Seminar 5: Understanding statistical inference

University of Kent
School of Mathematics, Statistics and Actuarial Science

What is Statistical Inference?

- The process of drawing conclusions about population parameters based on a sample taken from the population.

Motivating Example

- Suppose a company sold packets of nuts and the manager of the company is worried that they are wasting money and they are putting too many nuts into each bag compared to the advertised 30g per pack.
- What we want to be able to do is to collect a sample of packets of nuts and record the weight of the nuts contained in each bag.
- We then want to use a hypothesis test to determine if the mean weight of nuts in a bag is **significantly** more than the advertised weight of 30g.

Motivating Example

- We start by setting up null and alternative hypotheses.
- The null hypothesis is the scenario we are trying to provide the evidence against – so in this example it would be:
  - H_0 : mu = 30g
- The alternative is what we are trying to prove – in this case that the company are putting in too many nuts:
  - H1: mu > 30

Significance level

- You should decide your significance level before starting your hypothesis test.
- In general terms, the significance level is a measure of the strength of evidence that must be present in your sample before you will reject the null hypothesis and conclude that the effect is statistically significant.
- In statistical terminology, the significance level is the probability of rejecting the null hypothesis when the null hypothesis is true.

Significance level

- For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.
- Lower significance levels indicate that you require stronger evidence before you will reject the null hypothesis.

Motivating Example

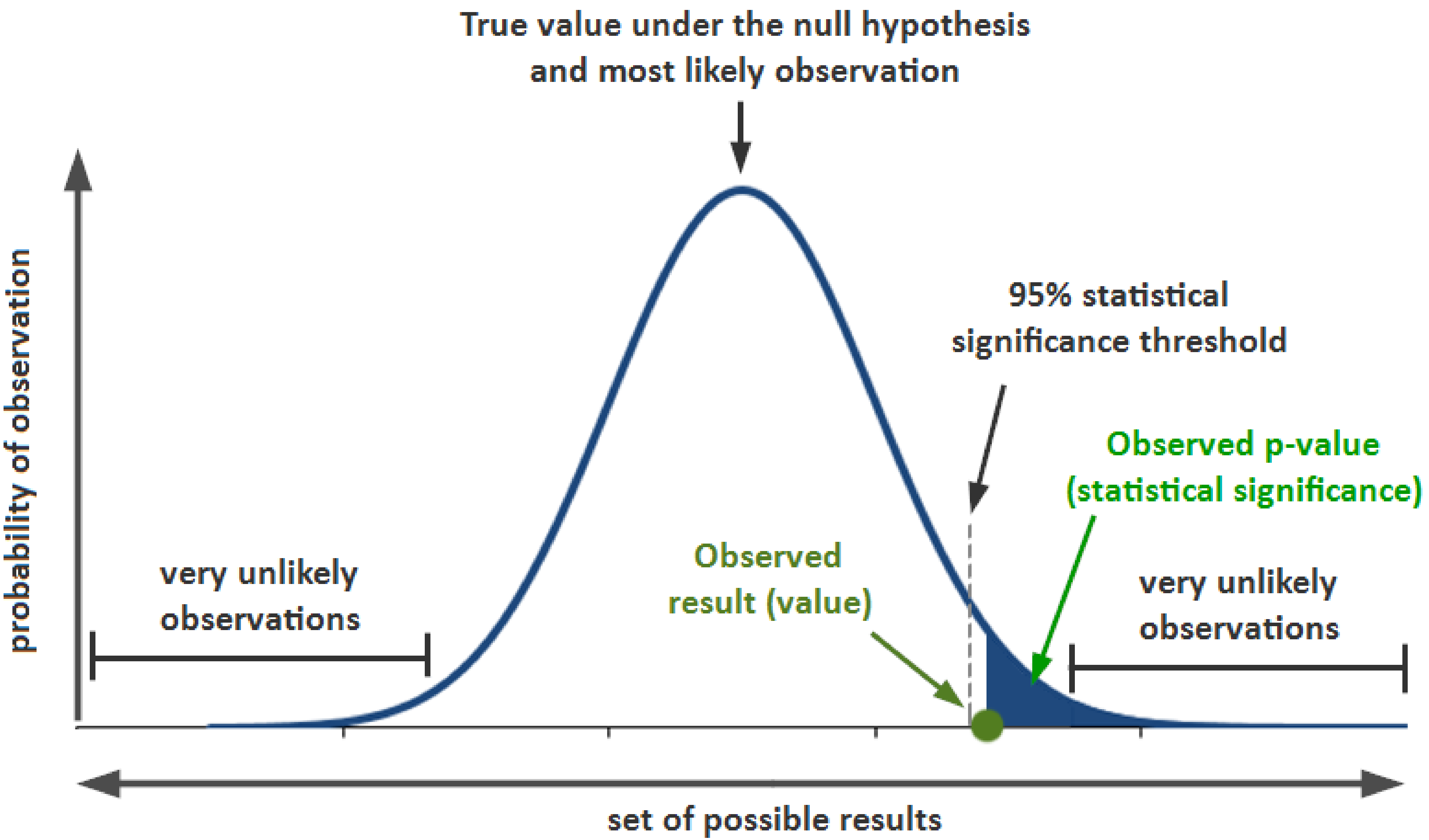
- Suppose that we now take a sample of 50 packets of nuts and weigh the contents.
- If all of the packets of nuts weigh less than 30g then it will be quite clear that the packets of nuts do not contain too many nuts as it would be very unlikely to select 50 bags of nuts with less than 30g if the mean weight of nuts was more than 30g.

Motivating Example

- Conversely if all of the packets have weights more than 30g it will be quite clear that the mean weight of nuts is more than 30g and the manager is right to be concerned.
- But, what if some of the bags weigh more than 30g and some weigh less than 30g?
- From our sample of 50 bags the mean weight of the bags was 30.6g.
- Does this provide enough evidence that the bags have too many nuts in?
- We can perform a hypothesis test to examine this question.

P-Values

- When performing a hypothesis test, the output from the test will be a **p-value**.
- The p-value helps you judge the strength of evidence (what the data are telling you about your population).
- The p-value evaluates how compatible your data are with the null hypothesis – i.e. how likely is the value you observed from your sample data if the null hypothesis is true.


True value under the null hypothesis
and most likely observation
probability of observation
95% statistical
significance threshold
Observed p-value
(statistical significance)
Observed
result (value)
very unlikely
observations
very unlikely
observations
set of possible results

Significance Levels and P-values

We compare the p-value to our pre-defined significance level.

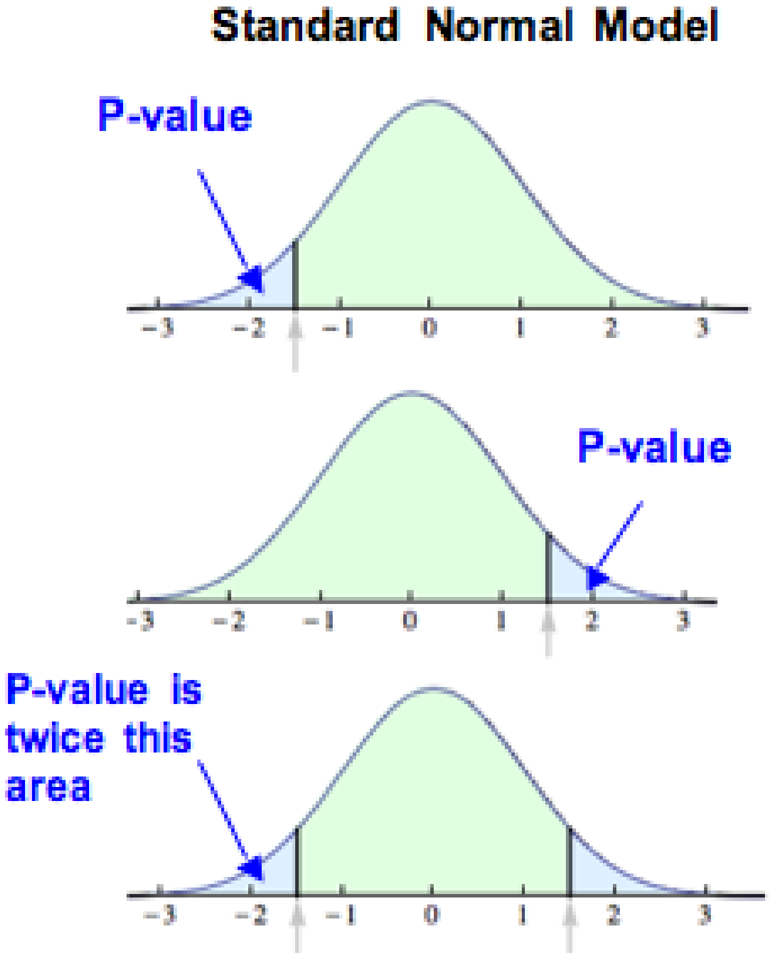
- If the p-value > significance level we have insufficient evidence to reject the null hypothesis
- If the p-value < significance level we can reject the null hypothesis and conclude that the alternative hypothesis is true.

Motivating Example

- Suppose the p-value from our hypothesis test is 0.16.
- This suggests that there is a 16% chance of getting a sample mean at least as large as 30.6 when the true value of the mean weight of nuts is 30g.
- **We compare the p-value to our pre-defined significance level.**
- Because the p-value (0.16) > significance level (0.05) we do not have enough evidence to reject the null hypothesis.

One-sided and two-sided tests

- Whether we perform a one or two-sided test will be determined by the form of the alternative hypothesis.
- In the motivating example we had a one-sided test:
    - $H_1\!: \mu > \mu_0$



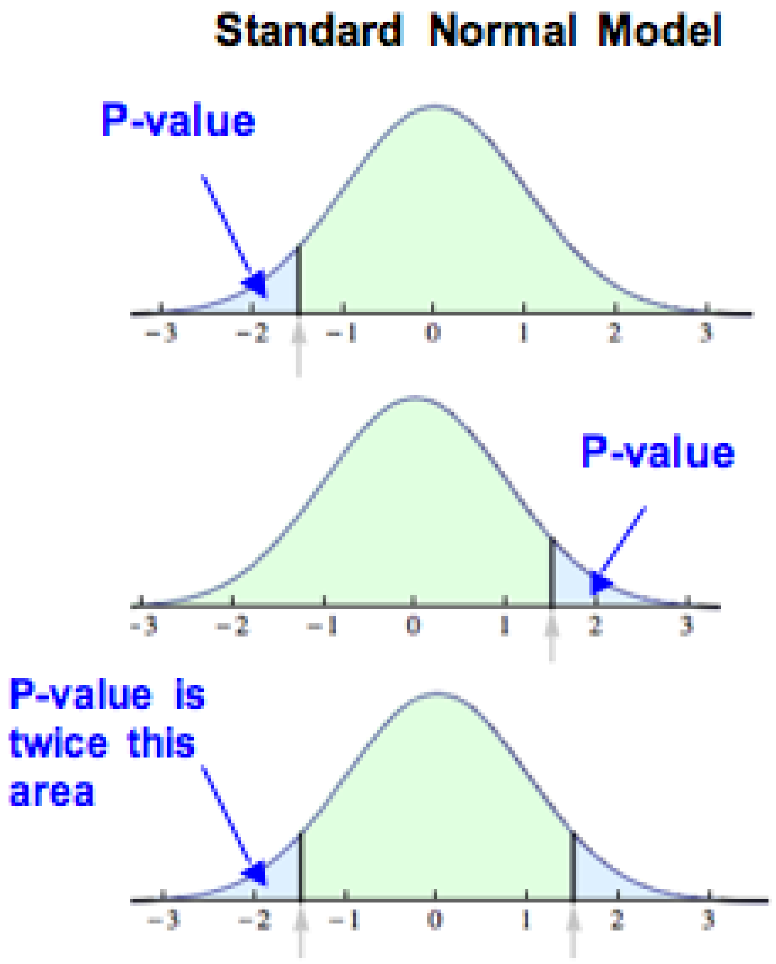
$H_1\!: \mu > \mu_0$

$H_1\!: \mu > \mu_0$

$H_1\!: \mu \neq \mu_0$

One-sided and two-sided tests

- For a two-sided test you split the significance level percentage between both tails of the distribution.
- Two-sided tests have reduced statistical power, however should be used as a default unless you are addressing a specific question of whether the true population is larger or smaller.



$H_1\!: \mu > \mu_0$

$H_1\!: \mu > \mu_0$

$H_1\!: \mu \neq \mu_0$

Hypothesis Testing using R

- During computer session 8 you will be performing hypothesis tests for the population mean and the population proportion. You will be testing whether these parameters take specific values.
  - One sample t-test
  - One sample proportion test

- During computer session 9 you will be performing hypothesis tests to compare population means and population proportions from two populations.
  - Two sample t-test
    - Independent and paired t-tests
  - Two sample proportion test