Office Use Only

MONASH University

Semester One 2019
Examination Period

Faculty of Business and Economics

EXAM CODES: **ETC2410-ETW2410-BEX2410**

TITLE OF PAPER: Introductory Econometrics - PAPER 1

EXAM DURATION: 2 hours writing time

READING TIME: 10 minutes

THIS PAPER IS FOR STUDENTS STUDYING AT: (tick where applicable)

☐ Caulfield ☑ Clayton ☐ Parkville ☐ Peninsula
☐ Monash Extension ☐ Off Campus Learning ☑ Malaysia ☐ Sth Africa
☐ Other (specify)

During an exam, you must not have in your possession any item/material that has not been authorised for your exam. This includes books, notes, paper, electronic device/s, mobile phone, smart watch/device, calculator, pencil case, or writing on any part of your body. Any authorised items are listed below. Items/materials on your desk, chair, in your clothing or otherwise on your person will be deemed to be in your possession.

No examination materials are to be removed from the room. This includes retaining, copying, memorising or noting down content of exam material for personal use or to share with any other person by any means following your exam.
Failure to comply with the above instructions, or attempting to cheat or cheating in an exam is a discipline offence under Part 7 of the Monash University (Council) Regulations, or a breach of instructions under Part 3 of the Monash University (Academic Board) Regulations.

AUTHORISED MATERIALS

OPEN BOOK ☐ **YES** ☑ **NO**

CALCULATORS ☑ **YES** ☐ **NO**
Only HP 10bII+ or Casio FX82 (any suffix) calculator permitted
SPECIFICALLY PERMITTED ITEMS ☑ **YES** ☐ **NO**
if yes, items permitted are: one A4 sheet of paper with hand written notes on both sides

Candidates must complete this section if required to write answers within this paper

STUDENT ID: __ __ __ __ __ __ __ __ DESK NUMBER: __ __ __ __ __

INSTRUCTIONS TO STUDENTS

- **Answer all FOUR questions. All questions are of equal value (15 marks). This paper is worth 60 marks in total and constitutes 60% of the final assessment.**
- **For multiple choice questions write the question number and only one letter (a), (b), (c), (d) or (e) for each question in your answer book (not on the question sheet).**
- **When testing a hypothesis, to obtain full marks you need to specify the null and the alternative hypotheses, the test statistic and its distribution under the null, and then perform the test and state your conclusion.**
- **If a question does not specify the level of significance of a hypothesis test explicitly, use 5%.**
- **Statistical tables are provided after Question 4.**

Question 1 (15 marks)

This question has 15 multiple choice questions. Make sure that you clearly specify the question number and only one letter for each multiple choice question in your answer book (**not on the question sheet**).

1. Consider two datasets. In dataset A, we have data on consumption expenditure, income and hours of work for every year from 2000 to 2017 for a group of individuals who were randomly selected in the year 2000. In data set B, we have data on consumption per capita, income per capita and unemployment rate for Australia, Indonesia, Malaysia, New Zealand, Thailand and Vietnam for every year from 2000 to 2017.

   (a) Both datasets are examples of time series data.
   (b) Both datasets are examples of cross-sectional data.
   (c) Both datasets are examples of panel data.
   (d) Dataset A is an example of panel data, dataset B is an example of time series data.
   (e) Dataset A is an example of cross-sectional data, dataset B is an example of time-series data.

   (1 mark)

2. Which of the following statements is NOT true?

   (a) Randomised controlled trials are the best means for measuring causal relationships.
   (b) In predictive modelling, the variables that are used as predictors need not cause the variable that they try to predict.
   (c) Correlation is not causation.
   (d) Time series observations are always *i.i.d.*
   (e) Time series data are ordered whereas cross section data are not.

   (1 mark)

3. Let x denote the weight of a newborn baby immediately after birth. x is a random variable with mean μ, i.e. $E(x) = \mu$, and variance σ^2, i.e. $E(x-\mu)^2 = \sigma^2$. We denote weights of 5 newborn babies selected at random by x_1, x_2, x_3, x_4 and x_5, and their sample average by $\bar{x}$. Which of the following statements is NOT true

   (a) $\sum_{i=1}^{5}(x_i - \bar{x}) = 0$
   (b) $\sum_{i=1}^{5} x_i = 5\bar{x}$
   (c) $E(\bar{x}) = \mu$
   (d) $\bar{x} = \mu$
   (e) $\bar{x}$ is a linear combination of x_1, x_2, x_3, x_4 and x_5

(1 mark)

4. Let x and y denote returns to two risky assets. We are told that $E(x) = E(y) = \mu$, and $Var(x) = Var(y) = \sigma^2$. If we invest half of our savings in one of these assets and the other half in the other asset, then the variance of the return to our investment will be

   (a) $\frac{\sigma^2}{4}$ if x and y are uncorrelated
   (b) $\frac{\sigma^2}{2}$ if x and y are uncorrelated
   (c) $\frac{\sigma^2}{2}$ always
   (d) σ^2 always
   (e) $\frac{(x-\mu)^2+(y-\mu)^2}{4}$ if x and y are uncorrelated

(1 mark)

Questions 5 and 6 refer to the following p.d.f.: According to an expert, the annual growth rate of the real GDP and the inflation rate for Malaysia in 2019 are governed by the following joint probability density function:

Inflation rate ↓ , GDP growth rate →	**4%**	**5%**	**6%**
1%	0.1	0.1	0.0
2%	0.1	0.2	0.0
3%	0.1	0.1	0.1
4%	0.0	0.1	0.1

5. The expected growth rate of real GDP in Malaysia in 2019 according to this expert is:

   (a) a random variable
   (b) 5.00% because $\frac{4+5+6}{3} = 5$
   (c) 4.90% because $4 \times 0.3 + 5 \times 0.5 + 6 \times 0.2 = 4.9$
   (d) 2.50% because $1 \times 0.2 + 2 \times 0.3 + 3 \times 0.3 + 4 \times 0.2 = 2.5$
   (e) 4.92% because

$$\frac{1}{4} \times \{(4 \times \frac{0.1}{0.1+0.1} + 5 \times \frac{0.1}{0.1+0.1}) + (4 \times \frac{0.1}{0.1+0.2} + 5 \times \frac{0.2}{0.1+0.2}) + (4 \times \frac{0.1}{0.1+0.1+0.1} + 5 \times \frac{0.1}{0.1+0.1+0.1} + 6 \times \frac{0.1}{0.1+0.1+0.1}) + (5 \times \frac{0.1}{0.1+0.1} + 6 \times \frac{0.1}{0.1+0.1})\}$$
$$= 4.92$$

(1 mark)

6. Conditional on 5% GDP growth rate, the expected inflation rate in Malaysia in 2019 according to this expert is:

   (a) a random variable
   (b) 2.50% because $\frac{1+2+3+4}{4} = 2.5$
   (c) 2.50% because $1 \times 0.2 + 2 \times 0.3 + 3 \times 0.3 + 4 \times 0.2 = 2.5$
   (d) 1.20% because $1 \times 0.1 + 2 \times 0.2 + 3 \times 0.1 + 4 \times 0.1 = 1.2$
   (e) 2.40% because $1 \times \frac{0.1}{0.5} + 2 \times \frac{0.2}{0.5} + 3 \times \frac{0.1}{0.5} + 4 \times \frac{0.1}{0.5} = 2.4$

(1 mark)

Questions 7, 8 and 9 refer to the multiple regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

which in matrix notation is

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times(k+1)}{\mathbf{X}} \underset{(k+1)\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\mathbf{u}}$$

7. $E(\mathbf{u} \mid \mathbf{X}) = \mathbf{0}$ implies that

   (a) $E(\mathbf{X}'\mathbf{u}) = \mathbf{0}$
   (b) $\mathbf{X}'\widehat{\mathbf{u}} = \mathbf{0}$ where $\widehat{\mathbf{u}}$ is the vector of OLS residuals of regression of $\mathbf{y}$ on $\mathbf{X}$
   (c) $Var(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n$ where $\mathbf{I}_n$ is the identity matrix of order n
   (d) $\mathbf{X}'\mathbf{X}$ is invertible
   (e) Columns of $\mathbf{X}$ are linearly independent

(1 mark)

8. Which one of the following statements is correct?

   (a) $\mathbf{X}\mathbf{y}$ is an $n \times 1$ vector
   (b) $\mathbf{X}'\mathbf{X}$ is an $n \times n$ matrix
   (c) $\mathbf{X}'\mathbf{u} = \mathbf{0}$
   (d) $\mathbf{X}'\mathbf{u}$ is a $(k+1) \times 1$ vector
   (e) $\mathbf{X}'\boldsymbol{\beta}$ is a $(k+1) \times 1$ vector

(1 mark)

9. Assuming that this model satisfies all assumptions of the Classical Linear Model (CLM) and denoting the OLS estimator of $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$, which of the following statements is NOT correct?

   (a) $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$
   (b) $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$
   (c) Conditional on $\mathbf{X}$, $\widehat{\boldsymbol{\beta}}$ is normally distributed
   (d) $\widehat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$
   (e) $\widehat{\boldsymbol{\beta}}$ is equal to $\boldsymbol{\beta}$

(1 mark)

10. We have chosen a random sample of 100 publicly listed companies and recorded their average share price, profits, revenues and total costs in 2017-2018 financial year. Note that profits = revenue - total cost. In a regression model with the share price as the dependent variable and a constant, profit, revenue and total cost as independent variables, the OLS estimator

   (a) cannot be computed because $\mathbf{X}'\mathbf{X}$ matrix is not invertible
   (b) will be biased because share price is not normally distributed
   (c) will be unbiased
   (d) will be BLUE
   (e) will be unbiased but not BLUE

(1 mark)

Questions 11 to 13 refer to the following problem: We would like to model the relationship between the price of an apartment with its area and its number of bedrooms. We postulate the following population regression model

$$price = \beta_0 + \beta_1 area + \beta_2 bedrooms + u.$$

Suppose all assumptions of the Classical Linear Model applies to this model. We have collected data on price (in 1000 dollars), area (in square metres) and number of bedrooms for 120 randomly selected apartments and estimated the parameters of this models using OLS. This resulted in 318.99, 1.35 and 62.37 for estimates of β_0, β_1 and β_2 respectively.

11. Which of the following equations reports the results appropriately?

   (a) $\widehat{price} = 318.99 + 1.35\ area + 62.37\ bedrooms$
   (b) $\widehat{price} = 318.99 + 1.35\ area + 62.37\ bedrooms + \hat{u}$
   (c) $\widehat{price} = 318.99 + 1.35\ area + 62.37\ bedrooms + u$
   (d) $price = 318.99 + 1.35\ area + 62.37\ bedrooms + u$
   (e) $E\left(price \mid area, bedrooms\right) = 318.99 + 1.35\ area + 62.37\ bedrooms$

(1 mark)

12. Which of the following statements is correct?

   (a) $E\left(price \mid area, bedrooms\right) = 318.99 + 1.35\ area + 62.37\ bedrooms$
   (b) $E\left(price \mid area, bedrooms\right) = 318.99 + 1.35\ area + 62.37\ bedrooms + u$
   (c) $E\left(price \mid area, bedrooms\right) = 318.99 + 1.35\ area + 62.37\ bedrooms + \hat{u}$
   (d) $E\left(price \mid area, bedrooms\right) = \beta_0 + \beta_1 area + \beta_2 bedrooms$
   (e) $E\left(price \mid area, bedrooms\right) = \beta_0 + \beta_1 area + \beta_2 bedrooms + u$

(1 mark)

13. The null hypothesis for testing that given the area of an apartment, its number of bedrooms is not a significant predictor of its price, is:

   (a) $H_0 : bedrooms = 0$
   (b) $H_0 : (bedrooms \mid area) = 0$
   (c) $H_0 : \widehat{\beta}_2 = 0$
   (d) $H_0 : \beta_2 = 0$
   (e) $H_0 : \widehat{\beta}_2 \neq 0$

(1 mark)

Questions 14 and 15 relate to the following econometric model: Some economists believe that the relationship between greenhouse gas emission and income is nonlinear. Denote a country's emission of CO_2 per capita by $CO2PC$ and its GDP per capita by $GDPPC$, and consider the following model:

$$CO2PC = \beta_0 + \beta_1 GDPPC + \beta_2 GDPPC^2 + u \quad (2)$$

14. The hypothesis that the relationship between $CO2PC$ and $GDPPC$ is linear versus the alternative that it is an inverted U shape relationship can be written as:

    (a) $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 < 0$
    (b) $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 > 0$
    (c) $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 < 0$
    (d) $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 > 0$
    (e) $H_0 : \beta_1 = \beta_2 = 0$ against H_1 : at least one of β_1 or β_2 not equal to zero

    (1 mark)

15. If we know that in the model shown in equation (2) $Var(u \mid GDPPC) = \sigma^2 GDPPC$, but all other assumptions of the Classical Linear Model are satisfied, then

    (a) we can still use the OLS estimator because it is unbiased, and we can use the usual OLS standard errors to perform t tests
    (b) we can still use the OLS estimator because it is unbiased, but we need to use heteroskedasticity robust standard errors to perform t tests
    (c) we cannot use the OLS estimator because the OLS estimator is biased in this case
    (d) we can still use the OLS estimator because it is the best linear unbiased estimator in this case
    (e) we can still use the OLS estimator because the OLS estimator is the same as the "weighted least squares" estimator in this case

    (1 mark)

Question 2 (15 marks)

2.a. Suppose we have a sample of n observations on a variable y. Show that if we run a regression of y on a constant only, the OLS estimate of the constant will be the sample average of y.

(3 marks)

2.b. From the World Development Indicators database, we have extracted data on the following variables for 121 countries in 2015:

Variable	**Definition**	**Range**
UNDER5	Mortality rate in children under 5 (per 1000 live births)	2.4 - 130.9
GDPPC	GDP per capita in PPP adjusted dollars (as defined in assignment 1)	626 - 80892
SANITATION	People using basic sanitation services (% of population)	7 - 100
WATER	People using basic drinking water services (% of population)	0 - 100

The "Range" column provides the range of these variable in our sample.

From these 121 countries, 35 are in sub-Saharan Africa. We have created a dummy variable called SUBSAHARA which is equal to 1 if the country is a sub-Saharan country and 0 otherwise. Using this data set, we have estimated the following regressions using OLS (standard errors are provided in parentheses below parameter estimates)

$$\widehat{UNDER5} = \underset{(2.1)}{17.2} + \underset{(3.8)}{59.6}\, SUBSAHARA \tag{3}$$

$$\widehat{UNDER5} = \underset{(14.5)}{159.0} - \underset{(2.2)}{7.2}\, \log(GDPPC) - \underset{(0.1)}{0.6}\, SANITATION - \underset{(0.1)}{0.2}\, WATER \tag{4}$$

i. From the information provided, compute the average under-5 mortality rate (a) for the 35 sub-Saharan countries, (b) for the remaining 86 countries, and (c) for all 121 countries in this sample.

(3 marks)

ii. Explain the estimated coefficients of $\log(GDPPC)$ in equation (4) in a way that a person with no econometric training would understand.

(2 marks)

iii. Suppose we want to test the hypothesis that after controlling for $\log(GDPPC)$, a 1 percentage point increase in the proportion of population with access to basic sanitation has the same effect on under-5 mortality as a 1 percentage point increase in the proportion of population with access to drinking water, against the alternative that these effects are not equal, at the 5% level of significance. Explain how we could do that. For full marks, you need to state the null, the alternative, the test statistic and its distribution under the null, any additional regressions that we may have to estimate to calculate the test statistic, and how to come up with a conclusion using this procedure. All of these need to be explained in the context of this question where appropriate.

(4 marks)

iv. We have added $SUBSAHARA$ to equation (4) and re-estimated it and obtained the following equation:

$$\begin{aligned}\widehat{UNDER5} = {} & \underset{(14.5)}{135.4} - \underset{(2.2)}{7.4}\, \log(GDPPC) - \underset{(0.1)}{0.4}\, SANITATION - \underset{(0.1)}{0.1}\, WATER \\ & + \underset{(5.8)}{18.2}\, SUBSAHARA\end{aligned} \tag{5}$$

Use this information to test the hypothesis that after controlling for GDP per capita and access to sanitation and water services, there is no difference between the mean of under-5 mortality in sub-Saharan countries and the rest of the world, against the alternative that sub-Saharan countries have a higher mean, at the 5% level of significance. Remember that you need to state all steps of hypothesis testing to obtain full marks.

(3 marks)

Question 3 (15 marks)

3.a. In predictive modelling, when we want to find the best subset of k explanatory variables $\{x_1, x_2, \ldots, x_k\}$ to predict a target variable y, we do not use R^2 to compare models. Explain why, and provide the formula of an alternative statistic (only one) that we can use for selecting the best predictive model, highlighting specifically how this statistic overcomes the deficiency of R^2 for model selection.

(3 marks)

3.b. We have randomly selected a sample of 249 employed men and collected the following information:

Variable	**Definition**	**Range**	**Median**
WAGE	hourly wage in dollars	7.5 - 125	30
EDUC	years of education	2 - 18	12
EXPER	years of experience	0 - 38	13

The "Range" and "Median" columns show the range and the median of each variable within our sample, and zero years of experience means people who have less than 6 months experience. Consider the following population regression model for the logarithm of wage given education and experience:

$$\log(WAGE) = \beta_0 + \beta_1 \ (EDUC - 12) + \beta_2 \ EXPER + \beta_3 \ EXPER^2 + u \tag{6}$$

We have estimated the following regression using OLS:

$$\begin{aligned} \log(\widehat{WAGE}) &= \underset{(0.066)}{2.837} + \underset{(0.010)}{0.095} \ (EDUC - 12) + \underset{(0.009)}{0.055} \ EXPER - \underset{(0.0003)}{0.001} \ EXPER^2 \quad (7) \\ R^2 &= 0.394, \ \text{standard error of the regression} = 0.420, \ n = 249 \end{aligned}$$

Note that we have subtracted 12 from years of education in order to make the results more readily interpretable.

i. Interpret the estimated coefficients in this regression, including its intercept.

(4 marks)

ii. Can we interpret the coefficient of $(EDUC - 12)$ as the estimate of the "return to education", i.e. proportional increase in wage caused by an extra year of education? Explain.

(2 marks)

iii. In order to test the hypothesis that the errors of this model are homoskedastic against a specific alternative, we have estimated the following auxiliary regression:

$$\begin{aligned} \hat{u}^2 &= \underset{(0.031)}{0.096} + \underset{(0.006)}{0.004} \ (EDUC - 12) + \underset{(0.002)}{0.005} \ EXPER \\ R^2 &= 0.039, \ \text{standard error of the regression} = 0.262, \ n = 249 \end{aligned}$$

where $\hat{u}$ is the estimated residual of equation (7). Use this information to perform the test at the 5% level of significance. Remember that you need to write down the null and the alternative and all steps of hypothesis testing to obtain full marks.

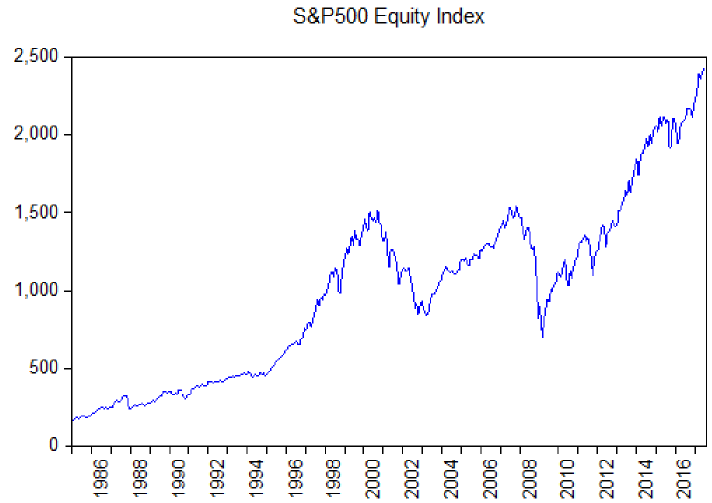
(4 marks)

iv. Suppose we are told that the conditional variance of the error in model (6) is proportional to experience, i.e. $Var(u \mid EDUC, EXPER) = \sigma^2 \times EXPER$. Explain how we can use this information to transform model (6) in such a way that the transformed model will have the same parameters but no heteroskedasticity.

(2 marks)

Question 4 (15 marks)

4.a. In finance, it is generally believed that asset prices are non-stationary time series. The following graph shows a time series plot of the Standard and Poor's U.S. Composite Stock Price Index (S&P500) from January 1985 to July 2017.



i. Define a covariance stationary time series.

(3 marks)

ii. With reference to the graph, explain why S&P500 is not a covariance stationary time series.

(1 mark)

4.b. In order to forecast the number of international visitors to Victoria, we postulated the following model:

$$\log(VIC_t) = \beta_0 + \beta_1\, t + \beta_2\, Q1_t + \beta_3\, Q2_t + \beta_4\, Q3_t + \beta_5\, AUD_{t-1} + u_t \tag{8}$$

where

VIC_t	:	number of short term international visitors arriving in Victoria at time t
$Q1_t$	:	dummy variable =1 if period t is the first quarter in a year, 0 otherwise
$Q2_t$	:	dummy variable =1 if period t is the second quarter in a year, 0 otherwise
$Q3_t$	:	dummy variable =1 if period t is the third quarter in a year, 0 otherwise
AUD_{t-1}	:	the value of 1 Australian dollar in terms of US dollar at time $t-1$,

and estimated it using quarterly data from 1991Q1 to 2008Q2:

$$\begin{aligned} \log(\widehat{VIC_t}) &= \underset{(0.037)}{10.71} + \underset{(0.0002)}{0.017}\ t - \underset{(0.017)}{0.027}\ Q1_t - \underset{(0.017)}{0.364}\ Q2_t - \underset{(0.017)}{0.300}\ Q3_t - \underset{(0.050)}{0.370}\ AUD_{t-1} \\ n &= 109,\ R^2 = 0.987 \end{aligned}$$

i. Explain what we can learn from the signs of the estimated coefficients of the three dummy variables and the sign of the coefficient of AUD_{t-1} (no need to interpret their magnitudes).

(2 marks)

ii. A tourism expert tells us that "as far as seasonal variation of tourism around its trend is concerned, there are only two seasons in Victoria, the cold season comprising $Q2 + Q3$, and the hot season comprising $Q1 + Q4$". Explain how you would use this data set to test this hypothesis. To obtain full marks, you need to specify the null and alternative hypotheses, the test statistic and its distribution under the null, and the regression or regressions that you need to run to test this hypothesis.

(4 marks)

iii. Since we are using time series data, we suspect that the errors of equation (8) might be serially correlated. To investigate that, we have obtained the correlogram of the residuals of the estimated model.

Correlogram of residuals of the estimated model
Sample: 1991Q1 2018Q2
Included observations: 109

Autocorrelation	Partial Correlation		AC	PAC
		1	0.583	0.583
		2	0.441	0.153
		3	0.358	0.079
		4	0.328	0.093
		5	0.223	-0.061
		6	0.122	-0.083
		7	0.012	-0.120
		8	-0.003	0.005
		9	-0.005	0.033
		10	-0.097	-0.112
		11	-0.127	-0.020
		12	-0.049	0.122

iii.a. What are the consequences of serial correlation in errors for the OLS estimator of the parameters and their usual OLS standard errors reported by statistical packages?

(2 marks)

iii.b. With reference to the correlogram of the residuals, explain how you would improve model (8). Justify your answer.

(3 marks)

END OF EXAMINATION

TABLE G.2 Critical Values of the *t* Distribution

		Significance Level				
1-Tailed:		.10	.05	.025	.01	.005
2-Tailed:		.20	.10	.05	.02	.01
Degrees of Freedom	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750
	40	1.303	1.684	2.021	2.423	2.704
	60	1.296	1.671	2.000	2.390	2.660
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

Examples: The 1% critical value for a one-tailed test with 25 *df* is 2.485. The 5% critical value for a two-tailed test with large (> 120) *df* is 1.96.

Source: This table was generated using the Stata® function invttail.

TABLE G.3a 10% Critical Values of the *F* Distribution										
	Numerator Degrees of Freedom									
Denominator Degrees of Freedom	1	2	3	4	5	6	7	8	9	10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

Example: The 10% critical value for numerator $df = 2$ and denominator $df = 40$ is 2.44.
Source: This table was generated using the Stata® function invFtail.

TABLE G.3b 5% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator $df(\infty)$ is 2.37.

Source: This table was generated using the Stata® function invFtail.

TABLE G.3c 1% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom: 1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Example: The 1% critical value for numerator $df = 3$ and denominator $df = 60$ is 4.13.
Source: This table was generated using the Stata® function invFtail.

TABLE G.4 Critical Values of the Chi-Square Distribution				
		Significance Level		
		.10	.05	.01
Degrees of Freedom	1	2.71	3.84	6.63
	2	4.61	5.99	9.21
	3	6.25	7.81	11.34
	4	7.78	9.49	13.28
	5	9.24	11.07	15.09
	6	10.64	12.59	16.81
	7	12.02	14.07	18.48
	8	13.36	15.51	20.09
	9	14.68	16.92	21.67
	10	15.99	18.31	23.21
	11	17.28	19.68	24.72
	12	18.55	21.03	26.22
	13	19.81	22.36	27.69
	14	21.06	23.68	29.14
	15	22.31	25.00	30.58
	16	23.54	26.30	32.00
	17	24.77	27.59	33.41
	18	25.99	28.87	34.81
	19	27.20	30.14	36.19
	20	28.41	31.41	37.57
	21	29.62	32.67	38.93
	22	30.81	33.92	40.29
	23	32.01	35.17	41.64
	24	33.20	36.42	42.98
	25	34.38	37.65	44.31
	26	35.56	38.89	45.64
	27	36.74	40.11	46.96
	28	37.92	41.34	48.28
	29	39.09	42.56	49.59
	30	40.26	43.77	50.89

Example: The 5% critical value with $df = 8$ is 15.51.

Source: This table was generated using the Stata® function invchi2tail.