ST 516: Foundations of Data Analytics

Hypothesis Tests

Hypothesis Tests: Overview

The goal of a hypothesis test is to use sample observations to answer a question about the plausibility of a certain value of the parameter of interest. To do this, we use the following approach:

1. Gather data
2. Determine how likely that data would be if our specified value of the parameter is the truth
   - If the observed data is reasonably likely when our specified value is true (that is, our data is *consistent* with the specified value), then we have no reason to discard that value.
   - If, however, the observed data is quite unlikely when our specified value of the parameter is the truth, then we are skeptical of that value and we decide to 'reject' it.

The formal setting and vocabulary of the hypothesis testing setting follow.

Hypothesis Test Setting

Hypothesis test setting:

- Population of Interest
- Variable of Interest
  - We will use X to denote a general variable of interest
- Parameter of Interest
  - We will use θ to denote a general parameter of interest
  - For instance, θ might be the population mean, or population median, or population variance, etc.
- Sample $X_1, \ldots, X_n$ from the population of interest.
- Goal: Use the sample observations to answer a question/make a decision about the plausibility of a certain value or range of values for the population parameter θ.

Hypothesis Test Setting: Example

- Population of interest: Dogs who visited the OSU Veterinary Clinic in 2014
- Variable of interest: $X =$ Body temperature
- Parameter of interest: Population mean μ (so the parameter of interest θ is μ in this example)
- Sample: Randomly select $n = 20$ dog visits in 2014
    - Simple random sample: all dog visits are equally likely to be selected.
    - Obtained sample temperatures: $X_1, X_2, \ldots, X_{20}$.
- Goal: Determine whether dogs visiting the OSU Veterinary clinic have normal body temperature (101°F) on average, or whether they tend to have temperatures different from normal (not equal to 101°F) on average.

Anatomy of a Hypothesis Test

A **hypothesis test procedure** (or, more commonly and briefly, just 'hypothesis test') consists of the following components:

- Null hypothesis
- Alternative hypothesis
- Test statistic
- Rejection region

More details and examples of each of these components follow in the next few slides.

Anatomy of a Hypothesis Test

- **Null hypothesis** $H_0 : \theta = \theta_0$ or $H_0 : \theta \in \Theta_0$ (where Θ_0 is a set of values): A specified value or range of values for the parameter of interest (generally values that represent 'uninteresting' results).
  - Example: $H_0 : \mu = 101$ (our null hypothesis is that the population mean body temperature is 101, which is the commonly reported 'normal' temperature for dogs, so in this case our null hypothesis is the single value $\mu_0 = 101$)
- **Alternative hypothesis** $H_A : \theta = \theta_A$ or $H_A : \theta \in \Theta_A$: A different specified value or range of values for the parameter of interest (generally values that represent 'interesting' results).
  - Example: $H_A : \mu \neq 101$ (our alternative hypothesis is that the population mean body temperature is not equal to 101, so in this case our alternative is the range of values $(-\infty, 101) \cup (101, \infty)$–every value except 101.)

Anatomy of a Hypothesis Test

- **Test statistic**: A statistic (function of sample observed values) that is used to decide between the null hypothesis and alternative hypothesis.

Anatomy of a Hypothesis Test

- **Test statistic**: A statistic (function of sample observed values) that is used to decide between the null hypothesis and alternative hypothesis.
  - Example: We could use the sample mean $\overline{X}$ as our test statistic.

    Recall that by the Central Limit Theorem, if the *population variance* is σ^2, then with a sample size $n = 20$, we have

$$\overline{X} \dot{\sim} \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$$

Anatomy of a Hypothesis Test

- **Test statistic**: A statistic (function of sample observed values) that is used to decide between the null hypothesis and alternative hypothesis.
  - Equivalently, we could use the *standardized* form of the sample mean *assuming the null hypothesis is true*. This is called the z-statistic:

$$Z = \frac{\overline{X} - \mu_0}{\sqrt{\sigma^2/n}}$$

When the null hypothesis is true, $Z \dot{\sim} \text{Normal}(0,1)$.

Anatomy of a Hypothesis Test

- **Rejection region**: A set of values for the test statistic that would lead us to reject the null hypothesis in favor of the alternative hypothesis.

Anatomy of a Hypothesis Test

- **Rejection region**: A set of values for the test statistic that would lead us to reject the null hypothesis in favor of the alternative hypothesis.
  - Example: As stated on the previous slide,

$$\overline{X} \stackrel{\cdot}{\sim} \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$$

Therefore, when the null hypothesis is true we expect the sample mean to be relatively close to the null hypothesis mean value $\mu_0 = 101$. If the sample mean $\overline{X}$ is much *larger* than 101 or much *smaller* than 101, we will therefore reject the null hypothesis in favor of $H_A: \mu \neq 101$.

Anatomy of a Hypothesis Test

- **Rejection region**: A set of values for the test statistic that would lead us to reject the null hypothesis in favor of the alternative hypothesis.
    - Equivalently, we expect our z-statistic

$$Z = \frac{\overline{X} - \mu_0}{\sqrt{\sigma^2/n}}$$

      to have small *absolute value* (that is, be close to zero) if the null hypothesis is true, and larger absolute value if the alternative hypothesis is true.

Hypothesis Test Decisions

Possible outcomes (decisions) for a hypothesis test:

- **Reject** H_0: Decide that the observed data (the sample values) are not consistent with the null hypothesis, and therefore it is unlikely that the null hypothesis is true.
- **Fail to reject** H_0: Decide that the observed data could be consistent with the null hypothesis, and therefore we cannot rule it out.
- NOTE: It is technically not considered correct to talk about 'accepting' a null hypothesis, because the null hypothesis may not be the *only* plausible value or range of values for the parameter.

Anatomy of a Hypothesis Test

The question now is: How far from $\mu_0 = 101$ should $\overline{X}$ be to lead us to reject the null hypothesis? Equivalently, how large should be absolute value of Z be to lead us to reject the null hypothesis?

Suppose the sample variance is $s^2 = 4$. We will use s^2 as our *estimate* of σ^2, the population variance, in the following computations of the z-statistics.

Would you reject $H_0 : \mu = 101$ in favor of $H_A : \mu \neq 101$ if your sample of $n = 20$ dogs had a sample mean body temperature of:

- $\overline{X} = 99.6$, corresponding to $Z = \frac{99.6-101}{\sqrt{4/20}} = -3.1305$?
- $\overline{X} = 104$, corresponding to $Z = \frac{104-101}{\sqrt{4/20}} = 6.7082$?
- $\overline{X} = 101.6$, corresponding to $Z = \frac{101.6-101}{\sqrt{4/20}} = 1.3416$?

Anatomy of a Hypothesis Test

We could arbitrarily decide on a rejection region, for instance:

Reject H_0 if $\overline{X} < 100$ or $\overline{X} > 102$ (equivalently, $Z < -2.236$ or $Z > 2.236$).

But would this be a *good* rejection region? To answer this question, we need to think about how we would like our hypothesis test to perform: what are the possible outcomes of the hypothesis test, and what criteria should we evaluate?

Hypothesis Test Decisions

Possible outcomes and errors for a hypothesis test:

	Truth	
Decision	H_0 True	H_A True
Reject H_0	Type I Error	Correct Decision
Fail to Reject H_0	Correct Decision	Type II Error

Hypothesis Test Decisions

Possible outcomes and errors for a hypothesis test:

Decision	Truth	
	H_0 True	H_A True
Reject H_0	Type I Error	Correct Decision
Fail to Reject H_0	Correct Decision	Type II Error

We would prefer a hypothesis test procedure that has a low probability of making errors.

Hypothesis Test Error Probabilities

We design hypothesis test procedures to control the probability of making a Type I error at a desired level.

Definition

The **significance level** of a test procedure, often denoted by α, is the probability of a Type I error (i.e. rejecting the null hypothesis when it is in fact true). The following are all different ways of writing this same definition.

$$\alpha = \mathrm{P}(\text{Type I Error}) = \mathrm{P}(\text{Reject } H_0 \text{ when } H_0 \text{ is true}) = \mathrm{P}_{H_0}(\text{Reject } H_0)$$

We typically design a test at significance levels like $\alpha = 0.05$ or $\alpha = 0.1$...but these aren't magic numbers. You can use any significance level that seems appropriate to the particular scientific setting.

Hypothesis Test Error Probabilities

- A level $\alpha = 0.05$ hypothesis test procedure has a 5% chance of rejecting the null hypothesis when the null hypothesis is in fact true.
- If the null hypothesis is rejected at level $\alpha = 0.05$, it would also be rejected at values $\alpha^* > 0.05$ (*at least for all reasonable test procedures*)
  - We reject the null hypothesis more often if we are willing to have a higher chance of Type I error.
  - If the observed data would lead us to reject the null hypothesis with a 5% chance of making a Type I error, we should be even more comfortable rejecting the null hypothesis if we are willing to have a 10% chance of making a Type I error.
- The significance level of a test procedure determines the *strength of evidence* required to reject the null hypothesis: a lower significance level means we need stronger evidence against the null to reject it.

Hypothesis Test Error Probabilities

What about the Type II Error probability?

We do not treat the two types of errors symmetrically:

- We design a test to *control* the Type I error probability (significance level)
- Among all tests that have the desired significance level, we want the one that has the lowest Type II error probability.

Note that the Type II error probability depends on what the true value of the parameter is. We will let θ_A denote a particular value of the parameter of interest.

In our example, the parameter of interest is the population mean ($\theta = \mu$, and so $\theta_A = \mu_A$ would be some particular value of $\mu \neq 101$: for example, $\theta_A = \mu_A = 102.3$.

Hypothesis Test Error Probabilities

Definition

The **power** of a test procedure is the probability of making a *correct* decision to reject the null hypothesis when a particular value θ_A of the alternative hypothesis is true, and is equal to 1 - the probability of a Type II error at that particular alternative value θ_A. The following are all different ways of writing this same definition.

$$\text{Power}(\theta_A) = \text{P}(\text{Reject } H_0 \text{ when } \theta = \theta_A) = \text{P}_{\theta_A}(\text{Reject } H_0)$$