

Why We Started The Initiative

We had a request from our Programmers regarding furthering their understanding of statistical tasks which they were being asked to perform in their day to day work.

Potential current level of applicational knowledge allows Programmers to implement the code correctly, however, our team wanted to have a deeper understanding of the question and interpretation of the results produced, e.g., knowing how to implement a “HR (95% CI)” but not understanding whether the numbers make sense – what HR of 15.9 means (risk of event being 15.9 times higher on experimental treatment when compared to standard of care)? So we asked the Programmers what **THEY** wanted to hear about...

What Programmers said

Intermediate/Advanced - PROC LIFETEST / KM plots

What to Expect - Confidence Intervals / Levels

Intermediate/Advanced - PROC PHREG

What to Expect - Covariates

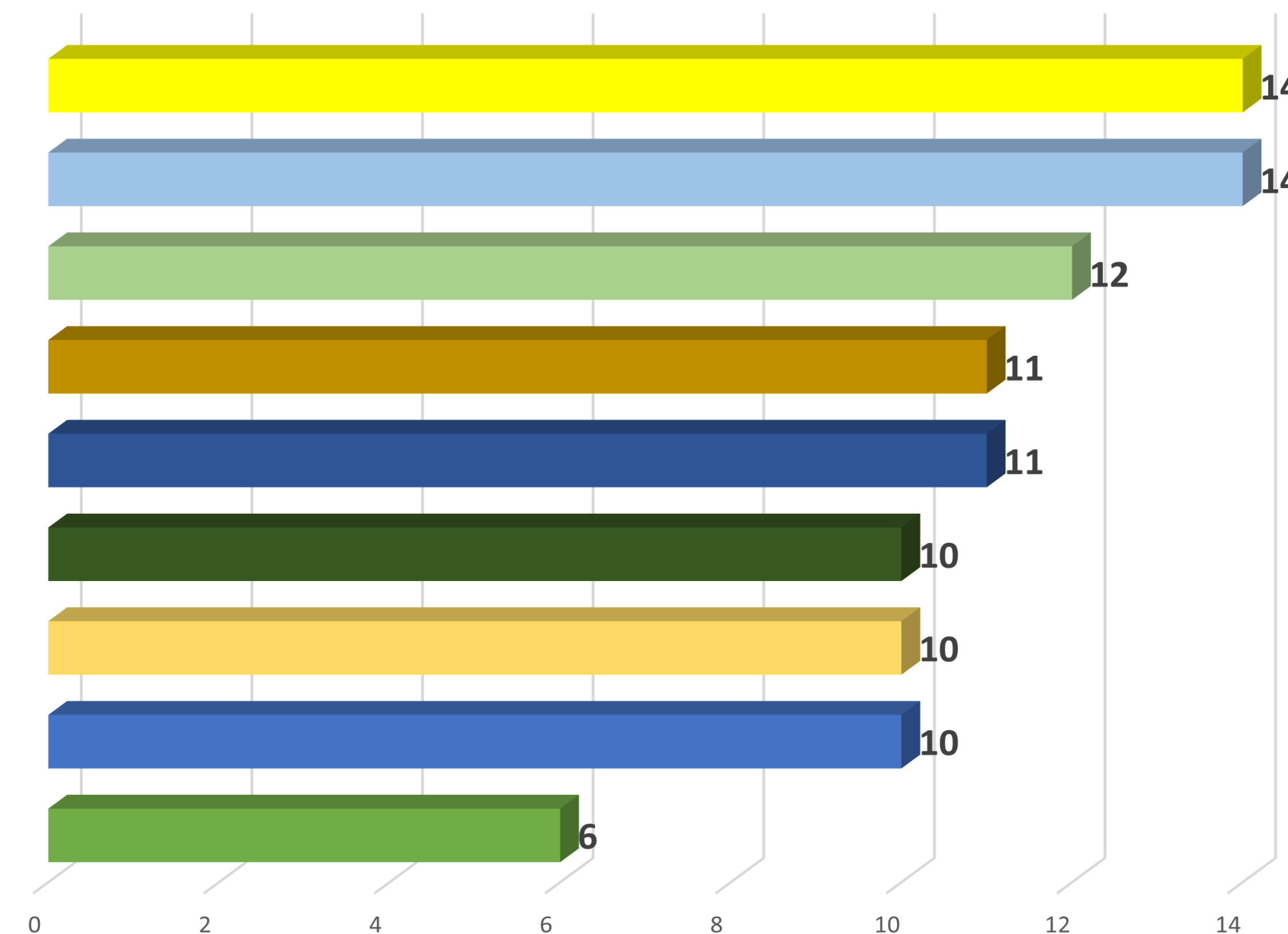
What to Expect - Standard Deviation / Error

Advanced - PROC MIXED

Intermediate/Advanced - Using CMH Method

Intermediate/Advanced - Event Rates

What to Expect - Percentiles / Quantiles



So... once we got the results, experienced Statisticians were identified to prepare draft presentations on topics of interest for our Programmers, varying in duration depending on complexity of the topic:

- ♦ **10-15 min** for ‘What to Expect’ topics
- ♦ **20-30 min** for ‘Intermediate’ topics
- ♦ **<1h/multiple sessions** for ‘Advanced’ topics.

Based on the information we collected, we drafted a pilot training session in ‘What to Expect’ group and delivered training to the programming team:

- ♦ **Topic:** What to Expect - Percentiles and Quantiles
- ♦ **Duration:** 15 minutes + question time
- ♦ **Feedback form:** 5-item questionnaire
 - ♦ 4 Likert scale style questions
 - ♦ 1 free text feedback question

Primary Author: **Presenting Author:**

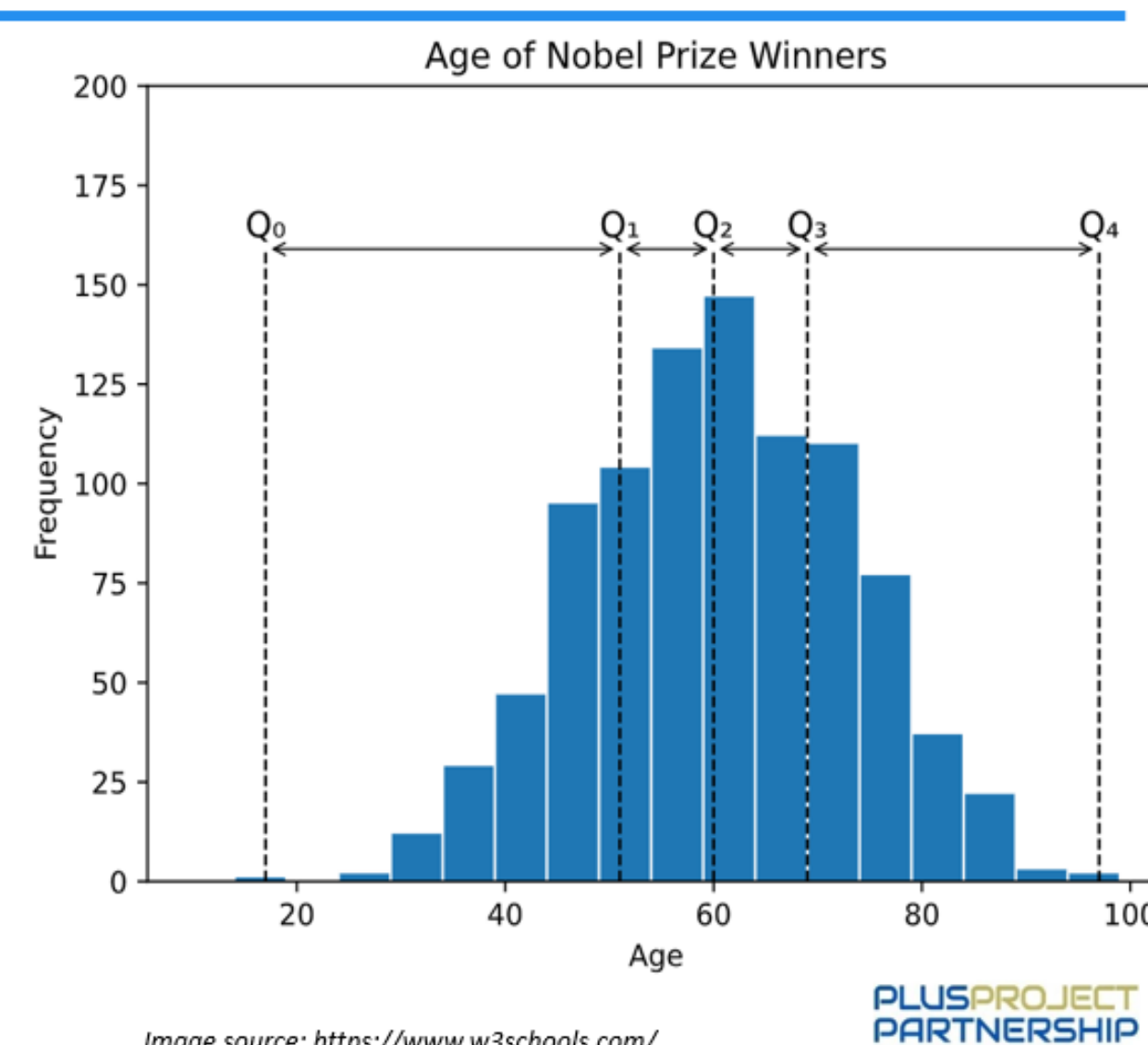
Alina Striha
Principal Statistician, MSc, BSc (hons)
alinastriha
Alina.Striha@Plus-Project.co.uk
www.plus-project.co.uk

Rachel Dunk
Programming Manager
rachel-d-010696107
Rachel.Dunk@Plus-Project.co.uk
www.plus-project.co.uk

What are Quartiles?

Quartiles are values that separate the data into four equal parts.

- The quartiles (Q_0, Q_1, Q_2, Q_3, Q_4) are the values that separate each quarter.
- Between Q_0 and Q_1 are the 25% lowest values in the data. Between Q_1 and Q_2 are the next 25%. And so on.
- Q_0 is the smallest value in the data.
- Q_1 is the value separating the first quarter from the second quarter of the data.
- Q_2 is the middle value (median), separating the bottom from the top half.
- Q_3 is the value separating the third quarter from the fourth quarter.
- Q_4 is the largest value in the data.



How to calculate quantiles/percentiles in SAS?

```
data test;
input assets @@;
datalines;
12 21 21 40 42 48 57 72
;
run;

proc univariate data=test;
ods select quantiles;
run;

proc means data = test n p50 q1 q3 QRANGE;
var value;
run;
```

- Other options for quantile statistics
- MEDIAN/P50 – 50th percentile
 - P1 – 1st percentile
 - P5 – 5th percentile
 - P10 – 10th percentile
 - Q1/P25 – lower quartile/25th percentile
 - Q3/P75 – upper quartile/75th percentile
 - P90 – 90th percentile
 - P95 – 95th percentile
 - P99 – 99th percentile
 - Q RANGE – Interquartile range (IQR)
 - RANGE – difference between min and max values

How to calculate quantiles/percentiles in R?

Use the R quantile() function to find the quantiles of the values:

```
values <- c(12,21,21,40,42,48,57,72)
quantile(values)
```

```
> values <- c(12,21,21,40,42,48,57,72)
> quantile(values)
0% 25% 50% 75% 100%
12.00 21.00 41.00 50.25 72.00
```

Use the R quantile() function to find the 65th percentile (0.65):

```
values <- c(13,21,21,40,42,48,57,72)
quantile(values, 0.65)
```

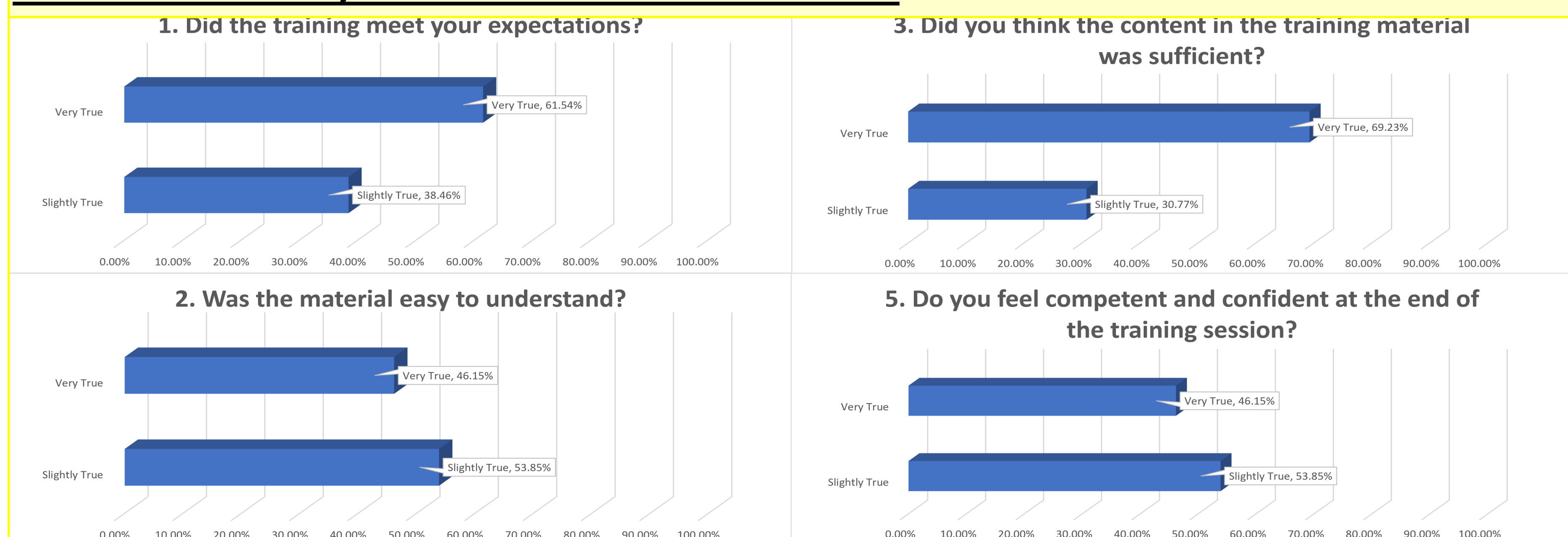
```
> values <- c(13,21,21,40,42,48,57,72)
> quantile(values, 0.65)
65%
45.3
```

quantile(values, probs = seq(0,1,1/3)) ← Tertiles

```
> quantile(values, probs = seq(0,1,1/3))
0% 33.33333% 66.66667% 100%
12.00000 27.33333 46.00000 72.00000
```

Substitute ‘1/3’ for any other breakdown, as required.

Pilot Feedback, Results and Conclusions



- ♦ Training meets expectations and content is easy to understand.
- ♦ Definitions and worked examples are a must-have.
- ♦ Having examples of SAS and R code as reference materials.
- ♦ Out of 17 respondents, 82% of the respondents found it easier to QC their work; 88% found unrealistic results to stand out during review and 76% had a reduction in the number of questions about derivations (based on Yes/Maybe response).

Next Steps

- ♦ Based on the responses provided, we tailored the subsequent sessions to reflect the feedback and have been rolling out the training on roughly two-monthly basis.
- ♦ Trainings are available for reference and new starter training at onboarding.

“I found the training valuable and aided understanding.”