

PubH 7405: REGRESSION ANALYSIS



CROSS-OVER EXPERIMENT DESIGNS

THE ROLE OF STUDY DESIGN

In a “standard” experimental design, a linear model for a continuous response/outcome is:

$$Y = \begin{bmatrix} \text{Overall} \\ \text{Constant} \end{bmatrix} + \begin{bmatrix} \text{Treatment} \\ \text{Effect} \end{bmatrix} + \begin{bmatrix} \text{Experimental} \\ \text{Error} \end{bmatrix}$$

The last component, ‘experimental error’, includes not only error specific to the experimental process but also includes “subject effect” (age, gender, etc...). Sometimes these subject effects are large making it difficult to assess “treatment effect”.

Blocking (to turn a completely randomized design into a randomized complete block design) would help. But it would only help to “reduce” subject effects, not to “eliminate” them: subjects in the same block are only similar, not identical – unless we have “blocks of size one”. And that the basic idea of “Cross-over Design**”, a very popular form in biomedical research.**

Cross-over is a very special design where we have “bloc” of size one; each subject serves as his/her own control receiving both treatment. Randomization decides the order. The outcome could be binary or on the continuous scale.

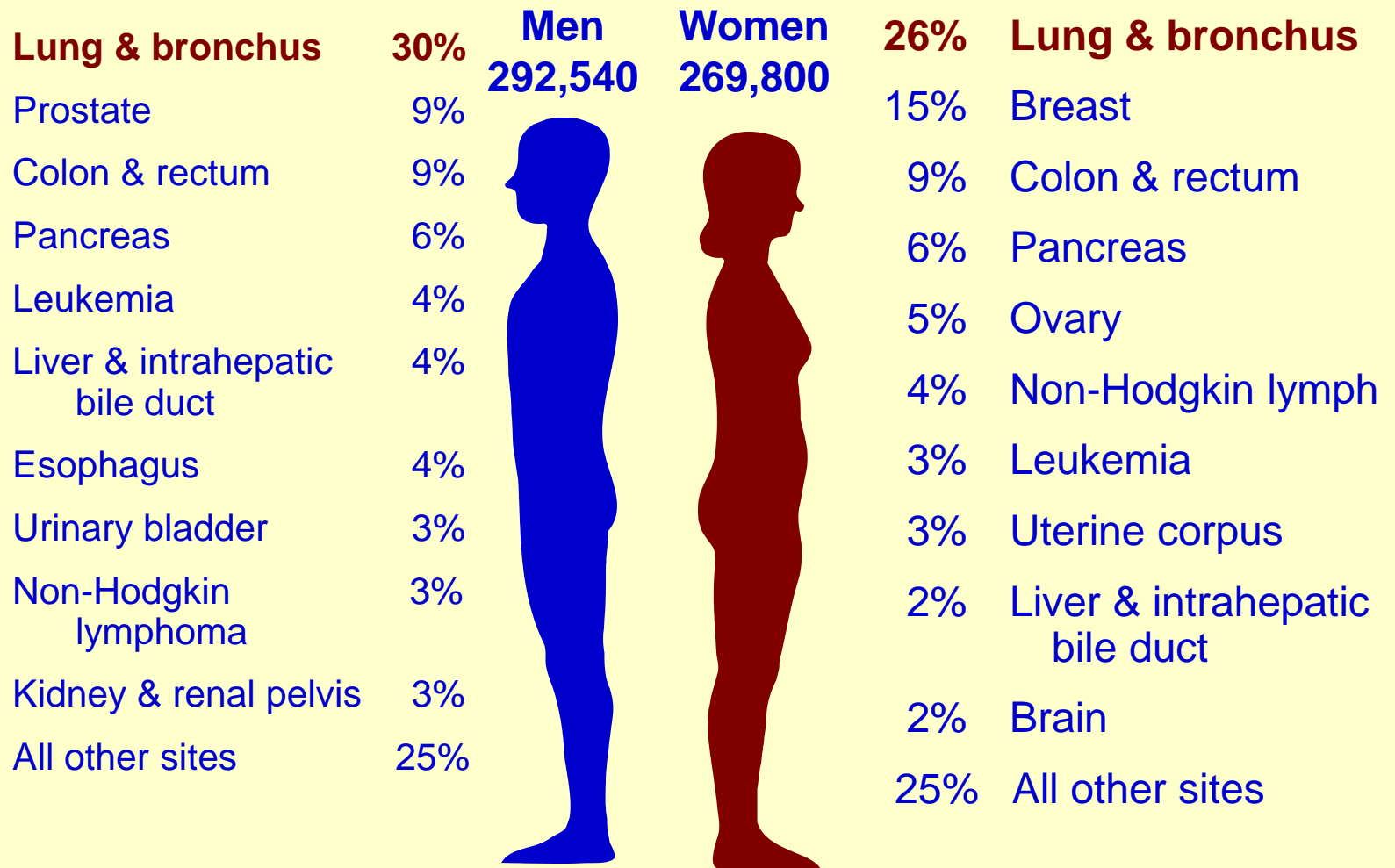
Let start with the case of a continuous outcome and, for illustration, consider a project we just completed here: a clinical trial to prevent lung cancers.

US Mortality, 2006

Rank	Cause of Death	No. of deaths	% of all deaths
1.	Heart Diseases	631,636	26.0
2.	Cancer	559,888	23.1
3.	Cerebrovascular diseases	137,119	5.7
4.	Chronic lower respiratory diseases	124,583	5.1
5.	Accidents (unintentional injuries)	121,599	5.0
6.	Diabetes mellitus	72,449	3.0
7.	Alzheimer disease	72,432	3.0
8.	Influenza & pneumonia	56,326	2.3
9.	Nephritis	45,344	1.9
10.	Septicemia	34,234	1.4

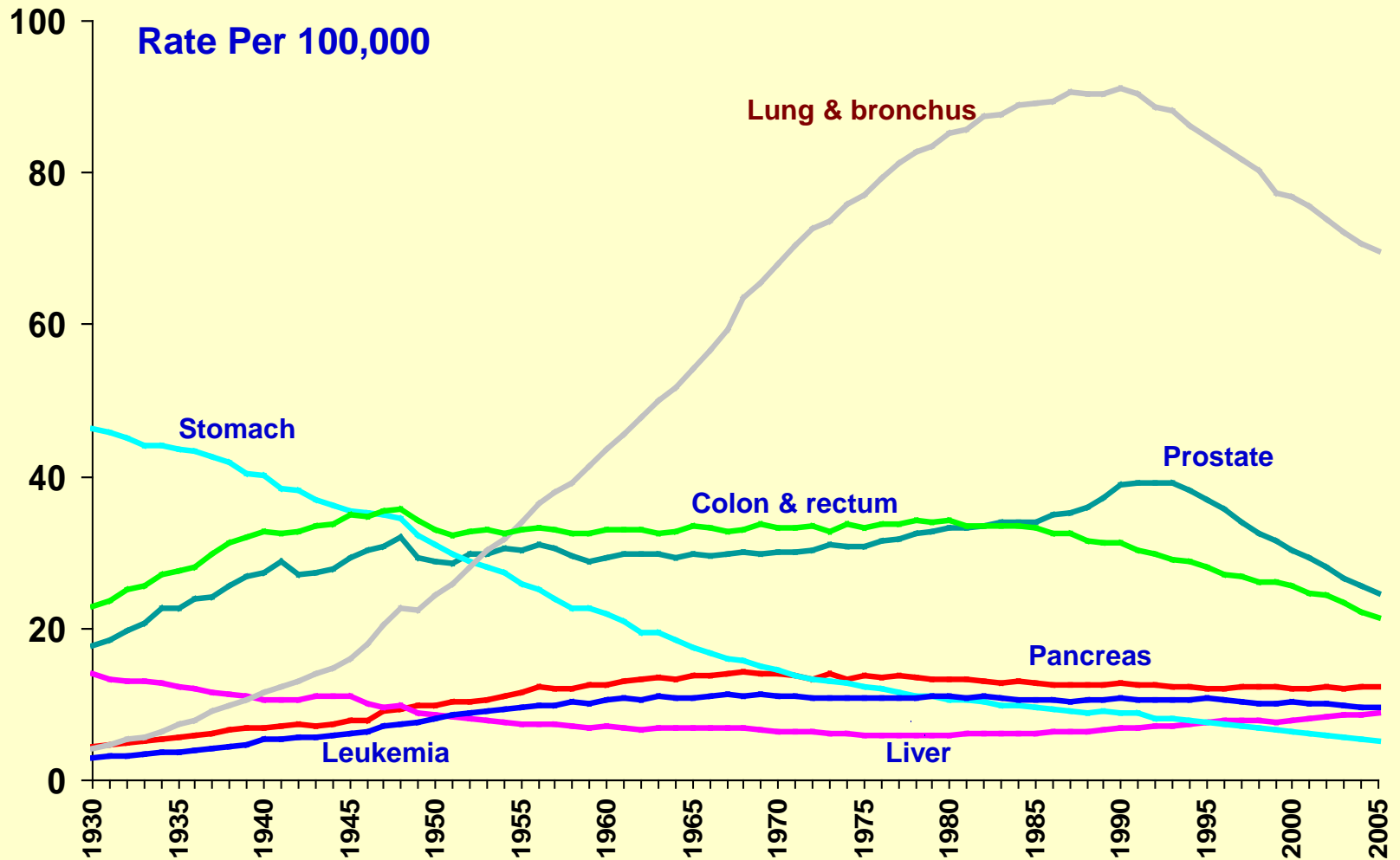
Source: US Mortality Data 2006, National Center for Health Statistics, Centers for Disease Control and Prevention, 2009.

2009 Estimated US Cancer Deaths



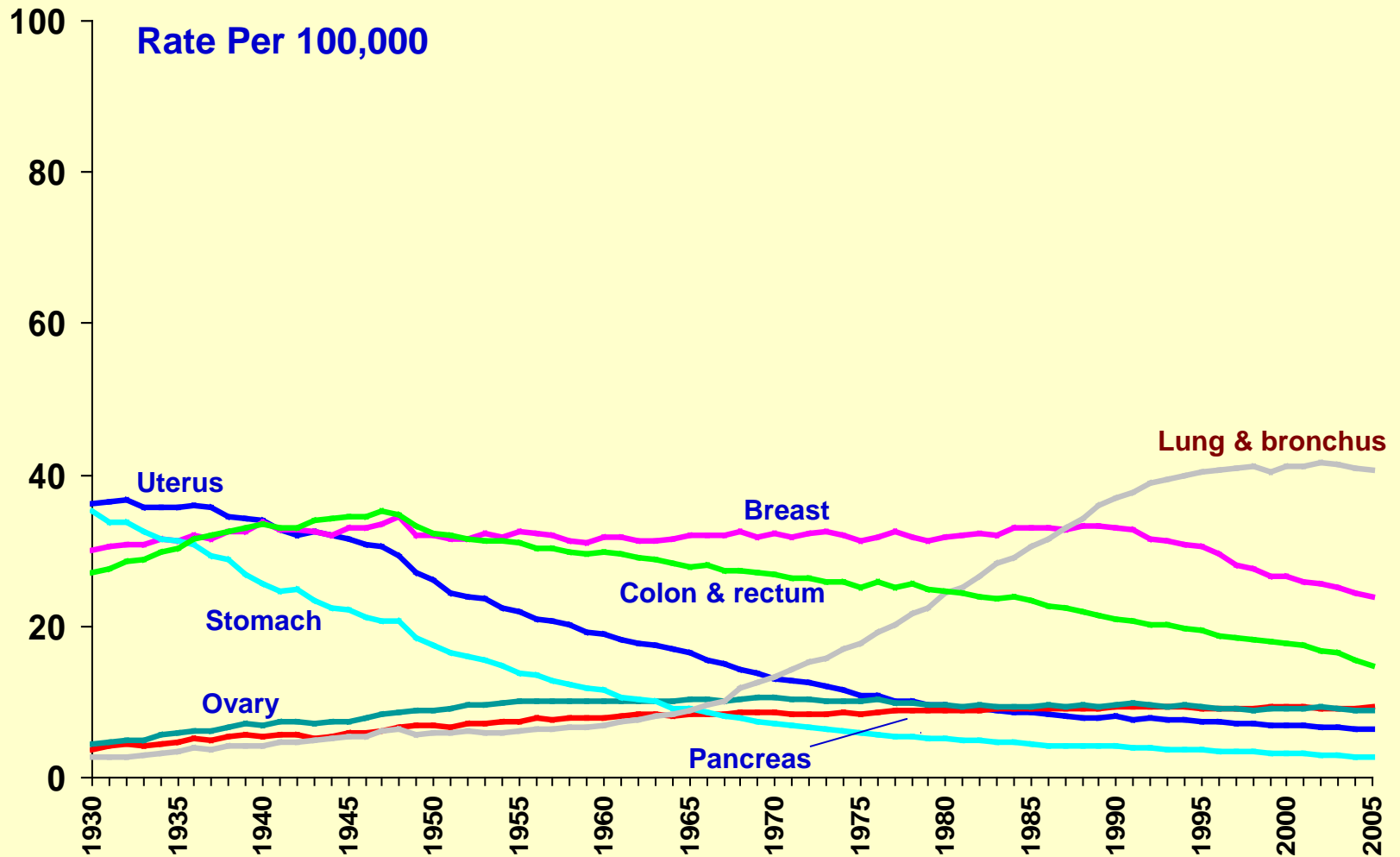
Source: American Cancer Society, 2009.

Cancer Death Rates Among Men, US, 1930-2005



Source: US Mortality Data 1960-2005, US Mortality Volumes 1930-1959, National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.

Cancer Death Rates Among Women, US, 1930-2005



Source: US Mortality Data 1960-2005, US Mortality Volumes 1930-1959, National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.

Lung cancer is the leading cause of cancer death in the United States and worldwide. Cigarette smoking causes approximately 90% of lung cancer. Despite anti-smoking campaigns over the past 40 years, over 45 million (22%) adult Americans are still current smokers. The development of a viable chemoprevention strategy targeting current smokers potentially could decrease lung cancer mortality.

Previous studies have shown

(1) that the tobacco specific nitrosamine 4(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) is a major lung carcinogen in tobacco smoke,

(2) that 2-phenethyl isothiocyanate (PEITC) is a potent inhibitor of NNK-induced lung carcinogenesis in rats and mice.

PEITC can be found in water cress, garden cress, broccoli, among other foods; but few people eat enough of those foods. We extract PEITC from foods then form pills having higher concentration.

Then we designed and conducted a placebo-controlled cross-over clinical trial to assess the effect of a PEITC supplement on changes of NNK metabolism in smokers. We hypothesize that there will be a 30% increase in urinary NNAL plus NNAL-Gluc among PEITC treated subjects (taking more toxin out).

In the most simple cross-over design, subjects are randomly divided into two groups (often of equal size); subjects in both groups/series take both treatments (experimental treatment and placebo/control) but in different “orders”.

Of course, “**order effects**” and “**carry-over effects**” are possible. And the cross-over designs are not always suitable. They are commonly used when treatment effects are not permanent; for example some treatments of rheumatism.

THE DESIGN

In our “PEITC trial”, measurements (urinary total NNAL) will be taken from each subject in the two supplementation sequences as seen in the following diagram:

Group 1: Period #1 (PEITC; A1) – washout – Period #2 (Placebo; B2)

Group 2: Period #1 (Placebo; B1) – washout – Period #2 (PEITC; A2)

The letter is used to denote supplementation or treatment (A for PEITC and B for Placebo) and the number, 1 or 2, denotes the period; e.g. “A1” for PEITC taken in period #1.

The “**washout periods**” are inserted in order to eliminate possible “**carry-over effects**” (The half-life of dietary PEITC in vivo is between 2-3 hours, with complete excretion within 1-2 days following ingestion).

There are more complicated designs in which three treatments, in different orders, are compared a three-sequence, three-period trial – with two washout periods.

REGRESSION MODELS

Group 1: Period #1 (PEITC; A1) – washout – Period #2 (Placebo; B2)

Group 2: Period #1 (Placebo; B1) – washout – Period #2 (PEITC; A2)

The mean of A1, A2, B1, and B2 can be modeled as follows:

Mean = $(\alpha)(\text{Treatment}) + (\beta)(\text{Order}) + \text{Others}$

Treatment is coded as (0 = Placebo, 1 = PEITC)

Order is coded as (0 = 2nd Period, 1 = 1st Period)

Others include all subjects' characteristics

OUTCOME VARIABLE

From the design:

Group 1: Period #1 (PEITC; A1) – washout – Period #2 (Placebo; B2)

Group 2: Period #1 (Placebo; B1) – washout – Period #2 (PEITC; A2)

Our data analysis could be based on the following “outcome variables” (Treatment - Placebo):

$X1 = A1 - B2$; and $X2 = A2 - B1$

This subtraction will cancel the “within-sequence” effects of all subject-specific factors.

This process will result in two independent samples (often with the same or similar sample size if there are no or minimal dropouts or missing data).

Recall the general model:

$$Y = \begin{bmatrix} \text{Overall} \\ \text{Constant} \end{bmatrix} + \begin{bmatrix} \text{Treatment} \\ \text{Effect} \end{bmatrix} + \begin{bmatrix} \text{Experimental} \\ \text{Error} \end{bmatrix}$$

The subtractions

$X1 = A1 - B2$; and $X2 = A2 - B1$

will cancel not only effects of all subject-specific factors; they cancel the “overall constant” as well, leaving **only two parameters in the means of $X1$ and $X2$:**

Mean of $X1 = [(\alpha)(1)+(\beta)(1)] - [(\alpha)(0)+(\beta)(0)] = \alpha + \beta$

Mean of $X2 = [(\alpha)(1)+(\beta)(0)] - [(\alpha)(0)+(\beta)(1)] = \alpha - \beta$

RESULTING LINEAR MODELS

Design:

Group 1: Period #1 (PEITC; A1) – washout – Period #2 (Placebo; B2)

Group 2: Period #1 (Placebo; B1) – washout – Period #2 (PEITC; A2)

Outcome variables:

$X1 = A1 - B2$; and $X2 = A2 - B1$

Resulting Linear regression models:

$X1$ is normally distributed as $N(\alpha + \beta, \sigma^2)$

$X2$ is normally distributed as $N(\alpha - \beta, \sigma^2)$

In this models, α represents the PEITC supplementation effect ($\alpha > 0$ if and only if PEITC increases the total NNAL) and β represents the period effect ($\beta > 0$ if and only if measurement from period 1 is larger than from period 2).

(1) The X 's do not really need to have normal distributions; the robustness comes from the fact that our analysis will be based on the normal distribution of the sample mean – not of the data, and the sample mean would be almost normally distribution for moderate to large sample sizes (Central Limit Theorem).

(2) Among the three parameters, α represents the PEITC effect and is the primary target, β could be of some interest; we have no in σ^2 (we have to handle it properly to make inferences on α (and β) valid and efficient.

DATA ANALYSIS

Design:

Group 1: Period #1 (PEITC; A1) – washout – Period #2 (Placebo; B2)

Group 2: Period #1 (Placebo; B1) – washout – Period #2 (PEITC; A2)

Outcome variables:

$X1 = A1 - B2$; and $X2 = A2 - B1$

From the model:

$X1$ is normally distributed as $N(\alpha + \beta, \sigma^2)$

$X2$ is normally distributed as $N(\alpha - \beta, \sigma^2)$

Let the sample means and sample variances be defined as usual and n the group size (total sample size is $2n$); then, we can easily prove the followings:

TREATMENT EFFECTS

\bar{x}_1 is *distributed* as normal $N(\alpha + \beta, \frac{\sigma^2}{n})$,

\bar{x}_2 is *distributed* as normal $N(\alpha - \beta, \frac{\sigma^2}{n})$

$$\mathbf{a} = \frac{\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2}{2}$$

\mathbf{a} is *distributed* as normal $N(\alpha, \frac{\sigma^2}{2n})$

ESTIMATION OF PARAMETERS

- (1) **Estimation of Variance.** We can pool data from the two sequences to estimate the common variance σ^2 by s_p^2 – the same pooled estimate used in two-sample t-test.
- (2) **Estimation of Treatment Effect.** Parameter α representing the PEITC effect, the difference between PEITC and the placebo, is estimated by a – the average of the two sample means. Its 95 percent confidence interval is given by:

$$a \pm t_{.975} \frac{s_p}{\sqrt{N}}$$

The t-coefficient goes with (N-2) degrees of freedom; without missing data, $N = 2n$ – total number of subjects.

TESTING FOR TREATMENT EFFECT

Testing for PEITC Treatment Effect: Null hypothesis of “no treatment effects” $H_0: \alpha = 0$ is tested using the “t test”, with $(N-2)$ degrees of freedom:

$$t = \frac{a}{s_p / \sqrt{N}}$$

It's kind of “one-sample t-test” but we use the degree of freedom associated with s_p . Alternatively, one can frame it as a two-sample t-test comparing the mean of X_1 versus the mean of $(-X_2)$ as seen from

X_1 is normally distributed as $N(\alpha + \beta, \sigma^2)$

X_2 is normally distributed as $N(\alpha - \beta, \sigma^2)$

ORDER EFFECTS

\bar{x}_1 is *distributed* as normal $N(\alpha + \beta, \frac{\sigma^2}{n})$,

\bar{x}_2 is *distributed* as normal $N(\alpha - \beta, \frac{\sigma^2}{n})$

$$\mathbf{b} = \frac{\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2}{2}$$

\mathbf{b} is *distributed* as normal $N(\beta, \frac{\sigma^2}{2n})$

ESTIMATION OF PARAMETERS

- (1) **Estimation of Variance.** We can pool data from the two sequences to estimate the common variance σ^2 by s_p^2 – the same pooled estimate used in two-sample t-test.
- (2) **Estimation of Order Effect:** Parameter β representing the order effect, the difference between Period 1 and Period 2, is estimated by b – half the difference of the two sample means. Its 95 percent confidence interval is given by:

$$b \pm t_{.975} \frac{s_p}{\sqrt{N}}$$

The t-coefficient goes with (N-2) degrees of freedom; without missing data, $N = 2n$ – total number of subjects.

TESTING FOR ORDER EFFECT

Testing for Order Effect: Null hypothesis of “no order effects” $H_0: \beta = 0$ is tested using the “t test”, with $(N-2)$ degrees of freedom:

$$t = \frac{b}{s_p / \sqrt{N}}$$

It's kind of “one-sample t-test” but we use the degree of freedom associated with s_p . Alternatively, one can frame it as a two-sample t-test comparing the mean of X_1 versus the mean of X_2 as seen from

X_1 is normally distributed as $N(\alpha + \beta, \sigma^2)$

X_2 is normally distributed as $N(\alpha - \beta, \sigma^2)$

Two-period crossover designs are often used in clinical trials in order to improved sensitivity of the trial by eliminating individual patient effects. They have been popular in dairy husbandry studies, long-term agricultural experiments, bioavailability and bioequivalence studies, nutrition experiments, arthritic and periodontal studies, and educational and psychological studies – where treatment effects are not permanent.

The response could be quantitative but quite often the response variable is binary , e.g. the response is whether or not relief from pain is obtained.

THE DESIGN

Recall the following design; the only difference is that, in this case, the four outcomes A1, A2, B1, and B2 are binary – say 1 if positive response and 0 otherwise:

Group 1: Period #1 (Trt A; A1) – washout – Period #2 (Trt B; B2)

Group 2: Period #1 (Trt B; B1) – washout – Period #2 (Trt A; A2)

The washout periods are optional and the group sizes could be different (due to dropouts) – but not by much.

In general, let Y be the outcome or dependent variable taking on values 0 and 1, and:

$$\pi = \Pr(Y=1)$$

Y is said to have the “**Bernoulli distribution**” (Binomial with $n = 1$). We have:

$$E(Y) = \pi$$

$$Var(Y) = \pi(1 - \pi)$$

Studies would involve some independent variables (treatment, order, etc...)

LOGISTIC REGRESSION

Let π be the probability (also the mean of the Bernoulli distribution) and X a covariate (let consider only one X for simplicity). The common step in the regression modeling process is to relate π and X using the Logistic Regression Model – as follows.

$$\pi = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

$$1 - \pi = \frac{1}{1 + e^{\beta_0 + \beta_1 x}}$$

$$\frac{\pi}{1 - \pi} = e^{\beta_0 + \beta_1 x}$$

$$\log \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 x$$

“Logistic Simple Linear Regression”

THE LOGISTIC MODELS

The Multiple Logistic Models for cross-over design are (J. J. Gart, Biometrika 1969):

$$\Pr(\mathbf{A1} = 1) = \frac{e^{\lambda_i + \alpha + \beta}}{1 + e^{\lambda_i + \alpha + \beta}}; \Pr(\mathbf{B2} = 1) = \frac{e^{\lambda_i - \alpha - \beta}}{1 + e^{\lambda_i + \alpha + \beta}}$$

$$\Pr(\mathbf{B1} = 1) = \frac{e^{\lambda_i - \alpha + \beta}}{1 + e^{\lambda_i - \alpha + \beta}}; \Pr(\mathbf{A2} = 1) = \frac{e^{\lambda_i + \alpha - \beta}}{1 + e^{\lambda_i + \alpha - \beta}}$$

In this models,

(1) λ 's represent the subjects effects varying from subject to subject; could be many terms here.

(2) α represents the new treatment effect, say PEITC supplementation, ($\alpha > 0$ if and only if PEITC is more effective) – **our main interest** - and

(3) β represents the period effect ($\beta > 0$ if and only if a treatment from period 1 is more effective than from period 2).

In this modeling:

- (1) We “code” binary covariates (Treatment and Order) as (+1/-1) instead of (0,1);**
- (2) All subject-specific covariates are lumped together with the Intercept.**

We want to eliminate the subjects' effects in drawing inferences on treatment and order effects. However, we cannot simply do some subtractions like ($X1 = A1 - B2$ and $X2 = A2 - B1$). For a continuous outcome, the difference of two normal variables is distributed as “normal”. But this is not true for the Bernoulli distribution.

A viable alternative is a “Conditional Analysis”, like the formation of the McNemar Chi-square test – used, for example, in the analysis of pair-matched case-control studies.

It was shown by Gart (1969) that optimum inferences about treatment and order effects, regarding subjects effects as nuisance, are based on those subjects with unlike responses in two periods; that are subjects whose pair of outcomes are either (0,1) or (1,0). This is similar to the argument leading to the McNemar Chi-square test.

1: Period 1 (Trt A; A1) – Period 2 (Trt B; B2)

2: Period 1 (Trt B; B1) – Period 2 (Trt A; A2)

The analysis will be conditioned on:

$A1+B2 = 1$, and

$B1+A2 = 1$

$$\Pr(\mathbf{A1} = \mathbf{1}) = \frac{e^{\lambda_i + \alpha + \beta}}{1 + e^{\lambda_i + \alpha + \beta}}; \Pr(\mathbf{B2} = \mathbf{1}) = \frac{e^{\lambda_i - \alpha - \beta}}{1 + e^{\lambda_i + \alpha + \beta}}$$

$$\Pr(\mathbf{B1} = \mathbf{1}) = \frac{e^{\lambda_i - \alpha + \beta}}{1 + e^{\lambda_i - \alpha + \beta}}; \Pr(\mathbf{A2} = \mathbf{1}) = \frac{e^{\lambda_i + \alpha - \beta}}{1 + e^{\lambda_i + \alpha - \beta}}$$

Recall The Bayes Theorem:

$$\begin{aligned} \Pr(\mathbf{A1} = \mathbf{1} | \mathbf{A1} + \mathbf{B2} = \mathbf{1}) &= \frac{\Pr(\mathbf{A1} = \mathbf{1}, \mathbf{B2} = \mathbf{0})}{\Pr(\mathbf{A1} = \mathbf{1}, \mathbf{B2} = \mathbf{0}) + \Pr(\mathbf{A1} = \mathbf{0}, \mathbf{B2} = \mathbf{1})} \\ &= \frac{\Pr(\mathbf{A1} = \mathbf{1})\Pr(\mathbf{B2} = \mathbf{0})}{\Pr(\mathbf{A1} = \mathbf{1})\Pr(\mathbf{B2} = \mathbf{0}) + \Pr(\mathbf{A1} = \mathbf{0})\Pr(\mathbf{B2} = \mathbf{1})} \end{aligned}$$

$$\begin{aligned}\Pr(\mathbf{A1} = 1 \mid \mathbf{A1} + \mathbf{B2} = 1) &= \frac{\Pr(\mathbf{A1} = 1, \mathbf{B2} = 0)}{\Pr(\mathbf{A1} = 1, \mathbf{B2} = 0) + \Pr(\mathbf{A1} = 0, \mathbf{B2} = 1)} \\ &= \frac{\Pr(\mathbf{A1} = 1)\Pr(\mathbf{B2} = 0)}{\Pr(\mathbf{A1} = 1)\Pr(\mathbf{B2} = 0) + \Pr(\mathbf{A1} = 0)\Pr(\mathbf{B2} = 1)} \\ &= \frac{1}{1 + e^{-2(\alpha+\beta)}}\end{aligned}$$

$$\Pr(\mathbf{A2} = 1 \mid \mathbf{B1} + \mathbf{A2} = 1) = \frac{1}{1 + e^{-2(\alpha-\beta)}}$$

$$\Pr(\mathbf{A1} = \mathbf{1} \mid \mathbf{A1} + \mathbf{B2} = \mathbf{1}) = \frac{\mathbf{1}}{\mathbf{1} + \mathbf{e}^{-2(\alpha+\beta)}}$$

$$\Pr(\mathbf{A2} = \mathbf{1} \mid \mathbf{B1} + \mathbf{A2} = \mathbf{1}) = \frac{\mathbf{1}}{\mathbf{1} + \mathbf{e}^{-2(\alpha-\beta)}}$$

$$\Pr(\mathbf{B1} = \mathbf{1} \mid \mathbf{B1} + \mathbf{A2} = \mathbf{1}) = \mathbf{1} - \frac{\mathbf{1}}{\mathbf{1} + \mathbf{e}^{-2(\alpha-\beta)}}$$

$$\Pr(\mathbf{B1} = \mathbf{1} \mid \mathbf{B1} + \mathbf{A2} = \mathbf{1}) = \frac{\mathbf{1}}{\mathbf{1} + \mathbf{e}^{-2(-\alpha+\beta)}}$$

Data #1: Frequencies of subjects with different Outcomes (0,1) and (1,0)

	Treatments (A,B)	Treatments (B,A)
Outcome A=1	y_{a1}	y_{a2}
Outcome B=1	y_{b2}	y_{b1}
Total	n_1	n_2

$$\Pr(A1 = 1 | A1 + B2 = 1) = \frac{1}{1 + e^{-2(\alpha+\beta)}} = p1$$

$$\Pr(A2 = 1 | B1 + A2 = 1) = \frac{1}{1 + e^{-2(\alpha-\beta)}} = p2$$

	Treatments (A,B)	Treatments (B,A)
Outcome A=1	y_{a1}	y_{a2}
Outcome B=1	y_{b2}	y_{b1}
Total	n_1	n_2

Results: With n_1 and n_2 fixed, y_{a1} and y_{a2} are distributed as Binomials $B(n_1, p1)$ and $B(n_2, p2)$

Note: If there are no Order Effects, then $p1 = p2$

Data #1: Frequencies of subjects with different Outcomes (0,1) and (1,0)

	Treatments (A,B)	Treatments (B,A)
Outcome A=1	y_{a1}	y_{a2}
Outcome B=1	y_{b2}	y_{b1}
Total	n_1	n_2

Testing for Order Effects

$H_0: \beta = 0$

Chi-square test; even Fisher's Exact Test

Data #2: The same set of data can also be assembled into a different 2x2 Table

	Treatments (A,B)	Treatments (B,A)
1st Outcome=1	y_{a1}	y_{b1}
2nd Outcome =1	y_{b2}	y_{a2}
Total	n_1	n_2

$$\Pr(A1 = 1 | A1 + B2 = 1) = \frac{1}{1 + e^{-2(\alpha+\beta)}} = p1$$

$$\Pr(B1 = 1 | B1 + A2 = 1) = \frac{1}{1 + e^{-2(-\alpha+\beta)}} = q2$$

	Treatments (A,B)	Treatments (B,A)
1st Outcome=1	y_{a1}	y_{b1}
2nd Outcome=1	y_{b2}	y_{a2}
Total	n_1	n_2

Results: With n_1 and n_2 fixed, y_{a1} and y_{b1} are distributed as Binomials $B(n_1, p1)$ and $B(n_2, q2)$

Note: If no Treatment Effects, then $p1 = q2$

Data #2: The same set of data can also be assembled into a different 2x2 Table

	Treatments (A,B)	Treatments (B,A)
1st Outcome=1	y_{a1}	y_{b1}
2nd Outcome =1	y_{b2}	y_{a2}
Total	n_1	n_2

Testing for Treatment Effects

$H_0: \alpha = 0$

Chi-square test; even Fisher's Exact Test

ESTIMATION OF PARAMETERS

$$\Pr(A1 = 1 | A1 + B2 = 1) = \frac{1}{1 + e^{-2(\alpha+\beta)}} = p1$$

$$\Pr(A2 = 1 | B1 + A2 = 1) = \frac{1}{1 + e^{-2(\alpha-\beta)}} = p2$$

	Treatments (A,B)	Treatments (B,A)
Outcome A=1	y_{a1}	y_{a2}
Outcome B=1	y_{b1}	y_{b2}
Total	n_1	n_2

Results: With n_1 and n_2 fixed, y_{a1} and y_{a2} are distributed as Binomials $B(n_1, p1)$ and $B(n_2, p2)$

Results: With n_1 and n_2 fixed, y_{a1} and y_{a2} are distributed as Binomials $B(n_1, p1)$ and $B(n_2, p2)$

(Conditional) Likelihood Function:

$$\mathbf{L} = \binom{\mathbf{n}_1}{\mathbf{y}_{a1}} \mathbf{p1}^{y_{a1}} (\mathbf{1} - \mathbf{p1})^{y_{b1}} \binom{\mathbf{n}_2}{\mathbf{y}_{a2}} \mathbf{p2}^{y_{a2}} (\mathbf{1} - \mathbf{p2})^{y_{b2}}$$

$$p1 = \frac{1}{1 + e^{-2(\alpha + \beta)}}$$

$$p2 = \frac{1}{1 + e^{-2(\alpha - \beta)}}$$

$$\begin{aligned} \mathbf{L} &= \binom{\mathbf{n}_1}{\mathbf{y}_{a1}} p1^{y_{a1}} (1 - p1)^{y_1} \binom{\mathbf{n}_2}{\mathbf{y}_{a2}} p2^{y_{a2}} (1 - p2)^{y_2} \\ &= \frac{\binom{\mathbf{n}_1}{\mathbf{y}_{a1}} \binom{\mathbf{n}_2}{\mathbf{y}_{a2}} [\exp\{-2y_{b1}(\alpha + \beta) + 2y_{b2}(\beta - \alpha)\}]}{[1 + e^{-2(\alpha + \beta)}]^{n_1} [1 + e^{2(\beta - \alpha)}]^{n_2}} \end{aligned}$$

RESULTS:

Estimates & Standard Errors

$$\hat{\alpha} = \mathbf{a} = \frac{1}{4} \ln \frac{y_{a1} y_{a2}}{y_{b1} y_{b2}}$$

$$\hat{\beta} = \mathbf{b} = \frac{1}{4} \ln \frac{y_{a1} y_{b2}}{y_{a2} y_{b1}}$$

$$\text{Var}(\mathbf{a}) = \text{Var}(\mathbf{b}) = \frac{1}{16} \left[\frac{\mathbf{n}_{ab}}{y_{a1} y_{b1}} + \frac{\mathbf{n}_{ba}}{y_{a2} y_{b2}} \right]$$

DUE AS HOMEWORK

21.1 We conducted a randomized, crossover trial to test whether 3,3'-diindolylmethane (DIM, a metabolite of I3C) excreted in the urine after consumption of raw Brassica vegetables with divergent glucobrassicin concentrations is a marker of I3C uptake from such foods. Twenty-five subjects were fed 50 g of either raw "Jade Cross" Brussels sprouts (high glucobrassicin concentration) or "Blue Dynasty" cabbage (low glucobrassicin concentration) once daily for 3 days. All urine was collected for 24 hours after vegetable consumption each day. After a washout period, subjects crossed over to the alternate vegetable. Data are in file "Brussels Sprouts"; use average of 3 days as our outcome. Estimate & Test for Treatment effects using both the t-test (hand calculation) and SAS program (handout).