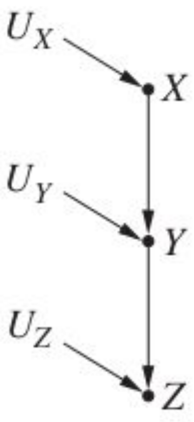
Seminario 8/9

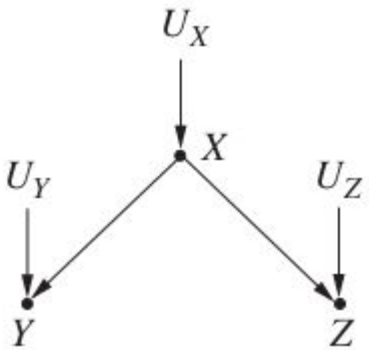
Repaso

Modelos vistos ….

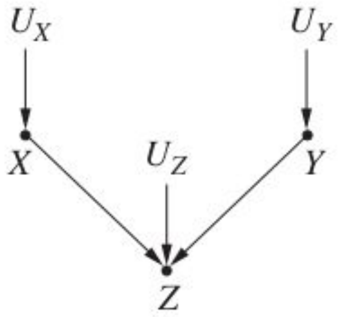
Chain



Fork



Collider

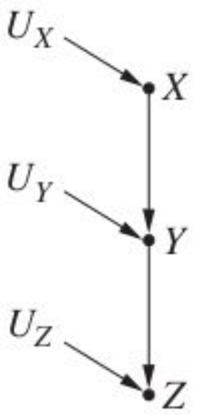


***Nota**: Ux,Uy,Uz son variables exógenas que representan efectos desconocidos o aleatorios que pueden perturbar la relación entre X,Y,Z*

Repaso

Propiedades de los modelos vistos

Chain

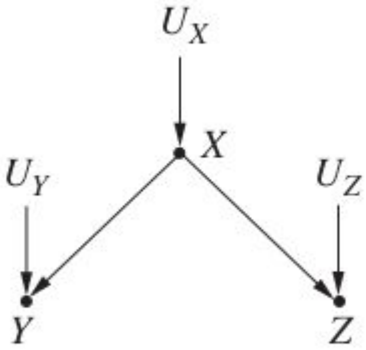


1. ***Z and Y are dependent***
   For some z, y, $P(Z = z|Y = y) \neq P(Z = z)$
2. ***Y and X are dependent***
   For some y, x, $P(Y = y|X = x) \neq P(Y = y)$
3. ***Z and X are likely dependent***
   For some z, x, $P(Z = z|X = x) \neq P(Z = z)$
4. ***Z and X are independent, conditional on Y***
   For all x, y, z, $P(Z = z|X = x, Y = y) = P(Z = z|Y = y)$

Para todos los modelos vamos a asumir que U_x, U_y, U_z son independientes

Repaso

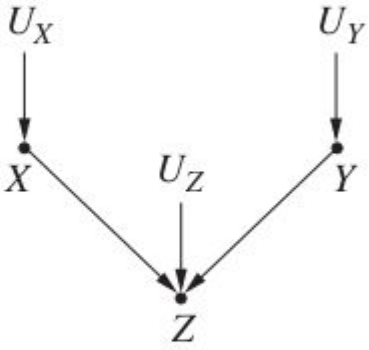
Fork



1. ***X and Y are dependent.***
   For some x, y, $P(X = x|Y = y) \neq P(X = x)$
2. ***X and Z are dependent.***
   For some x, z, $P(X = x|Z = z) \neq P(X = x)$
3. ***Z and Y are likely dependent.***
   For some z, y, $P(Z = z|Y = y) \neq P(Z = z)$
4. ***Y and Z are independent, conditional on X.***
   For all $4x, y, z$, $P(Y = y|Z = z, X = x) = P(Y = y|X = x)$

Repaso

Collider



1. ***X and Z are dependent.***
   For some x, z, $P(X = x|Z = z) \neq P(X = x)$
2. ***Y and Z are dependent.***
   For some y, z, $P(Y = y|Z = z) \neq P(Y = y)$
3. ***X and Y are independent.***
   For all x, y, $P(X = x|Y = y) = P(X = x)$
4. ***X and Y are dependent conditional on Z.***
   For some x, y, z, $P(X = x|Y = y, Z = z) \neq P(X = x|Z = z)$

Repaso

Resumen …

Rule 1 (Conditional Independence in Chains) *Two variables, X and Y, are conditionally independent given Z, if there is only one unidirectional path between X and Y and Z is any set of variables that intercepts that path.*

Rule 2 (Conditional Independence in Forks) *If a variable X is a common cause of variables Y and Z, and there is only one path between Y and Z, then Y and Z are independent conditional on X.*

Rule 3 (Conditional Independence in Colliders) *If a variable Z is the collision node between two variables X and Y, and there is only one path between X and Y, then X and Y are unconditionally independent but are dependent conditional on Z and any descendants of Z.*

Resumen

d-separation

d-separated: Dos nodos están d-separados cuando las variables que representan son independientes.

d-connected: Dos nodos están d-conectados cuando las variables que representan son dependientes o probablemente dependientes.

Dos nodos **X** y **Y** están d-separados si cada camino entre ellos(Si es que existe) están bloqueados.

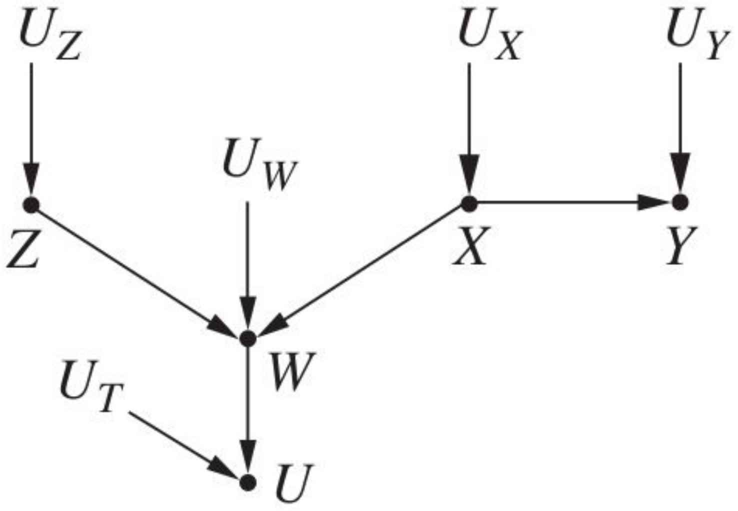
Resumen

Definition 2.4.1 **(d-separation)** *A path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.*

Resumen

Ejemplo



El camino entre Z y Y está bloqueado, por lo que Z,Y están d-separados. Si condicionamos W pasa a no estar bloqueado el camino por lo que pasan a estar d-conectados, luego si condicionamos {W,X} se vuelve a bloquear el camino por lo que pasan a estar conectados

Model Testing and Causal Search

Si tenemos un grafo G que creemos que podría haber generado un data set S, la d-separación nos diría que variables de G tienen que ser condicionalmente independientes dada otras variables. La independencia condicional es algo que podemos testear para un data set S. Si tenemos un grafo G, y notamos que A y B están d-separadas condicionando C, y luego estimamos las probabilidades usando al dataset S, llegando a que A y B **no son** independientes condicionado C, podríamos rechazar a G como modelo.

Model Testing and Causal Search

Definicion: Dos grafos G1 y G2 están en la misma clase de equivalencia si comparten el mismo esqueleto, es decir los mismos vértices(Sin importar la dirección) y si comparten las mismas v-estructuras, es decir colliders cuyos padres no son adyacentes.

Prop: Si dos grafos pertenecen a la misma clase de equivalencia entonces tienen las mismas condiciones de d-separación.

Si testeamos un modelo con un grafo G y como resultado no podemos rechazarlo, cualquier modelo con un grafo G2 asociado y con G2 perteneciente a la misma clase de equivalencia que G, tampoco lo vamos a poder rechazar.

Model Testing and Causal Search

La importancia de este resultado es que nos permite buscar a partir del dataset el modelo causal que lo generó. Por lo que no solo podemos partir de un modelo y generar el data set si no que a partir de un data set podemos rastrear un posible modelo que lo haya generado.

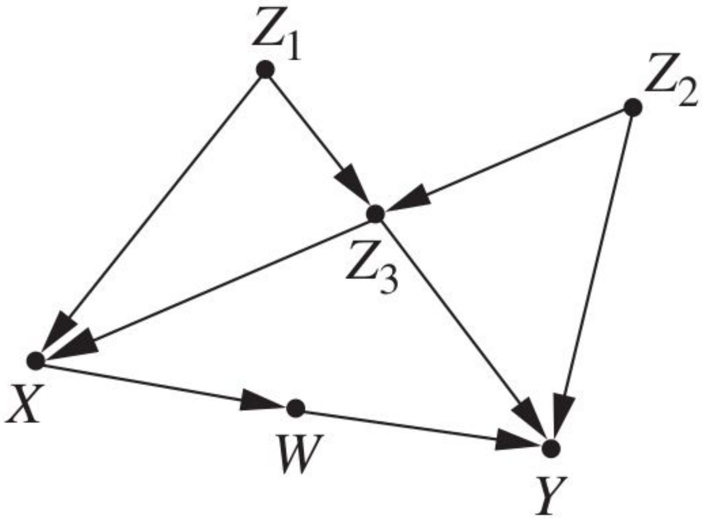
Model Testing and Causal Search

Ejercicio:

1. For each pair of nonadjacent nodes in this graph, find a set of variables that d-separates that pair.
2. Repeat question (1) assuming that only variables in the set $\{Z_3, W, X, Z_1\}$ can be measured.
3. Which of the arrows in Figure 2.9 can be reversed without being detected by any statistical test?
4. List all graphs that are observationally equivalent to the one in Figure 2.9.

Model Testing and Causal Search

Ejercicio:



Interventions

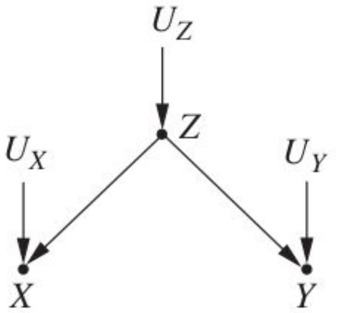
Intervening vs Conditioning

Cuando intervenimos en la variable en un modelo, fijamos el valor. Cambiamos el sistema y los valores de las otras variables cambian como resultado. Cuando condicionamos en una variable, no cambiamos nada; solo hacemos foco a un subconjunto de casos en el cual la variable toma el valor en el que estamos interesados. Lo que cambia es nuestra percepción del mundo pero no el mundo en sí. En el primer caso la variable no es más aleatoria, en el segundo sí lo es.

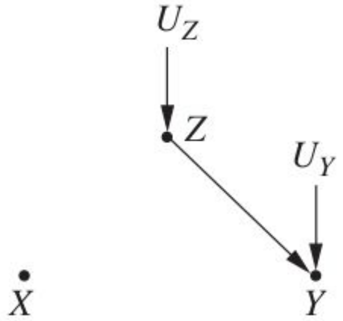
Interventions

Ejemplo: X representa la venta de helado, Y la cantidad de crímenes, Z la temperatura. Cuando intervenimos para fijar el valor de una variable, restringimos la tendencia de la variable de variar en función de otras variables. Por ejemplo si intervenimos en la venta de helado (Por ejemplo cerrando todas las heladerías) nos quedaría el modelo que teníamos antes pero sin la arista que va de Z a X. Cuando examinamos esto nos queda que la cantidad de crímenes es totalmente independiente a la venta de helados, ya que la venta de helados no está más asociada con la temperatura. En otras palabras, aunque cambiamos el valor en el que X es constante, esta variación no se va a ver reflejada en Y. Notar cómo intervenir en X, no da lo mismo que condicionar X, si vemos los valores en que X es alto probablemente Y también sea alto (Punto 3 de diapositivas de Fork).

Antes de la intervención



Después de la intervención



Interventions

Notación: Cuando fijamos el valor de una variable X a un valor x, lo anotamos con do(X = x). Por lo que $P(Y=y|X=x)$ es la probabilidad de que Y = y condicionado a que X = x, mientras que $P(Y = y|do(X=x))$ es la probabilidad de que Y=y cuando intervenimos para que X = x.

The Adjustment Formula

Volvamos al ejemplo de la paradoja de Simpson …

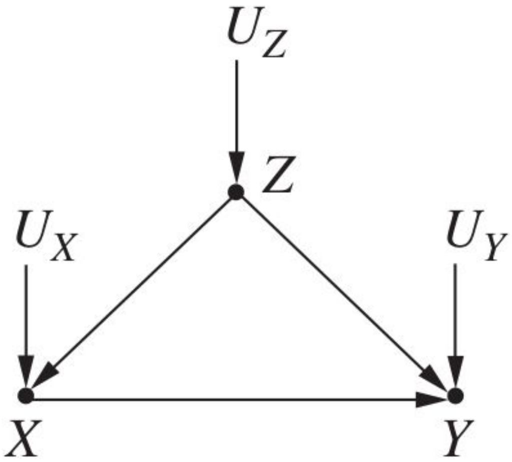
Queremos probar el efecto de una nueva droga, tenemos la variable Z que representa el género, X que representa si uso la droga y Y si se recuperó o no. Para descubrir qué tan efectiva es la droga, supongamos una intervención en el que le damos a toda la población la droga do(X=1) y comparamos los recuperados contra la otra intervención de que no le hubiésemos dado a nadie la droga do(X=0). Queremos calcular el *Avarage Causal Effect* (ACE)

$$\text{ACE} = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0))$$

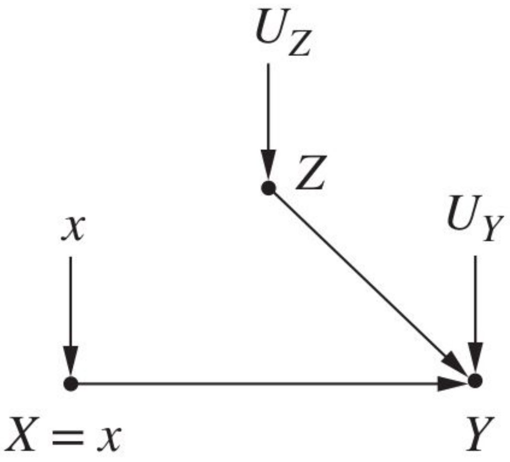
Recordar 1: En este caso, si nos fijabamos por género convenía tomar la droga y si nos fijabamos el total no.
Recordar 2: Efecto causal, es cómo afecta a que suceda B dado que sucedio A.

The Adjustment Formula

Modelo antes de la intervención



Modelo después de la intervención



The Adjustment Formula

Para diferenciar, las probabilidades en el nuevo modelo quedan escritas como P_m

Tenemos que:

1. $P_m(Z=z) = P(Z=z)$, esto se debe a que lo que determina que Z valga z no cambia si fijamos el valor de X. La proporción de hombres y mujeres no va a cambiar si intervenimos dándole la droga o no dándosela.
2. $P_m(Y=y|Z=z, X=x) = P(Y=y|Z=z, X=x)$ porque el proceso en el cual Y responde a X y Z, $Y=f(x, z, u_Y)$ no cambia si X toma su valor de forma natural o si lo fijamos.

The Adjustment Formula

Utilizando los dos resultados anteriores tenemos que:

$$
\begin{aligned}
P(Y = y|do(X = x)) & \\
&= P_m(Y = y|X = x) \qquad \text{(by definition)} \\
&= \sum_z P_m(Y = y|X = x, Z = z)P_m(Z = z|X = x) \\
&= \sum_z P_m(Y = y|X = x, Z = z)P_m(Z = z) \\
&= \sum_z P(Y = y|X = x, Z = z)P(Z = z)
\end{aligned}
$$

A este resultado se le llama fórmula de ajustamiento, calcula la asociación de cada valor de Z y luego promedia en estos valores. Este procedimiento se conoce como “ajustando para z” o “controlando para z”

The Adjustment Formula

Poniendo números al ejemplo:

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

$$P(Y = 1|do(X = 1)) = \frac{0.93(87 + 270)}{700} + \frac{0.73(263 + 80)}{700} = 0.832$$

$$P(Y = 1|do(X = 0)) = \frac{0.87(87 + 270)}{700} + \frac{0.69(263 + 80)}{700} = 0.7818$$

The Adjustment Formula

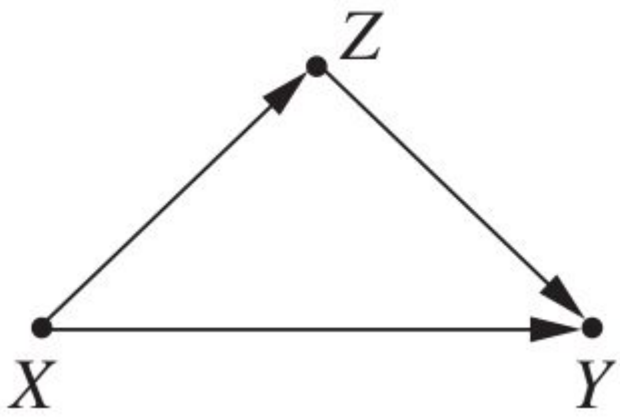
Comparando el efecto de los que tomaron la droga(X=1) con los que no la tomaron(X=0), obtenemos:

$$ACE = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0)) = 0.832 - 0.7818 = 0.0502$$

Con este resultado podemos decir que es mejor tomar la droga

The Adjustment Formula

Que pasaría en el ejemplo de la paradoja de Simpson, en que Z es la presión arterial



Cambia la dirección del vértice X,Z ya que tomar la droga influye en la presión arterial. Ahora intervenir en X ya no cambia el modelo. $P(Y = y|do(X = x)) = P(Y = y|X = x)$

Recordar: En este caso, si nos fijabamos por nivel de presión arterial no convenía tomar la droga y si nos fijabamos el total sí convenía

The Adjustment Formula

Rule 1 (The Causal Effect Rule) *Given a graph G in which a set of variables PA are designated as the parents of X, the causal effect of X on Y is given by*

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, PA = z)P(PA = z) \tag{3.6}$$

La ecuación 3.6 se puede reescribir como

$$P(y|do(x)) = \sum_z \frac{P(X = x, Y = y, PA = z)}{P(X = x|PA = z)}$$

El factor $P(X = x|PA = z)$ se le llama ***propensity score***

The Backdoor Criterion

Bajo qué condiciones, la estructura del grafo causal es suficiente para calcular el efecto causal de una variable en otra, dado un dataset?

The Backdoor Criterion

Definición 3.3.1 (The Backdoor Criterion) *Dado un par ordenado de variables (X,Y) en un DAG G, un conjunto de variables Z satisface el criterio de la puerta trasera en relación a (X,Y) si ningún nodo en Z es un descendiente de X, y Z bloquea cada camino entre X y Y que contenga un vértice que entra a X.*

Si un conjunto de variables Z satisface el criterio de la puerta trasera para X y Y, entonces el efecto causal de X en Y está dado por la fórmula

$$P(Y = y|do(X = x)) = \sum_{z} P(Y = y|X = x, Z = z)P(Z = z)$$

The Backdoor Criterion

- Cuando queremos averiguar el efecto causal de X en Y, queremos que los nodos en los que condicionamos bloqueen todos los caminos (“Puertas traseras”) entre X,Y en los que una de las puntas del camino sea un vértice entrando a X, porque dichos caminos hacen que X e Y sean dependientes, pero no transmiten la influencia causal que tiene X en Y. Por lo que al condicionar en este padre de X, eliminamos esta dependencia (Bloqueamos el camino).
- Los descendientes de X se verían afectados por una intervención sobre X y podrían afectar a su vez a Y; el condicionamiento sobre ellos bloquearía estos caminos. No queremos que pase esto.
- Al no condicionar los descendientes de X, tampoco condicionamos en algún posible collider de X y Y que nos generaría dependencia entre estos nodos(Desbloquearía el camino)

The Backdoor Criterion

La idea detrás del criterio de puerta trasera es que, las condiciones llevan a que:

1. Bloqueamos todos los caminos espurios entre X y Y
2. Dejamos todos los caminos dirigidos de X a Y sin alterar
3. No creamos nuevos caminos espurios

The Backdoor Criterion

Ejemplo:

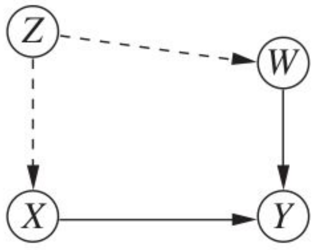


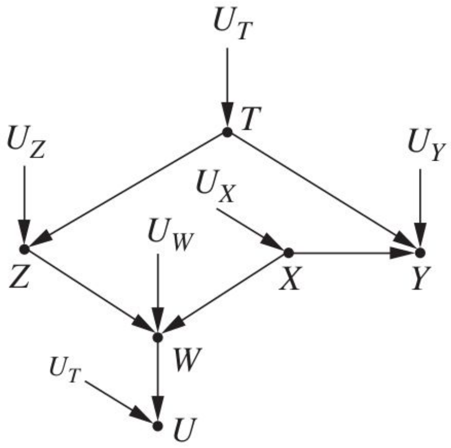
Figure 3.6 A graphical model representing the relationship between a new drug (X), recovery (Y), weight (W), and an unmeasured variable Z (socioeconomic status)

Sabemos que Z afecta a X y a W pero no la medimos. Aunque no hayamos observado Z, vemos que W que no es un descendiente de X bloquea el camino X <- Z -> W -> Y. Por lo tanto, W cumple el criterio de puerta trasera. Por lo que ajustar para W nos daría el efecto causal de X en Y.

$$P(Y = y|do(X = x)) = \sum_{w} P(Y = y|X = x, W = w)P(W = w)$$

The Backdoor Criterion

Otro ejemplo:



Queremos calcular el efecto de X en Y. ¿Qué variables tenemos que condicionar para calcular bien el efecto? La pregunta se reduce a encontrar el conjunto de variables que satisfacen el criterio de la puerta trasera, pero como no hay ningún camino del tipo “puerta trasera” entre X y Y, no hay ninguna variable que cumpla el criterio, por lo que ningún ajuste es necesario.

$$P(y|do(x)) = P(y|x)$$

The Backdoor Criterion

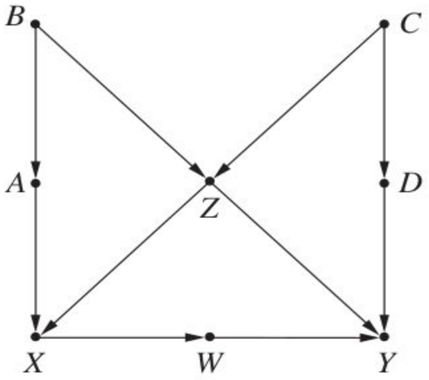


Figure 3.8 Causal graph used to illustrate the backdoor criterion in the following study questions

(a) *List all of the sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y.*
(b) *List all of the minimal sets of variables that satisfy the backdoor criterion to determine the causal effect of X on Y (i.e., any set of variables such that, if you removed any one of the variables from the set, it would no longer meet the criterion).*
(c) *List all minimal sets of variables that need be measured in order to identify the effect of D on Y. Repeat, for the effect of $\{W, D\}$ on Y.*

The Backdoor Criterion

Part (a)

Let us use the abbreviation "backdoor admissible" to denote a set of variables that satisfy the backdoor criterion of Definition 3.3.1; for the present model, a backdoor admissible set $\boldsymbol{Z}$ blocks all spurious paths between X and Y while leaving all directed paths from X to Y open. We can easily verify that the following sets satisfy the backdoor criterion to determine the causal effect of X on Y.

1. *Sets of 2 nodes:* $\{Z, A\}, \{Z, B\}, \{Z, C\}, \{Z, D\}$
2. *Sets of 3 nodes:* $\{Z, A, B\}, \{Z, A, C\}, \{Z, A, D\}, \{Z, B, C\}, \{Z, B, D\}, \{Z, C, D\}$
3. *Sets of 4 nodes:* $\{Z, A, B, C\}, \{Z, A, B, D\}, \{Z, A, C, D\}, \{Z, B, C, D\}$
4. *Sets of 5 nodes:* $\{Z, A, B, C, D\}$

Part (b)

According to (a), the following 4 sets are minimal, since in every other set, a node could be removed and still ensure that the backdoor criterion is satisfied:
$\{Z, A\}, \{Z, B\}, \{Z, C\}, \{Z, D\}$.

The Backdoor Criterion

Part (c)

For identifying the effect of D on Y:
We want to find a set $\mathbf{Z}$ that blocks all backdoor paths from D to Y. Notice that the set $\{C\}$ is one solution, so any other set that contains C is not minimal. Also, if the set does not contain C, then it must contain Z, otherwise, we have an open path $Y \leftarrow Z \leftarrow C \rightarrow D$. However, by including Z, the backdoor path $Y \leftarrow W \leftarrow X \leftarrow A \leftarrow B \rightarrow Z \leftarrow C \rightarrow D$ is open. To block this path, we add any of A, B, X, or W. So, there are a total of 5 minimal sets:
$\{C\}, \{Z, A\}, \{Z, B\}, \{Z, X\}, \{Z, W\}$.

For identifying the effect of $\{W, D\}$ on Y:
Again, we want to verify that the only open paths from $\{W, D\}$ to Y are the direct edges. So similar to above, if the set $\mathbf{Z}$ contains Z, then all backdoor paths from D to Y or from W to Y will be blocked. Also if the set $\mathbf{Z}$ does not contain Z, then it must contain C and X, otherwise, we have an open path through $D \leftarrow C \rightarrow Z \rightarrow Y$ or $W \leftarrow X \leftarrow Z \rightarrow Y$. So there are a total of 2 minimal sets as follows:
$\{Z\}, \{C, X\}$.

Nota: *http://dagitty.net/m331*