

Checkpoint 3 Solutions

Questions

1. State three properties of the expectation operator.

The chapter mentions the following properties:

$E[c] = c$: The expectation of a constant is the constant.

$E[aX] = aE[X]$: The expectation of a (variable multiplied by a scalar) is equivalent to the scalar multiplied by the expectation.

$E[aX + b] = E[aX] + E[b] = aE[X] + b$: Linearity of expectations + property 2

$E[a_1X_1 + \dots + a_nX_n] = \sum_i^n a_iE[X_i]$: The expectation of a sum is the sum of the expectations.

$E[W + H] = E[W] + E[H]$: Same as previous line.

$E[Z] = E[W - E[W]] = 0$: The expectation of a random variable created by subtracting the expectation of a variable from the variable is 0.

2. State the variance of a line and the variance of the sum of two random variables.

The variance of a line is $V[aX + b] = a^2V[X]$.

The variance of a sum of two random variables is $V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$. In the special case when X and Y are independent, the Covariance term drops out.

3. Define mean independence. Why is this useful for regression?

$E[u|x] = E[u], \forall i \in x$. The unconditional expectation of u is equivalent to the conditional expectation of u given x. Mean independence, along with a trivial renormalization, allows us to assume that $E[u|x] = 0, \forall i \in x$ which is a key identifying assumption in regression models because it says that the error term is uncorrelated with any variable that seeks to explain our outcome of interest. It implies that the population regression function is a linear function of x and allows us to consider the parameter of interest as a causal one.

4. Suppose we have two variables, **X** and **Y** and we run the regression of **Y** on **X**. Consider the model as $y_i = \beta_0 + \beta_1x_i + u_i$. Define in words what β_0 and β_1 are. What values do we need to identify them? You can answer with either mathematical definitions or in words.

β_0 is the intercept parameter. β_1 is the slope parameter. Both of these are population model concepts. In order to identify them, we need data on y and x such that we can create the sample analogs of the population parameters.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

5. Suppose we have heteroskedastic errors (which in practice is always true). Is OLS estimator for a parameter biased? Why or why not?

No. OLS is an unbiased estimator provided the following assumptions:

1. The population model is linear in parameters
2. We have random sampling from a population

3. The variance is finite and non-zero.
4. The zero conditional mean assumption holds.

None of these assumptions are related to the variance of the error term, so having heteroskedastic errors does not affect the unbiasedness of the OLS estimator. Heteroskedasticity does mean that OLS is no longer the Best Linear Unbiased Estimator (BLUE).