Hypothesis test

The ASTA team

Contents

1 Statistical inference: Hypothesis and test

1.1 Concept of hypothesis

- A **hypothesis** is a statement about a given population. Usually it is stated as a population parameter having a given value or being in a certain interval.
- Examples:
  - Quality control of products: The hypothesis is that the products e.g. have a certain weight, a given power consumption or a minimal durability.
  - Scientific hypothesis: There is no dependence between a company's age and level of return.

1.2 Significance test

- A significance test is used to investigate, whether data is contradicting the hypothesis or not.
- If the hypothesis says that a parameter has a certain value, then the test should tell whether the sample estimate is "far" away from this value.
- For example:
  - Waiting times in a queue. We sample n customers and count how many that have been waiting more than 5 minutes. The company policy is that at most 10% of the customers should wait more than 5 minutes. In a sample of size $n = 32$ we observe 4 with waiting time above 5 minutes, i.e. the estimated proportion is $\hat{\pi} = \frac{4}{32} = 12.5\%$. Is this "much more" than (i.e. significantly different from) 10%?
  - The blood alcohol level of a student is measured 4 times with the values $0.504, 0.500, 0.512, 0.524$, i.e. the estimated mean value is $\bar{y} = 0.51$. Is this "much different" than a limit of 0.5?

1.3 Null and alternative hypothesis

- **The null hypothesis** - denoted H_0 - usually specifies that a population parameter has some given value. E.g. if μ is the mean blood alcohol level we can state the null hypothesis
  - $H_0 : \mu = 0.5$.
- The **alternative hypothesis** - denoted H_a - usually specifies that the population parameter is contained in a given set of values different than the null hypothesis. E.g. if μ again is the population mean of a blood alcohol level measurement, then
  - the null hypothesis is $H_0 : \mu = 0.5$
  - the alternative hypothesis is $H_a : \mu \neq 0.5$.

1.4 Test statistic

- We consider a population parameter μ and write the null hypothesis

$$H_0 : \mu = \mu_0,$$

  where μ_0 is a known number, e.g. $\mu_0 = 0.5$.
- Based on a sample we have an estimate $\hat{\mu}$.
- A **test statistic** T will typically depend on $\hat{\mu}$ and μ_0 (we may write this as $T(\hat{\mu}, \mu_0)$) and measure "how far from μ_0 is $\hat{\mu}$?"
- Often we use $T(\hat{\mu}, \mu_0) =$ "the number of standard deviations from $\hat{\mu}$ to μ_0".
- For example it would be very unlikely to be more than 3 standard deviations from μ_0, i.e. in that case μ_0 is probably not the correct value of the population parameter.

1.5 P-value

- We consider
  - H_0: a null hypothesis.
  - H_a: an alternative hypothesis.
  - T: a test statistic, where the value calculated based on the current sample is denoted t_{obs}.
- To investigate the plausibility of H_0, we measure the evidence against H_0 by the so-called p-value:
  - The p-value is the probability of observing a more extreme value of T (if we were to repeat the experiment) than t_{obs} *under the assumption that H_0 is true.*
  - "Extremity" is measured relative to the alternative hypothesis; a value is considered extreme if it is "far from" H_0 and "closer to" H_a.
  - If the p-value is small then there is a small probability of observing t_{obs} if H_0 is true, and thus H_0 is not very probable for our sample and we put more support in H_a, so:

The smaller the p-value, the less we trust H_0.

- What is a small p-value? If it is below 5% we say it is **significant** at the 5% level.

1.6 Significance level

- We consider
  - H_0: a null hypothesis.
  - H_a: an alternative hypothesis.

- T: a test statistic, where the value calculated based on the current sample is denoted t_{obs} and the corresponding p-value is p_{obs}.

- Small values of p_{obs} are critical for H_0.
- In practice it can be necessary to decide whether or not we are going to reject H_0.
- The decision can be made if we previously have decided on a so-called **α-level**, where

  - α is a given percentage
  - we reject H_0, if p_{obs} is less than or equal to α
  - α is called the **significance level** of the test
  - typical choices of α are 5% or 1%.

1.7 Significance test for mean

1.7.1 Two-sided t-test for mean:

- We assume that data is a sample from `norm`(μ, σ).
- The estimates of the population parameters are $\hat{\mu} = \bar{y}$ and $\hat{\sigma} = s$ based on n observations.
- Null hypothesis: $H_0: \ \mu = \mu_0$, where μ_0 is a known value.
- **Two-sided alternative hypothesis**: $H_a: \ \mu \neq \mu_0$.
- Observed test statistic: $t_{obs} = \frac{\bar{y}-\mu_0}{se}$, where $se = \frac{s}{\sqrt{n}}$.
- I.e. t_{obs} measures, how many standard deviations (with $\pm$ sign) the empirical mean lies away from μ_0.
- If H_0 is true, then t_{obs} is an observation from the t-distribution with $df = n - 1$.
- P-value: Values bigger than $|t_{obs}|$ or less than $-|t_{obs}|$ puts more support in H_a than H_0.
- The p-value = 2 x "upper tail probability of $|t_{obs}|$". The probability is calculated in the t-distribution with df degrees of freedom.

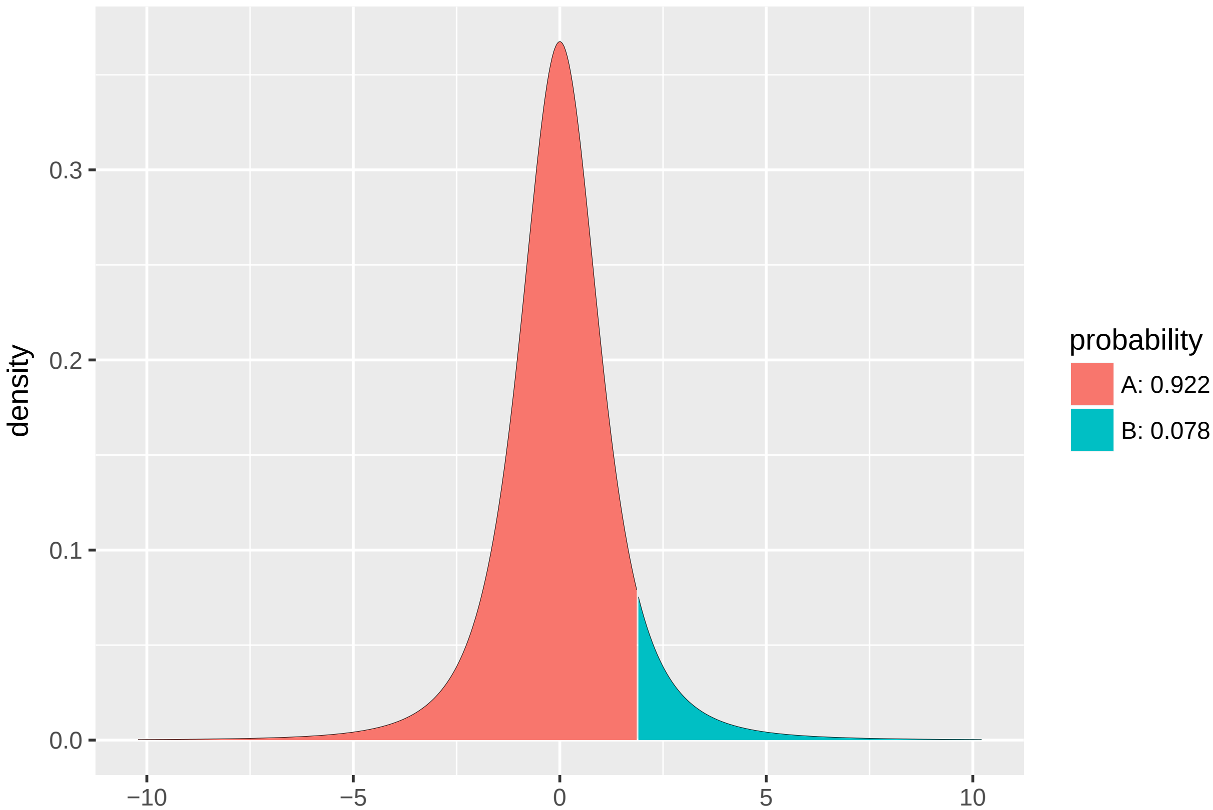
1.7.2 Example: Two-sided t-test

- Blood alcohol level measurements: $0.504, 0.500, 0.512, 0.524$.
- These are assumed to be a sample from a normal distribution.
- We calculate

  - $\bar{y} = 0.51$ and $s = 0.0106$
  - $se = \frac{s}{\sqrt{n}} = \frac{0.0106}{\sqrt{4}} = 0.0053$.
  - $H_0: \mu = 0.5$, i.e. $\mu_0 = 0.5$.
  - $t_{obs} = \frac{\bar{y}-\mu_0}{se} = \frac{0.51-0.5}{0.0053} = 1.89$.

- So we are almost 2 standard deviations from 0.5. Is this extreme in a t-distribution with 3 degrees of freedom?

```
library(mosaic)
1 - pdist("t", q = 1.89, df = 3)
```




```
## [1] 0.07757725
```

- The p-value is $2 \cdot 0.078$, i.e. more than 15%. On the basis of this we do not reject H_0.

1.8 One-sided t-test for mean

The book also discusses one-sided t-tests for the mean, but we will not use those in the course.

1.9 Agresti: Overview of t-test

TABLE 6.3: The Five Parts of Significance Tests for Population Means

1. **Assumptions**
   Quantitative variable
   Randomization
   Normal population (robust, especially for two-sided H_a, large n)
2. **Hypotheses**
   H_0: $\mu = \mu_0$
   H_a: $\mu \neq \mu_0$ (or H_a: $\mu > \mu_0$ or H_a: $\mu < \mu_0$)
3. **Test statistic**
   $t = \frac{\bar{y} - \mu_0}{se}$ where $se = \frac{s}{\sqrt{n}}$
4. ***P*-value**
   In t curve, use
   P = Two-tail probability for H_a: $\mu \neq \mu_0$
   P = Probability to right of observed t-value for H_a: $\mu > \mu_0$
   P = Probability to left of observed t-value for H_a: $\mu < \mu_0$
5. **Conclusion**
   Report P-value. Smaller P provides stronger evidence against H_0 and supporting H_a. Can reject H_0 if $P \leq \alpha$-level.

1.10 Significance test for proportion

- Consider a sample of size n, where we observe whether a given property is present or not.
- The relative frequency of the property in the population is π, which is estimated by $\hat{\pi}$.
- Null hypothesis: $H_0 : \pi = \pi_0$, where π_0 is a known number.
- **Two-sided alternative** hypothesis: $H_a : \pi \neq \pi_0$.
- *If H_0 is true* the standard error for $\hat{\pi}$ is given by $se_0 = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$.
- Observed test statistic: $z_{obs} = \frac{\hat{\pi}-\pi_0}{se_0}$
- I.e. z_{obs} measures, how many standard deviations (with $\pm$ sign) there is from $\hat{\pi}$ to π_0.

1.10.1 Approximate test

- If both $n\hat{\pi}$ and $n(1-\hat{\pi})$ are larger than 15 we know from previously that $\hat{\pi}$ follows a normal distribution (approximately), i.e.
  - If H_0 is true, then z_{obs} is an observation from the standard normal distribution.
- P-value for **two-sided** test: Values greater than $|z_{obs}|$ or less than $-|z_{obs}|$ point more towards H_a than H_0.

- The p-value=2 x "upper tail probability for $|z_{obs}|$". The probability is calculated in the standard normal distribution.

1.10.2 Example: Approximate test

- We consider a study from Florida Poll 2006:
  - In connection with problems financing public service a random sample of 1200 individuals were asked whether they preferred less service or tax increases.
  - 52% preferred tax increases. Is this enough to say that the proportion is significantly different from fifty-fifty?
- Sample with $n = 1200$ observations and estimated proportion $\hat{\pi} = 0.52$.
- Null hypothesis $H_0 : \pi = 0.5$.
- Alternative hypothesis $H_a : \pi \neq 0.5$.
- Standard error $se_0 = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{0.5\times0.5}{1200}} = 0.0144$
- Observed test statistic $z_{obs} = \frac{\hat{\pi}-\pi_0}{se_0} = \frac{0.52-0.5}{0.0144} = 1.39$
- "upper tail probability for 1.39" in the standard normal distribution is 0.0823, i.e. we have a p-value of $2\cdot$ 0.0823$\approx$ 16%.
- Conclusion: There is not sufficient evidence to reject H_0, i.e. we do not reject that the preference in the population is fifty-fifty.
- Note, the above calculations can also be performed automatically in **R** by (a little different results due to rounding errors in the manual calculation):

```
count <- 1200 * 0.52 # number of individuals preferring tax increase
prop.test(x = count, n = 1200, correct = F)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  count out of 1200
## X-squared = 1.92, df = 1, p-value = 0.1659
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.4917142 0.5481581
## sample estimates:
##    p
## 0.52
```

1.10.3 Binomial (exact) test

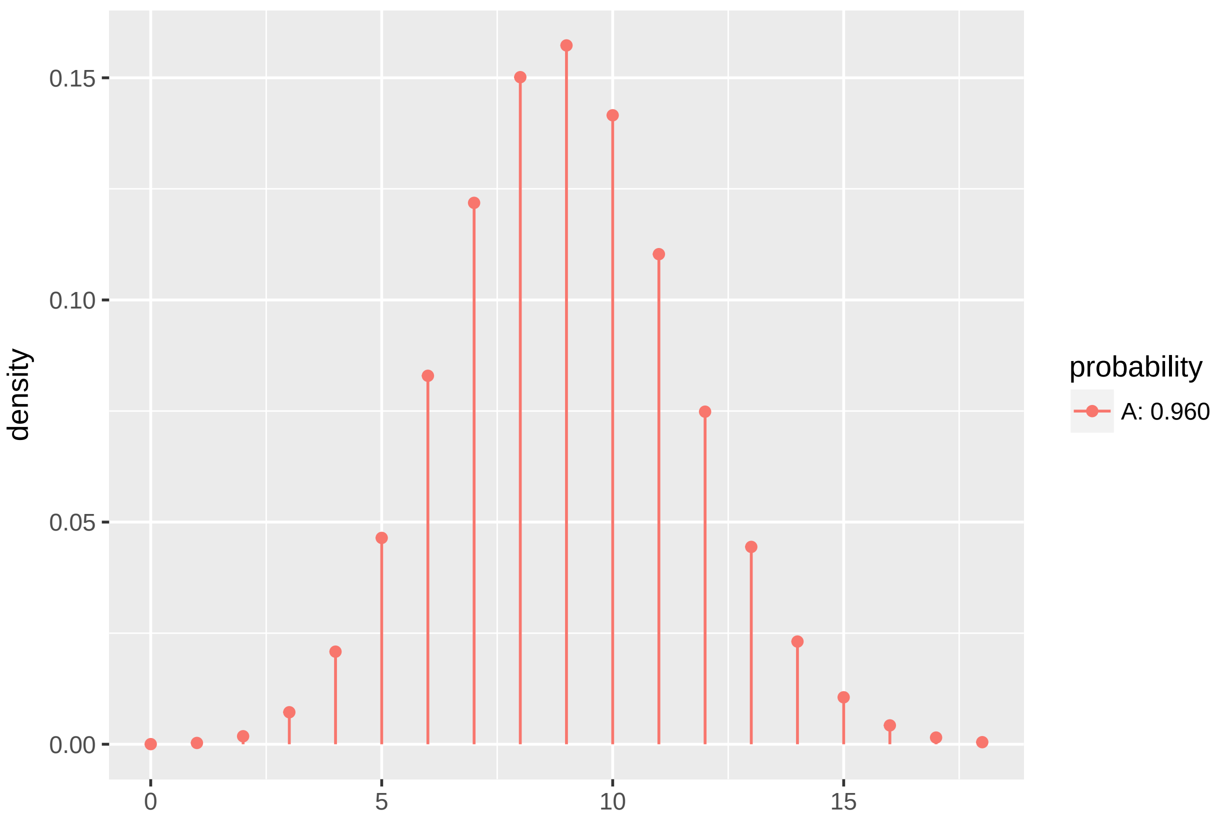
- Consider again a sample of size n, where we observe whether a given property is present or not.
- The relative frequency of the property in the population is π, which is estimated by $\hat{\pi}$.
- Let $y_+ = n\hat{\pi}$ be the frequency (total count) of the property in the sample.
- It can be shown that y_+ follows the **binomial distribution** with size parameter n and success probability π. We use $Bin(n, \pi)$ to denote this distribution.
- Null hypothesis: $H_0 : \pi = \pi_0$, where π_0 is a known number.
- Alternative hypothesis: $H_a : \pi \neq \pi_0$, where π_0 is a known number.

- P-value for **two-sided** binomial test:
  - If $y_+ \geq n\pi_0$: 2 x "upper tail probability for y_+" in the $Bin(n, \pi_0)$ distribution.
  - If $y_+ < n\pi_0$: 2 x "lower tail probability for y_+" in the $Bin(n, \pi_0)$ distribution.

1.10.4 Example: Binomial test

- Experiment with $n = 30$, where we have $y_+ = 14$ successes.
- We want to test $H_0 : \pi = 0.3$ vs. $H_a : \pi \neq 0.3$.
- Since $y_+ > n\pi_0 = 9$ we use the upper tail probability corresponding to the sum of the height of the red lines to the right of 14 in the graph below. (Notice, the graph continues on the right hand side to $n = 30$, but it has been cut off for illustrative purposes.)
- The upper tail probability from 14 and up (i.e. greater than 13) is:

```
lower_tail <- pdist("binom", q = 13, size = 30, prob = 0.3)
```




```
1 - lower_tail
```

```
## [1] 0.04005255
```

- The two-sided p-value is then 2 x 0.04 = 0.08.

1.10.5 Binomial test in R

- We return to the Chile data, where we again look at the variable `sex`.
- Let us test whether the proportion of females is different from 50 %, i.e., we look at $H_0: \ \pi = 0.5$ and $H_a: \ \pi \neq 0.5$, where π is the unknown population proportion of females.

```
Chile <- read.delim("https://asta.math.aau.dk/datasets?file=Chile.txt")
binom.test( ~ sex, data = Chile, p = 0.5, conf.level = 0.95)
```

```
## 
## 
## 
## data:  Chile$sex  [with success = F]
## number of successes = 1379, number of trials = 2700, p-value =
## 0.2727
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4916971 0.5297610
## sample estimates:
## probability of success
##              0.5107407
```

- The p-value for the binomial exact test is 27%, so there is no significant difference between the proportion of males and females.
- The approximate test has a p-value of 26%, which can be calculated by the command

```
prop.test( ~ sex, data = Chile, p = 0.5, conf.level = 0.95, correct = FALSE)
```

(note the additional argument `correct = FALSE`).

1.11 Agresti: Overview of tests for mean and proportion

TABLE 6.7: Summary of Significance Tests for Means and Proportions

Parameter	Mean	Proportion
1. Assumptions	Random sample, quantitative variable normal population	Random sample, categorical variable null expected counts at least 10
2. Hypotheses	H_0: $\mu = \mu_0$ H_a: $\mu \neq \mu_0$ H_a: $\mu > \mu_0$ H_a: $\mu < \mu_0$	H_0: $\pi = \pi_0$ H_a: $\pi \neq \pi_0$ H_a: $\pi > \pi_0$ H_a: $\pi < \pi_0$
3. Test statistic	$t = \frac{\bar{y} - \mu_0}{se}$ with $se = \frac{s}{\sqrt{n}}$, $df = n - 1$	$z = \frac{\hat{\pi} - \pi_0}{se_0}$ with $se_0 = \sqrt{\pi_0(1 - \pi_0)/n}$
4. P-value	Two-tail probability in sampling distribution for two-sided test (H_0: $\mu \neq \mu_0$ or H_a: $\pi \neq \pi_0$); One-tail probability for one-sided test	
5. Conclusion	Reject H_0 if P-value $\leq \alpha$-level such as 0.05	