Math 472 Homework Assignment 4

Problem 4.2.11. Let $X_1, X_2, \ldots, X_n, X_{n+1}$ be a random sample of size $n+1$, $n > 1$, from a distribution that is $N(\mu, \sigma^2)$. Let $\overline{X} = \sum_{i=1}^n X_i/n$ and $S^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$. Find the constant c so that the statistic $c(\overline{X} - X_{n+1})/S$ has a t-distribution. If $n = 8$, determine k such that

$$P(\overline{X} - kS < X_9 < \overline{X} + kS) = 0.80.$$

The observed interval $(\bar{x} - ks, \bar{x} + ks)$ is often called an 80% **prediction interval** for X_9.

Solution 4.2.11. The random variable $\overline{X} - X_{n+1}$ has a normal distribution since it is a linear combination of independent normally distributed random variables. Its expected value and variance are

$$\begin{aligned} E[\overline{X} - X_{n+1}] &= \mu - \mu = 0, \\ \operatorname{Var}[\overline{X} - X_{n+1}] &= \frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(\frac{n+1}{n}\right). \end{aligned}$$

Therefore, $\sqrt{n/(n+1)}(\overline{X} - X_{n+1})/\sigma$ has a standard normal distribution.

By Student's theorem, $\overline{X}$ and S^2 are independent random variables, and $(n-1)S^2/\sigma^2$ has a χ-square distribution with $r = n-1$ degrees of freedom. Since $\overline{X}$ and S^2 are functions of $X_1, \ldots, X_n$, we see that $\overline{X}$, S^2, and X_{n+1} are independent random variables, therefore $\overline{X} - X_{n+1}$ and S^2 are independent random variables. From these observations we conclude that

$$\frac{\sqrt{n/(n+1)}(\overline{X} - X_{n+1})/\sigma}{\sqrt{S^2/\sigma^2}} = \frac{\sqrt{n/(n+1)}(\overline{X} - X_{n+1})}{S}$$

has a t-distribution with $r = n-1$ degrees of freedom. It follows that

$$c = \sqrt{\frac{n}{n+1}}.$$

Let $n = 8$. Then $c = \sqrt{8/9} = 2\sqrt{2}/3$. Let T_7 be a t-distributed random variable with 7 degrees of freedom and let $t_{0.10,7}$ be the number defined by $P(T_7 > t_{0.10,7}) = 0.10$. Then $P(-t_{0.10,7} < T_7 < t_{0.10,7}) = 0.80$. This, combined with the observation that

$$-t_{0.10,7} < \frac{c(\overline{X} - X_9)}{S} < t_{0.10,7} \iff \overline{X} - \frac{t_{0.10,7}}{c}S < X_9 < \overline{X} + \frac{t_{0.10,7}}{c}S,$$

shows that

$$k = \frac{t_{0.10,7}}{c} = \frac{3t_{0.10,7}}{2\sqrt{2}} \doteq \frac{(3)(1.415)}{2.828} \doteq 1.501.$$

Problem 4.2.18. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, where both parameters μ and σ^2 are unknown. A *confidence interval* for σ^2 can be found as follows. We know that $(n-1)S^2/\sigma^2$ is a random variable with a $\chi^2(n-1)$ distribution. Thus we can find constants a and b so that $P((n-1)S^2/\sigma^2 < b) = 0.975$ and $P(a < (n-1)S^2/\sigma^2 < b) = 0.95$.

(a) Show that this second probability statement can be written as

$$P\left(\frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}\right) = 0.95.$$

(b) If $n = 9$ and $s^2 = 7.93$, find a 95% confidence interval for σ^2.

(c) If μ is known, how would you modify the preceding procedure for finding a confidence interval for σ^2?

Solution 4.2.18.

(a) With the numbers a and b chosen as above, the result follows immediately from the observation that

$$a < \frac{(n-1)S^2}{\sigma^2} < b \iff \frac{(n-1)S^2}{b} < \sigma^2 < \frac{(n-1)S^2}{a}.$$

(b) Let W be a χ^2-distributed random variable with 8 degrees of freedom. Then $P(W < 17.5345) \doteq 0.975$ and $P(2.1797 < W < 17.5345) \doteq 0.95$. With $n = 9$ we calculate the lower confidence limit to be $(n-1)s^2/b \doteq (8)(7.93)/(17.5345) \doteq 3.618$ and the upper confidence limit to be $(n-1)s^2/a \doteq (8)(7.93)/(2.1797) \doteq 29.104$. Therefore the 95% confidence interval for σ^2 is

$$(3.618, 29.104).$$

(c) If μ is known then we can replace $(n-1)S^2/\sigma^2$ with $\sum_{i=1}^n (X_i - \mu)^2/\sigma^2$, which has a χ-square distribution with n degrees of freedom, as a pivotal random variable.

Problem 4.2.25. To illustrate Exercise 4.2.24, let $X_1, \ldots, X_9$ and $Y_1, \ldots, Y_{12}$ represent two independent random samples from the respective normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. It is given that $\sigma_1^2 = 3\sigma_2^2$, but σ_2^2 is unknown. Define a random variable that has a t-distribution that can be used to find a 95% confidence interval for $\mu_1 - \mu_2$.

Solution 4.2.25. Let $\overline{X} = \sum_{i=1}^9 X_i/9$ and $\overline{Y} = \sum_{i=1}^{12} Y_i/12$. Since $\overline{X} - \overline{Y}$ is a linear combination of independent normal random variables, it has a normal distribution with

$$\begin{aligned} E[\overline{X}-\overline{Y}] &= \mu_1-\mu_2 \\ \operatorname{Var}[\overline{X}-\overline{Y}] &= \frac{\sigma_1^2}{9}+\frac{\sigma_2^2}{12} \\ &= \frac{\sigma_1^2}{9}+\frac{3\sigma_1^2}{12} \\ &= \sigma_1^2\left(\frac{1}{9}+\frac{1}{4}\right). \end{aligned}$$

Thus the random variable

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma_1\sqrt{1/9+1/4}} \tag{1}$$

has a standard normal distribution.

Since $8S_1^2/\sigma_1^2$ and $11S_2^2/\sigma_2^2 = 11S_2^2/(3\sigma_1^2)$ are independent random variables that have χ^2-distributions with respective degrees of freedom 8 and 11, the random variable

$$19S_p^2/\sigma_1^2 = 8S_1^2/\sigma_1^2 + 11S_2^2/(3\sigma_1^2) = (8S_1^2+11S_2^2/3)/\sigma_1^2$$

has a χ^2-distribution with 19 degrees of freedom. Therefore S_p/σ_1 is the square root of a χ^2-distributed random variable divided by its degrees of freedom. Since Student's Theorem implies that $\overline{X}$, $\overline{Y}$, S_1^2, and S_2^2 are independent random variables, we see that the random variable in equation (1) and S_p^2 are independent. Therefore

$$\frac{\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma_1\sqrt{1/9+1/4}}}{S_p/\sigma_1} = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_p\sqrt{1/9+1/4}}$$

has a t-distribution with 19 degrees of freedom that can be used for a 95% confidence interval for $\mu_1-\mu_2$, where $S_p^2 = (8S_1^2+11S_2^2/3)/19$.

Problem 4.2.27. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be two independent random samples from the respective normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, where the four parameters are unknown. To construct a *confidence interval for the ratio*, σ_1^2/σ_2^2, of the variances, form the quotient of the two independent χ^2 variables, each divided by its degrees of freedom, namely,

$$F = \frac{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)}{\frac{(n-1)S_2^2}{\sigma_1^2}/(n-1)} = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2}, \tag{2}$$

where S_1^2 and S_2^2 are the respective sample variances.

(a) What kind of distribution does F have?

(b) From the appropriate table, a and b can be found so that $P(F<b) = 0.975$ and $P(a<F<b) = 0.95$.

(c) Rewrite the second probability statement as

$$P\left[a\frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < b\frac{S_1^2}{S_2^2}\right] = 0.95. \tag{3}$$

The observed values, s_1^2 and s_2^2, can be inserted in these inequalities to provide a 95% confidence interval for σ_1^2/σ_2^2.

Solution 4.2.27.

(a) The random variable F in equation (2) has an F-distribution with $r_1 = m$ numerator degrees of freedom and $r_2 = n$ denominator degrees of freedom.

(b) Using the notation of Table V on page 661, $b = F_{0.025}(m, n)$ and $a = F_{0.975}(m, n)$.

(c) A direct calculation shows that

$$a < \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} < b \iff a\frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < b\frac{S_1^2}{S_2^2}.$$

Combining this calculation with $P(a < F < b) = 0.95$ leads immediately to equation (3). We therefore find a 95% confidence interval for the ratio σ_1^2/σ_2^2 is given by

$$\left(F_{0.975}(m, n)\frac{S_1^2}{S_2^2}, F_{0.0275}(m, n)\frac{S_1^2}{S_2^2}\right).$$

Problem 4.6.5. Assume that the weight of cereal in a "10-ounce box" is $N(\mu, \sigma)$. To test $H_0 : \mu = 10.1$ against $H_1 : \mu > 10.1$, we take a random sample of size $n = 16$ and observe that $\bar{x} = 10.4$ and $s = 0.4$.

(a) Do we accept or reject H_0 at the 5% significance level?
(b) What is the approximate p-value of this test?

Solution 4.6.5.

(a) The test statistic, $\sqrt{16}(\overline{X} - 10.1)/S$ has a Student's t-distribution with $r = 15$ degrees of freedom. The realization of the test statistic is $t = (4)(10.4 - 10.1)/(0.4) = 3$, and the critical value for the right-tailed alternative H_1 is $t_{0.05,15} \doteq 1.753$. Since $t > t_{0.05,15}$, we reject H_0 at the 5% significance level.

(b) The approximate p-value of this test is $p = P(T_{15} > 3) \doteq 0.0045$.

Problem 4.6.6. Each of 51 golfers hit three golf balls of brand X and three golf balls of brand Y in a random order. Let X_i and Y_i equal the averages of the distances traveled by the brand X and brand Y golf balls hit by the ith golfer, $i = 1, 2, \ldots, 51$. Let $W_i = X_i - Y_i$, $i = 1, 2, \ldots, 51$. Test $H_0 : \mu_W = 0$ against $H_1 : \mu_W > 0$, where μ_W is the mean of the differences. If $\bar{w} = 2.07$ and $s_W^2 = 84.63$, would H_0 be accepted or rejected at an $\alpha = 0.05$ significance level? What is the p-value of this test?

Solution 4.6.6. We may assume that the distribution of the test statistic $(\overline{W} - \mu_W)/(S_W/\sqrt{51})$ is approximately the standard normal distribution. Under the null hypothesis the realization of the test statistic is

$$z \doteq \frac{\bar{w} - 0}{\sqrt{s_w^2/n}} \doteq \frac{2.07}{\sqrt{84.63/51}} \doteq 1.607.$$

The critical value for the right-tailed alternative hypothesis is $z_{0.05} \doteq 1.645$, and since $z < z_{0.05}$ we accept H_0 at the $\alpha = 0.05$ significance level. The p-value of this test is $p \doteq P(Z > 1.607) \doteq 0.054$.

Problem 4.6.7. Among the data collected for the World Health Organization air quality monitoring project is a measure of suspended particles in $\mu g/m^3$. Let X and Y equal the concentration of suspended particles in $\mu g/m^3$ in the city center (commercial district) for Melbourne and Houston, respectively. Using $n = 13$ observations of X and $m = 16$ observations of Y, we test $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$.

(a) Define the test statistic and critical region, assuming that the unknown variances are equal. Let $\alpha = 0.05$.
(b) If $\bar{x} = 72.9$, $s_x = 25.6$, $\bar{y} = 81.7$, and $s_y = 28.3$, calculate the value of the test statistic and state your conclusion.

Solution 4.6.7.
(a) We are testing a difference in means and the sample sizes are small. This suggests using the random variable

$$T_{27} = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{1/13 + 1/16}},$$
$$\text{where } S_p^2 = \frac{12S_X^2 + 15S_Y^2}{27},$$

as our test statistic. The random variable T_{27} has an approximate Student's t-distribution with 27 degrees of freedom. We are testing the left-tailed alternative $\mu_X - \mu_Y < 0$, which leads to the critical region $t < -t_{0.05,27}$, where $t_{0.05,27} \doteq 1.703$.

(b) From the given data, the realization of S_p is

$$s_p = \sqrt{\frac{12s_x^2 + 15s_y^2}{27}} \doteq \sqrt{\frac{(12)(25.6)^2 + (15)(28.3)^2}{27}} \doteq 27.133$$

and, assuming H_0 is true, the realization of T_{27} is

$$t = \frac{(\bar{x} - \bar{y}) - 0}{s_p\sqrt{1/13 + 1/16}} \doteq \frac{72.9 - 81.7}{(27.133)(0.3734)} \doteq -0.8686.$$

Since $t > -1.703$, t lies outside the critical region. Therefore these data provide insufficient evidence to reject H_0 at the significance level $\alpha = 0.05$. The p-value of this test is $p \doteq P(T_{27} < -1.703) \doteq 0.196$.

Problem 4.6.8. Let p equal the proportion of drivers who use a seat belt in a country that does not have a mandatory seat belt law. It was claimed that $p = 0.14$. An advertising campaign was conducted to increase this proportion. Two months after the campaign, $y = 104$ out of a random sample of $n = 590$ drivers were wearing their seat belts. Was the campaign successful?

Solution 4.6.8. We will perform a large sample hypothesis test to test whether the value of p increased after the advertising campaign. We test whether the data support rejection of the null hypothesis $H_0 : p = 0.14$ in favor of the right-tailed alternative hypothesis $H_a : p > 0.14$. We will use the random variable

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}},$$

whose distribution is approximately the standard normal distribution, as the test statistic. We will test at the significance level of $\alpha = 0.05$. The approximate rejection region is $z > z_{0.05} \doteq 1.645$.

From the given data, the realized value of $\hat{p}$ is $\hat{p} = 104/590 \doteq 0.176$ and, assume the null hypothesis is true, the realized value of Z is

$$z \doteq \frac{0.176 - 0.14}{\sqrt{(0.14)(1-0.14)/590}} \doteq 2.539.$$

and the approximate p-value of this test is $p \doteq P(Z > 2.539) \doteq 0.0056$. Since $z > 1.645$ these data support rejection of the null hypothesis at the significance level of $\alpha = 0.05$. In fact, the p-value shows that these data would support rejection of H_0 in favor of H_a even at the significance level of $\alpha = 0.01$. These data certainly support the claim that the proportion of seat belt use increased following the advertising campaign.

Problem 4.6.9. In Exercise 4.2.18 we found a confidence interval for the variance σ^2 using the variance S^2 of a random sample of size n arising from $N(\mu, \sigma^2)$, where the mean μ is unknown. In testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$, use the critical region defined by $(n-1)S^2/\sigma_0^2 \geq c$. That is, reject H_0 and accept H_1 if $S^2 \geq c\sigma_0^2/(n-1)$. If $n = 13$ and the significance level $\alpha = 0.025$, determine c.

Solution 4.6.9. Let the random variable W have a χ^2-distribution with $r = n - 1 = 12$ degrees of freedom. By Student's Theorem, $12S^2/\sigma_0^2$ has the same distribution, therefore c is determined by the property $P(W \geq c) = 0.025$, which yields $c \doteq 23.337$.

Problem 4.6.10. In Exercise 4.2.27, in finding a confidence interval for the ratio of the variances of two normal distributions, we used a statistic S_1^2/S_2^2, which has an F-distribution when those two variances are equal. If we denote that statistic by F, we can test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 > \sigma_2^2$ using the critical region $F \geq c$. If $n = 13$, $m = 11$, and $\alpha = 0.05$, find c.

Solution 4.6.10. By Student's Theorem, the numerator degrees of freedom is $r_1 = n-1 = 12$ and the denominator degrees of freedom is $r_2 = m-1 = 10$. The value of c is determined by the property $P(F \geq c) = 0.05$. Using the notation of Table V on page 662 we find that $c = F_{0.05}(12, 10)$, and using the program R we find $c \doteq 2.913$.