Overview

Stephanie Ranft S2459825

January 7, 2021

Statistics 2
PSBE2-07

Exercises

Fisher Z Transformation

1. A regression model $Y_j = \beta_0 + \beta_1 X_j + \varepsilon_j$ has $n = 11$ observations. The sample correlation between X and Y is 0.60. We test the null hypothesis $H_0 : \rho = 0$ (the true correlation between the X and Y variables is zero).

   (a) What is the t-value to test the null hypothesis?

   (b) What is the p-value to test the null hypothesis? Summarise your results of the test.

   (c) What can you say about the results of the test with respect to the sample correlation coefficient?

2. A linear regression with 11 data points has an estimated β_1 of 4.5 and a sample correlation between the X and Y values of 0.60.

   (a) What is the t-value to test the null hypothesis that the correlation ρ is zero? Summarise your results of the test.

   (b) What is the t-value to test the null hypothesis that β_1 is zero? Summarise your results of the test.

   (c) What is the standard error of the estimate of β_1? What does this tell you about the reliability of the t-test you performed, and how might you improve the test?

   (d) How are these two tests similar/different?

3. X and Y are a bivariate normal distribution from which a sample of 40 observations is taken. The sample correlation between X and Y is 0.833. We test the null hypothesis $H_0 : \rho = 0.750$. The alternative hypothesis is $H_a : \rho_0 > 0.750$.

   (a) What is the Fisher transform r_z of the random variable r of the correlation r between X and Y?

   (b) What is the Fisher transform of the observed correlation?

   (c) What is the distribution of r_z?

   (d) What is the Fisher transform of the correlation ρ_0 assumed in the null hypothesis?

   (e) What is the z-value to test this null hypothesis?

   (f) What is the p-value for this test of the null hypothesis?

   (g) What is the 95% confidence interval for the true value of the Fisher transform of the correlation?

   (h) What is the 95% confidence interval for the true value of the correlation?

   (i) Summarise your results of the previous parts.

4. Using the data set Album Sales (JASP>data library>regression>album sales), we have the following outputs:

Table 1: Descriptive Statistics

	sales	adverts
Valid	200	200
Missing	0	0
Mean	193.200	614.412
Std. Deviation	80.699	485.655
Minimum	10.000	9.104
Maximum	360.000	2271.860

Table 2: Model Summary

Model	R	R^2	Adjusted R^2	RMSE
1	0.578	0.335		65.991

Table 3: Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
1	(Intercept)	134.140	7.537			< .001
	adverts	0.096	0.010			< .001

(a) Fill in the missing values (highlighted lilac). Explain the steps you take to calculate the appropriate values.

(b) Perform a hypothesis test for correlation (there are two ways of doing this). State the null and alternative hypotheses, compute your statistic and summarise your findings.

(c) Compute the confidence interval for the correlation. Explain all intermediary steps.

Solutions

1. Recall:
$$t = \frac{\text{estimate} - \text{hypothesised value}}{\text{standard error of the estimate}}.$$
For this question, we have the hypothesised value $\rho_0 = 0$, estimate $r = 0.60$, and the standard error of r is $se_r = \sqrt{\frac{1-r^2}{n-2}}$. Assume that the alternative hypothesis is $H_A : \rho > 0$.
$$\implies t = \frac{(r - \rho_0)\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.60\sqrt{9}}{\sqrt{0.64}} = \frac{0.6 \cdot 3}{0.8} = 2.25 \sim t_9 = 2.262 \implies p = \mathbb{P}(t_9 > 2.262) = 0.025502.$$
We reject the null hypothesis at a 5% level, i.e. the population correlation coefficient between variables X and Y is significantly greater than zero, and they are dependent.

2. The t-statistic for $H_0 : \beta_1 = 0$ is identical to the t-statistic for $H_0 : \rho = 0$; this question was meant to trick you. Sorry, not sorry ☺

   You can use the answer to the previous question to determine the standard error for β_1:
$$t = \frac{b_1 - \beta_1}{se_{\beta_1}} \implies 2.25 = \frac{4.5 - 0}{se_{\beta_1}} \implies se_{\beta_1} = \frac{4.5}{2.25} = 2.$$
We know that the standard error is always decreased by increasing our sample size.

3.
$$r_z = \frac{\ln\left(\frac{1+r}{1-r}\right)}{2} = \frac{\ln\left(\frac{1.833}{0.167}\right)}{2} = 1.198 \quad (1)$$
$$r_z \sim \mathcal{N}\left(\frac{\ln\left(\frac{1+\rho}{1-\rho}\right)}{2}, \frac{1}{n-3}\right) = \mathcal{N}\left(\frac{\ln\left(\frac{1.750}{0.250}\right)}{2}, \frac{1}{40-3}\right) = \mathcal{N}(0.973, 0.027) \quad (2)$$
$$\rho_{0z} = \frac{\ln\left(\frac{1.750}{0.250}\right)}{2} = 0.973 \quad (3)$$
$$z = (r_z - \rho_{0z})\sqrt{n-3} = (1.198 - 0.973)\sqrt{37} = 1.368 \sim \mathcal{N}(0,1) \quad (4)$$
$$p = \mathbb{P}(z > 1.368) = 0.085656 \quad (5)$$
$$\text{CI for } \rho_z\text{: } r_z \pm \frac{z^*}{\sqrt{n-3}} = 1.198 \pm \frac{1.645}{\sqrt{37}} = (0.928, 1.468) \quad (6)$$
$$\text{CI for } \rho\text{: } \left(\frac{e^{2\cdot \mathrm{LB}_{\rho_z}} - 1}{e^{2\cdot \mathrm{LB}_{\rho_z}} + 1}, \frac{e^{2\cdot \mathrm{UB}_{\rho_z}} - 1}{e^{2\cdot \mathrm{UB}_{\rho_z}} + 1}\right) = \left(\frac{e^{2\cdot 0.928} - 1}{e^{2\cdot 0.928} + 1}, \frac{e^{2\cdot 1.468} - 1}{e^{2\cdot 1.468} + 1}\right) = (0.730, 0.899) \quad (7)$$
We do not reject the null hypothesis at the 5% level as $p > 0.05$ and 0.75 is contained in the 90% confidence interval for ρ.

4. Stein's adjusted R^2
$$\bar{R}_S^2 = 1 - \left(\frac{n-1}{n-p-1}\right)\left(\frac{n-2}{n-p-2}\right)\left(\frac{n+1}{n}\right)\left[1 - R^2\right] \quad (8a)$$
Wherry's adjusted R^2
$$\bar{R}_W^2 = 1 - \left(\frac{n-1}{n-p-1}\right)\left[1 - R^2\right] \quad (8b)$$
$$\bar{R}_S^2 = 1 - \left(\frac{199}{198}\right)\left(\frac{198}{197}\right)\left(\frac{201}{200}\right)\left[1 - 0.578^2\right] = 0.324 \quad (9)$$
$$\bar{R}_W^2 = 1 - \left(\frac{199}{198}\right)\left[1 - 0.578^2\right] = 0.331 \quad (10)$$
$$t_{b_0} = \frac{143.14}{7.537} = 18.99; \qquad t_{b_1} = \frac{0.096}{0.010} = 9.6. \quad (11)$$
The standardised slope coefficient in simple linear regression is equal to $r = 0.578$. The p-value for the slope coefficient indicates that the variables are significantly correlated, i.e. $H_0 : \rho = 0$ is rejected as this is equal to $H_0 : \beta_1 = 0$. The 95% CI for ρ (two-sided alternative hypothesis) is $(0.478, 0.664)$.