Introductory Econometrics Notes and Exercises

Math Review A

Notes

Summation Proofs

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)$$

$$=\sum_{i=1}^{n}x_i^2-2\sum_{i=1}^{n}x_i\bar{x}+\sum_{i=1}^{n}\bar{x}^2$$

$$=\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+\bar{x}\sum_{i=1}^{n}\bar{x}$$

$$=\sum_{i=1}^{n}x_i^2-2n\bar{x}^2+n\bar{x}^2$$

$$=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$$

$$=\sum_{i=1}^{n}x_i^2-\bar{x}\sum_{i=1}^{n}x_i$$

$$=\sum_{i=1}^{n}(x_i^2-\bar{x}x_i)$$

$$=\sum_{i=1}^{n}x_i(x_i-\bar{x})$$

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}(x_iy_i-x_i\bar{y}-y_i\bar{x}+\bar{x}\bar{)}$$

$$=\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\bar{y}-\sum_{i=1}^{n}y_i\bar{x}+\sum_{i=1}^{n}\bar{x}\bar{y}$$

$$=\sum_{i=1}^{n}x_iy_i-\bar{y}\sum_{i=1}^{n}x_i-\bar{x}\sum_{i=1}^{n}y_i+\bar{y}\sum_{i=1}^{n}\bar{x}$$

$$=\sum_{i=1}^{n}x_iy_i-n\bar{y}\bar{x}-n\bar{x}\bar{y}+n\bar{y}\bar{x}$$

$$=\sum_{i=1}^{n}x_iy_i-n\bar{y}\bar{x}$$

$$=\sum_{i=1}^{n}x_iy_i-\bar{x}\sum_{i=1}^{n}y_i$$

$$=\sum_{i=1}^{n}(x_iy_i-\bar{x}y_i)$$

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

Natural Logorithm

1. $\ln(xy) = \ln(x) + \ln(y)$
2. $\ln(\frac{x}{y}) = \ln(x) - \ln(y)$
3. $\ln(x^c) = c\ln(x)$

The difference in lns can be used to approximate proportionate changes. Let x_0 and x_1 be positive values. Then, for small changes in x

$$\ln(x_1) - \ln(x_0) \approx \frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x_0}$$

Thus,

$$100 \cdot \Delta \ln(x) \approx \%\Delta x$$

Elasticity

The **elasticity** of y with respect to x equals

$$\frac{\%\Delta y}{\%\Delta x} = \frac{(y_1 - y_0)/y_0}{(x_1 - x_0)/x_0} = \frac{\Delta y/y_0}{\Delta x/x_0} = \frac{\Delta y}{\Delta x} \cdot \frac{x_0}{y_0}$$

Defining a linear model $y = \beta_0 + \beta_1 x$, the elasticity of y with respect to x equals

$$\frac{\%\Delta y}{\%\Delta x} = \frac{\Delta y/y}{\Delta x/x} = \frac{\Delta y}{\Delta x} \cdot \frac{x}{y} = \frac{\beta_1 \Delta x}{\Delta x} \cdot \frac{x}{\beta_0 + \beta_1 x} = \beta_1 \cdot \frac{x}{\beta_0 + \beta_1 x} \approx \frac{\Delta \ln(y)}{\Delta \ln(x)}$$

If we use the above approximation for both x and y, then the elasticity is approximately equal to $\frac{\ln(y)}{\ln(x)}$. Thus, a **constant elasticity model** is approximated by

$$\ln(y) = \beta_0 + \beta_1 \ln(x)$$

where β_1 is the approximate elasticity of y with respect to x.

A **semi-elasticity model** approximates the percentage change in y with respect to a unit change in x and takes the form

$$\ln(y) = \beta_0 + \beta_1 x$$

where β_1 is the semi-elasticity of y with respect to x. In other words, $\%\Delta y = 100\beta_1 \Delta x \rightarrow \beta_1 \approx \frac{\%\Delta y}{100\Delta x}$.

Another relationship of some interest is

$$y = \beta_0 + \beta_1 \ln(x)$$

Using calculus, we can derive

$$\Delta y = \beta_1 \Delta \ln(x)$$

and thus

$$\beta_1 = \frac{\Delta y}{\Delta \ln(x)} \approx \frac{\Delta y}{\frac{\%\Delta x}{100}}$$

In other words, $\beta_1/100$ is the unit change in y when x increases by 1%.

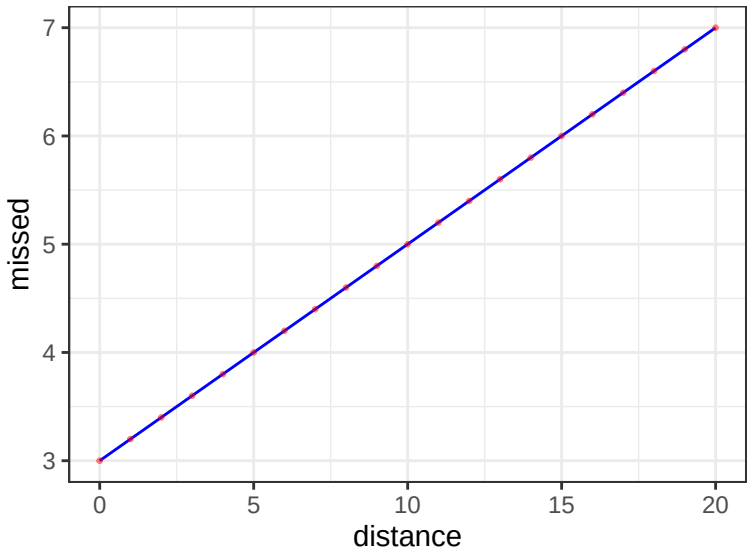
Exercises

1. (a) Mean = 566
   (b) Median = 505
   (c) Mean = 5.66 Median = 5.05
   (d) Mean = 586 Median = 505

2. (a) See below

   (b) 4 classes

   (c) 2 classes



3. -21 CDs. This suggests using linear functions to describe demand curves may not be realistic/a good idea. Some form of an elasticity model would likely be more suitable.

4. (a) A 0.8 percentage point decrease.

   (b) A 0.125% fall.

5. The correct terminology would be the stock return increased by 3 percentage points, a 20% increase in the return on the stock.

6. (a) 20%

   (b) $\approx$ 18.2321557%

7. (a) \$40134.84

   \$45935.80

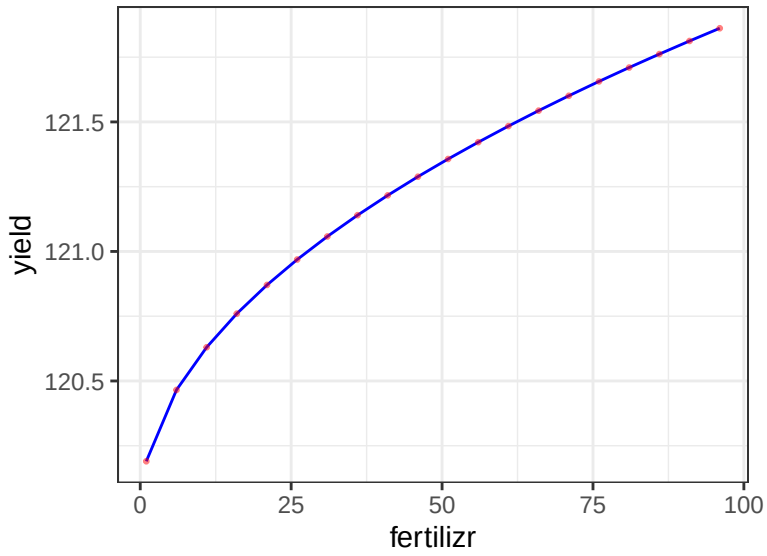
   (b) $\%\Delta\text{salary} = 100(.027)(5) = 13.5\%$

   (c) 7.0642847% error

8. The intercept indicates that, with no sales tax, the proportionate growth in employment would be .043 units. The slope indicates that for every unit increase in sales tax, we would expect the proportionate growth in employment to decrease by .78 units.

9. (a) See below

   (b) The most notable difference is the convexity/concavity of the functions. A linear model would have constant marginal returns to yield with respect to fertilizer while the given relationship displays diminishing marginal returns.



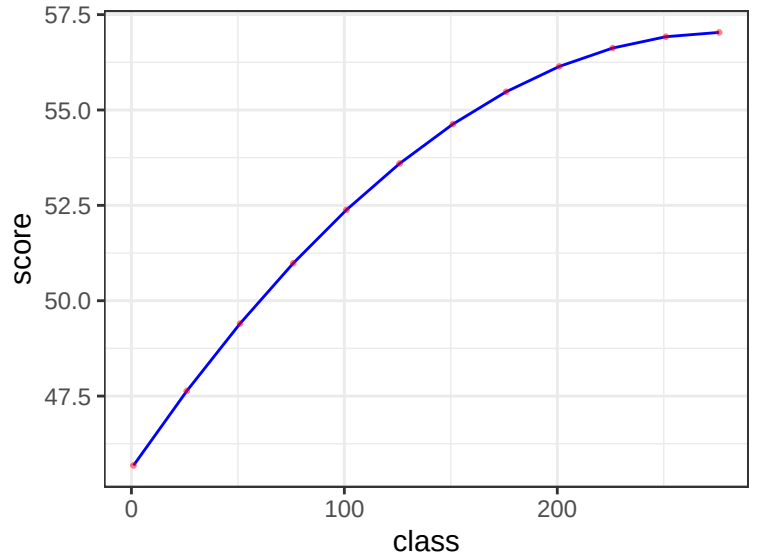
10. (a) It's not of much interest by itself. It suggests a class with 0 students would expect a test score of 45.6, which doesn't make sense or have any real meaning.

    (b)
    $$\frac{\partial score}{\partial class} = .082 - .000294 \cdot class = 0$$
    $$\rightarrow class^* = \frac{.082}{.000294} \approx 279 \text{ students}$$

    The highest achievable test score is about 57.

    (c) See below

    (d) No, this equation may give an idea about what one can expect for `score` given `class`, but it's unrealistic to expect exact results.



11. (a)

$$\bar{y} = \frac{y_1 + y_2}{2} = \frac{\beta_0 + \beta_1 x_1 + \beta_0 + \beta_1 x_2}{2}$$
$$= \frac{2\beta_0 + \beta_1(x_1 + x_2)}{2}$$
$$= \beta_0 + \beta_1 \frac{(x_1 + x_2)}{2} = \beta_0 + \beta_1 \bar{x}$$

(b)

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} \beta_0 + \beta_1 x_i}{n}$$
$$= \frac{n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i}{n}$$
$$= \beta_0 + \beta_1 \frac{\sum_{i=1}^{n} x_i}{n} = \beta_0 + \beta_1 \bar{x}$$

12. (a)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \left(\sum_{i=1}^{n_1} x_i + \sum_{i=n_1+1}^{n} x_i \right)$$
$$= \frac{1}{n} \left(n_1 \bar{x_1} + n_2 \bar{x_2} \right)$$
$$= \frac{n_1}{n} \bar{x_1} + = \frac{n_2}{n} \bar{x_2}$$
$$= w_1 \bar{x_1} + = w_2 \bar{x_2}$$

(b) Yes, they represent the relative portions of the sample space in $i = 1, \ldots, n_1$ and $i = n_1 + 1, \ldots, n$.

(c) The case in part (a) applies to all cases for $g \in \mathbb{Z}^+$.

13. (a) No, take the sample $\{x_1, x_2\} = \{1, 2\}$. Then,

$$\sum_{i=1}^{n} \frac{1}{x_i} = \frac{1}{2} + \frac{1}{2} = 1$$

while

$$\frac{1}{\sum_{i=1}^{n} x_i} = \frac{1}{2+2} = \frac{1}{4}$$

(b) No, see part (a) where $x_i = 2 \; \forall \; i$.

Math Review B

Notes

Experiments

An **experiment** is a procedure that can theoretically be conducted an infinite number of times and has a well-defined set of outcomes.

A **random variable** is a variable that takes on numerical values and has an outcome determined by an experiment.

Variables

A **Bernoulli** (or **binary) random variable** is a random variable that can only take on the values zero and one.

A **discrete random variable** is one that takes on only a finite or countably infinite number of values. A Bernoulli random variable is the simplest example of a discrete random variable.

A **continuous random variable** is a random variable that takes on any real value with *zero* probability.

Density Functions

A **probability density function (pdf)** summarizes the information concerning the possible outcomes of a random variable and the corresponding probabilities. A pdf of a random variable X is generally denoted $f(x)$ or $f_x \equiv P(X = x)$.

A **cumulative distribution function (cdf)** is a function that describes the cumulative probability that a random variable's value is less than (or equal to, if continuous) a given value. A cdf of a random variable X is generally denoted $F(x)$ or $F_x \equiv P(X \leq x)$

Cumulative Distribution Function Properties:

1. For any number c, $P(X > c) = 1 - F(c)$
2. For any numbers $a < b$, $P(a < X \leq b) = F(b) - F(a)$
3. For a continuous random variable X, $P(X \geq c) = P(X > c)$
4. For a continuous random variable X, $P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b)$

Independence

Let X and Y be discrete random variables. Then, (X, Y) have a **joint distribution**, which is fully described by the **joint probability density function** of (X, Y):

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

Two random variables X and Y are said to be **independent** if, and only if,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

or

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all x and y. The pdfs f_X and f_Y are often called the **marginal probability density functions** to distinguish the from the joint pdf $f_{X,Y}$.

Beyond the case of two random variables, the same concept applies. Random variables $X_1, X_2, \ldots, X_n$ are **independent random variables** if, and only if, their joint pdf is the product of the individual pdfs for any $(x_1, x_2, \ldots, x_n)$. This definition of independence holds for both continuous and discrete random variables.

Given independent outcomes with 'success' rate θ, the pdf ($X \sim \text{Binomial}(n, \theta)$) is equal to $f(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ where $\binom{n}{x} = {}^nC_x = \frac{n!}{x!(n-x)!}$.

Conditional Distributions

The **conditional distribution** of Y given X is summarized by the **conditional probability density function**, defined by

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

for all values of x such that $f_X(x) > 0$. The interpretation of the conditional probability density function is

$$f_{Y|X}(y|x) = P(Y = y|X = x)$$

Expected Values

If X is a random variable, the **expected value** (or **expectation**) of X, denoted $E[X]$ and sometimes μ_X or simply μ, is a weighted average of all possible values of X. The weights are determined by the probability distribution function. Sometimes, the expected value is called the *population mean*, especially when we want to emphasize that X represents some variable in a population. For a discrete random variable X that takes on values $\{x_1, \ldots, x_n\}$,

$$E[X] = x_1 f(x_1) + \cdots + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i)$$

If X is a continuous random variable, then $E[X]$ is defined through an integral as

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx,$$

which we assume is well-defined.

Expected Values Properties

1. For a constant c, $E[c] = c$
2. For any constants a and b, $E[aX + b] = aE[x] + b$
3. If $\{a_1, a_2, \ldots, a_n\}$ are constants and $\{X_1, X_2, \ldots, X_n\}$ are random variables, then

$$E[a_1 X_1 + a_2 X_2 + \cdots + a_n E[X_n] = E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E[X_i]$$

For $X \sim \text{Binomial}(n, \theta)$, we can rewrite X as $Y_1 + \cdots + Y_n$, where each $Y_i \sim \text{Bernoulli}(\theta)$. Then,

$$E[X] = \sum_{i=1}^{n} E[Y_i] = \sum_{i=1}^{n} \theta = n\theta$$

Variance

For a random variable X,

$$\text{Var}(x) = E[(X - \mu)^2]$$

The **variance** tells us the expected distance from X to its mean and is sometimes denoted σ_x^2 or just σ^2. For a Bernoulli random variable X

$$\sigma_x^2 = E[X^2] - E[X]^2 = \theta - \theta^2 = \theta(1 - \theta)$$

Variance Properties

1. $\text{Var}(x) = 0$ if, and only if, there is a constant c such that $P(X = c) = 1$, in which case $E[X] = c$. In other words, this first property says that the variance of any constant is zero and if a random variable has zero variance, then it is essentially constant.
2. For any constants a and b, $\text{Var}(aX + b) = a^2 \text{Var}(x)$
3. For any constants a and b,

$$\text{Var}(aX + bY) = a^2 \text{Var}(x) + 2ab\text{Cov}(X, Y) + b^2 \text{Var}(Y)$$

4. If $\{X_1, \ldots, X_n\}$ are pairwise uncorrelated random variables and $a_i : i = 1, \ldots, n$ are constants then

$$\text{Var}(a_1 X_1 + \cdots + a_n X_n) = a_1^2 \text{Var}(X_1) + \cdots + a_n^2 \text{Var}(X_n) = \text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i)$$

Standard Deviation

The **standard deviation** of a random variable, denoted $\text{sd}(X)$, is simply the positive square root of the variance: $\text{sd}(X) = +\sqrt{\text{Var}(X)}$. The standard deviation is sometimes denoted σ_X, or simply σ, when the random variable is understood.

Standard Deviation Properties

1. For any constant c, $\text{sd}(c) = 0$.
2. For any constants a and b,
$$\text{sd}(aX + b) = |a|\text{sd}(X) = |a|\sigma_X$$

Standardized Random Variables

If X is a random variable, we can redefine a random variable

$$Z \equiv \frac{X - \mu}{\sigma},$$

which we can write as $Z = aX + b$, where $a \equiv (1/\sigma)$ and $b \equiv -(\mu/\sigma)$. Then,

$$E[Z] = aE[X] + b = (\mu/\sigma) - (\mu/\sigma) = 0$$

and

$$\text{Var}(Z) = a^2\text{Var}(X) = 1$$

Thus, the random variable Z has a mean of zero and a variance (and therefore a standard deviation) equal to one. This procedure is sometimes known as standardizing the random variable X, and Z is called a **standardized random variable**.

Skewness and Kurtosis

We can use the standardized version of a random variable to define other features of the distribution of a random variable. These features are described by using what are called *higher order moments*. For example, the third moment of the standardized random variable Z is used to determine whether a distribution is symmetric about its mean. We can write

$$E[Z^3] = \frac{E[(X - \mu)^3]}{\sigma^3}$$

Generally, $\frac{E[(X-\mu)^3]}{\sigma^3}$ is viewed as a measure of **skewness** in the distribution of X. If X has a symmetric distribution about μ, then Z has a symmetric distribution about zero. That means the density of Z at any two points z and $-z$ is the same.

It also can be informative to compute the fourth moment of Z

$$E[Z^4] = \frac{E[(X - \mu)^4]}{\sigma^4}$$

The fourth moment $E[Z^4]$ is called a measure of **kurtosis** in the distribution of X. Generally, larger values mean that the tails in the distribution of X are thicker.

Covariance and Correlation

The **covariance** between two random variables X and Y, sometimes called the *population covariance* to emphasize that it concerns the relationship between two variables describing a population, is defined as the expected value of the product $(X - \mu_X)(Y - \mu_Y)$:

$$\text{Cov}(X, Y) \equiv E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

which is sometimes denoted σ_{XY}. If $\sigma_{XY} > 0$, then, on average, when X is above its mean, Y is also above its mean. If $\sigma_{XY} < 0$, then, on average, when X is above its mean, Y is below its mean. Note that

$$\begin{aligned}\text{Cov}(X,Y) &= E\left[(X-\mu_X)(Y-\mu_Y)\right] = E\left[(X-\mu_X)Y\right] \\ &= E\left[X(Y-\mu_Y)\right] = E[XY] - E[X]E[Y]\end{aligned}$$

Covariance measures the amount of linear dependence between two random variables. A positive covariance indicates that two random variables move in the same direction, while a negative covariance indicates they move in opposite directions.

Covariance Properties

1. If X and Y are independent, then $\text{Cov}(X,Y) = 0$

   This property stems from the fact that $E[XY] = E[X]E[Y]$ when X and Y are independent. It is important to remember that the converse of is not true: zero covariance between X and Y does not imply that X and Y are independent.

2. For any constants a_1, b_1, a_2, and b_2,

   $$\text{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\text{Cov}(X,Y)$$

3. From the **Cauchy-Schwartz inequality**:

   $$|\text{Cov}(X,Y)| \leq \text{sd}(X)\text{sd}(Y)$$

The fact that the covariance depends on units of measurement is a deficiency that is overcome by the **correlation coefficient** between X and Y:

$$\text{Corr}(X,Y) \equiv \frac{\text{Cov}(X,Y)}{\text{sd}(X)\text{sd}(Y)} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

the correlation coefficient between X and Y is sometimes denoted ρ_{XY} (and is sometimes called the *population correlation*).

Correlation Properties

1. $-1 \leq \text{Corr}(X,Y) \leq 1$
2. For any constants a_1, b_1, a_2, and b_2,

   $$\text{Corr}(a_1X + b_1, a_2Y + b_2) = \text{Corr}(X,Y)$$

   if $a_1a_2 > 0$ or

   $$\text{Corr}(a_1X + b_1, a_2Y + b_2) = -\text{Corr}(X,Y)$$

   if $a_1a_2 < 0$.

Conditional Expectation

The **conditional expectation** of a random variable is the expected or average value of one random variable, called the dependent or explained variable, that depends on the values of one or more other variables, called the independent or explanatory variables. When Y is a discrete random variable

$$E[Y|x] = \sum_{i=1}^{n} y_i f_{Y|x}(y_i|x)$$

When Y is a continuous random variable

$$E[Y|x] = \int_{-\infty}^{\infty} y_i f_{Y|x}(y_i|x)dy$$

Conditional Expecation Properties

1. $E[c(X)|X] = c(X)$, for any function $c(X)$.
2. For any functions $a(X)$ and $b(X)$,
$$E[a(X)Y + b(X)|X] = a(X)E[Y|X] + b(X)$$
3. If X and Y are independent, $E[Y|X] = E[Y]$.
4. From the **law of iterated expectations** $E[E[Y|X]] = E[Y]$.
5. From a more general version of the law of iterated expectation $E[Y|X] = E[E[Y|X, Z]|X]$.
6. If $E[Y|X] = E[Y]$, then $\text{Cov}(X, Y) = 0$.
7. If $E[Y^2] < \infty$ and $E[g(X)^2] < \infty$ for some function g, then $E[[Y - E[Y|X]]^2|X] \leq E[[Y - g(X)]^2|X]$ and $E[[Y - E[Y|X]]^2] \leq E[[Y - g(X)]^2]$. This property is very useful in predicting or forecasting contexts. The first inequality says that, if we measure prediction inaccuracy as the expected squared prediction error, conditional on X, then the conditional mean is better than any other function of X for predicting Y. The conditional mean also minimizes the unconditional expected squared prediction error.

Conditional Variance

Given random variables X and Y, the variance of Y, conditional on $X = x$, is simply the variance associated with the conditional distribution of Y, given $X = x : E[[Y - E[Y|x]]^2|x]$. The formula can be rewritten as
$$\text{Var}(Y|X = x) = E[Y^2|x] - E[Y|x]^2$$

Conditional Variance Properties

1. If X and Y are independent, then $\text{Var}(Y|X) = \text{Var}(Y)$

Normal Distribution

A **normal random variable** is a continuous random variable that can take on any value. Its probability density function has the familiar bell-shaped graph and can be written as
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp[-(x - \mu)^2/2\sigma^2], \ -\infty < x < \infty$$

We say that X has a **normal distribution** with expected value μ and variance σ^2, written as $X \sim \mathcal{N}(\mu, \sigma^2)$.Because the normal distribution is symmetric about μ, μ is also the median of X. The normal distribution is also sometimes called the Gaussian distribution after Carl Friedrich Gauss.

One special case of the normal distribution is the **standard normal distribution** where the mean is zero and the variance is unity. If a random variable Z has a Normal(0,1) distribution, then we say it has a standard normal distribution, and its pdf is given by
$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp(-z^2/2), \ -\infty < z < \infty$$

Normal Distribution Properties

1. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$
2. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$
3. If X and Y are jointly normally distributed, then they are independent if, and only if, $\text{Cov}(X, Y) = 0$
4. Any linear combination of independent, identically distributed normal random variables has a normal distribution.

Chi-Square Distribution

The chi-square distribution is obtained directly from independent, standard normal random variables. Let $Z_i, i = 1, 2, \ldots, n$ be independent random variables, each distributed as standard normal. Define a new random variable as the sum of the squares of the Z_i:

$$X = \sum_{i=1}^{n} Z_i^2$$

Then, X has what is known as a **chi-square distribution** with n **degrees of freedom**. We write this as $X \sim \chi_n^2$ where the expected value of X is n and the variance of X is $2n$.

t Distribution

A **t distribution** is obtained from a standard normal and a chi-square random variable. Let Z have a standard normal distribution and let X have a chi-square distribution with n degrees of freedom. Further, assume that Z and X are independent. Then, the random variable

$$T = \frac{Z}{\sqrt{X/n}}$$

has a t distribution with n degrees of freedom. This is denoted by $T \sim t_n$ where n comes from the degrees of freedom of the chi-square random variable in the denominator. The pdf of the t distribution has a shape similar to that of the standard normal distribution (maintaining a zero expected value), except that it is more spread out (with a variance of $n/(n-2)$ and therefore has more area in the tails. As the degrees of freedom gets large, the t distribution approaches the standard normal distribution.

F Distribution

To define an F random variable, let $X_1 \sim \chi_{k_1}^2$, and $X_2 \sim \chi_{k_2}^2$ and assume that X_1 and X_2 are independent. Then, the random variable

$$F = \frac{X_1/k_1}{X_2/k_2}$$

has an **F distribution** with (k_1, k_2) degrees of freedom. This is denoted as $F \sim F_{k_1,k_2}$

Exercises

1. His or her eventual SAT score is viewed as a random variable because his or her score is a variable that takes on numerical values and has an outcome determined by an experiment (the test). The test score is stochastic as it can change from day-to-day or depending on other various circumstances/conditions.
2. (a) $P(X \leq 6)$ =0.6914625
   (b) $P(X > 4) = 1 - P(X \leq 4)$ =0.6914625
   (c) $P(|X - 5| > 1) = P([X - 5 > 1] \text{ or } [X - 5 < -1]) = 1 - P(4 < X < 6) = 1 - (P(X < 6) - P(X < 4))$ =0.6170751
3. (a) 0.0009766 or 0.0976562%
   (b) 4.0722656 mutual funds
   (c) $P(\text{At Least One}) = 1 - P(\text{None}) = 1 - (1 - .5^{10})^{4170}$ =0.9829951
   Similarly, this equals $1 - \binom{4170}{0}(.5^{10})^0(1 - .5^{10})^{4170}$ =0.9829951
   (d) $P(X \geq 5) = 1 - P(X < 4)$ =0.3852852
4. $P(X \geq .6) = 1 - P(X < .6) = F(.6)$ =1.512

5. (a) $P(\text{At least one}) = 1 - P(\text{None}) = 1 - \binom{12}{0}(.2)^0(.8)^{12}$ =0.9312805

   (b) $P(X \geq 2) = 1 - P(X < 1) = 1 - \binom{12}{1}(.2)^1(.8)^{11}$ =0.7251221

6. $E[X] = \int_0^3 \frac{x^2}{9} x dx = \frac{1}{9}\int_0^3 x^3 dx = \frac{1}{9}\left[\frac{x^4}{4}\right]_0^3 = \frac{81}{36} = \frac{9}{4}$

7. $E[\text{Made FTs}] = .74 * 8$=5.92

8. $E[GPA] = 3.5(\frac{2}{9}) + 3(\frac{7}{9}) = \frac{28}{9} \approx$3.11

9. $E[\text{salary}] = 52.3 \times 1000$ =\$52300

   $\sigma_{\text{salary}} = |1000| \times 14.6$\$14600

10. (a) $E[GPA|SAT = 800] = .70 + .002(800)$ =2.3

    $E[GPA|SAT = 1400] = .70 + .002(800)$ =3.5 The difference in expected GPAs is fairly large, but the difference in SAT scores is also rather large. I don't feel these estimates are entirely unreasonable.

    (b) $E[GPA] = E[E[GPA|SAT]] = E[.70 + .002(1100)] = .70 + .002(1100)$ =2.9

    (c) No, we don't know any particular student's GPA given his or her SAT score. The provided formula only allows us to derive an expected GPA given an SAT score.

11. (a) $E[X] = 1/2(-1) + 1/2(1) = 0$

    $E[X^2] = 1/2(-1)^2 + 1/2(1)^2 = 1$

    (b) $E[X] = 1/2(1) + 1/2(2) = 3/2$

    $E[1/X] = 1/2(1) + 1/2(1/2) = 3/4$

    (c) From part(a), $E[X^2] = 1 \neq (E[X])^2 = 0^2 = 0$

    From part(b), $E[1/X] = 3/4 \neq (1/E[X]) = 2/3$

    (d)

$$E[F] = E\left[\frac{X_1/k_1}{X_2/k_2}\right]$$

    Because k_1 and k_2 are constants,

$$= \frac{k_2}{k_1} E\left[\frac{X_1}{X_2}\right]$$

    Using the fact that X_1 and X_2 are assumed independent

$$= \frac{k_2}{k_1} E[X_1] E[X_2^{-1}]$$

    Using the fact that X_1 is a chi-square random variable (and thus has a mean of k_1)

$$= \frac{k_2}{k_1} k_1 1 E[X_2^{-1}] = k_2 E[X_2^{-1}] = E\left[\frac{k_2}{X_2}\right] = E\left[\frac{1}{X_2/k_2}\right]$$

    As we showed in parts (a-c), $E\left[\frac{1}{X_2/k_2}\right]$ has a nonlinear 'internal' function, and thus, we cannot conclude that $E[F] = 1$.

Math Review C

Notes

Populations, Parameters, and Random Sampling

A **population** is any well-defined group of subjects, which could be individuals, firms, cities, or many other possibilities.

A **random sample** is a sample obtained by sampling randomly from the specified population. In particular, no unit is more likely to be selected than any other unit, and each draw is independent of all other draws.

When $\{Y_1, \ldots, Y_n\}$ is a random sample from the density $f(y;\theta)$, we also say that the Y_i are **independent, identically distributed** (or **i.i.d.**) random variables from $f(y;\theta)$.

Estimators

An **estimator** is a rule for combining data to produce a numerical value for a population parameter; the form of the rule does not depend on the particular sample obtained. More generally, an estimator W of a parameter θ can be expressed as an abstract mathematical formula:

$$W = h(Y_1, Y_2, \ldots, Y_n)$$

for some known function h of the random variables $Y_1, Y_2, \ldots, Y_n$

Unbiasedness

An estimator, W of θ, is an **unbiased estimator** if

$$E[W] = \theta$$

for all possible values of θ.

The distribution of an estimator is often called its **sampling distribution**, because this distribution describes the likelihood of various outcomes of W across different random samples.

If W is a **biased estimator** of θ, its bias is defined

$$\text{Bias}(W) \equiv E[W] - \theta$$

The **sample average** is an unbiased estimator of the population variance and is defined as

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Proof of Unbiasedness

$$E[\bar{Y}] = E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right]$$

$$= \frac{1}{n}E\left[\sum_{i=1}^{n} Y_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[Y_i] = \frac{1}{n}(n\mu) = \mu$$

The **sample variance** is an unbiased estimator of the population variance and is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Proof of Unbiasedness

$$E[S^2] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right]$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}Y_i^2 - 2\sum_{i=1}^{n}Y_i\bar{Y} + \sum_{i=1}^{n}\bar{Y}^2\right]$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}Y_i^2 - 2\bar{Y}\sum_{i=1}^{n}Y_i + \bar{Y}\sum_{i=1}^{n}\bar{Y}\right]$$

$$= \frac{1}{n-1}E\left[\sum_{i=1}^{n}Y_i^2 - 2n\bar{Y}^2 + n\bar{Y}^2\right]$$

$$= \frac{1}{n-1}\left\{\sum_{i=1}^{n}E[Y_i^2] - nE[\bar{Y}^2]\right\}$$

Using the facts that $\mathrm{Var}(\bar{Y}) = \frac{\sigma_Y}{n}$ and $\sigma_Y = E[Y_i^2] - E[Y_i]^2$,

$$= \frac{1}{n-1}\left\{\sum_{i=1}^{n}\left(\sigma_Y + E[Y_i]^2\right) - n\left(\frac{\sigma_Y}{n} + E[\bar{Y}]^2\right)\right\}$$

$$= \frac{1}{n-1}\left\{n\sigma_Y + n\mu_Y^2 - \sigma_Y - n\mu_Y^2\right\}$$

$$= \frac{1}{n-1}\left[(n-1)\sigma_Y\right] = \sigma_Y$$

The **sample covariance** is defined as

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

and is an unbiased and consistent estimator of σ_{XY}

The **sample correlation coefficient** is defined as

$$R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

and is a consistent but biased estimator of ρ_{XY}. Because S_{XY}, S_X, and S_Y are consistent for the corresponding population parameter, R_{XY} is a consistent estimator of the population correlation, ρ_{XY}. However, R_{XY} is a biased estimator for two reasons. First, S_X and S_Y are biased estimators of σ_X and σ_Y, respectively. Second, R_{XY} is a ratio of estimators, so it would not be unbiased, even if S_X and S_Y were.

Efficiency

The variance of an estimator is often called its **sampling variance** because it is the variance associated with a sampling distribution. Remember, the sampling variance is not a random variable; it is a constant, but it might be unknown.

An estimator, W_1 is **efficient** relative to W_2 when $\text{Var}(W_1) \leq \text{Var}(W_2)$ for all θ, with strict inequality for at least one value of θ.

One way to compare estimators that are not necessarily unbiased is to compute the **mean squared error (MSE)** of the estimators. If W is an estimator of θ, then the MSE of W is defined as

$$\text{MSE}(W) = E[(W - \theta)^2]$$

The MSE measures how far, on average, the estimator is away from θ. It can be shown that $\text{MSE}(W) = \text{Var}(W) + [\text{Bias}(W)]^2$, so that $\text{MSE}(W)$ depends on the variance and bias (if any is present). This allows us to compare two estimators when one or both are biased.

Consistency

Let W_n be an estimator of θ based on a sample $Y_1, Y_2, \ldots, Y_n$ of size n. Then, W_n is a **consistent estimator** of θ if for every $\epsilon > 0$,

$$P(|W_n - \theta| > \epsilon) \to 0 \text{ as } n \to \infty$$

When W_n is consistent, we also say that θ is the probability limit of W_n, written as $\text{plim}(W_n) = \theta$. Unlike unbiasedness—which is a feature of an estimator for a given sample size—consistency involves the behavior of the sampling distribution of the estimator as the sample size n gets large.

Asymptotic Unbiasedness ← Consistency + Bounded Variance

Consider an estimator W_n for a parameter θ. Asymptotic unbiasedness means that the bias of the estimator goes to zero as $n \to \infty$, which means that the expected value of the estimator converges to the true value of the parameter. Consistency is a stronger condition than this; it requires the estimator (not just its expected value) to converge to the true value of the parameter (with convergence interpreted in various ways). Since there is generally some non-zero variance in the estimator, it will not generally be equal to (or converge to) its expected value. Assuming the variance of the estimator is bounded, consistency ensures asymptotic unbiasedness, but asymptotic unbiasedness is not enough to get consistency. To put it another way, under some mild conditions, asymptotic unbiasedness is a necessary but not sufficient condition for consistency.

Asymptotic Unbiasedness + Vanishing Variance → Consistency

If you have an asymptotically unbiased estimator, and its variance converges to zero, this is sufficient to give weak consistency. (This follows from Markov's inequality, which ensures that convergence in mean-square implies convergence in probability). Intuitively, this reflects the fact that a vanishing variance means that the sequence of random variables is converging closer and closer to the expected value, and if the expected value converges to the true parameter (as it does under asymptotic unbiasedness) then the random variable is converging to the true parameter.

More simply, unbiased estimators are not necessarily consistent, but those whose variances shrink to zero as the sample size grows are *consistent*. For example, the sample variance and standard deviation formulas without **Bessel's correction** are biased estimators; however, they are also consistent because the converge in probability toward their population values as $n \to \infty$.

The **law of large numbers (LLN)** is a theorem that states the average from a random sample converges in probability to the population average. It also holds for stationary and weakly dependent time series. This result comes from the fact that $\text{Var}(\bar{Y}) = \frac{\sigma_Y}{n}$, which approaches 0 as $n \to \infty$.

Plim Properties

1. If θ is a parameter and $\gamma = g(\theta)$ is a newly-defined parameter for some continuous function $g(\theta)$. If $\text{plim}(W_n) = \theta$, then the estimator of $\gamma, G_n = g(W_n)$, has a plim defined by

$$\text{plim}(G_n) = \gamma$$

This is often stated as

$$\text{plim}(g(W_n)) = g\left(\text{plim}(W_n)\right)$$

2. If $\text{plim}(T_n) = \alpha$ and $\text{plim}(U_n) = \beta$, then
   (a) $\text{plim}(T_n + U_n) = \alpha + \beta$
   (b) $\text{plim}(T_n U_n) = \alpha\beta$
   (c) $\text{plim}(T_n/U_n) = \alpha/\beta$, if $\beta \neq 0$

Asymptotic Normality

Let $\{Z_n : n = 1, 2, \dots\}$ be a sequence of random variables, such that for numbers z,

$$P(Z_n \leq z) \to \Phi(z) \text{ as } n \to \infty$$

where $\Phi(z)$ is the standard normal cumulative distribution function. Then, Z_n is said to have an **asymptotic standard normal distribution**. This is sometimes written as $Z_n \overset{\text{a}}{\sim} \mathcal{N}(0, 1)$, where the "$a$" stands for "asymptotically" or "approximately."

The **central limit theorem (CLT)** states that the average from a random sample (and many other estimates that depend on the sample mean) for any population (with finite variance), when standardized, has an asymptotic standard normal distribution. More formally, for a random sample $\{Y_1, Y_2, \dots, Y_n\}$ with a mean μ and a variance σ^2. Then,

$$Z_n = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

has an asymptotic standard normal distribution. Note that Z_n is the standardized version of $\bar{Y_n}$: $E[\bar{Y_n}] = \mu$ has been subtracted off and divided by $\text{sd}(\bar{Y_n}) = \sigma/\sqrt{n}$.

Maximum Likelihood

The **maximum likelihood estimator** of θ, call it W, is the value of θ that maximizes the **likelihood function**

$$L(\theta; Y_1, Y_2, \dots, Y_n) = f(Y_1; \theta) f(Y_2; \theta) \dots f(Y_n; \theta)$$

which equals $P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n)$ in the discrete case. Usually, it is more convenient to work with the **log-likelihood function**, which is obtained by taking the natural log of the likelihood function:

$$\mathscr{L}(\theta) = \ln\left(L(\theta; Y_1, Y_2, \dots, Y_n)\right) = \sum_{i=1}^{n} \ln[f(Y_i, \theta)] = \sum_{i=1}^{n} \ell(\theta; X_i)$$

where we use the fact that the log of the product is the sum of the logs.

Least Squares

A least squares estimator is an estimator of a parameter that minimizes the sum of squared differences. That is, an estimator, W is a least squares estimator if it minimizes

$$\sum_{i=1}^{n} (W - \theta)^2$$

It should be noted that the principles of least squares, method of moments, and maximum likelihood often result in the same estimator. In other cases, the estimators are similar but not identical.

Confidence Intervals

A **condidence interval** is a rule used to construct a random interval so that a certain percentage of all data sets, determined by the confidence level, yields an interval that contains the population value. Thus, a 95% confidence interval of an estimator will contain the true population value 95% of the time. Theoretically, for the sample average the 95% confidence interval can be constructed as follows:

$$P\left(-1.96 < \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = .95$$

$$\rightarrow CI_{95} = [\bar{y} - 1.96(\sigma/\sqrt{n}), \bar{y} + 1.96(\sigma/\sqrt{n})]$$

where μ is the hypothesized population mean. In practice, however, σ is unknown and must be estimated with s. Unfortunately, this does not preserve the 95% level of confidence because s depends on the particular sample. In other words, the random interval $[\bar{Y} \pm 1.96(S/\sqrt{n})]$ no longer contains μ with probability .95 because the constant σ has been replaced with the random variable S. Thus, rather than using the standard normal distribution, we must rely on the t distribution. The t distribution arises from the fact that

$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

The denominator $S/\sqrt{N}$ is an estimate of the $\text{sd}(\bar{Y})$. In general, these estimates of **sampling standard deviations** are referred to as **standard errors**.

Hypothesis Testing

A **Type I error** is an error in which a true null hypothesis is rejected.

A **Type II error** is an error in which one fails to reject a false null hypothesis.

A **significance level** is the probability of Type I error, which is generally denoted

$$\alpha = P(\text{Reject } H_0 | H_0)$$

A p**-value** is the *largest* significance level at which we could carry out a test and still fail to reject the null hypothesis. Formally,

$$p - \text{value} = P(T > t | H_0) = 1 - \Phi(t)$$

where $\Phi(\cdot)$ is the standard normal cdf.

Exercises

1. (a)

$$E[\bar{Y}] = E[\frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)] = \frac{1}{4}\big\{E[Y_1] + E[Y_2] + E[Y_3] + E[Y_4]\big\} = \frac{1}{4}(4\mu) = \mu$$

$$\text{Var}(\bar{Y}) = \text{Var}\left(\frac{1}{4}(Y_1 + Y_2 + Y_3 + Y_4)\right) = \frac{1}{16}\sum_{i=1}^{4} Var(Y_i) = \frac{1}{16}\sum_{i=1}^{4}\sigma^2 = \frac{\sigma^2}{4}$$

(b)

$$E[W] = E\left[\frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4\right]$$
$$= \frac{1}{8}E[Y_1] + \frac{1}{8}E[Y_2] + \frac{1}{4}E[Y_3] + \frac{1}{2}E[Y_4]$$
$$= \frac{1}{8}\mu + \frac{1}{8}\mu + \frac{1}{4}\mu + \frac{1}{2}\mu = \mu$$

$$\text{Var}(W) = \text{Var}\left(\frac{1}{8}Y_1 + \frac{1}{8}Y_2 + \frac{1}{4}Y_3 + \frac{1}{2}Y_4\right)$$
$$= \frac{1}{64}\text{Var}(Y_1) + \frac{1}{64}\text{Var}(Y_2) + \frac{1}{16}\text{Var}(Y_3) + \frac{1}{4}\text{Var}(Y_4) = \frac{11}{32}\sigma^2$$

(c) I prefer $\bar{Y}$ over W because $\bar{Y}$ is a more efficient unbiased estimator.

Note: Parts (a and b) make use of the fact that the sample is iid and thus the variance of the sums of the variables is equal to the sum of the variances of the variables.

2. (a) $\sum_{i=1}^{n} a_i = 1$

(b)
$$\text{Var}(W_a) = \text{Var}\left(\sum_{i=1}^{n} a_i Y_i\right) = \sum_{i=1}^{n} \text{Var}(a_i Y_i) = a_1^2\sigma^2 + \cdots + a_n^2\sigma^2 = \sigma^2 \sum_{i=1}^{n} a_i^2$$

$$\min_{a_1,\ldots,a_n} \text{Var}(W_a) = a_1^2\sigma^2 + \cdots + a_n^2\sigma^2 \text{ s.t. } a_1 + \cdots + a_n = 1$$
$$\mathscr{L} = a_1^2\sigma^2 + \cdots + a_n^2\sigma^2 + \lambda(1 - a_1 - \cdots - a_n)$$
$$\frac{\partial \mathscr{L}}{\partial a_1} = 2a_1\sigma^2 - \lambda = 0$$
$$\cdots$$
$$\frac{\partial \mathscr{L}}{\partial a_n} = 2a_n\sigma^2 - \lambda = 0$$
$$\frac{\partial \mathscr{L}}{\partial \lambda} = 1 - a_1 - \cdots - a_n = 0$$
$$\rightarrow a_1 = \cdots = a_n \text{ and } \sum_{i=1}^{n} a_i = 1$$
$$\rightarrow \sum_{i=1}^{n} a_1 = 1$$
$$\rightarrow a_1^* = \cdots = a_n^* = \frac{1}{n}$$

3. (a) $E[W_1] = E\left[\frac{n-1}{n}(\bar{Y})\right] = \frac{n-1}{n}E[\bar{Y}] = \frac{n-1}{n}(\mu) \neq \mu$

$\text{bias}(W_1) = E[W_1] - \mu = \frac{n-1}{n}(\mu) - \mu = \frac{-\mu}{n}$

$E[W_2] = E\left[\frac{\bar{Y}}{2}\right] = \frac{1}{2}E[\bar{Y}] = \frac{\mu}{2}$

$\text{bias}(W_2) = E[W_2] - \mu = \frac{\mu}{2} - \mu = \frac{-\mu}{2}$

One important difference is that the bias of W_1 converges to 0 as $n \rightarrow \infty$, while the bias of W_2 is constant.

(b) $\text{plim}(W_1) = \text{plim}\left(\frac{n-1}{n}(\bar{Y})\right) = \text{plim}\left(\frac{n-1}{n}\right)\text{plim}(\bar{Y}) = 1 \cdot \mu = \mu$ $\text{plim}(W_2) = \text{plim}\left(\frac{\bar{Y}}{2}\right) = \text{plim}(\bar{Y})/\text{plim}(2) = \frac{\mu}{2}$

W_1 is consistent.

(c)

$$\text{Var}(W_1) = \text{Var}\left(\frac{n-1}{n}(\bar{Y})\right) = \left(\frac{n-1}{n}\right)^2 \text{Var}(\bar{Y}) = \left(\frac{(n-1)^2\sigma^2}{n^3}\right)$$

$$\text{Var}(W_2) = \text{Var}\left(\frac{\bar{Y}}{2}\right) = \frac{1}{4}\text{Var}(\bar{Y}) = \frac{\sigma^2}{4n}$$

(d) While $\bar{Y}$ is unbiased regardless of the value of μ, when μ is "close" to zero, the bias of W_1 is also close to zero (especially for large samples). Thus, it may be worthwhile to consider W_1 over $\bar{Y}$ if W_1 is efficient relative to $\bar{Y}$. We know $\text{Var}(\bar{Y}) = \frac{\sigma^2}{n}$ and $\text{Var}(W_1) = \left(\frac{(n-1)^2\sigma^2}{n^3}\right)$. Using these calculations to evaluate when W_1 is efficient relative to $\bar{Y}$:

$$\left(\frac{(n-1)^2\sigma^2}{n^3}\right) \leq \frac{\sigma^2}{n}$$

$$\left(\frac{n-1}{n}\right)^2 \leq 1$$

which holds for all positive values of n. Thus, W_1 is efficient relative to $\bar{Y}$ and, given the small amount of bias, it may be a better estimator of μ.

4. (a) $E[Z] = E\left[E[Z|X]\right] = E\left[E\left[\frac{Y}{X}|X\right]\right] = E\left[\frac{1}{X}E\left[Y|X\right]\right] = E\left[\frac{1}{X}\theta X\right] = E[\theta] = \theta$

(b) $E[W_1] = E\left[n^{-1}\sum_{i=1}^n (Y_i/X_i)\right] = n^{-1}\sum_{i=1}^n E\left[(Y_i/X_i)\right] = n^{-1}\sum_{i=1}^n \theta = n^{-1}(n\theta) = \theta$

(c) In general, the average of the ratios, Y_i/X_i, is not the ratio of the averages $\bar{Y}/\bar{X}$.

$$E[W_2|X_1,\ldots,X_2] = E\left[\frac{\bar{Y}}{\bar{X}}|X_1,\ldots,X_2\right]$$

$$= \frac{1}{\bar{X}}E[\bar{Y}|X_1,\ldots,X_n]$$

$$= \frac{1}{\bar{X}}E\left[n^{-1}\sum_{i=1}^n Y_i|X_1,\ldots,X_n\right]$$

$$= \frac{1}{n\bar{X}}\sum_{i=1}^n E[Y_i|X_1,\ldots,X_n]$$

$$= \frac{1}{n\bar{X}}\left(n\theta\bar{X}\right) = \theta$$

(d) $W_1 = n^{-1}\sum_{i=1}^n (Y_i/X_i)$ =0.4179674

$W_2 = \frac{\bar{Y}}{\bar{X}}$ =0.4180968

Yes, they are similar.

5. (a) G is not an unbiased estimator of γ because G has a nonlinear relationship with $\bar{Y}$. As we concluded in Math Review B, the expected value of the ratio is not the ratio of the expected value.

(b) $\text{plim}(G) = \text{plim}\left(\frac{\bar{Y}}{1-\bar{Y}}\right) = \text{plim}(\bar{Y})/\text{plim}(1-\bar{Y}) = \theta/(\text{plim}(1) - \text{plim}(\bar{Y})) = \frac{\theta}{1-\theta} = \gamma$

6. (a)

$$H_0 : \mu = 0$$

(b)

$$H_1 : \mu < 0$$

(c) $t = \frac{\bar{y}-\mu}{s/\sqrt{n}} = \frac{-32.8-0}{466.4/\sqrt{900}} \approx$-2.109777

$p \approx$0.0175767

We reject the null hypothesis at the 5% level but fail to reject H_0 at the 1% level.

(d) We've already shown there is a statistically significant difference at the 5% level but not at the 1% level. On the other hand, I would struggle to argue there is a practical significance when there is only a 32.8 ounce difference in alcohol consumption over an entire year.

(e) This analysis implicitly assumes all other factors that affect liquor consumption have remained the same. Factors such as such as income, or changes in price due to transportation costs, are assumed constant over the two years.

7. (a) CI_{95} = [-0.0096847,0.4896847]

(b)

$$H_0 : \mu = 0$$
$$H_1 : \mu > 0$$

(c) $t = \frac{\bar{d}-0}{s_d/\sqrt{n}} =$2.0615954

For a one-sided test, we would reject H_0 at the 5% level but fail to reject H_0 at the 1% level.

(d) $p =$0.0291614

8. (a) $\bar{Y} = \frac{188}{429} \approx$0.4382284

(b)

$$\text{sd}(\bar{Y}) = \frac{\sigma_\theta}{\sqrt{n}} = \sqrt{\frac{\theta(1-\theta)}{n}}$$

(c) $t = \frac{\bar{Y}-.5}{\text{se}(\bar{Y})} = \frac{\bar{Y}-.5}{\sqrt{\bar{Y}(1-\bar{Y})/n}} =$-2.5786184

$p =$0.0051263

Thus, we reject H_0 at the 1% level.

9. (a) $E[X] = 200(.65) = 130$

(b) $\text{sd}(X) = \sqrt{|200| \times .65(1-.65)} =$6.7453688

(c) $t = \frac{(115/200)-.65}{\sqrt{(.65)(.35)/200}} =$-2.2237479

$p =$0.0136449

(d) The value calculated in part(c) is a p-value which is the probability of rejecting a true null hypothesis. In the previous part, we would reject the dictator's claim at the 5

10. $CI_{95} = \left[.394 - 1.96\sqrt{(.394)(1-.394)/419}, .394 + 1.96\sqrt{(.394)(1-.394)/419}\right] =$ [0.3472121,0.4407879]

Based on his average up to the strike, there is not very strong evidence against $\theta = .400$, as this value is well within the 95% confidence interval.

11. $t = \frac{\bar{y}-0}{s/\sqrt{n}} = \frac{.132-0}{1.27/20} \approx$2.0787402

The difference is statistically greater than zero at the 5% level but not at the 1% level.

Chapter 1

Notes

Data Types

Nonexperimental data are not accumulated through controlled experiments on individuals, firms, or segments of the economy. Nonexperimental data are sometimes called **observational data**, or **retrospective data**, to emphasize the fact that the researcher is a passive collector of the data.

Experimental data are often collected in laboratory environments in the natural sciences, but they are more difficult to obtain in the social sciences. Although some social experiments can be devised, it is often impossible, prohibitively expensive, or morally repugnant to conduct the kinds of controlled experiments that would be needed to address economic issues.

An **empirical analysis** uses data to test a theory or to estimate a relationship.

A **cross-sectional data set** consists of a sample of individuals, households, firms, cities, states, countries, or a variety of other units, taken at a given point in time. An important feature of cross-sectional data is that we can often assume that they have been obtained by **random sampling** from the underlying population. Sometimes, however, the random sampling assumption is not appropriate for a variety of reasons such as respondents' willingness to answer or sampling from units that are large relative to the population. Nevertheless, random sampling is often assumed with cross-sectional data. The analysis of cross-sectional data is closely aligned with the applied microeconomics fields, such as labor economics, state and local public finance, industrial organization, urban economics, demography, and health economics

A **time series data set** consists of observations on a variable or several variables over time. Unlike the arrangement of cross-sectional data, the chronological ordering of observations in a time series conveys potentially important information. A key feature of time series data that makes them more difficult to analyze than cross-sectional data is that economic observations can rarely, if ever, be assumed to be independent across time.

A **pooled cross section** is a data configuration where independent cross sections, usually collected at different points in time, are combined to produce a single data set.

A **panel data** (or **longitudinal data**) **set** consists of a time series for each cross-sectional member in the data set. The key feature of panel data that distinguishes them from a pooled cross section is that the *same* cross-sectional units are followed over a given time period.

Causality

In econometrics, we're often concerned about finding a **causaul effect**, or A **ceteris paribus** (meaning all other relevant factors are held fixed) change in one variable that has an effect on another variable.

Exercises

Problems

1. (a) I would randomly assign (that is without other factors that affect student performance in mind) fourth grade students to varying class sizes and compare students' performances across the various groups.

   (b) I might expect a negative correlation between class size and test score because generally, larger classes have less funding per student and students in larger classes receive less individualized instruction. There are many additional factors positively correlated with student performance and negatively correlated with class size.

   (c) No, causality can only be established when ceteris-paribus is satisfied; however, this is not the case as some of the confounding factors that wouldn't be controlled for are listed in part(b).

2. (a) All else equal, do job training programs improve worker productivity?

   (b) No. For one, perhaps a firm that requires job training because it has less-skilled workers and wants to increase its production efficiency. There are many other factors like this on the firm side that make me believe that a firm's decision to train its workers will be independent of worker characteristics. Additionally, perhaps the firm does not require but offers job training. It's likely individuals who actually do the job training have different characteristics to those who do not. Some factors may include innate ability, intelligence, and motivation.

   (c) The quality of the equipment.

   (d) No, causality can only be established when ceteris-paribus is satisfied; however, this is not the case as some of the confounding factors that wouldn't be controlled for (such as the quality of the equipment).

3. No. Again, ceteris-paribus has not been established. Many other confounding factors correlated with "work" and "study" would not be controlled.

4. (a) Ideally, panel data that contains corporate tax rates and GSP.

   (b) Theoretically, it'd be possible to do a controlled experiment, but it would not be ethical. It would require randomly assigning individuals to varying levels of corporate tax rates and measuring the GSP within these groups.

   (c) If other factors impacting GSP growth and tax rates are sufficiently controlled for, then yes. Otherwise, such correlational analysis will likely be biased and not convincing.

Computer Exercises

C1)

```
#i
mean(wage1$educ)
```

```
## [1] 12.56274
```

```
min(wage1$educ)
```

```
## [1] 0
```

```
max(wage1$educ)
```

```
## [1] 18
```

```
#ii
mean(wage1$wage)
```

```
## [1] 5.896103
```

```
#It seems low for the year 2013

#iii
#Consumer Price Index (CPI) for the years 1976 and 2013.
cpi_1976 <- 55.6
cpi_2013 <- 230.280

#iv
mean(wage1$wage)*cpi_2013/cpi_1976
```

```
## [1] 24.42005
```

```r
#Yes, the average wage now seems more reasonable

#v
#women
sum(wage1$female)
```

```
## [1] 252
```

```r
#men
sum(wage1$female==0)
```

```
## [1] 274
```

```r
rm(cpi_1976,cpi_2008)
```

```
## Warning in rm(cpi_1976, cpi_2008): object 'cpi_2008' not found
```

```r
gc()
```

```
##           used (Mb) gc trigger  (Mb) max used  (Mb)
## Ncells 1100636 58.8    2289358 122.3  2289358 122.3
## Vcells 1917628 14.7    8388608  64.0  2651253  20.3
```

C2)

```r
#i
#number of women
sum(bwght$male==0)
```

```
## [1] 665
```

```r
#number of women who report smoking during pregnancy
sum(bwght$male==0 & bwght$cigs>0)
```

```
## [1] 112
```

```r
#ii and iii
mean(bwght$cigs)
```

```
## [1] 2.087176
```

```r
#No, this average includes males, a more descriptive average may be:
mean(bwght$cigs[bwght$male==0])
```

```
## [1] 2.090226
```

```r
#for all women and
mean(bwght$cigs[bwght$male==0 & bwght$cigs>0])
```

```
## [1] 12.41071
```

```r
#for those that reported smoking

#iv
mean(bwght$fatheduc, na.rm = T)
```

```
## [1] 13.18624
```

```r
#There are only 1192 observations because there are 196 missing values for fatheduc

#v
mean(bwght$faminc) * 1000
```

```
## [1] 29026.66
```

```
sd(bwght$faminc) * 1000
```

```
## [1] 18739.28
```

C3)

```
#i
min(meap01$math4)
```

```
## [1] 0
```

```
max(meap01$math4)
```

```
## [1] 100
```

```
#Without knowing much about the data, the range seems to make sense covering all 100% of pass rates

#ii
sum(meap01$math4 == 100)
```

```
## [1] 38
```

```
#iii
sum(meap01$math4 == 50)
```

```
## [1] 17
```

```
#iv
#math
mean(meap01$math4)
```

```
## [1] 71.909
```

```
#reading
mean(meap01$read4)
```

```
## [1] 60.06188
```

```
#The reading test seems harder to pass

#v
cor(meap01$math4,meap01$read4)
```

```
## [1] 0.8427281
```

```
#Those with higher pass rates on one exam tend to have higher pass rates on the other
#In other words, pass rates on the math and reading tests are highly correlated

#vi
mean(meap01$exppp)
```

```
## [1] 5194.865
```

```
sd(meap01$exppp)
```

```
## [1] 1091.89
```

```
#vii
#actual
500*100/5500
```

```
## [1] 9.090909
```

```
#ln approx
100 * (log(6000)-log(5500))
```

```
## [1] 8.701138
```

C4)

```
#i (reporting proportion instead of fraction)
mean(jtrain2$train)
```

```
## [1] 0.4157303
```

```
#ii
#receiving training
mean(jtrain2$re78[jtrain2$train==1])
```

```
## [1] 6.349145
```

```
#not receiving training
mean(jtrain2$re78[jtrain2$train==0])
```

```
## [1] 4.554802
```

```
#The difference does appear economically large

#iii
#receiving training
mean(jtrain2$unem78[jtrain2$train==1])
```

```
## [1] 0.2432432
```

```
#not receiving training
mean(jtrain2$unem78[jtrain2$train==0])
```

```
## [1] 0.3538462
```

```
#The proportion of unemployed who are not receiving training is about 11% larger

#Yes, the training seems effective, establishing ceteris paribus would make the results more convincing
```

C5)

```
#i
#min
min(fertil2$children)
```

```
## [1] 0
```

```
#max
max(fertil2$children)
```

```
## [1] 13
```

```
#mean
mean(fertil2$children)
```

```
## [1] 2.267828
```

```
#ii
sum(fertil2$children>0) / length(fertil2$children)
```

```
## [1] 0.7404265
```

```
#iii
#for those who have electricity
mean(fertil2$children[fertil2$electric==1 & !is.na(fertil2$electric)])
```

```
## [1] 1.898527
```

```
#for those who do not have electricity
mean(fertil2$children[fertil2$electric==0 & !is.na(fertil2$electric)])
```

```
## [1] 2.327729
```

```
#Those without electricity have more children on average (for this sample)

#iv
#No, ceteris paribus is not established
```

C6)

```
county_murders <- as_tibble(countymurders) %>%
  filter(year == 1996)

#i
n_distinct(county_murders$countyid)
```

```
## [1] 2197
```

```
#number with 0 murders
n_distinct(county_murders$countyid[county_murders$murders==0])
```

```
## [1] 1051
```

```
#percent with 0 murders
100 * n_distinct(county_murders$countyid[county_murders$murders==0]) / n_distinct(county_murders$county
```

```
## [1] 47.83796
```

```
#ii
#max number of murders
max(county_murders$murders)
```

```
## [1] 1403
```

```
#max number of executions
max(county_murders$execs)
```

```
## [1] 3
```

```
#mean number of executions
mean(county_murders$execs)
```

```
## [1] 0.01593081
```

```
#iii
cor(county_murders$murders,county_murders$execs)
```

```
## [1] 0.2095042
```

```
#iv
#No, I would suspect the positive correlation may be due to executions resulting from murders and other
rm(county_murders)
gc()
```

```
##           used (Mb) gc trigger  (Mb) max used  (Mb)
## Ncells 1127241 60.3    2289358 122.3  2289358 122.3
## Vcells 2666168 20.4    8388608  64.0  5275450  40.3
```

C7)

```
#i
#percent who report abusing alcohol
100 * mean(alcohol$abuse)
```

```
## [1] 9.916514
```

```
#employment rate
100 * mean(alcohol$employ)
```

```
## [1] 89.81877
```

```
#ii
100 * mean(alcohol$employ[alcohol$abuse==1])
```

```
## [1] 87.26899
```

```
#iii
100 * mean(alcohol$employ[alcohol$abuse==0])
```

```
## [1] 90.09946
```

```
#iv
#No, ceteris paribus is not established
```

C8)

```
#i
#assuming each obs/row corresponds to a unique student
NROW(econmath)
```

```
## [1] 856
```

```
#ii
#for those who took econ in high school
mean(econmath$score[econmath$econhs==1])
```

```
## [1] 72.07593
```

```
#for those who did not take econ in high school
mean(econmath$score[econmath$econhs==0])
```

```
## [1] 72.90792
```

```
#iii
#No, it simply tells us the average scores for those who did and did not take econ in hs
#It may signify some level of correlation but not causality

#iv
#Randomly assigning individuals to two groups (one who does take econ in high school
#and one that does not), then performing part(ii) can be used to obtain a good causal estimate
```