# NORMAL INVERSE FUNCTION IN TEACHING INFERENCE ABOUT POPULATION MEAN AND POPULATION PROPORTION

Xiaoyi Ji
Department of Mathematics
Utah Valley University, USA
xiaoyi.ji@uvu.edu

*If a simple random sample is drawn from a normal distribution or the sample size $n$ is large, the normal inverse function can be used to obtain a confidence interval of the population parameter $\mu$ or $p$. The traditional teaching method, finding a confidence interval indirectly through a standardized normal variable, is not necessary. The study shows that once the normal distributed random variables, $\tilde{X}$ or $\tilde{P}$, have been specified, the lower and upper quantiles of the distribution of the variable form an interval estimation of the population parameter $\mu$ or $p$, and the lower and upper quantiles can be directly computed by using normal inverse function in statistics or mathematics software packages. Two examples given in my teaching show, furthermore, the investigation of normal inverse function in teaching the inference about the population parameter provides a more quick, direct and natural manner to solve the problems.*

## INTRODUCTION

An algorithm for the inverse normal cumulative distribution function, called normal inverse function generally, widely exists in many computing software packages. However, it is limited to use for finding the critical value $z_{\alpha/2}$ or $z_\alpha$ when we teach inference about population mean and population proportion. The reason of the restriction is that if a problem is involved in a normal distribution, we usually work on the standardized normal distribution rather than directly solve the problem implicating the non-standard normal distribution itself. To compute a confidence interval of a mean $\mu$ or proportion $p$ of a population in introductory statistics textbooks, for instance, it is expected to find $z_{\alpha/2}$ at the first and then perform calculations $\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$ or $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ if the sample size $n$ is large enough. This paper investigates the method that the quantiles of the distribution of the specified random normal variable are used for inference about the population mean and population proportion, and the quantiles can be directly obtained by using the normal inverse function. In addition, the idea of quantiles of the distribution of the specified random normal variable can also apply to hypotheses tests to determine the rejection region. Furthermore, two examples in last section demonstrate that the normal inverse function and the normal distribution function can also apply to teach the problem of computing a probability of making a type II error. These methods had been used to my teaching in the courses of principles of statistics and mathematical statistics in two semesters. The investigation of normal inverse function on the inference about the population parameter provides a more quick, direct and natural manner to solve the problems in class teaching.

## NORMAL INVERSE FUNCTION IN TEACHING CONFIDENCE INTERVALS

The method: Using the quantiles of the distribution of a specified variable to construct a confidence interval and computing the quantiles directly by using the normal inverse function can be introduced to students after the following three definitions to be reviewed.

**Definition 1**. Let $X$ be a random variable. A $\alpha$ quantile of the distribution of a random variable $X$ is value $\xi_\alpha$ such that $P(X < \xi_\alpha) \le \alpha$ and $P(X \le \xi_\alpha) \ge \alpha$.

**Definition 2**. Let $X$ be a random normal variable with mean of $\mu$ and standard deviation of $\sigma$. The normal inverse function is defined in terms of the normal $c.d.f.$ as

$$x_\alpha = F^{-1}(\alpha;\ \mu, \sigma) = \{x: F(x;\ \mu, \sigma) = \alpha\}, \text{ where } \alpha = F(\alpha;\ \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt.$$

**Definition 3.** The normal inverse function in Definition 2 is defined to be in name *Norminv* or *NORM.INV* in statistics or mathematics computing software packages. That is, $x_\alpha = NORM.INV(\alpha, \mu, \sigma)$ in Excel.

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org 

With the assumption that a simple random sample is drawn from the normally distributed population or has enough large sample sizes throughout the study, normal inverse function in teaching confidence intervals in the different scenarios are summarized in the following cases along with the very simple proof in Case 2.1 to students.

**Case 2.1** Suppose that a simple random sample of size $n$ is drawn from a normal distribution with unknown mean of $\mu$ and standard deviation of $\sigma < \infty$, and $\bar{x}$ is the sample mean. Let $\tilde{X}$ be a random normal variable with mean of $\bar{x}$ and standard error of $\frac{\sigma}{\sqrt{n}}$ and $\tilde{x}_{\alpha/2}$ be the lower $\frac{\alpha}{2}$ quantile and $\tilde{x}_{1-\alpha/2}$ be the upper $\frac{\alpha}{2}$ quantile of the distribution of $\tilde{X}$. Then a $100(1-\alpha)\%$ confidence interval for $\mu$ is $(\tilde{x}_{\alpha/2}, \tilde{x}_{1-\alpha/2})$, and the lower bound $\tilde{x}_{\alpha/2}$ and the upper bound $\tilde{x}_{1-\alpha/2}$ can be directly computed by using the normal inverse function. That is, $\tilde{x}_{\alpha/2} = NORM.INV(\frac{\alpha}{2}, \bar{x}, \frac{\sigma}{\sqrt{n}})$ and $\tilde{x}_{1-\alpha/2} = NORM.INV(1-\frac{\alpha}{2}, \bar{x}, \frac{\sigma}{\sqrt{n}})$, where $NORM.INV$ is the inverse of the normal cumulative distribution in Excel in **Definition 3**.

Proof: Let a random variable $\tilde{X}$ follow the normal distribution with mean of $\bar{x}$ and standard deviation of $\sigma$. Then we have
$P\left(\tilde{X} < \bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = P\left(Z < -z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ and $P\left(\tilde{X} < \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = P\left(Z < z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$.

By definition 1, the equations in above show that the lower bound $\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ is $\frac{\alpha}{2}$ quantile and the upper bound $\bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ is $1-\frac{\alpha}{2}$ quantile of the random variable $\tilde{X}$, respectively.

Since the random variable $\tilde{X}$is normal distributed with mean of $\bar{x}$ and standard error of $\frac{\sigma}{\sqrt{n}}$, we can simply use the normal inverse function to compute the lower $\frac{\alpha}{2}$ quantile of $\tilde{x}_{\frac{\alpha}{2}}$ and the upper $1-\frac{\alpha}{2}$ quantile of $\tilde{x}_{1-\frac{\alpha}{2}}$. By definition 2 and definition 3, the lower and upper bounds of a $100(1-\alpha)\%$ confidence interval can be directly computed by using computing software packages. In Excel, for instance, $LB = \tilde{x}_{\alpha/2} = NORM.INV(\frac{\alpha}{2}, \bar{x}, \frac{\sigma}{\sqrt{n}})$ and $= \tilde{x}_{1-\alpha/2} = NORM.INV(1-\frac{\alpha}{2}, \bar{x}, \frac{\sigma}{\sqrt{n}})$.

**Case 2.2** Suppose that a large number of simple random samples of size $n$ is drawn from a population, and the population mean $\mu$ is unknown. Let $\bar{x}$ be the sample mean and $s$ be the sample standard deviation. Let $\tilde{X}$ be a random normal variable with mean of $\bar{x}$ and standard error of $\frac{s}{\sqrt{n}}$. Let $\tilde{x}_{\alpha/2}$ be the lower $\frac{\alpha}{2}$ quantile and $\tilde{x}_{1-\alpha/2}$ be the upper $\frac{\alpha}{2}$ quantile of the distribution of $\tilde{X}$. Then a $100(1-\alpha)\%$ confidence interval for μ is approximately given by $(\tilde{x}_{\alpha/2}, \tilde{x}_{1-\alpha/2})$, and the lower bound $\tilde{x}_{\alpha/2}$ and the upper bound $\tilde{x}_{1-\alpha/2}$ can be directly computed by using the normal inverse function. $LB = \tilde{x}_{\alpha/2} = NORM.INV(\frac{\alpha}{2}, \bar{x}, \frac{s}{\sqrt{n}})$ and $UB = \tilde{x}_{1-\alpha/2} = NORM.INV(1-\frac{\alpha}{2}, \bar{x}, \frac{s}{\sqrt{n}})$.

Similarly, to find a confidence interval for the population proportion, the specified random normal variables $\tilde{P}$ is introduced. The lower $\frac{\alpha}{2}$ quantile and upper $\frac{\alpha}{2}$ quantile of the distribution of the specified random normal variable $\tilde{P}$ form a $100(1-\alpha)\%$ confidence interval of the population proportion $p$. Therefore, the lower and upper $\frac{\alpha}{2}$ quantiles of the variable for a $100(1-\alpha)\%$ confidence interval of the population parameter $p$ can be directly computed by using the normal inverse function, which is described in the following cases.

**Case 2.3** Suppose that a large number of simple random samples of size $n$ is taken from a population with unknown population proportion of $p$, and $\hat{p}$ is the sample proportion. Let a random variable $\tilde{P}$ be normal distributed with mean of $\hat{p}$ and standard error of $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Let $\hat{p}_{\alpha/2}$ be the lower $\frac{\alpha}{2}$ quantile and $\tilde{p}_{1-\alpha/2}$ be the upper $\frac{\alpha}{2}$ quantile of the distribution of $\tilde{P}$. Then a $100(1-\alpha)\%$ confidence interval for $p$ is approximately given by $(\tilde{p}_{\alpha/2}, \tilde{p}_{1-\alpha/2})$, and the lower bound $\tilde{p}_{\alpha/2}$ and the upper bound $\tilde{p}_{1-\alpha/2}$ can be directly computed by using the normal inverse function, respectively. For instance, use NORM.INV function in Excel, $LB = \tilde{p}_{\alpha/2} = NORM.INV(\frac{\alpha}{2}, \hat{p}, \sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$ and $UB = \tilde{p}_{1-\frac{\alpha}{2}} = NORM.INV(1-\frac{\alpha}{2}, \hat{p}, \sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$.

**Case 2.4** Suppose that a large number of two simple random samples of sizes $n_1$ and $n_2$ are taken from the two populations, respectively. Suppose that the two population proportions of $p_1$ and $p_2$ are unknown, and $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions for two populations, respectively. Let a random variable $\widetilde{P_1 - P_2}$ be normal distributed with mean of $\hat{p}_1 - \hat{p}_2$ and standard error of $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$. Let $(\widetilde{p_1 - p_2})_{\alpha/2}$ be the lower $\frac{\alpha}{2}$ quantile and $(\widetilde{p_1 - p_2})_{1-\alpha/2}$ be the upper $\frac{\alpha}{2}$ quantile of the distribution of $\widetilde{P_1 - P_2}$. Then a $100(1-\alpha)\%$ confidence interval for the difference of the two population proportions $p_1 - p_2$ is approximately given by $((\widetilde{p_1 - p_2})_{\alpha/2}, (\widetilde{p_1 - p_2})_{1-\alpha/2})$, and the lower bound $(\widetilde{p_1 - p_2})_{\alpha/2}$ and the upper bound $(\widetilde{p_1 - p_2})_{1-\alpha/2}$ can be obtained by using the normal inverse function, respectively. That is,

$LB = (\widetilde{p_1 - p_2})_{\frac{\alpha}{2}} = NORM.INV\left(\frac{\alpha}{2}, \hat{p}_1 - \hat{p}_2, \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$, and

$UB = (\widetilde{p_1 - p_2})_{1-\frac{\alpha}{2}} = NORM.INV\left(1-\frac{\alpha}{2}, \hat{p}_1 - \hat{p}_2, \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$.

## NORMAL INVERSE FUNCTION IN TEACHING THE HYPOTHESES TESTS

Use technology of normal inverse function for determining the lower and upper bound of a confidence interval for the population parameter can also be expanded to the content in teaching the simple hypothesis test. Using the inverse function with respect to the specified variable to determine the rejection region is demonstrated in the following two types of cases.

**Case 3.1** Let a large number of simple random samples of size $n$ be drawn from a population with mean of $\mu$ and standard deviation unknown. Let $\bar{x}$ be the sample mean and $s$ be the sample standard deviation. Suppose we want to test the hypotheses as following at size $\alpha$, $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$. Then the critical values $\bar{x}_c$ for the rejection region are given by $\bar{x}_{c,lower} = \bar{x}_{\alpha/2} = NORM.INV\left(\frac{\alpha}{2}, \mu_0, \frac{s}{\sqrt{n}}\right)$, and $\bar{x}_{c,upper} = \bar{x}_{1-\alpha/2} = NORM.INV\left(1-\frac{\alpha}{2}, \mu_0, \frac{s}{\sqrt{n}}\right)$.

**Case 3.2** Suppose that a large number of simple random samples of size $n$ is taken from a population with the population proportion of $p$, and $\hat{p} = \frac{x}{n}$ is the sample proportion. Suppose we want to test the hypotheses as following at size $\alpha$, $H_0: p = p_0$ versus $H_1: p \neq p_0$. Then the critical values for the rejection region are given by $\hat{p}_{c,lower} = \hat{p}_{\alpha/2} = NORM.INV\left(\frac{\alpha}{2}, p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$, and $\hat{p}_{c,upper} = \hat{p}_{1-\alpha/2} = NORM.INV\left(1-\frac{\alpha}{2}, p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$.

## NORMAL INVERSE FUNCTION IN TEACHING COMPUTATION OF PROBABILITY OF TYPE II ERROR

When the probability of making type II error or the power of the test is taught in statistics textbooks, the z-score transformation is used twice. One, for instance, is to determine $\bar{x}_c$ by solving the equation such as $z_c = \frac{\bar{x}_c - \mu_0}{\sigma_{\bar{x}}}$ in testing the population mean, and another is to compute the probability of making a Type II error because the sampling distribution of $\bar{X}$ is normal but not standard normal in general. It is difficult for students to grasp the concepts and main tasks because of the tedious Fcomputation between two normal variables, standard normal $Z$ and sampling distribution $\bar{X}$ or $\hat{P}$. The examples given in the below provide the applications of the methods in Section 3. The process of problem solving with normal inverse function involved can reduce the tedious calculation works, and students can focus on the concepts of the rejection region and Type II error without getting distracted.

**Example 1** Let a simple random sample of large sample size of $n$ be drawn from a population with the population standard deviation of $\sigma < \infty$, and $\bar{x}$ be the sample mean and $s$ be the sample standard deviation. Suppose we want to test the hypotheses as following at size $\alpha$, $H_0: \mu = \mu_0$ versus $H_1: \mu < \mu_0$.

According to Case 3.1, the critical value for the rejection region with respect to $\bar{X}$ is $\bar{x}_\alpha = NORM.INV\left(\alpha, \mu_0, \frac{s}{\sqrt{n}}\right)$. If $\bar{x} < \bar{x}_\alpha$, then we reject $H_0$. So

$$P(Type\ II\ error) = P(\bar{X} > x_\alpha, \mu \neq \mu_0) = 1 - Norm.Dist(\bar{x}_\alpha, \mu, \frac{s}{\sqrt{n}}, true).$$

**Example 2** Suppose that a large number of simple random samples of size $n$ is taken from the population. Let $\hat{p} = \frac{x}{n}$ be the sample proportion, and $n\hat{p}(1-\hat{p}) \geq 10$ and $n \leq 0.05N$. Suppose we want to test the hypotheses as following at size $\alpha$, $H_0: p = p_0$ versus $H_1: p > p_0$.

According to the result from Case 3.3, the critical value for the rejection region with respect to $\hat{P}$ is $\hat{p}_{1-\alpha/2} = NORM.INV\left(1 - \frac{\alpha}{2}, p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$. If $\hat{p} > \hat{p}_{1-\alpha/2}$, then we reject $H_0$. So

$$P(Type\ II\ error) = P(\hat{P} < \hat{p}_{1-\alpha/2}, p \neq p_0) = Norm.Dist(\hat{p}_{1-\alpha/2}, p, \sqrt{\frac{p(1-p)}{n}}, true).$$

## CONCLUSION

The idea of using quantiles of the distribution of the specified random normal variable to construct a confidence interval enables students to use the currently existed computing technology, normal inverse function, to obtain an answer more quickly and less tediously, or directly implicating the non-standard normal distribution itself rather than a conversion to the standard normal. In addition, the inverse function of the specified random normal variable can also be extended to determine the critical values for hypotheses tests. Some processes and procedures are no longer necessary because technology has made them obsolete. Normal inverse function can also be used in teaching computation of probability of making type II error or the power of the test avoiding tedious computation. But above all, the method showed in the study gives a more direct and natural manner with the definitions involved only to solve the problems to students. It makes a great contribution in effectively teaching to instructors and productive learning to students in statistics courses.

## REFERENCES

Hogg, R. V., McKean, J. W., & Craig, A. (2013). *Introduction to mathematical statistics* (3rd ed.). New Jersey: Pearson.

Mathematician (2008). *Normal inverse cumulative distribution function.* www.mathworks.com/help/stats/NORM.INV.html

Moore, D. S. (2011). *Statistics: Concepts and controversies* (7th Ed.). New York: W. H. Freeman & Company.

Sullivan III, M. (2004). *Statistics: Informed decisions using data* (3rd Ed.). New Jersey: Pearson.

Tishkovskaya, S., & Lancaster, G. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education, 20*(2), 1-55.