Somers' D: A common currency for associations

Roger B. Newson
r.newson@imperial.ac.uk
http://www.imperial.ac.uk/nhli/r.newson/

Department of Primary Care and Public Health, Imperial College London

21st UK Stata Users' Group Meeting, 10–11 September, 2015
Downloadable from the conference website at
http://ideas.repec.org/s/boc/usug15.html

What is Somers' D?

- We assume that pairs of (X, Y)–pairs (X_i, Y_i) and (X_j, Y_j) are sampled from a population of (X, Y)–pairs, under a specified sampling scheme.
- **Kendall's** τ_a is defined (symmetrically) as the expectation

$$\tau_{XY} = \mathrm{E}[\mathrm{sign}(X_i - X_j)\mathrm{sign}(Y_i - Y_j)],$$

  or as the difference between the probabilities of **concordance** and **discordance** between the two (X, Y)–pairs.
- **Somers'** D is defined (asymmetrically) as the ratio

$$D(Y|X) = \tau_{XY}/\tau_{XX},$$

  or as the difference between the two corresponding *conditional* probabilities, given that one X–value is known to be larger than the other X–value.
- These definitions can be extended to cases where the X–values and/or the Y–values may be left–censored and/or right–censored.

What is Somers' D?

- We assume that pairs of (X, Y)–pairs (X_i, Y_i) and (X_j, Y_j) are sampled from a population of (X, Y)–pairs, under a specified sampling scheme.
- **Kendall's** τ_a is defined (symmetrically) as the expectation

$$\tau_{XY} = \mathrm{E}[\mathrm{sign}(X_i - X_j)\mathrm{sign}(Y_i - Y_j)],$$

  or as the difference between the probabilities of **concordance** and **discordance** between the two (X, Y)–pairs.
- **Somers'** D is defined (asymmetrically) as the ratio

$$D(Y|X) = \tau_{XY}/\tau_{XX},$$

  or as the difference between the two corresponding *conditional* probabilities, given that one X–value is known to be larger than the other X–value.
- These definitions can be extended to cases where the X–values and/or the Y–values may be left–censored and/or right–censored.

What is Somers' D?

- We assume that pairs of (X,Y)–pairs (X_i,Y_i) and (X_j,Y_j) are sampled from a population of (X,Y)–pairs, under a specified sampling scheme.
- **Kendall's** τ_a is defined (symmetrically) as the expectation

$$\tau_{XY} = \mathrm{E}[\mathrm{sign}(X_i - X_j)\mathrm{sign}(Y_i - Y_j)],$$

  or as the difference between the probabilities of **concordance** and **discordance** between the two (X,Y)–pairs.
- **Somers'** D is defined (asymmetrically) as the ratio

$$D(Y|X) = \tau_{XY}/\tau_{XX},$$

  or as the difference between the two corresponding *conditional* probabilities, given that one X–value is known to be larger than the other X–value.
- These definitions can be extended to cases where the X–values and/or the Y–values may be left–censored and/or right–censored.

What is Somers' D?

- We assume that pairs of (X, Y)–pairs (X_i, Y_i) and (X_j, Y_j) are sampled from a population of (X, Y)–pairs, under a specified sampling scheme.
- **Kendall's** τ_a is defined (symmetrically) as the expectation

$$\tau_{XY} = \mathrm{E}[\mathrm{sign}(X_i - X_j)\mathrm{sign}(Y_i - Y_j)],$$

  or as the difference between the probabilities of **concordance** and **discordance** between the two (X, Y)–pairs.
- **Somers'** D is defined (asymmetrically) as the ratio

$$D(Y|X) = \tau_{XY}/\tau_{XX},$$

  or as the difference between the two corresponding *conditional* probabilities, given that one X–value is known to be larger than the other X–value.
- These definitions can be extended to cases where the X–values and/or the Y–values may be left–censored and/or right–censored.

What is Somers' D?

- We assume that pairs of (X, Y)–pairs (X_i, Y_i) and (X_j, Y_j) are sampled from a population of (X, Y)–pairs, under a specified sampling scheme.
- **Kendall's** τ_a is defined (symmetrically) as the expectation

$$\tau_{XY} = \mathrm{E}[\mathrm{sign}(X_i - X_j)\mathrm{sign}(Y_i - Y_j)],$$

  or as the difference between the probabilities of **concordance** and **discordance** between the two (X, Y)–pairs.
- **Somers'** D is defined (asymmetrically) as the ratio

$$D(Y|X) = \tau_{XY}/\tau_{XX},$$

  or as the difference between the two corresponding *conditional* probabilities, given that one X–value is known to be larger than the other X–value.
- These definitions can be extended to cases where the X–values and/or the Y–values may be left–censored and/or right–censored.

Somers' *D* and the `somersd` package

- The package `somersd`[4] can be downloaded from SSC, and estimates Somers' *D* or Kendall's τ_a.
- These **rank parameters** can be interval–estimated under a wide range of sampling schemes, with or without censorship.
- The `cluster()` option allows for sampling clusters from a population of clusters.
- Sampling–probability weights are enabled by `pweights`, allowing us to estimate target–population parameters from a sampled population by direct standardization.
- The `wstrata()` and `bstrata()` options allow for restriction to comparisons within or between strata.
- And the parameter estimates are saved as Stata estimation results, using delta–jackknife variances and a choice of Normalizing and/or variance–stabilizing transformations.

Somers' D and the `somersd` package

- The package `somersd`[4] can be downloaded from SSC, and estimates Somers' D or Kendall's τ_a.
- These **rank parameters** can be interval–estimated under a wide range of sampling schemes, with or without censorship.
- The `cluster()` option allows for sampling clusters from a population of clusters.
- Sampling–probability weights are enabled by `pweights`, allowing us to estimate target–population parameters from a sampled population by direct standardization.
- The `wstrata()` and `bstrata()` options allow for restriction to comparisons within or between strata.
- And the parameter estimates are saved as Stata estimation results, using delta–jackknife variances and a choice of Normalizing and/or variance–stabilizing transformations.

Somers' *D* and the `somersd` package

- The package `somersd`[4] can be downloaded from SSC, and estimates Somers' *D* or Kendall's τ_a.
- These **rank parameters** can be interval–estimated under a wide range of sampling schemes, with or without censorship.
- The `cluster()` option allows for sampling clusters from a population of clusters.
- Sampling–probability weights are enabled by `pweights`, allowing us to estimate target–population parameters from a sampled population by direct standardization.
- The `wstrata()` and `bstrata()` options allow for restriction to comparisons within or between strata.
- And the parameter estimates are saved as Stata estimation results, using delta–jackknife variances and a choice of Normalizing and/or variance–stabilizing transformations.

Somers' *D* and the `somersd` package

- The package `somersd`[4] can be downloaded from SSC, and estimates Somers' *D* or Kendall's τ_a.
- These **rank parameters** can be interval–estimated under a wide range of sampling schemes, with or without censorship.
- The `cluster()` option allows for sampling clusters from a population of clusters.
- Sampling–probability weights are enabled by `pweights`, allowing us to estimate target–population parameters from a sampled population by direct standardization.
- The `wstrata()` and `bstrata()` options allow for restriction to comparisons within or between strata.
- And the parameter estimates are saved as Stata estimation results, using delta–jackknife variances and a choice of Normalizing and/or variance–stabilizing transformations.

Somers' *D* and the `somersd` package

- The package `somersd`[4] can be downloaded from SSC, and estimates Somers' *D* or Kendall's τ_a.
- These **rank parameters** can be interval–estimated under a wide range of sampling schemes, with or without censorship.
- The `cluster()` option allows for sampling clusters from a population of clusters.
- Sampling–probability weights are enabled by `pweights`, allowing us to estimate target–population parameters from a sampled population by direct standardization.
- The `wstrata()` and `bstrata()` options allow for restriction to comparisons within or between strata.
- And the parameter estimates are saved as Stata estimation results, using delta–jackknife variances and a choice of Normalizing and/or variance–stabilizing transformations.

Somers' D and the `somersd` package

- The package `somersd`[4] can be downloaded from SSC, and estimates Somers' D or Kendall's τ_a.
- These **rank parameters** can be interval–estimated under a wide range of sampling schemes, with or without censorship.
- The `cluster()` option allows for sampling clusters from a population of clusters.
- Sampling–probability weights are enabled by `pweights`, allowing us to estimate target–population parameters from a sampled population by direct standardization.
- The `wstrata()` and `bstrata()` options allow for restriction to comparisons within or between strata.
- And the parameter estimates are saved as Stata estimation results, using delta–jackknife variances and a choice of Normalizing and/or variance–stabilizing transformations.

Somers' *D* and the `somersd` package

- The package `somersd`[4] can be downloaded from SSC, and estimates Somers' *D* or Kendall's τ_a.
- These **rank parameters** can be interval–estimated under a wide range of sampling schemes, with or without censorship.
- The `cluster()` option allows for sampling clusters from a population of clusters.
- Sampling–probability weights are enabled by `pweights`, allowing us to estimate target–population parameters from a sampled population by direct standardization.
- The `wstrata()` and `bstrata()` options allow for restriction to comparisons within or between strata.
- And the parameter estimates are saved as Stata estimation results, using delta–jackknife variances and a choice of Normalizing and/or variance–stabilizing transformations.

A useful feature of Somers' D

- Somers' D has the useful feature that a larger Somers' D cannot be secondary to a smaller Somers' D in the same direction.
- That is to say, if a positive Somers' D between X and Y is secondary to a positive Somers' D between X and W, then, from the argument of Newson (1987)[5], we must have the inequality

$$D(Y|X) \leq D(W|X).$$

- *So*, if a confidence interval for $D(Y|X) - D(W|X)$ contains only positive values, then a positive association of X with Y *cannot* be caused by a positive association of X with W.
- Such a confidence interval can be produced using `lincom` or `nlcom` after `somersd`.
- This is especially useful if Y is an outcome, X is an exposure, and W is a positive predictor of X.

A useful feature of Somers' D

- Somers' D has the useful feature that a larger Somers' D cannot be secondary to a smaller Somers' D in the same direction.
- That is to say, if a positive Somers' D between X and Y is secondary to a positive Somers' D between X and W, then, from the argument of Newson (1987)[5], we must have the inequality

$$D(Y|X) \leq D(W|X).$$

- *So*, if a confidence interval for $D(Y|X) - D(W|X)$ contains only positive values, then a positive association of X with Y *cannot* be caused by a positive association of X with W.
- Such a confidence interval can be produced using `lincom` or `nlcom` after `somersd`.
- This is especially useful if Y is an outcome, X is an exposure, and W is a positive predictor of X.

A useful feature of Somers' D

- Somers' D has the useful feature that a larger Somers' D cannot be secondary to a smaller Somers' D in the same direction.
- That is to say, if a positive Somers' D between X and Y is secondary to a positive Somers' D between X and W, then, from the argument of Newson (1987)[5], we must have the inequality

$$D(Y|X) \leq D(W|X).$$

- *So*, if a confidence interval for $D(Y|X) - D(W|X)$ contains only positive values, then a positive association of X with Y *cannot* be caused by a positive association of X with W.
- Such a confidence interval can be produced using `lincom` or `nlcom` after `somersd`.
- This is especially useful if Y is an outcome, X is an exposure, and W is a positive predictor of X.

A useful feature of Somers' D

- Somers' D has the useful feature that a larger Somers' D cannot be secondary to a smaller Somers' D in the same direction.
- That is to say, if a positive Somers' D between X and Y is secondary to a positive Somers' D between X and W, then, from the argument of Newson (1987)[5], we must have the inequality

$$D(Y|X) \leq D(W|X).$$

- *So*, if a confidence interval for $D(Y|X) - D(W|X)$ contains only positive values, then a positive association of X with Y *cannot* be caused by a positive association of X with W.
- Such a confidence interval can be produced using `lincom` or `nlcom` after `somersd`.
- This is especially useful if Y is an outcome, X is an exposure, and W is a positive predictor of X.

A useful feature of Somers' D

- Somers' D has the useful feature that a larger Somers' D cannot be secondary to a smaller Somers' D in the same direction.
- That is to say, if a positive Somers' D between X and Y is secondary to a positive Somers' D between X and W, then, from the argument of Newson (1987)[5], we must have the inequality

$$D(Y|X) \leq D(W|X).$$

- *So*, if a confidence interval for $D(Y|X) - D(W|X)$ contains only positive values, then a positive association of X with Y *cannot* be caused by a positive association of X with W.
- Such a confidence interval can be produced using `lincom` or `nlcom` after `somersd`.
- This is especially useful if Y is an outcome, X is an exposure, and W is a positive predictor of X.

A useful feature of Somers' D

- Somers' D has the useful feature that a larger Somers' D cannot be secondary to a smaller Somers' D in the same direction.
- That is to say, if a positive Somers' D between X and Y is secondary to a positive Somers' D between X and W, then, from the argument of Newson (1987)[5], we must have the inequality

$$D(Y|X) \leq D(W|X).$$

- *So*, if a confidence interval for $D(Y|X) - D(W|X)$ contains only positive values, then a positive association of X with Y *cannot* be caused by a positive association of X with W.
- Such a confidence interval can be produced using `lincom` or `nlcom` after `somersd`.
- This is especially useful if Y is an outcome, X is an exposure, and W is a positive predictor of X.

Example: Propensity scores

- In particular, W may be a **propensity score**[1], predicting X from a list of multiple confounders $V_1, \ldots, V_K$.
- In the 21st–century Rubin method[6], the propensity score is a predictor from a regression model in the exposure, which we find in the joint distribution of the exposure and the confounders.
- We then add in the outcome data, and use the propensity score to estimate a propensity–adjusted exposure effect on the outcome.
- This adjusted exposure effect is typically interpreted as an exposure effect in a fantasy **target population**, with real–world marginal exposure and propensity distributions, but no exposure–propensity association.
- This is usually done using propensity weighting, propensity matching, or propensity stratification, in regression models of Y with respect to X. *However*…

Example: Propensity scores

- In particular, W may be a **propensity score**[1], predicting X from a list of multiple confounders $V_1, \ldots, V_K$.
- In the 21st–century Rubin method[6], the propensity score is a predictor from a regression model in the exposure, which we find in the joint distribution of the exposure and the confounders.
- We then add in the outcome data, and use the propensity score to estimate a propensity–adjusted exposure effect on the outcome.
- This adjusted exposure effect is typically interpreted as an exposure effect in a fantasy **target population**, with real–world marginal exposure and propensity distributions, but no exposure–propensity association.
- This is usually done using propensity weighting, propensity matching, or propensity stratification, in regression models of Y with respect to X. *However…*

Example: Propensity scores

- In particular, W may be a **propensity score**[1], predicting X from a list of multiple confounders $V_1, \ldots, V_K$.
- In the 21st–century Rubin method[6], the propensity score is a predictor from a regression model in the exposure, which we find in the joint distribution of the exposure and the confounders.
- We then add in the outcome data, and use the propensity score to estimate a propensity–adjusted exposure effect on the outcome.
- This adjusted exposure effect is typically interpreted as an exposure effect in a fantasy **target population**, with real–world marginal exposure and propensity distributions, but no exposure–propensity association.
- This is usually done using propensity weighting, propensity matching, or propensity stratification, in regression models of Y with respect to X. *However*…

Example: Propensity scores

- In particular, W may be a **propensity score**[1], predicting X from a list of multiple confounders $V_1, \ldots, V_K$.
- In the 21st–century Rubin method[6], the propensity score is a predictor from a regression model in the exposure, which we find in the joint distribution of the exposure and the confounders.
- We then add in the outcome data, and use the propensity score to estimate a propensity–adjusted exposure effect on the outcome.
- This adjusted exposure effect is typically interpreted as an exposure effect in a fantasy **target population**, with real–world marginal exposure and propensity distributions, but no exposure–propensity association.
- This is usually done using propensity weighting, propensity matching, or propensity stratification, in regression models of Y with respect to X. *However…*

Example: Propensity scores

- In particular, W may be a **propensity score**[1], predicting X from a list of multiple confounders $V_1, \ldots, V_K$.
- In the 21st–century Rubin method[6], the propensity score is a predictor from a regression model in the exposure, which we find in the joint distribution of the exposure and the confounders.
- We then add in the outcome data, and use the propensity score to estimate a propensity–adjusted exposure effect on the outcome.
- This adjusted exposure effect is typically interpreted as an exposure effect in a fantasy **target population**, with real–world marginal exposure and propensity distributions, but no exposure–propensity association.
- This is usually done using propensity weighting, propensity matching, or propensity stratification, in regression models of Y with respect to X. *However...*

Example: Propensity scores

- In particular, W may be a **propensity score**[1], predicting X from a list of multiple confounders $V_1, \ldots, V_K$.
- In the 21st–century Rubin method[6], the propensity score is a predictor from a regression model in the exposure, which we find in the joint distribution of the exposure and the confounders.
- We then add in the outcome data, and use the propensity score to estimate a propensity–adjusted exposure effect on the outcome.
- This adjusted exposure effect is typically interpreted as an exposure effect in a fantasy **target population**, with real–world marginal exposure and propensity distributions, but no exposure–propensity association.
- This is usually done using propensity weighting, propensity matching, or propensity stratification, in regression models of Y with respect to X. *However*. . .

Propensity scores and Somers' D

- ... a *possible* alternative candidate for the adjusted exposure effect is a **propensity–weighted Somers'** $D(Y|X)$.
- This uses **sampling–probability weights**, generated from the exposure X and the propensity score W, so that subjects with high exposure and low propensity, or low exposure and high propensity, are weighted upwards to remove the association between W and X.
- Before we add in the outcome data, we might want to be sure that our propensity weights are indeed removing this association.
- If the propensity–weighted $D(W|X)$ is close to zero, then we might be confident that a larger propensity–weighted $D(Y|X)$ cannot be secondary to it.
- And, if the propensity–weighted $D(W|X)$ is *not* close to zero, then we have diagnosed a problem with **non–overlap**, because there are not enough high–exposure low–propensity and low–exposure high–propensity subjects for us to weight upwards.

Propensity scores and Somers' D

- ... a *possible* alternative candidate for the adjusted exposure effect is a **propensity–weighted Somers'** $D(Y|X)$.
- This uses **sampling–probability weights**, generated from the exposure X and the propensity score W, so that subjects with high exposure and low propensity, or low exposure and high propensity, are weighted upwards to remove the association between W and X.
- Before we add in the outcome data, we might want to be sure that our propensity weights are indeed removing this association.
- If the propensity–weighted $D(W|X)$ is close to zero, then we might be confident that a larger propensity–weighted $D(Y|X)$ cannot be secondary to it.
- And, if the propensity–weighted $D(W|X)$ is *not* close to zero, then we have diagnosed a problem with **non–overlap**, because there are not enough high–exposure low–propensity and low–exposure high–propensity subjects for us to weight upwards.

Propensity scores and Somers' D

- ... a *possible* alternative candidate for the adjusted exposure effect is a **propensity–weighted Somers'** $D(Y|X)$.
- This uses **sampling–probability weights**, generated from the exposure X and the propensity score W, so that subjects with high exposure and low propensity, or low exposure and high propensity, are weighted upwards to remove the association between W and X.
- Before we add in the outcome data, we might want to be sure that our propensity weights are indeed removing this association.
- If the propensity–weighted $D(W|X)$ is close to zero, then we might be confident that a larger propensity–weighted $D(Y|X)$ cannot be secondary to it.
- And, if the propensity–weighted $D(W|X)$ is *not* close to zero, then we have diagnosed a problem with **non–overlap**, because there are not enough high–exposure low–propensity and low–exposure high–propensity subjects for us to weight upwards.

Propensity scores and Somers' D

- ... a *possible* alternative candidate for the adjusted exposure effect is a **propensity–weighted Somers'** $D(Y|X)$.
- This uses **sampling–probability weights**, generated from the exposure X and the propensity score W, so that subjects with high exposure and low propensity, or low exposure and high propensity, are weighted upwards to remove the association between W and X.
- Before we add in the outcome data, we might want to be sure that our propensity weights are indeed removing this association.
- If the propensity–weighted $D(W|X)$ is close to zero, then we might be confident that a larger propensity–weighted $D(Y|X)$ cannot be secondary to it.
- And, if the propensity–weighted $D(W|X)$ is *not* close to zero, then we have diagnosed a problem with **non–overlap**, because there are not enough high–exposure low–propensity and low–exposure high–propensity subjects for us to weight upwards.

Propensity scores and Somers' D

- … a *possible* alternative candidate for the adjusted exposure effect is a **propensity–weighted Somers'** $D(Y|X)$.
- This uses **sampling–probability weights**, generated from the exposure X and the propensity score W, so that subjects with high exposure and low propensity, or low exposure and high propensity, are weighted upwards to remove the association between W and X.
- Before we add in the outcome data, we might want to be sure that our propensity weights are indeed removing this association.
- If the propensity–weighted $D(W|X)$ is close to zero, then we might be confident that a larger propensity–weighted $D(Y|X)$ cannot be secondary to it.
- And, if the propensity–weighted $D(W|X)$ is *not* close to zero, then we have diagnosed a problem with **non–overlap**, because there are not enough high–exposure low–propensity and low–exposure high–propensity subjects for us to weight upwards.

Propensity scores and Somers' D

- . . . a *possible* alternative candidate for the adjusted exposure effect is a **propensity–weighted Somers'** $D(Y|X)$.
- This uses **sampling–probability weights**, generated from the exposure X and the propensity score W, so that subjects with high exposure and low propensity, or low exposure and high propensity, are weighted upwards to remove the association between W and X.
- Before we add in the outcome data, we might want to be sure that our propensity weights are indeed removing this association.
- If the propensity–weighted $D(W|X)$ is close to zero, then we might be confident that a larger propensity–weighted $D(Y|X)$ cannot be secondary to it.
- And, if the propensity–weighted $D(W|X)$ is *not* close to zero, then we have diagnosed a problem with **non–overlap**, because there are not enough high–exposure low–propensity and low–exposure high–propensity subjects for us to weight upwards.

Problem: Not everybody understands Somers' *D*

- Somers' *D* is expressed on a sensible scale from -1 to 1, as it is a difference between probabilities.
- *However*, an audience accustomed to other association measures, arising from specific models, may be culture–shocked when presented with arguments using Somers' *D*.
- *Fortunately*, under a range of familiar models, there is a one–to–one mapping between Somers' *D* and the association measure defined by the model.
- And, even better, these mappings are *nearly* linear (or at least log–linear), at least for Somers' *D* values between -0.5 and 0.5.
- This makes Somers' *D* a **common currency** for comparing associations measured using different models.
- We will now examine the currency conversions involved under four familiar example models.

Problem: Not everybody understands Somers' D

- Somers' D is expressed on a sensible scale from -1 to 1, as it is a difference between probabilities.
- *However*, an audience accustomed to other association measures, arising from specific models, may be culture–shocked when presented with arguments using Somers' D.
- *Fortunately*, under a range of familiar models, there is a one–to–one mapping between Somers' D and the association measure defined by the model.
- And, even better, these mappings are *nearly* linear (or at least log–linear), at least for Somers' D values between -0.5 and 0.5.
- This makes Somers' D a **common currency** for comparing associations measured using different models.
- We will now examine the currency conversions involved under four familiar example models.

Problem: Not everybody understands Somers' D

- Somers' D is expressed on a sensible scale from -1 to 1, as it is a difference between probabilities.
- *However*, an audience accustomed to other association measures, arising from specific models, may be culture–shocked when presented with arguments using Somers' D.
- *Fortunately*, under a range of familiar models, there is a one–to–one mapping between Somers' D and the association measure defined by the model.
- And, even better, these mappings are *nearly* linear (or at least log–linear), at least for Somers' D values between -0.5 and 0.5.
- This makes Somers' D a **common currency** for comparing associations measured using different models.
- We will now examine the currency conversions involved under four familiar example models.

Problem: Not everybody understands Somers' *D*

- Somers' *D* is expressed on a sensible scale from -1 to 1, as it is a difference between probabilities.
- *However*, an audience accustomed to other association measures, arising from specific models, may be culture–shocked when presented with arguments using Somers' *D*.
- *Fortunately*, under a range of familiar models, there is a one–to–one mapping between Somers' *D* and the association measure defined by the model.
- And, even better, these mappings are *nearly* linear (or at least log–linear), at least for Somers' *D* values between -0.5 and 0.5.
- This makes Somers' *D* a **common currency** for comparing associations measured using different models.
- We will now examine the currency conversions involved under four familiar example models.

Problem: Not everybody understands Somers' D

- Somers' D is expressed on a sensible scale from -1 to 1, as it is a difference between probabilities.
- *However*, an audience accustomed to other association measures, arising from specific models, may be culture–shocked when presented with arguments using Somers' D.
- *Fortunately*, under a range of familiar models, there is a one–to–one mapping between Somers' D and the association measure defined by the model.
- And, even better, these mappings are *nearly* linear (or at least log–linear), at least for Somers' D values between -0.5 and 0.5.
- This makes Somers' D a **common currency** for comparing associations measured using different models.
- We will now examine the currency conversions involved under four familiar example models.

Problem: Not everybody understands Somers' *D*

- Somers' *D* is expressed on a sensible scale from -1 to 1, as it is a difference between probabilities.
- *However*, an audience accustomed to other association measures, arising from specific models, may be culture–shocked when presented with arguments using Somers' *D*.
- *Fortunately*, under a range of familiar models, there is a one–to–one mapping between Somers' *D* and the association measure defined by the model.
- And, even better, these mappings are *nearly* linear (or at least log–linear), at least for Somers' *D* values between -0.5 and 0.5.
- This makes Somers' *D* a **common currency** for comparing associations measured using different models.
- We will now examine the currency conversions involved under four familiar example models.

Problem: Not everybody understands Somers' *D*

- Somers' *D* is expressed on a sensible scale from -1 to 1, as it is a difference between probabilities.
- *However*, an audience accustomed to other association measures, arising from specific models, may be culture–shocked when presented with arguments using Somers' *D*.
- *Fortunately*, under a range of familiar models, there is a one–to–one mapping between Somers' *D* and the association measure defined by the model.
- And, even better, these mappings are *nearly* linear (or at least log–linear), at least for Somers' *D* values between -0.5 and 0.5.
- This makes Somers' *D* a **common currency** for comparing associations measured using different models.
- We will now examine the currency conversions involved under four familiar example models.

Example 1: Binary X, binary Y

- In our first (trivial) example, we assume that the variables X and Y are both binary, with possible values 0 and 1.
- Somers' $D(Y|X)$ is then simply the difference between proportions

$$D(Y|X) = \Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0).$$

- *So*, the "scientific" rank parameter, measuring the strength of associations, is also the "practically–useful" regression parameter, measuring how much good we can do by intervening to change X.

Example 1: Binary X, binary Y

- In our first (trivial) example, we assume that the variables X and Y are both binary, with possible values 0 and 1.
- Somers' $D(Y|X)$ is then simply the difference between proportions

$$D(Y|X) = \Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0).$$

- *So*, the "scientific" rank parameter, measuring the strength of associations, is also the "practically–useful" regression parameter, measuring how much good we can do by intervening to change X.

Example 1: Binary X, binary Y

- In our first (trivial) example, we assume that the variables X and Y are both binary, with possible values 0 and 1.
- Somers' $D(Y|X)$ is then simply the difference between proportions

$$D(Y|X) = \Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0).$$

- *So*, the "scientific" rank parameter, measuring the strength of associations, is also the "practically–useful" regression parameter, measuring how much good we can do by intervening to change X.

Example 1: Binary X, binary Y

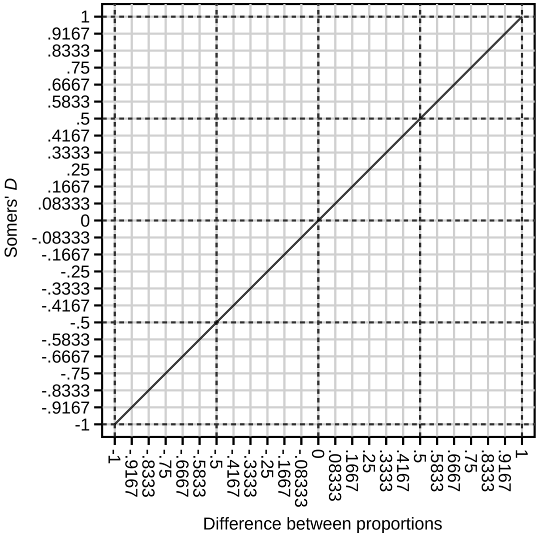
- In our first (trivial) example, we assume that the variables X and Y are both binary, with possible values 0 and 1.
- Somers' $D(Y|X)$ is then simply the difference between proportions

$$D(Y|X) = \Pr(Y = 1|X = 1) - \Pr(Y = 1|X = 0).$$

- *So*, the "scientific" rank parameter, measuring the strength of associations, is also the "practically–useful" regression parameter, measuring how much good we can do by intervening to change X.

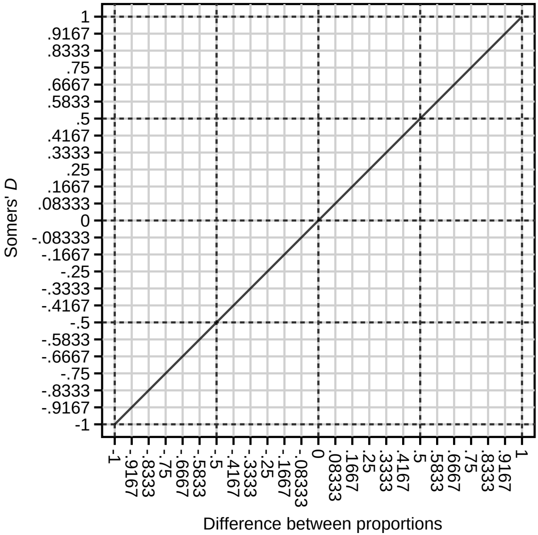
Somers' *D* plotted against the difference between proportions

- We plot Somers' *D* on the vertical axis against the difference between proportions on the horizontal axis.
- Note that the axes are labelled at multiples of 1/12, including $\pm 1/2$, $\pm 1/3$ and $\pm 1/4$.
- The reason for this will become clear in subsequent examples.



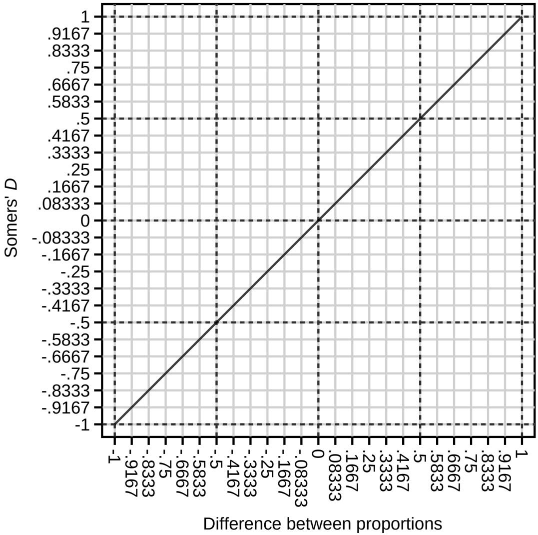
Somers' D plotted against the difference between proportions

- We plot Somers' D on the vertical axis against the difference between proportions on the horizontal axis.
- Note that the axes are labelled at multiples of 1/12, including $\pm 1/2$, $\pm 1/3$ and $\pm 1/4$.
- The reason for this will become clear in subsequent examples.



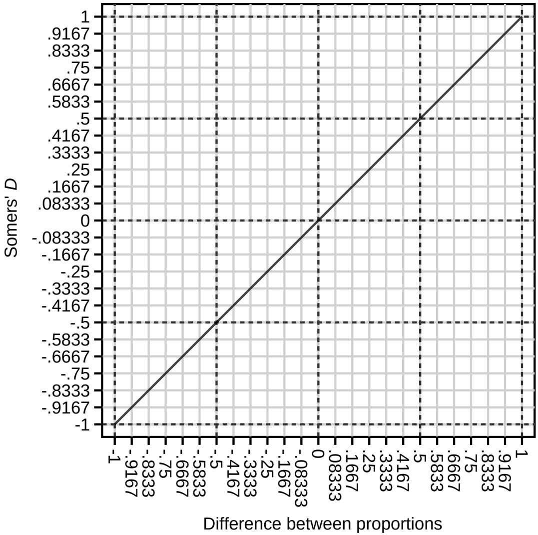
Somers' *D* plotted against the difference between proportions

- We plot Somers' *D* on the vertical axis against the difference between proportions on the horizontal axis.
- Note that the axes are labelled at multiples of 1/12, including $\pm 1/2$, $\pm 1/3$ and $\pm 1/4$.
- The reason for this will become clear in subsequent examples.



Somers' *D* plotted against the difference between proportions

- We plot Somers' *D* on the vertical axis against the difference between proportions on the horizontal axis.
- Note that the axes are labelled at multiples of 1/12, including $\pm 1/2$, $\pm 1/3$ and $\pm 1/4$.
- The reason for this will become clear in subsequent examples.



Example 2: Binary X, Normal Y, equal variances

- In this example, we assume that the variable X is binary, and that Y has Normal distributions conditionally on each X–value, with respective means μ_0 and μ_1, and a common standard deviation (SD) σ.
- Somers' $D(Y|X)$ is then given by

$$D(Y|X) = 2\Phi\left(\frac{\delta}{\sqrt{2}}\right) - 1,$$

where $\Phi(\cdot)$ is the standard Normal distribution function, and $\delta = (\mu_1 - \mu_0)/\sigma$ is the mean difference, expressed in units of the common SD.

- This relationship is not linear, but sigmoid.
- *However*, it implies that δ is *approximately* ± 0.954 SDs when Somers' D is ± 0.5, and *approximately* linear in between.

Example 2: Binary X, Normal Y, equal variances

- In this example, we assume that the variable X is binary, and that Y has Normal distributions conditionally on each X–value, with respective means μ_0 and μ_1, and a common standard deviation (SD) σ.
- Somers' $D(Y|X)$ is then given by

$$D(Y|X) = 2\Phi\left(\frac{\delta}{\sqrt{2}}\right) - 1,$$

  where $\Phi(\cdot)$ is the standard Normal distribution function, and $\delta = (\mu_1 - \mu_0)/\sigma$ is the mean difference, expressed in units of the common SD.
- This relationship is not linear, but sigmoid.
- *However*, it implies that δ is *approximately* ± 0.954 SDs when Somers' D is ± 0.5, and *approximately* linear in between.

Example 2: Binary X, Normal Y, equal variances

- In this example, we assume that the variable X is binary, and that Y has Normal distributions conditionally on each X–value, with respective means μ_0 and μ_1, and a common standard deviation (SD) σ.
- Somers' $D(Y|X)$ is then given by

$$D(Y|X) = 2\Phi\left(\frac{\delta}{\sqrt{2}}\right) - 1,$$

where $\Phi(\cdot)$ is the standard Normal distribution function, and $\delta = (\mu_1 - \mu_0)/\sigma$ is the mean difference, expressed in units of the common SD.

- This relationship is not linear, but sigmoid.
- *However*, it implies that δ is *approximately* ± 0.954 SDs when Somers' D is ± 0.5, and *approximately* linear in between.

Example 2: Binary X, Normal Y, equal variances

- In this example, we assume that the variable X is binary, and that Y has Normal distributions conditionally on each X–value, with respective means μ_0 and μ_1, and a common standard deviation (SD) σ.
- Somers' $D(Y|X)$ is then given by

$$D(Y|X) = 2\Phi\left(\frac{\delta}{\sqrt{2}}\right) - 1,$$

where $\Phi(\cdot)$ is the standard Normal distribution function, and $\delta = (\mu_1 - \mu_0)/\sigma$ is the mean difference, expressed in units of the common SD.
- This relationship is not linear, but sigmoid.
- *However*, it implies that δ is *approximately* ± 0.954 SDs when Somers' D is ± 0.5, and *approximately* linear in between.

Example 2: Binary X, Normal Y, equal variances

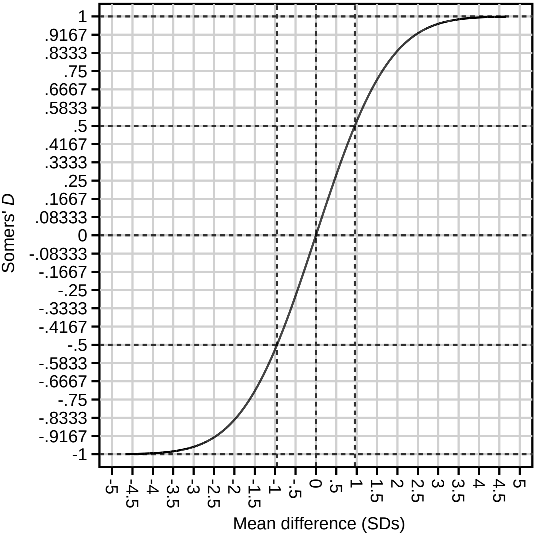
- In this example, we assume that the variable X is binary, and that Y has Normal distributions conditionally on each X–value, with respective means μ_0 and μ_1, and a common standard deviation (SD) σ.
- Somers' $D(Y|X)$ is then given by

$$D(Y|X) = 2\Phi\left(\frac{\delta}{\sqrt{2}}\right) - 1,$$

  where $\Phi(\cdot)$ is the standard Normal distribution function, and $\delta = (\mu_1 - \mu_0)/\sigma$ is the mean difference, expressed in units of the common SD.
- This relationship is not linear, but sigmoid.
- *However*, it implies that δ is *approximately* ± 0.954 SDs when Somers' D is ± 0.5, and *approximately* linear in between.

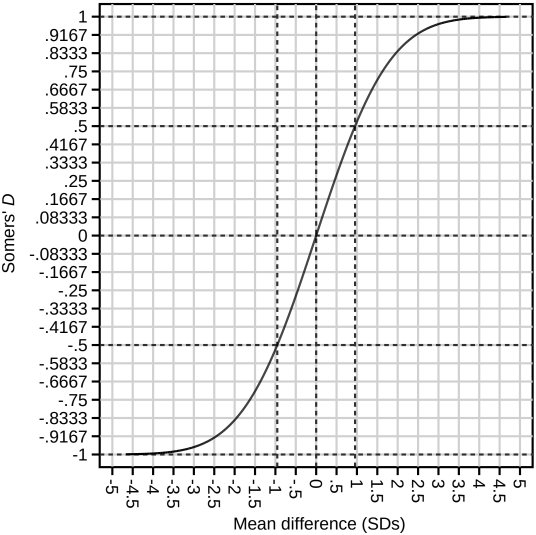
Somers' *D* plotted against the difference between means (in SDs)

- This time, the curve is sigmoid, as Somers' *D* is bounded between -1 and 1.
- *However*, there is a range of near–linearity between Somers' *D* values of -0.5 and 0.5.
- At these values, the mean differences are -0.954 and +0.954 SDs, respectively.



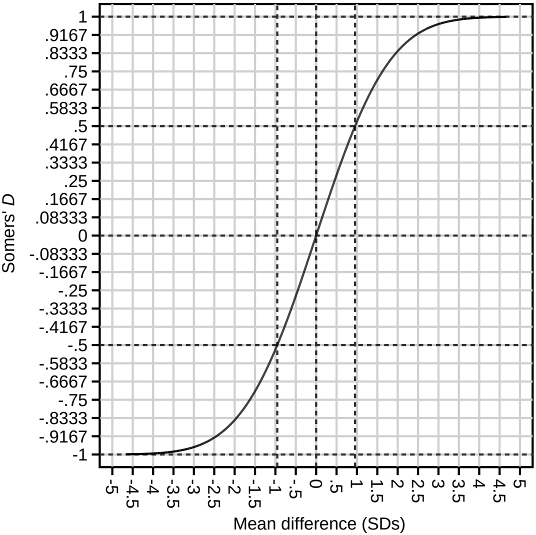
Somers' *D* plotted against the difference between means (in SDs)

- This time, the curve is sigmoid, as Somers' *D* is bounded between -1 and 1.
- *However*, there is a range of near–linearity between Somers' *D* values of -0.5 and 0.5.
- At these values, the mean differences are -0.954 and +0.954 SDs, respectively.



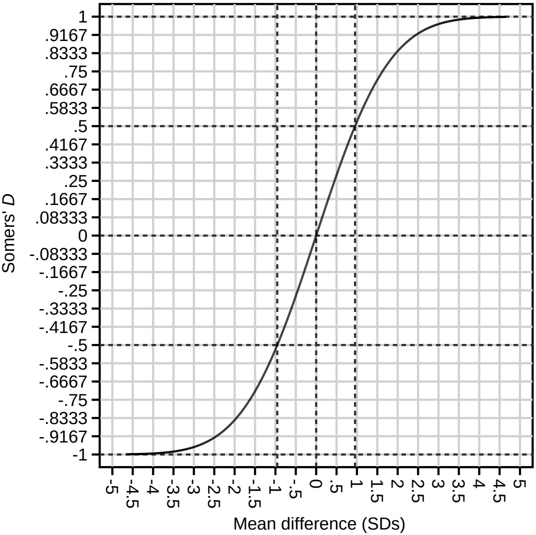
Somers' *D* plotted against the difference between means (in SDs)

- This time, the curve is sigmoid, as Somers' *D* is bounded between -1 and 1.
- *However*, there is a range of near–linearity between Somers' *D* values of -0.5 and 0.5.
- At these values, the mean differences are -0.954 and +0.954 SDs, respectively.



Somers' *D* plotted against the difference between means (in SDs)

- This time, the curve is sigmoid, as Somers' *D* is bounded between -1 and 1.
- *However*, there is a range of near–linearity between Somers' *D* values of -0.5 and 0.5.
- At these values, the mean differences are -0.954 and +0.954 SDs, respectively.



Alternative parameterization: The interquartile odds ratio

- Alternatively, under the same model, we might reverse the roles of X and Y, making X a binary disease–case status variable, Y a continuous logistic disease predictor, and $[D(Y|X) + 1]/2$ the area under the receiver–operating characteristic (ROC) curve[2].
- The interesting *model* parameter may then be the **interquartile odds ratio (IQOR)** between the 75th and 25th percentiles of the *control* distribution of Y, given by

$$\text{IQOR} = \exp\left\{\delta\left[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)\right]\right\},$$

  where δ is still the mean case–control difference (in SDs).
- This time, the curve is nearly log–linear between Somers' D values of -0.5 and 0.5.
- And, at these values, the IQORs are approximately 0.276 and 3.621, respectively.
- *So*, the IQORs corresponding to intermediate Somers' D values can be approximated well by weighted geometric means, interpolated log–linearly between 0.276 and 3.621.

Alternative parameterization: The interquartile odds ratio

- Alternatively, under the same model, we might reverse the roles of X and Y, making X a binary disease–case status variable, Y a continuous logistic disease predictor, and $[D(Y|X)+1]/2$ the area under the receiver–operating characteristic (ROC) curve[2].
- The interesting *model* parameter may then be the **interquartile odds ratio (IQOR)** between the 75th and 25th percentiles of the *control* distribution of Y, given by

$$\text{IQOR} = \exp\left\{\delta\left[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)\right]\right\},$$

  where δ is still the mean case–control difference (in SDs).
- This time, the curve is nearly log–linear between Somers' D values of -0.5 and 0.5.
- And, at these values, the IQORs are approximately 0.276 and 3.621, respectively.
- *So*, the IQORs corresponding to intermediate Somers' D values can be approximated well by weighted geometric means, interpolated log–linearly between 0.276 and 3.621.

Alternative parameterization: The interquartile odds ratio

- Alternatively, under the same model, we might reverse the roles of X and Y, making X a binary disease–case status variable, Y a continuous logistic disease predictor, and $[D(Y|X) + 1]/2$ the area under the receiver–operating characteristic (ROC) curve[2].
- The interesting *model* parameter may then be the **interquartile odds ratio (IQOR)** between the 75th and 25th percentiles of the *control* distribution of Y, given by

$$\text{IQOR} = \exp\left\{\delta\left[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)\right]\right\},$$

  where δ is still the mean case–control difference (in SDs).
- This time, the curve is nearly log–linear between Somers' D values of -0.5 and 0.5.
- And, at these values, the IQORs are approximately 0.276 and 3.621, respectively.
- *So*, the IQORs corresponding to intermediate Somers' D values can be approximated well by weighted geometric means, interpolated log–linearly between 0.276 and 3.621.

Alternative parameterization: The interquartile odds ratio

- Alternatively, under the same model, we might reverse the roles of X and Y, making X a binary disease–case status variable, Y a continuous logistic disease predictor, and $[D(Y|X) + 1]/2$ the area under the receiver–operating characteristic (ROC) curve[2].
- The interesting *model* parameter may then be the **interquartile odds ratio (IQOR)** between the 75th and 25th percentiles of the *control* distribution of Y, given by

$$\mathrm{IQOR} = \exp\left\{\delta\left[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)\right]\right\},$$

  where δ is still the mean case–control difference (in SDs).
- This time, the curve is nearly log–linear between Somers' D values of -0.5 and 0.5.
- And, at these values, the IQORs are approximately 0.276 and 3.621, respectively.
- *So*, the IQORs corresponding to intermediate Somers' D values can be approximated well by weighted geometric means, interpolated log–linearly between 0.276 and 3.621.

Alternative parameterization: The interquartile odds ratio

- Alternatively, under the same model, we might reverse the roles of X and Y, making X a binary disease–case status variable, Y a continuous logistic disease predictor, and $[D(Y|X)+1]/2$ the area under the receiver–operating characteristic (ROC) curve[2].
- The interesting *model* parameter may then be the **interquartile odds ratio (IQOR)** between the 75th and 25th percentiles of the *control* distribution of Y, given by

$$\mathrm{IQOR} = \exp\left\{\delta\left[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)\right]\right\},$$

  where δ is still the mean case–control difference (in SDs).
- This time, the curve is nearly log–linear between Somers' D values of -0.5 and 0.5.
- And, at these values, the IQORs are approximately 0.276 and 3.621, respectively.
- *So*, the IQORs corresponding to intermediate Somers' D values can be approximated well by weighted geometric means, interpolated log–linearly between 0.276 and 3.621.

Alternative parameterization: The interquartile odds ratio

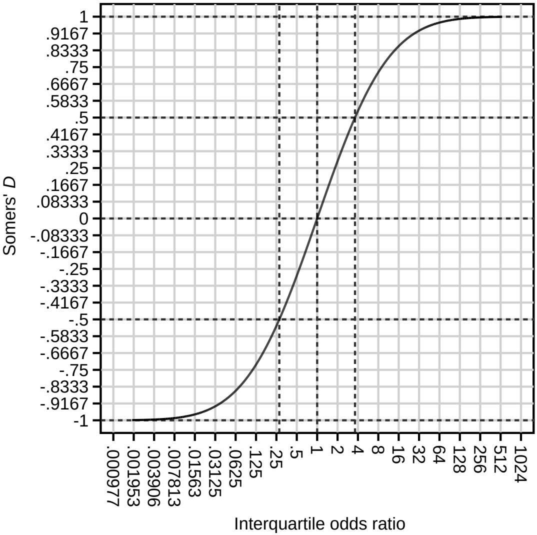
- Alternatively, under the same model, we might reverse the roles of X and Y, making X a binary disease–case status variable, Y a continuous logistic disease predictor, and $[D(Y|X)+1]/2$ the area under the receiver–operating characteristic (ROC) curve[2].
- The interesting *model* parameter may then be the **interquartile odds ratio (IQOR)** between the 75th and 25th percentiles of the *control* distribution of Y, given by

$$\mathrm{IQOR} = \exp\left\{\delta\left[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)\right]\right\},$$

  where δ is still the mean case–control difference (in SDs).
- This time, the curve is nearly log–linear between Somers' D values of -0.5 and 0.5.
- And, at these values, the IQORs are approximately 0.276 and 3.621, respectively.
- *So*, the IQORs corresponding to intermediate Somers' D values can be approximated well by weighted geometric means, interpolated log–linearly between 0.276 and 3.621.

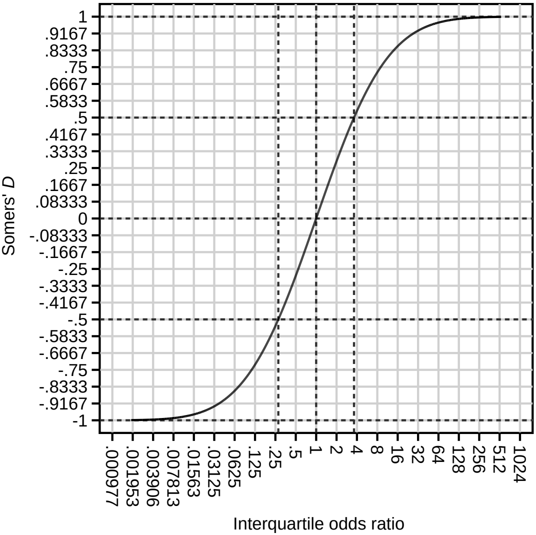
Somers' *D* plotted against the interquartile odds ratio (IQOR)

- Note that the IQOR (on the horizontal axis) is plotted on a binary log scale.
- *However*, after this transformation, there is again a range of near–linearity between Somers' *D* values of -0.5 and 0.5.
- And, at these values, the IQORs are 0.276 and 3.621, respectively.



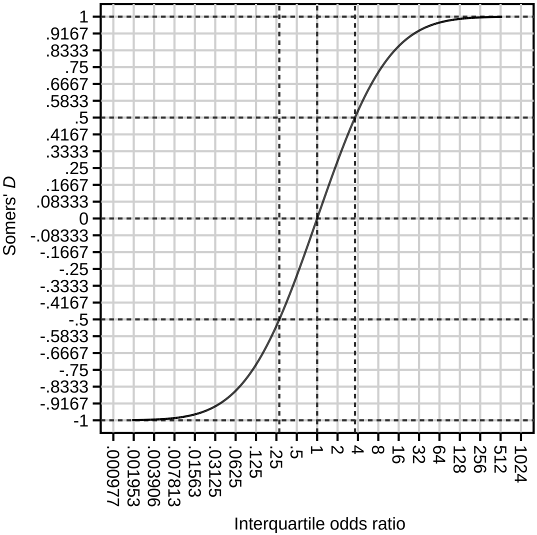
Somers' *D* plotted against the interquartile odds ratio (IQOR)

- Note that the IQOR (on the horizontal axis) is plotted on a binary log scale.
- *However*, after this transformation, there is again a range of near–linearity between Somers' *D* values of -0.5 and 0.5.
- And, at these values, the IQORs are 0.276 and 3.621, respectively.



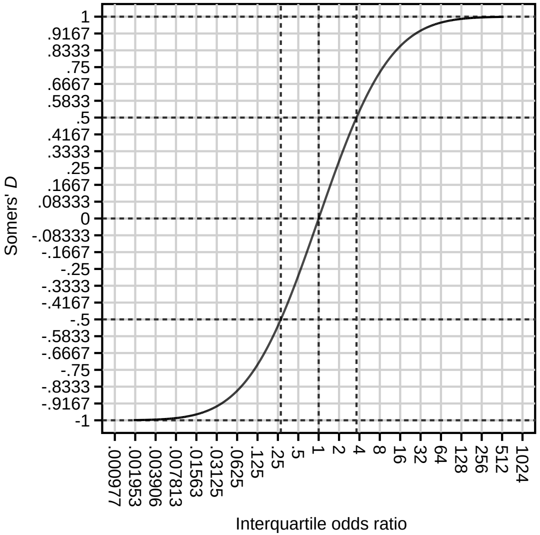
Somers' *D* plotted against the interquartile odds ratio (IQOR)

- Note that the IQOR (on the horizontal axis) is plotted on a binary log scale.
- *However*, after this transformation, there is again a range of near–linearity between Somers' *D* values of -0.5 and 0.5.
- And, at these values, the IQORs are 0.276 and 3.621, respectively.



Somers' *D* plotted against the interquartile odds ratio (IQOR)

- Note that the IQOR (on the horizontal axis) is plotted on a binary log scale.
- *However*, after this transformation, there is again a range of near–linearity between Somers' *D* values of -0.5 and 0.5.
- And, at these values, the IQORs are 0.276 and 3.621, respectively.



Example 3: Binary X, continuous lifetime Y, constant hazard ratio

- This time, Y is a positive–valued continuous lifetime variable, with a constant hazard ratio R between the subpopulations identified by $X = 1$ and $X = 0$ (as in a Cox regression).
- In the absence of censorship, Somers' $D(Y|X)$ is then given by

$$D(Y|X) = (1 - R)/(1 + R).$$

- *So*, the Somers' D values corresponding to hazard ratios of 3, 2, 1, 1/2 and 1/3 are -1/2, -1/3, 0, 1/3 and 1/2, respectively.
- This relationship is decreasing, and nearly log–linear for Somers' D between -0.5 and 0.5.
- Note that the hazard ratio (unlike Somers' D) is still the same in the *presence* of censorship.

Example 3: Binary X, continuous lifetime Y, constant hazard ratio

- This time, Y is a positive–valued continuous lifetime variable, with a constant hazard ratio R between the subpopulations identified by $X = 1$ and $X = 0$ (as in a Cox regression).
- In the absence of censorship, Somers' $D(Y|X)$ is then given by

$$D(Y|X) = (1 - R)/(1 + R).$$

- *So*, the Somers' D values corresponding to hazard ratios of 3, 2, 1, 1/2 and 1/3 are -1/2, -1/3, 0, 1/3 and 1/2, respectively.
- This relationship is decreasing, and nearly log–linear for Somers' D between -0.5 and 0.5.
- Note that the hazard ratio (unlike Somers' D) is still the same in the *presence* of censorship.

Example 3: Binary X, continuous lifetime Y, constant hazard ratio

- This time, Y is a positive–valued continuous lifetime variable, with a constant hazard ratio R between the subpopulations identified by $X = 1$ and $X = 0$ (as in a Cox regression).
- In the absence of censorship, Somers' $D(Y|X)$ is then given by

$$D(Y|X) = (1 - R)/(1 + R).$$

- *So*, the Somers' D values corresponding to hazard ratios of 3, 2, 1, 1/2 and 1/3 are -1/2, -1/3, 0, 1/3 and 1/2, respectively.
- This relationship is decreasing, and nearly log–linear for Somers' D between -0.5 and 0.5.
- Note that the hazard ratio (unlike Somers' D) is still the same in the *presence* of censorship.

Example 3: Binary X, continuous lifetime Y, constant hazard ratio

- This time, Y is a positive–valued continuous lifetime variable, with a constant hazard ratio R between the subpopulations identified by $X = 1$ and $X = 0$ (as in a Cox regression).
- In the absence of censorship, Somers' $D(Y|X)$ is then given by

$$D(Y|X) = (1 - R)/(1 + R).$$

- *So*, the Somers' D values corresponding to hazard ratios of 3, 2, 1, 1/2 and 1/3 are -1/2, -1/3, 0, 1/3 and 1/2, respectively.
- This relationship is decreasing, and nearly log–linear for Somers' D between -0.5 and 0.5.
- Note that the hazard ratio (unlike Somers' D) is still the same in the *presence* of censorship.

Example 3: Binary X, continuous lifetime Y, constant hazard ratio

- This time, Y is a positive–valued continuous lifetime variable, with a constant hazard ratio R between the subpopulations identified by $X = 1$ and $X = 0$ (as in a Cox regression).
- In the absence of censorship, Somers' $D(Y|X)$ is then given by

$$D(Y|X) = (1 - R)/(1 + R).$$

- *So*, the Somers' D values corresponding to hazard ratios of 3, 2, 1, 1/2 and 1/3 are -1/2, -1/3, 0, 1/3 and 1/2, respectively.
- This relationship is decreasing, and nearly log–linear for Somers' D between -0.5 and 0.5.
- Note that the hazard ratio (unlike Somers' D) is still the same in the *presence* of censorship.

Example 3: Binary X, continuous lifetime Y, constant hazard ratio

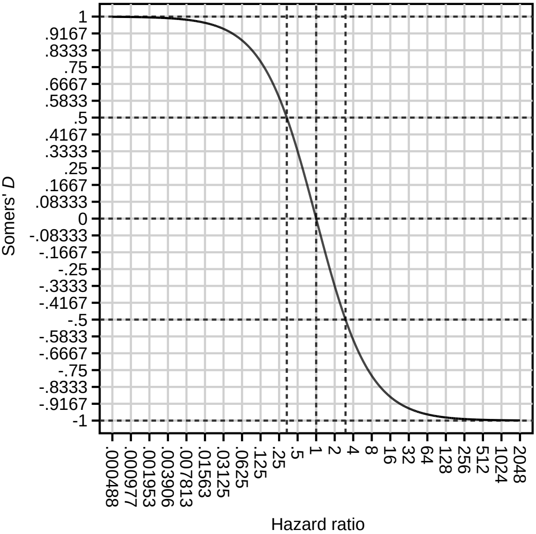
- This time, Y is a positive–valued continuous lifetime variable, with a constant hazard ratio R between the subpopulations identified by $X = 1$ and $X = 0$ (as in a Cox regression).
- In the absence of censorship, Somers' $D(Y|X)$ is then given by

$$D(Y|X) = (1 - R)/(1 + R).$$

- *So*, the Somers' D values corresponding to hazard ratios of 3, 2, 1, 1/2 and 1/3 are -1/2, -1/3, 0, 1/3 and 1/2, respectively.
- This relationship is decreasing, and nearly log–linear for Somers' D between -0.5 and 0.5.
- Note that the hazard ratio (unlike Somers' D) is still the same in the *presence* of censorship.

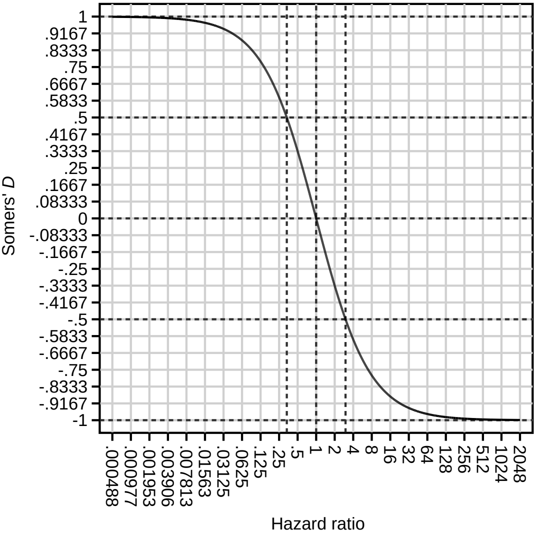
Somers' *D* plotted against the hazard ratio

- Again, the horizontal axis is on a binary log scale.
- The curve is decreasing, but nearly log–linear for Somers' *D* between -0.5 and 0.5.
- The respective hazard ratios, corresponding to these limits, are 3 and 1/3.
- This log–linearity range includes the typical smoking–related hazard ratio of 2, corresponding to an *uncensored* Somers' *D* of -1/3.



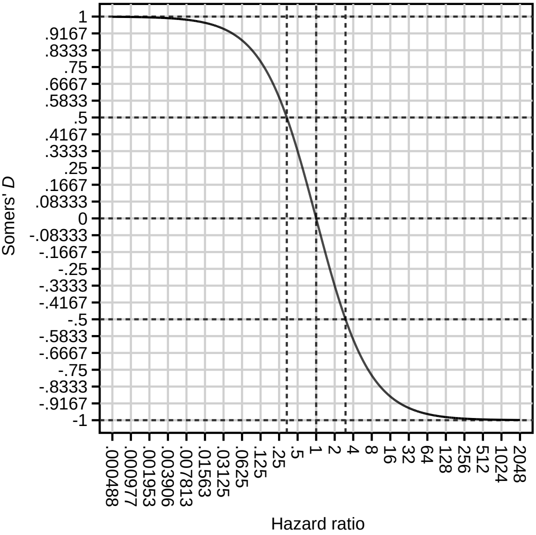
Somers' *D* plotted against the hazard ratio

- Again, the horizontal axis is on a binary log scale.
- The curve is decreasing, but nearly log–linear for Somers' *D* between -0.5 and 0.5.
- The respective hazard ratios, corresponding to these limits, are 3 and 1/3.
- This log–linearity range includes the typical smoking–related hazard ratio of 2, corresponding to an *uncensored* Somers' *D* of -1/3.



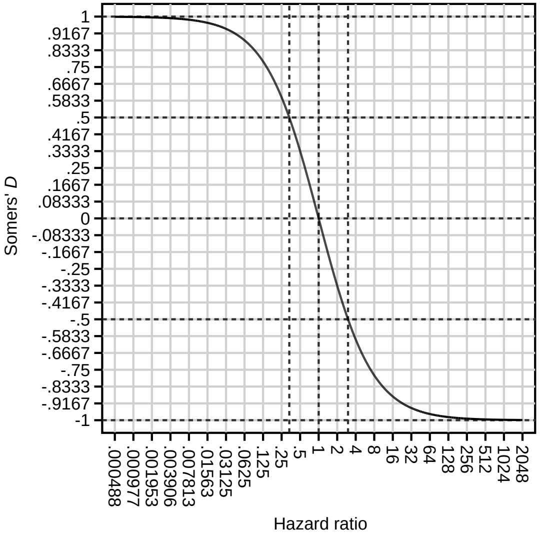
Somers' *D* plotted against the hazard ratio

- Again, the horizontal axis is on a binary log scale.
- The curve is decreasing, but nearly log–linear for Somers' *D* between -0.5 and 0.5.
- The respective hazard ratios, corresponding to these limits, are 3 and 1/3.
- This log–linearity range includes the typical smoking–related hazard ratio of 2, corresponding to an *uncensored* Somers' *D* of -1/3.



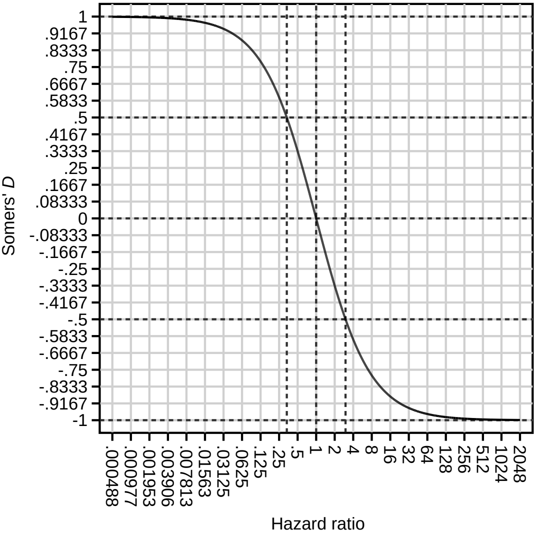
Somers' *D* plotted against the hazard ratio

- Again, the horizontal axis is on a binary log scale.
- The curve is decreasing, but nearly log–linear for Somers' *D* between -0.5 and 0.5.
- The respective hazard ratios, corresponding to these limits, are 3 and 1/3.
- This log–linearity range includes the typical smoking–related hazard ratio of 2, corresponding to an *uncensored* Somers' *D* of -1/3.



Somers' *D* plotted against the hazard ratio

- Again, the horizontal axis is on a binary log scale.
- The curve is decreasing, but nearly log–linear for Somers' *D* between -0.5 and 0.5.
- The respective hazard ratios, corresponding to these limits, are 3 and 1/3.
- This log–linearity range includes the typical smoking–related hazard ratio of 2, corresponding to an *uncensored* Somers' *D* of -1/3.



Example 4: Bivariate Normal X and Y

- This time, X and Y have a bivariate Normal joint distribution, with means μ_X and μ_Y, SDs σ_X and σ_Y, and a Pearson correlation coefficient ρ.
- Somers' $D(Y|X)$ is then equal to Kendall's tau-a, and is related to the Pearson correlation by **Greiner's relation**,

$$D(Y|X) = \tau_{XY} = \frac{2}{\pi}\arcsin(\rho).$$

- *So*, the Somers' D values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1 correspond to Pearson correlations of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1, respectively.
- This relationship, again, is nearly linear (with slope $2/\pi$) for Somers' D between -0.5 and 0.5.
- Note that Greiner's relation still holds if X and/or Y is derived using a Normalizing transformation.

Example 4: Bivariate Normal X and Y

- This time, X and Y have a bivariate Normal joint distribution, with means μ_X and μ_Y, SDs σ_X and σ_Y, and a Pearson correlation coefficient ρ.
- Somers' $D(Y|X)$ is then equal to Kendall's tau-a, and is related to the Pearson correlation by **Greiner's relation**,

$$D(Y|X) = \tau_{XY} = \frac{2}{\pi} \arcsin(\rho).$$

- *So*, the Somers' D values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1 correspond to Pearson correlations of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1, respectively.
- This relationship, again, is nearly linear (with slope $2/\pi$) for Somers' D between -0.5 and 0.5.
- Note that Greiner's relation still holds if X and/or Y is derived using a Normalizing transformation.

Example 4: Bivariate Normal X and Y

- This time, X and Y have a bivariate Normal joint distribution, with means μ_X and μ_Y, SDs σ_X and σ_Y, and a Pearson correlation coefficient ρ.
- Somers' $D(Y|X)$ is then equal to Kendall's tau-a, and is related to the Pearson correlation by **Greiner's relation**,

$$D(Y|X) = \tau_{XY} = \frac{2}{\pi} \arcsin(\rho).$$

- *So*, the Somers' D values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1 correspond to Pearson correlations of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1, respectively.
- This relationship, again, is nearly linear (with slope $2/\pi$) for Somers' D between -0.5 and 0.5.
- Note that Greiner's relation still holds if X and/or Y is derived using a Normalizing transformation.

Example 4: Bivariate Normal X and Y

- This time, X and Y have a bivariate Normal joint distribution, with means μ_X and μ_Y, SDs σ_X and σ_Y, and a Pearson correlation coefficient ρ.
- Somers' $D(Y|X)$ is then equal to Kendall's tau-a, and is related to the Pearson correlation by **Greiner's relation**,

$$D(Y|X) = \tau_{XY} = \frac{2}{\pi}\arcsin(\rho).$$

- *So*, the Somers' D values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1 correspond to Pearson correlations of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1, respectively.
- This relationship, again, is nearly linear (with slope $2/\pi$) for Somers' D between -0.5 and 0.5.
- Note that Greiner's relation still holds if X and/or Y is derived using a Normalizing transformation.

Example 4: Bivariate Normal X and Y

- This time, X and Y have a bivariate Normal joint distribution, with means μ_X and μ_Y, SDs σ_X and σ_Y, and a Pearson correlation coefficient ρ.
- Somers' $D(Y|X)$ is then equal to Kendall's tau-a, and is related to the Pearson correlation by **Greiner's relation**,

$$D(Y|X) = \tau_{XY} = \frac{2}{\pi} \arcsin(\rho).$$

- *So*, the Somers' D values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1 correspond to Pearson correlations of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1, respectively.
- This relationship, again, is nearly linear (with slope $2/\pi$) for Somers' D between -0.5 and 0.5.
- Note that Greiner's relation still holds if X and/or Y is derived using a Normalizing transformation.

Example 4: Bivariate Normal X and Y

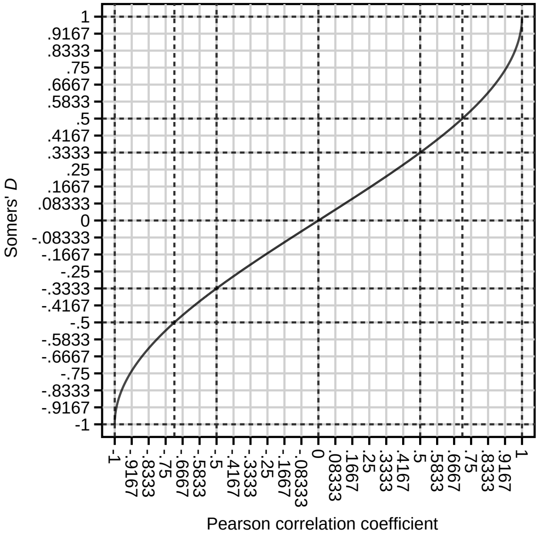
- This time, X and Y have a bivariate Normal joint distribution, with means μ_X and μ_Y, SDs σ_X and σ_Y, and a Pearson correlation coefficient ρ.
- Somers' $D(Y|X)$ is then equal to Kendall's tau-a, and is related to the Pearson correlation by **Greiner's relation**,

$$D(Y|X) = \tau_{XY} = \frac{2}{\pi} \arcsin(\rho).$$

- *So*, the Somers' D values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1 correspond to Pearson correlations of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1, respectively.
- This relationship, again, is nearly linear (with slope $2/\pi$) for Somers' D between -0.5 and 0.5.
- Note that Greiner's relation still holds if X and/or Y is derived using a Normalizing transformation.

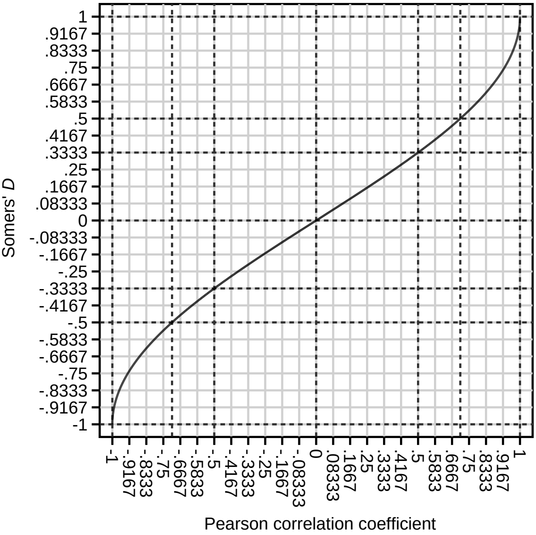
Somers' *D* plotted against the Pearson correlation coefficient

- The curve is nearly linear (with slope $2/\pi$) for Somers' *D* values between -0.5 and 0.5.
- The dashed lines on the vertical axis denote the Somers' *D* values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1.
- The dashed lines on the horizontal axis give the corresponding Pearson ρ values of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1.



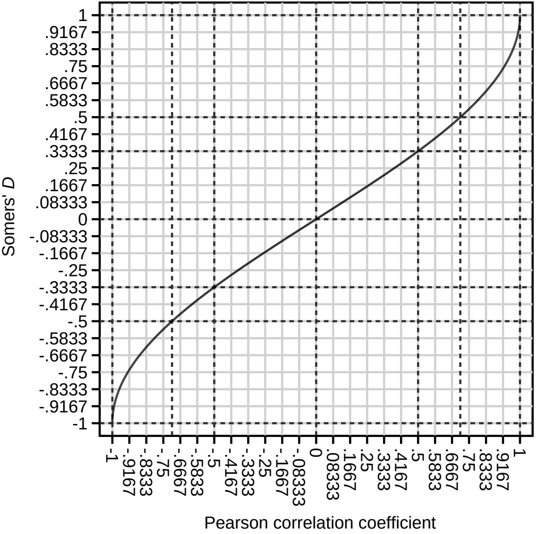
Somers' *D* plotted against the Pearson correlation coefficient

- The curve is nearly linear (with slope $2/\pi$) for Somers' *D* values between -0.5 and 0.5.
- The dashed lines on the vertical axis denote the Somers' *D* values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1.
- The dashed lines on the horizontal axis give the corresponding Pearson ρ values of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1.



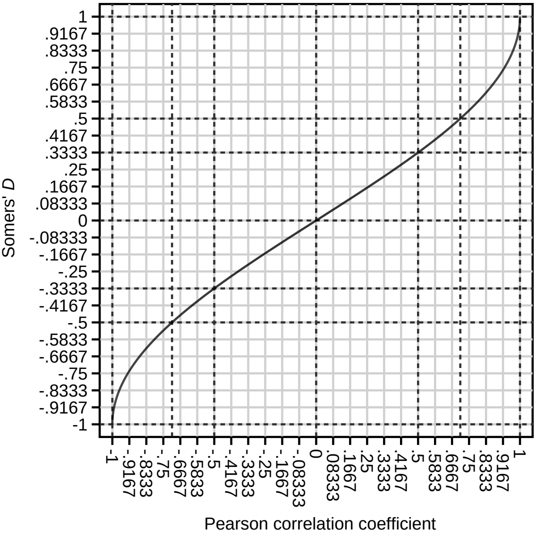
Somers' *D* plotted against the Pearson correlation coefficient

- The curve is nearly linear (with slope $2/\pi$) for Somers' *D* values between -0.5 and 0.5.
- The dashed lines on the vertical axis denote the Somers' *D* values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1.
- The dashed lines on the horizontal axis give the corresponding Pearson ρ values of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1.



Somers' D plotted against the Pearson correlation coefficient

- The curve is nearly linear (with slope $2/\pi$) for Somers' D values between -0.5 and 0.5.
- The dashed lines on the vertical axis denote the Somers' D values of 0, $\pm 1/3$, $\pm 1/2$, and ± 1.
- The dashed lines on the horizontal axis give the corresponding Pearson ρ values of 0, $\pm 1/2$, $\pm\sqrt{1/2}$, and ± 1.



Summary: Somers' *D* as a common currency between associations

- Somers' *D* defines a hierarchy between monotonic associations, stating (scientifically) which ones may *not* be secondary to which others, without committing ourselves to a particular model.
- *However*, a lot of statistically–minded scientists prefer to measure associations using model parameters, and answer *practical* questions about how much good they can do.
- *Fortunately*, under a range of familiar models, Somers' *D* is monotonic in the more familiar model parameter, and nearly linear (or log–linear) for Somers' *D* values between -0.5 and 0.5.
- *So*, given a Somers' *D* between -0.5 and 0.5, we can interpolate linearly (or log–linearly) to limit the range of proportion/mean differences, odds/hazard ratios, or Pearson correlations, which *may* be secondary to an association with that Somers' *D*.
- A more equation–intensive version of this story can be found in the "Miscellaneous documents" section of the author's website[3].

Summary: Somers' *D* as a common currency between associations

- Somers' *D* defines a hierarchy between monotonic associations, stating (scientifically) which ones may *not* be secondary to which others, without committing ourselves to a particular model.
- *However*, a lot of statistically–minded scientists prefer to measure associations using model parameters, and answer *practical* questions about how much good they can do.
- *Fortunately*, under a range of familiar models, Somers' *D* is monotonic in the more familiar model parameter, and nearly linear (or log–linear) for Somers' *D* values between -0.5 and 0.5.
- *So*, given a Somers' *D* between -0.5 and 0.5, we can interpolate linearly (or log–linearly) to limit the range of proportion/mean differences, odds/hazard ratios, or Pearson correlations, which *may* be secondary to an association with that Somers' *D*.
- A more equation–intensive version of this story can be found in the "Miscellaneous documents" section of the author's website[3].

Summary: Somers' *D* as a common currency between associations

- Somers' *D* defines a hierarchy between monotonic associations, stating (scientifically) which ones may *not* be secondary to which others, without committing ourselves to a particular model.
- *However*, a lot of statistically–minded scientists prefer to measure associations using model parameters, and answer *practical* questions about how much good they can do.
- *Fortunately*, under a range of familiar models, Somers' *D* is monotonic in the more familiar model parameter, and nearly linear (or log–linear) for Somers' *D* values between -0.5 and 0.5.
- *So*, given a Somers' *D* between -0.5 and 0.5, we can interpolate linearly (or log–linearly) to limit the range of proportion/mean differences, odds/hazard ratios, or Pearson correlations, which *may* be secondary to an association with that Somers' *D*.
- A more equation–intensive version of this story can be found in the "Miscellaneous documents" section of the author's website[3].

Summary: Somers' *D* as a common currency between associations

- Somers' *D* defines a hierarchy between monotonic associations, stating (scientifically) which ones may *not* be secondary to which others, without committing ourselves to a particular model.
- *However*, a lot of statistically–minded scientists prefer to measure associations using model parameters, and answer *practical* questions about how much good they can do.
- *Fortunately*, under a range of familiar models, Somers' *D* is monotonic in the more familiar model parameter, and nearly linear (or log–linear) for Somers' *D* values between -0.5 and 0.5.
- *So*, given a Somers' *D* between -0.5 and 0.5, we can interpolate linearly (or log–linearly) to limit the range of proportion/mean differences, odds/hazard ratios, or Pearson correlations, which *may* be secondary to an association with that Somers' *D*.
- A more equation–intensive version of this story can be found in the "Miscellaneous documents" section of the author's website[3].

Summary: Somers' *D* as a common currency between associations

- Somers' *D* defines a hierarchy between monotonic associations, stating (scientifically) which ones may *not* be secondary to which others, without committing ourselves to a particular model.
- *However*, a lot of statistically–minded scientists prefer to measure associations using model parameters, and answer *practical* questions about how much good they can do.
- *Fortunately*, under a range of familiar models, Somers' *D* is monotonic in the more familiar model parameter, and nearly linear (or log–linear) for Somers' *D* values between -0.5 and 0.5.
- *So*, given a Somers' *D* between -0.5 and 0.5, we can interpolate linearly (or log–linearly) to limit the range of proportion/mean differences, odds/hazard ratios, or Pearson correlations, which *may* be secondary to an association with that Somers' *D*.
- A more equation–intensive version of this story can be found in the "Miscellaneous documents" section of the author's website[3].

Summary: Somers' *D* as a common currency between associations

- Somers' *D* defines a hierarchy between monotonic associations, stating (scientifically) which ones may *not* be secondary to which others, without committing ourselves to a particular model.
- *However*, a lot of statistically–minded scientists prefer to measure associations using model parameters, and answer *practical* questions about how much good they can do.
- *Fortunately*, under a range of familiar models, Somers' *D* is monotonic in the more familiar model parameter, and nearly linear (or log–linear) for Somers' *D* values between -0.5 and 0.5.
- *So*, given a Somers' *D* between -0.5 and 0.5, we can interpolate linearly (or log–linearly) to limit the range of proportion/mean differences, odds/hazard ratios, or Pearson correlations, which *may* be secondary to an association with that Somers' *D*.
- A more equation–intensive version of this story can be found in the "Miscellaneous documents" section of the author's website[3].

References

[1] Guo, S. and Fraser, M. W. 2014. *Propensity score analysis. Second edition.* Los Angeles, CA: Sage.

[2] Hanley, J. A. and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**: 29–36.

[3] Newson, R. B. 2014. Interpretation of Somers' *D* under four simple models. Downloaded on 25 February 2015 from the author's website at *http://www.imperial.ac.uk/nhli/r.newson/papers.htm#miscellaneous_documents*

[4] Newson, R. 2006. Confidence intervals for rank statistics: Somers' *D* and extensions. *The Stata Journal* **6(3)**: 309–334.

[5] Newson, R. B. 1987. *An analysis of cinematographic cell division data using U–statistics.* DPhil thesis, Sussex University, Brighton, UK.

[6] Rubin, D. B. 2008. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics* **2(3)**: 808–840.

This presentation, and the do–file producing the examples, can be downloaded from the conference website at *http://ideas.repec.org/s/boc/usug15.html*

The packages described and used in this presentation can be downloaded from SSC, using the `ssc` command.