STAT 226 Lecture 21
Section 6.2 Cumulative Logit Models for Ordinal Responses

Yibi Huang

6.2 Cumulative Logit Models for Ordinal Responses

Suppose the response Y is multinomial with **ordered** categories

$$\{1,\ 2,\ \dots\ ,J\}.$$

Let $\pi_i = \mathrm{P}(Y = i)$. The **cumulative probabilities** are

$$\mathrm{P}(Y \le j) = \pi_1 + \cdots + \pi_j, \quad j = 1, 2, ..., J.$$

- Note $\mathrm{P}(Y \le 1) \le \mathrm{P}(Y \le 2) \le \ldots \le \mathrm{P}(Y \le J) = 1$
- If Y is not ordinal, it's nonsense to say "$Y \le j$."

The **cumulative logits** are

$$\text{logit}[\mathrm{P}(Y \le j)] = \log\left(\frac{\mathrm{P}(Y \le j)}{1 - \mathrm{P}(Y \le j)}\right) = \log\left(\frac{\mathrm{P}(Y \le j)}{\mathrm{P}(Y > j)}\right)$$
$$= \log\left(\frac{\pi_1 + \cdots + \pi_j}{\pi_{j+1} + \cdots + \pi_J}\right), \quad j = 1, ..., J - 1.$$

Cumulative Logit Models

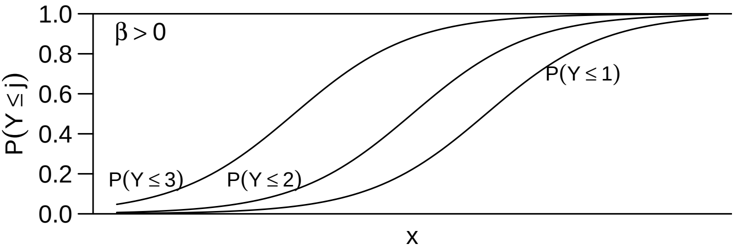
$$\text{logit}[P(Y \leq j\,|x)] = \alpha_j + \beta x, \quad j = 1, \ldots, J-1.$$

- separate intercept α_j for each cumulative logit
- *same slope* β for all cumulative logits

$\Rightarrow$ Curves of $P(Y \leq j\,|x)$ are "parallel", never cross each other.

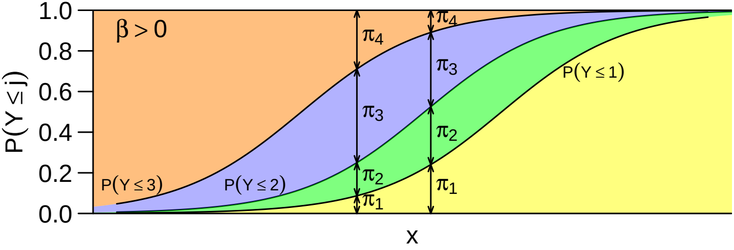
As long as $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_{J-1}$, we can ensure that

$$P(Y \leq j\,|x) = \frac{\exp(\alpha_j + \beta x)}{1 + \exp(\alpha_j + \beta x)} \leq \frac{\exp(\alpha_{j+1} + \beta x)}{1 + \exp(\alpha_{j+1} + \beta x)} = P(Y \leq j + 1|x)$$



Cumulative Logit Models

$$P(Y \le j|x) = \frac{\exp(\alpha_j + \beta x)}{1 + \exp(\alpha_j + \beta x)}, \quad j = 1, \ldots, J-1.$$



$$\pi_j(x) = P(Y = j|x) = P(Y \le j|x) - P(Y \le j-1|x)$$
$$= \frac{\exp(\alpha_j + \beta x)}{1 + \exp(\alpha_j + \beta x)} - \frac{\exp(\alpha_{j-1} + \beta x)}{1 + \exp(\alpha_{j-1} + \beta x)}$$

- If $\beta > 0$, as $x \uparrow$, Y is more likely at the **lower** categories ($Y \le j$)
- If $\beta < 0$, as $x \uparrow$, Y is more likely at the **higher** categories ($Y > j$)

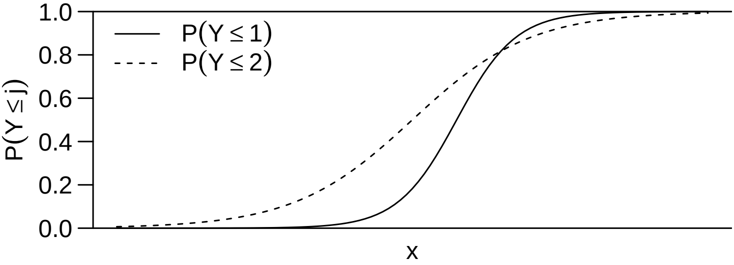
"Non-Parallel" Cumulative Logit Models

$$\text{logit}[P(Y \le j|x)] = \alpha_j + \beta_j x, \quad j = 1, \ldots, J-1.$$

- separate intercept α_j for each cumulative logit
- *separate slope* β_j for each cumulative logit

However, $P(Y \le j)$ curves in "non-parallel" cumulative logit models may cross each other and hence may not maintain that

$$P(Y \le j|x) \le P(Y \le j+1|x) \quad \text{for all } x$$



Properties of Cumulative Logit Models

$$\text{odds}(Y \le j|x) = \frac{P(Y \le j|x)}{P(Y > j|x)} = \exp(\alpha_j + \beta x), \quad j = 1, \ldots, J-1.$$

- $\exp(\beta)$ = multiplicative effect of 1-unit increase in x on odds that $(Y \le j)$ (instead of $(Y > j)$).

  $$\frac{\text{odds}(Y \le j|x_2)}{\text{odds}(Y \le j|x_1)} = \frac{\exp(\alpha_j + \beta x_2)}{\exp(\alpha_j + \beta x_1)} = \exp(\beta(x_2 - x_1)) \text{ for all } j$$

  So cumulative logit models are also called **proportional odds models**.
- ML estimates for coefficients (α_j, β) can be found via R function `vglm()` in the package `VGAM` w/ `cumulative family`.

Exercise 2.21 (Job Satisfaction and Income, ICDA, p.61)

Income	Job Satisfaction (Y)			
(x)	Dissat	Little	Moderate	Very
0-5K	2	4	13	3
5-15K	2	6	22	4
15-25K	0	1	15	8
>25K	0	3	13	8

Using x = income scores (3K, 10K, 20K, 35K), we fit the model

$$\text{logit}[P(Y \le j|x)] = \alpha_j + \beta x, \quad j = 1, 2, 3.$$

```
library(VGAM)
Income = c(3,10,20,35)
Diss = c(2,2,0,0)
Little = c(4,6,1,3)
Mod = c(13,22,15,13)
Very = c(3,4,8,8)
jobsat.cl1 = vglm(cbind(Diss,Little,Mod,Very) ~ Income,
                  family=cumulative(parallel=TRUE))
```

```
coef(jobsat.cl1)
(Intercept):1 (Intercept):2 (Intercept):3        Income
     -2.58287      -0.89698       2.07506      -0.04486
```

A more organized view of the fitted coefficients:

```
coef(jobsat.cl1, matrix=TRUE)
            logitlink(P[Y<=1]) logitlink(P[Y<=2]) logitlink(P[Y<=3])
(Intercept)           -2.58287           -0.89698            2.07506
Income                -0.04486           -0.04486           -0.04486
```

Fitted model:

$$\text{logit}[\widehat{P}(Y \le j\,|x)] = \log\left(\frac{P(Y \le j)}{P(Y > j)}\right)$$

$$= \begin{cases} -2.583 - 0.045x, & \text{for } j = 1 \text{ (Dissat)} \\ -0.897 - 0.045x, & \text{for } j = 2 \text{ (Dissat or little)} \\ 2.075 - 0.045x, & \text{for } j = 3 \text{ (Dissat or little or mod)} \end{cases}$$

Interpretation of Coefficients

Interpretation of β in the model

$$\text{logit}[\widehat{\mathrm{P}}(Y \leq j \,|x)] = \log\left(\frac{\mathrm{P}(Y \leq j)}{\mathrm{P}(Y > j)}\right) = \alpha_j + \beta x$$

$$\iff \frac{\mathrm{P}(Y \leq j)}{\mathrm{P}(Y > j)} = \exp(\alpha_j + \beta x)$$

For every 1-unit increase in x, the odds of $Y \leq j$ become $\exp(\beta)$ times as large.

Example(Job Satisfaction)

$$\log\left(\frac{P(Y \le j)}{P(Y > j)}\right) = \begin{cases} -2.583 - 0.045x, & \text{for } j = 1 \text{ (Dissat)} \\ -0.897 - 0.045x, & \text{for } j = 2 \text{ (Dissat or little)} \\ 2.075 - 0.045x, & \text{for } j = 3 \text{ (Dissat or little or mod)} \end{cases}$$

Estimated odds of being

$(Y \le 1)$ "dissat"	rather than	"little", "mod" or "very sat." $(Y > 1)$
$(Y \le 2)$ "dissat" or "little"	rather than	"mod" or "very sat." $(Y > 2)$
$(Y \le 3)$ "dissat" or "little" or "mod"	rather than	"very sat." $(Y > 3)$

all become $e^{-0.045} \approx 0.96$ times as large for each 1K increase in income.

If We Reverse the Order of Response Categories . . .

If we **reverse** the **order** of response categories, β would **change sign** but has same SE.

With "very sat." < "mod" < "little" < "dissat":

```
jobsat.cl1r = vglm(cbind(Very,Mod,Little,Diss) ~ Income,
                    family=cumulative(parallel=TRUE))
coef(jobsat.cl1r)
(Intercept):1 (Intercept):2 (Intercept):3        Income
     -2.07506       0.89698       2.58287       0.04486
```

$\widehat{\beta} = 0.045$, estimated odds of satisfaction above any given level is multiplied by

$$e^{10\widehat{\beta}} = 1.566 = 1/0.64.$$

for each 10K increase in income

Probabilities of Categories for Cumulative Logit Models

$$\log\left(\frac{P(Y \le j)}{P(Y > j)}\right) = \alpha_j + \beta x \iff P(Y \le j) = \frac{e^{\alpha_j + \beta x}}{1 + e^{\alpha_j + \beta x}} \quad j = 1, 2, \ldots, J-1.$$

The probability π_j for an individual category

$$\pi_1 = P(Y = 1) = P(Y \le 1) = \frac{e^{\alpha_1 + \beta x}}{1 + e^{\alpha_1 + \beta x}}$$

$$\pi_2 = P(Y = 2) = P(Y \le 2) - P(Y \le 1) = \frac{e^{\alpha_2 + \beta x}}{1 + e^{\alpha_2 + \beta x}} - \frac{e^{\alpha_1 + \beta x}}{1 + e^{\alpha_1 + \beta x}}$$

$$\pi_3 = P(Y = 3) = P(Y \le 3) - P(Y \le 2) = \frac{e^{\alpha_3 + \beta x}}{1 + e^{\alpha_3 + \beta x}} - \frac{e^{\alpha_2 + \beta x}}{1 + e^{\alpha_2 + \beta x}}$$

$$\vdots$$

$$\pi_J = P(Y = J) = P(Y \le J) - P(Y \le J - 1)$$

$$= 1 - P(Y \le J - 1) = 1 - \frac{e^{\alpha_{J-1} + \beta x}}{1 + e^{\alpha_{J-1} + \beta x}}$$

Probabilities of Categories (Job Satisfaction)

$$\log\left(\frac{P(Y \le j)}{P(Y > j)}\right) = \begin{cases} -2.583 - 0.045x, & \text{for } j = 1 \text{ (Dissat)} \\ -0.897 - 0.045x, & \text{for } j = 2 \text{ (Dissat or little)} \\ 2.075 - 0.045x, & \text{for } j = 3 \text{ (Dissat or little or mod)} \end{cases}$$

E.g., at $x = 20$ (K), estimated prob. of the 4 categories are

$$\widehat{\pi}_{dissat} = \widehat{\pi}_1 = \frac{e^{-2.583-0.045(20)}}{1+e^{-2.583-0.045(20)}} \approx 0.03$$

$$\widehat{\pi}_{little} = \widehat{\pi}_2 = \frac{e^{-0.897-0.045(20)}}{1+e^{-0.897-0.045(20)}} - \frac{e^{-2.583-0.045(20)}}{1+e^{-2.583-0.045(20)}} \approx 0.113$$

$$\widehat{\pi}_{mod} = \widehat{\pi}_3 = \frac{e^{2.075-0.045(20)}}{1+e^{2.075-0.045(20)}} - \frac{e^{-0.897-0.045(20)}}{1+e^{-0.897-0.045(20)}} \approx 0.622$$

$$\widehat{\pi}_{verysat} = \widehat{\pi}_4 = 1 - \frac{e^{2.075-0.045(20)}}{1+e^{2.075-0.045(20)}} \approx 0.235$$

Obtaining Probabilities of Categories in R

In R, we can obtain the prob. of being "diss", "little", "moderate", or "very satisfied" when income is 20K using `predict()`.

```
predict(jobsat.cl1, data.frame(Income=20), type="response")
     Diss Little   Mod   Very
1 0.02989 0.1127 0.622 0.2354
```

We see $\widehat{\pi}_{diss} \approx 0.03$, $\widehat{\pi}_{little} \approx 0.113$, $\widehat{\pi}_{mod} \approx 0.622$, and $\widehat{\pi}_{very} \approx 0.235$. Observe that $\widehat{\pi}_{diss} + \widehat{\pi}_{little} + \widehat{\pi}_{mod} + \widehat{\pi}_{very} = 1$.

Caution: without specifying `type="response"` in `predict()`, the predicted values would be the logits $\log\left(\frac{P(Y\le j)}{P(Y>j)}\right) = \alpha_j + \beta x$, not the probabilities $\widehat{\pi}_i$.

```
predict(jobsat.cl1, data.frame(Income=20))
  logitlink(P[Y<=1]) logitlink(P[Y<=2]) logitlink(P[Y<=3])
1              -3.48             -1.794              1.178
```

Wald Tests and Wald CIs for Parameters

Wald test of H_0: $\beta = 0$ (job satisfaction indep. of income):

$$z = \frac{\widehat{\beta} - 0}{\text{SE}(\widehat{\beta})} = \frac{-0.0449}{0.0175} = -2.56, \quad (z^2 = 6.57, df = 1)$$

P-value = 0.01038

$$\underline{\text{95\% CI for } \beta} : \ \widehat{\beta} \pm 1.96\text{SE}(\widehat{\beta}) = -0.0449 \pm 1.96 \times 0.0175$$

$$= (-0.079, -0.011)$$

$$\underline{\text{95\% CI for } e^{\beta}} : \ (e^{-0.079}, e^{-0.011}) = (0.924, 0.990)$$

```
coef(summary(jobsat.cl1))
              Estimate Std. Error z value    Pr(>|z|)
(Intercept):1 -2.58287     0.5584  -4.625 0.000003740
(Intercept):2 -0.89698     0.3550  -2.527 0.011511413
(Intercept):3  2.07506     0.4158   4.990 0.000000603
Income        -0.04486     0.0175  -2.563 0.010380707
```

LR Test for Parameters

LR test of H_0: $\beta = 0$ (job satisfaction indep. of income):

$$\text{LR statistic} = -2(L_0 - L_1) = -2((-21.358) - (-18.000)) = 6.718$$

P-value $= 0.0095$

```
lrtest(jobsat.cl1)
Likelihood ratio test

Model 1: cbind(Diss, Little, Mod, Very) ~ Income
Model 2: cbind(Diss, Little, Mod, Very) ~ 1
  #Df LogLik Df Chisq Pr(>Chisq)
1   8  -18.0
2   9  -21.4  1  6.72     0.0095
```

Pearson's X^2 v.s. Baseline Category Logit v.s. Cumulative Logit

For the Income and Job Satisfaction data, we obtained stronger evidence of association if we use a cumulative logits model treating Y (Job Satisfaction) as ordinal than obtained if we treat:

- Y as nominal (baseline category logit model) and X as ordinal:

  $$\log(\pi_j/\pi_4) = \alpha_j + \beta_j x.$$

  Recall P-value = 0.032 for LR test.
- X, Y both as nominal: Pearson's test of independence had
  $X^2 = 11.5$, $df = 9$, P-value $= 0.24$
  $G^2 = 13.47$, $df = 9$, P-value $= 0.14$

Deviance and Goodness of Fit

Deviance can be used to test Goodness of Fit in the same way.

For cumulative logit model for Job Satisfaction data

$$\text{Deviance} = 6.749, \quad df = 8, \quad P\text{-value} = 0.56$$

The Model fits data well.

```
deviance(jobsat.cl1)
[1] 6.749
df.residual(jobsat.cl1)
[1] 8
pchisq(deviance(jobsat.cl1),df=8,lower.tail=F)
[1] 0.5639
```

Remark. Generally, Goodness of fit test is appropriate if most of the fitted counts are ≥ 5. which is not the case for for the Job Satisfaction data. The P-value might not be reliable.

Example (Political Ideology and Party Affiliation)

Gender	Political Party	Political Ideology				
		very liberal	slightly liberal	moderate	slightly conservative	very conservative
Female	Democrat	44	47	118	23	32
	Republican	18	28	86	39	48
Male	Democrat	36	34	53	18	23
	Republican	12	18	62	45	51

Y = political ideology $(1 =$ very liberal$, \ldots, 5 =$ very conservative$)$

G = gender $(1 = M, 0 = F)$

P = political party $(1 =$ Republican$,\ 0 =$ Democrat$)$

Cumulative Logit Model:

$$\text{logit}[\mathrm{P}(Y \leq j)] = \alpha_j + \beta_G G + \beta_P P, \quad j = 1, 2, 3, 4.$$

```r
Gender = c("F","F","M","M")
Party = c("Dem","Rep","Dem","Rep")
VLib = c(44,18,36,12)
SLib = c(47,28,34,18)
Mod  = c(118,86,53,62)
SCon = c(23,39,18,45)
VCon = c(32,48,23,51)
ideo.cl1 = vglm(cbind(VLib,SLib,Mod,SCon,VCon) ~ Gender + Party,
                family=cumulative(parallel=TRUE))
coef(ideo.cl1)
(Intercept):1 (Intercept):2 (Intercept):3 (Intercept):4       GenderM
      -1.4518       -0.4583        1.2550        2.0890       -0.1169
     PartyRep
      -0.9636
```

Fitted Model:

$$\text{logit}[\widehat{\text{P}}(Y \leq j)] = \widehat{\alpha}_j - 0.1169G - 0.9636P, \qquad j = 1, 2, 3, 4.$$

```
coef(summary(ideo.cl1))
              Estimate Std. Error  z value  Pr(>|z|)
(Intercept):1  -1.4518     0.1228 -11.8182 3.144e-32
(Intercept):2  -0.4583     0.1058  -4.3334 1.468e-05
(Intercept):3   1.2550     0.1145  10.9560 6.220e-28
(Intercept):4   2.0890     0.1292  16.1737 7.726e-59
GenderM        -0.1169     0.1268  -0.9215 3.568e-01
PartyRep       -0.9636     0.1294  -7.4492 9.393e-14
```

- Controlling for Gender, estimated odds that a Republican is in liberal direction ($Y \leq j$) rather than conservative ($Y > j$) are

$$e^{\widehat{\beta}_P} = e^{-0.9636} \approx 0.38$$

  times the estimated odds for a Democrat, for all $j = 1, 2, 3, 4$.

- 95% Wald CI for e^{β_P} is

$$e^{\widehat{\beta}_P \pm 1.96\,\mathrm{SE}(\widehat{\beta}_P)} \approx e^{-0.9636 \pm 1.96 \times 0.1294} \approx (0.30, 0.49)$$

- Based on Wald test, Party effect is significant (controlling for Gender) but Gender is not significant (controlling for Party).

LR test of H_0: $\beta_G = 0$ (no Gender effect, given Party):

```
lrtest(ideo.cl1, "Gender")     # LR test for Gender effect
Likelihood ratio test

Model 1: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender + Party
Model 2: cbind(VLib, SLib, Mod, SCon, VCon) ~ Party
  #Df LogLik Df Chisq Pr(>Chisq)
1  10  -47.4
2  11  -47.8  1  0.84       0.36
```

LR test of H_0: $\beta_P = 0$ (no Party effect, given Gender):

```
lrtest(ideo.cl1, "Party")     # LR test for Party effect
Likelihood ratio test

Model 1: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender + Party
Model 2: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender
  #Df LogLik Df Chisq Pr(>Chisq)
1  10  -47.4
2  11  -75.8  1  56.9    4.7e-14
```

Interaction?

Model w/ `Gender*Party` interaction:

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_G G + \beta_P P + \beta_{GP} G * P, \quad j = 1, 2, 3, 4.$$

For H_0: $\beta_{GP} = 0$, LR statistic = 3.99, df = 1, P-value = 0.046

⇒ Evidence of Party effect depends on Gender (and vice versa)

```
ideo.cl2 = vglm(cbind(VLib,SLib,Mod,SCon,VCon) ~ Gender * Party ,
                 family=cumulative(parallel=TRUE))
lrtest(ideo.cl2,ideo.cl1)
Likelihood ratio test

Model 1: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender * Party
Model 2: cbind(VLib, SLib, Mod, SCon, VCon) ~ Gender + Party
  #Df LogLik Df Chisq Pr(>Chisq)
1   9  -45.4
2  10  -47.4  1  3.99      0.046
```

$$\text{logit}[\mathrm{P}(Y \le j)] = \alpha_j + \beta_G G + \beta_P P + \beta_{GP} GP$$

$$\Leftrightarrow \text{odds of } (Y \le j) = \frac{\mathrm{P}(Y \le j)}{\mathrm{P}(Y > j)} = \exp(\alpha_j + \beta_G G + \beta_P P + \beta_{GP} GP)$$

Odds ratio for Party effect is

$$\frac{\text{odds of } Y \le j \text{ given } P = 1, G}{\text{odds of } Y \le j \text{ given } P = 0, G} = \frac{\exp(\alpha_j + \beta_G G + \beta_P(1) + \beta_{GP} G(1))}{\exp(\alpha_j + \beta_G G + \beta_P(0) + \beta_{GP} G(0))}$$

$$= \begin{cases} \exp(\beta_P) & \text{for females } (G = 0) \\ \exp(\beta_P + \beta_{GP}) & \text{for males } (G = 1) \end{cases}$$

```
coef(ideo.cl2)
   (Intercept):1    (Intercept):2    (Intercept):3    (Intercept):4
         -1.5521          -0.5550           1.1647           2.0012
         GenderM         PartyRep GenderM:PartyRep
          0.1431          -0.7562          -0.5091
```

Fitted model w/ `Gender*Party` interaction:

$$\text{logit}[\widehat{\text{P}}(Y \le j)] = \widehat{\alpha}_j + 0.143G - 0.756P - 0.509G * P, \quad j = 1, 2, 3, 4.$$

Estimated odds ratio for Party effect is

$$\begin{cases} e^{-0.756} = 0.47 & \text{for females } (G = 0) \\ e^{-0.756-0.51} = e^{-1.266} = 0.28 & \text{for males } (G = 1) \end{cases}$$

Greater discrepancies between male Dem. and male Rep. than between female Dem. and female Rep.

Observed Percentages

		Political Ideology				
Gender	Party	Very Liberal	Slightly Liberal	Moderate	Slightly Conserve.	Very Conserve.
Female	Dem.	17%	18%	45%	9%	12%
	Rep.	8%	13%	39%	18%	22%
Male	Dem.	22%	21%	32%	11%	14%
	Rep.	6%	10%	33%	24%	27%

Fitted model w/ Gender×Party interaction:

$$\text{logit}[\widehat{\text{P}}(Y \leq j)] = \widehat{\alpha}_j + 0.143G - 0.756P - 0.509G * P, \quad j = 1, 2, 3, 4.$$

Estimated odds ratio for Gender effect is

$$\begin{cases} e^{0.143} = 1.15 & \text{for Dems } (P = 0) \\ e^{0.143-0.51} = e^{-0.336} = 0.69 & \text{for Reps } (P = 1) \end{cases}$$

- Among Dems, males were more liberal than females.
- Among Reps, males were more conservative than females.

Goodness of Fit

$$\text{Deviance} = 11.063, \quad df = 9 \text{ (why?)}, \quad P\text{-value} = 0.2714$$

The cumulative logits model w/ interaction fits data well.

```
deviance(ideo.cl2)
[1] 11.06
df.residual(ideo.cl2)
[1] 9
pchisq(deviance(ideo.cl2), df= 9, lower.tail=FALSE)
[1] 0.2714
```