Question 1: Multiple Choice (20 marks, 2 marks per question)

Answer Question 1 using the multiple choice form provided.

1. Which of the following is not one the broad categories in which econometric analyses typically fall under?

   a. Testing economic theory
   b. Evaluating and forecasting the impact of policy
   c. Exploring new phenomena using data
   d. All of the above are broad categories for econometric analyses

2. Suppose you have a sample average of $\bar{Y} = 12$ with sample variance $s_Y^2 = 2$ and sample size $n = 25$. What is the 90% confidence interval for the population mean?

   a. [7.05,16.95]
   b. [11.53,12.47]
   c. [10.10,12.90]
   d. [4.26,19.74]

3. You estimate a single linear regression:

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

   and obtain a 99% confidence interval for β_1 of [-0.50,0.10]. Which of the following statements is necessarily true?

   a. The OLS regression coefficient estimate must be $\hat{\beta}_1 = -0.25$
   b. We fail to reject the null that β_1 equals 0 in a two-sided test at the 5% level
   c. Because the confidence interval is not centered around 0, the distribution of X_i is skewed
   d. None of the above are necessarily true

4. Which of the following is correct about the value of $\bar{R}^2$ in a multiple linear regression model?

   a. It is bounded between 0 and 1
   b. It can be larger than R^2
   c. It can be negative
   d. It strictly increases as sample size n grows

5. Suppose you run the following regression:

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

Further suppose there is an omitted variable Z_i that is positively correlated with Y_i and negatively correlated with X_i. Finally, suppose we obtained a regression coefficient $\hat{\beta}_1 < 0$. Based on the omitted variable Z_i, which of the following is true?

   a. There is no omitted variable bias
   b. $\hat{\beta}_1$ exhibits a negative omitted variable bias, causing it to be too large in magnitude
   c. $\hat{\beta}_1$ exhibits a positive omitted variable bias, causing it to be too small in magnitude
   d. Unable to determine whether there is any omitted variable bias based on the information provided.

6. Suppose you run the following regression:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

Which is the correct interpretation of β_1?

   a. A 1% increase in X_i has an associated 1% increase in Y_i
   b. A 1 unit increase in X_i has an associated 1% increase in Y_i
   c. A 1% increase in X_i has an associated 1 unit increase in Y_i
   d. A 1 unit in X_i has an associated 1% increase in Y_i

7. Suppose your OLS estimates suffer from imperfect multicollinearity, but where the OLS assumptions hold. Your OLS estimates will be:

   a. biased and efficient
   b. unbiased and inefficient
   c. biased and inefficient
   d. unbiased and efficient

8. What problem does classical measurement error in a regressor create for regression analysis?

   a. it inflates the magnitudes of regression coefficients
   b. it causes regression coefficients to be biased toward 0
   c. it leads to severe multicollinearity between regressors
   d. it creates reverse causality in regression coefficients

9. Consider the following estimation results from a polynomial regression:

$$Y_i = \underset{(12.72)}{42} + \underset{(14.50)}{12.9}\, X_i - \underset{(17.92)}{0.34}\, X_i^2 + \underset{(8.22)}{0.01} X_i^3; \qquad \bar{R}^2 = 0.23, SER = 42.11$$

What is the conclusion that can be drawn from these regression results?

a. There must be omitted variable bias because only X_i is included in some form in all of the regressors

b. The model's adjusted R-Squared is very low, implying that more terms in the polynomial should be included to better fit the data

c. A scatter plot of Y_i on X_i is likely to exhibit less than 2 inflection points in curvature in the relationship, which can give rise to imperfect multicollinearity in this regression

d. The insignificance of some of the regressors is direct evidence of a break down in the conditional mean independence assumption, which is necessary for obtaining unbiased and efficient regression coefficient estimates

10. When testing a joint hypothesis with a multiple linear regression model, you should:

a. use t-statistics for each hypothesis and reject the joint null hypothesis if all of the restrictions fail

b. use the F-statistic and reject at least one of the hypotheses if the statistic exceeds the critical value

c. use t-statistics for each hypothesis and reject the null hypothesis once the statistic exceeds the critical value for a single hypothesis

d. use the F-statistic and reject all the hypotheses if the statistic exceeds the critical value

Question 2: Short Answer Questions (20 marks)

Answer Questions 2-5 using exam booklets. You do not have to answer the questions in the order in which they are asked.

a. Consider the following estimated single linear regression model using a sample of $n = 100$ observations:

$$Y = \underset{(0.2)}{10} - \underset{(0.1)}{0.5} \ln(X); \qquad \bar{R}^2 = 0.15$$

where the coefficient estimates' standard errors are in parantheses under the estimates. Interpret the regression coefficient on $\ln(X)$ and compute the 99% confidence interval on the coefficient. (6 marks)

b. Write down the formula of the Homoskedasticity-only F-statistic, explain the steps you would take to compute it, state the null and alternative hypotheses for the test that corresponds to the statistic, and briefly comment on the connection between the regression R^2 and F-statistic for joint hypothesis testing highlighted by the Homoskedasticity-only F-statistic. (6 marks)

c. Suppose you estimate a single linear regression model $Y = \beta_0 + \beta_1 X + u$ and obtain an OLS regression coefficient estimate of $\hat{\beta}_1$. Holding all other aspects of the data fixed except sample n size, explain why you are more likely to find that the OLS regression coefficient estimate $\hat{\beta}_1$ is statistically significantly different from 0 using a two-sided hypothesis test as n grows. Make explicit use of appropriate equations and/or formulas in answering the question. (8 marks)

Question 3: MetricsBars (20 marks)

A chocolate bar retailer asks you to evaluate the impact of a marketing campaign they ran for their MetricsBars chocolate bar. They provide you with a raw dataset called **marketing.csv**, which contains the following variables:

> $sales_i$: sales of MetricsBars in market i in dollars ($)
>
> $campaign_i$: a dummy variable equalling 1 if the marketing campaign was run in market i, and is equal to 0 otherwise.
>
> $income_i$: average individual income in market i in dollars ($)

These data are provided for a sample of $n = 1000$ markets, where 500 markets were randomly chosen to have the marketing campaign, and the other 500 markets did not have the marketing campaign. The sample average of $sales_i$ is $1,000,000 and the sample average of $income_i$ is $50,000.

MetricBars wants to inform the following question:

> *If household income increases from $50,000 to $70,000, how much more effective will the marketing campaign be at increasing MetricsBars sales in percentage terms?*

Starting from the raw data in **marketing.csv**, write down the pseudo-code in R you would develop to provide a relevant 95% confidence interval that empirically informs this question. List the steps you would include in your code, and if it helps in describing your answer, you may state explicit R code though this is not necessary for obtaining full marks. Be precise and explicitly describe any variable scaling you would use, regressions run in your code, hypothesis tests required, test statistics used, or any other calculations necessary for computing the 95% confidence interval.

Question 4: Wages, Experience, and Education (20 Marks)

The Department of Jobs and Small Business has approached you to study the relationship between wages, experience, and education. They have provided you a dataset which contains the following information from a random sample of $n = 801$ individuals:

> $wage_i$: hourly wage earned by individual i in dollars ($)
>
> $exper_i$: experience of individual i working measured as the number of years they have been working in the labour market
>
> $degree_i$: a dummy variable equalling 1 if individual i has a university degree and equals 0 if they do not have a university degree.
>
> age_i: age of individual i in years
>
> $urban_i$: dummy variable equalling 1 if individual i lives in an urban location, and equals 0 if they live in a regional location.

The Department wants to understand how wages change with work experience in the labour market. Figure 1 on the next page produces summary statistics for these data, and regression output from R for three different regressions that focus on the relationship between wages and experience. Based on this output, answer the following questions. Throughout assume a 5% level of significance in conducting hypothesis tests.

a. What percentage of individuals in the data set have a university degree? (2 marks)

b. Interpret the statistical significance, sign and magnitude of the regression coefficient estimate on $exper_i$ in Regression 1. (3 marks)

c. The regression coefficient on $exper_i$ changes substantially between Regressions 1 and 2. Carefully explain what might drive this large change in the regression coefficient on $exper_i$ between Regressions 1 and 2. (4 marks)

d. Interpret the statistical significance, sign and magnitude of the regression coefficient estimate on $degree_i$ in Regression 3. (3 marks)

e. Based on the regression output in Figure 1, is there evidence of a nonlinear relationship between $wage_i$ and $exper_i$? Carefully state the null and alternative for the relevant hypothesis test for conducting this test, and highlight what regression coefficient estimate(s), t-statistic(s), and p-value(s) in Figure 1 allow you to conduct such a test for a nonlinear relationship. (3 marks)

f. Based on Regression 3, holding age and urban status fixed, for an individual starting with 0 years experience and without a university degree, how many years will they have to work in the labour market to "catch up" in terms of expected wages to an individual with 0 years experience but with a university degree? Work with 3 digits after the decimal in conducting your calculations. (5 marks)

Figure 1: Summary Statistics and Estimation Results for the Wage Regressions

```
##### SUMMARY STATISTICS #####

> mydata=read.csv(file="q4_wage_dat.csv")

> summary(mydata)
      wage            exper             age            degree           urban       
 Min.   :34.85   Min.   : 8.00   Min.   :22.75   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:41.62   1st Qu.:13.00   1st Qu.:28.50   1st Qu.:0.0000   1st Qu.:0.0000  
 Median :56.75   Median :14.00   Median :30.25   Median :1.0000   Median :0.0000  
 Mean   :52.54   Mean   :14.39   Mean   :30.31   Mean   :0.7141   Mean   :0.4931  
 3rd Qu.:58.49   3rd Qu.:16.00   3rd Qu.:32.00   3rd Qu.:1.0000   3rd Qu.:1.0000  
 Max.   :63.86   Max.   :22.00   Max.   :37.75   Max.   :1.0000   Max.   :1.0000  


##### REGRESSION 1 #####

> reg1=lm(wage~exper,data=mydata)

> coeftest(reg1, vcov = vcovHC(reg1, "HC1"))

t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 46.61942    2.20113 21.1798 < 2.2e-16 ***
exper        0.41174    0.15095  2.7277  0.006517 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



##### REGRESSION 2 #####

> reg2=lm(wage~exper+age,data=mydata)

> coeftest(reg2, vcov = vcovHC(reg2, "HC1"))

t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 37.80714    3.68366 10.2635 < 2.2e-16 ***
exper        0.10760    0.17977  0.5985  0.549657    
age          0.43513    0.14670  2.9660  0.003107 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



##### REGRESSION 3 #####

> mydata$exper_sq=mydata$exper*mydata$exper

> reg3=lm(wage~exper+exper_sq+age+degree+urban,data=mydata)

> coeftest(reg3, vcov = vcovHC(reg3, "HC1"))

t test of coefficients:

              Estimate Std. Error  t value Pr(>|t|)    
(Intercept) 21.0229602  1.3044264  16.1166  < 2e-16 ***
exper        0.3823205  0.1760285   2.1719  0.03016 *  
exper_sq    -0.0143891  0.0062149  -2.3153  0.02085 *  
age          0.5014108  0.0170532  29.4028  < 2e-16 ***
degree      18.0537071  0.0786755 229.4705  < 2e-16 ***
urban        1.9569213  0.0727571  26.8966  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question 5: Understanding the Demand for Sunscreen (20 Marks)

The Department of Health wants to understand the relationship between sunscreen demand month-to-month, and the number of tourists visiting Australia. To conduct this analysis, you have been provided the following dataset:

$sunscreen_t$: total bottles of sunscreen sold in terms of 1000s of bottles in month t

$lag\_sunscreen_t$: one-month lag of $sunscreen_t$ (e.g., $lag\_sunscreen_t = sunscreen_{t-1}$)

$tourists_t$: total number of tourists visiting Australia in terms of 1000s of tourists in month t

$lag\_tourists_t$: one-month lag of $tourists_t$ (e.g., $lag\_tourists_t = tourists_{t-1}$)

$month_t$: month of year for month t, taking on one of 12 values in the following list: {Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec}

You have this information for $T = 135$ months. Figures 2 and 3 on the next two pages respectively present time series plots and regression output generated by R that analyzes these data. Based on this information, please answer the following questions:

a. Based on Figure 2 explain whether $sunscreen_t$ appears to be stationary and whether it exhibits seasonality. Similarly, based on Figure 2 explain whether $tourists_t$ appears to be stationary and whether it exhibits seasonality. (2 marks)

b. Explain what the residuals from in Regression 1 in Figure 3 would represent compared to the raw $sunscreen_t$ time series in the data set. (2 marks)

<u>Note</u>: For the remainder of the question, note that Figure 3 is displayed over pages 11 and 12 below.

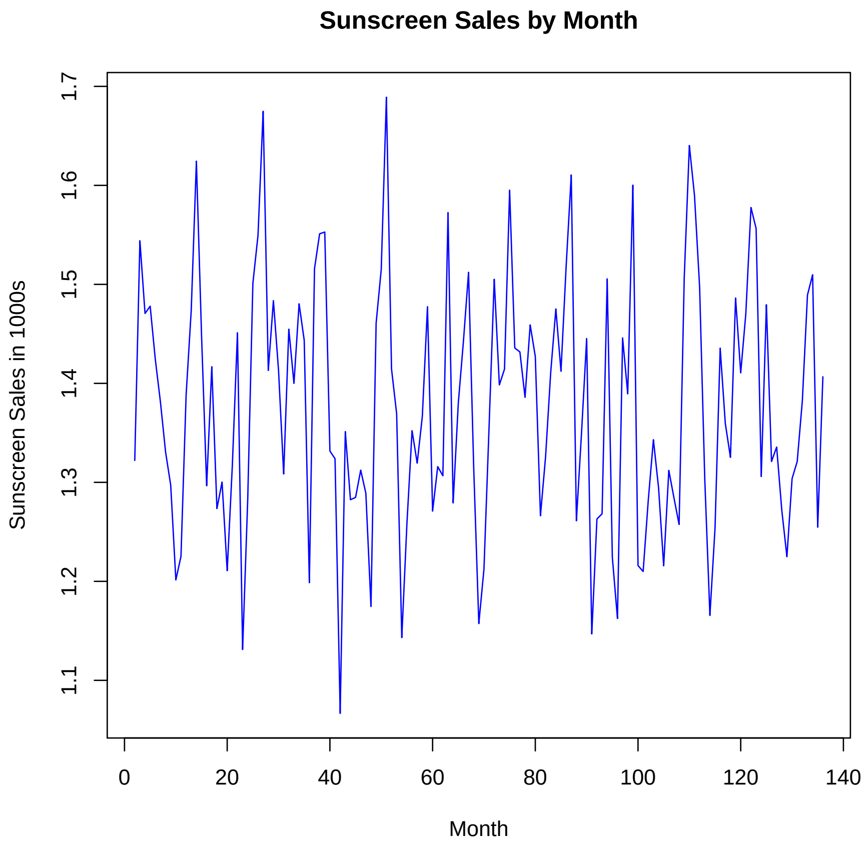
c. Explain how Regression 1 avoids a potential dummy variable trap based on the dummy variables created in the code, and the dummy variables included in the regression. (2 marks)

d. Suppose that the last data point at $T = 135$ in the sample is January, has a value of $sunscreen_{135} = 8$, and a value of $tourists_{135} = 12$. Based on Regression 2 in Figure 3, what would be your out-of-sample forecast for $sunscreen_t$ in period $T = 136$ and the 95% forecast interval assuming IID normal errors in the regression equation in Regression 2? Work with 3 digits after the decimal in conducting your calculations. (4 marks)

e. Compute the Bayes-Schwartz Information Criterion for Regressions 1 and 2 in Figure 3 and explain which is the preferable time series model based on this criterion. Work with 3 digits after the decimal in conducting your calculations. (6 marks)

f. Conduct a Granger Causality test to determine whether $tourists_t$ "Granger causes" $sunscreen_t$ based on the regression results in Regression 2 in Figure 3. Assume a 5% level of significance in conducting the test. Work with 3 digits after the decimal in conducting your calculations. (4 marks)

Figure 2: Time Series Plots of Sunscreen Sales and Tourists by Month



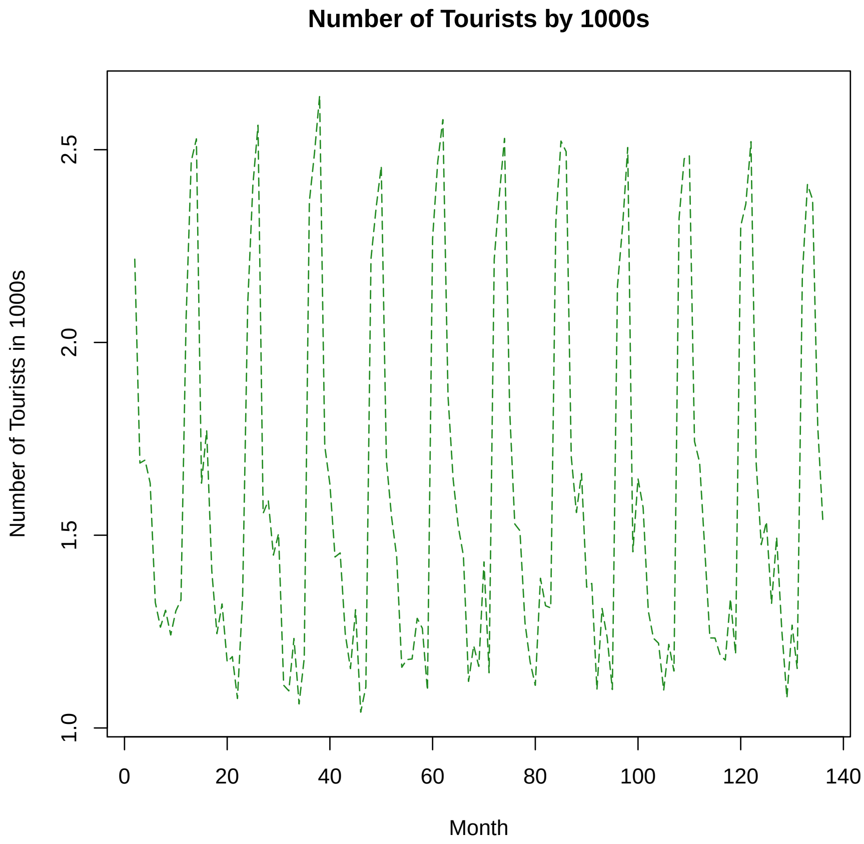


Figure 3: Estimation Results for the Time Series Regressions

```
> mydata=read.csv(file="q5_sunscreen_dat.csv")

##### CREATING DUMMY VARIABLES #####

> mydata$dJan=as.numeric(mydata$month=="Jan")
> mydata$dFeb=as.numeric(mydata$month=="Feb")
> mydata$dMar=as.numeric(mydata$month=="Mar")
> mydata$dApr=as.numeric(mydata$month=="Apr")
> mydata$dMay=as.numeric(mydata$month=="May")
> mydata$dJun=as.numeric(mydata$month=="Jun")
> mydata$dJul=as.numeric(mydata$month=="Jul")
> mydata$dAug=as.numeric(mydata$month=="Aug")
> mydata$dSep=as.numeric(mydata$month=="Sep")
> mydata$dOct=as.numeric(mydata$month=="Oct")
> mydata$dNov=as.numeric(mydata$month=="Nov")
> mydata$dDec=as.numeric(mydata$month=="Dec")



##### REGRESSION 1 ESTIMATES AND MODEL FIT #####

> reg1=lm(sunscreen~dJan+dFeb+dMar+dApr+dMay+dJun+dJul+dAug+dSep+dOct+dNov,data=mydata)

> coeftest(reg1, vcov = vcovHC(reg1, "HC1"))

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  1.318972   0.036075 36.5615 < 2.2e-16 ***
dJan         0.134728   0.040205  3.3510 0.0010698 **
dFeb         0.174234   0.048269  3.6096 0.0004446 ***
dMar         0.238514   0.048816  4.8860 3.137e-06 ***
dApr         0.041750   0.044544  0.9373 0.3504455
dMay         0.065165   0.044312  1.4706 0.1439548
dJun        -0.013259   0.053756 -0.2467 0.8055894
dJul         0.012884   0.046883  0.2748 0.7839271
dAug         0.011853   0.043214  0.2743 0.7843192
dSep        -0.036224   0.041502 -0.8728 0.3844647
dOct         0.026185   0.046862  0.5588 0.5773346
dNov         0.012821   0.049966  0.2566 0.7979189
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


# Regression 1 Fit
> summary(reg1)
# (note: regression output for reg1 with
# homoskedastic standard errors omitted)

Residual standard error: 0.1 on 123 degrees of freedom
Multiple R-squared:  0.4201,       Adjusted R-squared:  0.3682
F-statistic:   8.1 on 11 and 123 DF,  p-value: 1.63e-10
```

Estimation Results for the Time Series Regressions (Figure 3 continued)

```
##### REGRESSION 2 ESTIMATES AND MODEL FIT #####

> reg2=lm(sunscreen~lag_sunscreen+lag_tourists+
+           dJan+dFeb+dMar+dApr+dMay+dJun+dJul+dAug+dSep+dOct+dNov,data=mydata)

> coeftest(reg2, vcov = vcovHC(reg2, "HC1"))

t test of coefficients:

                Estimate Std. Error t value  Pr(>|t|)
(Intercept)    0.9413073  0.1013951  9.2836 8.207e-16 ***
lag_sunscreen  0.1776939  0.0767075  2.3165   0.02221 *
lag_tourists   0.1183732  0.0476024  2.4867   0.01426 *
dJan           0.0145053  0.0609275  0.2381   0.81223
dFeb           0.0281326  0.0610873  0.4605   0.64596
dMar           0.0560163  0.0787568  0.7113   0.47829
dApr          -0.0582770  0.0487922 -1.1944   0.23466
dMay           0.0104385  0.0456310  0.2288   0.81944
dJun          -0.0606499  0.0517203 -1.1727   0.24324
dJul           0.0011059  0.0449387  0.0246   0.98041
dAug           0.0049781  0.0426757  0.1166   0.90733
dSep          -0.0348500  0.0392062 -0.8889   0.37582
dOct           0.0310750  0.0436303  0.7122   0.47769
dNov           0.0057216  0.0480600  0.1191   0.90543
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


# Regression 2 Fit
> summary(reg2)
# (note: regression output for reg2 with
# homoskedastic standard errors omitted)

Residual standard error: 0.09623 on 121 degrees of freedom
Multiple R-squared:  0.4721,    Adjusted R-squared:  0.4153
F-statistic: 8.323 on 13 and 121 DF,  p-value: 8.566e-12
```

END OF EXAMINATION

Formula Sheet

Expected Values, Variances, Correlation

$E(c) = c$

$E(cx) = cE(x)$

$E(a + cx) = a + cE(x)$

$E(x + y) = E(x) + E(y)$

$E(c_1 x + c_2 y) = c_1 E(x) + c_2 E(y)$

$var(x) = \sigma^2 = E(x - E(x))^2$

$std(x) = \sigma = \sqrt{E(x - E(x))^2}$

$var(a + cx) = c^2 var(x)$

$cov(x, y) = E[(x - E(x))(y - E(y))]$

$corr(x, y) = \rho = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}$

$P(y = y_1 | x = x_1) = \frac{P(x=x_1, y=y_1)}{p(X=x_1)}$

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

$var(\bar{Y}) = \frac{\sigma_Y^2}{n}$

$std(\bar{Y}) = \frac{\sigma_Y}{\sqrt{n}}$

$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{N} (y_i - \bar{y})^2$

$s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^{N} (y_i - \bar{y})^2}$

$SE(\bar{y}) = \frac{s_y}{\sqrt{n}}$

$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

$r_{xy} = \frac{s_{xy}}{s_x s_y}$

Logarithms

$x = \ln(e^x)$

$\frac{d \ln(x)}{dx} = \frac{1}{x}$

$\ln(1/x) = -\ln(x)$

$\ln(ax) = \ln(a) + \ln(x)$

$\ln(x/a) = \ln(x) - \ln(a)$

$\ln(x^a) = a \ln(x)$

$\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x}$ (approximately equal for small Δx)

Quadratic Formula

The solution to the quadratic equation:

$$ax + bx^2 + c = 0$$

where a, b, and c are constants can be computed by the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Calculus

x^* that maximizes (minimizes) a strictly concave (convex) function, $f(x)$, solves $\frac{df(x)}{dx} = 0$

OLS Estimator for Single Linear Regression

$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s_{XY}}{s_X^2}$

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

$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{var((X_i - \mu_X)u_i))}{(var(X_i))^2}$

$\sigma^2_{\hat{\beta}_0} = \frac{1}{n}\frac{var(H_i u_i)}{(E(H_i^2))^2}$; where $H_i = 1 - (\frac{\mu_X}{E(X_i^2)})X_i$

$\hat{\beta}_1 \rightarrow \beta_1 + \rho_{Xu}\frac{\sigma_u}{\sigma_X}$

Testing Differences in Means

$H_0 : \mu_w - \mu_m = d_0; \quad vs. \quad H_1 : \mu_w - \mu_m \neq d_0$

$SE(\bar{Y}_w - \bar{Y}_m) = \sqrt{s_w^2/n_w + s_m^2/n_m}$

$t^{act} = \frac{(\bar{Y}_w - \bar{Y}_m) - d_0}{SE(\bar{Y}_w - \bar{Y}_m)}$

Single Hypothesis Testing in Regression Models

$t^{act} = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$

$H_0 : \beta_1 = \beta_{1,0}$ vs. $H_1 : \beta_1 \neq \beta_{1,0}$, p-value $= 2\Phi(-|t^{act}|)$

$H_0 : \beta_1 = \beta_{1,0}$ vs. $H_1 : \beta_1 < \beta_{1,0}$, p-value $= \Phi(t^{act})$

$H_0 : \beta_1 = \beta_{1,0}$ vs. $H_1 : \beta_1 > \beta_{1,0}$, p-value $= 1 - \Phi(t^{act})$

t^α is the critical value for a two-sided test with α significance level

$\alpha = 2\Phi(-|t^\alpha|)$

$(1-\alpha)$ CI: $[\hat{\beta}_1 - t^\alpha SE(\hat{\beta}_1), \hat{\beta}_1 + t^\alpha SE(\hat{\beta}_1)]$

For testing means, replace β with μ_X and $\hat{\beta}$ with $\bar{X}$

Joint Hypothesis Testing in Regression Models

$H_0 : \beta_j = \beta_{j,0},\ \beta_m = \beta_{m,0}, \dots$ for a total of q restrictions

H_1 : one or more of the q restrictions under H_0 does not hold

F-statistic for the test is distributed $F_{q,n-k-1}$

p-value $= \Pr[F_{q,n-k-1} > F^{act}] = 1 - G(F^{act}; q, n-k-1)$

For $q = 2$ restrictions, relationship between F-statistic and individual t-statistics for testing coefficients jointly equal 0:

$$F = \frac{1}{2}\left(\frac{(t_1^{act})^2 + (t_2^{act})^2 - 2\hat{\rho}_{t_1^{act},t_2^{act}} t_1^{act} t_2^{act}}{1 - \hat{\rho}_{t_1^{act},t_2^{act}}}\right)$$

Homoskedasticity–Only F-statistic with q restrictions

$$F^{act} = \frac{(SSR_{restricted} - SSR_{unrestricted})/q}{SSR_{unrestricted}/(n-k-1)} = \frac{(R^2_{unrestricted} - R^2_{restricted})/q}{(1 - R^2_{unrestricted})/(n-k-1)}$$

$F^{act}_{1,n-k-1} = (t^{act})^2$ if $q = 1$ restriction

Goodness of Fit in Regression Models

$SSR = \sum_{i=1}^n \hat{u}_i^2; \qquad ESS = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2; \qquad TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2$

$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$

$SER = s_{\hat{u}} = \sqrt{s^2_{\hat{u}}},\ s^2_{\hat{u}} = \frac{SSR}{n-k-1}$

$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{s^2_{\hat{u}}}{s^2_Y}$

Nonlinear Regression Partial Effects, Standard Errors, and CIs

$E[Y|X_1, X_2, \dots, X_k] = f(X_1, X_2, \dots, X_k)$

$\Delta\hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \dots, X_k) - \hat{f}(X_1, X_2, \dots, X_k); \qquad SE(\Delta\hat{Y}) = \frac{|\Delta\hat{Y}|}{\sqrt{F}}$

$(1-\alpha)$ CI: $[\Delta\hat{Y} - t^\alpha SE(\Delta\hat{Y}), \Delta\hat{Y} + t^\alpha SE(\Delta\hat{Y})]$

Time Series Regression

$\text{RMSFE} = \sqrt{E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]}$

$SE(Y_{T+1} - \hat{Y}_{T+1|T}) = \widehat{RMSFE} = \sqrt{var(\hat{u}_t)} = SER$

$(1-\alpha)$ CI: $[\hat{Y}_{T+1|T} - t^\alpha \times SE(Y_{T+1} - \hat{Y}_{T+1|T}), \hat{Y}_{T+1|T} + t^\alpha \times SE(Y_{T+1} - \hat{Y}_{T+1|T})]$

$\text{BIC}(K) = \ln\left[\frac{SSR(K)}{T}\right] + K\frac{\ln(T)}{T}$

$\text{AIC}(K) = \ln\left[\frac{SSR(K)}{T}\right] + K\frac{2}{T}$

F-statistic for the Granger Causality test has q and $T - \ell - p - 1$ degrees of freedom, where q is the number of restrictions imposed under the null, T is sample size, ℓ is the maximum lag length included in the time series model, p is the number of parameters in the time series model excluding the constant.

Statistical Distribution Tables

Critical Values of the t Distribution

	Significance Level					
	1- Tailed: / *2- Tailed:*	*.10* / *.20*	*.05* / *.10*	*.025* / *.05*	*.01* / *.02*	*.005* / *.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
	35	1.306	1.690	2.030	2.438	2.724
	36	1.306	1.688	2.028	2.434	2.719
	37	1.305	1.687	2.026	2.431	2.715
	38	1.304	1.686	2.024	2.429	2.712
	39	1.304	1.685	2.023	2.426	2.708
	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

95th Percentile for the F-distribution F_{ν_1,ν_2}

	Numerator ν_1											
Denominator ν_2: ν_2/ν_1	1	2	3	4	5	7	9	10	15	20	60	∞
1	161.45	199.50	215.71	224.58	230.16	236.77	240.54	241.88	245.95	248.01	252.2	254.31
2	18.51	19.00	19.16	19.25	19.30	19.35	19.41	19.40	19.43	19.45	19.48	19.50
3	10.13	9.55	9.28	9.12	9.01	8.89	8.81	8.79	8.70	8.66	8.57	8.53
4	7.71	6.94	6.59	6.39	6.26	6.09	6.00	5.96	5.86	5.80	5.69	5.63
5	6.61	5.79	5.41	5.19	5.05	4.88	4.77	4.74	4.62	4.56	4.43	4.37
6	5.99	5.14	4.76	4.53	4.39	4.21	4.10	4.06	3.94	3.87	3.74	3.67
7	5.59	4.74	4.35	4.12	3.97	3.79	3.68	3.64	3.51	3.44	3.30	3.23
8	5.32	4.46	4.07	3.84	3.69	3.50	3.39	3.35	3.22	3.15	3.01	2.93
9	5.12	4.26	3.86	3.63	3.48	3.29	3.18	3.14	3.01	2.94	2.79	2.71
10	4.96	4.10	3.71	3.48	3.33	3.14	3.02	2.98	2.85	2.77	2.62	2.54
15	4.54	3.68	3.29	3.06	2.90	2.71	2.59	2.54	2.40	2.33	2.16	2.07
20	4.35	3.49	3.10	2.87	2.71	2.51	2.39	2.35	2.20	2.12	1.92	1.84
30	4.17	3.32	2.92	2.69	2.53	2.33	2.21	2.16	2.01	1.93	1.74	1.62
40	4.08	3.23	2.84	2.61	2.45	2.25	2.12	2.08	1.92	1.84	1.64	1.51
50	4.03	3.18	2.79	2.56	2.40	2.20	2.07	2.03	1.87	1.78	1.58	1.44
60	4.00	3.15	2.76	2.53	2.37	2.17	2.04	1.99	1.84	1.75	1.53	1.39
120	3.92	3.07	2.68	2.45	2.29	2.09	1.95	1.91	1.75	1.66	1.43	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.01	1.88	1.83	1.67	1.57	1.32	1.00

Critical Values for the Chi-Squared Distribution

Degrees of Freedom	Critical Values		
	1%	5%	10%
1	6.64	3.84	2.71
2	9.21	5.99	4.61
3	11.35	7.81	6.25
4	13.28	9.49	7.78
5	15.09	11.07	9.24
6	16.81	12.59	10.65
7	18.48	14.07	12.02
8	20.09	15.51	13.36
9	21.67	16.92	14.68
10	23.21	18.31	15.99
11	24.73	19.68	17.28
12	26.22	21.0	18.55
13	27.69	22.4	19.81
14	29.14	23.7	21.06
15	30.58	25.0	22.31
16	32.00	26.3	23.54
17	33.41	27.6	24.77
18	34.81	28.9	25.99
19	36.19	30.1	27.20
20	37.57	31.4	28.41