STA260: PROBABILITY AND STATISTICS II
SPRING 2021

TUTORIAL 2 (TUT9101)

MAY 7, 2021

YUEHUAN HE

UNIVERSITY OF TORONTO

$\leq$ $\sigma^2 < 0.04$

7.19 Ammeters produced by a manufacturer are marketed under the specification that the standard deviation of gauge readings is no larger than .2 amp. One of these ammeters was used to make ten independent readings on a test circuit with constant current. If the sample variance of these ten measurements is .065 and it is reasonable to assume that the readings are normally distributed, do the results suggest that the ammeter used does not meet the marketing specifications? [*Hint:* Find the approximate probability that the sample variance will exceed .065 if the true population variance is .04.]

$\sigma = 0.2$ $\quad Y_1 \cdots Y_{10} \sim N(\mu, \sigma^2)$.

$$P(S^2 \geq 0.065) = P\left(\boxed{\frac{(n-1)S^2}{\sigma^2}} \geq \frac{0.065 \times (n-1)}{\sigma^2}\right).$$

sample variance

$$= P\left(\frac{(n-1)S^2}{\sigma^2} \geq \boxed{\frac{9 \times 0.065}{0.04}}\right)$$

14.625

≈ 0.1. $\quad < 0.1$

Whiteboard Page

chi square-Distribution

$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$	df
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9

Let $Y_1, Y_2, \ldots, Y_5$ be a random sample of size 5 from a normal population with mean 0 and variance 1 and let $\overline{Y} = (1/5)\sum_{i=1}^{5} Y_i$. Let Y_6 be another independent observation from the same population. What is the distribution of

a $W = \sum_{i=1}^{5} Y_i^2$? Why? $\chi^2(5)$

b $U = \sum_{i=1}^{5} (Y_i - \overline{Y})^2$? Why? $\chi^2(4)$

c $\sum_{i=1}^{5} (Y_i - \overline{Y})^2 + Y_6^2$? Why? $\chi^2(5)$.

7.38 Suppose that $Y_1, Y_2, \ldots, Y_5, Y_6, \overline{Y}, W$, and U are as defined in Exercise 7.37. What is the distribution of

a $\sqrt{5}Y_6/\sqrt{W}$? Why?

b $2Y_6/\sqrt{U}$? Why?

c $2\left(5\overline{Y}^2 + Y_6^2\right)/U$? Why?

7.38 Suppose that $Y_1, Y_2, \ldots, Y_5, Y_6, \overline{Y}, W$, and U are as defined in Exercise 7.37. What is the distribution of

a $\sqrt{5}Y_6/\sqrt{W}$? Why?

b $2Y_6/\sqrt{U}$? Why?

c $2\left(5\overline{Y}^2 + Y_6^2\right)/U$? Why?

t-distribution

Definition 7.2.20 (Definition 7.2 (page number 360) of text [3]:) Let $Z \sim N(0,1)$ and let W be a χ^2-distributed variable with n df. If Z and W are independent, then

$$t_n = \frac{Z}{\sqrt{W/n}} \qquad \frac{Z}{\sqrt{\frac{W}{n}}}$$

follows a t-distribution with n degrees of freedom. We write $T = t_n$. The probability density function of T is

$$f(x) = \frac{\Gamma((n+1)/2))}{\sqrt{\pi n}\Gamma(n/2)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \quad \text{for} \quad -\infty < x < \infty$$

a. $\dfrac{\sqrt{5} \cdot Y_6}{\sqrt{W}}$ $\qquad W \sim \chi^2(5)$. $\qquad Y_6 \sim N(0,1)$.

$$= \frac{1 \cdot Y_6}{\frac{\sqrt{W}}{\sqrt{5}}} = \frac{Y_6}{\sqrt{\frac{W}{5}}} \sim t_5.$$

$U \sim \chi^2(4)$.

b. $$\frac{2 \cdot Y_6}{\sqrt{U}} = \frac{Y_6}{\frac{\sqrt{U}}{2}} = \frac{Y_6}{\sqrt{\frac{U}{4}}} \sim t_4.$$

$\bar{Y} = \frac{1}{5}(Y_1 + Y_2 + \cdots + Y_5)$

$Y_1 \cdots Y_6 \sim N(0,1)$

7.38 Suppose that $Y_1, Y_2, \ldots, Y_5, Y_6, \bar{Y}, W$, and U are as defined in Exercise 7.37. What is the distribution of

a $\sqrt{5}Y_6/\sqrt{W}$? Why?

b $2Y_6/\sqrt{U}$? Why?

c $2\left(5\bar{Y}^2 + Y_6^2\right)/U$? Why?

$(\sqrt{5}\cdot\bar{Y})^2$

$\chi^2(2)$

b. $\dfrac{2\cdot\boxed{((\sqrt{5}\cdot\bar{Y})^2 + Y_6^2)}}{U}$

$\chi^2(4)$

F-Distribution

Definition 7.2.24 (Definition 7.3 (page 362) of text[3]:) Let W_1 and W_2 be independent χ^2-distributed random variables with n_1 and n_2 df, respectively. Then

$$F_{n_1,n_2} = \frac{W_1/n_1}{W_2/n_2}$$

is said to have F-distribution with n_1 numerator degrees of freedom and n_2 denominator degrees of freedom. The probability density function of F-distribution is

$$f_{n_1,n_2}(x) = \begin{cases} \frac{\Gamma[(n_1+n_2)/2](n_1/n_2)^{n_1/2}}{\Gamma(n_1/2)\Gamma(n_2/2)} \frac{x^{(n_1/2)-1}}{(1+n_1x/n_2)^{(n_1+n_2)/2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

$\dfrac{W_1}{n_1} \Big/ \dfrac{W_2}{n_2}$

$$= \frac{\dfrac{(\sqrt{5}\bar{Y})^2 + Y_6^2}{2}}{\dfrac{U}{4}}$$

$(\sqrt{5}\bar{Y})^2 + Y_6^2 \sim \chi^2(2)$

$U \sim \chi^2(4)$.

$\sim F_{2,4}$

$\bar{Y} \sim N(0, \frac{1}{5})$.

$\sqrt{5}\bar{Y} \sim N(0,1)$

$\bar{Y}, Y_6$ independent

$Y \sim N(\mu, \sigma^2)$ $aY \sim N(a\mu, a^2\sigma^2)$

Problem 3

***7.39** Suppose that independent samples (of sizes n_i) are taken from each of k populations and that population i is normally distributed with mean μ_i and variance σ^2, $i = 1, 2, \ldots, k$. That is, all populations are normally distributed with the *same* variance but with (possibly) different means. Let $\overline{X}_i$ and S_i^2, $i = 1, 2, \ldots, k$ be the respective sample means and variances. Let $\theta = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k$, where $c_1, c_2, \ldots, c_k$ are given constants.

a Give the distribution of $\hat{\theta} = c_1\overline{X}_1 + c_2\overline{X}_2 + \cdots + c_k\overline{X}_k$. Provide reasons for any claims that you make.

b Give the distribution of

$$\frac{\text{SSE}}{\sigma^2}, \quad \text{where SSE} = \sum_{i=1}^{k}(n_i - 1)S_i^2.$$

Provide reasons for any claims that you make.

$U_1 . U_2$ indenpedent

$U_1 \sim N(\mu_1 . \sigma_1^2)$

$U_2 \sim N(\mu_2, \sigma_2^2)$.

$U_1 + U_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

$U_1 - U_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.

a. $\overline{X}_i \sim N(\mu_i, \frac{\sigma^2}{n_i})$.

$\overline{X}_1 . \overline{X}_2 \ldots \overline{X}_k$ independent.

$c_i \overline{X}_i \sim N(\boxed{c_i \mu_i}, \frac{c_i^2 \sigma^2}{n_i})$.

$\sigma^2 \left(\sum_{i=1}^{k} \frac{c_i^2}{n_i} \right)$.

$\hat{\theta} \sim N\left(\sum_{i=1}^{k} c_i \mu_i, \sum_{i=1}^{k} \frac{c_i^2 \sigma^2}{n_i}\right)$.

***7.39** Suppose that independent samples (of sizes n_i) are taken from each of k populations and that population i is normally distributed with mean μ_i and variance σ^2, $i = 1, 2, \ldots, k$. That is, all populations are normally distributed with the *same* variance but with (possibly) different means. Let $\overline{X}_i$ and S_i^2, $i = 1, 2, \ldots, k$ be the respective sample means and variances. Let $\theta = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k$, where $c_1, c_2, \ldots, c_k$ are given constants.

a Give the distribution of $\hat{\theta} = c_1\overline{X}_1 + c_2\overline{X}_2 + \cdots + c_k\overline{X}_k$. Provide reasons for any claims that you make.

b Give the distribution of

$$\frac{\text{SSE}}{\sigma^2}, \quad \text{where SSE} = \sum_{i=1}^{k}(n_i - 1)S_i^2.$$

Provide reasons for any claims that you make.

$\chi^2(n_i - 1)$.

b. $$\frac{SSE}{\sigma^2} = \frac{\sum_{i=1}^{k} (n_i - 1) S_i^2}{\sigma^2} = \sum_{i=1}^{k} \boxed{\frac{(n_i - 1) S_i^2}{\sigma^2}}$$

$\sim \chi^2(n_1 + n_2 + \cdots + n_k - k)$.

$\sum_{i=1}^{k} (n_i - 1)$

problem 7.2.1

$\theta = E(\hat\theta)$.

$\sqrt{V(\hat\theta)} = \sqrt{\left(\sum_{i=1}^{k} \frac{c_i^2}{n_i}\right)\sigma^2}$

Whiteboard Page

c Give the distribution of

$$\frac{\hat\theta - \theta}{\sqrt{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \cdots + \frac{c_k^2}{n_k}\right)\text{MSE}}}, \quad \text{where MSE} = \frac{\text{SSE}}{n_1 + n_2 + \cdots + n_k - k}.$$

Provide reasons for any claims that you make.

t-distribution

Definition 7.2.20 (Definition 7.2 (page number 360) of text [3]:) Let $Z \sim N(0,1)$ and let W be a χ^2-distributed variable with n df. If Z and W are independent, then

$$t_n = \frac{Z}{\sqrt{W/n}}$$

follows a t-distribution with n degrees of freedom. We write $T = t_n$. The probability density function of T is

$$f(x) = \frac{\Gamma((n+1)/2))}{\sqrt{\pi n}\,\Gamma(n/2)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \quad \text{for} \quad -\infty < x < \infty$$

$$= \frac{\dfrac{\hat\theta - \theta}{\sigma\sqrt{\sum_{i=1}^{k}\frac{c_i^2}{n_i}}}}{\dfrac{\sqrt{\text{MSE}}}{\sigma}}$$

$\dfrac{\hat\theta - \theta}{\sigma\sqrt{\sum_{i=1}^{k}\frac{c_i^2}{n_i}}} \sim N(0,1)$.

$t_{n_1+n_2+\cdots+n_k-k}$.

$\dfrac{\sqrt{\text{MSE}}}{\sigma} = \sqrt{\dfrac{\text{SSE}}{(n_1+n_2+\cdots+n_k-k)\sigma^2}} = \sqrt{\dfrac{\text{SSE}/\sigma^2}{(n_1+n_2+\cdots n_k-k)}}$

$\dfrac{\text{SSE}}{\sigma^2} \sim \chi^2(n_1+n_2+\cdots+n_k-k)$

$\sigma = 2.5$

7.43 An anthropologist wishes to estimate the average height of men for a certain race of people. If the population standard deviation is assumed to be 2.5 inches and if she randomly samples 100 men, find the probability that the difference between the sample mean and the true population mean will not exceed .5 inch.

$n = 100$

$\bar{Y}$

$$P(|\bar{Y} - \mu| \le 0.5)$$

$$= P\left(\boxed{\frac{|\bar{Y} - \mu|}{\frac{\sigma}{\sqrt{n}}}} \le \boxed{\frac{0.5}{\frac{\sigma}{\sqrt{n}}}}\right)$$

10

$$= P(|Z| \le 2) = 0.9544$$

The Central Limit Theorem

Theorem 7.2.32 (Theorem 7.4 (page 372) of [3]:) Let $Y_1, Y_2, \ldots, Y_n$ be independent and indentically distributed random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2 < \infty$. Define

$$U_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$$

That is,

$$\lim_{n\to\infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \qquad \text{for all } u$$

Remark 7.2.33 The Central Limit Theorem (CLT) states that the sample mean from any probability distribution (as long as mean and variance are finite) will have an approximate normal distribution, if the sample is sufficiently large. "Large n" means $n \ge 30$ in general, but in some cases may even be much less. The larger the sample size, the more nearly normally distributed is the population of all possible sample means. For fairly symmetric distributions, $n > 15$ will be sufficient.

UNIVERSITY OF TORONTO

7.45 Workers employed in a large service industry have an average wage of \$7.00 per hour with a standard deviation of \$.50. The industry has 64 workers of a certain ethnic group. These workers have an average wage of \$6.90 per hour. Is it reasonable to assume that the wage rate of the ethnic group is equivalent to that of a random sample of workers from those employed in the service industry? [*Hint:* Calculate the probability of obtaining a sample mean less than or equal to \$6.90 per hour.]

$n = 64$

$$P(\bar{Y} \leq 6.9) = P\left(\boxed{\frac{\bar{Y} - 7}{\frac{0.5}{\sqrt{64}}}} \leq \frac{6.9 - 7}{\frac{0.5}{\sqrt{64}}}\right)$$

$$= P(Z \leq -1.6)$$

$$= 0.0548$$

Δ

UNIVERSITY OF TORONTO

THANK YOU FOR ATTENDING TODAY’S TUTORIAL

Feedbacks: yuehuan.he@utoronto.ca