University of Illinois at Urbana-Champaign
ECON471
Introduction to Applied Econometrics
Exam 2
November 10, 2020

H. Arvin-Rad

Instructions.

This is a closed book exam. Show your work. Unsupported or unreadable answers receive no credit. Answer all questions in parts I and II. In part I, each multiple-choice question has 3 points. In part II, the weight of each part of a question is the number in parenthesis next to the part. The formula sheet and statistical tables are attached to the exam.

Please hand-write your answers in the space provided for each part. If you don't have access to a printer to get a printout of the exam, make sure you clearly label each question and the parts in that question.

Please join the zoom meeting while taking the test. The password for the meeting is: EXAM2.

Upload your completed exam in a single file by 10:30pm, CST. You can only submit your exam once.

Please take a picture of your student ID and attach it to your exam, and sign the code of honor below. **Exams uploaded without the attached photo ID or no signing of the code of honor will be penalized.**

Violations of academic integrity as given in the Code on Campus Affairs will be taken extremely seriously. Students found cheating will be penalized according to the Code's guidelines.

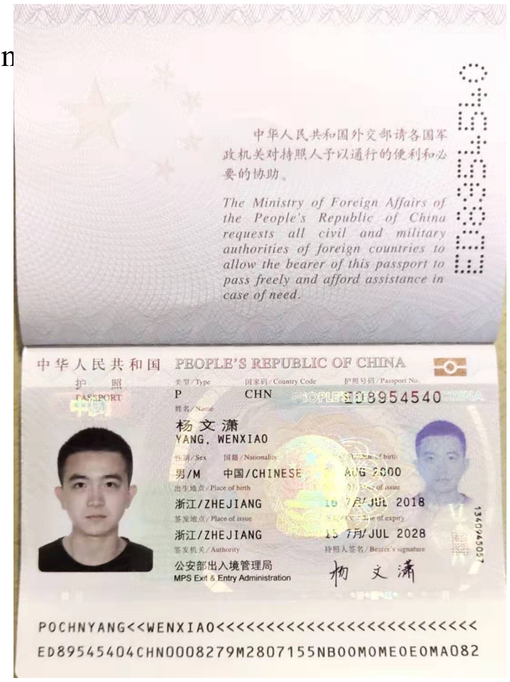
Please sign the honor code:
On my honor as a student, I have neither received nor given aid on

Signature: On my honor as a student I have neither received nor given aid on this exam.

Name: Wenxiao Yang

NetID: wenxiao 5

Section: B3



中华人民共和国外交部请各国军政机关对持照人予以通行的便利和必要的协助。

The Ministry of Foreign Affairs of the People's Republic of China requests all civil and military authorities of foreign countries to allow the bearer of this passport to pass freely and afford assistance in case of need.

中华人民共和国 PEOPLE'S REPUBLIC OF CHINA
护照 PASSPORT

类型/Type	国家码/Country Code	护照号码/Passport No.
P	CHN	ED8954540

姓名/Name
杨文潇
YANG, WENXIAO

性别/Sex	国籍/Nationality	出生日期/Date of birth
男/M	中国/CHINESE	AUG 2000
出生地点/Place of birth		签发日期/Date of issue
浙江/ZHEJIANG		16 7月/JUL 2018
签发地点/Place of issue		有效期至/Date of expiry
浙江/ZHEJIANG		15 7月/JUL 2028
签发机关/Authority		持照人签名/Bearer's signature
公安部出入境管理局 MPS Exit & Entry Administration		杨文潇

1340945057

POCHNYANG<<WENXIAO<<<<<<<<<<<<<<<<<<<<<<<<<<<
ED89545404CHN0008279M2807155NBOOMOMEOEOMA082

I am a new transfer student
I haven't got my ID card
This is the picture of my passport.

$$\frac{SSR_r - SSR_{ur} / q}{SSR_{ur} / n-k-1}$$

Part I (30 points). Multiple Choice Questions. Please enter your answers for this part in the multiple-choice answer sheet (page 5) by circling the letter of your choice.

d

1.The rationale behind the F test is that if the null hypothesis is true, by imposing the null hypothesis restrictions on the OLS estimation the sum of squared residuals

~~a~~. falls by a significant amount

~~b~~. rises by a significant amount

~~c~~. falls by an insignificant amount

[✓ d. rises by an insignificant amount]

e

2. which of the following is an advantage of multivariate regression over univariate regression?

a. One can control for other determinants of y that may be correlated with the primary x variable of interest.

b. The zero conditional mean assumption is more likely to hold with multivariate regression.

c. One can often reduce bias by moving to multivariate regression.

d. One can often obtain a model that better fits the data with more right-hand-side variables in the model.

e. All of the above are advantages of multivariate regression over univariate regression.

b.

3. Suppose you estimate a regression model and obtain $\hat{\beta}_1 = -.56$ and a p-value of .086 for testing $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. The p-value for testing $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 < 0$ is

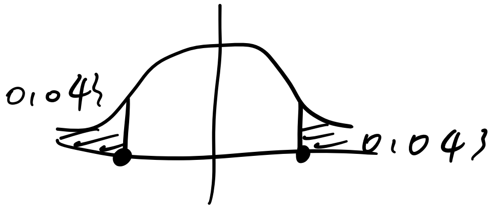
a. -.086

[✓ b. .043]

c. .172

d. .086

e. None of the above



d.

4. Suppose you regress y on x and the square of x.

a. Estimates will be unreliable

b. It doesn't make sense to use the square of x as a regressor

c. The regression will not run because these two regressors are perfectly correlated

[✓ d. There should be no problem with this.]

b.

5. If $\hat{\beta}_j$, an unbiased estimator of β_j, is consistent, then the:

a. distribution of $\hat{\beta}_j$ becomes more and more loosely distributed around β_j as the sample size grows.

b. distribution of $\hat{\beta}_j$ becomes more and more tightly distributed around β_j as the sample size grows.

c. distribution of $\hat{\beta}_j$ tends toward a ~~standard~~ normal distribution as the sample size grows.

d. distribution of $\hat{\beta}_j$ remains unaffected as the sample size grows.

~~d.~~ A

6. Which of the following statements is true when the dependent variable, y, takes only positive values?

a. Models using log(y) as the dependent variable will satisfy CLM assumptions more closely than models using the level of y.

b. Taking log of a variable often expands its range.

$(100\cdot y) = \beta_0 + \beta_1(100\cdot X) \Rightarrow \beta_1' = \beta_1$

c. Taking log of variables make OLS estimates more sensitive to extreme values.

d. Taking logarithmic form of variables make the slope coefficients more responsive to rescaling.

$\log y(100) = \beta_0 + \beta_1(X\cdot 100) \Rightarrow \beta_1' = \frac{\beta_1}{100}$ less responsive.

~~d.~~ C

7. Consider the following two regression models:

Model 1: $Y_i = \beta_0 + \beta_1 X_{i1} + u_i$

$R_1^2 < R_2^2$

Model 2: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$

$$\bar{R}^2 = 1 - \frac{\frac{SSR}{n-k-1}}{\frac{SST}{n-1}} = 1 - \underbrace{(1-R^2)}_{1>2}\underbrace{\frac{n-1}{n-k-1}}_{2>1.}$$

The same sample of size n was used to estimate both models. Suppose that both models have identical R^2 values. Which of the following statements is true?

一样的 R^2

a. The two models will also have identical values of adjusted R^2

b. Model 2 must have a higher value of adjusted R^2

c. Model 2 must have a lower value of adjusted R^2

d. It is not possible to determine which model will have the higher adjusted R^2 without knowing the sample size n

Model 1: higher

$$\tilde{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k-1}$$

8. In the following equation, gdp refers to gross domestic product, and FDI refers to foreign direct investment.

b.

$$\log(gdp) = 2.65 + 0.527\log(bankcredit) + 0.222FDI$$
$$(0.13) \quad (0.022) \qquad\qquad (0.017)$$

$\% gdp = 0.527 \% b$

$= 22.2 \, FDI$

Which of the following statements is then true?

~~a.~~ If gdp increases by 1%, bank credit increases by 0.527%, the level of FDI remaining constant.

✓ b. If bank credit increases by 1%, gdp increases by 0.527%, the level of FDI remaining constant.

~~c.~~ If gdp increases by 1%, bank credit increases by log(0.527)%, the level of FDI remaining constant.

~~d.~~ If bank credit increases by 1%, gdp increases by log(0.527)%, the level of FDI remaining constant.

~~b.~~ d

9. Suppose that the error term u is independent of the explanatory variables, and it takes on the values -2, -1, 0, 1, 2 with equal probability of 1/5. This violates MLR6: $u \sim N(0, \sigma^2)$

~~a.~~ The zero conditional mean assumption

b. One of the Gauss-Markov assumptions MLR1-5

~~c.~~ The no perfect collinearity assumption

d. One of the Classical Linear Model assumptions /MLR 1-6.

d.

10. You have estimated the relationship between tests cores (*TestScore*) and the student-teacher ratio (*STR*). The regression output is $\widehat{TestScore} = 698.9 - 2.28 \times STR$, and the standard error on the slope is 0.48. The *F*- statistic for testing the overall significance of the regression model is approximately

a. 0.96

b. 1.96

c. 4.75

d. 22.56

Multiple-Choice Answer Sheet

Please circle the letter of your choice for each multiple-choice question.

1. a b c [✓ d] e
2. a b c d [✓ e]
3. a [✓ b] c d e
4. a b c [✓ d] e
5. a [✓ b] c d e
6. a b c [✓ d] e
7. a b c [✓ d] e
8. a [✓ b] c d e
9. a [✓ b] c d e
10. a b c [✓ d] e

Part II(70 points)

1. (30 points) An equation explaining SAT score is specified as follows:

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female + \beta_4 black + \beta_5 female * black + u$$

The variable *sat* is the combined SAT score; *hsize* is size of the student's high school graduating class, in hundreds; *female* is a gender dummy variable equal to one for females, and zero otherwise; and *black* is a race dummy variable equal to one for blacks, and zero otherwise.

We estimate the model using the data in GPA2 (with standard errors in parentheses):

$$\widehat{sat} = 1028.10 + 19.30\, hsize - 2.19\, hsize^2 - 45.09\, female$$

(6.29) (3.83) (0.53) (24.29)

$$-169.81\, black + 62.31\, female * black$$

(82.71) (18.15)

$$n = 4137, \quad R^2 = .0858.$$

a. (9) Is there strong evidence that $hsize^2$ should be included in the model? From this equation, what is the optimal high school size?

t-statistic of $hsize^2 = \frac{-2.19}{0.53} = -4.13208.$

$\Rightarrow$ significant $\Rightarrow$ should be included.

$\frac{\Delta sat}{\Delta hsize} = 19.30 - 2 \times 2.19\, hsize$

$\Rightarrow$ optimal hsize is $\frac{19.30}{2 \times 2.19} = 4.40639$

b. (6) Holding *hsize* fixed, what is the estimated difference in SAT score between nonblack females and nonblack males? Test the null hypothesis that there is no difference between their scores, against the alternative that there is a difference.

estimated difference: nonblack females are 45.09 lower than noblack males in average

$H_0: \beta_3 = 0.$ $\quad$ $t\text{-statistic} = \dfrac{-45.09}{24.29} = -1.85632.$

$|t\text{-statistic}| < t_{\text{critical } 0.05} = 1.960$

$\Rightarrow$ accept H_0

$\Rightarrow$ no difference (in $\alpha = 0.05$.)

c. (6) What is the estimated difference in SAT score between nonblack males and black males? Test the null hypothesis that there is no difference between their scores, against the alternative that there is a difference.

estimated difference: nonblack males are 169.81 higher than black males in average.

$H_0: \beta_4 = 0$ $\quad$ $t\text{-statistic} = \dfrac{-169.81}{82.71} = -2.0531$

$|t\text{-statistic}| > t_{\text{critical } 0.05} = 1.960.$

$\Rightarrow$ reject H_0

$\Rightarrow$ there is a difference (in $\alpha = 0.05$.)

β_3	β_4	β_5
✓	✓	✓
✓	×	×

d. (9) What is the estimated difference in SAT score between black females and nonblack females? What would you need to do to test whether the difference is statistically significant?

$\beta_4 + \hat{\beta}_5 = -169.81 + 62.31 = -107.5$

estimated difference: black femals are 107.5 lower than nonblack females in average

$H_0 : \beta_4 + \beta_5 = 0$

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female + \beta_5 ((female - 1) \cdot black)$$

Use the same sample to regress the equation above. and get the R^2_{new}

$$F\text{-statistic} = \frac{(R^2 - R^2_{new})/1}{(1 - R^2)/4131}$$

if F-statistic < 3.84

accept H_0, i.e the difference is insignificant.

otherwise reject H_0, i.e. the difference is significant.

2. (20 points) The monthly salary (*WAGE*), age (*AGE*), number of years of education beyond the eight grade (*EDUC*), and the number of years of experience (*EXPER*) were obtained for a random sample of 49 workers in large manufacturing firm. The estimated relation between *WAGE* and the characteristics of a worker is as follows (with standard errors in parentheses):

$$\widehat{WAGE} = \underset{(423.47)}{632.244} + \underset{(34.86)}{142.510}\,EDUC + \underset{(14.30)}{43.225}EXPER - \underset{(8.70)}{1.913}\,AGE$$

a. (9) The value of adjusted R^2 is 0.277. Using this information, test the model for overall significance. (Note: you have all the information you need to perform the test.)

$H_0: \beta_{educ} = \beta_{exper} = \beta_{age} = 0.$

$$0.277 = 1 - (1-R^2)\,\frac{49-1}{49-3-1}$$

$$\Rightarrow R^2 = 0.322$$

$$F\text{-statistic} = \frac{\overset{0.322}{\cancel{R^2}}/3}{(1-\underset{0.322}{\cancel{R^2}})/49-4} = \overset{7.12}{\cancel{5.746888}}.$$

$> F_{critical}$

$\Rightarrow$ the model is overall significant.

b. (8) Test the coefficients of *EDUC* and *AGE* individually for statistical significance by calculating the p-value for each case. What do you conclude?

EDUC: $t\text{-statistic} = \frac{142.510}{34.86} = 4.088067$

$p\text{-value} < 0.01$

AGE: $t\text{-statistic} = -\frac{1.913}{8.70} = -0.21989$

$p\text{-value} > 0.2$

$\Rightarrow$ coefficient of EDUC is significant

coefficient of AGE is insignificant.

c. (3) How do you explain the negative sign of the coefficient of *AGE*?

① in large manufacturing firm, when people get older, they may work less, which causes less wage.

② the number of sample is small, the data source may have problem.

③ and as we discussed above, coefficient of AGE is insignificant, the coefficient may can't accuratly express the relation between AGE and wage.

$H_0 = \beta_1+\beta_2=1.$

$y_i = \beta_0 + (1-\beta_2)x_{i1} + \beta_2 x_{i2} + u_i$

3. (12 points) In the model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ describe two alternative ways of testing the hypothesis $\beta_1 + \beta_2 = 1$. Be sure to do this using an *F* test and a *t* test. For each test, describe regression(s) to run, the statistic to compute, the distribution of the statistic under the null hypothesis (including the degrees of freedom), and the criterion for the rejection of the null hypothesis.

F-test:

regress initial model.

$\beta_1 = 1-\beta_2$

get R^2 of initial model, n: number of samples.

$\hat{y} = \hat{\beta}_0 + X_1 + \hat{\beta}_2(X_2 - X_1)$

$y_i - x_{i1} = \beta_0 + \beta_2(x_{i2} - x_{i1}) + u_i$

regress the equation above with the same samples of initial model.

get its R^2: R_F^2

F-statistic $= \dfrac{(\overset{SSR_F - SSR_{ur}}{\cancel{R^2 - R_F^2}})/1}{\underset{SSR_{ur}}{\cancel{(1-R^2)}}/n-3}$

distribution: $F_{1,n-3}$

get critical value of $F_{1,n-3}$ in $\alpha = 0.05$

if F-statistic > critical value, reject H_0.

---

t-test: $\hat{y} = \hat{\beta}_0 + (\widehat{\beta_1+\beta_2})X_1 + \hat{\beta}_2(X_2 - X_1)$

regress the equation above

get $\widehat{\beta_1+\beta_2}$ and its standard error $se(\widehat{\beta_1+\beta_2})$

t-statistc $= \dfrac{\widehat{\beta_1+\beta_2} - 1}{se(\widehat{\beta_1+\beta_2})}$, distribution: t_{n-3}

get critical value of t_{n-3} in $\alpha = 0.05$

if $|$t-statistic$|$ > critical value, reject H.

4. (8 points) To test the effectiveness of a job training program on the subsequent wages of workers, we estimate the model

$\log(wage) = \beta_0 + \beta_1 train + u,$

where $train$ is a binary variable equal to unity if a worker participated in the program. Think of the term u as containing unobserved worker ability. If less able workers have a greater chance of being selected for the program, and you use an OLS analysis, what can you say about the likely bias (or inconsistency) in the OLS estimator of β_1 when you omit ability from the model because you don't have data on ability? Explain.

less able → greater chance ⇒ $Cov(train, ability) < 0$

$$\tilde{\beta}_1 = \beta_1 + \frac{\sum (train_i - \overline{train}) u_i}{\sum (train_i - \overline{train})^2}$$

assume $u = \beta_2 \, ability + v$, $\beta_2 > 0$

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \frac{Cov(train, ability)}{Var(train)} \beta_2$$

inconsistency: $\frac{Cov(train, ability)}{Var(train)} \cdot \beta_2 < 0.$

since $Cov(train, ability) < 0$, $Var(train), \beta_2 > 0$.

Formula Sheet

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$SST = \sum (Y_i - \bar{Y})^2 \qquad SSE = \sum (\hat{Y}_i - \bar{Y})^2 \qquad SSR = \sum (Y_i - \hat{Y}_i)^2$$

$$R^2 = \frac{SSE}{SST}$$

$$r_{x,y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2}\sqrt{\sum (Y_i - \bar{Y})^2}}.$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \bar{X})^2}$$

$$\hat{\sigma}^2 = \frac{SSR}{n-k-1}$$

$$\widehat{Var(\widehat{\beta_1})} = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2}$$

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum (X_i - \bar{X})^2}}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i$$

$\hat{\beta}_j = \frac{\sum \hat{r}_{ij} y_i}{\sum \hat{r}_{ij}^2}$, $j = 1, 2, ..., k,$ where the $\hat{r}_{ij}$ are the OLS residuals from the regression of x_j on all the other regressors.

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$

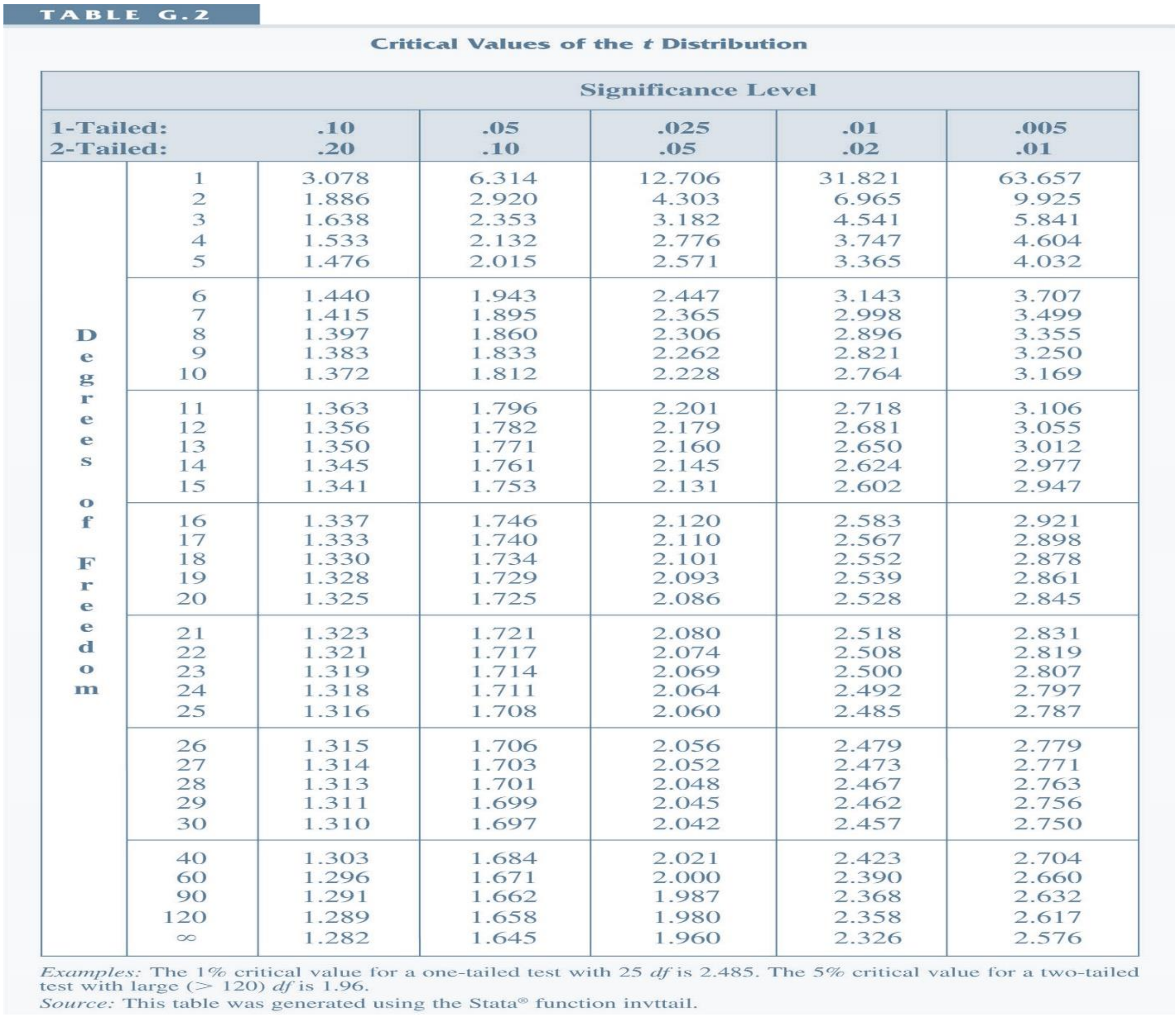


TABLE G.2

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.

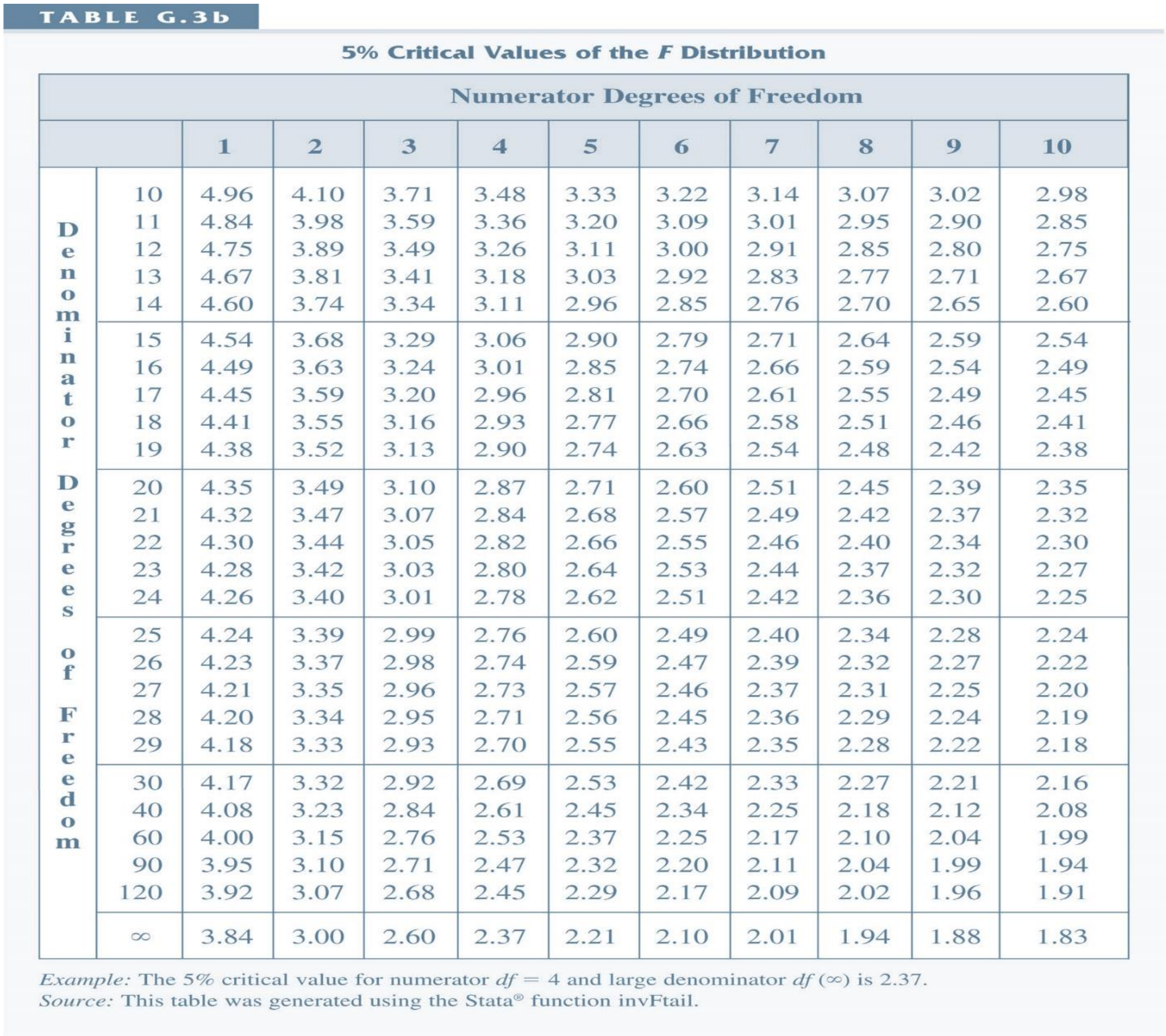


TABLE G.3b

5% Critical Values of the *F* Distribution

		Numerator Degrees of Freedom									
		1	2	3	4	5	6	7	8	9	10
Denominator Degrees of Freedom	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 (∞) is 2.37.
Source: This table was generated using the Stata® function invFtail.