Comments on homework (Week 5)

- No difference in means (average) between the two populations (H_0)
- The direction of difference can be inferred from the means.
- Confidence intervals show the range of the magnitude of difference at a specified level of confidence.
- Confidence intervals allow to test the hypothesis (95% CI for 5% significance level)

Some useful formulas (t-test last week)

- T-test statistic (two-sample): $t = \frac{\hat{\Delta}}{\hat{\sigma}_{\hat{\Delta}}} = \frac{\bar{Y}_1 - \bar{Y}_2}{\hat{\sigma}_{\hat{\Delta}}}$
- Estimated standard error (two-sample): $\hat{\sigma}_{\hat{\Delta}} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- Estimated common population standard deviation (assuming equal variances): $\hat{\sigma} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$
- Confidence intervals: $\hat{\Delta} \pm z_{a/2}\hat{\sigma}_{\hat{\Delta}}$
- Confidence intervals (95%, standard normal/t-distribution with df>120): $\hat{\Delta} \pm 1.96\hat{\sigma}_{\hat{\Delta}}$

Z-test (inference in proportions)

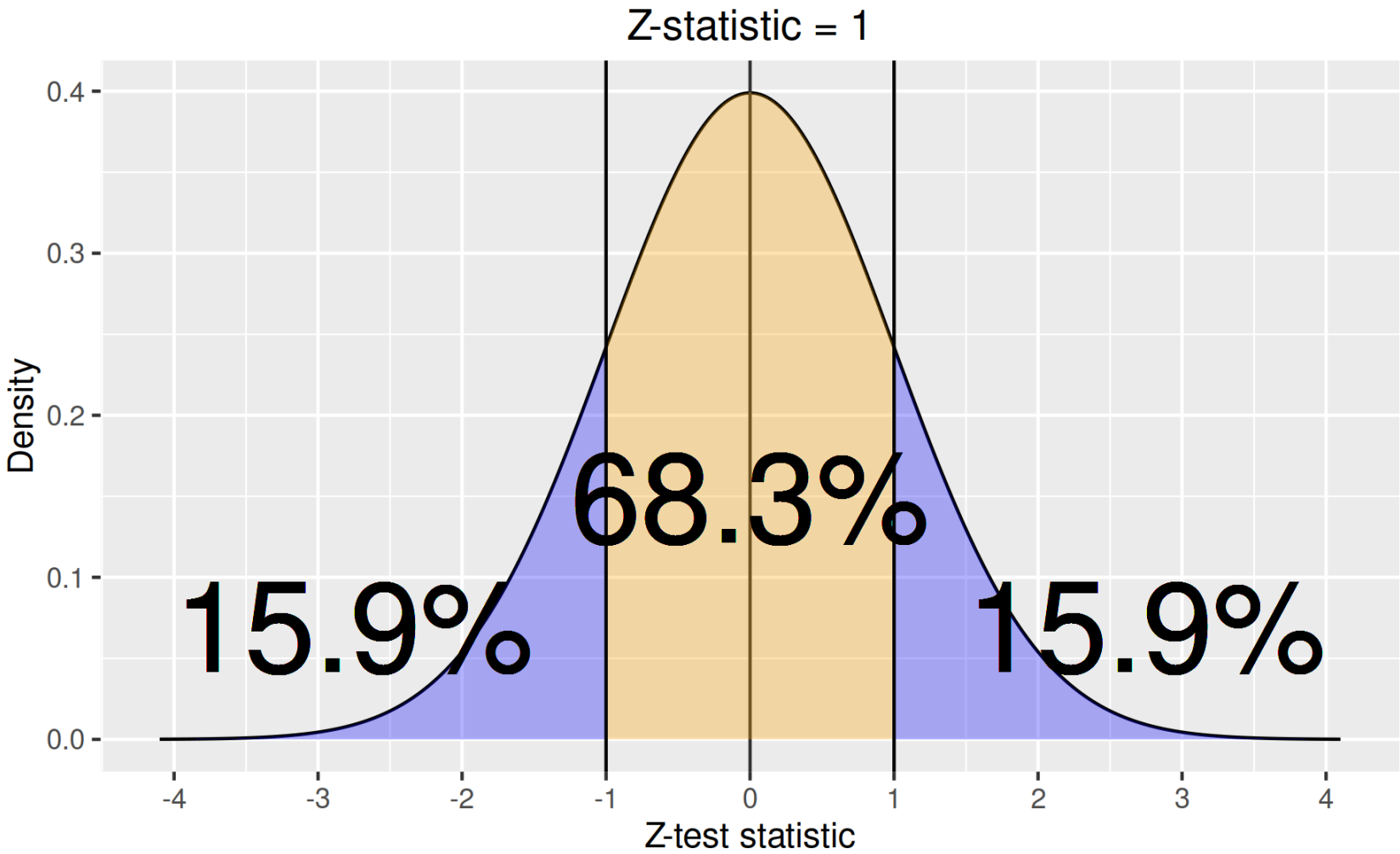
Recall the conceptual framework we use for statistical inference:

1. **Type** of data (**level** of measurement)
2. **Assumptions** about the data (often assumed implicitly)
3. Statistical **hypotheses**:

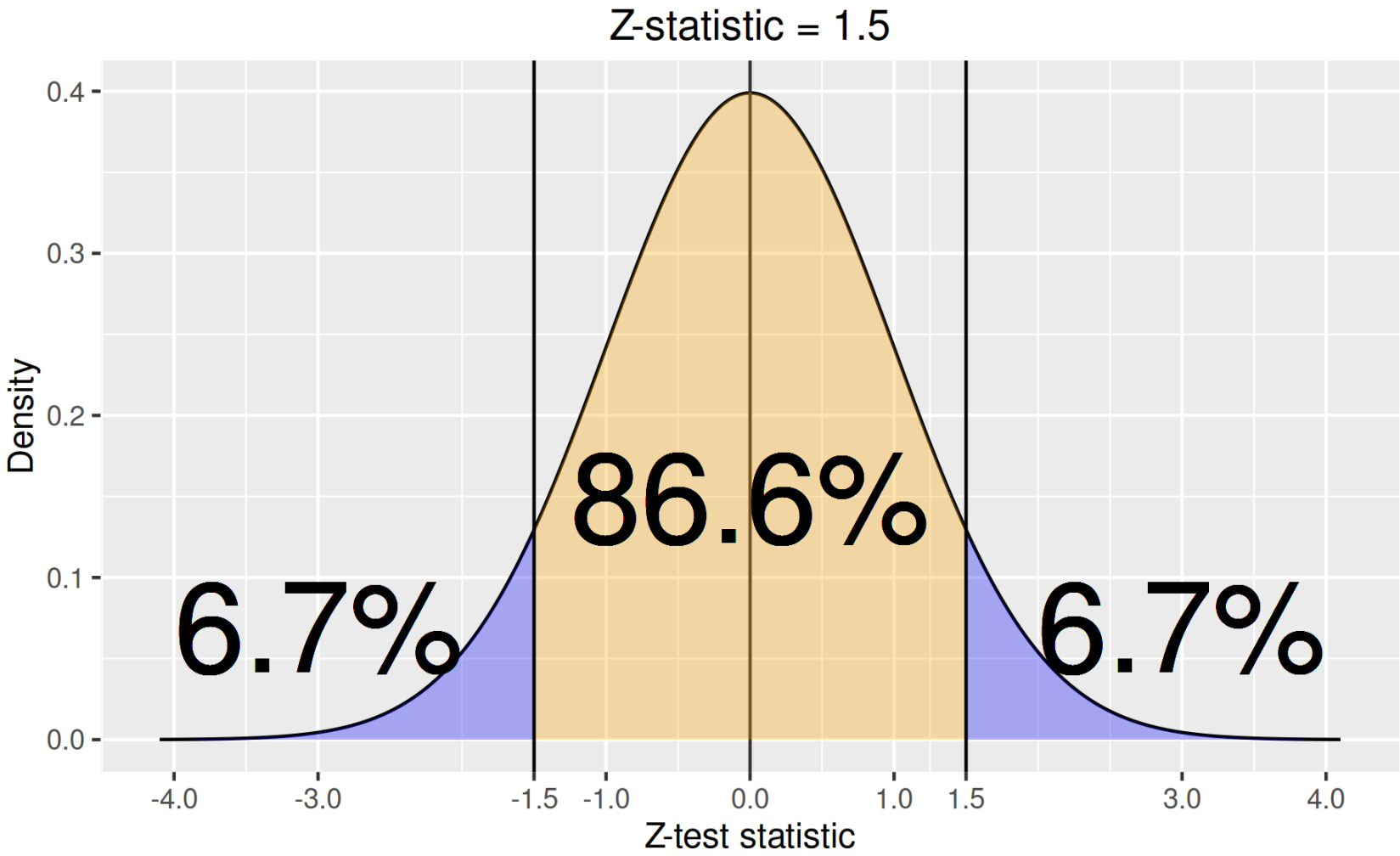
- Null hypothesis (H_0)
- Alternative hypothesis (H_1)

4. **Test statistic**
5. **P-value** (using sampling distribution of the test statistic)
6. Substantive **conclusion** (inference in the population)

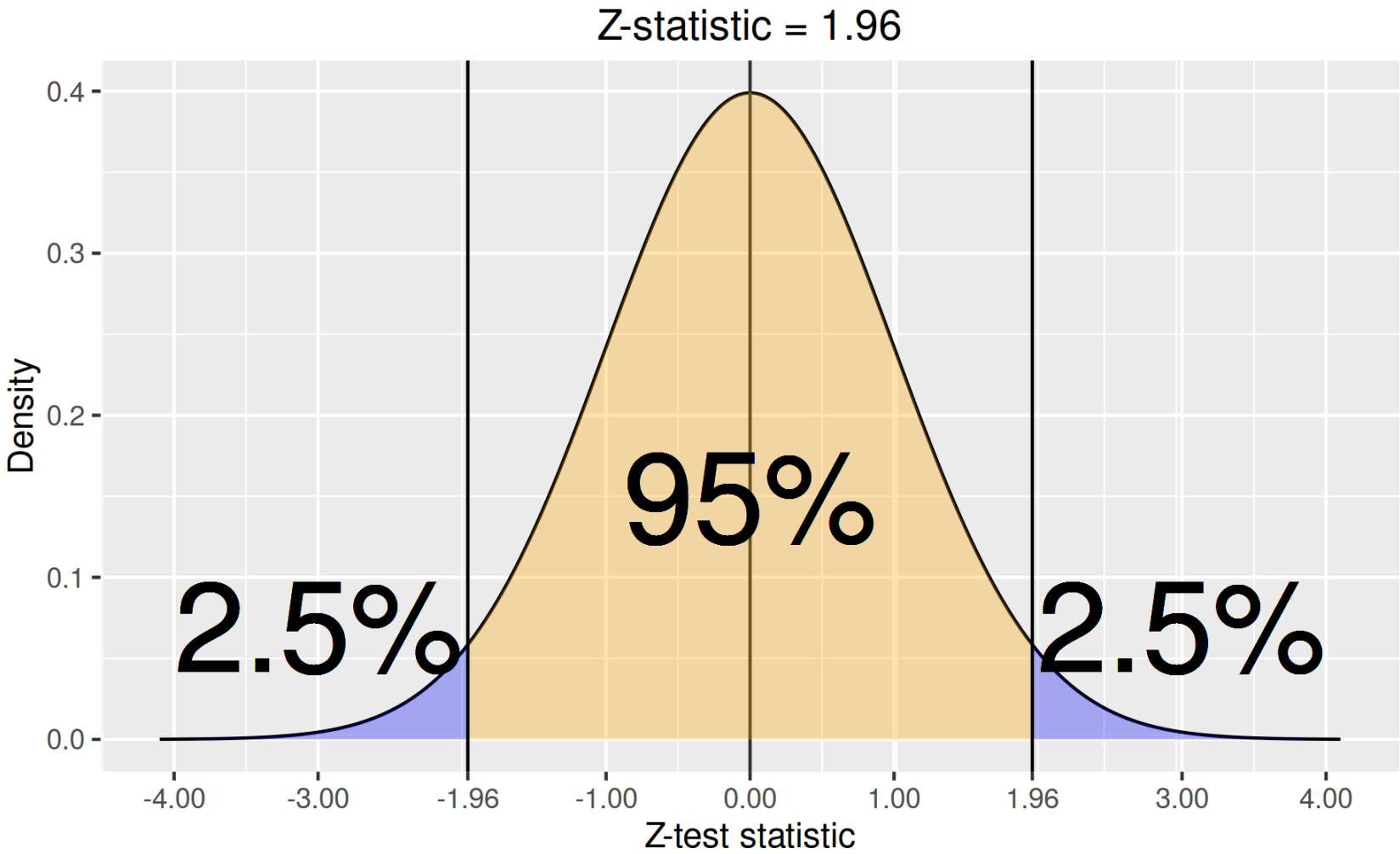
Standard Normal Distribution Probabilities



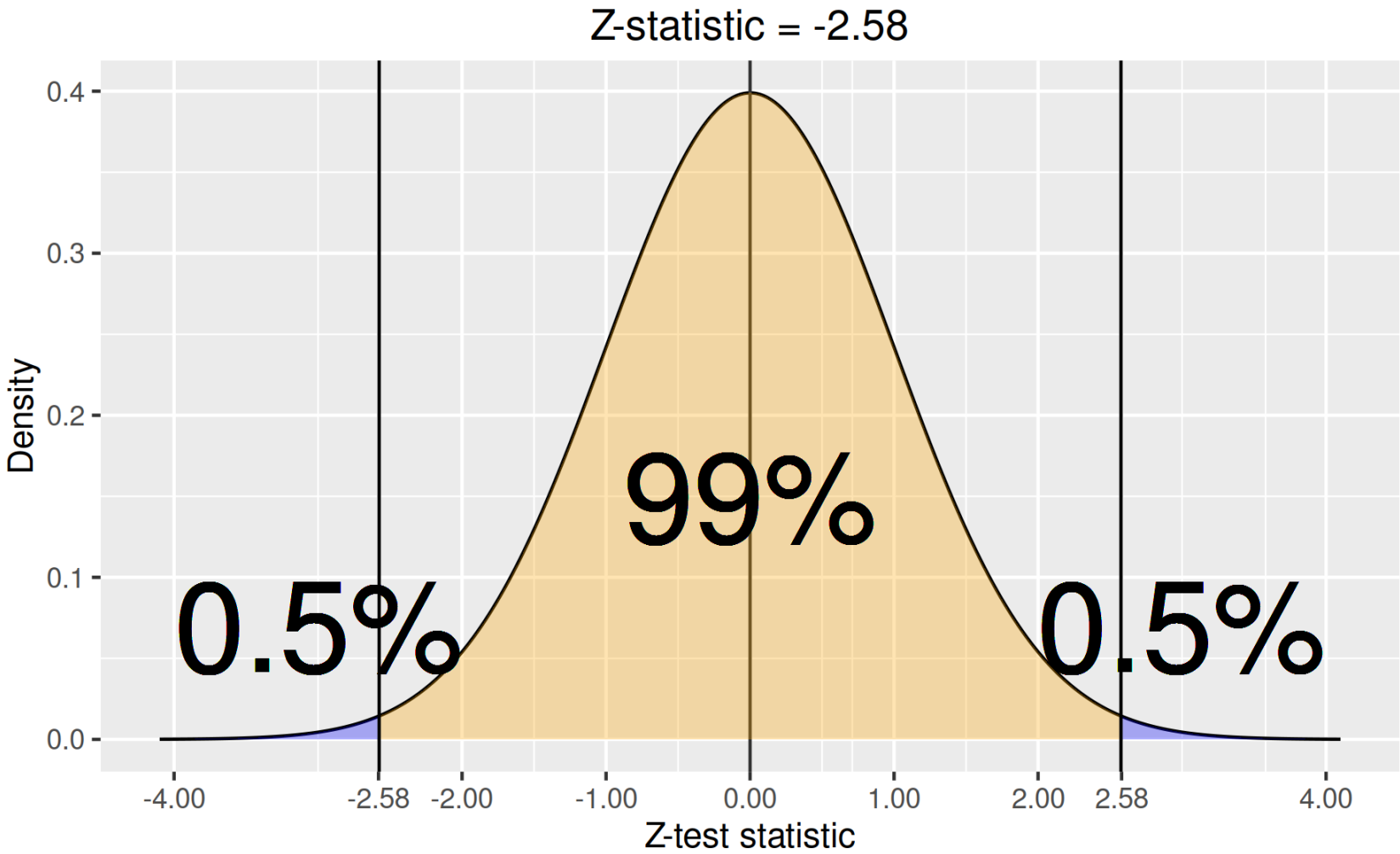
Standard Normal Distribution Probabilities



Standard Normal Distribution Probabilities



Standard Normal Distribution Probabilities



Some useful formulas (Z-test this week)

- Z-test statistic: $z = \frac{\hat{\Delta}}{\hat{\sigma}_{\hat{\Delta}}} = \frac{\hat{\pi} - \pi_0}{\hat{\sigma}_{\hat{\Delta}}}$
- Estimated standard error: $\hat{\sigma}_{\hat{\Delta}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$
- Z-test statistic (two-sample): $z = \frac{\hat{\Delta}}{\hat{\sigma}_{\hat{\Delta}}} = \frac{\hat{\pi_1} - \hat{\pi_2}}{\hat{\sigma}_{\hat{\Delta}}}$
- Estimated standard error (two-sample): $\hat{\sigma}_{\hat{\Delta}} = \sqrt{\hat{\pi}(1-\hat{\pi})(\frac{1}{n_1} + \frac{1}{n_2})}$
- Estimated standard error (two-sample) for confidence intervals: $\hat{\sigma}_{\hat{\Delta}} = \sqrt{\frac{\hat{\pi_1}(1-\hat{\pi_1})}{n_1} + \frac{\hat{\pi_2}(1-\hat{\pi_2})}{n_2}}$
- Confidence intervals: $\hat{\Delta} \pm z_{a/2}\hat{\sigma}_{\hat{\Delta}}$
- Confidence intervals (95%, standard normal): $\hat{\Delta} \pm 1.96\hat{\sigma}_{\hat{\Delta}}$

Exercise 1

Decided to vote Lib Dem

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No (vote other party)	2387	56.2	74.0	74.0
	Yes (vote Lib Dem)	839	19.8	26.0	100.0
	Total	3226	76.0	100.0	
Missing	Not asked	1021	24.0		
Total		4247	100.0		

Standard Normal Distribution Probabilities