ECON 6511: Econometrics
Midterm Exam
Points: 60, Time: 210 minutes

Instructions

- Use 3 or more decimal places unless otherwise stated
- No notes or cellphones

Section A: Multiple-choice (15 points)

1. The model: $y_t = \beta_0 + \beta_1 c_t + u_t$, $t = 1, 2, \ldots, n$, is an example of a(n):

   (a) autoregressive conditional heteroskedasticity model.
   (b) static model.
   (c) finite distributed lag model.
   (d) infinite distributed lag model.

   Answer: (b)

2. Which of the following is an assumption on which time series regression is based?

   (a) A time series process follows a model that is nonlinear in parameters.
   (b) In a time series process, no independent variable is a perfect linear combination of the others.
   (c) In a time series process, at least one independent variable is a constant.
   (d) For each time period, the expected value of the error u_t, given the explanatory variables for all time periods, is positive.

   Answer: (b)

3. If an explanatory variable is strictly exogenous it implies that:

   (a) changes in the lag of the variable does not affect future values of the dependent variable.
   (b) the variable is correlated with the error term in all future time periods.
   (c) the variable cannot react to what has happened to the dependent variable in the past.
   (d) the conditional mean of the error term given the variable is zero.

   Answer: (c)

4. A study which observes whether a particular occurrence (e.g. a new policy) influences some outcome is referred to as a(n):

   (a) event study.

   (b) exponential study.

   (c) laboratory study.

   (d) comparative study.

   Answer: (a)

5. A covariance stationary time series is weakly dependent if:

   (a) the correlation between the independent variable at time t and the dependent variable at time $t+h$ goes to ∞ as $h \to 0$.

   (b) the correlation between the independent variable at time t and the dependent variable at time $t+h$ goes to 0 as $h \to \infty$.

   (c) the correlation between the independent variable at time t and the independent variable at time $t+h$ goes to ∞ as $h \to 0$.

   (d) the correlation between the independent variable at time t and the independent variable at time $t+h$ goes to 0 as $h \to \infty$.

   Answer: (d)

6. The model $y_t = e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2}$, $t = 1, 2, \ldots$, where e_t is an i.i.d. sequence with zero mean and variance σ_e^2 represents a(n):

   (a) static model.

   (b) moving average process of order one.

   (c) moving average process of order two.

   (d) autoregressive process of order two.

   Answer: (c)

7. The model $y_t = y_{t-1} + e_t$, $t = 1, 2, \ldots$ represents a:

   (a) AR(2) process.

   (b) MA(1) process.

   (c) random walk process.

   (d) random walk with a drift process.

Answer: (c)

8. If a process is said to be integrated of order one, or I(1), then:

   (a) it is covariance stationary

   (b) averages of such processes already satisfy the standard limit theorems

   (c) the first difference of the process is weakly dependent

   (d) it does not have a unit root

   Answer: (c)

9. In the presence of serial correlation:

   (a) estimated standard errors remain valid.

   (b) estimated test statistics remain valid.

   (c) estimated OLS values are not BLUE.

   (d) estimated variance does not differ from the case of no serial correlation.

   Answer: (c)

10. Which of the following identifies an advantage of first differencing a time-series?

    (a) First differencing eliminates most of the serial correlation.

    (b) First differencing eliminates most of the heteroskedastcicty.

    (c) First differencing eliminates most of the multicollinearity.

    (d) First differencing eliminates the possibility of spurious regression.

    Answer: (a)

11. A regression model exhibits unobserved heterogeneity if

    (a) there are unobserved factors affecting the dependent variable that change over time

    (b) there are unobserved factors affecting the dependent variable that do not change over time

    (c) there are unobserved factors which are correlated with the observed independent variables

    (d) there are no unobserved factors affecting the dependent variable

    Answer: (b)

12. Ordinary least squares estimation is subject to (heterogeneity) bias if

(a) the regression model exhibits heteroskedasticty

(b) the unobserved effect is correlated with the observed explanatory variables

(c) the regression model includes a lagged dependent variable

(d) the explanatory variables do not change over time

Answer: (b)

13. What should be the degrees of freedom (df) for fixed effects estimation if the data set includes N cross sectional units over T time periods and the regression model has k independent variables?

(a) $N - kT$

(b) $NT - k$

(c) $NT - N - k$

(d) $N - T - k$

Answer: (c)

14. The general approach to obtaining fully robust standard errors and test statistics in the context of panel data is known as

(a) confounding

(b) differencing

(c) clustering

(d) attenuating

Answer: (c)

15. Which of the following assumptions is needed for the usual standard errors to be valid when differencing with more than two time periods?

(a) The regression model exhibits heteroskedasticty.

(b) The differenced idiosyncratic error or Δu_{it} is uncorrelated over time.

(c) The unobserved factors affecting the dependent variable are time-constant.

(d) The regression model includes a lagged independent variable.

Answer: (b)

Section B: Written Answer (45 points)

1. **(20 points)** Consider the following model that relates wages to productivity:

$$\log(hrwage_t) = \beta_0 + \beta_1 \log(outphr_t) + u_t,$$

where the variable *hrwage* is average hourly wage in the U.S. economy, and the variable *outphr* is output per hour. The data are annual data between 1947 and 1987 with $n = 41$. The output from this regression and many other specifications are included below.

(a) Interpret (with one perfect sentence) what the estimate of β_1 in the above model means.
Answer: A one percent increase in productivity (output per hour) increases wages by 0.689 percent.

(b) **Formally** test the hypothesis that $\beta_1 = 1$ against a two-sided alternative, i.e. test whether productivity growth is fully passed onto workers in the form of higher wages.
Answer: The null and alternative hypotheses are: $H_0 : \beta_1 = 1$ and $H_1 : \beta_1 \neq 1$. The t statistic is $\frac{0.689-1}{0.039} = -7.974$ so we can reject the null hypothesis that productivity increases are fully passed onto workers. The critical value is approximately 2.021.

(c) Included in the attached output are regressions of both $\log(hrwage)$ and $\log(output)$ regressed on a simple linear time trend, t. What do these regressions suggest about the likely bias in the estimate of β_1 above? Explain.
Answer: Both the left- and right-hand side variables appear to be increasing over time, which means the coefficient on $\log(outphr_t)$ is likely to be biased.

(d) Consider a second specification that includes a time trend:

$$\log(hrwage_t) = \beta_0 + \beta_1 \log(outphr_t) + \beta_2 t + u_t$$

Interpret the coefficients on $\log(outphr_t)$ and t. Are they both statistically significant?
Answer: The coefficient on $\log(outphr_t)$ implies that a one percent increase in productivity leads to a 1.64 percent increase in wages. The coefficient on t indicates that wages decrease by 1.8 percent per year. Regarding part (c), the coefficient on $\log(outphr_t)$ in the model without t is given by $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \delta_1$ where $\hat{\delta}_1$ specifies the relationship between $\log(outphr_t)$ and t. Since $\hat{\beta}_2 < 0$ and $\hat{\delta}_1 > 0$, the coefficient on $\log(outphr_t)$ in part (c) is biased downward.

(e) Explain how you would linearly de-trend the variable $\log(hrwage)$, i.e. remove its time trend?
Answer: You could regress $\log(hrwage)$ on a linear time trend, t, and obtain the residuals from the regression. The residuals would have the trend removed.

(f) After removing time-trends, suppose the first order autocorrelations of $\log(hrwage)$ and $\log(outphr)$ are $\hat{\rho} = 0.967$ and $\hat{\rho} = 0.945$ respectively. What do these findings imply? What is an appropriate remedy?

Answer: Since the first order autocorrelations are so close to 1, these data are highly persistent (likely have unit roots). The appropriate remedy is to difference the data.

(g) Consider a model in first differences:

$$\Delta\log(hrwage_t) = \beta_0 + \beta_1 \Delta\log(outphr_t) + \Delta u_t$$

where $ghrwage = \Delta\log(hrwage_t)$ and $goutphr = \Delta\log(outphr_t)$ in the Stata output. Interpret the coefficient on $\Delta\log(outphr_t)$. **Formally** test the hypothesis that $\beta_1 = 1$ against a two-sided alternative.

Answer: The coefficient implies that a one percent increase in productivity leads to a 0.81 percent increase in wages. The null and alternative hypotheses are: $H_0 : \beta_1 = 1$ and $H_1 : \beta_1 \neq 1$. The t statistic is $\frac{0.809-1}{0.173} = -1.104$, which is less than the critical value of 2.021 so we cannot reject the null hypothesis that productivity increases are fully passed onto workers.

(h) Consider a final model:

$$ghrwage_t = \beta_0 + \beta_1 goutphr_t + \beta_2 goutphr_{t-1} + u_t$$

that includes one lag of output growth. Is the lagged value statistically significant?

Answer: Yes, the t statistic is 2.76 and the p-value is less than 0.01.

(i) Test $\beta_1 + \beta_2 = 1$ against a two-sided alternative.

Answer: The null and alternative hypotheses are: $H_0 : \theta = \beta_1 + \beta_2 = 1$ and $H_1 : \theta \neq 1$. Rearranging to obtain $\beta_1 = \theta - \beta_2$ and substituting into the regression equation we obtain:

$$ghrwage_t = \beta_0 + \theta goutphr_t + \beta_2(goutphr_{t-1} - goutphr_t) + u_t$$

Answer: The estimate of θ is 1.186 and the standard error of $\hat{\theta}$ is 0.203. The t statistic is $\frac{1.186-1}{0.203} = 0.916$, so we fail to reject the null hypothesis that productivity increases are fully passed on to workers.

(j) Does $goutphr_{t-2}$ need to be in the model? Explain.

Answer: The regression output indicates that the second lag is not statistically significant. The t statistic is only 0.41 so we do not need to include this second lag.

```
. reg lhrwage loutphr

      Source |       SS       df       MS              Number of obs =      41
-------------+------------------------------           F(  1,    39) =  311.95
       Model |  .956010689     1  .956010689           Prob > F      =  0.0000
    Residual |  .119521723    39  .00306466            R-squared     =  0.8889
-------------+------------------------------           Adj R-squared =  0.8860
       Total |  1.07553241    40  .02688831            Root MSE      =  .05536

------------------------------------------------------------------------------
     lhrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     loutphr |   .6891012    .039016    17.66   0.000     .6101839    .7680185
       _cons |  -1.534398   .1716633    -8.94   0.000     -1.88162   -1.187176
------------------------------------------------------------------------------
```

```
. reg loutphr t

      Source |       SS       df       MS              Number of obs =      41
-------------+------------------------------           F(  1,    39) =  800.97
       Model |  1.91976939     1  1.91976939           Prob > F      =  0.0000
    Residual |   .09347595    39  .002396819           R-squared     =  0.9536
-------------+------------------------------           Adj R-squared =  0.9524
       Total |  2.01324534    40  .050331133           Root MSE      =  .04896

------------------------------------------------------------------------------
     loutphr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0182881   .0006462    28.30   0.000     .0169811    .0195951
       _cons |   4.010183   .0155758   257.46   0.000     3.978678    4.041688
------------------------------------------------------------------------------
```

```
. reg lhrwage t

      Source |       SS       df       MS              Number of obs =      41
-------------+------------------------------           F(  1,    39) =  109.61
       Model |   .79327836     1   .79327836           Prob > F      =  0.0000
    Residual |  .282254052    39  .007237283           R-squared     =  0.7376
-------------+------------------------------           Adj R-squared =  0.7308
       Total |  1.07553241    40   .02688831           Root MSE      =  .08507

------------------------------------------------------------------------------
     lhrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |   .0117559   .0011229    10.47   0.000     .0094847    .0140272
       _cons |   1.246799   .0270657    46.07   0.000     1.192053    1.301544
------------------------------------------------------------------------------
```

```
. reg lhrwage loutphr t

      Source |       SS       df       MS              Number of obs =      41
-------------+------------------------------           F(  2,    38) =  641.22
       Model |  1.04458064     2  .522290318           Prob > F      =  0.0000
    Residual |  .030951776    38   .00081452           R-squared     =  0.9712
-------------+------------------------------           Adj R-squared =  0.9697
       Total |  1.07553241    40   .02688831           Root MSE      =  .02854

------------------------------------------------------------------------------
     lhrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     loutphr |   1.639639   .0933471    17.56   0.000     1.450668    1.828611
           t |    -.01823   .0017482   -10.43   0.000     -.021769   -.0146909
       _cons |  -5.328454   .3744492   -14.23   0.000    -6.086487   -4.570421
------------------------------------------------------------------------------
```

```
. reg ghrwage goutphr

      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  1,    38) =   21.77
       Model |  .006255013     1  .006255013           Prob > F      =  0.0000
    Residual |   .01091799    38  .000287316           R-squared     =  0.3642
-------------+------------------------------           Adj R-squared =  0.3475
       Total |  .017173003    39  .000440333           Root MSE      =  .01695

------------------------------------------------------------------------------
     ghrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     goutphr |    .809316   .1734537     4.67   0.000     .4581773    1.160455
       _cons |  -.0036621     .00422    -0.87   0.391    -.0122051    .0048808
------------------------------------------------------------------------------
```

```
. reg ghrwage goutphr goutph_1

      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  2,    36) =   17.51
       Model |  .008456778     2  .004228389           Prob > F      =  0.0000
    Residual |  .008692555    36   .00024146           R-squared     =  0.4931
-------------+------------------------------           Adj R-squared =  0.4650
       Total |  .017149333    38  .000451298           Root MSE      =  .01554

------------------------------------------------------------------------------
     ghrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     goutphr |   .7283641   .1672223     4.36   0.000     .3892217    1.067507
    goutph_1 |   .4576351   .1656126     2.76   0.009     .1217571     .793513
       _cons |   -.010425   .0045439    -2.29   0.028    -.0196404   -.0012096
------------------------------------------------------------------------------
```

```
. gen new_var = goutph_1 - goutphr
(2 missing values generated)

. reg ghrwage goutphr new_var

      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  2,    36) =   17.51
       Model |  .008456778     2  .004228389           Prob > F      =  0.0000
    Residual |  .008692555    36   .00024146           R-squared     =  0.4931
-------------+------------------------------           Adj R-squared =  0.4650
       Total |  .017149333    38  .000451298           Root MSE      =  .01554

------------------------------------------------------------------------------
     ghrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     goutphr |   1.185999   .2031416     5.84   0.000     .7740089     1.59799
     new_var |   .4576351   .1656126     2.76   0.009     .1217571     .793513
       _cons |   -.010425   .0045439    -2.29   0.028    -.0196404   -.0012096
------------------------------------------------------------------------------
```

```
. reg ghrwage goutphr goutph_1 goutph_2
```

Source	SS	df	MS
Model	.008048615	3	.002682872
Residual	.007575626	34	.000222813
Total	.015624241	37	.000422277

Number of obs =	38
F(3, 34) =	12.04
Prob > F =	0.0000
R-squared =	0.5151
Adj R-squared =	0.4724
Root MSE =	.01493

ghrwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
goutphr	.7464273	.1615037	4.62	0.000	.4182123	1.074642
goutph_1	.3740461	.1665312	2.25	0.031	.0356139	.7124783
goutph_2	.0653486	.1597067	0.41	0.685	-.2592144	.3899116
_cons	-.0112838	.0048409	-2.33	0.026	-.0211217	-.0014458

2. **(15 points)** Consider the following model that relates traffic deaths per 100 million miles to two policies that attempt to reduce drunk driving. The two laws are *open container laws* – which make it illegal for travel with an open alcoholic drink and *administrative per se laws* – which enable courts to suspend driver licenses immediately after an arrest. Consider a cross-sectional model relating the death rate (*dthrte*) in a state to dummy variables indicating whether the state has any of these laws (*open* and *admin*):

$$dthrte_i = \beta_0 + \beta_1 open_i + \beta_2 admin_i + u_i.$$

This equation is estimated separately for 1985 where *open*85 and *admin*85 are the dummy variables for 1985. Results are provided below.

(a) How do you interpret the coefficients on *open*85 and *admn*85? Do the signs of the coefficients make sense? Explain.

Answer: The adoption of open contain laws reduces traffic deaths by 0.213 per 100 million miles, while the adoption of an administrative per se law increases traffic deaths by 0.229 deaths. The sign of the first coefficient makes sense but the second does not. However, neither of the coefficients are statistically significant at conventional levels.

(b) Are the estimates likely to be unbiased? Explain.

Answer: There are many factors that are omitted from the simple cross sectional approach. States decide through legislative processes whether they need such laws and these decisions are likely to be related to the number of drunk driving fatalities. Therefore, the coefficients are likely to be biased.

(c) Consider an alternative unobserved effects model using panel data with $T = 2$ periods:

$$dthrte_{it} = \beta_0 + \beta_1 d2_t + \beta_2 open_{it} + \beta_3 admin_{it} + a_i + u_{it}.$$

where $d2_t$ is a dummy variable equal to 1 in the second period. List two factors that might be included in a_i.

Answer: a_i includes characteristics of state i that do not change between periods. Attitudes toward and average quantity of alcohol consumed are likely factors. Age and gender distributions may also affect the death rates. These are unlikely to change much across time within a state.

(d) Explain how we can obtain unbiased or at least consistent estimates of β_1 and β_2 assuming we have data from states before and after these laws were passed.

Answer: A first differenced model would allow us to control for characteristics contained in a_i.

(e) Write down the Stata code that you would use to generate the change in *dthrte* between two periods assuming the data was organized so that you had one observation per state and period.

Answer: The following command in Stata should work: "gen delta_dthrte = dthrte - dthrte[_n-1]."

(f) Results from the following model:

$$\Delta dthrte_i = \beta_0 + \beta_1 \Delta open_i + \beta_2 \Delta admin_i + \Delta u_i.$$

using two years of data (1985 and 1990) are included below. In 1985, 19 states had open container laws, while 22 states had such laws in 1990. In 1985, 21 states had per se laws; the number had grown to 29 by 1990. Interpret the estimated coefficient on *copen*, which is equivalent to $\Delta open_i$ in the above model. Interpret the estimate of β_0.

Answer: The coefficient indicates that adopting an open container law lowered the traffic fatality rate by 0.420 (which is quite large since the average death rate was 2.7 in 1985). The estimate of $\beta_0 = 0.497$ indicates that the death rate fell by almost 0.5 fatalities between the two time periods regardless of whether they adopted either of the two laws.

(g) Does the sign on *cadmin* make sense? Is it statistically significant?

Answer: The sign of *cadmin* makes sense (i.e. the increased likelihood license suspesion lowers fatalities) but it is not statistically significant.

(h) What do we need to assume for the estimates in part (f) to be "correct?"

Answer: We need to assume that there were no other laws or changes that occurred during the same time period that might have also affected fatalities, such as the adoption of seat belt laws, motorcycle helmet laws, and maximum speed limits.

```
. reg dthrte85 open85 admn85
```

Source	SS	df	MS
Model	1.17595166	2	.58797583
Residual	17.0240489	48	.354667685
Total	18.2000005	50	.36400001

Number of obs =	51
F(2, 48) =	1.66
Prob > F =	0.2013
R-squared =	0.0646
Adj R-squared =	0.0256
Root MSE =	.59554

dthrte85	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
open85	-.2131043	.1725004	-1.24	0.223	-.5599395	.1337309
admn85	.229711	.1694617	1.36	0.182	-.1110146	.5704366
_cons	2.684805	.1257889	21.34	0.000	2.431889	2.93772

```
. reg cdthrte copen cadmn
```

Source	SS	df	MS
Model	.762579785	2	.381289893
Residual	5.66369475	48	.117993641
Total	6.42627453	50	.128525491

Number of obs =	51
F(2, 48) =	3.23
Prob > F =	0.0482
R-squared =	0.1187
Adj R-squared =	0.0819
Root MSE =	.3435

cdthrte	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
copen	-.4196787	.2055948	-2.04	0.047	-.8330547	-.0063028
cadmn	-.1506024	.1168223	-1.29	0.204	-.3854894	.0842846
_cons	-.4967872	.0524256	-9.48	0.000	-.6021959	-.3913784

3. **(15 points)** Consider the following model of rental prices in one of 64 college towns for the years 1980 and 1990. The idea is to see whether a stronger presence of students affects rental rates. The unobserved effects model is:

$$\log(rent_{it}) = \beta_0 + \delta_0 y90_t + \beta_1 \log(pop_{it}) + \beta_2 \log(avginc_{it}) + \beta_3 pctstu_{it} + a_i + u_{it},$$

where *pop* is city population, *avginc* is average income, and *pctstu* is student population as a percentage (0 to 100) of city population (during the school year). Results by pooled OLS, random effects, and fixed effects are given below.

(a) Using pooled OLS, what is the interpretation of the 1990 dummy variable? What does the estimate of β_3 imply?

Answer: Nominal rental rates are 26.2 percent higher in 1990 compared to 1980. A 10 percent increase in the proportion of students increases rental rates by 5 percent.

(b) Are the standard errors for the pooled OLS estimates likely to be valid? Explain.

Answer: Since there are likely to be unobserved factors contained in a_i, there will be serial correlation so the standard errors will not be valid.

(c) What does the estimate of β_3 using fixed effects imply? Why is it so different to the pooled OLS estimate?

Answer: A 10 percent increase in the proportion of students increases rental rates by 11 percent (this is double the pooled OLS estimate). The fixed effect estimate controls for omitted variables that likely biased the coefficient on *pctstu* down in the pooled OLS regression.

(d) List one factor that might be contained in a_i and explain how it might be correlated with *pctstu*.

Answer: The presence of public transportation might be positively correlated with rents but negatively correlated with the proportion of students living in the town. Omitting this (dummy) variable would bias the coefficient on *pctstu* downwards.

(e) What does the estimate of β_3 under random effects imply about the likely magnitude of $\hat{\theta}$ (the quasi demeaning parameter)?

Answer: Since coefficients are quite different to the fixed effects estimates, $\hat{\theta}$ is unlikely to be too close to one (in fact, it is 0.61).

(f) Which model do you prefer and why?

Answer: None of the variables are constant over time so we can estimate coefficients for each variable using fixed effects. The ability to estimate coefficients on time constant variables is the chief advantage of the random effects model. Since we are always worried about omitted variable bias, the fixed effects model should be preferred.

(g) Suppose you estimated the model using OLS but included dummy variables for each city as opposed to time demeaning or quasi demeaning the data. What would you expect to get for your estimate of β_3?

Answer: This will give identical coefficients and standard errors to the fixed effects regression so the estimated coefficient on *pctstu* will be 0.011.

	Pooled OLS	Fixed Effects	Random Effects
$\log(pop)$	0.041	0.072	0.056
	(0.023)	(0.088)	(0.029)
$\log(avginc)$	0.571**	0.310**	0.447**
	(0.053)	(0.066)	(0.052)
$pctstu$	0.005**	0.011**	0.005**
	(0.001)	(0.004)	(0.001)
$y90$	0.262**	0.386**	0.323**
	(0.035)	(0.037)	(0.029)
Constant	-0.569	1.409	0.445
	(0.535)	(1.167)	(0.566)
Observations	128	128	128
R^2	0.861	0.977	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$

Formulae

- For simple regression model: $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$
- Estimated slope parameter when regression equation passes through origin: $\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$
- $SSR = \sum_{i=1}^{n} \hat{u}_i^2$
- $SSE = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$
- $SST_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$
- $R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$
- For simple regression model: $se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$
- $\widehat{Var}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}$
- t statistic: $t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} = t_{df}$
- 95% confidence interval: $P\Big(\hat{\beta}_j - c_{0.05} \cdot se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c_{0.05} \cdot se(\hat{\beta}_j)\Big) = 0.95$
- F statistic $= \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$
- R^2 form of F statistic $= \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n-k-1)}$
- $\overline{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$
- Chow test statistic: $F = \frac{[SSR_P - (SSR_1 + SSR_2)]/(k+1)}{(SSR_1 + SSR_2)/[n - 2(k+1)]}$

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

(continued)

z	0	1	2	3	4	5	6	7	8	9
−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
−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.

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.

F Values for $\alpha = 0.05$

	d_1								
d_2	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.3	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88