

TEACHER'S CORNER

An Elementary Proof of Independence of Least Squares Estimators of Regression Coefficients and of Variance in Linear Regression

Alexander R. de Leon

University of Calgary

Joyce Raymund B. Punzalan

University of the Philippines Diliman

This pedagogical note extends to the linear regression context an elementary proof of independence of the sample mean and sample variance when sampling from a normal population. The proof is simple enough for classroom demonstration in an elementary statistical inference course as an interesting exercise in multivariable transformation; it can be used as well to highlight the link between classical random sample and non-iid (i.e., independent and identical distribution) regression settings.

Keywords: χ^2 distribution; Conditional density; Moment-generating function; Multivariable transformation; Normal distribution

1. Introduction

Consider the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where x_i 's are fixed constants and ε_i 's are independent normal random errors, each with mean zero and variance $\sigma^2 > 0$. It is well-known that the usual least squares estimators

$$\hat{\beta}_1 = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{S_{xx}} \right) Y_i = \sum_{i=1}^n a_{1i} Y_i, \quad (2)$$

$$\hat{\beta}_0 = \sum_{i=1}^n \left(\frac{1}{n} - a_{1i}\bar{x} \right) Y_i = \sum_{i=1}^n a_{0i} Y_i, \quad (3)$$

with $\bar{x} = \sum_{i=1}^n x_i / n$ and $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$, are jointly independent of $(n-2)\hat{\sigma}^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$ and that $(n-2)\hat{\sigma}^2 / \sigma^2 \sim \chi_{n-2}^2$ where χ_v^2 is the χ^2 -distribution with v degrees of freedom. This independence is needed, for example, so that the pivotal quantity $\sqrt{S_{xx}}(\hat{\beta}_1 - \beta_1) / \hat{\sigma}$ for constructing a confidence interval for β_1 has a t -distribution with $n-2$ degrees of freedom (Montgomery et al., 2006: 23).

While usually available in most elementary textbooks in mathematical statistics, the proof of this important result, if not long and tedious, usually requires additional results not typically covered in an introductory course on statistical inference. Hogg et al. (2004: 493; see also Montgomery et al., 2006: 555), for example, proves this result using a version of Cochran's Theorem on independence of quadratic forms. Casella and Berger (2002: 569) makes use of a special result on independence of linear combinations of normal random variables. Mood et al. (1974: 488) relies on the moment-generating function (mgf) technique, which entails evaluating a complicated integral.

In this note, we adapt Schuster's (1973) proof of independence of the sample mean and sample variance when sampling from a normal distribution, to the simple linear regression model (1) to prove that $\hat{\beta}_0$ and $\hat{\beta}_1$ are jointly independent of $\hat{\sigma}^2$. The proof is simple enough to be used in an elementary statistical inference course since it does not require any special results on quadratic forms and normal random variables nor entail evaluating a complicated integral. The proof is also likely to prove instructive in a classroom setting, as it involves an interesting exercise in multi-variable transformation quite unlike those usually found in problems discussed in the classroom. First, the transformation entails working with conditional densities. Second, a common integration technique in probability which is not usually taught in calculus classes, is given prominence. Last, independence is argued using the conditional mgf, a useful result that is normally not emphasized in mathematical statistics classes.

2. Proof of Independence

We derive the conditional mgf of $\hat{\theta} = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 / \sigma^2$, given $\hat{\beta}_0 = b_0$ and $\hat{\beta}_1 = b_1$. To do this, observe that by independence we get the following conditional density of Y_1 and Y_2 , given $Y_3 = y_3, \dots, Y_n = y_n$:

$$f_{Y_1, Y_2 | Y_3, \dots, Y_n}(y_1, y_2 | y_3, \dots, y_n) = f_{Y_1}(y_1) f_{Y_2}(y_2), \quad (4)$$

where $f_{Y_i}(\cdot)$ is the density of Y_i . In (4), make the bivariate transformation (Casella and Berger, 2002: 185)

$$\left. \begin{aligned} \hat{\beta}_0 &= a_{01}Y_1 + a_{02}Y_2 + \sum_{i=3}^n a_{0i}y_i \\ \hat{\beta}_1 &= a_{11}Y_1 + a_{12}Y_2 + \sum_{i=3}^n a_{1i}y_i \end{aligned} \right\} \Leftrightarrow \begin{cases} Y_1 = h_1(\hat{\beta}_0, \hat{\beta}_1) \\ Y_2 = h_2(\hat{\beta}_0, \hat{\beta}_1) \end{cases} \quad (5)$$

and let J be its Jacobian. Note that it is not necessary to know the actual forms of inverses $h_1(\cdot)$ and $h_2(\cdot)$, and hence, of J ; note also that while we arbitrarily chose Y_1 and Y_2 for convenience, any pair from Y_1, \dots, Y_n would work as well. The conditional joint density of $\hat{\beta}_0$ and $\hat{\beta}_1$, given $Y_3 = y_3, \dots, Y_n = y_n$, is then

$$f_{\hat{\beta}_0, \hat{\beta}_1 | Y_3, \dots, Y_n}(b_0, b_1 | y_3, \dots, y_n) = |J| f_{Y_1}(h_1(b_0, b_1)) f_{Y_2}(h_2(b_0, b_1)) \quad (6)$$

The conditional joint density of Y_3, \dots, Y_n , given $\hat{\beta}_0 = b_0$ and $\hat{\beta}_1 = b_1$, follows as

$$f_{Y_3, \dots, Y_n | \hat{\beta}_0, \hat{\beta}_1}(y_3, \dots, y_n | b_0, b_1) = \frac{f_{\hat{\beta}_0, \hat{\beta}_1 | Y_3, \dots, Y_n}(b_0, b_1 | y_3, \dots, y_n) \prod_{i=3}^n f_{Y_i}(y_i)}{f_{\hat{\beta}_0, \hat{\beta}_1}(b_0, b_1)} \quad (7)$$

where $f_{\hat{\beta}_0, \hat{\beta}_1}(\cdot)$ is the joint density of $\hat{\beta}_0$ and $\hat{\beta}_1$. Since $\hat{\beta}_0$ and $\hat{\beta}_1$ are jointly bivariate normal with respective means β_0 and β_1 , respective variances $\sigma^2 \sum_{i=1}^n x_i^2 / (nS_{xx})$ and σ^2 / S_{xx} , and covariance $-\sigma^2 \bar{x} / S_{xx}$, using the fact that $f_{Y_i}(\cdot)$ is the normal $(\beta_0 + \beta_1 x_i, \sigma^2)$ density, we get

$$f_{Y_3, \dots, Y_n | \hat{\beta}_0, \hat{\beta}_1}(y_3, \dots, y_n | b_0, b_1) = |J| \frac{\sqrt{nS_{xx}}}{(2\pi\sigma^2)^{(n-2)/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2\right\} \quad (8)$$

where $y_1 = h_1(b_0, b_1)$ and $y_2 = h_2(b_0, b_1)$. This is a proper density defined over the $(n-2)$ -dimensional real plane for any $\sigma^2 > 0$, and hence, is equal to 1 when integrated over this space. Because $\hat{\theta}$ depends only on Y_3, \dots, Y_n , given $\hat{\beta}_0 = b_0$ and $\hat{\beta}_1 = b_1$, the required conditional mgf of $\hat{\theta}$ is obtained from (8). This mgf is thus

$$E\left(e^{t\hat{\theta}} \mid \hat{\beta}_0 = b_0, \hat{\beta}_1 = b_1\right) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |J| \sqrt{nS_{xx}} (2\pi\sigma^2)^{-(n-2)/2} \\ \times \exp\left\{-\frac{1}{2\left(\frac{\sigma^2}{1-2t}\right)} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2\right\} dy_3, \dots, dy_n. \quad (9)$$

The integral in (9) is easily evaluated using an integration technique we call “garbage in-garbage out,” which entails simply making the integrand in (9) the density in (8) by the inclusion of the unit constant $(1-2t)^{-(n-2)/2} \times (1-2t)^{(n-2)/2}$. This is done by taking $\sigma^2 > 0$ in (8) as $\sigma^2/(1-2t) > 0$ with $t < 1/2$, which then simplifies (9) as

$$E\left(e^{t\hat{\theta}} \mid \hat{\beta}_0 = b_0, \hat{\beta}_1 = b_1\right) = (1-2t)^{-(n-2)/2} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |J| \sqrt{nS_{xx}} \left(\frac{2\pi\sigma^2}{1-2t}\right)^{-(n-2)/2} \\ \times \exp\left\{-\frac{1}{2\left(\frac{\sigma^2}{1-2t}\right)} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2\right\} dy_3, \dots, dy_n \quad (10) \\ = (1-2t)^{-(n-2)/2}.$$

Note that this is the mgf of χ_{n-2}^2 , and is free of b_0 and b_1 . This implies that it is also the unconditional mgf of $\hat{\theta}$, and hence, it follows that $\hat{\beta}_0$ and $\hat{\beta}_1$ are jointly independent of $\hat{\sigma}^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 / (n-2)$, and $\hat{\theta} = (n-2)\hat{\sigma}^2 / \sigma^2 \sim \chi_{n-2}^2$. Note that this result specializes to that in Shuster (1973) in the iid case with $\beta_1=0$.

3. Conclusion

The above proof has an interesting pedagogical use. It can serve as a useful teaching device for highlighting the relationship between classical inference based on random samples and the non-iid regression setting. As the basic building block of statistical inference, a random sample comprises independent random variables Y_1, \dots, Y_n having a common distribution. One usual extension of this notion is the linear regression setting, which relaxes the assumption of identical

distributions and allows for means $E(Y_i)$ to vary in a linear fashion as a function of a fixed explanatory variable x_i . In (1), Y_1, \dots, Y_n are still independent but not identically distributed, since $Y_1 \sim \text{normal}(\beta_0 + \beta_1 x_1, \sigma^2), \dots, Y_n \sim \text{normal}(\beta_0 + \beta_1 x_n, \sigma^2)$. It is then to be expected that mean estimators $\hat{\beta}_0 + \hat{\beta}_1 x_i$ in this case would be independent of variance estimator $\hat{\sigma}^2$, in light of their independence in the iid case. By extending Schuster's (1973) proof to the regression setting, beginning students are able to clearly see this connection. They are then better able to understand how the simplest statistical models are made more complex and to see how certain properties of simple models are preserved in more complicated ones.

While it is easy to extend our proof to the multiple linear regression case—a topic discussed in introductory regression analysis courses—it may not be as pedagogically attractive since, given students' backgrounds, arguably more straightforward, albeit more advanced, approaches exist (e.g., via matrices and properties of multivariate normal distributions); see Montgomery et al. (2006).

REFERENCES

- CASELLA, G. and R.L. BERGER, 2002, *Statistical Inference*, 2nd ed., Duxbury, New York.
- HOGG, R.V., J.W. MCKEAN, and A. CRAIG, 2004, *Introduction to Mathematical Statistics* 5th ed., Macmillan, New York.
- MONTGOMERY, D.C., E.A. PECK, and G.G. VINING, 2006, *Introduction to Linear Regression Analysis*, 4th ed., Wiley.
- MOOD, A.M., F.A. GRAYBILL, and D.C. BOES, 1974, *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.
- SHUSTER, J., 1973, A simple method of teaching the independence of \bar{X} and s^2 , *The American Statistician*, 27: 29-30.

ACKNOWLEDGMENT

This work was partially supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

