Econ4570/6560
Econometrics/Introduction to Econometrics
Slide 1: Introduction and Review of Probability

Huaming Peng

Stock and Watson Chapter 1-2

Lecture outline

- What is econometrics?
- Review of Probability

What is Econometrics?

- Definition from Stock and Watson:

Econometrics is the science and art of using economic theory and statistical techniques to analyze economic data.

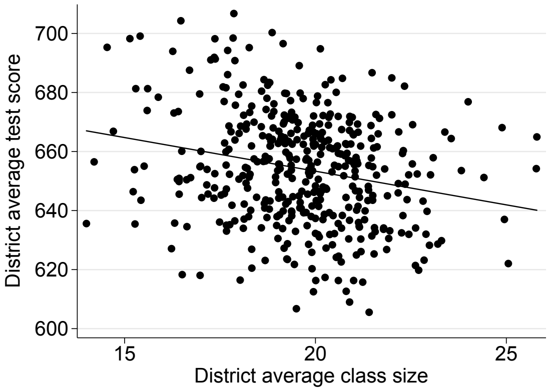
- In this course you will learn econometric techniques that you can use to answer economic questions using data on individuals, firms, municipalities, states or countries observed at one or multiple points in time.
- Focus will be on causality:

What is the causal effect of a change in X on Y?

What is Econometrics?
Example of questions addressed in this course

Does reducing class size improve test scores?

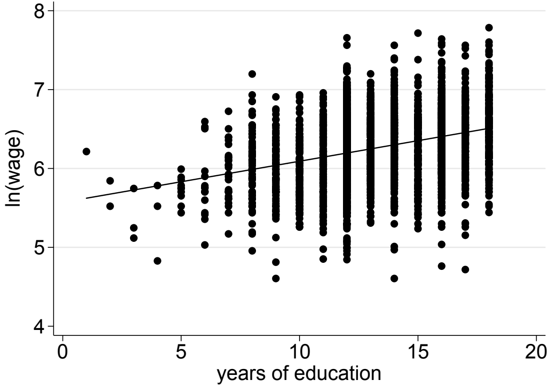
- Using data on 420 California school districts with information on class size and test scores, we will analyze whether reducing class size improves student's test scores



What is Econometrics?
Example of questions addressed in this course

What are the returns to education?

- Using data on 3,010 full-time working men in the US we will analyze whether obtaining more years of education increase wages.

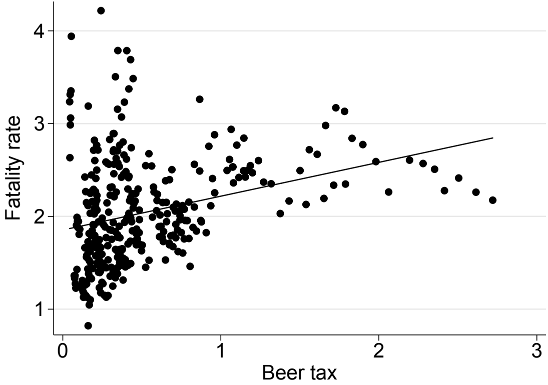


What is Econometrics?

Example of questions addressed in this course

Does increasing the tax on beer reduce traffic fatalities?

- Using data on 48 U.S. states for the years 1982-1988, we will analyze whether there is an effect of the tax on beer on the traffic fatality rate.



Review of Probability

Today we will discuss:

- A random variable and its probability distribution
- Measures of the shape of a probability distribution
  - Mean, variance, skewness and kurtosis
- Two random variables and their joint distribution
  - Joint distribution, marginal distribution, conditional distribution
  - Law of iterated expectations
  - Means, variances and covariances of sums of random variables
- Often used probability distributions in econometrics
  - Normal, Chi-Squared, Student t and F-distributions

A random variable

Some definitions:

Outcomes are the mutually exclusive potential results of a random process

- Your grade on the exam, the number of days it will snow next week

Random variable is a numerical summary of a random outcome

- The number of days it will snow next week is random and takes on a numerical value (0,1,2,3,4,5,6 or 7).

- There are two types of random variables:

Discrete random variable takes on discrete number of values, like 0,1,2,...

Continuous random variable takes on a continuum of possible values

Probability distribution of a discrete random variable

- Each outcome of a discrete random variable occurs with a certain probability

A Probability distribution of a discrete random variable is the list of possible values of the variable and the probability that each value will occur.

- Let random variable S be the number of days it will snow in the last week of January

Probability distribution of S

Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01

Cumulative distribution of a discrete random variable

A cumulative probability distribution is the probability that the random variable is less than or equal to a particular value

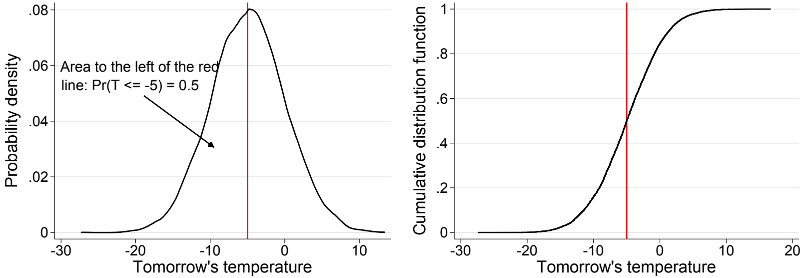
- The probability that it will snow less than or equal to s days, $F(s) = Pr(S \leq s)$ is the cumulative probability distribution of S evaluated at s
- A cumulative probability distribution is also referred to as a cumulative distribution or a CDF.

(cumulative) Probability distribution of S

Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01
CDF	0.20	0.45	0.65	0.80	0.90	0.95	0.99	1

Probability distribution of a continuous random variable

- Tomorrow's temperature is an example of a continuous random variable
- The CDF is defined similar to a discrete random variable.
- A probability distribution that lists all values and the probability of each value is not suitable for a continuous random variable.
- Instead the probability is summarized in a probability density function (PDF/ density)



Measures of the shape of a probability distribution

Expected value

The expected value or mean of a random variable is the average value over many repeated trails or occurrences.

Suppose a discrete random value Y takes on k possible values

$$E\left(Y\right) = \sum_{i=1}^{k} y_i \cdot Pr(Y = y_i) = \mu_Y$$

Number of days it will snow in the last week of January (S)								
Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01

$$E\left(S\right) = 0{\cdot}0.2 + 1{\cdot}0.25 + 2{\cdot}0.2 + 3{\cdot}0.15 + 4{\cdot}0.1 + 5{\cdot}0.05 + 6{\cdot}0.04 + 7{\cdot}0.01 = 2.06$$

Expected value of a continuous random variable

$$E\left(Y\right) = \int_{-\infty}^{\infty} y \cdot f(y)dy = \mu_Y$$

Measures of the shape of a probability distribution

Variance

The variance of a random variable Y is the expected value of the square of the deviation of Y from its mean.

- The variance is a measure of the spread of a probability distribution.
- Suppose a discreet random variable Y takes on k possible values

$$\sigma_Y^2 = Var(Y) = E\left[(Y - \mu_Y)^2\right] = \sum_{i=1}^{k} (y_i - \mu_Y)^2 \cdot Pr(Y = y_i)$$

Number of days it will snow in the last week of January (S)								
Outcome	0	1	2	3	4	5	6	7
Probability	0.20	0.25	0.20	0.15	0.10	0.05	0.04	0.01

$$\begin{aligned} Var(S) = \; & (0-2.06)^2 \cdot 0.2 + (1-2.06)^2 \cdot 0.25 + (2-2.06)^2 \cdot 0.2 + (3-2.06)^2 \cdot 0.15 \\ & + (4-2.06)^2 \cdot 0.1 + (5-2.06)^2 \cdot 0.05 + (6-2.06)^2 \cdot 0.04 + (7-2.06)^2 \cdot 0.01 \\ & 2.94 \end{aligned}$$

The standard deviation $\sigma_Y = \sqrt{Var(Y)}$ has the same units as Y

Mean and variance of a Bernoulli random variable

A Bernoulli random variable is a binary random variable with two possible outcomes, 0 and 1.

- For example, let B be a random variable which equals 1 if you pass the exam and 0 if you don't pass

$$B = \begin{cases} 1 & \textit{with probability } p \\ 0 & \textit{with probability } (1-p) \end{cases}$$

- The expected value of B is

$$E(B) = \mu_B = \sum_{i=1}^{k} b_i \cdot Pr(B = b_i) = 1 \cdot p + 0 \cdot (1-p) = p$$

- The variance of B is

$$\begin{aligned} Var(B) = \sigma_B^2 &= \sum_{i=1}^{k} (b_i - \mu_B)^2 \cdot Pr(B = b_i) \\ &= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p \\ &= p(1-p) \end{aligned}$$

- The standard deviation of B is $\sigma_B = \sqrt{p(1-p)}$

Mean and variance of a linear function of a random variable

- In this course we will consider random variables (say X and Y) that are related by a linear function

$$Y = a + b \cdot X$$

- Suppose $E(X) = \mu_X$ and $Var(X) = \sigma_X^2$
- This implies that the expected value of Y equals

$$E(Y) = \mu_Y = E(a + b \cdot X) = a + b \cdot E(X) = a + b \cdot \mu_X$$

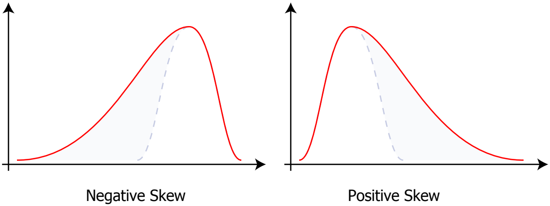
- The variance of Y equals

$$\begin{aligned}
Var(Y) = \sigma_Y^2 &= E\left[(Y - \mu_Y)^2\right] \\
&= E\left[((a + bX) - (a + b\mu_X))^2\right] \\
&= E\left[b^2(X - \mu_X)^2\right] \\
&= b^2 E\left[(X - \mu_X)^2\right] \\
&= b^2 \cdot \sigma_X^2
\end{aligned}$$

Measures of the shape of a probability distribution: Skewness

Skewness is a measure of the lack of symmetry of a distribution

$$Skewness = \frac{E\left[(Y-\mu_Y)^3\right]}{\sigma_Y^3}$$

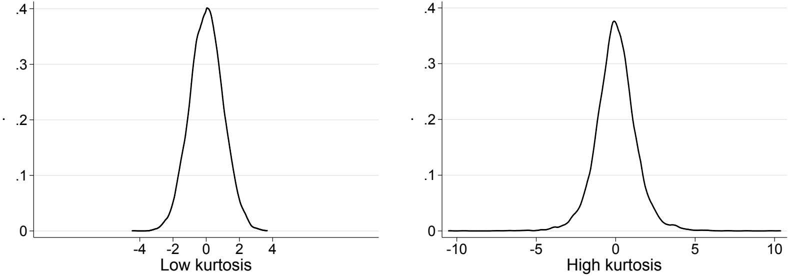


- For a symmetric distribution positive values of $(Y-\mu_Y)^3$ are offset by negative values (equally likely) and skewness is 0
- For a negatively (positively) skewed distribution negative (positive) values of $(Y-\mu_Y)^3$ are more likely and the skewness is negative (positive).
- Skewness is unit free.

Measures of the shape of a probability distribution: Kurtosis

Kurtosis is a measure of how much mass is in the tails of a distribution

$$Kurtosis = \frac{E\left[(Y - \mu_Y)^4\right]}{\sigma_Y^4}$$



- If a random variable has extreme values “outliers” the kurtosis will be high
- The Kurtosis is unit free and cannot be negative

Two random variables and their joint distribution

- Most of the interesting questions in economics involve 2 or more random variables
- Answering these questions requires understanding the concepts of joint, marginal and conditional probability distribution.

The joint probability distribution of two random variables X and Y can be written as $Pr(X = x, \; Y = y)$

- Let Y equal 1 if it snows and 0 if it does not snow.
- Let X equal 1 if it is very cold and 0 if it is not very cold.

Joint probability distribution of X and Y

	Very cold ($X = 1$)	Not very cold ($X = 0$)	*Total*
Snow ($Y = 1$)	0.15	0.07	*0.22*
No snow ($Y = 0$)	0.15	0.63	*0.78*
Total	*0.30*	*0.70*	*1.00*

Two random variables and the marginal distributions

The marginal probability distribution of a random variable is just another name for its probability distribution

- The marginal distribution of Y can be computed from the joint distribution of X and Y by adding up the probabilities of all possible outcomes for which Y takes a specific value

$$Pr\,(Y = y) = \sum_{i=1}^{l} Pr(X = x_i,\ Y = y)$$

- The probability that it will snow

$$Pr(Y = 1) = Pr(X = 1,\ Y = 1) + Pr(X = 0,\ Y = 1) = 0.22$$

Joint probability distribution of X and Y			
	Very cold ($X = 1$)	Not very cold ($X = 0$)	*Total*
Snow ($Y = 1$)	0.15	0.07	*0.22*
No snow ($Y = 0$)	0.15	0.63	*0.78*
Total	*0.30*	*0.70*	*1.00*

Two random variables and the conditional distribution

The conditional distribution is the distribution of a random variable conditional on another random variable taking on a specific value.

- The conditional probability that it snows given that is it very cold

$$Pr(Y = 1 \mid X = 1) = \frac{0.15}{0.3} = 0.5$$

- In general the conditional distribution of Y given X is

$$Pr\left(Y = y \mid X = x\right) = \frac{Pr\left(X = x,\ Y = y\right)}{Pr\left(X = x\right)}$$

- The conditional expectation of Y given X is

$$E\left(Y \mid X = x\right) = \sum_{i=1}^{k} y_i Pr(Y = y_i \mid X = x)$$

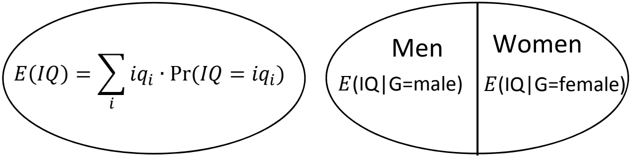
The expected value of snow given that it is very cold equals

$$E\left(Y \mid X = 1\right) = 1 \cdot Pr(Y = 1 \mid X = 1) + 0 \cdot Pr(Y = 0 \mid X = 1)$$
$$1 \cdot 0.5 + 0 \cdot 0.5 \qquad = 0.5$$

The law of iterated expectations

Law of iterated expectations states that the mean of Y is the weighted average of the conditional expectation of Y given X, weighted by the probability distribution of X.

$$E(Y) = E[E(Y \mid X)] = \sum_i E(Y \mid X = x_i) \cdot Pr(X = x_i)$$



If we use the second circle to compute $E(IQ)$:

$$\begin{aligned} E(IQ) &= E[E(IQ \mid G)] \\ &= \sum_i E(IQ \mid G = g_i)\, Pr(G = g_i) \\ &= E(IQ \mid G = m) \cdot Pr(G = m) + E(IQ \mid G = f) \cdot Pr(G = f) \end{aligned}$$

Independence

Independence: Two random variables X and Y are independent if the conditional distribution of Y given X does not depend on X

$$Pr\left(Y=y \mid X=x\right)=Pr\left(Y=y\right)$$

- If X and Y are independent this also implies

$$\begin{aligned} Pr\left(X=x,\ Y=y\right) &= Pr\left(X=x\right)\cdot Pr\left(Y=y|X=x\right) \\ &= Pr\left(X=x\right)\cdot Pr\left(Y=y\right) \end{aligned}$$

Mean independence: The conditional mean of Y given X equals the unconditional mean of Y

$$E\left(Y \mid X\right)=E\left(Y\right)$$

- For example if the expected value of snow (Y) does not depend on whether it is very cold (X)

$$E\left(Y \mid X=1\right)=E\left(Y \mid X=0\right)=E\left(Y\right)$$

Covariance

The covariance is a measure of the extend to which two random variables X and Y move together,

$$
\begin{aligned}
Cov(X,Y) = \sigma_{XY} &= E\left[(X-\mu_X)\cdot(Y-\mu_Y)\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{l}(x_j-\mu_X)(y_i-\mu_Y)\cdot Pr\left(X=x_j, Y=y_i\right)
\end{aligned}
$$

	Very cold ($X=1$)	Not very cold ($X=0$)	*Total*
Snow ($Y=1$)	0.15	0.07	*0.22*
No snow ($Y=0$)	0.15	0.63	*0.78*
Total	*0.30*	*0.70*	*1.00*

Example: the covariance between snow (Y) and it being very cold (X):

$$
\begin{aligned}
Cov(X,Y) &= (1-0.3)(1-0.22)\cdot 0.15 \\
&\quad +(1-0.3)(0-0.22)\cdot 0.15 \\
&\quad +(0-0.3)(1-0.22)\cdot 0.07 \\
&\quad +(0-0.3)(0-0.22)\cdot 0.63 \\
&= 0.084
\end{aligned}
$$

Correlation

- The units of the covariance of X and Y are the units of X multiplied by the units of Y
- This makes it hard to interpret the size of the covariance.

The correlation between X and Y is unit free:

$$Corr\,(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- A correlation is always between -1 and 1 and X and Y are uncorrelated if $Corr\,(X, Y) = 0$
- If the conditional mean of Y does not depend on X, X and Y are uncorrelated

$$if\ E\,(Y \mid X) = E\,(Y)\,,\ then\ Cov(X, Y) = 0\ \&\ Corr\,(X, Y) = 0$$

- If X and Y are uncorrelated this does not necessarily imply mean Independence!

Means, Variances and covariances of sums of random variables

Let

$$Z = aX + bY$$

The mean of Z equals

$$E(Z) = E(aX + bY) = aE(X) + bE(Y)$$

The variance of Z equals

$$\begin{aligned}
Var(Z) = Var(aX + bY) &= E\left\{[(aX + bY) - (a\mu_X + b\mu_Y)]^2\right\} \\
&= E\left\{[a(X - \mu_X) + b(Y - \mu_Y)]^2\right\} \\
&= E\left\{ \begin{array}{c} a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 \\ +2ab(X - \mu_X)(Y - \mu_Y) \end{array} \right\} \\
&= \begin{array}{c} a^2 E\left[(X - \mu_X)^2\right] + b^2 E\left[(Y - \mu_Y)^2\right] \\ +2abE[(X - \mu_X)(Y - \mu_Y)] \end{array} \\
&= a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)
\end{aligned}$$

Means, Variances, and Covariances of Sums of Random Variables

KEY CONCEPT 2.3

Let X, Y, and V be random variables, let μ_X and σ_X^2 be the mean and variance of X, let σ_{XY} be the covariance between X and Y (and so forth for the other variables), and let a, b, and c be constants. Equations (2.29) through (2.35) follow from the definitions of the mean, variance, and covariance:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \tag{2.29}$$

$$\text{var}(a + bY) = b^2\sigma_Y^2, \tag{2.30}$$

$$\text{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \tag{2.31}$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \tag{2.32}$$

$$\text{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \tag{2.33}$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y, \tag{2.34}$$

$$|\text{corr}(X, Y)| \leq 1 \text{ and } |\sigma_{XY}| \leq \sqrt{\sigma_X^2\sigma_Y^2} \text{ (correlation inequality).} \tag{2.35}$$

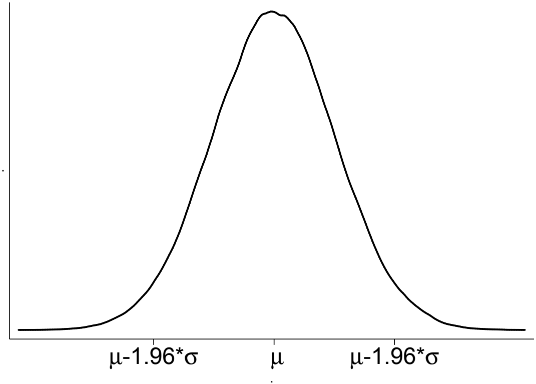
Examples of often used probability distributions in Econometrics

The Normal distribution

The most often encountered probability density function in econometrics is the Normal distribution:

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)\right]$$

- A normal distribution with mean μ and standard deviation σ is denoted as $N(\mu, \sigma)$

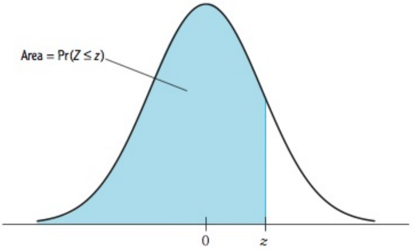


The Normal distribution

A standard normal distribution $N(0,1)$ has $\mu = 0$ and $\sigma = 1$

A random variable with a $N(0,1)$ distribution is often denoted by Z and the CDF is denoted by $\Phi(z) = Pr(Z \leq z)$

TABLE 1 The Cumulative Standard Normal Distribution Function, $\Phi(z) = \Pr(Z \leq z)$



	Second Decimal Value of z									
z	0	1	2	3	4	5	6	7	8	9
–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

The Normal distribution

To look up probabilities of a general normally distributed random variable

$$Y \sim N(\mu, \sigma)$$

we must first standardize Y to obtain the standard normal random variable Z

$$Z = \frac{(Y - \mu)}{\sigma}$$

- For example let $Y \sim N(5, 2)$

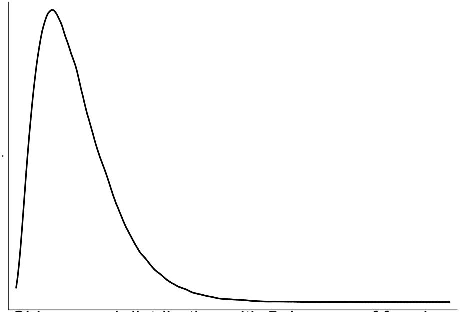
$$\begin{aligned} Pr(Y \leq 0) &= Pr\left(\frac{(Y-5)}{2} \leq \frac{(0-5)}{2}\right) \\ &= Pr(Z \leq -2.5) \\ &= 0.0062 \end{aligned}$$

The Chi-Squared distribution

The chi-squared distribution is the distribution of the sum of m squared independent standard normal random variables

- Let $Z_1, Z_2, \ldots, Z_m$ be m independent standard normal random variables
- The sum of the squares of these random variables has a chi-squared distribution with m degrees of freedom

$$\sum_{i=1}^{m} Z_i^2 \sim \chi_m^2$$



Chi-squared distribution with 5 degrees of freedom

The Chi-Squared distribution

806 Appendix

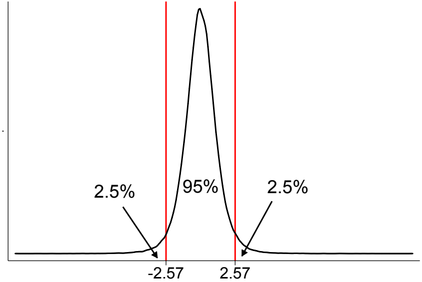
TABLE 3 Critical Values for the χ^2 Distribution

	Significance Level		
Degrees of Freedom	**10%**	**5%**	**1%**
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67

The Student t distribution

Let Z be a standard normal random variable and W a Chi-Squared distributed random variable with m degrees of freedom

The Student t-distribution with m degrees of freedom is the distribution the random variable $\frac{Z}{\sqrt{W/m}}$



Student t distribution with 5 degrees of freedom

- The t distribution has fatter tails than the standard normal distribution.
- When $m \geq 30$ it is well approximated by the standard normal distribution.

The Student t distribution

- The Student t distribution is often used when testing hypotheses in econometrics
- Appendix Table 2 shows selected percentiles of the t_m distribution
- For example with $m = 5$;

$$Pr\left(t < -2.57\right) = 0.025;\ Pr\left(|t| > 2.57\right) = 0.05;\ Pr\left(t > 2.57\right) = 0.025$$

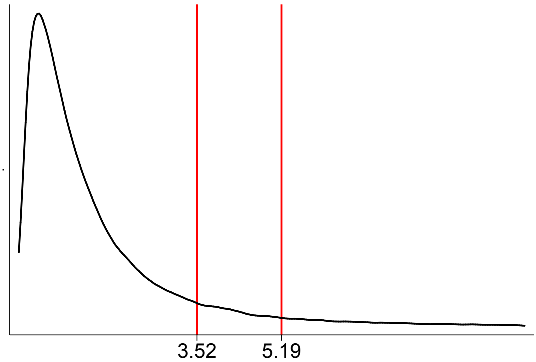
TABLE 2 Critical Values for Two-Sided and One-Sided Tests Using the Student t Distribution

Degrees of Freedom	Significance Level				
	20% (2-Sided) 10% (1-Sided)	10% (2-Sided) 5% (1-Sided)	5% (2-Sided) 2.5% (1-Sided)	2% (2-Sided) 1% (1-Sided)	1% (2-Sided) 0.5% (1-Sided)
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03

The F-distribution

Let W a chi-squared random variable with m degrees of freedom and V a chi-squared random variable with n degrees of freedom.

The F-distribution with m and n degrees of freedom $F_{m,n}$ is the distribution of the random variable $\frac{W/m}{V/n}$



F distribution with m=4 and n=5 degrees of freedom

The F-distribution

The 90th, 95th & 99th percentiles of the $F_{m,n}$ distribution are shown in Table 5

- For example with $m = 4$ and $n = 5$;

$$Pr\,(F > 3.52) = 0.10; \;\; Pr\,(F > 5.19) = 0.05$$

TABLE 5A Critical Values for the F_{n_1,n_2} Distribution—10% Significance Level

Denominator Degrees of Freedom (n_2)	Numerator Degrees of Freedom (n_1)									
	1	2	3	4	5	6	7	8	9	10
1	39.86	49.50	53.59	55.83	57.24	58.20	58.90	59.44	59.86	60.20
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30

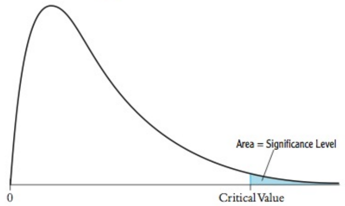
TABLE 5B Critical Values for the F_{n_1,n_2} Distribution—5% Significance Level

Denominator Degrees of Freedom (n_2)	Numerator Degrees of Freedom (n_1)									
	1	2	3	4	5	6	7	8	9	10
1	161.40	199.50	215.70	224.60	230.20	234.00	236.80	238.90	240.50	241.90
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.39	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74

The F-distribution

- A special case of the F distribution which is often used in econometrics is when $F_{m,n}$ can be approximated by $F_{m,\infty}$
- In this limiting case the denominator is the mean of infinitely many squared standard normal random variables, which equals 1.
- Appendix Table 4 shows the 90th, 95th and 99th percentiles of the $F_{m,\infty}$ distribution.

TABLE 4 Critical Values for the $F_{m,\infty}$ Distribution



	Significance Level		
Degrees of Freedom	10%	5%	1%
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78