Curtin College

DIPLOMA OF INFORMATION TECHNOLOGY

IPDA1005 INTRODUCTION TO PROBABILITY AND DATA ANALYSIS

Acknowledgement

We respectfully acknowledge the Elders and custodians of the Whadjuk Nyungar nation, past and present, their descendants and kin. Curtin College Bentley Campus enjoys the privilege of being located in Whadjuk / Nyungar Boodjar (country) on the site where the Derbal Yerrigan (Swan River) and the Djarlgarra (Canning River) meet. The area is of great cultural significance and sustains the life and well being of the traditional custodians past and present.

Outline

1. Estimating sample size for estimating μ

2. Hypothesis Testing
   - Null and Alternative Hypotheses
   - The p-value
   - The Critical Values Method

3. Matched Pairs

Estimating Sample Size for estimating μ

- A confidence interval for a population mean has limits given by

$$x \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

- To estimate a sample size that will achieve a margin of error *e* with confidence level *100(1 − α)%,* we solve for *n*:

$$e = \frac{t_{n-1,\frac{\alpha}{2}} s}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{t_{n-1,\frac{\alpha}{2}} s}{e} \Rightarrow n = \left(\frac{t_{n-1,\frac{\alpha}{2}} s}{e}\right)^2$$

- There are 2 practical problems with this formula:
  - the value $t_{n-1,\frac{\alpha}{2}}$ requires us to already know the sample size we are trying to estimate
  - we don't have a sample standard deviation because haven't selected the sample yet.

Estimating Sample Size for estimating μ

- Bearing in mind that sample size estimation is approximate at best:
  - we replace $t_{n-1,\frac{\alpha}{2}}$ with $z_{\frac{\alpha}{2}}$
  - we estimate s from a trial sample, or a sample from a similar population.
- Our formula is now –

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma^*}{e}\right)^2$$

- where σ∗ is an estimate of population standard deviation (however obtained)
- The result should be **rounded up** to an integer since we want the error to be no more than the target *e*.

Estimating Sample Size for estimating μ - example

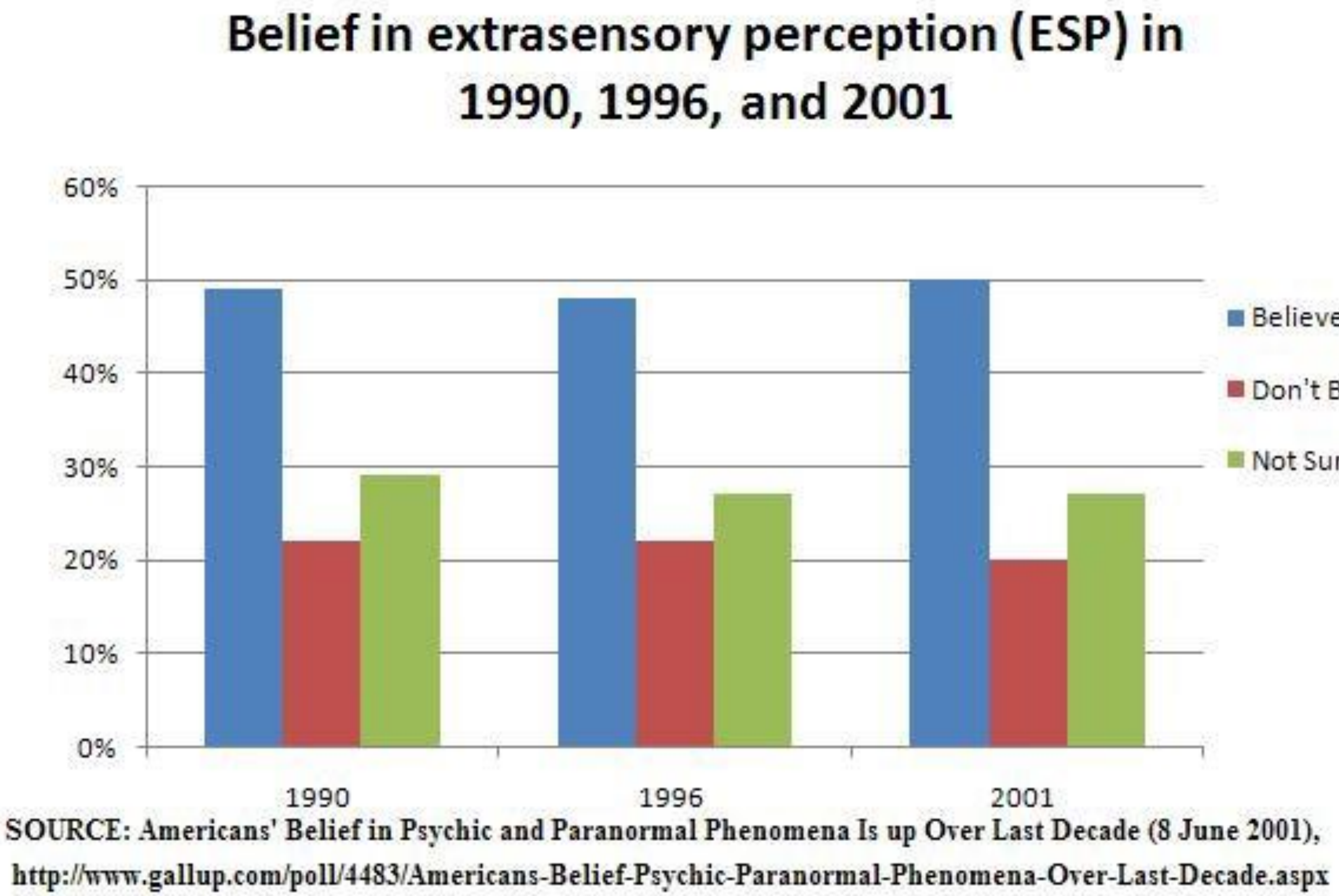
- **Example 1**
- Suppose we need to construct a 98% confidence interval for μ with a margin of error no more than .0001.
- It is known that $\sigma = 0.0002$.
- The required sample size is calculated as follows:

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma^*}{e}\right)^2 = \left(\frac{2.326(.0002)}{.0001}\right)^2 = 21.64.$$

- So the required sample size is 22

Hypothesis Testing

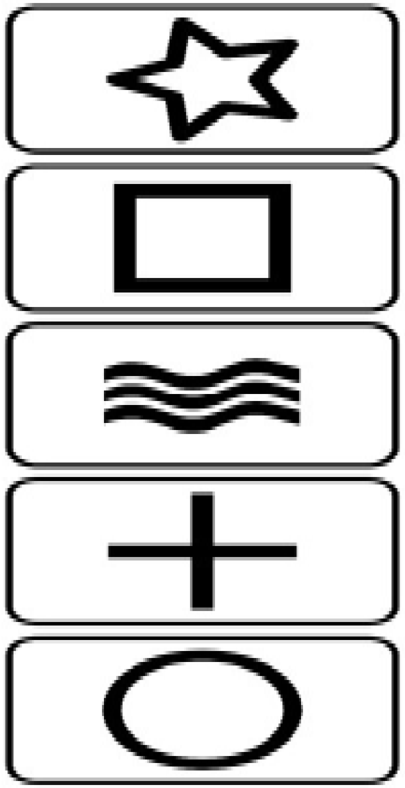
- The second main type of statistical inference is hypothesis testing (sometimes called *significance testing*), in which sample data are used to evaluate a claim about a population parameter.
- The basic idea is this:
  - We assume the truth of a claim, and on that basis calculate the chance that the observed sample result could have occurred.
  - If the observed result is *relatively likely to occur* when the claim is true, then we say the claim is *consistent* with the data. That is, the data provide insufficient reason to reject the claim
  - But if the observed result is *unlikely to occur* when the claim is true, then we say the claim is *inconsistent* with the data, and we have reason to reject the claim.



- From Lock et al. (2012) Statistics: Unlocking the Power of Data. Wiley: New York

Formulating hypotheses - ESP example

- One way of testing for ESP is with Zener cards.
- There are five cards, each with a different symbol.
- Subjects draw a card at random and try to mentally communicate the symbol to someone in the next room who then guesses the symbol.
- If there is no such thing as ESP, what should be the population proportion of correct guesses?



Hypothesis Testing

- A hypothesis test requires a statistic that follows a known sampling distribution. If we start from the assumption that ESP is not true, and hence that the guesses are purely random, we would expect that P(successful guess) = 0.2.
- If p = 0.2 and we intend to test n times, then $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$
- The hypothesis that ESP is not true (and hence p = 0.2) serves as our initial hypothesis here because:
  - it indicates a specific sampling distribution for which we can calculate probabilities, and
  - we may prefer to accept ESP only if we find strong evidence in its favour.
- We call our initial hypothesis the null hypothesis, with notation H_0.
- If ESP were real, we would expect P(success) > 0.2. This is our alternative hypothesis, with notation H_1 or H_a.

Hypothesis Testing - Null and Alternative Hypotheses

- for the ESP example, our hypothesis statements are:

$$H_0: p = 0.2 \qquad\qquad H_1: p > 0.2$$

- where p is the population proportion of correct guesses.
- Note that we defined what we meant by p. If you don't define it, you must write it out in words in both hypotheses.
- H_0 is what we will accept initially, and keep accepting if we don't find strong enough evidence against it. (This is like the situation in Australian law courts.)
- H_1 is what we will believe if we reject H_0 because we did find sufficiently strong evidence against it.

Hypothesis Testing - Null and Alternative Hypotheses

- **Example**: Nancy claims that the coin in her pocket is fair, meaning, on average, it comes up heads half the time. She seeks evidence by tossing the coin lots of times and counting the number of heads.
- "Coin is fair" implies a specific value for *P(Head)*, so we can use this as H_0. "Coin is not fair" could not serve as H_0 because it does not give us a value for p and so does not indicate a sampling distribution against which to test a statistic.
- In this example the hypotheses will be:

$$H_0: p = 0.5 \qquad\qquad H_1: p \neq 0.5$$

- where p is the population proportion of heads arising from tossing Nancy's coin.
- The choice of alternative hypothesis may depend on
  - what we fear may be true
  - what we would really like to have evidence in favour of
  - the specific wording of the research question

Hypothesis Testing

- Georges-Louis Leclerc, Comte de Buffon (1707 – 1788) tossed a coin 4040 times and got 2048 heads
- The sample proportion is

$$\hat{p} = \frac{2048}{4040} \approx 0.5069307$$

- Is that far enough away from p = 0.5 to doubt the fairness of a coin?
- We will do a complete hypothesis test on the fairness of a coin using this data.

Hypothesis Testing

- Our hypotheses are $H_0: p = 0.5 \qquad H_1: p \neq 0.5$ where p is the population proportion of heads in random tosses of the coin.
- We proceed as if H_0 were true.
- *If*:
  - H_0 is true
  - and *n* is large
- then $\hat{p} \sim N\left(0.5, \frac{0.5\times0.5}{n}\right)$. Equivalently, $\frac{\hat{p}-0.5}{\sqrt{\frac{0.5\times0.5}{n}}} \sim N(0,1)$
- n = 4040 is a very large sample size, so the size condition is met.
- If H_0 is true, the standardised $\hat{p}$ follows a standard Normal distribution

Hypothesis Testing - test statistic

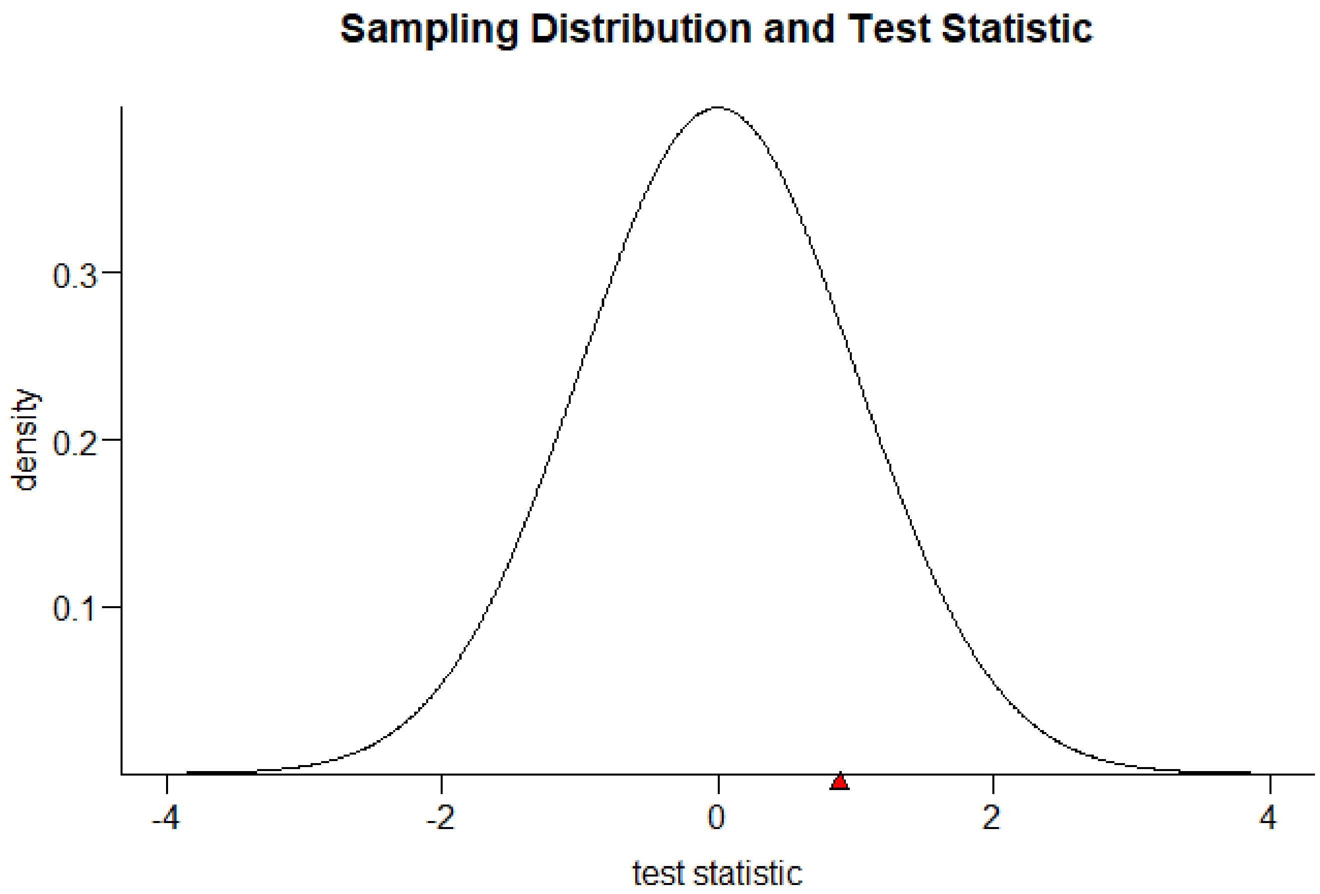
- The observed proportion was $\hat{p} = \dfrac{2048}{4040} = 0.5069307$
- $SE(\hat{p}) = \sqrt{\dfrac{p(1-p)}{n}} == \sqrt{\dfrac{0.5 \times 0.5}{4040}} = 0.00786646$
- So the standardised $\hat{p}$ is $\dfrac{\hat{p}-p}{SE(\hat{p})} = \dfrac{0.5069307-0.5}{0.00786646} = 0.8810$

- This is our test statistic since we measure (or test) its extremeness against the sampling distribution of standardised sample proportions, which is Standard Normal.
- You should already realise that a Z-score of 0.881 is not very extreme.
- However, we need a formal measure for this.

Hypothesis Testing - test statistic

- The observed proportion was $\hat{p} = \dfrac{2048}{4040} = 0.5069307$
- $SE(\hat{p}) = \sqrt{\dfrac{\text{p(1−p)}}{n}} == \sqrt{\dfrac{0.5\times 0.5}{4040}} = 0.00786646$
- So the standardised $\hat{p}$ is $\dfrac{\hat{p}-p}{SE(\hat{p})} = \dfrac{0.5069307-0.5}{0.00786646} = \ 0.8810$
- This is our test statistic since we measure (or test) its extremeness against the sampling distribution of standardised sample proportions, which is Standard Normal.
- You should already realise that a Z-score of 0.881 is not very extreme.
- However, we need a formal measure for this.

Hypothesis Testing - test statistic

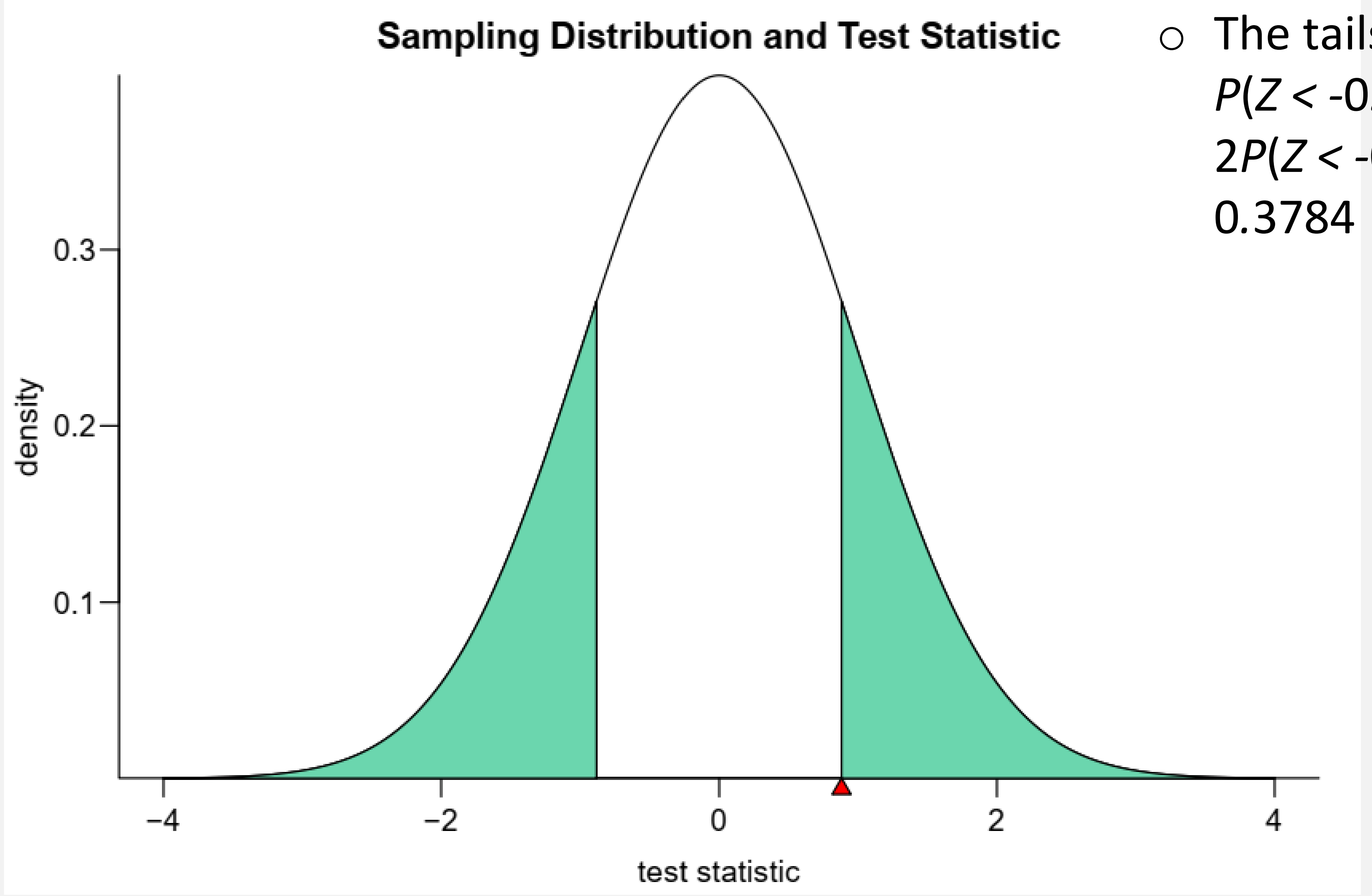


Test statistic = 0.881

Hypothesis Testing

- Our formal measure is the proportion of the sampling distribution that is at least as extreme as the statistic we observed.
- The meaning of “extreme” is indicated by H_1. We have a two-sided H_1, where alternative values of p could be either less than or greater than .5, so we consider both extremes of the sampling distribution.
- Our test statistic is 0.881, so the relevant proportion of the sampling
- distribution is where Z > |0.881|
- This is illustrated on the next slide.

Hypothesis Testing – p-value



- The tails of the distribution are $P(Z < -0.881) + P(Z > 0.881) = 2P(Z < -0.881) = 2 \times 0.1892 = 0.3784$

Hypothesis Testing

- In hypothesis testing our evidence against H_0 is the chance of getting a statistic at least as extreme as the observed test statistic. We have calculated the relevant proportion of the sampling distribution to be *0.3784*, but if we are to interpret this as a probability we must be satisfied that we were equally likely to obtain any of the possible samples of size *n* from the population. That is, we need simple random sampling (SRS).
- Therefore, SRS is a crucial assumption for hypothesis tests.
- If our sample was randomly selected (in the sense of SRS), then under H_0 there was about a 38% chance of getting a result at least as extreme as what we observed. This is called the p-value of the test.
- 38% is a relatively high chance, so we would not treat this as providing much evidence against H_0.

p-value and Statistical Significance

To summarise what we have learned so far:

- A null hypothesis H_0 is a claim that we initially (and provisionally) adopt until we encounter strong evidence against it.
- Evidence against a null hypothesis is the extremeness of the test statistic in relation to its sampling distribution.
- The measure of extremeness is the smallness of the p-value, which is the probability of getting a test statistic at least as extreme as the one actually observed.
- Small p-values are evidence against H_0 because they say that the observed result is unlikely to occur when H_0 is true.
- Large p-values give little or no evidence against H_0 because they say that results like the one observed are fairly likely to occur when H_0 were true.

p-value and Statistical Significance

- Our conclusion in a significance test comes down to:
  - p-value small → reject H_0 → conclude H_1
  - p-value large → fail to reject H_0 → cannot conclude H_1
- Our criterion for when a p-value is small enough for us to reject H_0 is called the significance level, and is denoted by α.
- We compare the p-value with α and reject the null hypothesis in favour of the alternative when p-value < α. Such a result is "statistically significant".
- When p-value > α, our conclusion is that we don't reject H_0. We avoid saying anything positive about H_0, because we do not have evidence that it is true. We can say that it is "plausible".
- It is possible that we may fail to reject two inconsistent null hypotheses (such as p = 0.5 and p = 0.4), and we clearly cannot say that we accept both.

Hypothesis test conclusions

Hypothesis test conclusions have two or three elements:

- The decision about H_0, with reasons (basically, comparison with the significance level)
- Interpretation of that decision in the context of the real-world problem.
- If H_0 is rejected - the message from the sample, e.g., estimate of population proportion or population mean.

In the coin example, with p-value = 0.3784, we conclude:

”Since p-value .05 we do not reject H_0. There is little evidence that the coin is unfair.”

There are many ways of saying the second sentence. They should all offer an evaluation of evidence and avoid unequivocal statements about the truth or falsity of H_0

Hypothesis testing - assumptions

The crucial evidence in a hypothesis test is the p-value. Its use as evidence depends on -

- H_0 being assumed true
- the sampling distribution actually occurring (so any relevant conditions must be met)
- sample selection being simple random

If we do not check that sampling was random, and that sampling distribution conditions have been met, then a small p-value will merely indicate a likely problem somewhere among these three elements. To treat the p-value as evidence against H_0 we must be satisfied regarding the sampling and sampling distribution conditions.

Steps for Hypothesis Testing

1. Translate the research question into null and alternative hypotheses, carefully defining the population parameter(s) being tested.
2. State the significance level you intend to use.
3. State the conditions required by the test you plan to use, and check that the conditions are met.
4. Calculate the test statistic.
5. Calculate the p-value.
6. Conclusion I: make your decision about H0, with reasons.
7. Conclusion II: interpret that decision in the practical context of the research question.
8. Conclusion III: if H0 was rejected, give an estimate of the population parameter.

Birth Ratio Example

A random sample from the USA was collected in 2005 to investigate whether newborns are more likely to be boys. It was found that among 25468 firstborn children, 13173 were boys. Do the data provide evidence that firstborn children are more likely to be boys?

- The research question asks whether the data provide evidence that something is the case. A hypothesis test cannot positively support H_0, so a prevalence of boys will be our alternative hypothesis.

  H_0 : p = 0.5 H_1 : p > 0.5

- where p is the proportion of boys among firstborn children in the USA.
- The hypothesis test requires a large simple random sample. The sample size of n = 25468 is very large indeed, and were are given that sampling was random.

A random sample from the USA was collected in 2005 to investigate whether newborns are more likely to be boys. It was found that among 25468 firstborn children, 13173 were boys. Do the data provide evidence that firstborn children are more likely to be boys?

- The research question asks whether the data provide evidence that something is the case. A hypothesis test cannot positively support H_0, so a prevalence of boys will be our alternative hypothesis.

  H_0 : p = 0.5 H_1 : p > 0.5
- where p is the proportion of boys among firstborn children in the USA.
- The hypothesis test requires a large simple random sample. The sample size of n = 25468 is very large indeed, and were are given that sampling was random.

Birth Ratio Example

A random sample from the USA was collected in 2005 to investigate whether newborns are more likely to be boys. It was found that among 25468 firstborn children, 13173 were boys. Do the data provide evidence that firstborn children are more likely to be boys?

- The research question asks whether the data provide evidence that something is the case. A hypothesis test cannot positively support H_0, so a prevalence of boys will be our alternative hypothesis.

  H_0 : p = 0.5 H_1 : p > 0.5

- where p is the proportion of boys among firstborn children in the USA.
- The hypothesis test requires a large simple random sample. The sample size of n = 25468 is very large indeed, and were are given that sampling was random.

Hypothesis Testing Using Critical Values

- An older way of performing the hypothesis tests is the critical value method, which does not use p-values.
- The critical value method rejects H_0 : p = p0 in favour of H_1 if
  - $z \geq z_\alpha$ whenH_1: p $> p_0$
  - $z \leq z_\alpha$ when H_1: p $< p_0$
  - $|z| \geq z_{\alpha/2}$ when H_1: p $\neq p_0$
- The set of test statistic values that lead to the rejection of H_0 is called the rejection region or the critical region.
- In the coin-tossing example, the test statistic of 0.881 does not fall in the rejection region $(-\infty, -1.96) \cup (1.96, \infty)$, so we do not reject H_0.
- Choice of p-value or critical value method makes no difference to the
- basic decision to reject or not-reject H_0.
- **Exercise**: Repeat the birth ratio example using a critical value approach

An Example

- **Example**: It is hypothesized that no more than 6% of the parts being produced in a manufacturing plant are defective. For a random sample of 200 parts, 19 are found to be defective. Using 1%, 5% and 10% levels of significance, are the data consistent with this hypothesis?
- **Solution**: The research question asks about consistency, which indicates the null hypothesis.
  - H_0 : p $\leq$ 0.06 H_1 : p $>$ 0.06
- where p is the proportion of defective parts produced in the manufacturing plant.
- Even though the hypothesis is p ≤ 0.06, we use the value at the boundary between H_0 and H_1, namely 0.06. Thus, it makes no difference whether we state H_0 as p ≤ 0.06 or p = 0.06.
- The hypothesis test requires a sufficiently large random sample. We are given that sampling was random, and both np_0 and $n(1 - p_0)$ are larger than 10, the sample size is adequate. Thus the relevant conditions are met.

An Example

$$\hat{p} = \frac{19}{200} = 0.095$$
$$z = \frac{.095 - .06}{\sqrt{\frac{.06(.94)}{200}}} = 2.08.$$

- As $z_{.01} = 2.326$, $z_{.05} = 1.645$ and $z_{.1} = 1.282$, we would reject the null hypothesis at the 5% and 10% levels of significance, but not at the 1% level.
- On a p-value approach, the p-value is P (Z > 2.08) ≈ 0.019, so the same decisions about H_0 would follow.
- For the 5% and 10% significance levels we would conclude, “the data provide significant evidence that the defect rate exceeds 6%, with the estimated defect rate being 9.5%.”
- At a 1% signifiance level we would conclude, “the data are consistent
- with a defect rate not exceeding 6%.”

Exercises

- A state's Department of Juvenile Corrections believes that less than 20% of offenders admitted to its training schools have been convicted of car theft. Scrutiny of a random sample of 125 admission records reveals that 19 of them were for car theft.

a) Do these figures lend support to the Department's belief? Use a 5% level of significance.

b) Find the *p*-value for the test.

c) Construct a 95% confidence interval for the proportion of car thieves among the offenders admitted to the training schools.

Exercises

- It is thought that 40% of the adult population are coffee drinkers. A random sample of size 240 contained 80 coffee drinkers.

a) Test the hypothesis that 40% of the adult population are coffee drinkers against the one-sided alternative that the proportion is less than 40%. Use a 5% level of significance.

b) Find the p-value for the test.

c) Construct a 95% confidence interval for the proportion of coffee drinkers among adults.

d) How many people must be questioned if we are to obtain an estimate of the proportion of coffee drinkers which we can be 95% sure is within 2% of the true value?

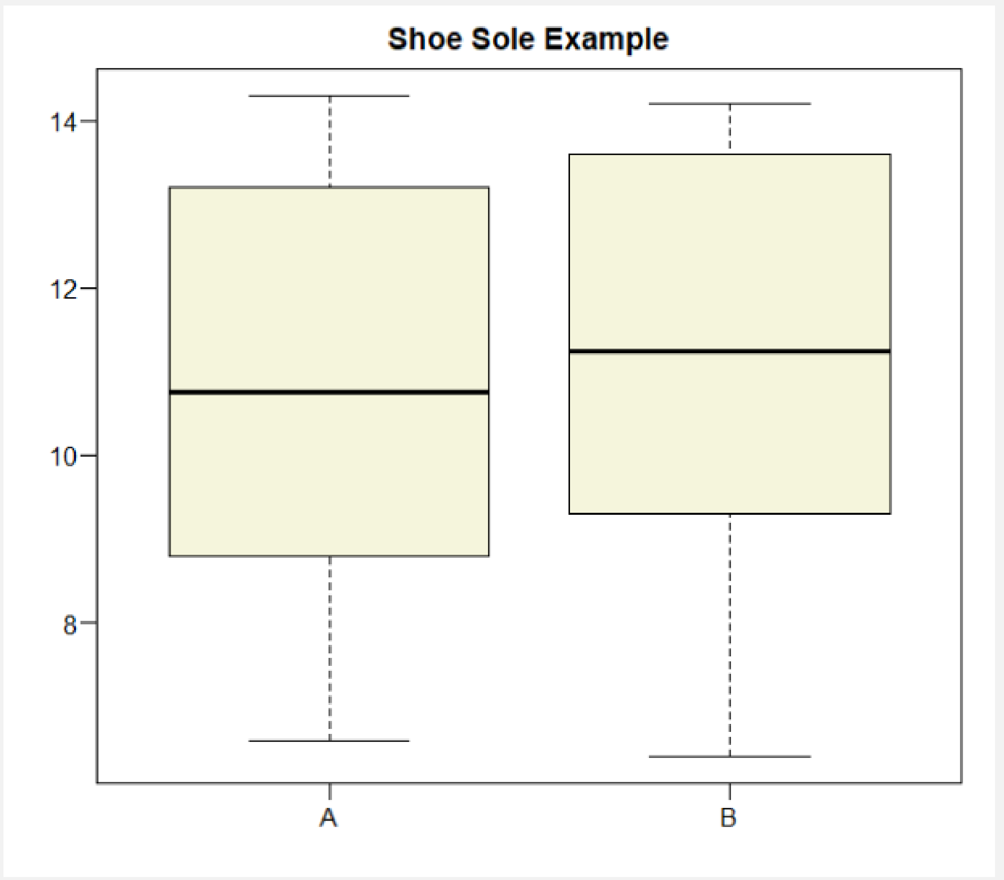
e) Repeat the hypothesis test in with a two-sided alternative using $\alpha = 0.01$

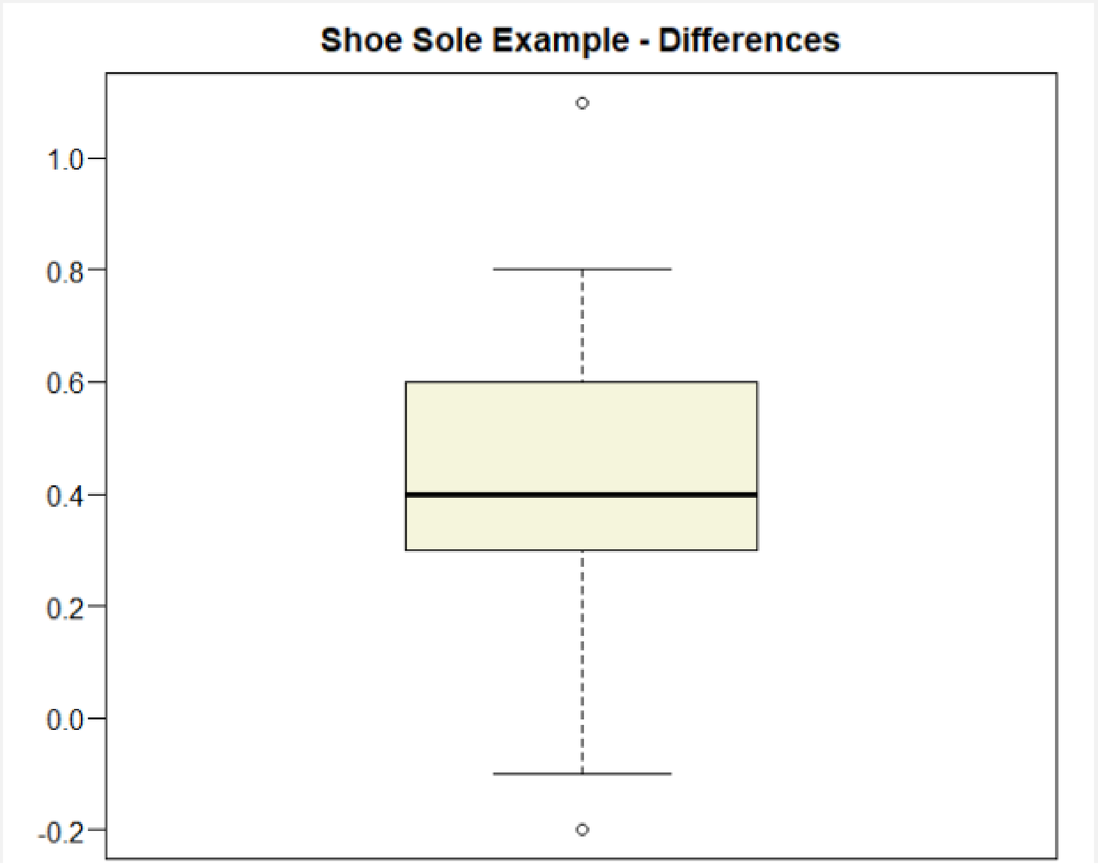
Matched Pairs

- If an investigation aims to show that a treatment causes an effect, a matched pairs design may be used.
- In this experimental design, subjects are matched in pairs and each treatment is given to one subject in each pair, or before-and-after measurements are made on the same subject.
  - In testing the efficacy of a drug to reduce cholesterol levels, we might compare the difference between cholesterol levels before and after the drug was administered.
  - In testing the durability of two different formulations of shoe sole material, we might randomly assign one material to the left shoe of an individual and the other to the right shoe, and then compare the difference in durability.
- Note that the two data sets obtained here are not independent. Rather, each observation in one “sample” is associated with a specific observation in the other “sample” (matched subject, same person, etc)

Examples

A	13.2	8.2	10.9	14.3	10.7	6.6	9.5	10.8	8.8	13.3
B	14	8.8	11.2	14.2	11.8	6.4	9.8	11.3	9.3	13.6





Examples

A	13.2	8.2	10.9	14.3	10.7	6.6	9.5	10.8	8.8	13.3
B	14	8.8	11.2	14.2	11.8	6.4	9.8	11.3	9.3	13.6
$B - A$	0.8	0.6	0.3	-0.1	1.1	-0.2	0.3	0.5	0.5	0.3

- Here $n = 10$, and the mean and the standard deviation of the differences are given by $\bar{x} = 0.41$ and $s = 0.3872$.
- $t_{9,.025} = 2.262$, so a 95% confidence interval for the true mean difference is given by

$$0.41 \pm 2.262 \frac{0.3872}{\sqrt{10}} = (0.13, 0.69).$$

Examples

- The following are average weekly losses of work-hours due to accidents in 10 industrial plants before and after a certain safety program was implemented.
- Construct a 90% confidence interval for the mean reduction in losses of work-hours.

Table: Losses of Work-hours

Before	45	73	46	124	33	57	83	34	26	17
After	36	60	44	119	35	51	77	29	24	11

- **Solution**:
- Calculate the pairwise differences

Before	45	73	46	124	33	57	83	34	26	17
After	36	60	44	119	35	51	77	29	24	11
Difference	9	13	2	5	-2	6	6	5	2	6

Examples

- The difference data set *d* has sample mean $\bar{d} = 5.2$ and sample standard deviation $s_d = 4.08$.
- The sample size is n = 10.
- $t_{9,.05}$= 1.833, so a 90% confidence interval for the true mean reduction is given by

$$5.2 \pm 1.833\frac{4.08}{\sqrt{10}} = 5.2 \pm 2.365 = (2.8, 7.6)$$

- **Exercise**: The following is the hours of sleep 6 patients had before and after taking a sleep-inducing drug. Construct a 99% confidence interval for the mean increase in sleeping hours.

Patient	1	2	3	4	5	6
Before	5	4	6	5.5	6	6.5
After	6.5	4.5	5	6	7.5	4