1. Road signs with gunshot holes

   (a) Poisson sampling
   because the total number of signs per state is random

   (b) Either $\chi^2 = 25.46$, χ^2_1 distribution, or $Z = \pm 5.05$, N(0,1) distribution
   Available in both SAS and R output

   (c) $\chi^2 = 4.71$, χ^2_1 distribution.
   Note; This is based on only the data on shot-at signs, with log-time offset
   Available in the SAS output, but not the R output

   (d) $Y_{ij} \sim Pois(\exp(o_{ij} * (\mu + \alpha_i)))$. i indexes states, j indexes road segment (mile)
   Notes: 1) Y_{ij}: count of shot-at-signs in a road segment
   o_{ij}: length of that road segment (1 mile by definition)
   2) If you answered using $\boldsymbol{X\beta}$, I deducted 2 points because $\boldsymbol{X\beta}$ describes any model!
   3) If you added an additional $+\epsilon$, I deducted 2 points, because that is a hangover from normal models

   (e) estimate log-ratio: $\log \frac{\mu_{UT}}{\mu_{NV}} = 0.0198$
   Note: available in SAS output, not in R output

   (f) se $= 0.33 = 0.244 * \sqrt{OD}$, where the overdispersion factor $= 1.86$
   Note: available in SAS output, not in R output

2. Combines

   (a) 3 sizes of eu: fields, field-parts, and field-bits

   (b) slope $\rightarrow$ fields, design $\rightarrow$ field-parts, and speed $\rightarrow$ field-bits

   (c) The non-zero columns of the $\boldsymbol{Z}$ matrix are:

fields		parts		
1	0	1	0	0
1	0	1	0	0
0	1	0	1	0
0	1	0	0	1

   Notes: The first two observations are from the same field and field-part (because same design). The third observation is a different field and field-part. The fourth is the same field as the third but a different field-part (because different design). Some common issues were including columns for the errors (not part of $\boldsymbol{Z}$) and including parts of the $\boldsymbol{X}$ matrix.

(d)

Source	df
slope	2
field(slope)	9
design	2
slope*design	4
part(field, slope)	18
speed	3
speed*slope	6
speed*design	6
speed*slope*design	12
error	81
c. total	143

Note: This stumped everyone and really stumped a few.

3. Vitamin A - study 1

(a)

Source	df
Age group	2
Subj(age)	27
Type	1
Age*type	2
Error	267
c. total	299

(b) Fixed: age group, type, and age*type interaction
Random: subject(age)
Note: subject(age) is random because it is an error term

(c)

$$\begin{aligned} E\,MS &= E\frac{nm}{t-1}\sum(\overline{y}_{i\ldots}-\overline{y}_{\ldots.})^2 \\ &= \frac{nm}{t-1}E\sum\left[(\mu+\alpha_i+\overline{\beta}_{.}+\overline{\alpha\beta}_{i.}+\overline{\gamma}_{i.}+\overline{\varepsilon}_{i\ldots})-(\mu+\overline{\alpha}_{.}+\overline{\beta}_{.}+\overline{\alpha\beta}_{..}+\overline{\gamma}_{..}+\overline{\varepsilon}_{\ldots.})\right]^2 \\ &= \frac{nm}{t-1}\left[\sum(\alpha_i-\overline{\alpha}_{.}+\overline{\alpha\beta}_{i.}-\overline{\alpha\beta}_{..})^2+E\sum(\overline{\gamma}_{i.}-\overline{\gamma}_{..})^2+E\sum(\overline{\varepsilon}_{i\ldots}-\overline{\varepsilon}_{\ldots.})^2\right] \\ &= \frac{nm}{t-1}\left[Q(t)+\frac{t-1}{n}\sigma_u^2+\frac{t-1}{nm}\sigma_e^2\right] \\ &= \frac{nm}{t-1}Q(t)+10\sigma_u^2+\sigma_e^2 \end{aligned}$$

(d)

$$\begin{aligned} Var C\,\overline{Y} &= \sum C_i^2 \text{Var}\,\overline{Y} \\ &= (1+1.09+1.69)\left(\frac{\sigma_u^2}{10}+\frac{\sigma_e^2}{100}\right) \\ &= 2.78\left(\frac{\sigma_u^2}{10}+\frac{\sigma_e^2}{100}\right) \end{aligned}$$

(e) E MS for subj(trt) = $10\sigma_u^2 + \sigma_e^2$, so the estimate of the desired quantity is $\frac{2.78}{100} MS_{subj(trt)}$.

4. Vitamin A - study 2

   (a) σ_e^2.
   Note: A few people calculated the MS. That's not the expected value.

   (b) No. Those observations provide information about the variability between subjects.
   Notes: Those observations provide no information about the error variance (because there is only one observation per subject). They provide no information about the age group mean (because the 1 df is "used" to estimate the subject effect).

   (c) $\hat{\sigma}_u^2 = 249.3$
   The calculations: $MS_{subj} = 57298/42 = 1,364.2$, $MS_{error} = 4650.6/198 = 23.5$, $1364.2 = 23.5 + 5.3776\hat{\sigma}_u^2$, and solve for $\hat{\sigma}_u^2$.

   (d) This is a test of $\sigma_u^2 = 0$. F = 1364.7/23.5 = 50.8. Central F distribution with 42,198 df.

   (e) The appropriate denominator is $\sigma_e^2 + 5.7089\sigma_u^2$, which is estimated as $23.5 + 5.7089 * 249.3 = 1446.7$.

   (f) To get the correct coefficient for σ_u^2, you need to multiply MS_{subj} by $\frac{5.7089}{5.3776} = 1.0616$. The desired linear combination of Mean Squares is 1.0616 MS_{subj} − 0.0616 MS_{error}. Using the Cochran-Satterthwaite approximation, you get

$$\begin{aligned} \hat{\nu} &= \frac{[1.0616 * 1364.2 - 0.0616 * 23.5]^2}{[1.0616^2 * 1364.2^2/42 + 0.0616^2 * 23.5^2/198]} \\ &= \frac{2,093,222.2}{49,937.1 + 200.1} \\ &= 41.7 \end{aligned}$$

   Note: different amounts of round off will give slightly different answers. If you were close and doing the right thing, you got full credit.