Statistical Methods

Lecture 9 – Testing Statistical Hypotheses

Luke Dickens

Autumn 2020

Hypothesis tests and significance levels

Tests concerning mean of normal population with known variance

The t-test for mean of normal population with unknown variance

Hypothesis tests for population proportions

Statistical inference and hypotheses

Statistical inference:

- is the process of drawing conclusions about the characteristics of a population based on information from a sample
- may involve testing hypotheses relating to parameters that describe the population

These hypotheses usually state that a population parameter has a value within a given region. The statistician must decide whether each hypothesis is consistent with the information in a sample.

Definition: A **statistical hypothesis** is a statement about the nature of a population. It is often stated in terms of a population parameter (from [Ros17])

Null and alternate hypotheses

To construct a hypothesis test about a population parameter, we need a **null hypothesis**, H_0 and an associated **alternative hypothesis**, H_1. Typically:

- $\mathbf{H_0}$**:** makes a statement that the value of a population parameter lies in a given region.
- $\mathbf{H_1}$**:** states the converse of H_0, namely that the population parameter does not lie in that region.
- A hypothesis test attempts to refute H_0 and thus establish H_1.

H_0 is rejected if it appears inconsistent with the sample data, i.e. there is convincing evidence that H_0 is not true:

- If so then H_1 is accepted as true.
- If not, then H_0 is not rejected.

Example Null/alternative hypotheses

One example for a pair of hypotheses from [Ros17] are as follows:

- A new tobacco curing process claims to reduce mean nicotine per cigarette below 1.5 milligrammes
- A researcher is skeptical an wants to test this
- Given mean nicotine in new cigarettes is μ, she sets
    - $\mathbf{H_0}$: $\mu \leq 1.5$ mg
    - $\mathbf{H_1}$: $\mu > 1.5$ mg
- If sample is inconsistent with H_0 she can refute the claim from the tobacco manufacturer

Conversely, if the tobacco manufacturer wanted to establish their claim with a hypothesis test, they would need reversed conditions:

- $\mathbf{H_0}$: $\mu \geq 1.5$ mg
- $\mathbf{H_1}$: $\mu < 1.5$ mg

Test statistics and critical regions

Our test involves calculating a **test statistic** using the sample, then deciing whether this test statistic justifies rejection of H_0.

Definition: A **test statistic** is a statistic whose value is determined from the sample data. Depending on the value of this test statistic, the null hypothesis will be rejected or not [Ros17].

Rejection occurs if the test statistic lies inside a **critical region**, that represents inconsistency with H_0.

Definition: The **critical region** is the set of values of the test statistic for which the H_0 is rejected.

More formally, if TS is the RV of our test statistic and C is the critical region then:

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } TS \in C \\ \text{Do not reject } H_0 & \text{if } TS \notin C \end{array}$$

Interpreting a test result

A hypothesis test has two possible outcomes:

- **Rejecting H_0** – a strong statement that H_0 is not consistent with observed data.
- **Not rejecting H_0** – a weak statement that H_0 is consistent with observed data.

Not rejecting H_0 does not prove, confirm, establish or suggest H_0.

Hypothesis tests depend on RVs and two different types of errors can result:

- **type I error** the test rejects H_0 when H_0 is true
- **type II error** the test does not reject H_0 when H_0 is false

Significance levels

Remark: The objective of a hypothesis test is not to determine whether the null hypothesis is true. Rather it determines if H_0 is consistent with the data. Given this, H_0 should only be rejected if it would be very unlikely to result in the sample data.

[Ros17] describes what he calls *the classical procedure* as:

- specify a small value α, the **level of significance** of the test
- require that, under the test, whenever H_0 is true, its probability of being rejected is less than or equal to α
- commonly chosen values are $\alpha = 0.10, 0.05,$ and 0.01

It is normally good practice to set α ahead of time, prior to any data collection.

Example test statistic and critical region

Returning to the tobacco example from [Ros17]:

- Assume that the SD of the nicotine content is 0.8 mg
- The significance level α is set
- Determine a sample of size n
- Additional assumptions, e.g. normality of the sample mean
- Determines a critical region:*

$$C = \left\{ \bar{X} \in \mathbb{R} \;\middle|\; \bar{X} \geq 1.5 + \frac{1.315}{\sqrt{n}} \right\}$$

- Meaning that the null hypothesis is:

$$\begin{array}{ll} \text{Rejected} & \text{if } \bar{X} \geq 1.5 + \frac{1.315}{\sqrt{n}} \\ \text{Not rejected} & \text{otherwise} \end{array}$$

*The equation uses **set builder notation** to define the critical region. If you aren't familiar with this, please see **here** for a description.

Hypothesis tests and significance levels

Tests concerning mean of normal population with known variance

The t-test for mean of normal population with unknown variance

Hypothesis tests for population proportions

Test for point mean: hypotheses

Assume $X_1, \ldots, X_n$ is a sample of size n from a normal distribution having **unknown mean,** μ, and **known variance** σ^2. Our null hypothesis is that μ is equal to an exact (point) value. Our alternative hypothesis is that it is not.

We can therefore state these hypothesis as:

$$\mathbf{H_0} : \mu = \mu_0$$
$$\mathbf{H_1} : \mu \neq \mu_0$$

This set up is suitable to test whether it is reasonable to conclude that the mean μ is not equal to μ_0 (if we reject H_0).

Failing to reject H_0 **does not mean** we accept that $\mu = \mu_0$.

Test for point mean: critical region

The natural point estimator for μ is the sample mean:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

A natural condition under which to reject H_0 is if $\bar{X}$ is a long way from μ_0, i.e. for some $c > 0$ we define critical region:

$$C = \left\{ \bar{X} \in \mathbb{R} \;\middle|\; |\bar{X} - \mu_0| \geq c \right\}$$

Given significance level α, we must choose c such that:

$$\Pr\left(\bar{X} \in C \;\middle|\; \mu = \mu_0 \right) = \Pr\left(|\bar{X} - \mu_0| \geq c \;\middle|\; \mu = \mu_0 \right) = \alpha$$

In words: If the population mean equals μ_0 then the probability that $\bar{X}$ is at least c distant from μ_0 is α

Test for point mean: modelling H_0

Under H_0 the population mean $\mu = \mu_0$ and sample mean $\bar{X}$ is a normal RV with mean μ_0 and SD $\sigma/\sqrt{n}$ (from known population variance σ^2), so we can define standardised variable:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\sqrt{n}}{\sigma}\left(\bar{X} - \mu_0\right)$$

We can therefore rewrite our probability from the last slide:

$$\Pr\left(|\bar{X} - \mu_0| \geq c \;\middle|\; \mu = \mu_0\right) = \alpha$$
$$\implies \quad \Pr\left(|Z| \geq \sqrt{n}\frac{c}{\sigma}\right) = \alpha$$

Test for point mean: choosing c

From symmetry of the standard normal we have:

$$\Pr\left(Z \geq \sqrt{n}\frac{c}{\sigma}\right) = \frac{\alpha}{2}$$

And so $\sqrt{n}\frac{c}{\sigma} = z_{\alpha/2}$ or equally $c = \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$

That is we reject H_0 if

$$|\bar{X} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

Or equally, reject H_0 if

$$\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| \geq z_{\alpha/2}$$

“Correct” significance level

The most appropriate significance level depends on the application. A smaller α may be better if:

- rejection of H_0 is associated with a high cost/effort change
- initially H_0 is highly convincing or strongly motivated by prior knowledge/understanding

Instead, we may invert the *classical* process, and calculate the strictest significance level that would still lead to rejection.

The p-value

We can invert the "classical" process so far described by letting the data determine α, e.g. for our point mean test we instead:

- calculate the value of test statistic $\nu = \sqrt{n}\left(\bar{X} - \mu_0\right)/\sigma$
- set our critical value c to the calculated value ν
- calculate the p-value

$$\Pr\left(\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| \geq \nu \;\middle|\; \mu = \mu_0\right) = \text{p-value}$$

- p is the strictest significance level, α, that would reject H_0

A small p-value (e.g. $p < 0.05$) is a strong indicator that the null hypothesis is not true. The smaller the p-value, the greater the evidence for the falsity of H_0.

As before, we have sample $X_1, \ldots, X_n$ from a normal distribution having **unknown mean,** μ, and **known variance** σ^2. Our null hypothesis is that μ is less than or equal than μ_0. Our alternative hypothesis is that it is greater

We can therefore state these hypothesis as:

$$\mathbf{H_0} : \mu \leq \mu_0$$
$$\mathbf{H_1} : \mu > \mu_0$$

There is a similar test where the null hypothesis is that μ is greater than or equal to some value μ_0 (details in [Ros17, Sec. 9.3]).

One sided mean test: critical region

To construct the critical region we consider when the sample mean, $\bar{X}$ is a long way above the maximum value allowed by H_0, i.e.

$$\bar{X} - \mu_0 > c$$

To evaluate c given α, we must find when:

$$\Pr\left(\bar{X} - \mu_0 > c \;\middle|\; H_0\right) = \alpha$$

What value should we consider for μ under H_0?

Use $\mu = \mu_0$ as this leads to the most conservative estimate, giving

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } \sqrt{n}\frac{\bar{X}-\mu_0}{\sigma} \leq z_\alpha \\ \text{Do no reject } H_0 & \text{otherwise} \end{array}$$

Note: One sided, so we use z_α not $z_{\alpha/2}$.

One sided mean test: p-value

Similar to the point mean, we can alternatively compute the value of the test statistic:

$$\sqrt{n}\left(\bar{X} - \mu_0\right)/\sigma$$

Call this value ν then we evaluate:

$$\text{p-value} = \Pr\left(Z \geq \nu\right)$$

and then reject H_0 at any significance level greater than p.

Tests for mean with known variance overview

Below is a summary of two sided and one sided tests, including both significance level and p-value approaches for:

Unknown population mean μ, known variance σ, sample data $X_1, \ldots, X_n$ and $\bar{X} = \sum_i X_i/n$.

$\mathbf{H_0}$	$\mathbf{H_1}$	**TS**	**test at Significance level** α	**p value if TS** $= \nu$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}\frac{\bar{X}-\mu_0}{\sigma}$	Reject H_0 if $\lvert TS \rvert \geq z_{\alpha/2}$ Do not reject otherwise	$2\Pr(Z \geq \lvert\nu\rvert)$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}\frac{\bar{X}-\mu_0}{\sigma}$	Reject H_0 if $TS \geq z_\alpha$ Do not reject otherwise	$\Pr(Z \geq \nu)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}\frac{\bar{X}-\mu_0}{\sigma}$	Reject H_0 if $TS \leq z_\alpha$ Do not reject otherwise	$\Pr(Z \leq \nu)$

Hypothesis tests and significance levels

Tests concerning mean of normal population with known variance

The t-test for mean of normal population with unknown variance

Hypothesis tests for population proportions

Assume $X_1, \ldots, X_n$ is a sample of size n from a normal distribution having **unknown mean,** μ, and **unknown variance** σ^2. Our null hypothesis is that μ is equal to an exact (point) value. Our alternative hypothesis is that it is not.

We can therefore state these hypothesis as:

$$\mathbf{H_0} : \mu = \mu_0$$
$$\mathbf{H_1} : \mu \neq \mu_0$$

The only difference to what we have seen before is that the population variance σ^2 is no longer known.

With known variance, we defined a standard normal test statistic,

$$Z = \frac{\sqrt{n}}{\sigma}\left(\bar{X} - \mu_0\right)$$

and used this to derive rejection criterion at significance level α.

Here instead we must approximate σ with

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

Replacing σ with S for the test statistic gives:

$$T_{n-1} = \frac{\sqrt{n}}{S}\left(\bar{X} - \mu_0\right)$$

T_{n-1} has a t-distribution with $n-1$ degrees of freedom.

t-Test for point mean: rejection criterion

To justify rejection at significance level α we instead use:

$$\Pr\left(|T_{n-1}| > t_{n-1,\alpha/2}\right) = \alpha$$

From this, at a significance level of α with:

$$\mathbf{H_0} : \mu = \mu_0 \quad \text{and} \quad \mathbf{H_1} : \mu \neq \mu_0$$

Thus we should:

Reject H_0	if $\lvert T_{n-1}\rvert \geq t_{n-1,\alpha/2}$
Do no reject H_0	otherwise

Below is a summary of two sided and one sided t-tests, including both significance level and p-value approaches for:

Unknown population mean μ and variance σ, sample data $X_1, \ldots, X_n$, $\bar{X} = \sum_i X_i/n$ and $S^2 = \sum_i (X_i - \bar{X})^2/(n-1)$.

$\mathbf{H_0}$	$\mathbf{H_1}$	**TS**	**test at Significance level** α	**p value if TS** $= \nu$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}\frac{\bar{X}-\mu_0}{S}$	Reject H_0 if $\lvert TS \rvert \geq t_{n-1,\alpha/2}$ Do not reject otherwise	$2\,\Pr\left(T_{n-1} \geq \lvert\nu\rvert\right)$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}\frac{\bar{X}-\mu_0}{S}$	Reject H_0 if $TS \geq t_{n-1,\alpha}$ Do not reject otherwise	$\Pr\left(T_{n-1} \geq \nu\right)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}\frac{\bar{X}-\mu_0}{S}$	Reject H_0 if $TS \leq t_{n-1,\alpha}$ Do not reject otherwise	$\Pr\left(T_{n-1} \leq \nu\right)$

Hypothesis tests and significance levels

Tests concerning mean of normal population with known variance

The t-test for mean of normal population with unknown variance

Hypothesis tests for population proportions

One-sided proportion test: hypothesis

Consider a large population, where **unknown proportion** p have a characteristic of interest. From a sample of size n we determine X have the characteristic, and wish to test the null hypothesis:

$$\mathbf{H_0} : p \leq p_0$$

Against the alternative hypothesis

$$\mathbf{H_1} : p > p_0$$

We know:

- X is drawn from Binomial(n, p)
- we want to reject H_0 when X (equally $\hat{p}$) is sufficiently large.

One-sided proportion test: p-value

If the observed value of X is x, then the p-value of the sample equals the probability that:

- under H_0 we observe a value at least as large as x
- where the null hypothesis, H_0 is that $p \leq p_0$

The most conservative estimate of this is:

$$\Pr(B \geq x) = 1 - \Pr(B < x)$$

where B is RV with distribution Binomial(n, p_0)

All we need is some way to calculate the cumulative probability function for binomial distribution Binomial(n, p_0).

One-sided proportion test: final comments

Computing p:
For large enough n, we can consider the normal approximation for RV B with distribution Binomial(n, p_0). This requires the continuity correction, then conversion to a standard normal:

$$\Pr\left(B \geq x\right) \approx \Pr\left(Z \geq \frac{x - \frac{1}{2} - np_0}{\sqrt{np_0(1-p_0)}}\right)$$

Otherwise we'll need computational tools, e.g `scipy.stats.binom.cdf`.

Testing the converse, e.g. $\mathbf{H_0}$:$p \geq p_0$ and $\mathbf{H_1}$:$p < p_0$, is a straightforward modification of this.

Consider again large population, where **unknown proportion** p have a characteristic of interest. From a sample of size n we determine X have the characteristic, and wish to test the null hypothesis:

$$\mathbf{H_0} : p = p_0$$

Against the alternative hypothesis

$$\mathbf{H_1} : p \neq p_0$$

We know:

- X is drawn from Binomial(n, p)
- we want to reject H_0 when X (equally $\hat{p}$) is either sufficiently large **OR** sufficiently small

We want the total probability of rejection to be less than or equal to α under the null hypothesis.

The idea: Under H_0 we reject with probabilty $\alpha/2$ for too large values, and probability $\alpha/2$ for too small values. In total, we reject extreme values with probability α.

Dual conditions: And so, if we observe $X = x$ we reject if either:

$$\begin{aligned} & \Pr(B \le x) \le \frac{\alpha}{2} \\ \text{or} \qquad & \Pr(B \ge x) \le \frac{\alpha}{2} \end{aligned}$$

where B is RV with distribution Binomial(n, p_0)

Two-sided proportion test: test evaluation

Given our two rejection possibilities and that B is RV with distribution Binomial(n, p_0), the significance-level-α test can be stated as: Reject H_0 if

$$\min\Big(\Pr(B \leq x), \Pr(B \geq x)\Big) \leq \frac{\alpha}{2}$$

Or equivalently if

$$2\min\Big(\Pr(B \leq x), \Pr(B \geq x)\Big) \leq \alpha$$

And we calculate the p-value as

$$\text{p-value} = 2\min\Big(\Pr(B \leq x), \Pr(B \geq x)\Big)$$

Below is a summary of two sided and one sided tests concerining proportion p for:

The number of population elements in sample of size n with specified characteristic X, B is RV with distribution Binomial(n, p_0).

$\mathbf{H_0}$	$\mathbf{H_1}$	**TS**	**p value if TS** $= \nu$
$p \leq p_0$	$p > p_0$	X	$\Pr(B \geq \nu)$
$p \geq p_0$	$p < p_0$	X	$\Pr(B \leq \nu)$
$p = p_0$	$p \neq p_0$	X	$2\min(\Pr(B \leq \nu), \Pr(B \geq \nu))$

References I

[Ros17] Sheldon M. Ross, *Introductory Statistics*, 4 ed., Academic Press, 2017.