Barcelona School of Economics

Assignment 4

Foundations of Econometrics
Group 7

al·la Gelpi, Mariajosé Argote, Samuel Fraley

October 23, 2025

Question 1

Class slide 3(33) (Unit 3) illustrated, via simulation, the effects of collinearity. The script used to generate the sample is included in file `data33.R`.

Part (a)

(i) Estimate the regression model included in the slide, presenting OLS estimates and the 95% confidence intervals for each parameter; Include the output in your answer.

Answer:

Table 1: Regression Results with 95% CI

	Estimate	Std..Error	Lower	Upper
(Intercept)	7.9365	1.6426	4.5863	11.2867
x2	0.5953	0.0747	0.4429	0.7476
x3	0.2996	0.5264	-0.7740	1.3733
x4	0.7818	0.5277	-0.2945	1.8582

(ii) Using `confidenceEllipse()` function, or equivalent, draw the 95% confidence region for parameters β_3, β_4. Include also in the drawing the confidence intervals for each parameter.

Answer:

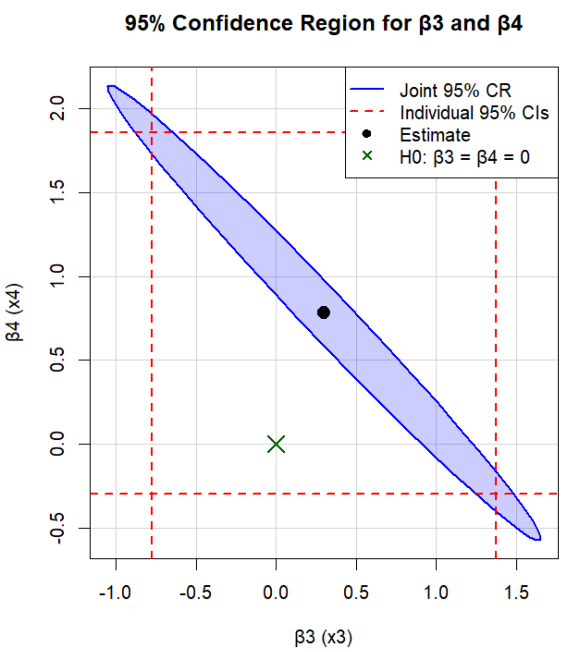

Figure 1: 95% Confidence Region for β_3 and β_4

(iii) Describe what the confidence region you just drew provides.

Answer:

The confidence region we drew illustrates the joint confiendence region, while the red lines show the individual confidence intervals. Its represent all the pair of values $\{x_i, y_i\}$ such that the joint test $H_0 : \beta_3 = x_i,\ \beta_4 = y_i;\quad H_1 : \text{No } H_0$, won't be rejected at a confidence level 95%. It also illsutrates the correlation between the two parameters, as the confidence region is elongated along a diagonal line. The negative slope provides insight into the negative correlation between the two parameters.

(iv) Use the figure of confidence intervals and confidence region to show the difference between testing statistical significance of regressors separately or jointly, and explain why this is so relevant under the presence of collinear regressors.

Answer:

The figure illustrates that joint testing can show us when collinear variables are jointly significant, even if they are not individually significant.

Individual tests: Testing $H_0 : \beta_3 = 0$ and $H_0 : \beta_4 = 0$ separately uses the rectangular region formed by the individual 95% confidence intervals (red dashed lines). From the regression output, both intervals include zero: x_3 CI $= [-0.7740, 1.3733]$ and x_4 CI $= [-0.2945, 1.8582]$. Since the rectangle contains the origin $(0, 0)$, we fail to reject both null hypotheses individually—neither x_3 nor x_4 appears statistically significant.

Joint test: Testing $H_0 : \beta_3 = \beta_4 = 0$ jointly uses the 95% confidence ellipse. The ellipse is much smaller than the rectangle due to the negative correlation between $\hat{\beta}_3$ and $\hat{\beta}_4$. If the origin $(0, 0)$ falls outside the ellipse, we reject the joint null hypothesis, meaning x_3 and x_4 are jointly significant even though neither is individually significant.

Collinearity: When regressors are collinear, their coefficients are negatively correlated because if one increases, the other must decrease to fit the same data. This creates the tilted ellipse. The individual tests ignore this correlation and use the wider rectangular region, making them overly conservative. The joint test correctly accounts for correlation, revealing that while we cannot precisely determine which variable drives the effect, we can confidently say that together they have a significant impact on y. This demonstrates why collinearity makes individual t-tests unreliable while joint F-tests remain valid.

Part (b)

Modify the script used to estimate now the same regression with data generated from the same dgp but now using a sample of 3500 observations.

(i) Surprised with how the estimates have changed? Rigorously justify.

Answer:

Table 2: Regression Results with 95% CI and n= 3500

	Estimate	Std..Error	Lower	Upper
(Intercept)	10.2294	0.1811	9.8744	10.5845
x2	0.4902	0.0078	0.4750	0.5054
x3	0.5512	0.0685	0.4169	0.6855
x4	0.4475	0.0680	0.3143	0.5808

Not surprised. The estimates have converged closer to the true parameter values. For $n = 35$, the estimates were $\hat{\beta}_3 = 0.2996$ and $\hat{\beta}_4 = 0.7818$, showing considerable sampling variability. With $n = 3500$, the estimates are much closer to 0.5 for both parameters, so now the estimates are more precise.

By the Law of Large Numbers, as sample size increases, the OLS estimators converge in probability to their true values. This is consistency: $\text{plim}(\hat{\beta}_j) = \beta_j$ as $n \to \infty$. The larger sample provides more information about the true relationship, reducing the influence of random sampling variation. The collinearity between x_3 and x_4 still exists (since $x_4 = x_3 + \text{noise}$ in both samples), but with more observations, the estimator can better distinguish their individual effects on y.

(ii) Surprised of the change of the 95% confidence intervals? Rigorously justify.

Answer:

Not surprised. As in (i) having more information about the population let you make better prediction because there's less variance, so less $se(\hat{\beta}_k)$. That's why the intervals in this case are smaller.

(iii) Surprised of the change of the 95% confidence region for parameters β_3, β_4? Rigorously justify.

Answer:

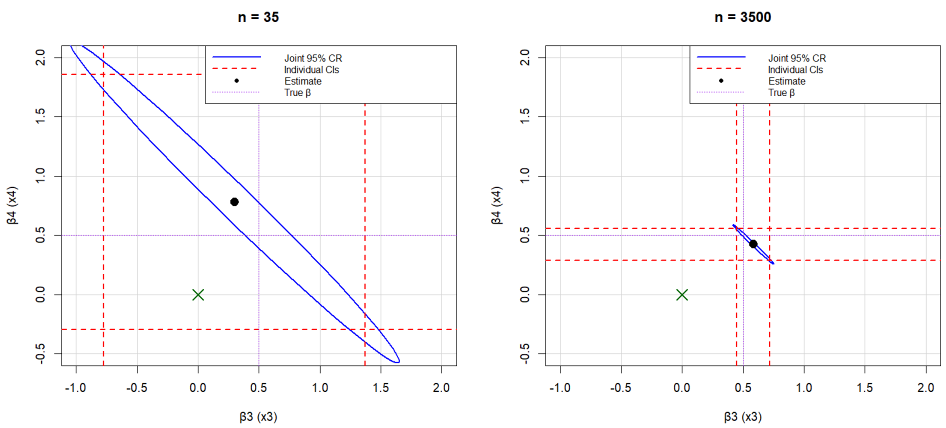

Figure 2: Original and Modified Script Results

We see that the ellipse it has become shorter. That's not surprising, because this ellipse is based in the variance-covariance matrix of x_3 and x_4, and as the variance and covariance decrease when the size of the samples grow, the entire ellipse becomes smaller.

(iv) Using the variance decomposition expression for $\text{var}(\hat{\beta}_3 \mid X)$, or $\text{var}(\hat{\beta}_4 \mid X)$, discuss why increasing n can explain the changes observed. Be specific.

Answer:

The variance decomposition is:

$$var(\hat{\beta}_k \mid X) = \sigma^2 (X'X)^{-1}_{kk} = \sigma^2 \tfrac{1}{SST_k} \tfrac{1}{(1-R_k^2)}.$$

Where $SST_k = \sum_{i=0}^{n} (x_{ik} - \bar{x}_k)^2$.

If n increases then SST_k also increases because it is the sum of the square of the errors of all the observations. The fact that SST_k increases means that $\frac{1}{SST_k}$ decreases and so the $var(\hat{\beta}_k \mid X)$.

Because the $se(\hat{\beta}_k) = \sqrt{var(\hat{\beta}_k \mid X)}$, $se(\hat{\beta}_k)$ will also decrease, so the intervalse, defined as:

$$[\hat{\beta}_k - t_{2.5\%(n-K)} se(\hat{\beta}_k), \hat{\beta}_k + t_{2.5\%(n-K)} se(\hat{\beta}_k)]$$

Will become shorter.

Also, as the ellipse is $z'Az$ with:

$$A = [\sigma^2 R(X'X)^{-1} R^{-1}]^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Where $\sigma^2 R(X'X)^{-1} R^{-1}]^{-1}$ is the variance-covariance matrix of β_3 and β_4. We know that as bigger a_{11} the horizontal axis is shorter, and as bigger a_{22} the vertical axis is shorter. Because A is the inverse matrix of the variance-covariance matrix of β_3 and β_4, if $var(\hat{\beta}_3 \mid X)$ decrease a_{11} increase so x-axis become shorter. By the same reasoning if $var(\hat{\beta}_4 \mid X)$ decrease a_{22} increase so y-axis of the ellipse become shorter.

Part (c)

Now, go back to the original script generating 35 observations, and modify the script so that now $x_{i3} + 2x_{i4} = 0$.

(i) Run the script again. Include the output in your answer.

Answer:

Table 3: Regression Results with 95% CI

	Estimate	Std..Error	Lower	Upper
(Intercept)	10.1838	2.2007	5.7012	14.6665
x2	0.4697	0.0999	0.2661	0.6733
x3	0.1985	0.1015	-0.0083	0.4052
x4	NA	NA	NA	NA

(ii) How many estimates did you get an estimate for β_3? And for β_4? You should be able to show, using the proper derivation, that in fact you got an infinite number of estimates for β_3 and β_4.

Answer:

What happens here is that there are 2 regressors with perfect collinearity, and so $det(X'X) = 0$ and $(X'X)^{-1}$ doesn't exist. That makes that we can't compute $\hat{\beta}_k$, and thats why R put a NA in x_4 so we can compute the others.

R gives us one estimate for $\hat{\beta}_3$ and none for $\hat{\beta}_4$, but really, as $x_{i3} = -2x_{i4}$ we have the regression:

$$
\begin{aligned}
y_i &= \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} - \beta_4 0.5 x_{i3} + \varepsilon_i \\
&= \beta_1 + \beta_2 x_{i2} + (\beta_3 - 0.5\beta_4) x_{i3} + \varepsilon_i
\end{aligned}
$$

So we have a regression with three independent regressors, and so we can find a unique estimate $\hat{\beta}'$ of $\beta' = \beta_3 - 0.5\beta_4$, thus $\hat{\beta}' = \hat{\beta}_3 - 0.5\hat{\beta}_4$ so there are infinite values of $\hat{\beta}_3$, $\hat{\beta}_4$ such that the equallity holds.

Question 2

Data file `microsoft.csv` includes monthly data from May 1986 to April 2013 on RP_{msft} (excess return of Microsoft stock), $RP_{s\&p}$ (excess return on the S&P500 portfolio), $Dprod$ (variation of Industrial production), $Dinflation$ (change in inflation rate), $Dterm$ (change in interest rate) and $m1$ (an indicator variable that takes value 1 if t is the month of January and 0 otherwise). The following regression is set to measure the reaction of the excess return of Microsoft stocks to changes in macroeconomic variables:

$$RP_{msft,t} = \beta_1 + \beta_2 RP_{s\&p,t} + \beta_3 Dprod_t + \beta_4 Dinflation_t + \beta_5 Dterm_t + \beta_6 m1_t + \epsilon_t$$

(a) Estimate the model above by OLS. Present the complete output (estimates, standard errors, p-values) as your answer.

Answer:

Dep. Variable:	RPmsoft	**R-squared:**	0.213
Model:	OLS	**Adj. R-squared:**	0.201
Method:	Least Squares	**F-statistic:**	17.26
Date:	Wed, 22 Oct 2025	**Prob (F-statistic):**	4.08e-15
Time:	12:54:20	**Log-Likelihood:**	-1276.7
No. Observations:	324	**AIC:**	2565.
Df Residuals:	318	**BIC:**	2588.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-0.9291	0.760	-1.223	0.222	-2.424	0.566
RPsandp	1.3232	0.152	8.678	0.000	1.023	1.623
Dprod	-1.5216	1.283	-1.186	0.237	-4.046	1.003
Dinflation	0.4716	2.351	0.201	0.841	-4.154	5.097
Dterm	4.1587	2.487	1.672	0.095	-0.735	9.052
m1	5.4352	2.869	1.894	0.059	-0.210	11.081

Omnibus:	203.965	**Durbin-Watson:**	2.141
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1809.211
Skew:	-2.541	**Prob(JB):**	0.00
Kurtosis:	13.401	**Cond. No.**	21.2

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

(b) The January effect states that on average, every else equal, the returns (or excess returns) are larger in the month of January than the rest of the months. Test, at

$\alpha = 1\%$, the presence of the January effect using the exact t-test statistic. Would you say the data supports the presence of this effect?

Answer:

We test for the presence of the January effect by examining whether the coefficient on the January dummy variable (m_1) is statistically significant.

Hypotheses:

$$H_0 : \beta_6 = 0 \quad \text{(No January effect)}$$
$$H_1 : \beta_6 \neq 0 \quad \text{(January effect exists)}$$

We use a two-tailed t-test at significance level $\alpha = 0.01$.

Test Statistic:

$$t = \frac{\hat{\beta}_6}{\text{SE}(\hat{\beta}_6)} = \frac{5.44}{2.87} = 1.89$$

With 318 degrees of freedom, the critical values are ± 2.59. So, reject H_0 if $|t| > 2.59$ or equivalently if $p < 0.01$. The t-statistic is 1.89 with a p-value of 0.059. Since $|1.89| < 2.59$ and $p = 0.059 > 0.01$, we fail to reject the null hypothesis. Therefore, we do not have sufficient evidence at the 1% significance level to conclude that there is a January effect in Microsoft stock returns.

(c) Aside from normality, list the assumptions needed to justify the use of the t test statistic. Justify your answer.

Answer:

To justify the use of the exact t-test statistic (aside from normality), the following assumptions from the Classical Linear Regression Model are required:

(a) **Linearity:** The model is correctly specified as linear in parameters:

$$y = X\beta + \epsilon$$

(b) **Strict Exogeneity:** $\mathbb{E}[\epsilon|X] = 0$. The error terms have zero conditional mean given all regressors.

(c) **No Perfect Multicollinearity:** The matrix X has full column rank, so $(X'X)$ is invertible and $\hat{\beta} = (X'X)^{-1}X'y$ exists.

(d) **Homoskedasticity:** $\text{Var}(\epsilon_i|X) = \sigma^2$ for all i. The variance of errors is constant across observations.

(e) **No Autocorrelation:** $\text{Cov}(\epsilon_i, \epsilon_j|X) = 0$ for all $i \neq j$. Error terms are uncorrelated.

Justification:

To see that the t-statistic is a good statistic we need to define it's distribution under the null, and be able to observe a value of this statistic given a sample. The first part is in which we will need the assumptions:

(a) **Linearity:** Because this way we see that $\hat{\beta} = \beta + A\varepsilon$ where $A = (X'X)^{-1}$.

(b) **Strict Exogeneity:** Because assiming that $\mathbb{E}[\epsilon|X] = 0$, we can see that $\mathbb{E}[\hat{\beta}|X] = \beta$

(c) **No Perfect Multicollinearity and Homoskedasticity:** These two let us see that the $\text{Var}(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$.

These assumptions plus the normality let us demonstrate that $\hat{\beta}_k \underset{H_0}{\sim} N(\beta_k, \sigma^2(X'X)^{-1}_{kk})$.

(d) For the test you performed in questions (2b), you were asked to use the t test statistic. Using this test requires, among others, for disturbances to be normally distributed. One of the available tests of normality of the distribution of a given random variable is the Jarque-Bera test. Under the null hypothesis of normality, the Jarque-Bera (JB) test statistic is:

$$JB \equiv \frac{n}{6}\left[sk^2 + \frac{(kur-3)^2}{4}\right] \overset{a}{\sim} \chi^2(2),$$

where sk is the sample coefficient of skewness of the variable and kur is its sample coefficient of kurtosis. Using a significance level of 1%, draw (by hand absolutely fine) the distribution of JB under H_0 and the corresponding acceptance and rejection regions. Provide an intuition for the location of the acceptance region.

Answer:

The acceptance region is the closest region to 0 because as Jarque-Berta test is a test of normality of the distribution, we will accept the test if $sk \approx 0$ and $kur \approx 3$ (a normal distribution has a teoretical skewness of 0 and a teoretical kurtosis of 3). We see that if $sk \approx 0$ and $kur \approx 3$ then $JB - value \approx 0$.

hw4/JB_drawing.jpeg

(e) Now, we want to test for the presence of normality in our disturbances using the JB test. Ideally, to test normality of disturbances, JB test should be applied to a sample of disturbances, but, given that they are unobservable, the JB test is usually applied to our OLS residuals. Explain how, if all the assumptions regarding the dgp for consistency of OLS estimator are met, it would be justified for $\hat{\epsilon}_t$'s to take the place of ϵ_t's to perform the test.

Answer:

The JB test should ideally be applied to the true disturbances ϵ_t, but since these are unobservable, we use the OLS residuals $\hat{\epsilon}_t$ instead. This substitution is justified under the consistency assumptions for OLS.

Justification:

If all assumptions for consistency of the OLS estimator are met, then:

(a) **Consistency of $\hat{\beta}$:** As $n \to \infty$, $\hat{\beta} \xrightarrow{p} \beta$ (converges in probability to the true parameter)

(b) **Residuals approximate errors:** The OLS residuals are:

$$\hat{\epsilon}_t = y_t - X_t'\hat{\beta} = (X_t'\beta + \epsilon_t) - X_t'\hat{\beta} = \epsilon_t - X_t'(\hat{\beta} - \beta)$$

(c) **Vanishing difference:** Since $(\hat{\beta} - \beta) \xrightarrow{p} 0$ as $n \to \infty$, the difference between $\hat{\epsilon}_t$ and ϵ_t becomes negligible in large samples.

(d) **Convergence of moments:** The sample skewness and kurtosis computed from $\{\hat{\epsilon}_t\}$ converge to the population skewness and kurtosis of the true errors $\{\epsilon_t\}$ as $n \to \infty$.

Conclusion: With $n = 324$ observations, the residuals provide a reliable approximation to the true errors. The JB test on $\{\hat{\epsilon}_t\}$ is asymptotically equivalent to the (infeasible) test on $\{\epsilon_t\}$, making it valid for testing normality of the disturbances.

(f) Perform the JB test on the OLS residuals. What do you conclude? Comment.

Answer:

From the regression output in part (a), we calculate the JB test statistic on the OLS residuals:

Test Results:

- Sample size: $n = 324$
- Skewness: $sk = -2.541$
- Kurtosis: $kur = 13.401$
- JB statistic: $JB = \frac{324}{6}\left[(-2.541)^2 + \frac{(13.401-3)^2}{4}\right] = 1809.21$
- Critical value: $\chi^2_{0.01}(2) = 9.21$
- p-value: < 0.0001

Decision: Since $JB = 1809.21 \gg 9.21$, we strongly reject the null hypothesis at the 1% significance level.

Conclusion: The residuals are not normally distributed. The distribution exhibits severe negative skewness ($sk = -2.54$) and extreme excess kurtosis ($kur = 13.40$ vs. normal $= 3$), indicating heavy tails and asymmetry.

Comment: This result is typical for financial returns data. Stock returns commonly exhibit:

- Negative skewness: Market crashes are more sudden and severe than rallies
- Excess kurtosis (fat tails): Extreme events occur more frequently than predicted by normal distribution

This violation of the normality assumption suggests caution when interpreting the exact t-test from part (b). However, with $n = 324$, the Central Limit Theorem provides robustness, and asymptotic tests may be more appropriate.

(g) Repeat the test performed in (2b) using the asymptotic T-test statistic. Use a 1% significance level.

Answer:

We repeat the test from part (2b) using the asymptotic test, which relies on the standard normal distribution instead of the t-distribution.

Hypotheses:

$$H_0 : \beta_6 = 0 \quad \text{(No January effect)}$$
$$H_1 : \beta_6 \neq 0 \quad \text{(January effect exists)}$$

Test Statistic:

Under H_0 and as $n \to \infty$, the asymptotic distribution is:

$$T = \frac{\hat{\beta}_6}{\text{SE}(\hat{\beta}_6)} \xrightarrow{d} N(0,1)$$

From the regression output in part (a):

$$T = \frac{5.4352}{2.869} = 1.894$$

Critical Value:

For a two-tailed test at $\alpha = 0.01$ using the standard normal distribution:

$$z_{0.005} = \Phi^{-1}(0.995) = 2.576$$

where Φ^{-1} is the inverse of the standard normal CDF.

Decision Rule: Reject H_0 if $|T| > 2.576$

P-value:

$$p = 2 \times P(Z > 1.894) = 2 \times (1 - \Phi(1.894)) = 2 \times 0.0291 = 0.0582$$

Result: Since $|1.894| < 2.576$ and $p = 0.0582 > 0.01$, we fail to reject H_0 at the 1%

(h) Is the use of asymptotic tests justified in this case? Rigorously argue.

Answer:

Yes, the use of asymptotic tests is justified in this case.

Arguments in favor:

(a) **Large sample size:** With $n = 324$ observations, we have a sufficiently large sample for asymptotic approximations to be accurate. The large sample ensures that the limiting distributions provide good approximations to the finite-sample distributions.

(b) **Violation of normality:** The JB test in part (e) strongly rejected normality of the residuals (JB = 1809.21). The exact t-test requires $\epsilon|X \sim N(0, \sigma^2 I)$ for its finite-sample validity. Since this assumption is violated, the exact t-distribution may not actually be the correct distribution of our test statistic under H_0.

(c) **Central Limit Theorem:** Under mild regularity conditions (finite variance, independence), the CLT ensures that:

$$\sqrt{n}(\hat{\beta}_6 - \beta_6) \xrightarrow{d} N(0, V)$$

regardless of the distribution of ϵ_t. This means:

$$\frac{\hat{\beta}_6 - \beta_6}{\text{SE}(\hat{\beta}_6)} \xrightarrow{d} N(0, 1)$$

even without normality of errors.

(d) **Consistency of variance estimator:** With $n = 324$, the estimator $\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k}$ consistently estimates σ^2, so:

$$\text{SE}(\hat{\beta}_6) = \sqrt{\hat{\sigma}^2[(X'X)^{-1}]_{66}} \xrightarrow{p} \sigma\sqrt{[(X'X)^{-1}]_{66}}$$

(e) **Negligible difference between distributions:** For $n = 324$, the t-distribution with 318 degrees of freedom is virtually indistinguishable from the standard normal:

- $t_{0.005}(318) = 2.588$ vs. $z_{0.005} = 2.576$
- Difference in critical values: only 0.012

Conclusion: Given the large sample size, the violation of normality, and the reliance of the CLT only on weak conditions, asymptotic tests are not only justified but actually more appropriate than exact tests in this context. The asymptotic approximation provides valid inference without requiring the normality assumption that is clearly violated in our data.

(i) Consider the following statement: "Using the exact t test statistic leads to slightly more conservative inference, because we get larger acceptance regions and larger p-values than if we used the asymptotic version." Do you agree? Rigorously argue.

Answer:

Yes, we agree, since these larger values make it harder to reject the null hypothesis, leading to more conservative inference. We also know that the t-distribution has heavier tails than the normal distribution, especially for smaller sample sizes. This means that for a given significance level, the critical values from the t-distribution are larger than those from the normal distribution.

Example: Exact t-test needs $|t| > 2.588$ to reject, Asymptotic test needs $|t| > 2.576$ to reject, Since $2.588 > 2.576$, the t-test requires a bigger test statistic to reject.

For p values of our test statistic of 1.894: Exact t-test gives p-value = 0.0591 Asymptotic test gives p-value = 0.0582 Higher p-value = harder to reject