simon-5502-14-slides

What did we learn in biostats-2, module 02?

Three models

- All use a continuous dependent (outcome) variables
- All include multiple independent variables
- Multiple linear regression (Week 02)
  - All independent variables are continuous
- Analysis of covariance (Week 03)
  - Mix of continuous and categorical independent variables
- Multi-factor analysis of variance (Week 04)
  - All independent variables are categorical

Speaker notes

Over the next three weeks, you will learn about three different models for predicting a continuous outcome. All of these models will incorporate multiple independent variables.

Why three models?

- Historical precedents
- Different issues
  - Multicollinearity
  - Mediator variable
  - Risk adjustment
  - Moderator variable
  - Interactions

Speaker notes

If it seems arbitrary to have such different names, you are right. Some of this is historical. If you invent something, you get the right to name it, and different inventors for multiple linear regression, analysis of covariance, and multifactor analysis of variance just decided to use different names.

There are also issues that arise far more often in one model than the other. Multicollinearity, for example, is reserved primarily for multiple linear regression. You'll hear discussion about multicollinearity this week as well as mediator variables. Risk adjustment is a topic for next week under analysis of covariance, and you will see discussion of moderator variables and interactions two weeks from now under multifactor analysis of variance. In theory, these issues are not exclusive to one model or another. It's just easier to talk about them within a particular context.

The general linear model

- Single model that unites all three models.
- Use of indicator variables for categorical data
- Not the same as general **IZED** linear model
  - SAS: proc glm versus proc glim
  - R: lm() versus glm()

Speaker notes

It didn’t take long for researchers to discover a common linkage between multiple linear regression, analysis of covariance, and multifactor analysis of variance. In particular, you can treat categorical variables as if they were continuous through the use of indicator variables. You’ll see more discussion about this later.

One bad thing about statistics are all the terms that sound almost the same but mean something quite different. You may have noticed “moderator” and “mediator” from an earlier slide. I can never remember which is which. You’ll also learn about sensitivity and specificity in week 7 (Diagnostic testing) and these are also so easily confused.

But the worst, by far, are the general linear model and the generalized linear model. The IZED added to the end of “general” makes it an entirely different model.

You may see a bit about the generalized linear model in week 7 (logistic regression).

In any case, I have really preferred the use of a single model, the general linear model, in place of multiple linear regression, analysis of covariance, and multifactor analysis of variance. I am a “lumper” and not a “splitter”. For what it is worth, SPSS forces you to use the individual approaches at times instead of the single general linear model.

Arguments for the lm() function

- formula = $dependent - variable$ ~ $independent - variables$
  - $independent - variables$ can be numeric, factors, or strings
- data =
- subset =
- na.action =
  - na.fail
  - na.omit
  - na.exclude
- other arguments

Speaker notes

This week I will show how you fit a multiple linear regression model using the lm function. Ditto for analysis of covariance and multi-factor analysis of variance. Not this week, though.

What did we learn in biostats-2, module 03?

What is a covariate?

- Variable not of direct interest
  - Relationship to outcome is already established
  - Still must be accounted for
- Examples
  - Smoking in a cancer study
  - Gestational in a neonatology study
- A covariate can be continuous or categorical

Speaker notes

This is a repeat of what I said earlier. A covariate is not of direct interest. Testing the relationship of the covariate with the outcome variable is not of great interest. Often this is because the relationship between the covariate and the outcome has already been established.

Even though it is not of direct interest, you feel an obligation to account for the covariate. It plays an important role and failure to measure and adjust for the covariate makes your research appear naive.

A covariate can be continuous or categorical, but more often the former.

What is covariate imbalance?

- Difference in mean value between treatment and control group
  - Often a problem in observational studies
  - Sometimes a problem in randomized studies

Speaker notes

Covariate imbalance is an issue in many research studies. It occurs in a comparison of an outcome between a treatment group and a control group (or maybe an exposure group and a control group). You want the treatment group to be identical to the control group in every way except for the treatment itself. But sometimes a covariate also differs between the treatment group and the control group. If that happens then the outcome could be influenced not by the treatment but by the covariate.

This is often a problem in observational studies.

It can happen in randomized studies at times as well. In theory, randomization will insure balance between the treatment and control group. Patients with large values of the covariate appear in the treatment and control group with equal likelihood. Patients with small values of the covariate appear in the treatment and control group with equal likelihood.

Sometimes, though, randomization doesn’t work. It relies on the law of large numbers and this doesn’t hold for some sample sizes. In particular, studies with less than 20 observations are fairly likely to see covariate imbalance, even with randomization. Even with larger sample sizes, sometimes you just get a bad batch of random numbers.

Why is covariate imbalance an issue?

- Biased estimates
  - Comparing apples to oranges
- Harms study credibility

Speaker notes

If a covariate is imbalanced, it can produce biased estimates. If younger patients are more likely to be in the treatment group, for example, and younger patients tend to have better outcomes, then you don’t know if the outcome variable is influenced by age instead of treatment. The variables are tangled up. This is comparable to the issue of collinearity in multiple linear regression.

Now the bias can go in either direction. Sometimes covariate imbalance will produce an artificial difference in the outcome between treatment and control. But it is also possible that a covariate imbalance masks a difference between the treatment and control.

You will find many times that the covariate imbalance does not produce any serious bias, but you still need to account for it. Failure to control for or to adjust for covariate imbalance will hurt the credibility of your study.

Examples of covariate imbalance

- Age in a study of smoking and Down's syndrome
- Smoking in a study of artillery assignment and sperm count

Speaker notes

A study of Down’s syndrome births had a covariate imbalance between women who smoked during pregnancy, the exposure group, and women who did not smoke during pregnancy, the control group. The average age in the exposure group was much lower than the average age in the control group.

I was involved in a study of lead exposure on male fertility. The exposed group were soldiers who worked on an artillery crew. These big guns could shoot missiles out at great speed, but when the missiles flew out, a cloud of lead dust washed back into the crew. Now, I should note the linkage between lead exposure and fertility is not well established. The control group were soldiers working in an office setting, far away from the big guns. It turns out that the proportion of smokers varied quite a bit between the exposed group and the control. When you are out in the field with explosives all around, smoking was totally banned. Now the artillery crew could still smoke while off duty, but the office workers had more opportunities to smoke on and off duty. This was in the 1990s before workplace bans on smoking were very common.

Now, I also have to admit that the link between smoking and male fertility is also not well established. But there was enough evidence to at least raise some concern about the covariate imbalance.

Covariate imbalance versus confounding

- Covariate imbalance is simpler
- Confounder definition relies on causation arguments

Speaker notes

I’m in the minority here, because I like the term “covariate imbalance” and most other researchers talk about “confounding.”

I like talking about covariate imbalance because it is simpler. Calculate the mean of the covariate in the treatment group and compare it to the mean of the covariate in the control group.

The definition of confounder involves complex arguments about causation. When I find myself needing to use a term like “confounder” or “confounding”, I find myself qualifying it. I’ll say “potential confounder” or “possible confounder.”

Preventing covariate imbalance

- Randomization
- Matching
- Stratification

Speaker notes

If you can, you should try to prevent covariate imbalance during the design of a research study. Randomization, if you can do it, is a great way to reduce the risk of covariate imbalance. It doesn’t always work, but it does work quite often. One of the nicest things about randomization is that it prevents covariate imbalance among those variables that you have measured, but it also prevents covariate imbalance among covariates that you didn’t measure, either because it was difficult or impossible to measure them.

Matching is another research strategy that helps to prevent covariate imbalance. For every treatment subject of a certain age, find a control subject with a closely matched age. Make sure that males are matched with other males and females are matched with other females.

You might choose other variables to match on. Just make sure that the covariates that you match on are important influences on the outcome. And don’t choose so many variables to match on that you have difficulty finding good matches.

Stratification also can help. Divide your patients into broad categories such as young, middle-aged, and older. Then randomly assign to treatment and control within these broad categories.

Adjusting for covariate imbalance

- Propensity score models
- Analysis of covariance

Speaker notes

If you did not or could not design the study to minimize covariate imbalance, you still have the option of adjusting for covariate imbalance.

Later on in this semester, you will learn about propensity score models. This week, you will learn about analysis of covariance.

Variables cannot be in the causal pathway

- Fixed at time of randomization
- Temporally preceding exposure
- Example: bottles given during a breast feeding study

Speaker notes

Covariates must be fixed and in place at the time when you flip the coin to choose whether they get into the treatment group or the control group.

If you are studying an exposure, then the covariate must be a variable that precedes the exposure in time.

Variables that are intermediate between the treatment/exposure and the assessment of the outcome are said to be in the causal pathway.

If you adjust for variables in the causal pathway, that may reduce or even eliminate the effect of your treatment or exposure.

I saw an example of this in a study I helped with. It was an examination of breast feeding patterns in pre-term infants. It is difficult to maintain breast feeding in a pre-term infant because the mother goes home from the hospital before the baby does. A rule of thumb is that the number of weeks that a pre-term baby has to stay in the hospital is roughly equal to the number of weeks early that the baby arrived.

In this study, mothers of newborn infants were randomly assigned to a treatment or control group. In both groups, mothers were encouraged to breast feed when they were in the hospital visiting their baby. They were given breat pumps to collect milk when they were at home. The difference was that in the control group, infant were fed by bottle when the mothers were not around. In the treatment group, the infants were fed through an nasogastral (ng) tube. The intervention was designed to avoid having the infants becoming habituated to the artificial nipple of the bottle and then having trouble latching onto the mother’s nipple.

The intervention was quite successful. There were a few minor mistakes made during the experiement, though. Sometimes an infant in the treatment group was given a few bottles at the hospital instead of being fed exclusively through the ng tube. That’s not surprising and not too much of a cause for concern.

The researchers did measure the number of bottles given in the birth hospital in both the treatment and control group and the average number of bottles was quite a bit lower than the treatment group, even though it was still a bit above zero.

I decided, on a lark, to put the number of bottles received in as a covariate in my analysis. To my initial horror, the effect of the treatment disappeared when you adjusted for the number of bottles.

But I quickly realized that this, if anything, reinforced the conclusions of the study. The number of bottles given was part of the causal pathway. It occured after random assignment, and the intervention was deliberately designed to influence this intermediate variable. So the adjustment actually helped to explain why the intervention worked.

Adjusting for baseline measurements

- Baseline = measurements prior to intervention
  - Done to improve precision
  - Can use baseline as a covariate
- Change score is an alternative
  - Also known as difference in differences (DID) model
  - Possible regression to the mean

Speaker notes

Many research studies include baseline measurements, measurement of the same outcome measure that you plan to use to compare the treatment and control groups. It often helps to make an assessment of the outcome BEFORE you implement any intervention. This is done mostly to improve precision, but it can also help control bias. If the intervention and control groups differ on the baseline measure, that is an indiciation that one group is more seriously ill at the start of the research than the other group.

You can consider the baseline value to be a covariate. Your outcome at baseline certainly does have some influence on your outcome at the end of the study. But there is no direct interest in the baseline measure itself.

A common alternative to using the baseline measure as a covariate is to compute change scores, the difference between the baseline measure of the outcome and the measure of the outcome at the end of the study. This is measuring the relative decline or improvement in health.

Use of change scores or difference in difference models is controversial. I like this approach but most of the research community is criticizes this choice. One criticism is that regression to the mean can possibly mess up the change score analysis.

Regression to the mean is the tendency for extremely low scores at one time point often are not quite as extreme when they are measured again, even if there is no change or intevention going on. Similarly extremely high scores at one time point often are not as extreme when measured again.

What did we learn in module 04?

Mathematical model, 1

- Decompose μ_{ij} into $\mu + \alpha_i + \beta_j$
  - α_i is the deviation for the ith level of first factor
  - β_j is the deviation for the jth level of second factor
  - Require $\alpha_1 = 0$ and $\beta_1 = 0$
  - μ is the mean for the reference levels

Speaker notes

The mathematical model for two factor analysis of variance is a bit more complex than a single factor analysis of variance. You have a mean at the reference levels, mu, You also have deviations from the overall mean associated with the first factor (alpha), deviations from the overall mean associated with the second factor (beta)

There are a total of a and b categories for the two categorical independent variables.

Mathematical model, 2

- $Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$
  - i=1,…,a levels of the first categorical variable
  - j=1,…,b levels of the second categorical variable
  - k=1,…,n replicates with first and second categories
- Note: $\mu, \alpha_i, \beta_j, \epsilon_{ijk}$ are population values

Speaker notes

The mathematical model for two factor analysis of variance is a bit more complex than a single factor analysis of variance. You have an overall mean, mu, and deviations from the overall mean associated with the first factor (alpha), deviations from the overall mean associated with the second factor (beta) and an error term (epsilon).

There are a total of a and b categories for the two categorical independent variables.

Mathematical model, 3

- $H_0 : \ \alpha_i = 0$ for all i

- $H_0 : \ \beta_j = 0$ for all j

Speaker notes

There are two hypotheses. The first, testing that all the alphas equal zero is effectively testing whether the first factor has no impact on the outcome. Testing that all the betas equal zero is effectively testing whether the second factor has no impact on the outcome.

Parameter estimates for the two factor model

```
# A tibble: 14 × 5
   term        estimate std.error statistic p.value
   <chr>          <dbl>     <dbl>     <dbl> <glue>
 1 (Intercept)    4.72      1.50      3.14  p = 0.005
 2 MoonDuring     2.5       0.984     2.54  p = 0.019
 3 MoonAfter      0.542     0.984     0.550 p = 0.588
 4 MonthSep       4.03      1.97      2.05  p = 0.053
 5 MonthOct       3.73      1.97      1.90  p = 0.071
 6 MonthNov       3.70      1.97      1.88  p = 0.073
 7 MonthDec       2.63      1.97      1.34  p = 0.195
 8 MonthJan       5.03      1.97      2.56  p = 0.018
 9 MonthFeb       6.93      1.97      3.52  p = 0.002
10 MonthMar       8.57      1.97      4.35  p < 0.001
11 MonthApr      13.0       1.97      6.61  p < 0.001
12 MonthMay       8.60      1.97      4.37  p < 0.001
```

Analysis of variance table comparing the two factor model to the null model

```
# A tibble: 2 × 7
  term                    df.residual   rss    df sumsq statistic p.value
  <chr>                         <dbl> <dbl> <dbl> <dbl>     <dbl> <glue>
1 Admission ~ 1                    35  625.    NA   NA       NA    <NA>
2 Admission ~ Moon + Month         22  128.    13  497.       6.58 p < 0.001
```

Speaker notes

Analysis of variance table comparing the two factor model to the one factor model

```
# A tibble: 2 × 7
  term                    df.residual   rss    df sumsq statistic p.value
  <chr>                         <dbl> <dbl> <dbl> <dbl>     <dbl> <glue>
1 Admission ~ Moon                 33  583.    NA   NA       NA    <NA>
2 Admission ~ Moon + Month         22  128.    11  456.       7.13 p < 0.001
```

R-squared values

```
# A tibble: 3 × 3
  model  r.squared deviance
  <glue>     <dbl>    <dbl>
1 m1         0          625.
2 m2         0.0664     583.
3 m3         0.795      128.
```

Tukey post hoc test

```
# A tibble: 3 × 7
  term  contrast      null.value estimate conf.low conf.high adj.p.value
  <chr> <chr>              <dbl>    <dbl>    <dbl>     <dbl> <chr>
1 Moon  After-Before           0    0.542  -1.93        3.01 0.847
2 Moon  During-Before          0    2.50    0.0280      4.97 0.047
3 Moon  During-After           0    1.96   -0.514       4.43 0.138
```

Speaker notes

Use the Tukey posthoc test because the sample sizes are equal across the moon phases. The results are a bit ambiguous because before and after are not statistically different, after and during are not statistically different but before and during are statistically different. This is probably due to a lack of precision and an extra year’s worth of data would help quite a bit.

The analogy I use is travel time. My wife and I live in Leawood. Our son lives in Lee’s Summit. A repair shop we all use is in Olathe. It is not far from Leawood to Olathe. It is not far from Leawood to Lee’s Summit. But it is far from Lee’s Summit to Olathe.

What did we learn in module 05?

Mathematical model, 1

- $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
  - i=1,…,a j=1,…,b, k=1,…,n
- If 1 is the reference category
  - $\alpha_1 = 0$
  - $\beta_1 = 0$
  - $(\alpha\beta)_{1j} = 0$
  - $(\alpha\beta)_{i1} = 0$

Speaker notes

You may see papers or books that present the mathematical model for an interaction. The model I present is a balanced model with the first category having levels one through a, the second category having levels one through b and for each combination of categories there are n observations.

If you set the first level as the reference category for each category, then you need to set some of these parameters to zero.

Mathematical model, 2

- $SS_A = \Sigma_i nb(\bar{Y}_{i..} - \bar{Y}_{...})^2$
- $SS_B = \Sigma_i na(\bar{Y}_{.j.} - \bar{Y}_{...})^2$
- $SS_{AB} = \Sigma_i \Sigma_j n(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$
- $SS_E = \Sigma_i \Sigma_j \Sigma_k (Y_{ijk} - \bar{Y}_{ij.})^2$
- $SS_T = \Sigma_i \Sigma_j \Sigma_k (Y_{ijk} - \bar{Y}_{...})^2$

Speaker notes

The dot notation may be a bit confusing until you get used to it, but $\bar{Y}_{i\cdot\cdot}$ is the average within the ith group, averaging across the subscripts j and k. $\bar{Y}_{\cdot j\cdot}$ is the average within the ith group, averaging across the subscripts i and k. $\bar{Y}_{ij\cdot}$ is the average within the combination of the ith group and the jth group, averaging across the subscript k. Finally, $\bar{Y}_{\cdot\cdot\cdot}$ is an overall mean and the average across all three subscripts.

Test for an interaction

- SS_{AB} has (a-1)(b-1) degrees of freedom
- SS_E has ab(n-1) degrees of freedom
- Accept H_0 if $F = \frac{MS_{AB}}{MS_E}$ is close to one
  - In R, fit a model without an interaction
  - Compare to a model with interaction
  - Using the anova function

Speaker notes

The formal test for an interaction uses an F ratio and you accept the null hypothesis if that F ratio is close to one. You would reject the null hypothesis if the F ratio is much larger than one.

It is not easy to get R to display all the sums of squares and mean squares that I defined above. Instead, compute two models-one without an interaction and one with an interaction. Compare those two models using the anova function.

What did we learn in module 06?

Comparing two binary outcomes

- Is there a difference in the proportion of deaths between male passengers and female passengers on the Titanic?
- Is there difference in the proportion of patients finishing the full three doses of HPV vaccine between Black women and White women?
- Does using a ng tube for feeding in pre-term infants increase the probability of successful breast feeding at six months?

Speaker notes

Most of the statistics, you have seen so far involve a continuous outcome. You can, however, use a binary outcome. Here are three examples comparing a binary outcome between two groups.

Other comparisons involving a binary outcome

- Is there are difference in the proportion of deaths between first class, second class, and third class passengers?
- Does age influence the proportion of women finishing the full three doses of HPV vaccine?
- Controlling for the mother's age, does using a ng tube for feeding in pre-term infants increase the probability of successful breast feeding at six months?

Speaker notes

Here are some more complex comparisons involving a binary outcome. The first example involves a comparison of three proportions, not two. The next example involves a continuous predictor of a binary outcome. The final example involves a comparison of binary outcomes in two groups, but controlling for a third variable.

Hypothesis framework

- $H_0: \pi_1 = \pi_2$
- $H_1: \pi_1 = \pi_2$
- Compute $\hat{p}_1$ and $\hat{p}_2$ from samples
- Accept H_0 if $\hat{p}_1 - \hat{p}_2$ is close to zero.
  - $T = (\hat{p}_1 - \hat{p}_2)/s.e.$
  - 95% CI: $(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} s.e.$

Speaker notes

The hypothesis to test two proportions uses the symbols π_1 and π_2 to represent the proportions in a population.

The Titanic dataset

```
Rows: 1,313
Columns: 5
$ Name     <chr> "Allen, Miss Elisabeth Walton", "Allison, Miss Helen
Loraine"…
$ PClass   <chr> "1st", "1st", "1st", "1st", "1st", "1st", "1st", "1st",
"1st"…
$ Age      <dbl> 29.00, 2.00, 30.00, 25.00, 0.92, 47.00, 63.00, 39.00, 58.00,
…
$ Sex      <chr> "female", "female", "male", "female", "male", "male",
"female…
$ Survived <dbl> 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1,
1…
```

Counts and percentages

```
        Survived
Sex      Yes  No
  female 308 154
  male   142 709
```

```
        Survived
Sex             Yes        No
  female 0.6666667 0.3333333
  male   0.1668625 0.8331375
```

Test for difference in proportions

```
# A tibble: 1 × 9
  estimate1 estimate2 statistic  p.value parameter conf.low conf.high method
      <dbl>     <dbl>     <dbl>    <dbl>     <dbl>    <dbl>     <dbl> <chr>
1     0.667     0.167      332. 3.43e-74         1    0.450     0.550 2-sample
…
# ℹ 1 more variable: alternative <chr>
```

Chi-square test of independence, 1 of 2

- Equivalent to test of two proportions
- Lay out data in two by two table

	$No\ event$	$Event$
$Treatment$	O_{11}	O_{12}
$Control$	O_{21}	O_{22}

Speaker notes

Chi-square test of independence, 2 of 2

	$No\ event$	$Event$
$Treatment$	$E_{11} = n_1(1 - \hat{p}_{.})$	$E_{12} = n_1\hat{p}_{.}$
$Control$	$E_{21} = n_2(1 - \hat{p}_{.})$	$E_{22} = n_2\hat{p}_{.}$

- $X^2 = \Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

Speaker notes

Expected counts for Titanic

Observed counts

```
        Survived
Sex      Yes  No
  female 308 154
  male   142 709
```

Expected counts

```
        Survived
Sex           Yes       No
  female 158.3397 303.6603
  male   291.6603 559.3397
```

Chisquare test for Titanic

```
# A tibble: 1 × 4
  statistic  p.value parameter method
      <dbl>    <dbl>     <int> <chr>
1      332. 3.43e-74         1 Pearson's Chi-squared test
```

Odds ratio calculation

	No event	Event	Odds
Group1	a	b	
Group2	c	d	

- Odds for group 1 = b/a
- Odds for group 2 = d/c
- Odds for group 1 = $\frac{d/c}{b/a} = \frac{ad}{bc}$
- s.e.(log or) = $\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$

Titanic data

	Survived	Died	Total
Female	308	154	462
Male	142	709	851
Total	450	863	1,313

Speaker notes

Titanic data, odds of death

	Survived	Died	Total	Odds	
Female	308	154	462	2	to 1 against
Male	142	709	851	4.993	to 1 in favor
Total	450	863	1,313		

Odds ratio = 4.993 / 0.5 = 9.986

Speaker notes

Clearly, a male passenger on the Titanic was more likely to die than a female passenger. But how much more likely? You can compute the odds ratio or the relative risk to answer this question.

The odds ratio compares the relative odds of death in each group. For females, the odds were exactly 2 to 1 against dying (154/308=0.5). For males, the odds were almost 5 to 1 in favor of death (709/142=4.993). The odds ratio is 9.986 (4.993/0.5). There is a ten fold greater odds of death for males than for females.

Odds ratio for survival by sex

```
$data
        Survived
Sex      Yes  No Total
  female 308 154   462
  male   142 709   851
  Total  450 863  1313

$measure
        odds ratio with 95% C.I.
Sex      estimate    lower    upper
  female 1.000000       NA       NA
  male   9.956188 7.662525 13.00928

$p.value
        two-sided
```

What did we learn in module 07?

What is a diagnostic test?

- Indication of disease
  - Rapid, convenient, and/or inexpensive
- Gold standard
  - Indication of same disease
  - Slow, inconvenient, and/or expensive

Speaker notes

A diagnostic test is a procedure which gives a rapid, convenient and/or inexpensive indication of whether a patient has a certain disease.

In research settings, a diagnostic test is often compared to a gold standard. This is a measurement that is slower, less convenient, or more expensive than the diagnostic test, but which also gives a definitive indication of disease status. The gold standard might involve invasive procedures like a biopsy or could mean waiting for several years until the disease status becomes obvious.

Examples of diagnostic tests, 1 of 2

- Yale-Brown obsessive-compulsive scale
  - Do you often feel sad or depressed?
- SCOFF questionnaire
  - Five yes/no questions
  - Two or more yes responses

Speaker notes

Some examples of diagnostic tests are:

The Yale-Brown obsessive-compulsive scale, a simple yes/no answer to the following question: Do you often feel sad or depressed? In a study of stroke patients at the Royal Liverpool and Broadgreen University Hospitals (BMJ 2001; 323: 1159), this test was shown to perform well compared to a more complex measure, the Montgomery Asberg depression rating scale.

The SCOFF questionnaire asks five yes/no questions to determine whether a patient has an eating disorder.

- Do you ever make yourself sick because you feel uncomfortably full?
- Do you worry you have lost control over how much you eat?
- Have you recently lost more than one stone in a 3 month period?
- Do you believe yourself to be fat when others say you are thin?
- Would you say that food dominates your life?

Two or more yes answers is considered a positive test. In a study of 341 consecutive patients at two general practices in southwest London (BMJ 2002; 325: 755-756), these patients were given the SCOFF questionnaire and then a formal interview based on Diagnostic and Statistical Manual of Mental Disorders, (fourth edition). The interview lasted 10-15 minutes and the interviewer did not know that score on the SCOFF questionnaire. The SCOFF questionnaire produced results that were comparable to the formal interview.

More examples

- Rectal bleeding as a sign of colorectal cancer
- Electrocardiogram, QTc dispersion

Speaker notes

Patients with rectal bleeding will sometimes develop colorectal cancer. In a study at a network of practices in Belgium (BMJ 2000; 321; 998-999), 386 patients presented with rectal bleeding between 1993 and 1994. After following these patients for 18 to 30 months, only a few developed colorectal cancer.

A standard electrocardiogram can produce a measure called QTc dispersion. In a study of 49 patients with peripheral vascular disease (BMJ 1996; 312: 874-878), all were assessed for their QTc dispersion values. These patients were then followed for 52 to 77 months. During this time, there were 12 cardiac deaths, 3 non-cardiac deaths, and 34 survivors. A value of QTc dispersion of 60 ms or more did quite well in predicting cardiac death.

What did we learn in module 08?

Survival analysis

- Time to event models
  - Death
  - Relapse
  - Rehospitalization
  - Failure of medical device
  - Pregnancy
- Not every patient experiences the event
  - These are censored observations

Speaker notes

Survival analysis models are more properly called time to event models. You follow a group of patients from a certain time point and note the amount of time until they die.

Or the amount of time until they relapse. Or the amount of time until they need to be rehospitalized. Or the amount time until a medical device that you implanted in them fails.

I should note that while almost all of the events in a time to event model are bad, there are a few exceptions. In a study of couples with fertility problems, you might use a time to event model to study the time to pregnancy, a very happy event for any couple with fertility problems.

Mortality is the context under which time to event models were derived, so the term survival analysis has been used even when the event is different.

A key feature of survival analysis is that not every patient experiences the event. You should be glad that not everyone that you recruit for a clinical trial dies, but this adds a layer of complexity to the analysis.

The reasons for not experiencing the event can vary. Everyone dies, but maybe the event is death from cancer and if you patient gets hit by a bus, that patient does not experience the event. If the event is death from any cause, you still have to end the study in some time frame, and not everyone will die in that time frame. Not every patient gets rehospitalized, not every patient relapses, not every medical device fails.

A patient may drop out of a study, and you no longer are able to tell from that point onward whether that patient would have experienced the event sometime during the rest of your study.

When your patient does not die during the study, this is not a missing value. You have partial information. You know that the patient was alive for a certain amount of time. When you end the study within a certain time frame, you know that your patient dies at a time beyond the end date of your study. If your patient drops out after six months, you know that the patient survived for more than six months.

First fruit fly experiment, 1

```
data_dictionary: fly1.txt
description: |
  This dataset provides a simple example of what survival and censoring. It
provides an inuitive explanation of estimation of survival probabilities.
vars:
  day:
    label: Time until death
    unit: days
```

Speaker notes

The following data represents survival time for a group of fruit flies and is a subset of a larger data set found at the Data and Story Library (DASL). The data set has been slightly modified to simplify some of these explanations.

There are 25 flies in the sample, with the first fly dying on day 37 and the last fly dying on day 96.

First fruit fly experiment, 2

37, 40, 43, 44, 45, 47, 49, 54, 56, 58, 59, 60, 61, 62, 68, 70, 71, 72, 73, 75, 77, 79, 89, 94, 96

Speaker notes

If you wanted to estimate the survival probability for this data, you would draw a curve that decreases by 4% (1/25) every time a fly dies.

First fruit fly experiment, 3

	day	p
1	37	96%
2	40	92%
3	43	88%
4	44	84%
5	45	80%
6	47	76%
7	49	72%
8	54	68%
9	56	64%

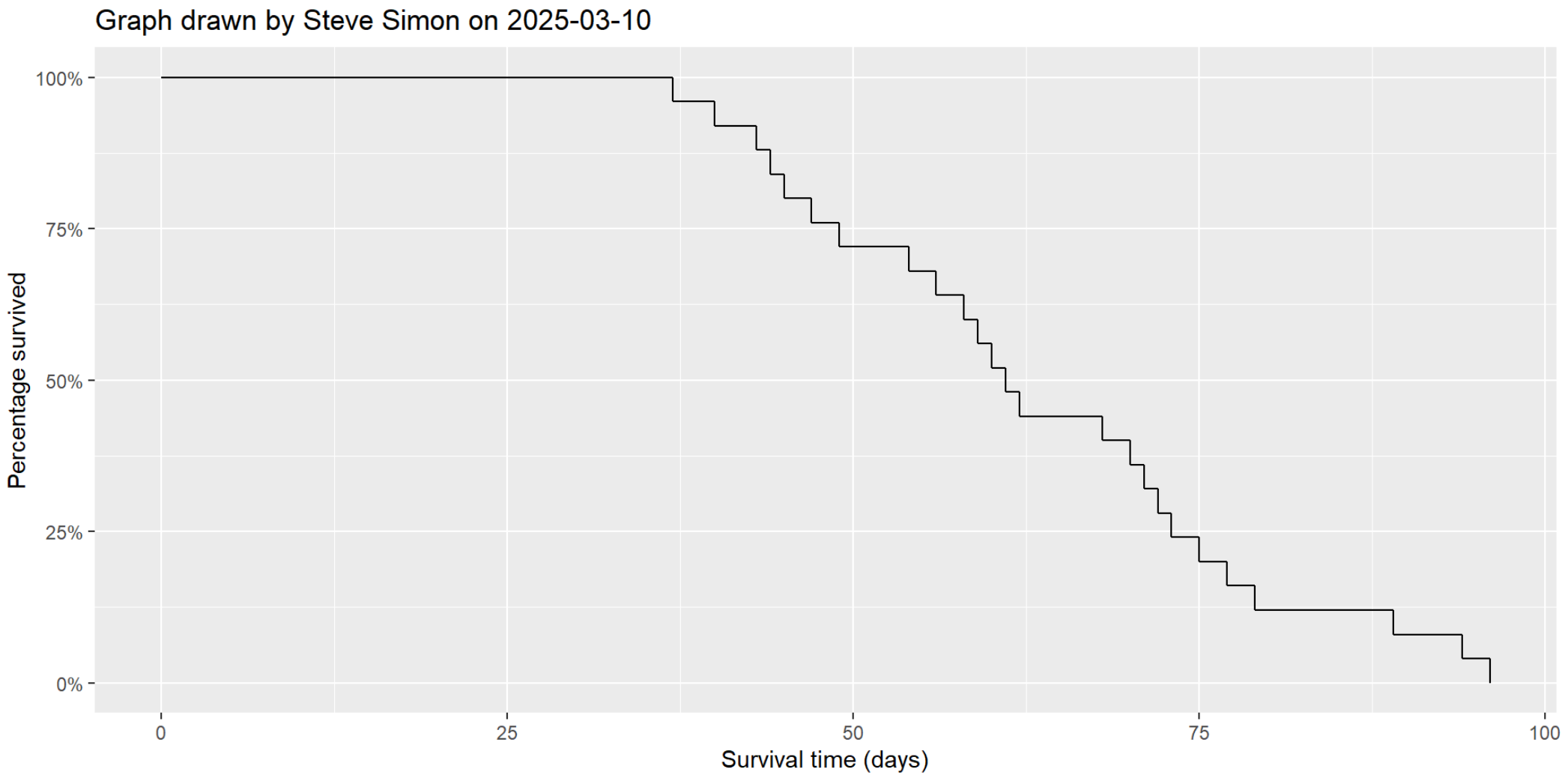
	day	p
10	58	60%
11	59	56%
12	60	52%
13	61	48%
14	62	44%
15	68	40%
16	70	36%
17	71	32%
18	72	28%

	day	p
19	73	24%
20	75	20%
21	77	16%
22	79	12%
23	89	8%
24	94	4%
25	96	0%

Speaker notes

The probability of survival drops by 4% (1/25) at each day of death.

First fruit fly experiment, 4



Speaker notes

Here’s a graph of these probabilities over time.

By tradition and for some rather technical reasons, you should use a stair step pattern rather than a diagonal line to connect adjacent survival probabilities.

Second fruit fly experiment, 1

```
37, 40, 43, 44, 45, 47, 49, 54, 56, 58, 59, 60, 61, 62, 68, ??, ??, ??, ??,
??, ??, ??, ??, ??, ??, ??
```

Speaker notes

Now let’s alter the experiment. Suppose that totally by accident, a technician leaves the screen cover open on day 70 and all the flies escape. This includes the fly who was going to die on the afternoon of the 70th day anyway. Oh the sadness of it all; the poor fly has the briefest of tastes of freedom then ends up shriveled up on a window sill.

You’re probably worried that the whole experiment has been ruined. But don’t be so pessimistic. You still have complete information on survival of the fruit flies up to their 70th day of life.

Second fruit fly experiment, 2

	day	event
1	37	1
2	40	1
3	43	1
4	44	1
5	45	1
6	47	1
7	49	1
8	54	1
9	56	1

	day	event
10	58	1
11	59	1
12	60	1
13	61	1
14	62	1
15	68	1
16	70	0
17	70	0
18	70	0

	day	event
19	70	0
20	70	0
21	70	0
22	70	0
23	70	0
24	70	0
25	70	0

Speaker notes

Here’s how you would code the data for importing into SPSS or any other software.

Second fruit fly experiment, 3

	day	event	p
1	37	1	96%
2	40	1	92%
3	43	1	88%
4	44	1	84%
5	45	1	80%
6	47	1	76%
7	49	1	72%
8	54	1	68%
9	56	1	64%

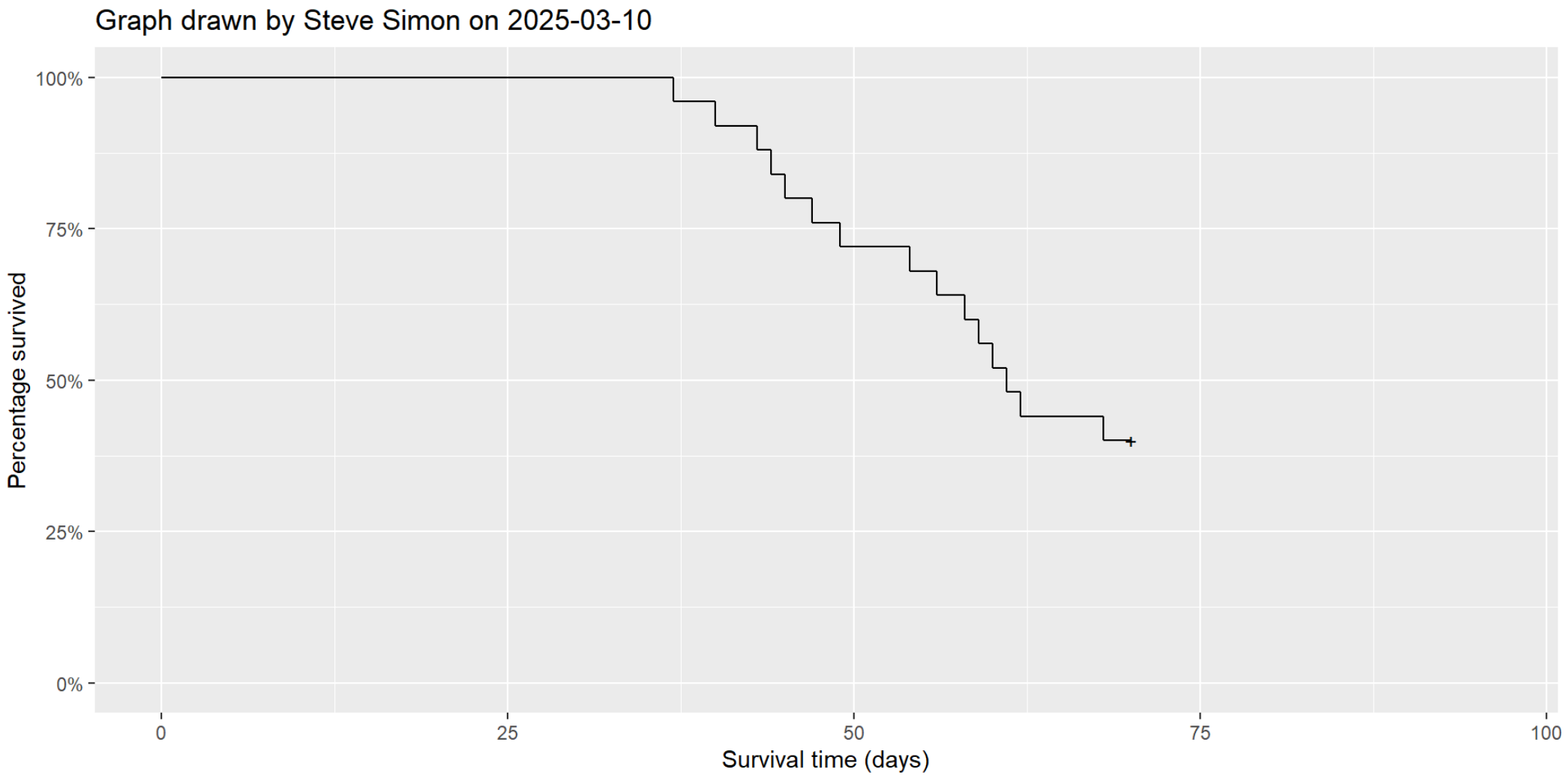
	day	event	p
10	58	1	60%
11	59	1	56%
12	60	1	52%
13	61	1	48%
14	62	1	44%
15	68	1	40%
16	70	0	
17	70	0	
18	70	0	

	day	event	p
19	70	0	
20	70	0	
21	70	0	
22	70	0	
23	70	0	
24	70	0	
25	70	0	

Speaker notes

We clearly have enough data to make several important statements about survival probability. For example, the median survival time is 61 days because roughly half of the flies had died before this day.

Second fruit fly experiment, 4



Speaker notes

Here is a graph of the survival probabilities of the second experiment. This graph is identical to the graph in the first experiment up to day 70 after which you can no longer estimate survival probabilities.

By the way, you might be tempted to ignore the ten flies who escaped. But that would seriously bias your results. All of these flies were tough and hardy flies who lived well beyond the median day of death. If you pretended that they didn't exist, you would seriously underestimate the survival probabilities. The median survival time, for example, of the 15 flies who did not escape, for example, is only 54 days which is much smaller than the actual median.

Third fruit fly experiment, 1

37, 40, 43, 44, 45, 47, 49, 54, 56, 58, 59, 60, 61, 62, 68, ??, 71, ??, ??, 75, ??, ??, 89, ??, 96

Speaker notes

Let’s look at a third experiment, where the screen cover is left open and all but four of the remaining flies escape. It turns out that those four remaining flies who didn’t bug out will allow us to still get reasonable estimates of survival probabilities beyond 70 days.

Third fruit fly experiment, 2

	day	event
1	37	1
2	40	1
3	43	1
4	44	1
5	45	1
6	47	1
7	49	1
8	54	1
9	56	1

	day	event
10	58	1
11	59	1
12	60	1
13	61	1
14	62	1
15	68	1
16	70	0
17	71	1
18	70	0

	day	event
19	70	0
20	75	1
21	70	0
22	70	0
23	89	1
24	70	0
25	96	1

Speaker notes

Here is how you would code the data for importing into SPSS.

Third fruit fly experiment, 3

	day	event	p
1	37	1	96%
2	40	1	92%
3	43	1	88%
4	44	1	84%
5	45	1	80%
6	47	1	76%
7	49	1	72%
8	54	1	68%
9	56	1	64%

	day	event	p
10	58	1	60%
11	59	1	56%
12	60	1	52%
13	61	1	48%
14	62	1	44%
15	68	1	40%
16	70	0	
17	71	1	30%
18	70	0	

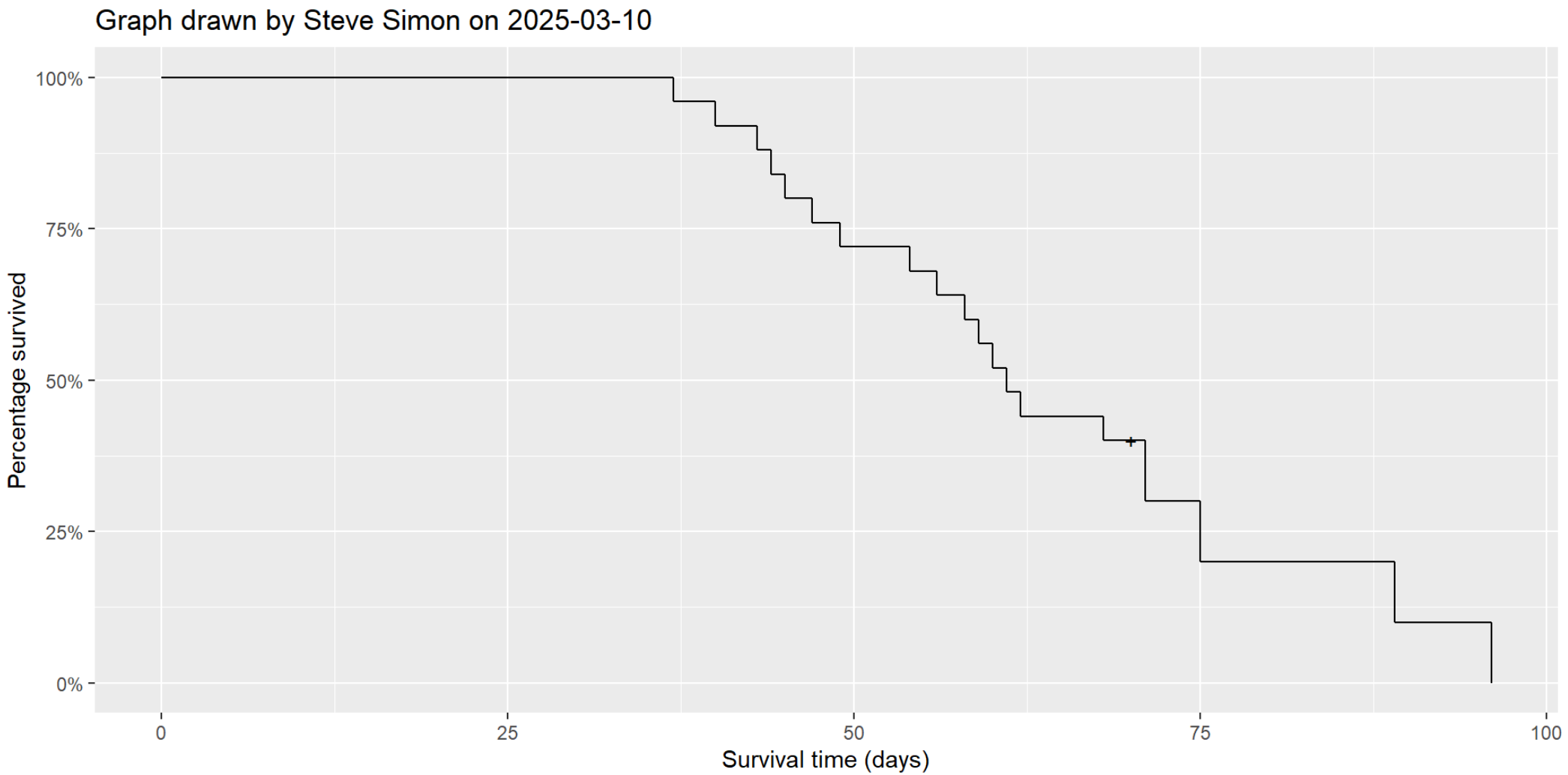
	day	event	p
19	70	0	
20	75	1	20%
21	70	0	
22	70	0	
23	89	1	10%
24	70	0	
25	96	1	0%

Speaker notes

What you do with the six escaped flies is to allocate their survival probabilities equally among the four flies who didn’t bug out. This places a great responsibility among each of those four remaining flies since each one is now responsible for 10% of the remaining survival probability, their original 4% plus 6% more which represents a fourth of the 24% survival probability that was lost with the six escaping flies.

Another way of looking at this is that the six flies who escaped influence the denominator of the survival probabilities up to day 70 and then totally drop out of the calculations for any further survival probabilities. Because the denominator has been reduced, the jumps at each remaining death are much larger.

Third fruit fly experiment, 4



Speaker notes

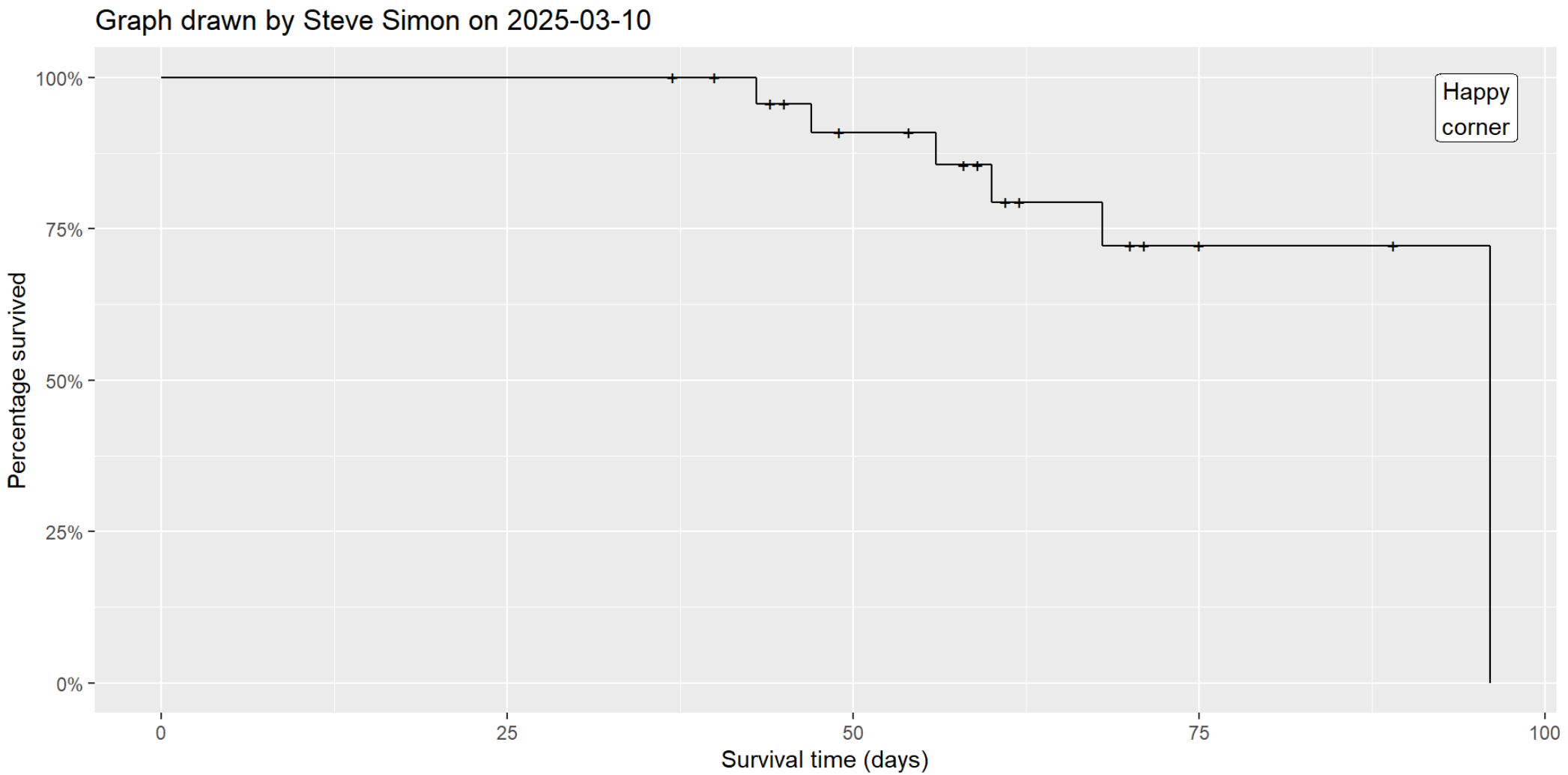
Here is a graph of the survival probability estimates from the third experiment.

These survival probabilities differ only slightly from the survival probabilities in the original experiment. This works out because the mechanism that caused us to lose information on six of the fruit flies was independent of their ultimate survival.

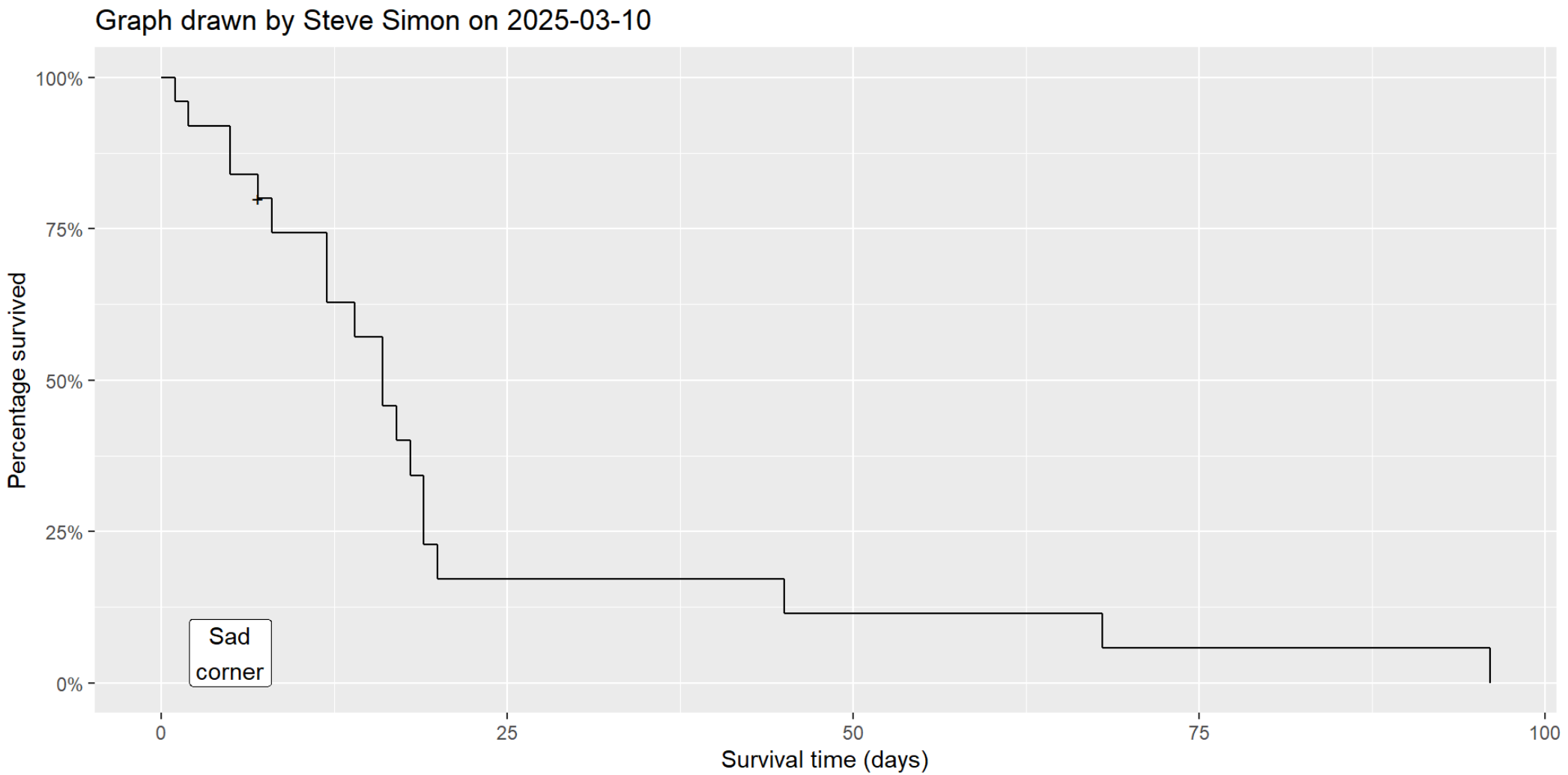
If the censoring mechanism were somehow related to survival prognosis, then you would have the possibility of serious bias in your estimates. Suppose for example, that only the toughest of flies (those with the most days left in their short lives) would have been able to escape. The flies destined to die on days 70, 71, 72, and 73, were already on their deathbeds and unable to fly at all, much les make a difficult escape. Then these censored values would not be randomly interspersed among the remaining survival times, but would constitute some of the larger values. But since these larger values would remain unobserved, you would underestimate survival probabilities beyond the 70th day.

This is known as informative censoring, and it happens more often that you might expect. Suppose someone drops out of a cancer mortality study because they are abandoning the drugs being studied in favor of laetrile treatments down in Mexico. Usually, this is a sign that the current drugs are not working well, so a censored observation here might represent a patient with a poorer prognosis. Excluding these patients would lead to an overestimate of survival probabilities.

Interpreting Kaplan-Meier plots, 1



Interpreting Kaplan-Meier plots, 2



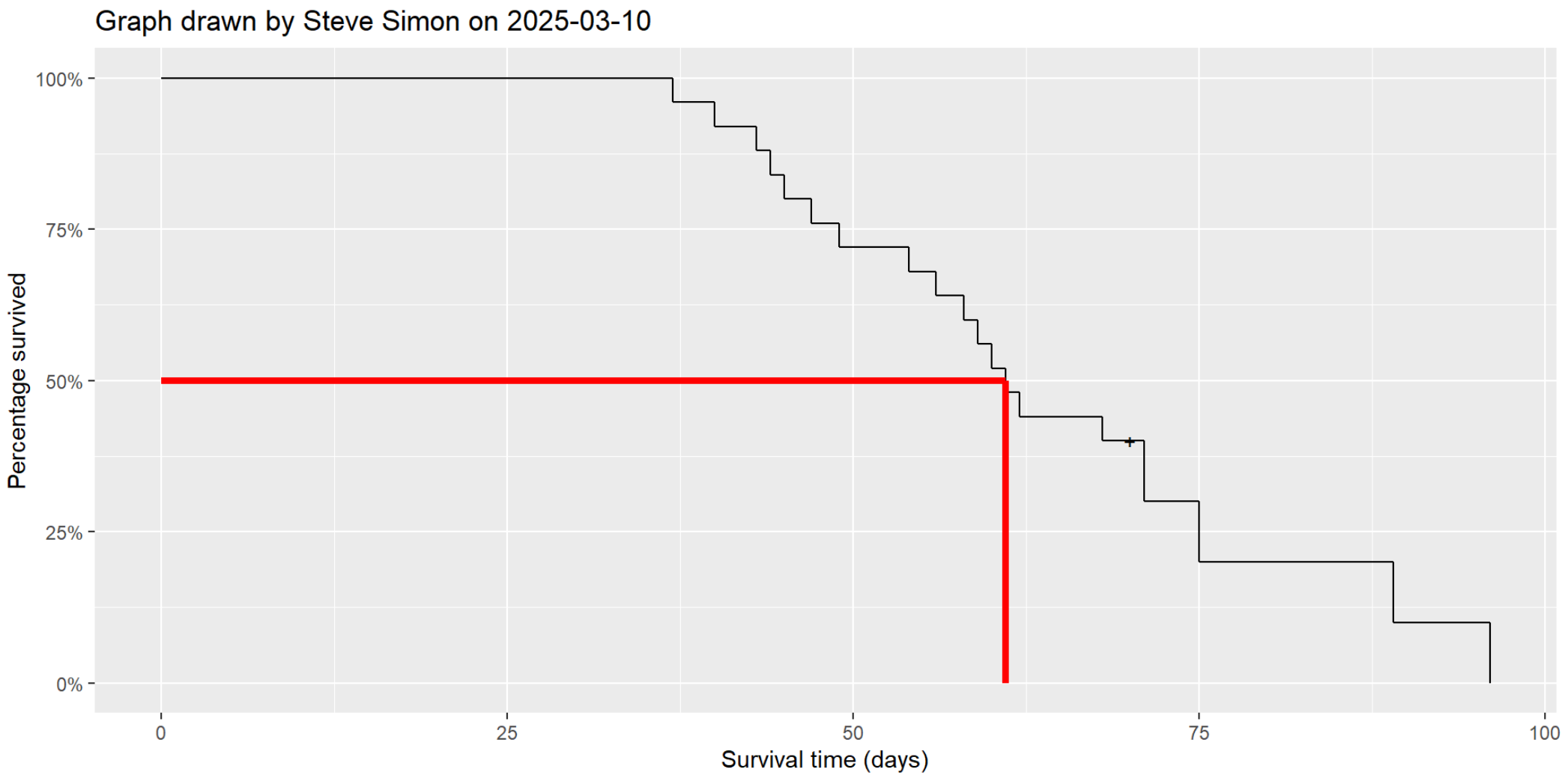
Speaker notes

When you see a survival curve in a research paper, there are three ways to interpret it.

First, presuming that the event in question is a sad event (such as death, relapse), then the upper right hand corner is the happy corner. Most of you patients go for a very long time with only a small proportion suffering the negative event.

In contrast, the lower left corner is the sad corner. Most of your patients experience the bad event, and they experience it very quickly.

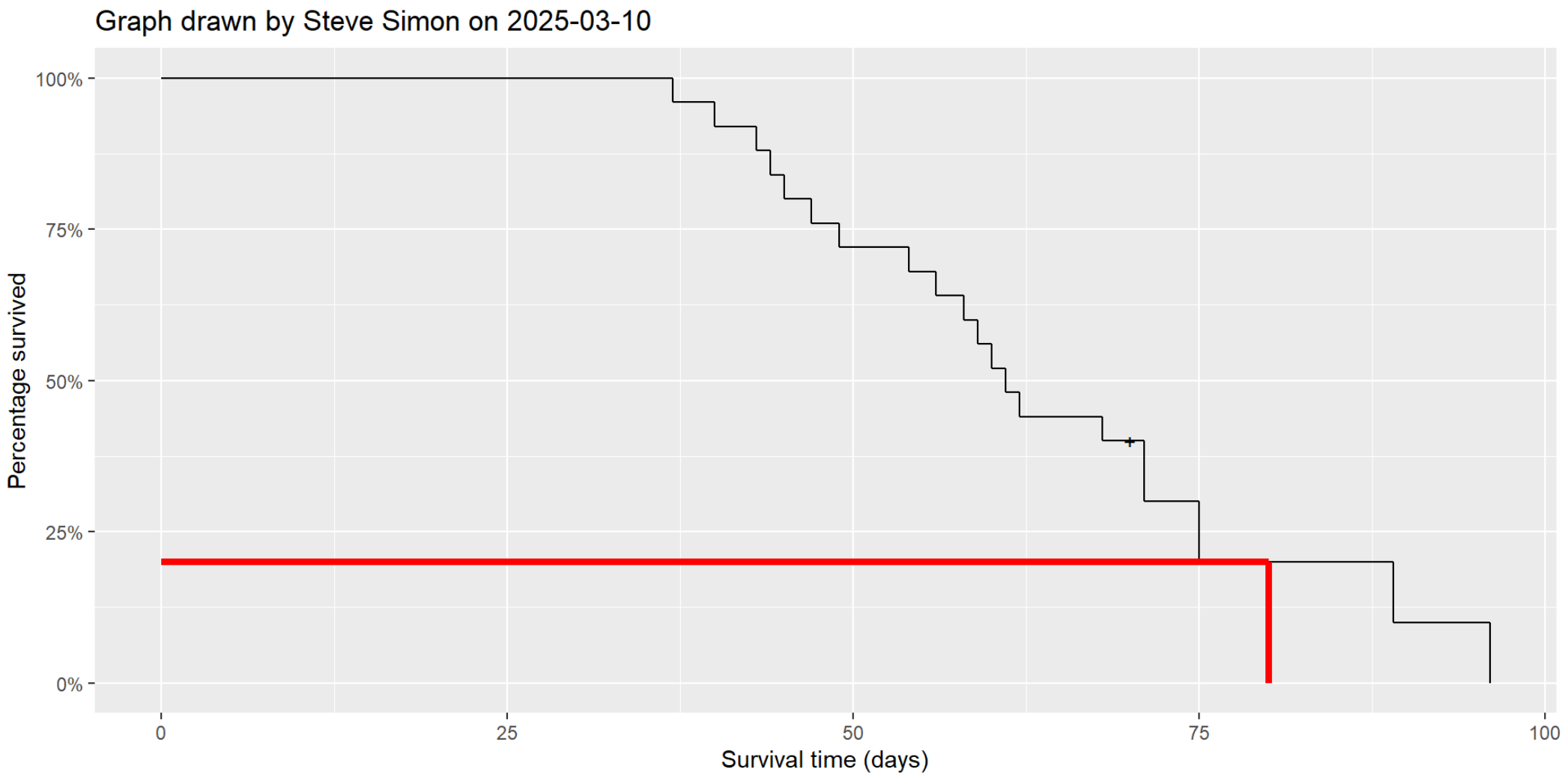
Interpreting Kaplan-Meier plots, 3



Speaker notes

Next, you can get an estimate of the median (or other percentiles) by projecting horizontally until you intersect with the survival curve and then head down to get your estimate. In the survival curve we have just looked at, you would estimate the median survival as slightly more than 60 days.

Interpreting Kaplan-Meier plots, 4



Speaker notes

You can also estimate probabilities for survival at any given time by projecting up from the time and then moving to the left to estimate the probability. In the example below, you can see that the 80 day survival probability is a little bit less than 25%.

What did we learn in module 09?

Meta-analysis

- Quantitative pooling of results from multiple studies
  - Multi-center study
    - Each center has a different protocol
    - Some centers do not share results
- Contrast to systematic overview
  - Careful review of multiple studies
  - May or may not include quantitative pooling
- Contrast to scoping review
  - “Researchers may conduct scoping reviews instead of systematic reviews where the purpose of the review is to identify knowledge gaps, scope a body of literature, clarify concepts or to investigate research conduct.” Munn 2018

Speaker notes

Meta-analysis is the quantitative pooling of results from multiple independently published research studies. I joke about how meta-analysis is a multi-center research study but with a couple of qualifications. First, each center gets to use a different protocol. Second, some centers do not share their results with you. This hints at a couple of important issues I will talk about in detail: heterogeneity and publication bias. When I describe it as a chaotic multi-center trial, it sounds terrible. Well, maybe, but we have still learned a lot from meta-analytic studies in a broad range of scientific and medical areas.

A more commonly used term is “systematic overview” which is a superset of meta-analysis. A systematic overview is systematic meaning that it uses a careful and transparently documented approach to identify all research studies associated with a particular issue. It may or may not include a quantitative analysis. Thus all meta-analyses are systematic reviews but not all systematic reviews are meta-analyses.

There’s a new term, scoping review which I am less familiar with. Here’s a nice quote from a paper by Zachary Munn et al in BMC Medical Research Methodology. It sounds almost like a scoping review is a systematic overview with the pre-specified intent to stop before any serious meta-analytic intents.

I am going to focus only on meta-analysis because I am a statistician and this is a statistics course, so we all love anything quantitative. But never forget that there is more to research than just its quantitative aspects.

Case study: Declining sperm counts

Speaker notes

In 1992, the British Medical Journal published a controversial meta-analysis. This study (BMJ 1992: 305(6854); 609-13) reviewed 61 papers published from 1938 and 1991 and showed that there was a significant decrease in sperm count and in seminal volume over this period of time. For example, a linear regression model on the pooled data provided an estimated average count of 113 million per ml in 1940 and 66 million per ml in 1990.

Several researchers (Fertil Steril 1996: 65(5); 1044-6 and Fertil Steril 1995: 63(4); 887-93) noted heterogeneity in this meta-analysis, a mixing of apples and oranges. Studies before 1970 were dominated by studies in the United States and particularly studies in New York. Studies after 1970 included many other locations including third world countries. Thus the early studies were United States apples. The later studies were international oranges. There was also substantial variation in collection methods, especially in the extent to which the subjects adhered to a minimum abstinence period.

The original meta-analysis and the criticisms of it highlight both the greatest weakness and the greatest strength of meta-analysis.

Meta-analysis is the quantitative pooling of data from studies with sometimes small and sometimes large disparities. Think of it as a multi-center trial where each center gets to use its own protocol and where some of the centers are left out.

On the other hand, a meta-analysis lays all the cards on the table. Sitting out in the open are all the methods for selecting studies, abstracting information, and combining the findings. Meta-analysis allows objective criticism of these overt methods and even allows replication of the research.

Contrast this to an invited editorial or commentary that provides a subjective summary of a research area. Even when the subjective summary is done well, you cannot effectively replicate the findings. Since a subjective review is a black box, the only way, it seems, to repudiate a subjective summary is to attack the messenger.

Major issues in meta-analysis

- Heterogeneity
  - Were apples combined with oranges?
- Publication bias
  - Were some apples left on the tree?
- Study quality
  - Were all the apples rotten?
- Interpretability
  - Did the pile of apples amount to more than just a hill of beans?

Speaker notes

There are four major issues that you should be aware of: heterogeneity, publication bias, study quality, and interpretability. We will tackle each of these in some detail.

What did we learn in module 10?

Talk given to first year medical students

- Topic also relevant to this class.
- Only a few minor changes
  - Different format for the “programming” assignment

Speaker notes

I like to re-use material, so this week's lecture is an adaptation of a talk I have given as a guest lecture for a special class of first year medical students.

There is no programming in this module, so I will change the homework assignment a bit.

Who am I?

Steve Simon

- PhD Statistics, 1982, U Iowa
- Teach in Biomedical and Health Informatics
  - Previous jobs at CMH, CDC
- Part-time independent statistical consultant (P.Mean Consulting)
- Married to a Pediatric Cardiologist (retired)
- Run 5K and 4 mile races

Speaker notes

Let me introduce myself. I am Steve Simon. I got a PhD in Statistics almost 40 years ago from the University of Iowa. The world has changed a lot since then, but I have tried to keep up. Today, if you want to sound trendy, you are a “data scientist”. I teach in the Department of Biomedical and Health Informatics at UMKC. I have had previous jobs at Children’s mercy Hospital, and the Centers for Disease Control and Prevention. I’m also a part-time statistical consultant. I have a sole proprietorship, P.Mean Consulting. That’s short for Professor Mean. For people who don’t get the joke, I point out that Professor Mean is not just your average Professor. Related to this talk, I should point out that I am obsessed with computers.

Obsessed with computers since 1972

List of computer skills

- Bibliographic software: EndNotes, Knowledge Finder, Mendeley, Zotero.
- Cloud storage: Box, Dropbox, iCloud, OneDrive.
- Database software: Microsoft Access, Microsoft SQL Server, Oracle, PC-File, SQLite.
- Electronic Health Records software: i2b2.
- Graphics software: ACDSee, Metafile Companion, Photoshop Elements, SigmaPlot.
- Internet systems: File Transfer Protocol, Gopher, Telnet, USENET, WordPress, World Wide Web.
- Mathematical software: MathCAD, Mathematica, MathType.
- Operating systems: Linux (Raspbian), MS-DOS, OS/2, Windows
- Presentation software: Powerpoint.
- Programming languages: BASIC, C, C++, FORTRAN, Pascal, Perl, PL/1, Python, Visual BASIC.
- Spreadsheets: Excel, Lotus 1-2-3, SuperCalc.
- Statistical software: AMOS, BMDP, IMSL, JMP, LogXact, MINITAB, nQuery Advisor, OpenBUGS, R (including RStudio and tidyverse), RATS, S-Plus, S-Plus/Wavelets, SAS (including SAS University), Stan, SPIDA, SPSS, STATA, Statgraphics, StatXact, Systat, WinBUGS.
- Utility software: DBMS/COPY, Norton Anti-virus, Notepad++, TextPad, WinZip.
- Word processing: LaTeX (MIKTeX), RMarkdown (including blogdown, bookdown, and pagedown), Word, Word Perfect.

Figure 1. Section on computer skills from my resume

Speaker notes

I should add that I am a bit of a computer geek. Here’s a list of computer skills that I put on my resume. I deliberately made this too small to read so that you wouldn’t recognize that half of the computer skills I mention have been obsolete for several deades. The computers I used in the 1970s were nothing like today’s computers.

Worked with health care applications since 1987

Speaker notes

I’ve also been working with a variety of health care professionals for many decades. I have learned a lot along the way, but I am not a doctor. Not an MD doctor anyway. When I talk about medical examples, keep that in mind. I don’t always describe things accurately from a medical perspective, and I’m always forgetting which one is the bad cholesterol.

Is it ldl or hdl? Does anyone know?

Quiz questions (1/3)

Why does Joel Best call statistics a social construct?

- Statistics are misquoted often on social media.
- Statistics are selected, shaped, and presented by human beings.
- Statistics are used to promote socialism.
- Statistics are dehumanizing.

Speaker notes

I want to list a few quiz questions relating to this lecture. Remember these questions when I get to the slide that discusses them.

Quiz questions (2/3)

What is the main philosophical foundation of empiricism?

- Everything can be reduced to a mathematical equation.
- Experiments can reveal the realities of the world.
- Some questions are impossible to answer.
- We construct our own reality based on our own lived experiences

Speaker notes

Here’s the second question.

Quiz questions (3/3)

What is a major problem with data science?

- Data scientists rely on large amounts of data with uneven quality.
- Models developed by data scientists can lead to loss of privacy.
- Prediction models are a black box that can hide discriminatory intent.
- All of the above.

Speaker notes

Here’s the third question. You don’t need to answer these questions now. Just be ready to answer them after the lecture.

First poll question

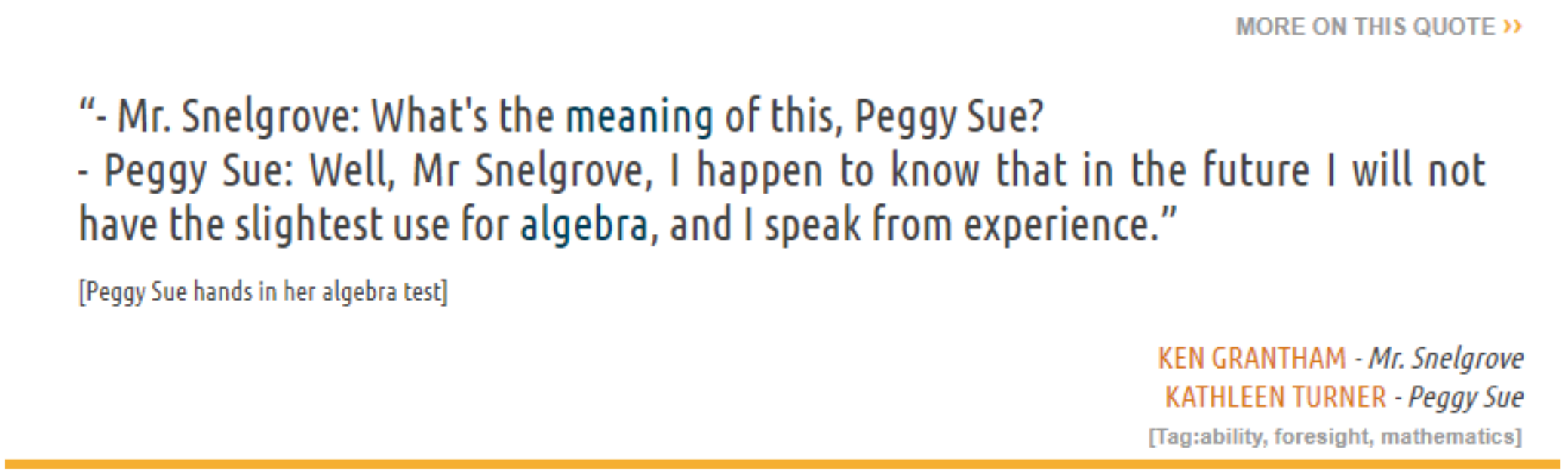


Figure 2. Quote from "Peggy Sue Got Married"

Speaker notes

I want to get a quick feel for your background and interests. Here's a quote from a romantic comedy starring Kathleen Turner from 1986. A forty year old woman, played by Kathleen Turner, travels back in time to her high school senior year, 1960. She has an amusing interchange with her high school math teacher.

"I happend to know that in the future, I will not have the slightest use for algebra, and I speak from experience."

Think back to your high school algebra class.

1. Do you remember any important formulas from that class?
2. Did you hate, hate, hate high school algebra?
3. Did you love high school algebra?

Big question: Will you use high school algebra in your future?

Source: https://www.moviequotes.com/s-movie/peggy-sue-got-married/

Second poll question

Figure 3. Images of various computers

Speaker notes

How many of these computational devices have you used?

1. Laptop computer
2. Desktop computer
3. Tablet computer
4. Smart phone
5. Gaming console
6. Smart watch

Big question: What impact does the current generation’s immersion of computing have on society?

Images found at

https://commons.wikimedia.org/wiki/File:Black_laptop_computer_open_frontal.svg https://www.pcmag.com/picks/the-best-desktop-computers https://www.cleverfiles.com/howto/what-is-tablet-computer.html https://www.theverge.com/21420196/best-budget-smartphone-cheap https://www.vulture.com/article/best-video-game-console-2020-ps5-xbox-series-x-nintendo-switch.html https://www.homernews.com/marketplace/fitnus-smartwatch-review-legit-fitness-tracker-smart-watch/

Are Statisticians Gods?

I’m helping someone who wants an alternative statistical analysis to the one used by the principal investigator. I’m happy to help and will offer advice about why my approach may be better, but I was warned that the PI considers the analysis chosen to be ordained by the “**Statistical Gods**” at her place of work.

Speaker notes

True story. I was asked to review a report from a federal agency, but was warned that negative comments, even if accurate, might not be well received because the agency's work was ordained by their Statistic Gods. At first, I thought this was amusing. If I could get the title of "Statistical God", I could double my hourly consulting rate. But deep down this story really bothered me. It implies that Statistical skills are supernatural, and only available to a select few special people. I learned Statistics through hard work, and you can learn Statistics through hard work also. Some people will learn it faster than others because they have good aptitudes in mathematics and programming. But there is nothing other than time to stop you.

What did we learn in module 11?

Hierarchical data

- Moving beyond the independence assumption
- Correlation within clusters

Speaker notes

Throughout this class, I have discussed the assumptions that you need for the t-test, the chi-square test, the ANOVA test, and so forth. Every single time, I mention the assumption of independence. It’s often one that you can only check qualitatively. I mention special cases where you can’t assume independence. In this lecture, I want to talk about one of those special cases: hierarchical data.

Hierarchical data has some additional grouping factor, often called a cluster. Measurements made within a cluster are correlated with one another, violating the assumption of independence.

Examples of hierarchical data, 1 of 2

- Body parts
  - Left eye/right eye
  - Teeth
  - Skin patches
- Human families
- Animal litters

Speaker notes

A simple example of hierarchical data is when you select a group of patients and then make measurements on two or more parts of their body. You might, for example, put an eye drop medication in the left eye and a placebo drop in the right eye. You might apply different types of sealents on different teeth in a mouth. You might put different food allergens on different parts of a patient’s back.

You might select families from a population and make measurements on two or more members of the same family. Since family members share the same environment and have very similar genetics, any comparison made within a family is likely to be more precise.

Likewise, measurements on the animals from the same litter will be precise because of a shared inter-uterine environment prior to birth and shared feeding from the same mother before weaning.

Examples of hierarchical data, 2 of 2

- Clinics/hospitals
- Communities
- Repeated measurements

Speaker notes

Patients treated at the same clinic or the same hospital will often have similar outcomes. This might be caused by the location of the clinic, which determines the types of patients that come in. it might also be caused by subtle treatment practices that are agreed upon within a clinic but which might vary from one clinic to another.

You might select entire communities and then sample people within each community. You will see some level of similarity within each community because of demographic similarities or because of common dietary or cultural practices.

Often, you take measurements repeatedly on an individual under different experimental conditions.

Longitudinal data (topic for next module)

- Measurements taken at different times
  - Emphasis in changes over time

Speaker notes

A special case that I want to handle separately is longitudinal data. This is similar to repeated measures data. With longitudinal data, often the emphasis is in changes that occur over time. Repeated measurements, in contrast, emphasize different treatments with the hope that the time gaps between the measurements are small enough that you don’t see changes over time other than the changes caused by differences in what you measure and how you measure it.

Between and within cluster comparisons

- Positive correlation
  - Improves precision of within cluster comparisons
  - Hurts precision of between cluster comparisons
- Example with litters
  - Medication administered during pregnancy
  - Medication administered after birth

Basic notation, 1 of 2

- Y_{ij}
  - i defines cluster
    - i=1,…,a
  - j defines individual within cluster
    - j=1,…,n

Basic notation, 2 of 2

- $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$
  - μ unknown constant
  - α_i is normally distributed
    - $SD(\alpha_i) = \sigma_{between}$
  - ϵ_{ij} is normally distributed
    - $SD(\epsilon_i) = \sigma_{within}$

Some basic results

- $SD(Y_{ij}) = \sigma_{total} = \sqrt{\sigma^2_{between} + \sigma^2_{within}}$
- $SD(\bar{Y}_{..}) = \sqrt{\frac{\sigma^2_{between}}{a} + \frac{\sigma^2_{within}}{an}}$
- $Corr(Y_{ij}, Y_{ik}) = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}}$
  - Intraclass correlation (ICC)

Expected mean squares, 1 of 2

- $MS(between) = \frac{1}{a-1}\Sigma n(\bar{Y}_{i.} - \bar{Y}_{..})^2$
  - $E[MS(between)] = n\sigma^2_{between} + \sigma^2_{within}$

Expected mean squares, 2 of 2

- $MS(within) = \frac{1}{a(n-1)}\Sigma\Sigma(\bar{Y}_{ij} - \bar{Y}_{i.})^2$
  - $E[MS(within)] = \sigma^2_{within}$

Variance components estimates

- $\hat{\sigma}^2_{between} = \frac{MS(between) - MS(within)}{n}$
- $\hat{\sigma}^2_{within} = MS(within)$

What did we learn in module 12?

Longitudinal data

- Measurements taken at different times
  - Emphasis in changes over time

Speaker notes

In the previous module, I talked about hierarchical models and mentioned a particular case, longitudinal data, that I want to talk more about in this presentation.

Longitudinal data is similar to repeated measures data. With both, you measure the same subject repeatedly. With longitudinal data, often the emphasis is in changes that occur over time. Repeated measurements, in contrast, emphasize different treatments with the hope that the time gaps between the measurements are small enough that you don’t see changes over time.

The differences between longitudinal data, repeated measures data, or hierarchical data are subtle. Perhaps these are distinctions without a difference. I decided to separate out longitudinal data for a different module perhaps more out of the desire to split a complex topic into smaller bite-sized pieces.

Random intercepts model, 1

- Simplest pattern for longitudinal data
- $Y_{ij},\ i = 1, \ldots, n;\ j = 1, \ldots, k$
  - n subjects, k time points
- t_j, time of jth measurement
  - First time is often zero

Speaker notes

The simplest longitudinal model has n subjects and k time points. The first time point is often set to zero. The times are often evenly spaced, but they don’t have to be.

Random intercepts model, 2

- $Y_{ij} = \beta_0 + u_{0i} + \beta_1 t_j + \epsilon_{ij}$
  - β_0 and β_1 are unknown constants
  - u_{0i} and ϵ_{ij} are normally distributed
    - $SD(u_{0i}) = \sigma_{intercept}$
    - $SD(\epsilon_{ij}) = \sigma_{error}$

Speaker notes

There are two sources of random variation in the random intercepts model, u_{0i} and ϵ_{ij}.

Random intercepts model, 3

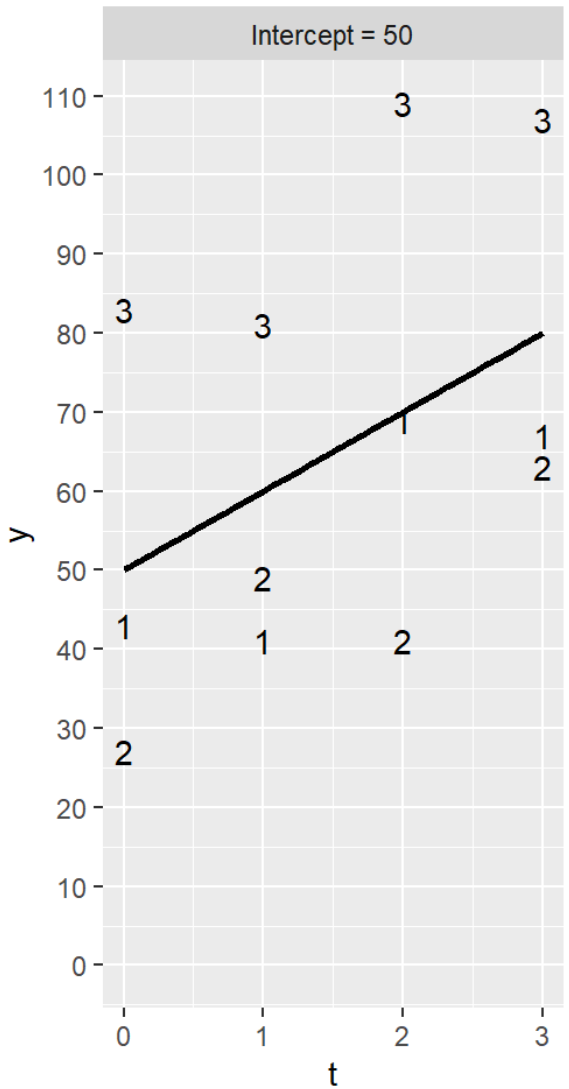
- $SD(Y_{ij}) = \sqrt{\sigma^2_{intercept} + \sigma^2_{error}}$
- $Corr(Y_{ij}, Y_{im}) = \frac{\sigma^2_{intercept}}{\sigma^2_{intercept} + \sigma^2_{error}}$

Speaker notes

The standard deviation for any individual observation combines the standard deviation for the random intercepts and the standard deviation for the error terms. They combine in a Pythagorean way.

The correlation of two measurements on the same patient is comparable to a measure we defined in the last module, the intraclass correlation.

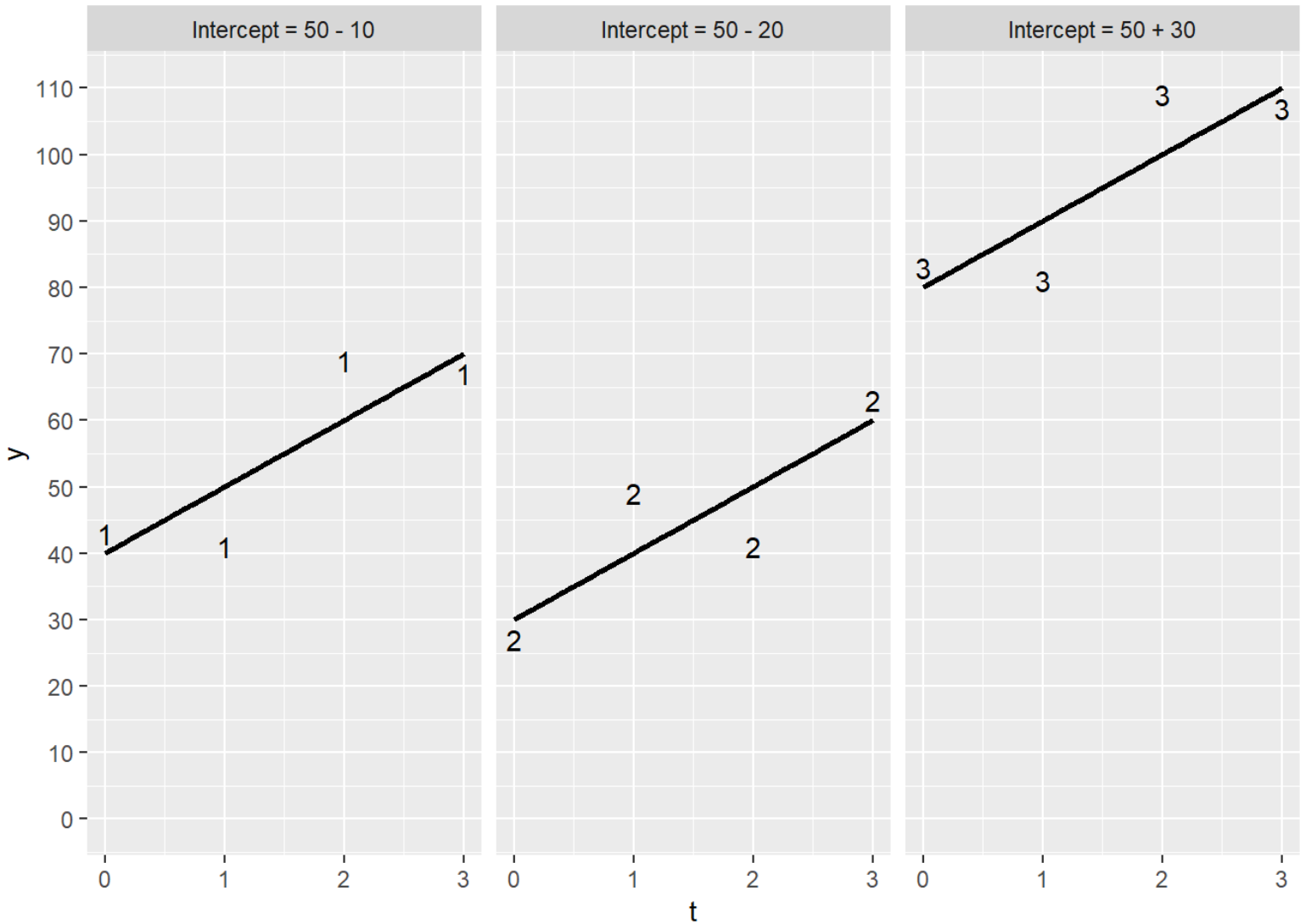
Random intercepts illustrated, 1



Speaker notes

This graph shows a single line. It does reasonably well, but there is a fair amount of variation. There is something worth examining more closely. The third group has values well above the regression line. The first two groups have values slightly below the regression line.

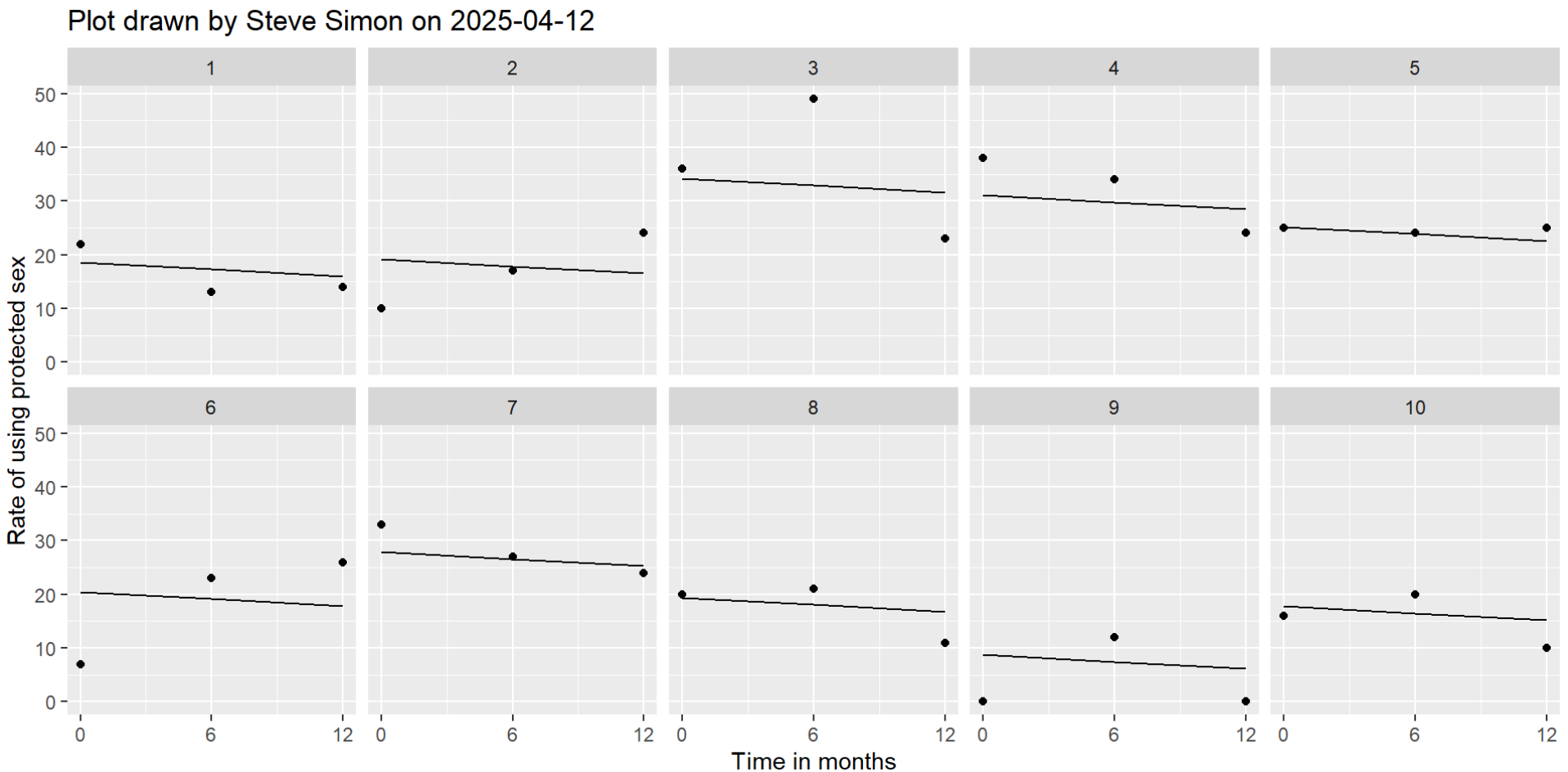
Random intercepts illustrated, 2



Speaker notes

If you fit a separate intercept for each line, you get a lot closer to the data.

Illustration of random intercepts with real data



What did we learn in module 13?

A simple example of Bayesian data analysis.

- ECMO study
- Treatment versus control, mortality endpoint
  - Treatment: 28 of 29 babies survived
  - Control: 6 of 10 babies survived
  - Source: Jim Albert in the Journal of Statistics Education (1995, vol. 3 no. 3).

Speaker notes

Bayesian data analysis seems hard, and it is. Even for me, I struggle with understanding Bayesian data analysis. In fairness, I must admit that much of my discomfort is just lack of experience with Bayesian methods. In fact, I've found that in some ways, Bayesian data analysis is simpler than classical data analysis. You, too, can understand Bayesian data analysis, even if you'll never be an expert at it. There's a wonderful example of Bayesian data analysis at work that is simple and fun. It's taken directly from an article by Jim Albert in the Journal of Statistics Education (1995, vol. 3 no. 3) which is available on the web at www.amstat.org/publications/jse/v3n3/albert.html.

I want to use his second example, involving a comparison of ECMO to conventional therapy in the treatment of babies with severe respiratory failure. In this study, 28 of 29 babies assigned to ECMO survived and 6 of 10 babies assigned to conventional therapy survived. Refer to the Albert article for the source of the original data. Before I show how Jim Albert tackled a Bayesian analysis of this data, let me review the general paradigm of Bayesian data analysis.

Wikipedia introduction

- $P(H|E) = P(E|H)\ P(H) / P(E)$
  - H = hypothesis
  - E = evidence
  - P(H) = prior
  - $P(E|H)$ = likelihood
  - $P(H|E)$ = posterior

Speaker notes

Wikipedia gives a nice general introduction to the concept of Bayesian data analysis with the following formula:

$P(H|E) = P(E|H)\ P(H) / P(E)$

where H represents a particular hypothesis, and E represents evidence (data). P, of course, stands for probability.

Prior distribution

- Degree of belief
  - Based on previous studies
  - Subjective opinion (!?!)
- Examples of subjective opinions
  - Simpler is better
  - Be cautious about subgroup analysis
  - Biological mechanism adds evidence
- Flat or non-informative prior

Speaker notes

The first step is to specify P(H), which is called the prior probability. Specifying the prior probability is probably the one aspect of Bayesian data analysis that causes the most controversy. The prior probability represents the degree of belief that you have in a particular hypothesis prior to collection of your data. The prior distribution can incorporate data from previous related studies or it can incorporate subjective impressions of the researcher. What!?! you're saying right now. Aren't statistics supposed to remove the need for subjective opinions? There is a lot that can be written about this, but I would just like to note a few things.

First, it is impossible to totally remove subjective opinion from a data analysis. You can't do research without adopting some informal rules. These rules may be reasonable, they may be supported to some extent by empirical data, but they are still applied in a largely subjective fashion. Here are some of the subjective beliefs that I use in my work:

You should always prefer a simple model to a complex model if both predict the data with the same level of precision.

You should be cautious about any subgroup finding that was not pre-specified in the research protocol.

if you can find a plausible biological mechanism, that adds credibility to your results.

Advocates of Bayesian data analysis will point out that use of prior distributions will force you to explicitly state some of the subjective opinions that you bring with you to the data analysis.

Second, the use of a range of prior distributions can help resolve controversies involving conflicting beliefs. For example, an important research question is whether a research finding should "close the book" to further research. If data indicates a negative result, and this result is negative even using an optimistic prior probability, then all researchers, even those with the most optimistic hopes for the therapy, should move on. Similarly, if the data indicates a positive result, and this result is positive even using a pessimistic prior probability, then it's time for everyone to adopt the new therapy. Now, you shouldn't let the research agenda be held hostage by extremely optimistic or pessimistic priors, but if any reasonable prior indicates the same final result, then any reasonable person should close the book on this research area.

Third, while Bayesian data analysis allows you to incorporate subjective opinions into your prior probability, it does not require you to incorporate subjectivity. Many Bayesian data analyses use what it called a diffuse or non-informative prior distribution. This is a prior distribution that is neither optimistic nor pessimistic, but spreads the probability more or less evenly across all hypotheses.