Multiple Logistic Regression

Response: Y binary, $\pi = \mathrm{P}(Y = 1)$

Explanatory variables: $x_1, x_2, \ldots, x_k$
can be quantitative, qualitative (dummy variables), or both.

Model form is

$$\text{logit}(\pi) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

or equivalently

$$\pi = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}$$

β_i = partial effect of x_i controlling for other variables in model

e^{β_i} = conditional odds ratio at $x_i + 1$ vs at x_i **keeping other x's fixed**
= multiplicative effect on odds of 1-unit increase in x_i
w/ other x's fixed

Example (Horseshoe Crabs)

In addition to `Width` (X), consider adding a categorical predictor — `Color`, coded 1-4 as

1 = medium light, 2 = medium, 3 = medium dark, 4 = dark

For a categorical predictor, need to create a **dummy variable** (= indicator variable) for each category:

$$c_1 = \begin{cases} 1 & \text{medium light} \\ 0 & \text{o/w} \end{cases}, \quad c_2 = \begin{cases} 1 & \text{medium} \\ 0 & \text{o/w} \end{cases},$$

$$c_3 = \begin{cases} 1 & \text{medium dark} \\ 0 & \text{o/w} \end{cases}, \quad c_4 = \begin{cases} 1 & \text{dark} \\ 0 & \text{o/w} \end{cases}$$

$$\boxed{\text{Model: } \text{logit}(\pi) = \alpha + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x}$$

- $c_1 + c_2 + c_3 + c_4 = 1$ always true, so one of them is redundant.
- To account for redundancies, most software set one of $\beta_1, \beta_2, \beta_3, \beta_4$ to 0

Example (Horseshoe Crabs)

Model 1:

$$\text{logit}(\pi) = \alpha + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x$$
$$= \begin{cases} \alpha + \beta x & \text{if med. light } (c_2 = c_3 = c_4 = 0) \\ \alpha + \beta_2 + \beta x & \text{if medium } (c_2 = 1, c_3 = c_4 = 0) \\ \alpha + \beta_3 + \beta x & \text{if med. dark } (c_2 = 0, c_3 = 1, c_4 = 0) \\ \alpha + \beta_4 + \beta x & \text{if dark } (c_2 = c_3 = 0, c_4 = 1) \end{cases}$$

- Here we set $\beta_1 = 0$
- The category with no dummy var. in the model (or with coefficient $\beta_i = 0$) is called the baseline category. In Model 1, the baseline category is the color medium light (Color = 1).

Below "odds" = odds having at least one satellite

$$\text{odds} = \frac{\pi}{1-\pi} = e^{\alpha + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x}$$
$$= \begin{cases} e^{\alpha + \beta x} & \text{if med. light } (c_2 = c_3 = c_4 = 0) \\ e^{\alpha + \beta_2 + \beta x} & \text{if medium } (c_2 = 1, c_3 = c_4 = 0) \\ e^{\alpha + \beta_3 + \beta x} & \text{if med. dark } (c_2 = 0, c_3 = 1, c_4 = 0) \\ e^{\alpha + \beta_4 + \beta x} & \text{if dark } (c_2 = c_3 = 0, c_4 = 1) \end{cases}$$

For female crabs of the same width,

$$\frac{\text{odds for a medium crab}}{\text{odds for a medium light crab}} = \frac{e^{\alpha + \beta_2 + \beta x}}{e^{\alpha + \beta x}} = e^{\beta_2}$$

- Likewise,
  - e^{β_3} = odds ratio of (med. dark v.s. med. light)
  - e^{β_4} = odds ratio of (dark v.s. med. light)
- Observe e^{β_i}'s are odds ratios of a category v.s. the baseline category (medium light), for crabs of the same width.
- Observe the effect of `Color` does not change with `Width`

Example (Horseshoe Crabs)

$$\text{Model 1:} \quad \text{odds} = \frac{\pi}{1-\pi} = e^{\alpha+\beta_2 c_2+\beta_3 c_3+\beta_4 c_4+\beta x}$$

For female crabs of same color but different width x_1, x_2,

$$\frac{\text{odds for crabs of Width } x_1}{\text{odds for crabs of Width } x_2} = \frac{e^{\alpha+\beta_2 c_2+\beta_3 c_3+\beta_4 c_4+\beta x_1}}{e^{\alpha+\beta_2 c_2+\beta_3 c_3+\beta_4 c_4+\beta x_2}} = e^{\beta(x_1-x_2)}$$

$\Rightarrow$ Width have the same effect for all colors.

As neither the effect of color change with width,
nor the effect of width change with color,
we said Model 1 assumes **no interaction** between color and width effects.

R regards `Color` (coded 1-4) as a numeric variable.
The R command `as.factor()` can create the dummy variables.

```
> C = as.factor(Color)
> crabs.fit1 = glm(has.sate ~ C + Weight, family = binomial)
> crabs.fit1$coef
 (Intercept)           C2           C3           C4        Width
-11.38519276   0.07241694  -0.22379766  -1.32991913   0.46795598
```

The fitted model is

$$\text{logit}(\widehat{\pi}) = -11.39 + 0.07c_2 - 0.22c_3 - 1.33c_4 + 0.468x$$

For a medium light female ($c_2 = c_3 = c_4 = 0$) of width $x = 25$ cm,

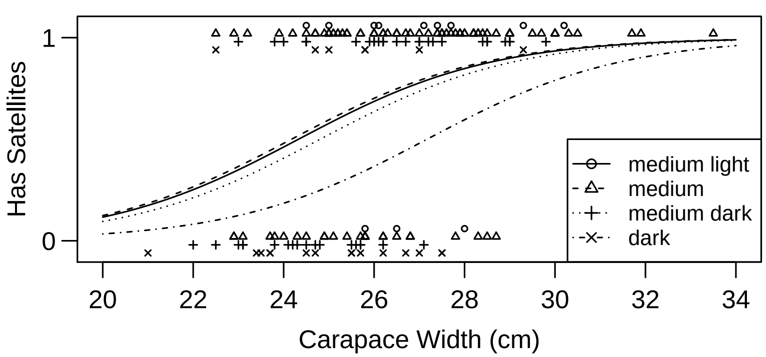
$$\widehat{\pi} = \frac{\exp(-11.39 + 0.468 \times 25)}{1 + \exp(-11.39 + 0.468 \times 25)} \approx 0.58$$

For a dark female ($c_2 = c_3 = 0, c_4 = 1$) of width $x = 25$ cm,

$$\widehat{\pi} = \frac{\exp(-11.39 + (-1.33)(1) + 0.468 \times 25)}{1 + \exp(-11.39 + (-1.33)(1) + 0.468 \times 25)} \approx 0.265.$$

$$\text{logit}(\widehat{\pi}) = -11.39 + 0.07c_2 - 0.22c_3 - 1.33c_4 + 0.468x$$
$$= \begin{cases} -11.39 + 0.468x & \text{if medium light} \\ -11.32 + 0.468x & \text{if medium} \\ -11.61 + 0.468x & \text{if medium dark} \\ -12.72 + 0.468x & \text{if dark} \end{cases}$$

Observe the four curves have the same shape because they have identical coefficient for Width.



Medium v.s. Medium Light Crabs

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -11.38519    2.87346  -3.962 7.43e-05 ***
C2            0.07242    0.73989   0.098    0.922
C3           -0.22380    0.77708  -0.288    0.773
C4           -1.32992    0.85252  -1.560    0.119
Width         0.46796    0.10554   4.434 9.26e-06 ***
```

- Interpretation of β_2: estimated odds of having satellite(s) for medium crabs are $e^{\widehat{\beta}_2} = e^{0.07} \approx 1.07$ times the estimated odds for medium light crabs of the same width.
- $H_0: \ \beta_2 = 0$ means medium and medium light crabs do not differ in their chance of having satellite(s) given width. To test

  $$H_0 : \beta_2 = 0 \quad \text{v.s.} \quad H_a : \beta_2 \neq 0$$

  $$\text{Wald statistic } z = \frac{\widehat{\beta}_2}{SE} = \frac{0.072}{0.74} = 0.098, \ P\text{-value} = 0.922.$$

  Conclusion: Medium light and medium crabs of the same width don't differ significantly in the prob. of having satellites.

95% LR CI for β_2 is $(-1.54, 1.45)$, which contains 0.
So LR test also fail to reject H_0: $\beta_2 = 0$.

```
> confint(crabs.fit1,test="Chisq")
                 2.5 %     97.5 %
(Intercept) -17.3084388 -5.9859523
C2           -1.5396596  1.4516138
C3           -1.8918959  1.2396603
C4           -3.1356611  0.2737758
Width         0.2712817  0.6870436
```

What about (medium dark v.s. medium light) crabs?
What about (dark v.s. medium light) crabs?

What about Medium v.s. Dark Crabs?

For medium and dark crabs of the same width, the odds ratio is

$$\frac{\text{odds for a medium crab}}{\text{odds for a dark crab}} = \frac{e^{\alpha+\beta_2+\beta x}}{e^{\alpha+\beta_4+\beta x}} = e^{\beta_2-\beta_4}.$$

Estimated odds of having satellite(s) for a medium crab is

$$e^{\widehat{\beta}_2-\widehat{\beta}_4} = e^{0.07-(-1.33)} = e^{1.4} \approx 4.06$$

times the estimated odds for a dark crabs of the same width.

However, to test $H_0 : \beta_2 = \beta_4$, need SE for $\widehat{\beta}_2 - \widehat{\beta}_4$, which is not provided in R.

The simplest solution is to change the baseline category. Say, use dark color as the baseline and model as

$$\text{Model 1a}: \ \text{logit}(\pi) = \alpha' + \beta'_1 c_1 + \beta'_2 c_2 + \beta'_3 c_3 + \beta x$$

Change of Baseline

$$\begin{aligned} \text{Model 1}: \ \text{logit}(\pi) &= \alpha \qquad\qquad + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x \\ \text{Model 1a}: \ \text{logit}(\pi) &= \alpha' + \beta'_1 c_1 + \beta'_2 c_2 + \beta'_3 c_3 \qquad\qquad + \beta x \end{aligned}$$

		logit(π) for	
Color	(c_1, c_2, c_3, c_4)	Model 1	Model 1a
med. light	$(1,0,0,0)$	$\alpha \quad + \beta x$	$\alpha' + \beta'_1 + \beta x$
medium	$(0,1,0,0)$	$\alpha + \beta_2 + \beta x$	$\alpha' + \beta'_2 + \beta x$
med. dark	$(0,0,1,0)$	$\alpha + \beta_3 + \beta x$	$\alpha' + \beta'_3 + \beta x$
dark	$(0,0,0,1)$	$\alpha + \beta_4 + \beta x$	$\alpha' \quad + \beta x$

The two models are equivalent, just a change of parameters.

$$\alpha' = \alpha + \beta_4, \quad \beta'_i = \beta_i - \beta_4 \quad \text{for } i = 1, 2, 3$$

Testing $\beta_2 = \beta_4$ in Model 1 is equivalent to testing $\beta'_2 = 0$ in Model 1a.


```
> C1 = as.numeric(Color==1)
> C2 = as.numeric(Color==2)
> C3 = as.numeric(Color==3)
> crabs.fit1a = glm(has.sate ~ C1+C2+C3 + Width, family = binomial)
> summary(crabs.fit1a)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.7151     2.7617  -4.604 4.14e-06 ***
C1            1.3299     0.8525   1.560   0.1188
C2            1.4023     0.5484   2.557   0.0106 *
C3            1.1061     0.5921   1.868   0.0617 .
Width         0.4680     0.1055   4.434 9.26e-06 ***
```

- $\widehat{\beta}'_2 = 1.4023$, which is equal to $\widehat{\beta}_2 - \widehat{\beta}_4$
- Wald test of H_0: $\beta'_2 = 0$ gives P-value 0.0106

Conclusion: Medium and dark crabs of the same width differ significantly in the prob. of having satellites.

```
> drop1(crabs.fit1a,test="Chisq")
Single term deletions
Model:
has.sate ~ C1 + C2 + C3 + Width
       Df Deviance    AIC     LRT  Pr(>Chi)
<none>      187.46 197.46
C1      1   190.07 198.07  2.6154  0.105831
C2      1   194.37 202.37  6.9101  0.008571 **
C3      1   191.11 199.11  3.6518  0.056010 .
Width   1   212.06 220.06 24.6038 7.041e-07 ***
```

LR test of $\beta_2' = 0$ gives P-value 0.0086, same conclusion as Wald test

```
> confint(crabs.fit1a)
Waiting for profiling to be done...
                  2.5 %     97.5 %
(Intercept) -18.45674069 -7.5788795
C1           -0.27377584  3.1356611
C2            0.35269965  2.5260703
C3           -0.02792233  2.3138635
Width         0.27128167  0.6870436
```

95% for β_2' is $(0.353, 2.526) \Longrightarrow$ estimated odds for medium crabs are at least $e^{0.353} \approx 1.42$, at most $e^{2.526} \approx 12.5$ times the est. odds for dark crabs of the same width.

Likelihood Ratio Test for Model Comparison

Likelihood Ratio Test can be used to do model comparison between a *simpler model* and a *more complex model*.

- The simpler model must be **a special case** of the more complex model.
  If not, CANNOT use LRT to do model comparison
- H_0: the simpler model is correct
  H_a: the complex model is correct, the simpler model is not
- Rejecting H_0 means the simpler model doesn't fit the data well, compared to the more complex model
- Not rejecting H_0 means the simpler model fits the data nearly as well as the more complex model

Likelihood Ratio Test for Model Comparison

- Likelihood ratio (LR) statistic $= -2(L_0 - L_1)$, where
  $L_0 =$ max. log-likelihood for the simpler model,
  $L_1 =$ max. log-likelihood for the complex model
- In general, $L_0 \leq L_1$. Under H_0, $L_0 \approx L_1$.
- Large sample distribution of LR statistic is **Chi-squared** with

  $$d.f. = \text{diff. in number of parameters for the 2 models}$$

Likelihood Ratio Test for Model Comparison

Rather than reporting the max. log-likelihood for a model, R reports

$$\text{Deviance} = -2(\text{max. log-likelihood} + C)$$

in which C is a constant depends only on the data but not the model. So

$$\begin{aligned} \text{LR statistic} &= -2(L_0 - L_1) \\ &= -2(L_0 + C) - [-2(L_1 + C)] \\ &= \text{diff. in deviance for the two models} \end{aligned}$$

- We will introduce deviance in Chapter 5
- d.f. for a deviance is
  $$(\text{num. of observations}) - (\text{num. of parameters})$$
- so d.f. for a LR statistic = diff. in d.f. for the two deviances
- LR test for model comparison is also called "*analysis of deviance*"

```
> summary(crabs.fit1)
Call:
glm(formula = has.sate ~ C + Width, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1124  -0.9848   0.5243   0.8513   2.1413

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -11.38519    2.87346  -3.962 7.43e-05 ***
C2            0.07242    0.73989   0.098    0.922
C3           -0.22380    0.77708  -0.288    0.773
C4           -1.32992    0.85252  -1.560    0.119
Width         0.46796    0.10554   4.434 9.26e-06 ***
---
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 187.46  on 168  degrees of freedom
AIC: 197.46

Number of Fisher Scoring iterations: 4
```

For Model 1, deviance $= 187.46$ with d.f. $= 173 - 5 = 168$
($n = 173$ for horseshoe crabs data)

Example (Horseshoe Crabs)

Do We Need Color in the Model?

$$H_0: \ \beta_2 = \beta_3 = \beta_4 = 0 \quad \text{(given width, } Y \text{ indep. of color)}$$

i.e.,

$$H_0 : \text{logit}(\pi) = \alpha + \beta x \qquad \text{(simpler model)}$$
$$H_a : \text{logit}(\pi) = \alpha + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x \quad \text{(complex model)}$$

```
> anova(crabs.logit, crabs.fit1, test="Chisq")
Analysis of Deviance Table

Model 1: has.sate ~ Width
Model 2: has.sate ~ C + Width
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       171     194.45
2       168     187.46  3   6.9956  0.07204 .
```

The LR statistic = diff. of deviance $= 194.45 - 187.46 = 6.99$
with $df = 171 - 168 = 3$, P-value$= 0.072$
$\Longrightarrow$ Some evidence (not strong) of a color effect given width.

R command `drop1` on a model performs LRT comparing

H_0 : the model w/ one term deleted
H_a : the model itself

for each term in the model, e.g., the P-value for for `Width` in the R output below is LRT for comparing

$$H_0 : \text{logit}(\pi) = \alpha + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4$$
$$H_a : \text{logit}(\pi) = \alpha + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x$$

```
> drop1(crabs.fit1, test="Chisq")
Single term deletions

Model:
has.sate ~ C + Width
       Df Deviance    AIC     LRT  Pr(>Chi)
<none>      187.46 197.46
C       3   194.45 198.45  6.9956   0.07204 .
Width   1   212.06 220.06 24.6038 7.041e-07 ***
```

Some evidence (not strong) of a color effect given width.
There is strong evidence of width effect.

Other simpler models might be adequate.
Plot of the four curves on Slide 9 suggests that maybe only **dark** crabs are different from others.

$$\text{Model 2: } \text{logit}(\pi) = \alpha + \beta_4 c_4 + \beta x, \quad \text{where } c_4 = \begin{cases} 1 & \text{dark} \\ 0 & \text{o/w} \end{cases}$$

Fitting gives $\widehat{\beta}_4 = -1.300$ (SE $= 0.5259$).
Odds of satellites for a dark crab is estimated to be $e^{-1.300} = 0.27$ times the odds a non-dark crab of the same width.

```
> crabs.fit2 = glm(has.sate ~ I(Color==4) + Width, family = binomial)
> summary(crabs.fit2)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -11.6790     2.6925  -4.338 1.44e-05 ***
I(Color == 4)TRUE  -1.3005     0.5259  -2.473   0.0134 *
Width               0.4782     0.1041   4.592 4.39e-06 ***
```

Compare model with 1 dummy for color to full model with 3 dummies.

H$_0$: $\text{logit}(\pi) = \alpha + \beta_4 c_4 + \beta x$ (simple model)

H$_a$: $\text{logit}(\pi) = \alpha + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x$ (more complex model)

Note H$_0$ is $\beta_2 = \beta_3 = 0$ in more complex model.

```
> anova(crabs.fit2, crabs.fit1, test="Chisq")
Analysis of Deviance Table

Model 1: has.sate ~ I(Color == 4) + Width
Model 2: has.sate ~ C + Width
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       170     187.96
2       168     187.46  2  0.50085   0.7785
```

LR stat = diff. in deviances = $187.96 - 187.45 = 0.50$
$df = 170 - 168 = 2$, P-value = 0.7785

Simpler model is adequate.

Ordinal Factors

- Color of horseshoe crabs is ordinal (from light to dark). Models with dummy variables treat color as nominal.
- To treat as quantitative, assign scores such as (1,2,3,4) and model trend.

Model 3: $\text{logit}(\pi) = \alpha + \gamma c + \beta x$, c: color, x: width

```
> crabs.fit3 = glm(has.sate ~ Color + Width, family = binomial)
> summary(crabs.fit3)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.0708     2.8068  -3.588 0.000333 ***
Color        -0.5090     0.2237  -2.276 0.022860 *
Width         0.4583     0.1040   4.406 1.05e-05 ***
```

The fitted model is $\text{logit}(\pi) = -10.071 - 0.509c + 0.458x$.

Controlling for width, odds of having satellite(s) is estimated to decrease by a factor of $e^{\widehat{\gamma}} = e^{-0.509} = 0.601$ for each 1-category increase in shell darkness.

Does model treating color as nominal fit as well as model treating it as qualitative?

H$_0$: $\text{logit}(\pi) = \alpha + \gamma c + \beta x$ (simpler (ordinal) model)

H$_a$: $\text{logit}(\pi) = \alpha + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x$ (more complex model)

```
> anova(crabs.fit3, crabs.fit1, test="Chisq")
Analysis of Deviance Table

Model 1: has.sate ~ Color + Width
Model 2: has.sate ~ C + Width
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       170     189.12
2       168     187.46  2   1.6641   0.4351
```

LR stat = diff. in deviances = $189.12 - 187.46 = 1.66$
$df = 170 - 168 = 2$, P-value = 0.4351

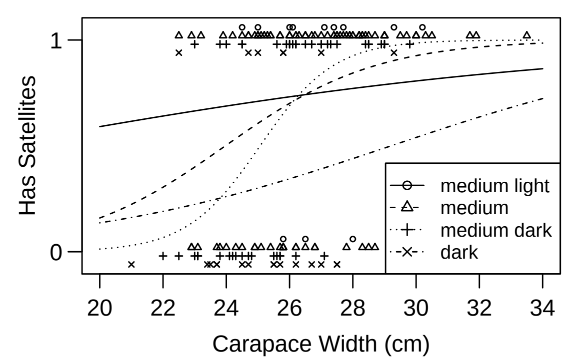
Simpler model is adequate.

Models Allowing Interactions

$$\text{logit}(\pi) = \alpha + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4 + \beta x + \gamma_2 c_2 x + \gamma_3 c_3 x + \gamma_4 c_4 x$$

$$= \begin{cases} \alpha + \beta x & \text{if medium light} \\ \alpha + \beta_2 + (\beta + \gamma_2)x & \text{if medium} \\ \alpha + \beta_3 + (\beta + \gamma_3)x & \text{if medium dark} \\ \alpha + \beta_4 + (\beta + \gamma_4)x & \text{if dark} \end{cases}$$

Different colors have different coefficient for "Width."



```
> crabs.fit4 = glm(has.sate ~ C + Width + C:Width, family = binomial)
> summary(crabs.fit4)
Call:
glm(formula = has.sate ~ C + Width + C:Width, family = binomial)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.75261   11.46409  -0.153    0.878
C2           -8.28735   12.00363  -0.690    0.490
C3          -19.76545   13.34251  -1.481    0.139
C4           -4.10122   13.27532  -0.309    0.757
Width         0.10600    0.42656   0.248    0.804
C2:Width      0.31287    0.44794   0.698    0.485
C3:Width      0.75237    0.50435   1.492    0.136
C4:Width      0.09443    0.50042   0.189    0.850

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 183.08  on 165  degrees of freedom
```


Testing H_0: no interaction ($\gamma_2 = \gamma_3 = \gamma_4 = 0$)

```
> anova(crabs.fit1,crabs.fit4,test="Chisq")
Analysis of Deviance Table

Model 1: has.sate ~ C + Width
Model 2: has.sate ~ C + Width + C:Width
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       168     187.46
2       165     183.08  3   4.3764   0.2236
```

LR stat = diff. in deviances $= 187.46 - 183.08 = 4.3764$
$df = 168 - 165 = 3$, P-value $= 0.2236$

Simpler model is adequate (no interaction).