STAT 226 Lecture 10-11

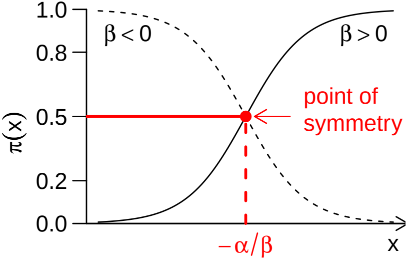
Section 4.1-4.2 Simple Logistic Regression

Yibi Huang

Simple Logistic Regression

Simple logistic regression has a single explanatory variable x and models the success probability $\pi(x)$ for the binomial response as

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}.$$



- If $\beta = 0$, then $\pi(x) = \dfrac{e^{\alpha}}{1 + e^{\alpha}}$ doesn't change with x
- bigger $|\beta|$, steeper curve
- **point of symmetry**:

$$\pi(x) = \frac{1}{2} \iff e^{\alpha+\beta x} = 1 = e^{0}$$
$$\iff \alpha + \beta x = 0 \iff \boxed{x = -\frac{\alpha}{\beta}}.$$

Example: Horseshoe Crabs

- See Section 3.3.3 and 4.1.3 for data info
- 5-min horseshoe crabs video: http://y2u.be/iYvWssvg1YU
- You can load the data by running the R command below

```
crabs = read.table(
  "https://www.stat.uchicago.edu/~yibi/s226/horseshoecrabs.txt",
  header=TRUE
)
```

```
    Color Spine Width Weight Satellites
1       2     3  28.3  3.050          8
2       3     3  22.5  1.550          0
3       1     1  26.0  2.300          9
4       3     3  24.8  2.100          0
5       3     3  26.0  2.600          4
... (omitted) ...
173     2     2  24.5  2.000          0
```

Variables of the Horseshoe Crabs Data

```
    Color Spine Width Weight Satellites
1       2     3  28.3  3.050          8
2       3     3  22.5  1.550          0
3       1     1  26.0  2.300          9
... (omitted) ...
173     2     2  24.5  2.000          0
```

One case (one row) is data for one female horseshoe crab

- `Satellites`: number of satellites (males) cling to a female
- `Width`: shell width (cm);
- `Weight`: weight in kg;
- `Color` (1 = medium light; 2 = medium; 3 = medium dark; 4 = dark);
- `Spine`: spine condition (1, both good; 2, one broken; 3, both broken);

Example: Horseshoe Crabs

$$Y = \begin{cases} 1 & \text{if female crab has satellite(s)} \\ 0 & \text{if no satellites} \end{cases}$$

$$X = \text{carapace width (cm) of female crab}$$

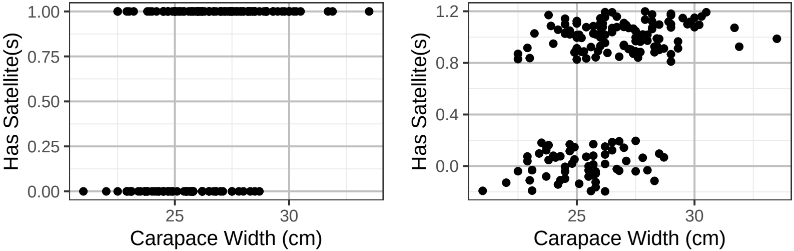
```r
crabs$has.sate = as.numeric(crabs$Satellites>0)
crabs.logit = glm(has.sate ~ Width, family = binomial, data=crabs)
```

If not specified, R uses the `logit` link by default.

```r
crabs.logit$coef
(Intercept)       Width
   -12.3508      0.4972
```

The fitted model is $\widehat{\pi}(x) = \dfrac{e^{-12.351+0.497x}}{1+e^{-12.351+0.497x}}$.

```r
library(ggplot2)
ggplot(crabs, aes(x=Width, y=has.sate)) + geom_point() +
  labs(x="Carapace Width (cm)", y="Has Satellite(s)")
ggplot(crabs, aes(x=Width, y=jitter(has.sate))) + geom_point() +
  labs(x="Carapace Width (cm)", y="Has Satellite(s)")
```



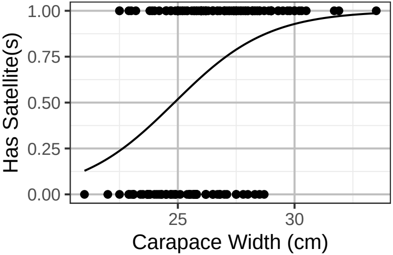
There are multiple observations (crabs) at same points (left plot).

To see them, we can **jitter** their Y values by adding a small amount of noise (right plot).

Adding the Fitted Logistic Curve (1)

One can manually add the fitted logistic curve $\widehat{\pi}(x) = \dfrac{e^{-12.351+0.497x}}{1 + e^{-12.351+0.497x}}$ using geom_function().

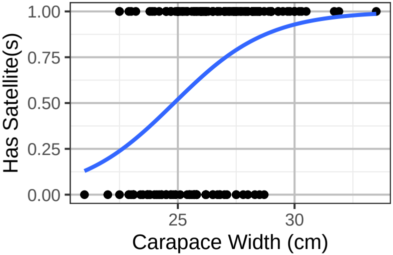
```
ggplot(crabs, aes(x=Width, y=has.sate)) + geom_point() +
  labs(x="Carapace Width (cm)", y="Has Satellite(s)") +
  geom_function(fun = function(x){
    exp(-12.351+0.497*x)/(1+exp(-12.351+0.497*x))
    })
```



Adding the Fitted Logistic Curve (2)

Alternatively, one can add the fitted logistic curve using `geom_smooth()`.

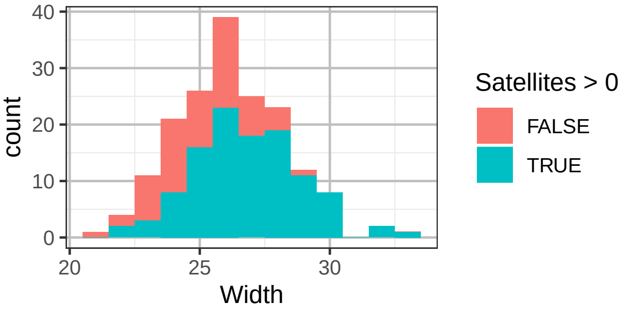
```
ggplot(crabs, aes(x=Width, y=has.sate)) + geom_point() +
  labs(x="Carapace Width (cm)", y="Has Satellite(s)") +
  geom_smooth(method='glm',method.args= list(family="binomial"), se=F)
```



It’s hard to visually assess how well the curve fits the data.

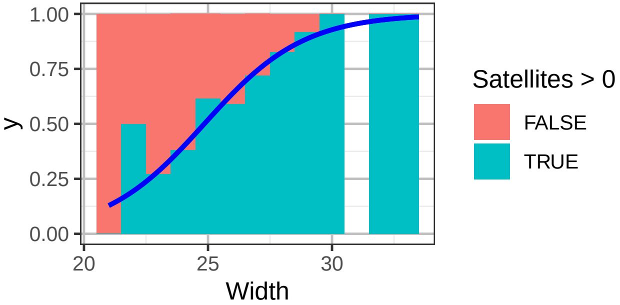
To better access the fit visually, one can group crabs of similar width and compute sample proportions for each group.

```
crabs$wd.grp = cut(crabs$Width, breaks= 21:34-0.5)
wd.table = xtabs(~wd.grp+ (Satellites > 0), data=crabs)
wd.table
             Satellites > 0
wd.grp        FALSE TRUE
  (20.5,21.5]     1    0
  (21.5,22.5]     2    2
  (22.5,23.5]     8    3
  (23.5,24.5]    13    8
  (24.5,25.5]    10   16
  (25.5,26.5]    16   23
  (26.5,27.5]     7   18
  (27.5,28.5]     4   19
  (28.5,29.5]     1   11
  (29.5,30.5]     0    8
  (30.5,31.5]     0    0
  (31.5,32.5]     0    2
  (32.5,33.5]     0    1
ggplot(crabs, aes(x=Width, fill=Satellites>0)) +
  geom_histogram(breaks=21:34-0.5)
```



estimated $\widehat{\pi}(x)$ based on the proportion of females of width in each interval that has at lease one satellite.

```
prop.table(wd.table,1)
             Satellites > 0
wd.grp            FALSE    TRUE
  (20.5,21.5] 1.00000 0.00000
  (21.5,22.5] 0.50000 0.50000
  (22.5,23.5] 0.72727 0.27273
  (23.5,24.5] 0.61905 0.38095
  (24.5,25.5] 0.38462 0.61538
  (25.5,26.5] 0.41026 0.58974
  (26.5,27.5] 0.28000 0.72000
  (27.5,28.5] 0.17391 0.82609
  (28.5,29.5] 0.08333 0.91667
  (29.5,30.5] 0.00000 1.00000
  (30.5,31.5]
  (31.5,32.5] 0.00000 1.00000
  (32.5,33.5] 0.00000 1.00000
ggplot(crabs, aes(x=Width, fill=Satellites>0)) +
  geom_histogram(binwidth=1, position="fill") +
  geom_function(fun = function(x){
    exp(-12.351+0.497*x)/(1+exp(-12.351+0.497*x))
    },
    lwd=1,color="blue"
  )
```



4.1.2 Linear Approximation Interpretations

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}, \quad \Rightarrow \quad 1-\pi(x) = \frac{1}{1+e^{\alpha+\beta x}}$$

One can show that

$$\frac{d}{dx}\pi(x) = \frac{\beta e^{\alpha+\beta x}}{(1+e^{\alpha+\beta x})^2} = \beta\pi(x)(1-\pi(x)).$$

i.e., **the slope of $\pi(x)$ at x is** $\boxed{\beta\pi(x)(1-\pi(x))}$.

- At x with $\pi(x) = \frac{1}{2}$, slope $= \beta \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{\beta}{4}$.
- At x with $\pi(x) = 0.1$ or 0.9, slope $= \beta \cdot 0.1 \cdot 0.9 = 0.09\beta$.
- **Steepest slope** at where $\pi(x) = 1/2$,
  i.e., **at point of symmetry** $x = -\frac{\alpha}{\beta}$.
- If x increases by Δx, then π increases by $\approx \beta\pi(1-\pi)\Delta x$.

Fitted Model:

$$\widehat{\pi}(x) = \frac{\exp(\widehat{\alpha} + \widehat{\beta}x)}{1 + \exp(\widehat{\alpha} + \widehat{\beta}x)} = \frac{\exp(-12.351 + 0.497x)}{1 + \exp(-12.351 + 0.497x)}$$

- $\widehat{\beta} = 0.497 > 0$, so $\widehat{\pi}(x)$ increases as Width (x) increases
- Point of symmetry:

$$\widehat{\pi}(x) = \frac{1}{2} \text{ when } x = -\frac{\widehat{\alpha}}{\widehat{\beta}} = -\frac{-12.351}{0.497} = 24.85\,\text{cm}$$

- Steepest slope at point of symmetry $x = 24.85$ cm with slope

$$\widehat{\beta}\pi(1-\pi) = 0.497 \times \frac{1}{2} \times \frac{1}{2} \approx 0.124$$

  If Width (x) increases by 1 cm, then π increases by 0.124 (actual $\widehat{\pi}$ at $x = 25.85$ is 0.623).

- At $x = 33.5$ (max. width), $\widehat{\pi} \approx 0.987$, estimated slope is

$$\widehat{\beta}\widehat{\pi}(x)(1 - \widehat{\pi}(x)) = 0.497 \cdot (0.987) \cdot (1 - 0.987) \approx 0.0064.$$

  $\Rightarrow$ Rate of change varies with x.

Predictions

The probability that an average-size female crab (w/ Width at $\bar{x} = 26.3$ cm) has satellite(s) is estimated to be

$$\widehat{\pi}(x) = \frac{e^{-12.351+0.497\times 26.3}}{1 + e^{-12.351+0.497\times 26.3}} \approx 0.67$$

R provides two kinds of predicted values.

The first one gives $\widehat{\alpha} + \widehat{\beta}x = -12.351 + 0.497 \times 26.3 \approx 0.72$.

```
predict(crabs.logit, data.frame(Width=26.3),type="link")
     1
0.7263
```

The second one gives $\widehat{\pi}(x) = \frac{\exp(\widehat{\alpha}+\widehat{\beta}x)}{1+\exp(\widehat{\alpha}+\widehat{\beta}x)}$ as computed above.

```
predict(crabs.logit, data.frame(Width=26.3),type="response")
    1
0.674
```

Remarks

Fitting linear probability model $\pi(x) = \alpha + \beta x$ (binomial w/ *identity* link) *fails* in the crabs example.

```
glm(has.sate ~ Width, family=binomial(link="identity"), data=crabs)
Error:  no valid set of coefficients has been found: please
supply starting values
```

If we pretend $Y \sim$ Normal and fit a least square regression model

$$Y = \alpha + \beta x + \varepsilon,$$

```
lm(has.sate ~ Width, data=crabs)$coef
(Intercept)       Width
   -1.76553     0.09153
```

We get the model $\widehat{Y} = -1.7655 + 0.09153x$.

At $x = 33.5$ cm, the predicted value (estimated prob. of satellites) is

$$-1.7655 + 0.09153 \times 33.5 = 1.30 \quad !?!$$

Odds Ratio Interpretation of Logistic Models

Since $\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x$, odds are

$$\text{odds} = \frac{\pi}{1-\pi} = \begin{cases} e^{\alpha+\beta x} & \text{at } x \\ e^{\alpha+\beta(x+1)} = e^{\beta} e^{\alpha+\beta x} & \text{at } x+1 \end{cases}$$

So

$$\frac{\text{odds at } (x+1)}{\text{odds at } x} = \frac{e^{\beta}\cancel{e^{\alpha+\beta x}}}{\cancel{e^{\alpha+\beta x}}} = e^{\beta}$$

More generally,

$$\frac{\text{odds at } (x+\Delta x)}{\text{odds at } x} = \frac{e^{\beta\Delta x}\cancel{e^{\alpha+\beta x}}}{\cancel{e^{\alpha+\beta x}}} = e^{\beta\Delta x}$$

If $\beta = 0$, then $e^{\beta} = 1$ and odds do not depend on x.

Example (Horseshoe Crabs)

$$\widehat{\beta} = 0.497 \quad \Longrightarrow \quad e^{\widehat{\beta}} = e^{0.497} \approx 1.64.$$

Odds of having satellite(s) are estimated to increase by a factor of 1.64 for each 1 cm increase in width.

If width increases by 0.1 cm, then odds are estimated to increase by a factor of

$$e^{(0.497)(0.1)} = e^{0.0497} = 1.051.$$

Inference for Simple Logistic Regression

- Wald tests and Wald CIs for β
- LR tests and LR CIs for β
- Confidence interval for prediction

Wald tests and Wald CIs for β

Wald Tests for β

The **Wald statistic** for testing H_0: $\beta = c$ is

$$z = \frac{\widehat{\beta} - c}{\mathrm{SE}(\widehat{\beta})} \sim N(0, 1) \quad \text{under } \mathrm{H}_0\text{: } \beta = c$$

We omit the formula for $\mathrm{SE}(\widehat{\beta})$. The value can be found in R.

Example (Horseshoe Crabs)

```
summary(crabs.logit)$coef
            Estimate Std. Error z value    Pr(>|z|)
(Intercept) -12.3508     2.6287  -4.698 0.000002622
Width         0.4972     0.1017   4.887 0.000001021
```

The column `Std.Error` gives the desired SE.

Remark: The SE of $\widehat{\beta}$ depends on the unknown true value of β. The SE in the Wald statistic is evaluated at $\beta = \widehat{\beta}$, not at the value $\beta = c$ under H_0.

```
summary(crabs.logit)$coef
            Estimate Std. Error z value    Pr(>|z|)
(Intercept) -12.3508     2.6287  -4.698 0.000002622
Width         0.4972     0.1017   4.887 0.000001021
```

R summary output gives the Wald statistics `z value` for testing H$_0$: $\beta = 0$ and the corresponding 2-sided P-values.

$$\texttt{z value} = \frac{\texttt{Estimate}}{\texttt{Std.Error}} = \frac{\widehat{\beta}}{\mathrm{SE}(\widehat{\beta})} \approx \frac{0.4972}{0.1017} \approx 4.887.$$

To test H$_0$: $\beta = 0.2$,

$$\text{Wald statistic } z = \frac{\widehat{\beta} - 0.2}{\mathrm{SE}(\widehat{\beta})} = \frac{0.4972 - 0.2}{0.1017} \approx 2.922$$

The two-sided P-value is about 0.0035.

```
2*pnorm(2.922, lower.tail=FALSE)
[1] 0.003478
```

Wald CIs for Regression Coefficients

Wald $(1-\alpha)100\%$ CIs for β are

$$\widehat{\beta} \pm z_{\alpha/2}\mathrm{SE}(\widehat{\beta}).$$

```
summary(crabs.logit)$coef
            Estimate Std. Error z value    Pr(>|z|)
(Intercept) -12.3508     2.6287  -4.698 0.000002622
Width         0.4972     0.1017   4.887 0.000001021
```

95% CI for β:

$$0.497 \pm (1.96)(0.102) = 0.497 \pm 0.200 = (0.297, 0.697)$$

95% CI for e^{β}: $(e^{0.297}, e^{0.697}) = (1.35, 2.01)$

$\Longrightarrow$ The odds that a female crab has a satellite are estimated to become 1.35 to 2.01 as large for every 1 cm increment in Width.

Wald CI for β and e^{β} in R:

R command command `confint.default()` gives the Wald CIs.

95% Wald CI for β:

```
confint.default(crabs.logit, level=0.95)
                2.5 %  97.5 %
(Intercept) -17.5030 -7.1986
Width         0.2978  0.6966
```

95% Wald CI for e^{β}:

```
exp(confint.default(crabs.logit, level=0.95))
                     2.5 %    97.5 %
(Intercept) 0.00000002503 0.0007476
Width       1.34693628236 2.0069749
```

Likelihood Ratio tests and CIs for β

Likelihood Ratio Tests for β

To test H$_0$: $\beta = 0$ vs H$_a$: $\beta \neq 0$

$$\begin{aligned} \ell_0 &= \text{ max. likelihood when } \beta = 0, \\ \ell_1 &= \text{ max. likelihood over all possible } \beta \end{aligned}$$

The likelihood ratio test statistic is

$$\begin{aligned} LRT &= -2\log\left(\ell_0/\ell_1\right) \\ &= -2\left[\log(\ell_0) - \log(\ell_1)\right] \\ &= -2(L_0 - L_1) \sim \chi^2_1 \quad \text{when sample size is large} \end{aligned}$$

where $L_i = \log(\ell_i)$.

Example (Horseshoe Crabs)

- under H$_a$: $\beta \neq 0$, $\pi(x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$, $L_1 = -97.226$
- under H$_0$: $\beta = 0$, $\pi(x) = \frac{e^{\alpha}}{1+e^{\alpha}}$, $L_0 = -112.879$

```
logLik(crabs.logit)
'log Lik.' -97.2263 (df=2)
logLik(glm(has.sate ~ 1, family = binomial, data=crabs))
'log Lik.' -112.879 (df=1)
```

$$LRT = -2(L_0 - L_1) = -2(-112.879 - (-97.226))$$

$$= 31.306, \quad \text{df} = 1,$$

$$P\text{-value} = 2.2 \times 10^{-8}$$

```
pchisq(31.306, df=1, lower.tail=FALSE)
[1] 0.0000000220397
```

Likelihood Ratio Tests for β Using drop1()

The drop1() command in R can perform LR tests for coefficients.

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

Model:
has.sate ~ Width
       Df Deviance   AIC   LRT    Pr(>Chi)
<none>        194.4 198.4
Width   1     225.8 227.8 31.31 0.000000022
```

- Observe drop1() reports LRT $= 31.3$, P-value $= 2.2 \times 10^{-8}$, agreeing with our calculation
- drop1() doesn't report the max. log-likelihood of the models, but the "*Deviances*" instead. What is "*deviances*"?

Deviance

The `summary()` output of a GLM model also reports the "*deviance*" (shown as `Residual deviance`), not the max. log-likelihood.

```
> summary(crabs.logit)
            Estimate Std. Error z value   Pr(>|z|)
(Intercept) -12.3508     2.6287  -4.698 0.00000262 ***
Width         0.4972     0.1017   4.887 0.00000102 ***
---
    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 194.45  on 171  degrees of freedom
AIC: 198.45
```

We will introduce "*deviance*" in Section 3.4.3 & 5.2. For now, just keep in mind that

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

where the $\mathit{constant}$ just depends on data but not the model.

Likelihood Ratio Statistic = Diff. in Deviance

As Deviance $= -2$(max. log-likelihood) $+ \mathit{constant}$,

$$\text{Diff. in Deviance} = \text{Deviance}_0 - \text{Deviance}_1$$
$$= -2(L_0 - \text{constant}) - [-2(L_1 - \text{constant})]$$
$$= -2(L_0 - L_1) = \text{LR statistic}$$

```r
drop1(crabs.logit, test="Chisq")
```

```
Model:
has.sate ~ Width
       Df Deviance    AIC    LRT      Pr(>Chi)
<none>      194.45 198.45
Width   1   225.76 227.76 31.306 0.00000002204
```

Deviance$_0$ $= 225.76$, Deviance$_1$ $= 194.45$

LRT = Deviance$_0$– Deviance$_1$ $= 225.76 - 194.45 = 31.31$.

Likelihood Ratio Tests v.s. Wald Tests

For very large n, Wald and LR tests are approx. equivalent, but for small to moderate n, the LR test is more reliable and powerful.

Safer to use LR CI than Wald CI

$(1-\alpha)100\%$ Likelihood Ratio (LR) CI for β is set of β^* for which P-value $> \alpha$ in LR test of H$_0$: $\beta = \beta^*$, computed by `confint()` in R.

95% Likelihood Ratio (LR) CI for β:

```
confint(crabs.logit, level=0.95)
Waiting for profiling to be done...
                 2.5 %    97.5 %
(Intercept) -17.810009 -7.457247
Width         0.308381  0.709017
```

95% Likelihood Ratio (LR) CI for e^{β}:

```
exp(confint(crabs.logit, level=0.95))
Waiting for profiling to be done...
                      2.5 %      97.5 %
(Intercept) 0.0000000184167 0.000577243
Width       1.3612190148900 2.031992299
```

For crabs example, 95% LR CI for e^{β} is $(1.36, 2.03)$.

The odds that a female crab has a satellite are estimated to become 1.36 to 2.03 as large for every 1 cm increment in Width.

(Wald) Confidence Intervals For $\pi(x)$

Prediction/Fitted Values

The estimated probability of having a satellite for a female crab with 30 cm wide carapace is

$$\widehat{\pi}(x) = \frac{e^{-12.35+0.4972\times 30}}{1+e^{-12.35-0.4972\times 30}} \approx 0.9286$$

```
predict(crabs.logit, data.frame(Width=30),type="response")
       1
0.928648
```

Caution: Without `type="response"`, `predict()` would give predicted values for $\widehat{\alpha} + \widehat{\beta}x \approx -12.35 + 0.4972 \times 30 = 2.566$ rather than for $\widehat{\pi}(x) = \frac{\exp(\widehat{\alpha}+\widehat{\beta}x)}{1+\exp(\widehat{\alpha}+\widehat{\beta}x)}$ as computed above.

```
predict(crabs.logit, data.frame(Width=30))
     1
2.5661
```

(Wald) Confidence Intervals For $\pi(x)$

To compute the Wald CI for $\pi(x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$, we first compute the CI for $\alpha + \beta x$, which is

$$\widehat{\alpha} + \widehat{\beta}x \pm z_{\alpha/2}\text{SE}(\widehat{\alpha} + \widehat{\beta}x)$$

where $\text{SE}(\widehat{\alpha} + \widehat{\beta}x)$ can be obtained by adding `se.fit=TRUE` within `predict()` with `type="link"`

```
predict(crabs.logit, data.frame(Width=30), type="link", se.fit=TRUE)
$fit
     1
2.5661

$se.fit
[1] 0.463043

$residual.scale
[1] 1
```

(Wald) Confidence Intervals For $\pi(x)$ (Cont'd)

The 95% CI for $\alpha + \beta x$ when $x = 30$ is then

$$2.566 \pm 1.96 \times 0.463 \approx (1.659, 3.474)$$

The 95% CI for $\pi(x) = \frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}$ when $x = 30$ is then

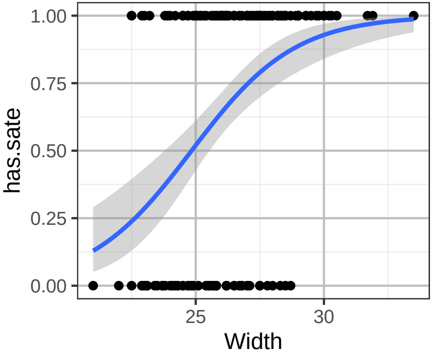
$$\left(\frac{e^{1.659}}{1 + e^{1.659}}, \frac{e^{3.474}}{1 + e^{3.474}} \right) = (0.84, 0.97)$$

- Note that the estimated $\widehat{\pi(x)} \approx 0.9286$ is not the mid-point of the 95% CI (0.84,0.97)
- This is a Wald type CI.

Plot of (Wald) Confidence Intervals For $\pi(x)$

The gray error band given by `geom_smooth()` is exactly the 95% CI for $\pi(x)$ as computed above.

```
ggplot(crabs, aes(x=Width, y = has.sate)) + geom_point() +
  geom_smooth(method = "glm", method.args = list(family = "binomial"))
```



Caution

We can NOT use the SE from `type="response"` since $\widehat{\pi}(x) = \frac{\exp(\widehat{\alpha}+\widehat{\beta}x)}{1+\exp(\widehat{\alpha}+\widehat{\beta}x)}$ is not approx. normal and hence we cannot calculate the 95% CI of $\pi(x)$ as

$$\widehat{\pi}(x) \pm 1.96(\text{SE from type=“response"})$$

```
predict(crabs.logit, data.frame(Width=30),type="response", se.fit=TRUE)
$fit
       1
0.928648

$se.fit
        1
0.0306818

$residual.scale
[1] 1
```

On the contrary, $\widehat{\alpha}+\widehat{\beta}x$ is approx. normal and hence we can calculate the 95% CI for $\alpha + \beta x$ as

$$\widehat{\alpha} + \widehat{\beta}x \pm (\text{SE from type=“link"}).$$