Topic 2

#1 (a) $E[h(\hat{p})] \approx E[h(p) + h'(p)(\hat{p}-p)]$
$= h(p) + h'(p) E(\hat{p}-p)$

Note $E(\hat{p}) = E\left(\frac{T}{n}\right) = \frac{np}{n} = p$

$\Rightarrow E(\hat{p}-p) = 0$

Therefore
$E[h(\hat{p})] \approx h(p).$

$var[h(\hat{p})] \approx var[h(p) + h'(p)(\hat{p}-p)]$
$= [h'(p)]^2 \, var(\hat{p}-p)$
$= [h'(p)]^2 \, var(\hat{p})$
$= \frac{[h'(p)]^2 \, p(1-p)}{n}$

$= \frac{[h'(p)]^2 \sigma^2}{n}$

(b) Set
$\frac{[h'(p)]^2 \, p(1-p)}{n} = c_0$ ← free of p.

$\rightarrow [h'(p)]^2 = \frac{n c_0}{p(1-p)}$

$\rightarrow h'(p) = \frac{c_1}{\sqrt{p(1-p)}}$, $c_1 = \sqrt{n c_0}$ ← free of p.

Therefore,

$$h(p) = \int \frac{c_1}{\sqrt{p(1-p)}}\, dp + c_2$$

(c_2 ↑ free of p)

Take $\sqrt{p} = \sin u$

$\Rightarrow p = \sin^2 u$

$dp = 2 \sin u \cos u \, du$

$$= \int \frac{2c_1 \sin u \cos u \, du}{\sin u \sqrt{1 - \sin^2(u)}} + c_2$$

$$= 2c_1 u + c_2$$

$$= 2c_1 \sin^{-1}(\sqrt{p}) + c_2$$

Take $c_1 = \frac{1}{2}$ and $c_2 = 0$ to get

$$h(p) = \sin^{-1}\sqrt{p} \; .$$

(c) By the δ-method

$$\sqrt{n}\,[h(\hat{p}) - h(p)] \xrightarrow{d} N\,[0, \underbrace{[h'(p)]^2 p(1-p)}]$$

$= \sigma_\phi^2$, say.

Find $h'(p)$.

If $y = \sin^{-1}(\sqrt{p})$, then

$$\sin y = \sqrt{p}$$

$$\rightarrow \frac{d}{dp} \sin y = \frac{d}{dp} \sqrt{p}$$

$$\Rightarrow \cos y \frac{dy}{dp} = \frac{1}{2\sqrt{p}}$$

Since $\sin^2 y + \cos^2 y = 1$, we have

$$\cos y = \sqrt{1 - \sin^2 y}$$
$$= \sqrt{1-p}$$

$$\Rightarrow \sqrt{1-p} \frac{dy}{dp} = \frac{1}{2\sqrt{p}}$$

$$\Rightarrow \frac{dy}{dp} = \frac{1}{2\sqrt{p(1-p)}} = h'(p).$$

$$\rightarrow \sigma_h^2 = [h'(p)]^2 p(1-p)$$
$$= \left[\frac{1}{2\sqrt{p(1-p)}}\right]^2 p(1-p) = \frac{1}{4}.$$

Therefore

$$\sqrt{n}\left[h(\hat{p}) - h(p)\right] \xrightarrow{d} N(0, \tfrac{1}{4}),$$

as claimed.

We have

"approximately normal"

$$h(\hat{p}) \sim AN\left[h(p), \frac{1}{4n}\right],$$

for large n.

$$\Rightarrow Z_n \equiv \frac{h(\hat{p}) - h(p)}{1/2\sqrt{n}} \sim AN(0,1)$$

So that

$$pr\left(-z_{\alpha/2} < Z_n < z_{\alpha/2}\right) \simeq 1-\alpha$$

$$\Rightarrow pr\left(-z_{\alpha/2} < \frac{h(\hat{p}) - h(p)}{1/2\sqrt{n}} < z_{\alpha/2}\right) \simeq 1-\alpha$$

$$\Rightarrow pr\left(\frac{-z_{\alpha/2}}{2\sqrt{n}} < h(\hat{p}) - h(p) < \frac{z_{\alpha/2}}{2\sqrt{n}}\right) \simeq 1-\alpha$$

$$\Rightarrow pr\left(h(\hat{p}) - \frac{z_{\alpha/2}}{2\sqrt{n}} < h(p) < h(\hat{p}) + \frac{z_{\alpha/2}}{2\sqrt{n}}\right) \simeq 1-\alpha.$$

Find the inverse transformation

$$x = h(p) = \sin^{-1}\sqrt{p} .$$

$$\Rightarrow \sin x = \sqrt{p}$$

$$\Rightarrow \sin^2 x = p$$

$$\Rightarrow h^{-1}(p) = \sin^2 p$$

Note $\frac{d h^{-1}(p)}{dp} = 2 \sin p \cos p > 0$

$$\forall p \in (0,1)$$

Therefore $h^{-1}(p) = \sin^2 p$ is increasing over $(0,1)$.

Therefore,

$$\Pr\left[\sin^2\left[h(\hat{p}) - \frac{z_{\alpha/2}}{2\sqrt{n}}\right] < p < \sin^2\left[h(\hat{p}) + \frac{z_{\alpha/2}}{2\sqrt{n}}\right] \right]$$

$$\approx 1 - \alpha$$

Showing that

$$\left(\sin^2\left[h(\hat{p}) - \frac{z_{\alpha/2}}{2\sqrt{n}}\right], \sin^2\left[h(\hat{p}) + \frac{z_{\alpha/2}}{2\sqrt{n}}\right] \right)$$

is an approximate $100(1-\alpha)\%$ CI for p.

(d) Examples include
- coverage probability
- interval length (average, median)
- Distal/Mesial noncoverage
- Minimum coverage
- proportion of intervals with endpoints outside parameter space

(e) Bayesian interval (e.g. beta prior)
Score interval
LR interval
Agresti-Coull interval
Exact interval
Continuity corrected interval

~~[illegible]~~ Day 2 Part 2

(a) $$P(T>t) = \int_0^\infty [1-\bar{G}(t)]^z \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\alpha z} dz$$

$$= \int_0^\infty \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp\left(-(\alpha - \log \bar{G}) z\right) dz$$

$\bar{G} = 1 - G$

$$= \left[\frac{\alpha}{\alpha - \log \bar{G}(t)}\right]^\alpha.$$

(b) $$P(T>t) = \left(\frac{\alpha}{\alpha + \theta t}\right)^\alpha$$

(c) $$\frac{d}{dt} P(T>t) = \alpha \left(\frac{\alpha}{\alpha + \theta t}\right)^{\alpha-1} \left(-\frac{\alpha\theta}{(\alpha+\theta t)^2}\right)$$

$$= -\left(\frac{\alpha}{\alpha+\theta t}\right)^\alpha \frac{\alpha\theta}{(\alpha+\theta t)}$$

So

$$f(t) = \left(\frac{\alpha\theta}{\alpha+\theta t}\right) P(T>t).$$

(d) $$h(t) = \frac{\left(\frac{\alpha\theta}{\alpha+\theta t}\right) P(T>t)}{P(T>t)}$$

$$= \frac{\alpha\theta}{\alpha+\theta t}.$$

This is <u>decreasing</u> in t.

Day 2. Problem 3.

3. $X_1, \cdots, X_n \stackrel{i.i.d}{\sim} f(x|\theta)$.

(a) For both $j=1$ and 2, $E(X^2) = \theta^2$.

(b) If $j=1$, the likelihood function is

$$L(\theta|\underline{X}) = \left(\frac{1}{\sqrt{2\pi\theta^2}}\right)^n \exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^n X_i^2\right)$$

Because the MLE for $Var(X)$ for $N(0, \theta^2)$ is T_1/n, where $T_1 = \sum_{i=1}^n X_i^2$. By the invariance property of MLE, the MLE for θ is $\hat{\theta}_1 = \sqrt{T_1/n}$

If $j=2$ the likelihood function is

$$L(\theta|\underline{X}) = \left(\frac{1}{\sqrt{2}\theta}\right)^n \exp\left(-\frac{\sqrt{2}}{\theta}\sum_{i=1}^n |X_i|\right)$$

and

$$\log f(\underline{x}|\theta) = -n\log\sqrt{2} - n\log\theta - \frac{\sqrt{2}}{\theta}\sum_{i=1}^n |X_i|.$$

Solving $\dfrac{\partial \log f(\underline{x}|\theta)}{\partial\theta} = -\dfrac{n}{\theta} + \dfrac{\sqrt{2}}{\theta^2}\sum_{i=1}^n |X_i| = 0$ for θ

gives $\hat{\theta}_2 = \dfrac{T_2}{n}$, where $T_2 = \sqrt{2}\sum_{i=1}^n |X_i|$.

The MLE for j is given by

$$\begin{aligned}\hat{j} &= I\{f(\underline{x}|\hat{\theta}_1) \geq f(\underline{x}|\hat{\theta}_2)\} + 2I\{f(\underline{x}|\hat{\theta}_1) < f(\underline{x}|\hat{\theta}_2)\} \\ &= I\left(\sqrt{T_1}/T_2 \leq \sqrt{e/(n\pi)}\right) + 2I\left(\sqrt{T_1}/T_2 > \sqrt{e/(n\pi)}\right).\end{aligned}$$

The MLE for θ is given by

$$\hat{\theta} = \hat{\theta}_1 I(\hat{j}=1) + \hat{\theta}_2 I(\hat{j}=2).$$

#

(c) The likelihood ratio statistic is

$$\lambda(\underset{\sim}{X}) = I(\hat{j}=1) + \frac{L(\hat{\theta}_1, j=1 \mid \underset{\sim}{X})}{L(\hat{\theta}_2, j=2 \mid \underset{\sim}{X})} I(\hat{j}=2)$$

$$= I\left(\sqrt{T_1}/T_2 \le \sqrt{e/(n\pi)}\right) + \left(\frac{\sqrt{e}\, T_2}{\sqrt{n\pi T_1}}\right)^n I\left(\sqrt{T_1}/T_2 > \sqrt{e/(n\pi)}\right)$$

A size α test based on $\lambda(\underset{\sim}{X})$ is ~~$\delta(\underset{\sim}{X}) = I\left(\frac{\sqrt{T_1}}{T_1}\right.$~~
$\delta(\underset{\sim}{X}) = I(T_2/\sqrt{T_1} < c_{\alpha,n})$, where $c_{\alpha,n}$ satisfies
$\sup_{\theta \in R} P\left(\frac{T_2}{\sqrt{T_1}} < c_{\alpha,n}\right) = \alpha$ when $X_1, \cdots, X_n$ iid $N(0, \theta^2)$.

#

Day 2

(#4) (a) $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$, for $i = 1, 2, \ldots, n = 54$.

- Y_i ← response for i^{th} unit
- $\beta_0, \beta_1, \beta_2, \beta_3$ ← regression parameters
- $\varepsilon_i \sim iid\ N(0, \sigma^2)$ (random errors)

(b) Body Type I → same as Body Type III

0.4249

GEST Type I = ?

NB: Type I SS add to Model SS.

GEST Type I + BRAIN Type I + Body Type I = 7.5231

⇒ GEST Type I = 7.5231 − 0.3017 − 0.4249 = **6.7965**

GEST Type III SS = $\dfrac{\hat{\beta}_1^2}{(X^T X)^{-1}_{22}}$

We know

$\widehat{se}(\hat{\beta}_1) = \sqrt{MSE\ (X^T X)^{-1}_{22}}$

$$\Rightarrow \quad 0.0844 = \sqrt{\frac{7.3624}{50}(X^TX)^{-1}_{22}}$$

$$\rightarrow \quad (0.0844)^2 = \frac{7.3624}{50}(X^TX)^{-1}_{22}$$

$$\Rightarrow \quad (X^TX)^{-1}_{22} = \frac{50\,(0.0844)^2}{7.3624}$$

$$= 0.0484$$

$$\rightarrow \quad \text{Type III GEST SS} = \frac{(-0.2685)^2}{0.0484} = \boxed{1.4902}$$

Summarizing:

	Type I SS	Type III
GEST	6.7965	1.4902
BRAIN	given	given
BODY	0.4249	given

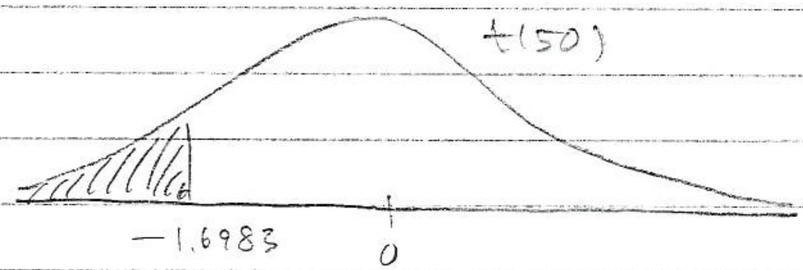
(C) We want to test

$$H_0: \beta_3 = 0$$
$$H_1: \beta_3 < 0$$

using $\alpha = 0.05$. The t-statistic is:

$$t = \frac{\hat{\beta}_3}{\widehat{se}(\hat{\beta}_3)} = \frac{-0.1019}{0.0600} = -1.6983$$

The p-value for this test is



$$p = pr(t(50) < -1.6983) = 0.0478$$

pt(−1.6983, 50) R command

Because the p-value is smaller than $\alpha = 0.05$, we reject H_0. At the 5 percent significance level, we conclude that there is an additional negative relationship between BODY & SLEEP, after adjusting for the effects of GEST & BRAIN.

(d) We want to test

$$H_0: \beta_2 = \beta_3 = 0$$
$$H_1: \text{not } H_0$$

at the $\alpha = 0.05$ level. Use a reduced-versus-full F test

reduced model → H_0: $Y = \beta_0 + \beta_1 GEST + e$

(4)

full model $\longrightarrow H_1: Y = \gamma_0 + \gamma_1 GEST + \gamma_2 BRAIN + \gamma_3 BODY + \epsilon$

F-statistic

$$F = \frac{(SS_{FULL} - SS_{RED})/2}{MSE_{full}}$$

$$= \frac{(0.3017 + 0.4249)/2}{7.3624/50}$$

$$= 2.467.$$

The p-value is ~~1 - pf(2.467, 2, 50) in R.~~

$$p = pr[F_{2,50} > 2.467] = 0.0808.$$

Because the p-value $> \alpha = 0.05$, we do not reject H_0. That is, BRAIN & BODY do not significantly add to a linear regression model that includes GEST.

e) From part (d), we do not reject H_0 so we stick w/ the simpler model

$$E(Y | x_1, x_2, x_3) = \beta_0 + \beta_1 x_1$$

An estimate of $var(T \mid x_1)$ is the MSE from the reduced model.

ANOVA table → reduced model

Source	df	SS	MS
Model	1	6.7965	
Error	52	8.089	0.1556
Total	53	14.8855	

Model SS: $7.5231 - 0.3017 - 0.4249$
(BRAIN) (BODY)

$= 6.7965$

An estimate of $var(T \mid x_1)$ is

$$\hat{\sigma}^2 = MSE$$
$$= 0.1556.$$

Day 2

(1)

#5 The 1st question essentially asks us to test

$$H_0: \beta_3 = 1$$
$$H_1: \beta_3 \neq 1$$

From the output

$$t = \frac{\hat{\beta}_3 - 1}{\widehat{se}(\hat{\beta}_3)} = \frac{1.0737 - 1}{0.2943} < 1$$

$\Rightarrow$ do not reject H_0; stick w/ simpler model.

Note that

$$\lim_{x \to \infty} \frac{\beta_1}{1 + \exp(\ldots)} = \beta_1$$

max. or maximum reaction rate.

A point estimate for β_1 is

$$\hat{\beta}_1 = 628.87$$

A 95% CI for β_1 is

$$\hat{\beta}_1 \pm t_{5,0.025}\, \widehat{se}(\hat{\beta}_1)$$

$\Rightarrow$ 628.87 ± 2.57 (54.53)

$\Rightarrow$ (488.73, 769.01).

We are 95% confident that the mean max. reaction force is between 488.73 and 769.01 newtons.

We would like to estimate

$$\theta = E(T \mid x=750) = \frac{\beta_1}{1+\left(\frac{\beta_2}{750}\right)} \equiv g(\beta), \text{ say.}$$

A point estimate of θ is

$$\hat{\theta} = \frac{628.27}{1+\left(\frac{363.10}{750}\right)} = 423.32 \text{ nM/mg/hr.}$$

Use δ-method to get a large sample CI for $\theta = g(\beta_1, \beta_2)$

We know

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \sim AN\left[\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, V\right],$$

where

$$V = \begin{bmatrix} var(\hat{\beta}_1) & cov(\hat{\beta}_1, \hat{\beta}_2) \\ symm & var(\hat{\beta}_2) \end{bmatrix}.$$

Therefore, by the δ-method

$$\hat{\theta} \sim AN\left(\theta, \underbrace{\frac{\partial g(\beta)}{\partial \beta'} V \frac{\partial g(\beta)}{\partial \beta}}_{\equiv \sigma^2_{\hat{\theta}}, \text{ say.}}\right)$$

$$g(s,t) = \frac{s}{1+\frac{t}{750}} = \frac{750s}{750+t}$$

$$\frac{\partial g(s,t)}{\partial s} = \frac{750}{750+t}$$

$$\frac{\partial g(s,t)}{\partial t} = \frac{-750s}{(750+t)^2}$$

$$\Rightarrow \frac{\partial g(\beta)}{\partial \beta} = \begin{pmatrix} \frac{750}{750+\beta_2} \\ \frac{-750\beta_1}{(750+\beta_2)^2} \end{pmatrix}_{2\times 1}$$

An estimate of this is

$$= \begin{pmatrix} \frac{750}{750+363.10} \\ \frac{-750(628.07)}{(750+363.10)^2} \end{pmatrix} = \begin{pmatrix} 0.6738 \\ -0.3803 \end{pmatrix}$$

We can estimate V by

$$\begin{bmatrix} (54.53)^2 & 4228.4 \\ 4228.4 & (89.34)^2 \end{bmatrix}$$

Note

$$4238.4 = 0.87\,(54.53)\,(89.34)$$

$\widehat{\text{corr}}$, $\widehat{se}(\hat\beta_1)$, $\widehat{se}(\hat\beta_2)$

Therefore

$$\hat\theta \sim AN\left(\theta,\ \sigma^2_{\hat\theta}\right),$$

where $\sigma^2_{\hat\theta}$ is estimated by

$$(0.6738,\ -0.8803)\begin{bmatrix} (54.53)^2 & 4238.4 \\ 4238.4 & (89.34)^2 \end{bmatrix}\begin{pmatrix} 0.6738 \\ -0.8803 \end{pmatrix}$$

$$(391.69,\ -170.28)\begin{pmatrix} 0.6738 \\ -0.8803 \end{pmatrix} = \underline{\underline{332.21}}$$

$$\Rightarrow \hat\theta \sim AN(\theta,\ 332.21)$$

Therefore

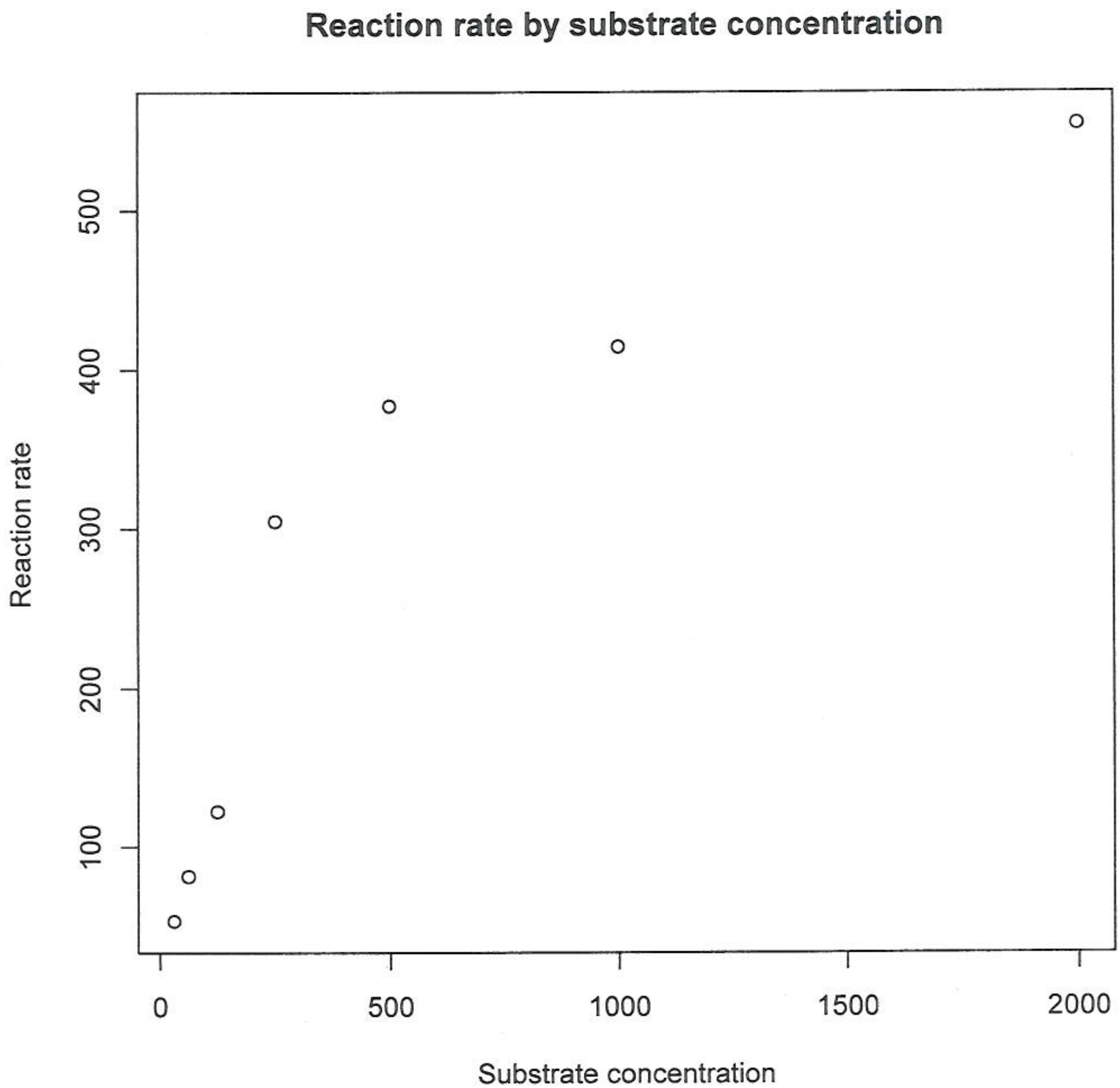
$$423.32 \pm 1.96\sqrt{332.21} \quad \text{is}$$

$\underline{\underline{(387.6,\ 459.0)}}$

or approximate 95% CI for $\theta = E(Y \mid x = 20)$.

(5)

We are 95% confident that the mean reaction rate when the substrate concentration is 750 μM is between 387.6 and 459.0 μM/mg/hr.



Reaction rate by substrate concentration
Reaction rate
100
200
300
400
500
0
500
1000
1500
2000
Substrate concentration

```
##############################
# Author: Joshua M. Tebbs
# Date: 1 April 2010
##############################

> reac<- c(53.01,81.42,122.11,304.57,376.87,414.13,553.46)
> subs <- c(31.25,62.5,125,250,500,1000,2000)
>
> #Plot the data
> plot(subs,reac,main="Reaction rate by substrate concentration",xlab="Substrate
concentration",ylab="Reaction rate")
>
> #Three-parameter model
> mod.3 <- nls(reac ~
beta1/(1+(beta2/subs)^(beta3)),start=list(beta1=500,beta2=200,beta3=1),trace=T)
17452.98 :  500 200   1
6857.979 :  603.627571 323.555612   1.029904
6715.411 :  603.358752 327.750493   1.070850
6715.092 :  602.411733 326.123946   1.072700
6715.076 :  602.149339 325.760607   1.073472
6715.075 :  602.078895 325.660483   1.073664
6715.075 :  602.060072 325.633813   1.073716
6715.075 :  602.055026 325.626657   1.073729
> summary(mod.3,correlation=TRUE)

Formula: reac ~ beta1/(1 + (beta2/subs)^(beta3))

Parameters:
      Estimate Std. Error t value Pr(>|t|)
beta1 602.0550   102.2422   5.889  0.00416 **
beta2 325.6267   141.5942   2.300  0.08296 .
beta3   1.0737     0.2943   3.649  0.02180 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40.97 on 4 degrees of freedom

Correlation of Parameter Estimates:
      beta1 beta2
beta2  0.95
beta3 -0.85 -0.82

Number of iterations to convergence: 7
Achieved convergence tolerance: 6.87e-06

> #Two-parameter model
> mod.2 <- nls(reac ~ beta1/(1+(beta2/subs)),start=list(beta1=500,beta2=200),trace=T)
17452.98 :  500 200
7006.661 :  610.0481 330.2598
6800.922 :  627.8438 362.0152
6800.678 :  628.2807 363.1148
6800.678 :  628.2749 363.1045
> summary(mod.2,correlation=TRUE)

Formula: reac ~ beta1/(1 + (beta2/subs))

Parameters:
      Estimate Std. Error t value Pr(>|t|)
beta1   628.27      54.53  11.522 8.63e-05 ***
beta2   363.10      89.34   4.064  0.00969 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.88 on 5 degrees of freedom

Correlation of Parameter Estimates:
      beta1
beta2 0.87

Number of iterations to convergence: 4
Achieved convergence tolerance: 5.704e-07
```

```
DATA injured;
INPUT x y;
cards;
31.25 53.01
62.5 81.42
125 122.11
250 304.57
500 376.87
1000 414.13
2000 553.46
;
run;

/* 3-parameter model */

PROC NLIN DATA = injured METHOD=GAUSS HOUGAARD;
   PARAMETERS  beta1=2010 beta2= 10 beta3=0.5;
   MODEL y    = beta1/(1+(beta2/x)**beta3);
run;

/* 2-parameter model */

PROC NLIN DATA = injured METHOD=GAUSS HOUGAARD;
   PARAMETERS  beta1=2010 beta2= 10;
   MODEL y    = beta1/(1+(beta2/x));
run;
```

The NLIN Procedure
Dependent Variable y
Method: Gauss-Newton

Iterative Phase

Iter	beta1	beta2	beta3	Sum of Squares
0	2010.0	10.0000	0.5000	12979342
1	890.8	26.1130	0.2105	626877
2	961.5	60.7221	0.2040	623028
3	1031.2	133.9	0.2031	611929
4	1158.0	404.4	0.2099	606898
5	1423.3	2013.6	0.2686	468568
6	1250.8	917.1	0.3251	386961
7	906.5	397.7	0.4877	141779
8	698.4	465.1	0.7398	11555.9
9	585.2	249.0	1.0022	11187.9
10	602.8	321.6	1.0618	6746.0
11	602.0	325.6	1.0745	6715.1
12	602.0	325.6	1.0738	6715.1
13	602.0	325.6	1.0738	6715.1

14	602.1	325.6	1.0737	6715.1

NOTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	14
Subiterations	9
Average Subiterations	0.642857
R	6.713E-6
PPC(beta2)	5.755E-6
RPC(beta2)	0.000021
Object	8.13E-10
Objective	6715.075
Observations Read	7
Observations Used	7
Observations Missing	0

NOTE: An intercept was not specified for this model.

The NLIN Procedure

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	3	730251	243417	145.00	0.0002
Error	4	6715.1	1678.8		
Uncorrected Total	7	736966			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence	Limits	Skewness
beta1	602.1	102.2	318.2	885.9	1.9821
beta2	325.6	141.6	-67.4919	718.7	3.0036
beta3	1.0737	0.2943	0.2567	1.8908	0.5425

Approximate Correlation Matrix

	beta1	beta2	beta3
beta1	1.0000000	0.9498455	-0.8526529
beta2	0.9498455	1.0000000	-0.8158150
beta3	-0.8526529	-0.8158150	1.0000000

The NLIN Procedure
Dependent Variable y
Method: Gauss-Newton

Iterative Phase

Iter	beta1	beta2	Sum of Squares

0	2010.0	10.0000	17677989
1	428.4	19.7982	191152
2	454.8	80.0124	58972.5
3	546.3	201.1	15431.1
4	610.3	320.0	7225.5
5	627.4	360.6	6802.1
6	628.3	363.1	6800.7
7	628.3	363.1	6800.7

NOTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	7
R	9.625E-7
PPC(beta2)	5.296E-7
RPC(beta2)	0.000048
Object	8.47E-9
Objective	6800.678
Observations Read	7
Observations Used	7
Observations Missing	0

NOTE: An intercept was not specified for this model.

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	2	730165	365082	268.42	<.0001
Error	5	6800.7	1360.1		
Uncorrected Total	7	736966			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits		Skewness
beta1	628.3	54.5270	488.1	768.4	0.4071
beta2	363.1	89.3448	133.4	592.8	0.7718

Day 2

(1)

(#6) (a) A statistical model is

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where Y_{ij} = response for j^{th} experimental unit in i^{th} trt. group

$$\mu_i = E(Y_{ij}) \quad \forall j = 1, 2, \ldots, n$$

ϵ_{ij} = random error with a $N(0, c_i\sigma^2)$ dist^n

↑ known.

(b) We know

$$\frac{(n-1)S_i^2}{c_i\sigma^2} \sim \chi^2_{(n-1)} \quad \forall i = 1, 2, \ldots, k$$

Here,

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i\cdot})^2$$

is the usual sample variance for the i^{th} treatment; $i = 1, 2, \ldots, k$

Because the samples are independent,

$$\sum_{i=1}^{k} \frac{(n-1) S_i^2}{c_i \sigma^2} \sim \chi^2_{[k(n-1)]}$$

↑ df add

$$\Rightarrow E\left[\sum_{i=1}^{k} \frac{(n-1) S_i^2}{c_i \sigma^2}\right] = k(n-1)$$

$$\Rightarrow E\left[\frac{1}{k(n-1)} \sum_{i=1}^{k} \frac{(n-1) S_i^2}{c_i}\right] = \sigma^2$$

Note that

$$(n-1) S_i^2 = \sum_{j=1}^{n} (Y_{ij} - \overline{Y}_{i+})^2$$

Therefore

$$E\left[\frac{1}{k(n-1)} \sum_{i=1}^{k} c_i^{-1} \sum_{j=1}^{n} (Y_{ij} - \overline{Y}_{i+})^2\right] = \sigma^2$$

Showing that $\hat{\sigma}^2$ is an unbiased estimator of σ^2.

(c) We know

$$\overline{Y}_{i+} \sim N\left(\mu_i, \frac{c_i \sigma^2}{n}\right)$$

for $i = 1, 2, \ldots, k$.

Consider the linear combination

$$\hat{\theta} = \sum_{i=1}^{k} a_i \bar{Y}_{i+}$$

Note that

$$E(\hat{\theta}) = E\left(\sum_{i=1}^{k} a_i \bar{Y}_{i+}\right)$$

$$= \sum_{i=1}^{k} a_i E(\bar{Y}_{i+})$$

$$= \sum_{i=1}^{k} a_i \mu_i = \theta$$

Also,

$$\mathrm{var}(\hat{\theta}) = \mathrm{var}\left(\sum_{i=1}^{k} a_i \bar{Y}_{i+}\right)$$

$$= \sum_{i=1}^{k} a_i^2 \mathrm{var}(\bar{Y}_{i+})$$

$$= \sum_{i=1}^{k} a_i^2 \left(c_i \frac{\sigma^2}{n}\right)$$

$$= \frac{\sigma^2}{n} \sum_{i=1}^{k} a_i^2 c_i$$

Because linear combinations of normal r.v.s are normal

$$\hat{\theta} \sim N\left(\theta, \frac{\sigma^2}{n} \sum_{i=1}^{k} a_i^2 c_i\right)$$

Therefore,

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\frac{\sigma^2}{n}\sum_{i=1}^{k} a_i^2 c_i}} \sim N(0,1)$$

Also, recall,

$$W \equiv \sum_{i=1}^{k} \frac{(n-1)S_i^2}{c_i \sigma^2} \sim \chi^2_{[k(n-1)]}$$

Note that Z and W are independent because

Z = function of $\bar{Y}_{i+}$ $i = 1, 2, \ldots, k$
W = " " S_i^2 $i = 1, 2, \ldots, k$

Therefore,

$$\frac{Z}{\sqrt{\frac{W}{k(n-1)}}} = \frac{\dfrac{\hat{\theta} - \theta}{\sqrt{\frac{\sigma^2}{n}\sum_{i=1}^{k} a_i^2 c_i}}}{\sqrt{\dfrac{\sum_{i=1}^{k} \frac{(n-1)S_i^2}{c_i \sigma^2}}{k(n-1)}}}$$

$$\sim \frac{\text{"}N(0,1)\text{"}}{\sqrt{\frac{\text{"}\chi^2\text{"}}{df}}} \sim t(df)$$

But, the σ^2's cancel in the last quantity, leaving us with

$$t = \frac{\hat{\theta} - \theta}{\sqrt{\frac{\hat{\sigma}^2}{n} \sum_{i=1}^{k} a_i^2 c_i}} \sim t[k(n-1)].$$

$$\Rightarrow \Pr\left(-t_{k(n-1), \alpha/2} < \frac{\hat{\theta} - \theta}{\sqrt{\frac{\hat{\sigma}^2}{n} \sum_{i=1}^{k} a_i^2 c_i}} < t_{k(n-1), \alpha/2} \right) = 1 - \alpha$$

After algebra, we have

$$\Pr\left(\hat{\theta} - t_{k(n-1), \alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n} \sum_{i=1}^{k} a_i^2 c_i} < \theta < \hat{\theta} + t_{k(n-1), \alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n} \sum_{i=1}^{k} a_i^2 c_i} \right) = 1 - \alpha$$

Showing that

$$\hat{\theta} \pm t_{k(n-1), \alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n} \sum_{i=1}^{k} a_i^2 c_i}$$

is a $100(1-\alpha)\%$ CI for θ.

(d) We have

$c_1 = 1 \qquad n = 5$

$c_2 = 2$

$c_3 = 15$

To compare the median group to the average of the low/high groups, we are interested in the contrast

$$\theta = \mu_2 - \frac{1}{2}(\mu_1 + \mu_3)$$

$$= -\frac{1}{2}\mu_1 + \mu_2 - \frac{1}{2}\mu_3$$

So that

$a_1 = -\frac{1}{2}$

$a_2 = 1$

$a_3 = -\frac{1}{2}$

R gives

$\bar{y}_{1+} = 3.908 \qquad \bar{y}_{2+} = 9.674 \qquad \bar{y}_{3+} = 24.926$

$s_1^2 = 0.909 \qquad s_2^2 = 1.373 \qquad s_3^2 = 15.864$

$$\hat{\theta} = -\frac{1}{2}\bar{y}_{1+} + \bar{y}_{2+} - \frac{1}{2}\bar{y}_{3+}$$

$$= -\frac{1}{2}(3.908) + 9.674 - \frac{1}{2}(24.926)$$

$$= -4.743$$

$$\hat{\sigma}^2 = \frac{1}{k(n-1)} \sum_{i=1}^{3} c_i^{-1} \sum_{j=1}^{5} (y_{ij} - \bar{y}_{i+})^2$$

$$= \frac{1}{3(5-1)} \sum_{i=1}^{3} c_i^{-1} (5-1) s_i^2$$

$$= \frac{1}{3} \sum_{i=1}^{3} \frac{s_i^2}{c_i}$$

$$= \frac{1}{3}\left[\frac{0.909}{1} + \frac{1.373}{2} + \frac{15.864}{15} \right]$$

$$= 0.884.$$

$$t_{k(n-1), \alpha/2} = t_{3(5-1), \frac{\alpha}{2}} = t_{12, 0.05}$$

$$= 1.782$$

↑ qt(0.95, 12) in R.

Therefore the 90% CI for θ is

$$-4.743 \pm 1.782 \sqrt{\frac{0.884}{5}\left[(-\tfrac{1}{2})^2(1) + (1)^2(2) + (-\tfrac{1}{2})^2(15)\right]}$$

$$-4.743 \pm 1.835 \longrightarrow (-6.578, -2.908)$$

We are 90% confident that θ is between -6.578 and -2.908 grams. The medium group is significantly different from the average of the low & high groups.

```
## Problem 6

> low<-c(3.89,4.87,3.26,2.70,4.82)
> med<-c(8.54,9.32,8.76,11.30,10.45)
> hi<-c(20.39,24.22,30.91,22.78,26.33)
>
> mean(low)
[1] 3.908
> mean(med)
[1] 9.674
> mean(hi)
[1] 24.926
>
> var(low)
[1] 0.90917
> var(med)
[1] 1.37318
> var(hi)
[1] 15.86463
> qt(0.95,12)
[1] 1.782288
>
```