# Lecture 1: General Introduction and Review on Basic Statistics

Shuo Liu

UCLA Summer School Econ 103

June 25, 2017

# Outline

# Course Information

- **You need to enroll in both Econ 103 and Econ 103L**;
- **Lecture (Econ 103)**: Mondays and Wednesdays, 1pm-3:05pm, Dodd 121;
- **Lecture Lab (Econ 103L)**: Wednesdays 4pm-4:50pm, Dodd 121;
- **Instructor office hour**: Mondays 3:30pm-5:30pm, Alper Room (2nd floor, Bunche Hall);
- **(Teaching Assistant) Yun Feng**, office hour: Monday 10-11am, Thursday 12am-1pm; **Alexander Graupner**, office hour: Tuesday 11am-1pm; Location: Alper Room (2nd floor, Bunche Hall);
- **Required textbook**: Principles of Econometrics. Hill, R. C., Griffiths, W. E. and G. C. Lim, 4th Edition, 2011. **Recommended for Learning Stata**: Using Stata for Principles of Econometrics. Adkins, L. C. and R. C. Hill, 4th Edition, 2011.

# Course Information

- **Two Assignments**: due on class July 5th (W) and July 26th (W). Please write manually/print and hand in the hardcopy;
- **On-Class Midterm**: July 12nd (W) 1pm-3pm, Dodd 121. **Cover**: Everything till (including) July 5th;
- **On-Class Final**: August 2nd (W) 1pm-3pm, Dodd 121. **Cover**: all topics starting from July 10th (midterms topics on "simple linear model" will not be specifically covered, but you should still have a general sense of the basic definitions (not require derivation) and how to read basic Stata output table);
- **No Makeup Exams**.

# Why Study Econometrics

Econometrics use theory and data from economics, **along with tools from statistics**, to depict (both quantitatively and qualitatively) the relationship between economic variables.

**Step 1**: Question of Interest (based on Economics theory);

- How does personal income determine/correlate to personal food expenditure?

**Step 2**: Collect available data on related economic variables and observe the pattern;

# General Introduction

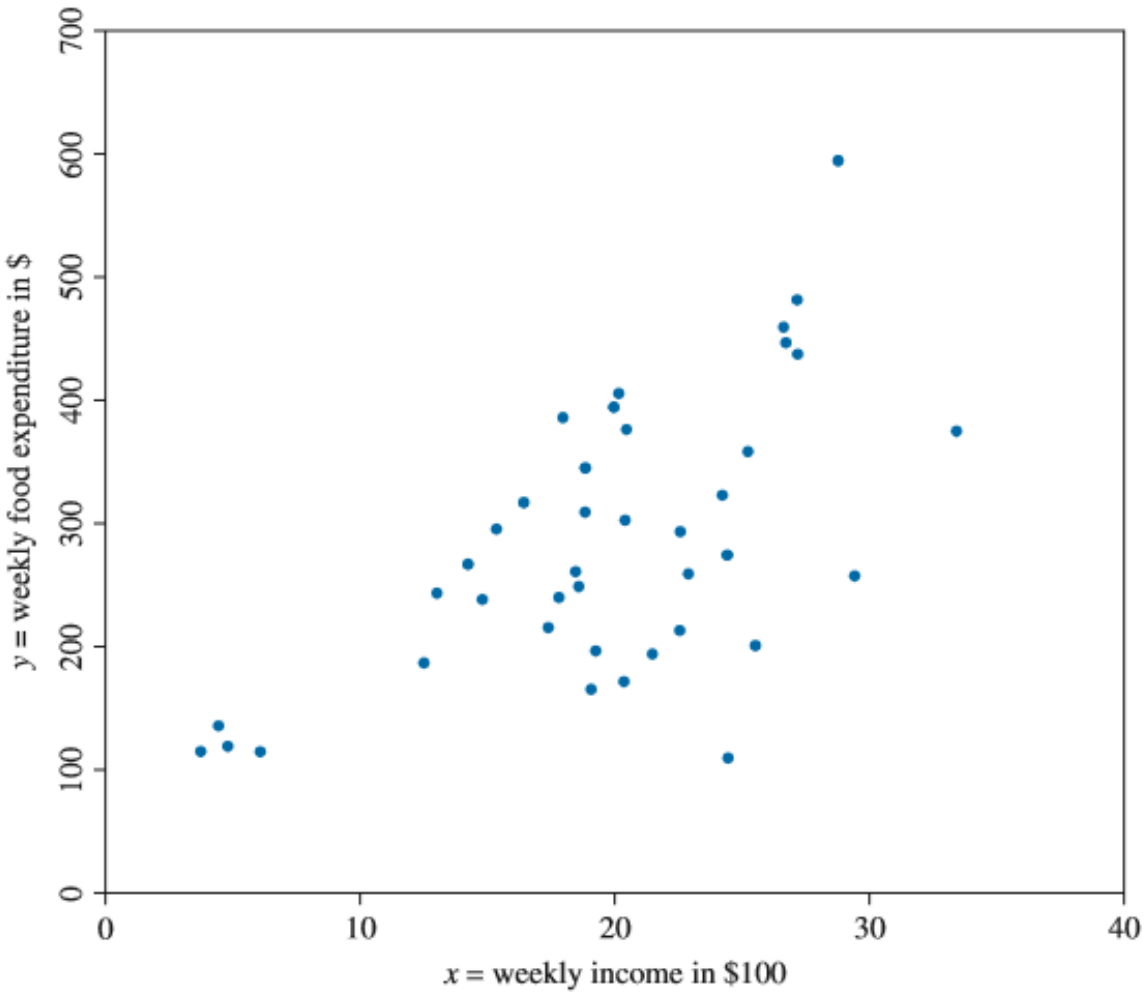



Figure: Linear Relationship: Data on Food Expenditure and Income

# Estimation of Nonlinear Relationship

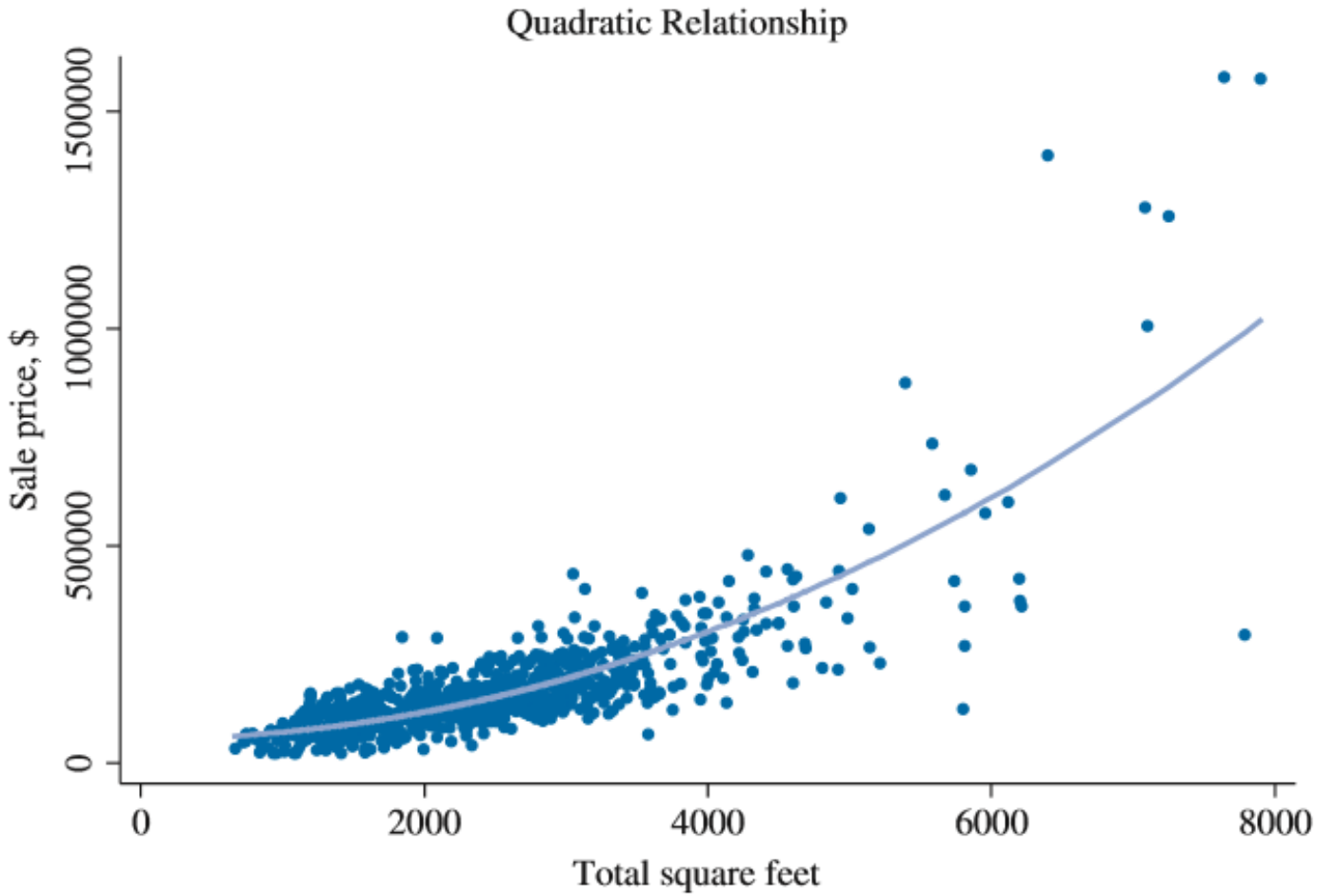



Figure: Quadratic Relationship: Data on House Price and House Square Feet

**Step 3**: Construct an econometric model (based on data pattern, economics theory and "model selection techniques/benchmarks" from Statistics). For example, $y$ is personal food expenditure, $x$ is personal income, we **assume** the model is

$$y = \alpha + \beta x + e$$

**Step 4**: Use Econometrics methods to estimate the unknown parameters ($\alpha$ and $\beta$) and obtain estimators ($a$ and $b$). For example, we get $a = 1, b = 2$

$$y = 1 + 2x$$

We do **point estimation** and **interval estimation**.

**Step 5**: Conduct hypothesis testing to test **(1)** the obtained estimators $a$ and $b$, to evaluate whether they are good or not; **(2)** the whole model, to evaluate whether the whole model is good or not.

**Step 6**: Interpret the result and use the estimated econometric model to do prediction.

# Examples of Questions of Interest (Which Historical Data Should We Collect?)

- How does demand change with the price of the good?
- What is the effect of a new marketing campaign on sales?
- How does class size affect education outcomes (e.g. test scores)?
- How much an additional year of schooling increases wages?
- What is the relationship between credit scores and loan default rates? (need to use "indicator variable")
- How much does output grow if the Fed cuts interest rates by 1%?
- Do more policemen reduce crime?

# Economic Data Types

**Time-Series Data**: Values of one economic variable in different time periods. Usually same economic quantity is recorded at a regular time interval.

***Table 1.1*** **Annual GDP (Billions of Real 2005 Dollars)**

| Year | GDP |
|---|---|
| 2001 | 11347.2 |
| 2002 | 11553.0 |
| 2003 | 11840.7 |
| 2004 | 12263.8 |
| 2005 | 12638.4 |
| 2006 | 12976.2 |
| 2007 | 13254.1 |
| 2008 | 13312.2 |

# Economic Data Types

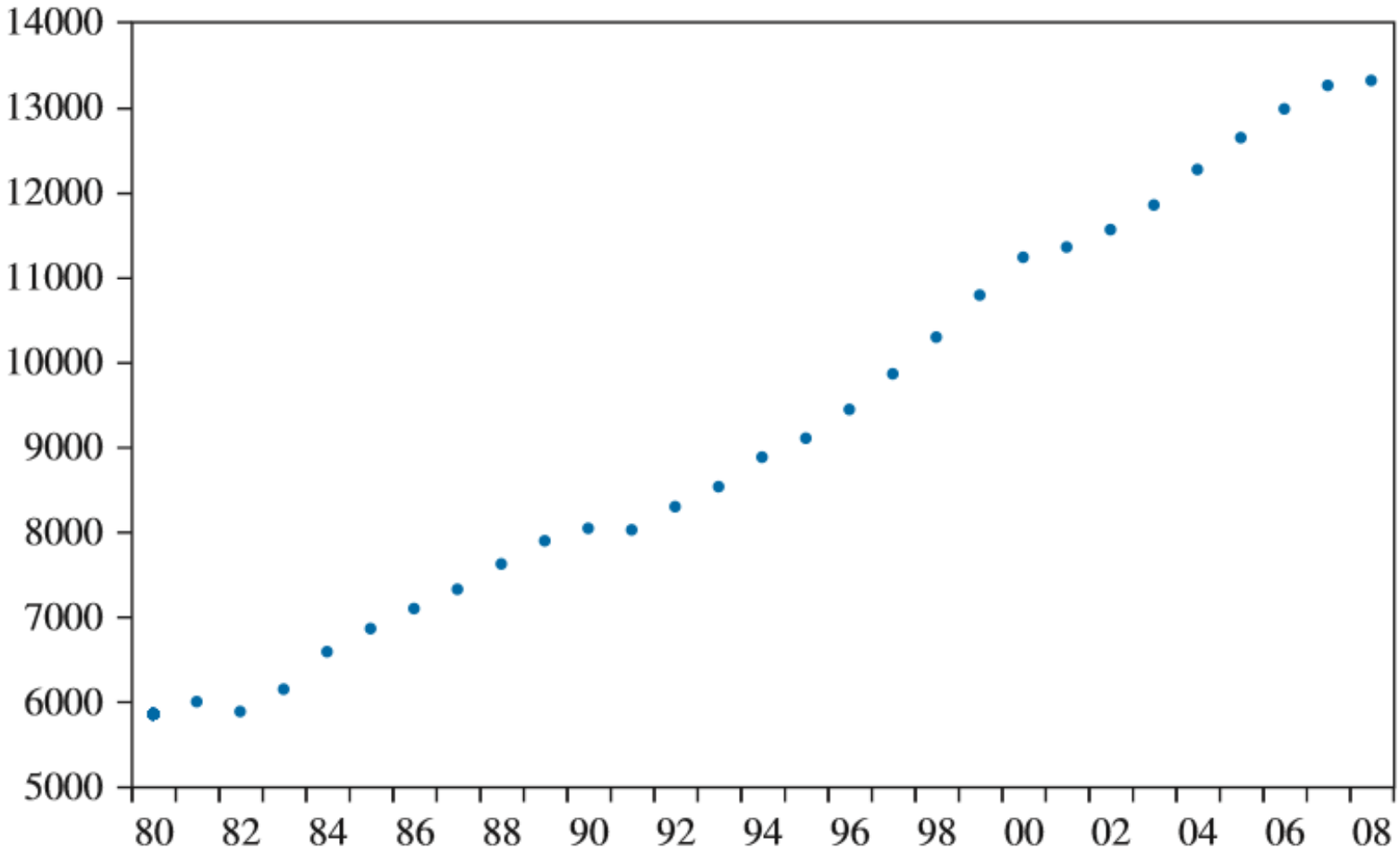



Figure: Real U.S. GDP (in 2005 dollars), 1980-2008

# Economic Data Types

**Cross-Sectional Data**: Values of more-than-one economic variables in a fixed time period.

***Table 1.2*** **Cross Section Data: CPS August 2009**

| | Variables | | | | | |
|---|---|---|---|---|---|---|
| Individual | *RACE* | *EDUCATION* | *MARITAL_STATUS* | *SEX* | *HOURS* | *WAGE* |
| 1 | White | 10th Grade | Never Married | Male | 2 | 8.00 |
| 2 | White | Assoc Degree | Married | Male | 40 | 10.81 |
| 3 | Other | Some College No Degree | Divorced | Male | 38 | 10.23 |
| 4 | White | High School Grad or GED | Married | Female | 32 | 11.50 |
| 5 | White | Some College No Degree | Never Married | Male | 50 | 12.50 |
| 6 | White | High School Grad or GED | Divorced | Female | 20 | 7.00 |
| 7 | White | High School Grad or GED | Married | Female | 10 | 8.00 |
| 8 | White | 5th or 6th Grade | Never Married | Female | 15 | 9.30 |
| 9 | White | High School Grad or GED | Married | Female | 40 | 20.00 |

# Economic Data Types

**Panel Data**: Values of more-than-one economic variables in different time periods. Also known as "longtitudinal" data.

*Table 1.3* **Panel Data from Two Rice Farms**

| *FIRM* | *YEAR* | *PROD* | *AREA* | *LABOR* | *FERT* |
|---|---|---|---|---|---|
| 1 | 1990 | 7.87 | 2.50 | 160 | 207.5 |
| 1 | 1991 | 7.18 | 2.50 | 138 | 295.5 |
| 1 | 1992 | 8.92 | 2.50 | 140 | 362.5 |
| 1 | 1993 | 7.31 | 2.50 | 127 | 338.0 |
| 1 | 1994 | 7.54 | 2.50 | 145 | 337.5 |
| 1 | 1995 | 4.51 | 2.50 | 123 | 207.2 |
| 1 | 1996 | 4.37 | 2.25 | 123 | 345.0 |
| 1 | 1997 | 7.27 | 2.15 | 87 | 222.8 |
| 2 | 1990 | 10.35 | 3.80 | 184 | 303.5 |
| 2 | 1991 | 10.21 | 3.80 | 151 | 206.0 |
| 2 | 1992 | 13.29 | 3.80 | 185 | 374.5 |
| 2 | 1993 | 18.58 | 3.80 | 262 | 421.0 |
| 2 | 1994 | 17.07 | 3.80 | 174 | 595.7 |
| 2 | 1995 | 16.61 | 4.25 | 244 | 234.8 |
| 2 | 1996 | 12.28 | 4.25 | 159 | 479.0 |
| 2 | 1997 | 14.20 | 3.75 | 133 | 170.0 |

# Some Useful Links of Economic Data

- Resources for Economist (RFE) (http://www.rfe.org )
- National Bureau of Economic Research (NBER) (http://www.nber.org/data/ )
- EconEdLink (http://www.econedlink.org/datalinks/ )
- Economagic (macro time series data) (http://www.economagic.com/ )
- FRED (The Federal Reserve Bank of St.Louis) is a good resource and system to obtain the macroeconomy and financial data (https://fred.stlouisfed.org )
- Current Population Survey (https://www.census.gov/cps )
- Panel Study of Income Dynamics (PSID) is a very good source of panel data on household survey (household income, expenditure, and etc) (https://psidon )
- Penn World Table is mainly used to obtain international economic/finance data (http://cid.econ.ucdavis.edu/data.html )

**Data Softwares**: MATLAB (most comprehensive: computation, simulation); R (Statistics model estimation); Stata (especially good to deal with panel data); SAS, Python, and etc.

# Econometric Models

- Econometric models are usually reduced-form models (at least for this course);
- **General Form**:

$$Y = f(X_1, X_2, \cdots) + e \tag{1}$$

where $e$ is called random error. It is a "noise" component that obscures our understanding of the "hypothesized" "true" relationship.

$\Longleftrightarrow$

$$E(Y|X_1, X_2, \cdots) = f(X_1, X_2, \cdots) \tag{2}$$

**Comments**

1. The "true" relationship $f$ applies for the whole population;
2. We use the realizations $\{y_i, x_{1i}, x_{2i}, \cdots\}_{i=1}^{n}$ (or sample of the population) to estimate $f$.

# Econometric Models

- **Note that**: reduced-form econometric models usually can only show the correlation (**but not the causal relationship**) between economic variables.
- **Examples of Cross-Sectional Models:** $\{y_i, x_i\}_{i=1}^n$

$$Y = \alpha + \beta X + e \tag{3}$$

$$Y = \alpha + \beta X^2 + e \tag{4}$$

$$Y = \alpha + \beta \sqrt{X} + e \tag{5}$$

- All the examples above are **linear** models: linear in parameters. One nonlinear example: $Y = \alpha + \sqrt{\beta} X + e$.

# Review on Basic Statistics (from Econ 41)

**1. Random Variable**: a variable whose value is unknown until it is observed. The support of a random variable is the collection of possible values it takes.

- **Discrete random variable**: it can only take a limited or countable number of values. **Examples**: the possible values of $X$ are $x = 1, 2, 3$ (the support of $X$ is $\{1, 2, 3\}$); natural numbers $x \in N \equiv \{1, 2, \cdots\}$, and etc.
- **Continuous random variable**: the number of values is uncountable and support is often interval(s) of real numbers. **Examples**: $x \in R$, $x \in R^+$, $x \in [0, 1]$, $x \in [0, 2] \cup [3, 4]$, and etc.

**One specific example of discrete random variable**: The support of $X$ is $\{0, 1\}$. $X$ is **indicator variable**, which is used to represent qualitative characteristics such as gender (male or female), race (Asian or not Asian), and etc, depending on specific question you focus on.

# Review on Basic Statistics (from Econ 41)

**2. Probability Distribution**:

- Use **probabiltiy density function (pdf)**(for discrete random variable, also called "probability mass function (pmf)") to indicate the probability of each possible value occurring, $f_X(x) = P(X = x)$;
- Use **cumulative distribution function (cdf)** to indicate the probability that $X$ is less than or equal to a specific value $x$, $F_X(x) = P(X \leq x)$.

**Discrete Example**: $S = \{1, 2, 3\}$,

$$X = \begin{cases} 1 & w.p. \quad \frac{1}{6}; \\ 2 & w.p. \quad \frac{2}{6}; \\ 3 & w.p. \quad \frac{3}{6}. \end{cases} \qquad \sum_{x=1}^{3} f_X(x) = 1$$

**Continous Example**: $S = [0, 1]$, $f_X(x) = 3x^2 \quad \forall x \in [0, 1]$, and $\int_S f_X(x)dx = \int_0^1 f_X(x)dx = 1$, and $P(X \leq x) = \int_0^x f_X(s)ds = F_X(x)$.

# Review on Basic Statistics (from Econ 41)

## 3. First Operation: Expectation

| Target | Discrete | Continuous |
|---|---|---|
| $E(X)$ | $E(X) = \sum_S x.f_X(x)$ | $E(X) = \int_S x.f_X(x)dx$ |
| $E(g(X))$ | $E(g(X)) = \sum_S g(x).f_X(x)$ | $E(g(X)) = \int_S g(x).f_X(x)dx$ |

Calculate $E(X)$ and $E(g(X))$ for the two examples on last page, and $g(X) = X^2$

## Second Operation: Variance

| Target | Discrete | Continuous |
|---|---|---|
| $Var(X)$ | $\sum_S (x - E(X))^2.f_X(x)$ | $\int_S (x - E(X))^2.f_X(x)dx$ |
| $Var(g(X))$ | $\sum_S (g(x) - E(g(X)))^2.f_X(x)$ | $\int_S (g(x) - E(g(X)))^2.f_X(x)dx$ |

Calculate $Var(X)$ and $Var(g(X))$ for the two examples on last page, and $g(X) = X^2$

# Review on Basic Statistics (from Econ 41)

**Standard Deviation:**

$$\sigma_X = \sqrt{Var(X)}$$

**Alternative Notations:**

$$\mu_X = E(X) \quad \sigma_X^2 = Var(X)$$

**Important**: $\mu_X$, $\sigma_X^2$ and $\sigma_X$ are population moments, to estimate these population moments, we use sample statistics calculated from the sample data $\{x_i\}_{i=1}^n$:

- **Sample mean**: $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$;
- **Sample variance**: $\bar{\sigma_X^2} = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}$.

# Review on Basic Statistics (from Econ 41)

**4. Joint Probability Distribution of Two Random Variables $X$ and $Y$:** $S$ is the set of all possible values of $(x, y)$, $S = S^x \times S^y$, joint pdf is $P(X = x, Y = y) = f_{X,Y}(x, y)$ and joint cdf is $P(X \leq x, Y \leq y) = F_{X,Y}(x, y) \quad \forall (x, y) \in S$.

| **Target** | **Discrete** | **Continuous** |
|---|---|---|
| joint cdf | $\sum_{s \leq x} \sum_{t \leq y} f_{X,Y}(s, t)$ | $\int_{s \leq x} \int_{t \leq y} f_{X,Y}(s, t) dsdt$ |
| marginal pdf $f_X(x)$ | $\sum_{y \in S^y} f_{X,Y}(x, y)$ | $\int_{S^y} f_{X,Y}(x, t) dt$ |
| marginal pdf $f_Y(y)$ | $\sum_{x \in S^x} f_{X,Y}(x, y)$ | $\int_{S^x} f_{X,Y}(s, y) ds$ |
| marginal cdf $F_X(x)$ | $\sum_{s \leq x} \sum_{t \in S^y} f_{X,Y}(s, t)$ | $\int_{s \leq x} \int_{S^y} f_{X,Y}(s, t) dsdt$ |
| marginal cdf $F_Y(y)$ | $\sum_{s \in S^x} \sum_{t \leq y} f_{X,Y}(s, t)$ | $\int_{S^x} \int_{t \leq y} f_{X,Y}(s, t) dsdt$ |

**Conditional pdf**:

$$f_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f_{X,Y}(x, y)}{f_X(x)} \tag{6}$$

**Comments:** use marginal pdfs of $X$ and $Y$ to calculate $E(X) = \mu_X$, $Var(X) = \sigma_X^2$ and $E(Y) = \mu_Y$, $Var(Y) = \sigma_Y^2$.

# Review on Basic Statistics (from Econ 41)

**Discrete Example:**

| **Support** | $x = 1$ | $x = 2$ |
|---|---|---|
| $y = 1$ | 0.16 | 0.215 |
| $y = 2$ | 0.225 | 0.18 |
| $y = 3$ | 0.115 | 0.105 |

**Calculate:** $f_X(x)$, $f_Y(y)$, $F_X(x)$, $F_Y(y)$, $f_{Y|X}(y|x)$

**Covariance:**

$$\sigma_{XY} = Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E(XY) - E(X)E(Y) \tag{7}$$

**Correlation Coefficient:**

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad \text{and} \quad -1 \leq \rho_{XY} \leq 1 \tag{8}$$

# Review on Basic Statistics (from Econ 41)

Properties of Operations **Expectation, Variance and Covariance**:
Assume $a$ and $b$ are any constants, $X$ and $Y$ are random variables

- $E(a) = a$, $E(aX + b) = aE(X) + b$, $E(X + Y) = E(X) + E(Y)$;
- $Var(a) = 0$, $Var(aX + b) = Var(aX) = a^2 Var(X)$;
- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$;
- $Cov(aX, bY) = aCov(X, bY) = abCov(X, Y)$, $Cov(X, X) = Var(X)$.

**If $X$ and $Y$ are independent**:

- $P(X = x, Y = y) = P(X = x)P(Y = y)$, $f_{X,Y}(x, y) = f_X(x)f_Y(y)$;
- $P(X = x|Y = y) = f_{X|Y}(x|y) = P(X = x) = f_X(x)$;
- $E(XY) = E(X)E(Y)$;
- $Var(X + Y) = Var(X) + Var(Y)$;
- $X \perp\!\!\!\perp Y \rightarrow Cov(X, Y) = 0$, but the reverse is not true.

# Review on Basic Statistics (from Econ 41)

**Important Specific Distributions**:

1. **Normal Distribution** $X \sim N(\mu, \sigma^2)$;
2. **Standard Normal Distribution**:

$$X \sim N(\mu, \sigma^2) \Longrightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

**Comments**:

- Any **linear combination of independent** normal random variables is still normal (for non-independent ones, linear combination may not be still normal);
- The reason to transfer to standard normal distribution $Z$ is that, we have the **table of values of cdf** associated with standard normal distribution, i.e. $P(Z \leq z)$ for $\forall z \in R$.

# Review on Basic Statistics (from Econ 41)

**Table of $Z \sim N(0,1)$:**

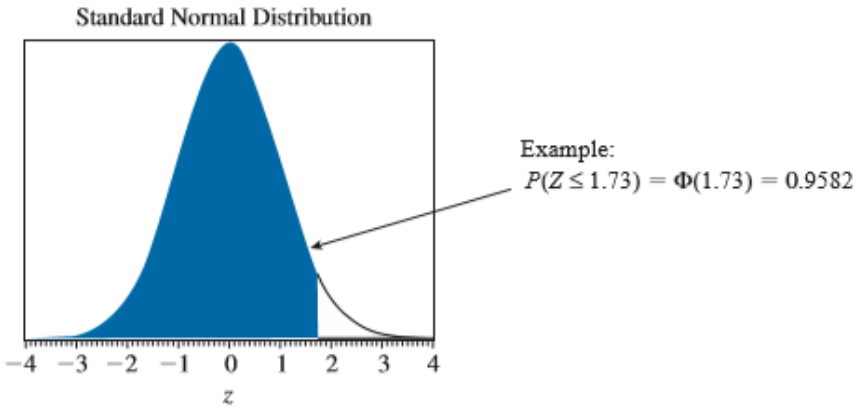



***Table 1*** **Cumulative Probabilities for the Standard Normal Distribution**
**$\Phi(z) = P(Z \le z)$**

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 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 |

# Review on Basic Statistics (from Econ 41)

**3. Chi-Square Distribution**: assume $Z_1, Z_2, \cdots, Z_m$ are $m$ randomly draws from standard normal distribution, that is $Z_i \sim N(0,1), \forall i = 1, 2, \cdots, m$. We call $Z_1, Z_2, \cdots, Z_m$ are i.i.d. (independently identically distributed). Then

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

where $m$ is called the "degree of freedom", which uniquely determines/fixes a Chi-Square distribution.

**4. Student t distribution**: assume $Z \sim N(0,1)$ ($Z$ is another standard normal distributed random variable, that is also independent of $Z_1, Z_2, \cdots, Z_m$ above), and thus $Z$ and $W$ are independent. Then

$$\frac{Z}{\sqrt{W/m}} \sim t_m$$

$m$ still uniquely determines/fixes the shape of Student t distribution $t_m$, which we will call **t distribution with $m$ degree of freedom**.

# Review on Basic Statistics (from Econ 41)

**Table of Student t distribution:**

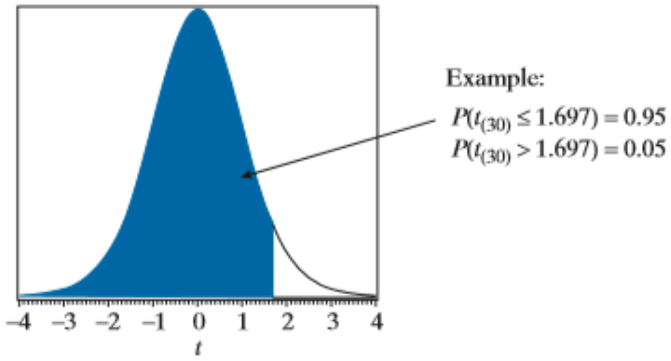
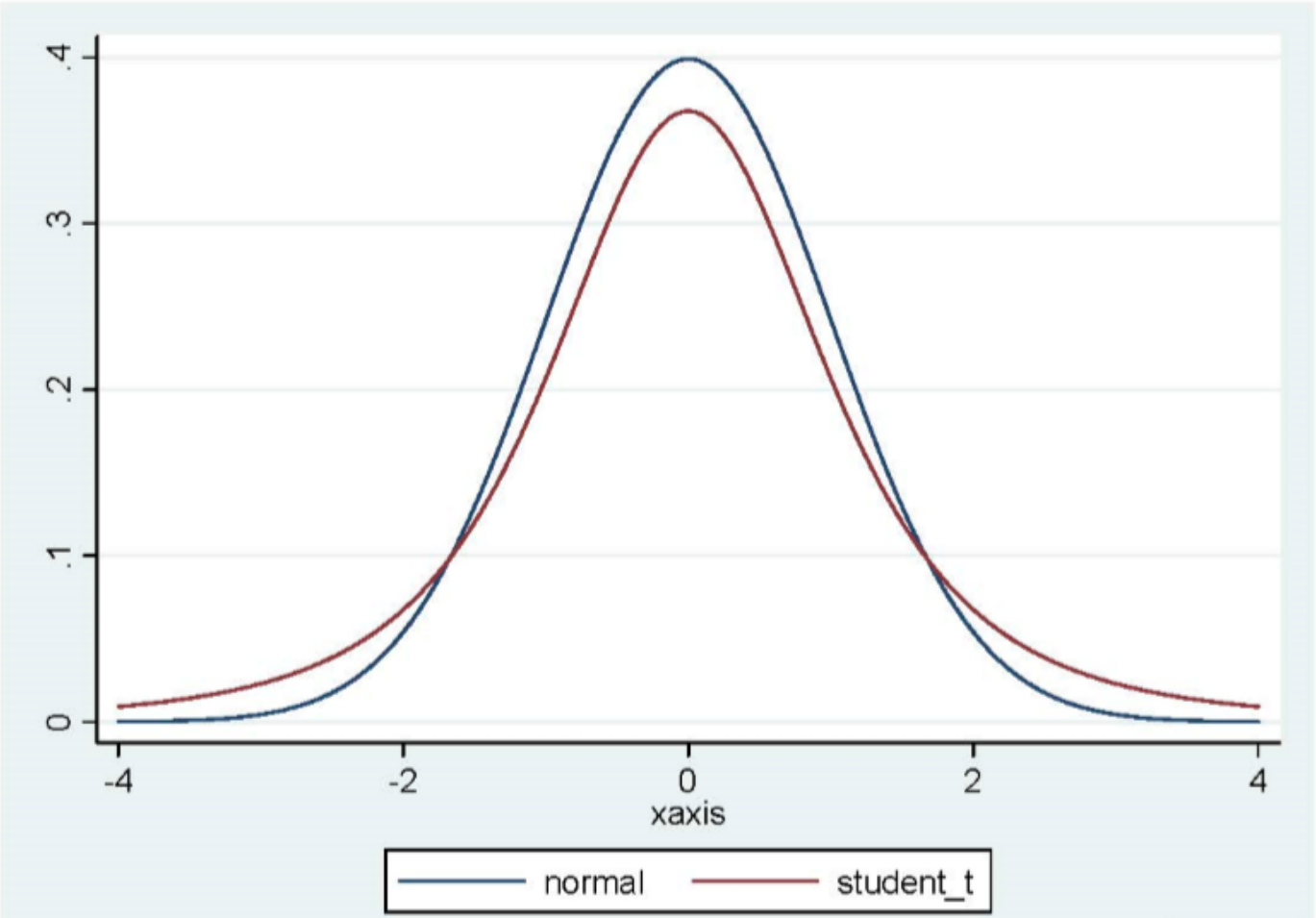



***Table 2*** **Percentiles of the *t*-distribution**

| df | $t_{(0.90,df)}$ | $t_{(0.95,df)}$ | $t_{(0.975,df)}$ | $t_{(0.99,df)}$ | $t_{(0.995,df)}$ |
|---|---|---|---|---|---|
| 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 |

# Review on Basic Statistics (from Econ 41)

**Compare the graphs of Standard Normal distribution and Student t distribution**: when $m \to \infty$, $t_m \sim N(0,1)$



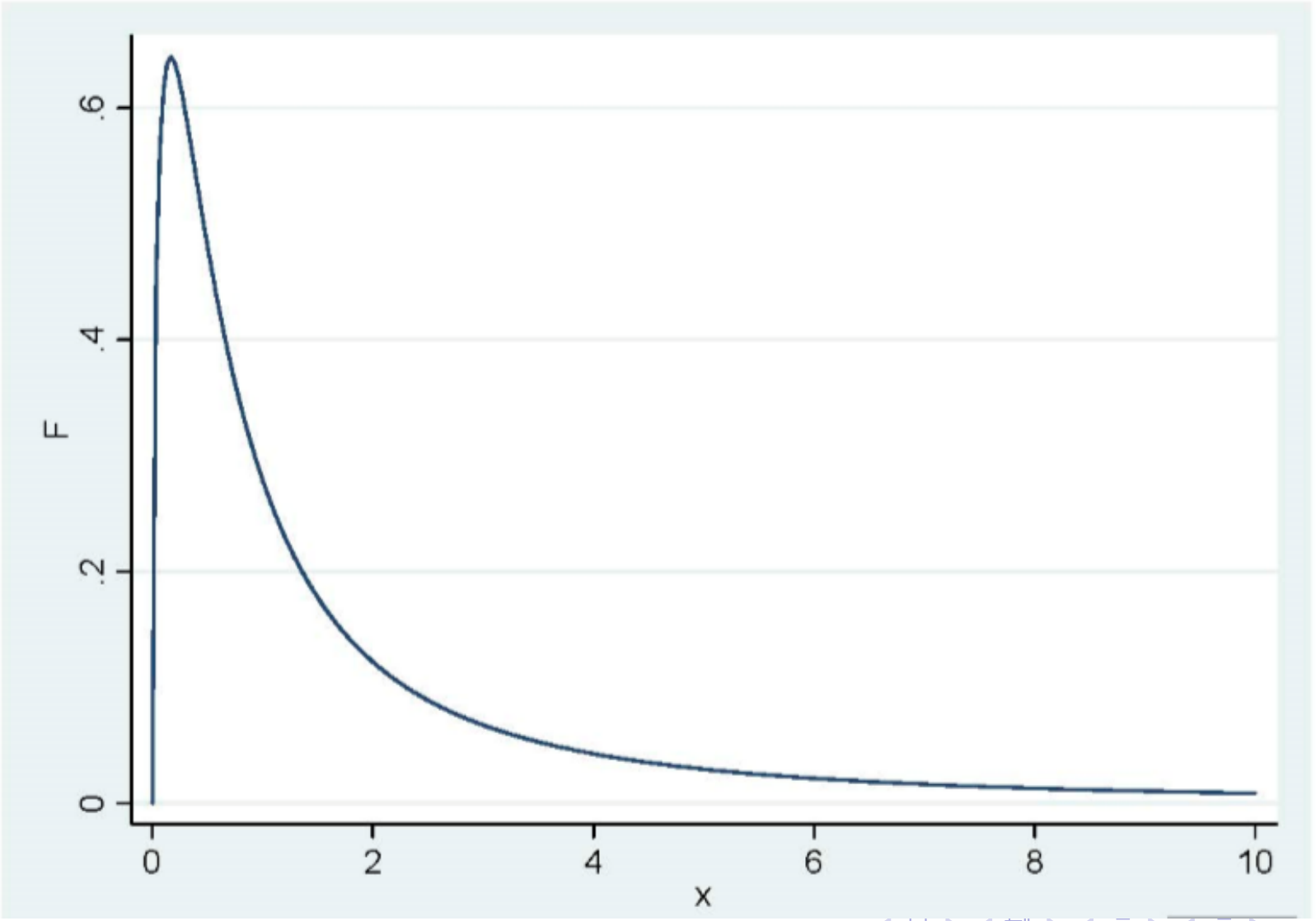
**5. F distribution**: if $W \sim \chi^2_m$, $V \sim \chi^2_n$, $W$ **and** $V$ **are independent**. Then

$$\frac{W/m}{V/n} \sim F_{m,n}$$

the **two degree of freedom** $m$ and $n$ **together** determines/fixes the shape of F distribution $F_{m,n}$.

- $mF_{m,\infty} = \chi^2_m$ and $F_{1,n} = t^2_n$

# Review on Basic Statistics (from Econ 41)



# Review on Basic Statistics (from Econ 41)

**Table of F distribution:**

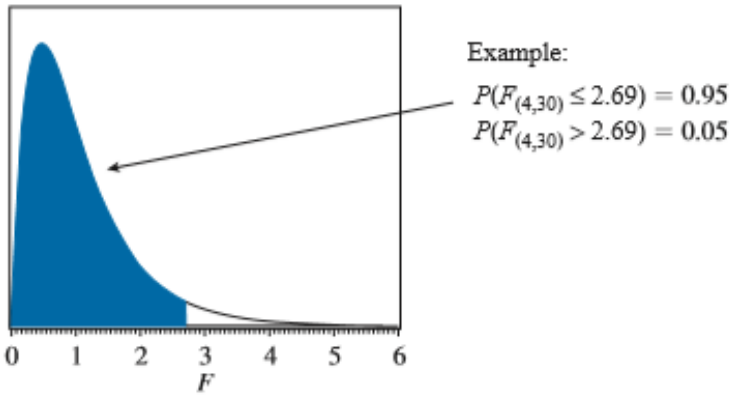



***Table 4*** **95th Percentile for the *F*-distribution**

| $v_2/v_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 | 238.88 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 |

# Review on Basic Statistics (from Econ 41)

**Example of Computing Probabilities using Table**: assume that adult male heights are normally distributed with mean 60 inches and standard deviation 4 inches. If you are 55 inches tall, what percentage of men are shorter than you? (if you randomly meet another man, what is the probability that that man's height is lower than you)

- Let $X$ denote the height of men. $X \sim N(60, 16)$.
- **Our target**: $P(X < 55) =?$
- **Always want to transform firstly**:

$$Z = \frac{X - 60}{4} \sim N(0, 1)$$

then our target becomes

$$P(X < 55) = P(\frac{X - 60}{4} < \frac{55 - 60}{4}) = P(Z < \frac{55 - 60}{4}) = P(Z < -1.25)$$

- **Check $Z$ table.**

# Review on Basic Statistics (from Econ 41)

**Table of Z:**

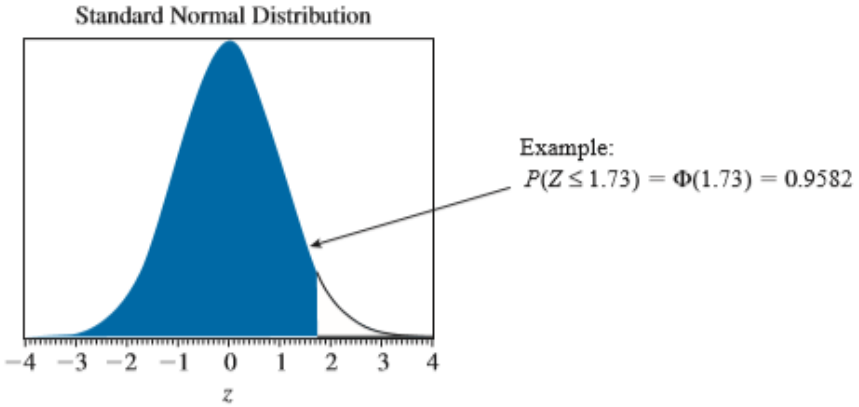



***Table 1*** **Cumulative Probabilities for the Standard Normal Distribution**
$\Phi(z) = P(Z \le z)$

| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 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 |

# Review on Basic Statistics (from Econ 41)

So the answer is:

$$\begin{aligned} P(Z < -1.25) &= \Phi(-1.25) \\ &= 1 - \Phi(1.25) \\ &= 1 - P(Z < 1.25) \\ &= 1 - 0.8944 \\ &= 0.1056 \\ &= 10.56\% \end{aligned}$$