the standard error of the estimator $\hat{\theta}$?

(b) What is the distribution of $\hat{\theta}$?

(c) If the sample variances are given by S_1^2, S_2^2, and S_3^2, respectively, consider

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2}{n_1 + n_2 + n_3 - 3}.$$

(i) What is the distribution of $(n_1 + n_2 + n_3 - 3)S_p^2/\sigma^2$?

(ii) What is the distribution of

$$T = \frac{\hat{\theta} - \theta}{S_p\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}}?$$

(d) Give a confidence interval for θ with confidence coefficient $1 - \alpha$.

(e) Develop a test for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

Solution

(a) Since the samples are independent of each other, the variance of $\hat{\theta}$ is

$$\begin{aligned}
\text{Var}\left(\hat{\theta}\right) &= \text{Var}\left(a_1\bar{X} + a_2\bar{Y} + a_3\bar{W}\right) && (1)\\
&= a_1^2\text{Var}\left(\bar{X}\right) + a_2^2\text{Var}\left(\bar{Y}\right) + a_3^2\text{Var}\left(\bar{W}\right) && (2)\\
&= a_1^2\frac{\sigma^2}{n_1} + a_2^2\frac{\sigma^2}{n_2} + a_3\frac{\sigma^2}{n_3} && (3)\\
&= \sigma^2\left(\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}\right). && (4)
\end{aligned}$$

The standard error of $\hat{\theta}$ is

$$\begin{aligned}
\mathrm{SE}\left(\hat{\theta}\right) &= \sqrt{\mathrm{Var}\left(\hat{\theta}\right)} && (5) \\
&= \sigma\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}. && (6)
\end{aligned}$$

(b) By Theorem 5.3.1 (Casella and Berger § 5.3),

$$\begin{aligned}
\bar{X} &\sim n\left(\mu_1, \frac{\sigma^2}{n_1}\right) && (7) \\
\bar{Y} &\sim n\left(\mu_2, \frac{\sigma^2}{n_2}\right) && (8) \\
\bar{W} &\sim n\left(\mu_3, \frac{\sigma^2}{n_3}\right). && (9)
\end{aligned}$$

Using the method of moment-generating functions, it is easy to see that

$$\begin{aligned}
a_1\bar{X} &\sim n\left(a_1\mu_1, \frac{a_1^2\sigma^2}{n_1}\right) && (10) \\
a_2\bar{Y} &\sim n\left(a_2\mu_2, \frac{a_2^2\sigma^2}{n_1}\right) && (11) \\
a_3\bar{W} &\sim n\left(a_3\mu_3, \frac{a_3^2\sigma^2}{n_1}\right). && (12)
\end{aligned}$$

Thus $\hat{\theta}$ is normal with mean $a_1\mu_1 + a_2\mu_2 + a_3\mu_3 = \theta$, and variance $\sigma^2\left(\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}\right)$.

(c) **(i)** Notice that we can write

$$(n_1 + n_2 + n_3 - 3)S_p^2/\sigma^2 = \frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} + \frac{(n_3-1)S_3^2}{\sigma^2}. \qquad (13)$$

By Theorem 5.3.1 (Casella and Berger § 5.3),

$$\begin{aligned}
\frac{(n_1-1)S_1^2}{\sigma^2} &\sim \chi^2(n_1 - 1) && (14) \\
\frac{(n_2-1)S_2^2}{\sigma^2} &\sim \chi^2(n_2 - 1) && (15) \\
\frac{(n_3-1)S_3^2}{\sigma^2} &\sim \chi^2(n_3 - 1) && (16)
\end{aligned}$$

Note that the above variables are independent. Now, a sum of k independent chi-squared random variables (with degrees of freedom p_i) is again a chi-squared random variable with $\sum_{i=1}^{k} p_i$ degrees of freedom. Thus,

$$(n_1 + n_2 + n_3 - 3)S_p^2/\sigma^2 \sim \chi^2\left(n_1 + n_2 + n_3 - 3\right).$$

(ii) We can write

$$T = \frac{\hat{\theta} - \theta}{S_p\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}} \tag{17}$$

$$= \frac{\hat{\theta} - \theta}{\sqrt{\frac{1}{n_1+n_2+n_3-3}(n_1 + n_2 + n_3 - 3)S_p^2/\sigma^2}\left(\sigma\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}\right)} \tag{18}$$

$$= \left(\frac{\hat{\theta} - \theta}{\sigma\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}}\right)\frac{1}{\sqrt{\frac{1}{n_1+n_2+n_3-3}(n_1 + n_2 + n_3 - 3)S_p^2/\sigma^2}}. \tag{19}$$

Note that $Z = \frac{\hat{\theta}-\theta}{\sigma\sqrt{\frac{a_1^2}{n_1}+\frac{a_2^2}{n_2}+\frac{a_3^2}{n_3}}}$ is a standard normal random variable and $X = (n_1 + n_2 + n_3 - 3)S_p^2/\sigma^2$ is a chi-squared random variable with $n_1 + n_2 + n_3 - 3$ degrees of freedom. Thus, we can write T as

$$T = \frac{Z}{\sqrt{X/(n_1 + n_2 + n_3 - 3)}} \tag{20}$$

If Z and X are independent, then T has a t distribution with $n_1+n_2+n_3-3$ degrees of freedom. To show that Z and X are independent, we only need to show that $\hat{\theta}$ and S_p^2 are independent. Now, from Theorem 5.3.1 (Casella and Berger § 5.3) we know that $\bar{X}$ is independent of S_1^2, $\bar{Y}$ is independent of S_2^2, and $\bar{W}$ is independent of S_3^2. Since $\hat{\theta}$ is a linear combination of $\bar{X}, \bar{Y}$, and $\bar{W}$ and S_p^2 is a linear combination of S_1^2, S_2^2, and S_3^2, it follows that $\hat{\theta}$ and S_p^2 are independent. Therefore, Z and X are independent. Thus, $T \sim t(n_1 + n_2 + n_3 - 3)$.

(d) Since the variances of the distributions are unknown, the confidence interval for θ with confidence coefficient $1 - \alpha$ is

$$\hat{\theta} \pm t_{\alpha/2;n_1+n_2+n_3-3}S_p\sqrt{\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \frac{a_3^2}{n_3}}.$$

(e) The hypotheses $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$ can be tested using the T statistic from the previous problem. Since the alternative hypothesis contains a "not equals“ sign, the test will be two-sides with rejection region

$$|T| > t_{\alpha/2;n_1+n_2+n_3-3}.$$

■

Wackerly 10.127

A merchant figures her weekly profit to be a function of three variables: retail sales (denoted by X), wholesale sales (denoted by Y), and overhead costs (denoted by W). The variables X, Y, and W are regarded as independent, normally distributed random variables with means μ_1, μ_2, and μ_3 and variances $\sigma^2, a\sigma^2$, and $b\sigma^2$, respectively, for known constants a and b but unknown σ^2. The merchant's expected profit per week is $\mu_1 + \mu_2 - \mu_3$. If the merchant has made independent observations of X, Y, and W for the past n weeks, construct a test of $H_0 : \mu_1 + \mu_2 - \mu_3 = k$ against the alternative $H_a : \mu_1 + \mu_2 - \mu_3 \neq k$, for a given constant k. You may specify $\alpha = 0.05$.

Solution

Let $\theta = \mu_1 + \mu_2 - \mu_3$. We can rewrite the hypotheses as $H_0 : \theta = k$ versus $H_a : \theta \neq k$. The MLE of θ is $\hat{\theta} = \bar{X} + \bar{Y} - \bar{W}$, which has a normal distribution with mean $\mu_1 + \mu_2 - \mu_3$ and variance $\frac{\sigma^2}{n}(1+a+b)$. The distribution of $\hat{\theta} - \theta$ is normal with mean 0 and variance $\frac{\sigma^2}{n}(1+a+b)$. Therefore, the distribution of

$$Z = \frac{\hat{\theta} - \theta}{\sigma\sqrt{\frac{1}{n}(1+a+b)}}$$

is standard normal. Next, define

$$S_p^2 = \frac{(n-1)S_x^2 + \frac{(n-1)}{a}S_y^2 + \frac{n-1}{b}S_w^2}{3n-3},$$

where S_x^2, S_y^2, and S_w^2 are the respective sample variances. Then

$$\chi = (3n-3)S_p^2/\sigma^2 \tag{21}$$

$$= \frac{(n-1)S_x^2}{\sigma^2} + \frac{(n-1)S_y^2}{a\sigma^2} + \frac{(n-1)S_w^2}{b\sigma^2} \tag{22}$$

has a chi-squared distribution with $3n-3$ degrees of freedom. By applying Theorem 5.3.1 (Casella and Berger § 5.3), we conclude that Z and χ are independent. So the distribution of

$$\frac{Z}{\sqrt{\chi/(3n-3)}} = \left(\frac{1}{S_p/\sigma}\right)\left(\frac{\hat{\theta}-\theta}{\sigma\sqrt{\frac{1}{n}(1+a+b)}}\right) \tag{23}$$

$$= \frac{\hat{\theta}-\theta}{S_p\sqrt{\frac{1}{n}(1+a+b)}} \tag{24}$$

is t with $3n-3$ degrees of freedom. So we reject H_0 if

$$\left|\frac{\hat{\theta}-\theta}{S_p\sqrt{\frac{1}{n}(1+a+b)}}\right| > t_{0.025;3n-3}.$$

■

Wackerly 10.128

A reading exam is given to the sixth graders at three large elementary schools. The scores on the exam at each school are regarded as having normal distributions with unknown means μ_1, μ_2, and μ_3, respectively, and unknown common variance σ^2 ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$). Using the data in the accompanying table on independent random samples from each school, test to see if evidence exists of a difference between μ_1 and μ_2. Use $\alpha = 0.05$.

Solution

This problem can be solved using the theory developed in problem 10.126. Letting $\theta = \mu_1 - \mu_2 + 0\mu_3 = \mu_1 - \mu_2$, our hypothesis test becomes $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. We will reject H_0 if the test statistic

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{10} + \frac{1}{10}}}$$

satisfies $|T| > t_{0.025;27} = 2.051831$. Using the data from the table, we find that

$$\begin{aligned} \bar{x} - \bar{y} &= 60 - 50 && (25) \\ &= 10, && (26) \end{aligned}$$

and

$$\begin{aligned} S_p^2 &= \frac{9(S_x^2 + S_y^2 + S_w^2)}{27} && (27) \\ &= \frac{9}{27}\left(\frac{\sum x_i^2 - n\bar{x}^2}{9} + \frac{\sum y_i^2 - n\bar{y}^2}{9} + \frac{\sum w_i^2 - n\bar{w}^2}{9}\right) && (28) \\ &= \frac{1}{27}[36950 - 10(60^2) + 25850 - 10(50^2) + 49900 - 10(70^2)] && (29) \\ &= \frac{2700}{27} && (30) \\ &= 100. && (31) \end{aligned}$$

Thus,

$$\begin{aligned} T &= \frac{10}{10\sqrt{2/10}} && (32) \\ &= \sqrt{5} && (33) \\ &\approx 2.23607. && (34) \end{aligned}$$

Since $|T| = 2.23607 > 2.051831$, we reject H_0 and conclude that there is evidence supporting the hypothesis that there is a difference between μ_1 and μ_2.

■

Wackerly 10.129

Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function given by

$$f(y|\theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_1} e^{-(y-\theta_2)/\theta_1}, & y > \theta_2 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the likelihood ratio test for testing $H_0 : \theta_1 = \theta_{1,0}$ versus $H_a : \theta_1 > \theta_{1,0}$ with θ_2 unknown.

Solution

The likelihood function for the sample is

$$L(\theta_1, \theta_2|\mathbf{y}) = \prod_{i=1}^{n} f(y_i|\theta_1, \theta_2) \tag{35}$$

$$= \begin{cases} \left(\frac{1}{\theta_1}\right)^n \exp\left\{-\frac{1}{\theta_1}\sum_{i=1}^{n}(y_i - \theta_2)\right\}, & \theta_2 \leq y_{(1)} \\ 0, & \text{elsewhere} \end{cases} \tag{36}$$

The log-likelihood function for $\theta_2 < y_{(1)}$ is then

$$\mathscr{L}(\theta_1, \theta_2|\mathbf{y}) = -n\log(\theta_1) - \frac{1}{\theta_1}\sum_{i=1}^{n}(y_i - \theta_2). \tag{37}$$

For other values of θ_2, it doesn't exist. Now we will use these two functions to find the MLEs of θ_1 and θ_2 (which we will denote by $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively). Examining the likelihood function, we see that on the interval $-\infty < \theta_2 < y_{(1)}$, it is an increasing function of θ_2. Thus, the MLE of θ_2 is $\hat{\theta}_2 = y_{(1)}$. To find the MLE of θ_1, we solve

$$\frac{d}{d\theta_1}\mathscr{L}(\theta_1, \hat{\theta}_2|\mathbf{y}) = -\frac{n}{\theta_1} + \frac{1}{\theta_1^2}\sum_{i=1}^{n}(y_i - y_{(1)}) \tag{38}$$

$$= 0 \tag{39}$$

for θ_1. It is easy to see that the solution is $\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{(1)})$.

The likelihood ratio test statistic for the hypotheses in this problem is

$$\lambda_1(\mathbf{y}) = \frac{L(\theta_{1,0}, \hat{\theta}_2|\mathbf{y})}{L(\hat{\theta}_1, \hat{\theta}_2|\mathbf{y})} \tag{40}$$

$$= \left(\frac{\hat{\theta}_1}{\theta_{1,0}}\right)^n \frac{\exp\left\{-\frac{1}{\theta_{1,0}}\sum(y_i - y_{(1)})\right\}}{\exp\left\{-\frac{1}{\hat{\theta}_1}\sum(y_i - y_{(1)})\right\}} \tag{41}$$

$$= \left(\frac{\sum(y_i - y_{(1)})}{n\theta_{1,0}}\right)^n \exp\left\{-\frac{1}{\theta_{1,0}}\sum(y_i - y_{(1)}) + n\right\}. \tag{42}$$

The likelihood ratio test is to reject H_0 whenever $\lambda_1(\mathbf{y}) < c$, where $0 \leq c \leq 1$.

■

Wackerly 10.130

Refer to Exercise 10.129. Find the likelihood ratio test for testing $H_0 : \theta_2 = \theta_{2,0}$ versus $H_a : \theta_2 > \theta_{2,0}$, with θ_1 unknown.

Solution

If θ_1 and θ_2 are restricted to values specified by the null hypothesis, then L is maximized when $\theta_2 = \theta_{2,0}$ and $\theta_1 = \frac{1}{n}\sum(y_i - \theta_{2,0})$. Thus, the likelihood test statistic for this hypothesis test is

$$\lambda_2(\mathbf{y}) = \frac{L\left(\frac{1}{n}\sum(y_i - \theta_{2,0}), \theta_{2,0}|\mathbf{y}\right)}{L(\hat{\theta}_1, \hat{\theta}_2|\mathbf{y})} \tag{43}$$

$$= \left[\frac{\frac{1}{n}\sum(y_i - y_{(1)})}{\frac{1}{n}\sum(y_i - \theta_{2,0})}\right]^n \frac{\exp\left\{-\frac{\sum(y_i-\theta_{2,0})}{\frac{1}{n}\sum(y_i-\theta_{2,0})}\right\}}{\exp\left\{-\frac{\sum(y_i-y_{(1)})}{\frac{1}{n}\sum(y_i-y_{(1)})}\right\}} \tag{44}$$

$$= \left[\frac{\sum(y_i - y_{(1)})}{\sum(y_i - \theta_{2,0})}\right]^n \frac{\exp\{-n\}}{\exp\{-n\}} \tag{45}$$

$$= \left[\frac{\sum(y_i - y_{(1)})}{\sum(y_i - \theta_{2,0})}\right]^n. \tag{46}$$

Again, the likelihood ratio test is to reject H_0 whenever $\lambda_1(\mathbf{y}) < c$, where $0 \le c \le 1$.

■