University of Illinois at Urbana-Champaign

ECON471

Introduction to Applied Econometrics

Final Exam

December 16, 2020

Instructions:

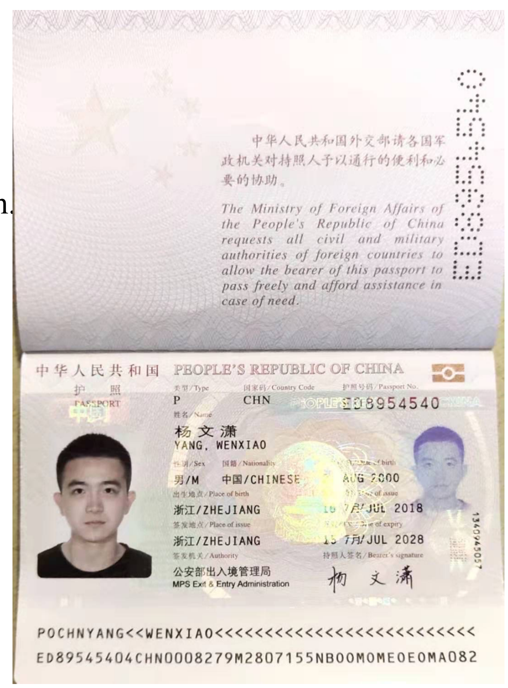
This is a closed book exam. Show your work. Unsupported or unreadable answers receive no credit. Answer all questions in parts I and II. Each multiple-choice question has a weight of 2.5. In part II, the weight of each part in a question is shown by the number in parenthesis next to the part.

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 with your webcam on while you are taking the test. The password for the meeting is: FINAL.

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

The statistical tables and the formula sheet are attached to the exam.



Please sign the honor code:

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

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

I am a new transfer student

I haven't got my ID card

This is the picture of my passport.

Part I (30 points): Multiple Choice Questions. Enter your answers on the multiple-choice answer sheet (page 5) by circling the letter of your choice.

$\beta_0 + \beta_1 X_i = \beta_0 + \beta_1 X_i + \beta_2 + \beta_3 X_i$

d

1. In the regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + u_i$, where X is a continuous variable and D is a binary variable, to test that the two regressions are identical, you must use the

 a. t-statistic for $\beta_2 = 0$.
 b. F-statistic for the joint hypothesis that $\beta_0 = 0, \beta_1 = 0$.
 c. t-statistic for $\beta_3 = 0$.
 d. F-statistic for the joint hypothesis that $\beta_2 = 0, \beta_3 = 0$.

b

2. In the simple regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$, the normal approximation to the sampling distribution of $\hat{\beta}_1$ is powerful because

 a. many explanatory variables in real life are normally distributed.
 b. it allows econometricians to develop methods for statistical inference.
 c. many other distributions are not symmetric.
 d. it implies that OLS is the BLUE estimator for β_1.

c

3. You have collected data for the 50 U.S. states and estimated the following relationship between the change in the unemployment rate from the previous year $\widehat{\Delta ur}$ and the growth rate of the respective state real GDP (g_y). The results are as follows

$$\widehat{\Delta ur} = \underset{(0.12)}{2.81} - \underset{(0.04)}{0.23} \times g_y,\ R^2 = 0.36,\ \hat{\sigma} = 0.78$$

t_{50-2}

Assuming that the estimator has a normal distribution, the 95% confidence interval for the slope is approximately the interval

$-2.0\ldots < \frac{-0.23 - \beta}{0.04} < 2.0\ldots$

 a. [-0.45, 0.45]
 b. [-0.31, 0.15]
 c. [-0.31, -0.15]
 d. [-0.33, -0.13]

$0.08 + 0.23$

$-0.08 + 0.23$

b

4. Omitted variable bias

 a. will always be present as long as the regression $R^2 < 1$.
 b. is always there but is negligible in almost all economic examples.
 c. exists if the omitted variable is correlated with the included regressor but is not a determinant of the dependent variable.
 d. exists if the omitted variable is correlated with the included regressor and is a determinant of the dependent variable.

d

5. The following OLS assumption is most likely violated by omitted variables bias:

   a. $E(u_i \mid X_i) = 0$
   b. (X_i, Y_i), i=1,2,..., n are randomly drawn observations
   c. there are no perfect collinearity
   d. there is homoskedasticity

c

6. Consider the following multiple regression models (a) to (d) below. *female* = 1 if the individual is a female, and is zero otherwise; *male* is a binary variable which takes on the value one if the individual is male, and is zero otherwise; *married* is a binary variable which is one for married individuals and is zero otherwise, and *single* is (1-married). Regressing weekly earnings (*Earn*) on a set of explanatory variables, you will experience perfect multicollinearity in the following cases except:

   ~~a.~~ $\widehat{Earn}_i = \hat{\beta}_0 + \hat{\beta}_1 female + \hat{\beta}_2 male + \hat{\beta}_3 X3_i$
   ~~b.~~ $\widehat{Earn}_i = \hat{\beta}_0 + \hat{\beta}_1 married + \hat{\beta}_2 single + \hat{\beta}_3 X3_i$
   c. $\widehat{Earn}_i = \hat{\beta}_0 + \hat{\beta}_1 female + \hat{\beta}_3 X3_i$
   ~~d.~~ $\widehat{Earn}_i = \hat{\beta}_0 + \hat{\beta}_1 female + \hat{\beta}_2 married + \hat{\beta}_3 single + \hat{\beta}_4 X3_i$

c

7. A regression model in which $.01\beta_1$ represents the expected change in Y in response to a 1% increase in X_1 is

$$Y_n - Y = \beta_1 \ln \frac{X_n}{X_1} = \beta_1 \ln\left(\frac{X_n - X_1}{X_1} + 1\right)$$

$$100(Y_n - Y) = \beta_1 \underline{\left(100 \frac{X_n - X_1}{X_1}\right)}$$

$x\%$

   a. $Y = \beta_0 + \beta_1 X_1 + u$.
   b. $\ln(Y) = \beta_0 + \beta_1 X_1 + u$.
   c. $Y = \beta_0 + \beta_1 \ln(X_1) + u$.
   d. $\ln(Y) = \beta_0 + \beta_1 \ln(X_1) + u$.

c

8. In the simple linear regression, the least squares assumption $E(u_i \mid X_i) = 0$ means that

   ~~a.~~ the regression error and X are independent.
   ~~b.~~ $E(Y_i) = E(u_i)$.
   c. $E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i$.
   ~~d.~~ u_i has a normal distribution.

a

9. Let $\hat{\beta}_j$, $j = 0, 1, 2$, be the OLS estimators of the parameters in the regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. Let $x_2^* = c_2 x_2$, and $y^* = c_1 y$. If $\hat{\beta}_j^*$, $j = 0,1,2$, are the OLS estimators from the regression of y^* on x_1, and x_2^*, then

$\frac{1}{C_1} y^* = \beta_0 + \beta_1 X_1 + \beta_2 \frac{X_2^*}{C_2} + u.$

$y^* = C_1 \beta_0 + C_1 \beta_1 + \frac{C_1}{C_2} \beta_2 X_2^*$

a. $\hat{\beta}_0^* = c_1\hat{\beta}_0$, $\hat{\beta}_1^* = c_1\hat{\beta}_1$, $\hat{\beta}_2^* = \frac{c_1}{c_2}\hat{\beta}_2$

b. $\hat{\beta}_j^* = \frac{c_1}{c_2}\hat{\beta}_j$, $j = 0, 1, 2$

c. $\hat{\beta}_0^* = c_1\hat{\beta}_0$, $\hat{\beta}_1^* = \hat{\beta}_1$, $\hat{\beta}_2^* = \frac{c_1}{c_2}\hat{\beta}_2$

d. $\hat{\beta}_0^* = c_2\hat{\beta}_0$, $\hat{\beta}_1^* = \frac{c_1}{c_2}\hat{\beta}_1$, $\hat{\beta}_2^* = c_2\hat{\beta}_2$

d

10. To test whether or not the population regression function is linear rather than a polynomial of order *r*,

a. check whether the regression R^2 for the polynomial regression is higher than that of the linear regression.
b. compare the *SST* from both regressions.
c. look at the pattern of the coefficients: if they change from positive to negative to positive, etc., then the polynomial regression should be used.
d. use the test of (*r*-1) restrictions using the *F*-statistic.

11. In the probit model $\Pr(Y = 1 | X_1, X_2, ..., X_k) = \Phi(\beta_0 + \beta_1 X_1 + \beta_x X_2 + ... + \beta_k X_k)$,

a

a. the β's do not have a simple interpretation.
b. the slopes tell you the effect of a unit increase in *X* on the probability of *Y*.
c. β_0 cannot be negative since probabilities have to lie between 0 and 1.
d. β_0 is the probability of observing *Y* when all *X*'s are 0.

b

12. The following problems could be analyzed using probit and logit estimation with the exception of whether or not

a. a college student decides to study abroad for one semester.
b. being a female has an effect on earnings.
c. a college student will attend a certain college after being accepted.
d. applicants will default on a loan.

Mutiple-Choice Answer Sheet

Circle the letter of your answer 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
11.	a	b	c	d	e
12.	a	b	c	d	e

Part II (70 points) Short Answer Questions: Answer all questions in this part.

1. (21 points) Data were collected from a random sample of 220 home sales from a community in 2015. Let *price* denote the selling price (in \$1000), *bdrms* denote the number of bedrooms, *bath* denote the number of bathrooms, *sqrft* denote the size of the house (in square feet), *lotsize* denote the lot size (in square feet), *age* denote the age of the house (in years), and *poor* denote a binary variable that is equal to one if the house is reported as "poor." An estimated regression yields

$$\widehat{price} = 119.2 + 0.485\, bdrms + 23.4\, bath + 0.156\, sqrft + 0.002\, lotsize + 0.090\, age - 48.8\, poor$$

(23.9) (2.61) (8.94) (0.011) (0.00048) (0.311) (10.5)

$\bar{R}^2 = 0.72$, $\hat{\sigma} = 41.5$,

where the standard errors are in parentheses.

a. (2) Suppose that a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?

expected increase $= (-0.485 + 23.4) \times 1000\,\$ = 22915\,\$$

b. (3) Suppose that a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?

expected increase $= (23.4 + 100 \times 0.156 + 100 \times 0.002) \times 1000\,\$$
$= 39200\,\$$

c. (2) What is the loss in value if a homeowner lets his house run down so that its condition becomes "poor"?

loss in value $= 48.8 \times 1000\,\$ = 48800\,\$$ in average.

d. (2) Compute the R^2 for the regression.

$\bar{R}^2 = 1 - (1-R^2)\dfrac{220-1}{220-7}$ $\quad 1-R^2 = (1-\bar{R}^2) \times \dfrac{213}{219}$

$\Rightarrow R^2 = 0.72767$.

e. (4) Interpret the coefficient of *bdrms*. Is the coefficient on *bdrms* statistically significant from zero? Carry out a test.

coefficient of bdrms : $\beta_1 = 0.485$,

which means when other factors fixed.
if the house has one more bedroom, the value of
the house will increase $0.485 \times 1000\$ = 485\$$ in average.

Test : $H_0: \beta_1 = 0$ $H_1: \beta_1 > 0$. $df = 213$, $\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \sim t_{213}$

t-statistics: $t = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{0.485}{2.61} = 0.1858 < t_{critical} = 1.645$ $(\alpha = 5\%)$

$\Rightarrow$ accept H_0 $\Rightarrow$ β_1 is insignificant from zero.

f. (3) Typically, five-bedroom houses sell for much more than two-bedroom houses. Is this consistent with your answer to part e? Explain.

Yes, it is not inconsistent with my answer to part e.
the part e shows that when other factors fixed, the number of bedrooms has insignificant effect on house value.
However, normally, the houses with five-bedroom have much more square feets than the two-bedroom houses.

g. (5) Test for the overall significance of the model.

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

H_1: H_0 is not true.

$$\frac{R^2/6}{(1-R^2)/213} \sim F_{6,213}$$

df = 6 , 213

$$\text{F-statistics} = \frac{R^2/6}{(1-R^2)/213} = 94.857 > F_{critical} = 2.1 \ (\alpha = 5\%).$$

$\Rightarrow$ reject H_0 $\Rightarrow$ overall significant.

2. (25 points) The variable *smoker* is a binary variable equal to one if an individual smokes. It has been argued that workplace smoking bans induce smokers to quit smoking by reducing their opportunities to smoke. Using data on a sample of 10,000 U.S. workers, we estimate both a linear probability model and a probit model for *smoker*, with *smkban, age, hsdrop, hsgrad, colsome, colgrad, black,* and *female* as explanatory variables, where

smkban=1 if there is workplace smoking ban
age = worker's age, in years
hsdrop = 1 if high school dropout
hsgrad = 1 if high school graduate
colsome = 1 if some college education
colgrad = 1 if college graduate
black = 1 if black
hispanic = 1 if hispanic
female = 1 if female.

The omitted education status is "Masters degree or higher."

The results are as shown in the table:

Regression	**Linear Probability Model (1)**	**Linear Probability Model (2)**	**Probit**
intercept	0.290 (0.045)	-0.014 (0.041)	-1.735 (0.053)
smkban	-0.078 (0.009)	-0.047 (0.009)	-0.159 (0.029)
age		0.0097 (0.0018)	0.035 (0.007)
age^2		-0.00013 (0.00002)	-0.00047 (0.00008)
hsdrop		0.323 (0.019)	1.142 (0.072)
hsgrad		0.233 (0.013)	0.883 (0.060)
colsome		0.164 (0.013	0.677 (0.061)
colgrad		0.045 (0.012)	0.235 (0.065)
black		-0.028 (0.016)	-0.084 (0.053)
hispanic		-0.105 (0.014)	-0.338 (0.048)
female		-0.033 (0.009)	-0.112 (0.028)

a. (5) Using the linear probability model (1), find the difference in the probability of smoking between workers affected by a workplace smoking ban and workers not affected by a workplace smoking ban. Is this difference statistically significant? Report a bound on the p-value.

the probability of smoking for workers affected by a workplace smoking ban is 0.078 lower than the workers not affected in average.

it is statistically significant:

$H_0: \beta_1 = 0, H_1: \beta_1 < 0 \quad \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \sim t_{9998}$

$\text{t-statistics} = \frac{-0.078}{0.009} = -8.667 \Rightarrow \text{p-value} < 0.005$

$\Rightarrow$ reject $H_0 \Rightarrow$ significant

b. (6) Using the linear probability model (2), find the difference in the probability of smoking between workers affected by a workplace smoking ban and workers not affected by a workplace smoking ban. Suggest a reason, based on the estimated models, explaining the change in the estimated effect of a working ban between (a) and (b).

the probability of smoking for workers affected by a workplace smoking ban is 0.047 lower than the workers not affected in average.

$LPM^{(1)}$ only have one explanatory variable: smkban which is related to other omitted explanatory variables, and some effects of these variables are also contained in the β_1 in $LPM^{(1)}$. Now in $LPM^{(2)}$, we have 10 explanatory variables, the related variables of smkban are shown separetly, which low down the absolute value of β_1.

c. (4) Interpret the coefficients of the binary variables related to the level of education in the linear probability model (2).

$\beta_4 - \beta_7 = 0.323 - 0.045 = 0.278$

$\beta_5 - \beta_7 = 0.233 - 0.045 = 0.188$

$\beta_6 - \beta_7 = 0.164 - 0.045 = 0.119$

the college graduate workers have the lowest probability of smoking in average.

and the workers with some college educations are 0.119 higher probability to smoke.

the high school graduate workers are 0.188 higher prob to smoke than

the high school dropout workers are 0.278 higher than college graduate. College graduate workers

d. (5) Mr. A is white, non-Hispanic, 20 years old, and a high school dropout. Using the probit model, and assuming that Mr. A is not subject to a workplace smoking ban, calculate the probability that Mr. A smokes. Carry out the calculation again assuming that he is subject to a workplace smoking ban. What is the effect of smoking ban on the probability of smoking?

No ban:

$-1.735 + 0.035 \times 20 - 0.00047 \times 20^2 + 1.142 = -0.081.$

Probability: Mr.A smokes $= \int_{-\infty}^{-0.081} \phi(z)dz = \Phi(-0.081) \approx \underline{0.4681.}$

Ban:

$-1.735 - 0.159 + 0.035 \times 20 - 0.00047 \times 20^2 + 1.142 = -0.24.$

Probability: Mr. A smokes $= \int_{-\infty}^{-0.24} \phi(z)dz = \Phi(-0.24) \approx \underline{0.4052}$

the effect of ban will make the prob of smoking <u>0.0629 lower</u>

e. (5) Repeat (d) for the linear probability model (2). Do the probit and linear probability model results differ?

No ban:

Probability: Mr.A smokes $= -0.014 + 0.0097 \times 20 - 0.00013 \times 20^2 + 0.323 = \underline{0.451.}$

Ban:

Probability: Mr.A smokes $= -0.014 - 0.047 + 0.0097 \times 20 - 0.00013 \times 20^2 + 0.323 = \underline{0.404}$

the effect of ban will make the prob of smoking <u>0.047 lower</u>

there are some differences between probit and linear but not big.

3. (6 points) A researcher uses a finite distributed lag model to estimate dynamic causal effects of U.S. economic activity on Canada. The results for the sample period 1961:I-1995:IV are:

$$\widehat{urcan}_t = \underset{(0.83)}{-1.42} + \underset{(0.457)}{0.717} \times urus_t + \underset{(0.557)}{0.262} \times urus_{t-1} + \underset{(0.398)}{0.023} \times urus_{t-2} - \underset{(0.405)}{0.083} \times urus_{t-3}$$

$$- \underset{(0.504)}{0.726} \times urus_{t-4} + \underset{(0.385)}{1.267} \times urus_{t-5}$$

where *urcan* is the Canadian unemployment rate, and *urus* is the United States unemployment rate. Calculate the long-run multiplier and give a careful interpretation of it.

log-run multiplier $= 0.717 + 0.262 + 0.023 - 0.083 - 0.726 + 1.267$
$= 1.46$

if the United states unemployment rate increases 1 in long-run
the Canadian unemployment rate will increases 1.46 in average.

4. (9 points) Suppose $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$ satisfies assumptions MLR.1 to MLR.4. It is also known that $Var(u_i|x_i) = \sigma_i^2 = \sigma^2 x_i^2$.

a. (3) Are the OLS estimators of βs unbiased, consistent, and efficient (no need for a formal proof, but carefully justify your answers)?

heteroskedasticity violates MLR.5
heteroskedasticity is no need for unbiased and consistent.
the $E(u_i | X_i)$ is still 0. so β_s is still unbiased and consistent.
but GMT can't be satisfied now, which means
OLS is no longer the BLUE, ⟹ not efficient

b. (4) Derive step by step how you transform the model in order to obtain the weighted least squares (WLS) estimators of βs. What are the regressors in the transformed model?

$$Var(u_i | x_i) = \sigma^2 x_i^2 \Rightarrow h_i(x_i) = x_i^2$$

$$\left[\frac{y_i}{\sqrt{h_i}}\right] = \beta_0\left[\frac{1}{\sqrt{h_i}}\right] + \beta_1\left[\frac{x_i}{\sqrt{h_i}}\right] + \beta_2\left[\frac{x_i^2}{\sqrt{h_i}}\right] + \left[\frac{u_i}{\sqrt{h_i}}\right]$$

$$\frac{y_i}{x_i} = \beta_0 \frac{1}{x_i} + \beta_1 + \beta_2 x_i + \left(\frac{u_i}{x_i}\right)$$

$$y_i^* = \beta_1 + \beta_0 x_{0i}^* + \beta_2 x_i + u_i^*$$

$$, \; y_i^* = \frac{y_i}{x_i}, \quad x_{0i}^* = \frac{1}{x_i}, \quad u_i^* = \frac{u_i}{x_i}$$

regressors are $\frac{1}{x_i}$, x_i.

c. (2) Why is the estimators in part (b) at least as efficient as the estimators in part (a)? Explain

$$Var(u_i^* | x_i) = Var\left(\frac{u_i}{x_i} \,\Big|\, x_i\right) = \frac{1}{x_i^2} Var(u_i | x_i) = \underline{\sigma^2}$$

the new estimators satisfy homoskedasticity.

$\Rightarrow$ OLS is BLUE

$\Rightarrow$ hence (b) can't be worse than (a)

(b) at least efficient as (a).

5. (9 points) The relationships between the monthly expenditure on housing (E) and household income (Y) for three age groups are as follows:

Age less than 30: $E = \alpha_1 + \beta_1 Y + u_1$

Age 31 to 55: : $E = \alpha_2 + \beta_2 Y + u_2$

Age 56 or over: : $E = \alpha_3 + \beta_3 Y + u_3$

a. (3)What is the economic interpretation of the hypothesis $\beta_1 = \beta_2 = \beta_3$?

the effect of household income (Y) on monthly expediture on housing (E) is same for different age groups (less than 30, 30 to 55, 56 or over):

if the income increase 1 \$ the expditure will all increase β \$. $(\beta_1 = \beta_2 = \beta_3 = \beta)$

b. (6) Describe step by step how the dummy variable apprpach could be used to test the hypothesis in part a. State the null and alternative hypotheses, the test statistic, its distribution, and the criterion for rejecting the null hypothesis.

Use CHOW Test

① use all data to regress $E = \alpha_0 + \beta_0 Y + u_0$

get $\hat{\alpha}_0$, $\hat{\beta}_0$, SSR_0

② regress the equations in the three groups.

get $\hat{\alpha}_i$, $\hat{\beta}_i$, SSR_i, $i = 1, 2, 3$

③ test $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_0$, H_1: H_0 is not true.

$\beta_1 = \beta_2 = \beta_3 = \beta_0$

F-statistics: $\dfrac{[SSR_0 - (SSR_1 + SSR_2 + SSR_3)]/2}{(SSR_1 + SSR_2 + SSR_3)/(n-6)} \sim F_{2, n-6}$

compute the F-statistics and the crtical value of $F_{2,n-6}$ in $\alpha\%$. If F-statistics > critical value, we can reject H_0, otherwise accept H_0.

Formula Sheet

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

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

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

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

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

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

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

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

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

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

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

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

Probability density function of a standard normal random variable:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Cumulative distribution function of a standard logistic random variable:

$$\Lambda(z) = \frac{e^z}{1 + e^z}$$

Probability density function of a standard logistic random variable:

$$\lambda(z) = \frac{e^z}{(1 + e^z)^2}$$

TABLE G.1 Cumulative Areas under the Standard Normal Distribution

z	0	1	2	3	4	5	6	7	8	9
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776

(continued)

TABLE G.1 *(Continued)*

z	0	1	2	3	4	5	6	7	8	9
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Examples: If $Z \sim \text{Normal}(0,1)$, then $P(Z \leq -1.32) = .0934$ and $P(Z \leq 1.84) = .9671$.

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

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**
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

(Row labels: **Degrees of Freedom**)

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

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 (∞) is 2.37.
Source: This table was generated using the Stata® function invFtail.