Precept 4: Describing data + testing hypotheses

Alex Kindel

Princeton University

4 October 2018

Today's precept

- Deviations and errors and variances, oh my!
- Parametric hypothesis tests
- Random permutation tests

Learning goals

The things we've learned so far are powerful tools for describing relationships in data.

Today, we'll practice testing hypotheses: informally by plotting your data, and two ways formally: with parametric hypothesis tests, and with nonparametric random permutation tests.

But first, we'll quickly review some of the key probability concepts you'll need in order to use these tools.

Variance review

Variance, σ^2 is a measure of spread, dispersion, or deviation from the central tendency of a distribution.

The **standard deviation**, σ, is just the square root of the variance.

We use these measures as a standard way of talking about how spread out a distribution is.

Population and sample

The **population** variance is the variance of a population. In statistical inference, this quantity is often *unknown* and must be estimated from data.

The **sample** variance is the variance of a sample. We often use the sample variance to *estimate* the population variance from data.

We sometimes use S to indicate the sample equivalent of σ.

Sampling distributions

We use sampling distributions to talk about the properties of estimators over repeated samples.

The **standard error** is the standard deviation of the sampling distribution. We use the standard error to quantify our uncertainty about the possible values of the estimand.

For example, if we're using the sample variance to estimate the population variance, we would compute the standard error of the sampling distribution of the sample variance to quantify our uncertainty about our estimate.

Summary

Population

- Population variance σ^2
- Population standard deviation σ

Sample

- Sample variance S^2
- Sample standard deviation S

Sampling distribution of an estimator

- Standard error $SE[\hat{\theta}]$

Hypothesis testing in 7 steps

- What is the null hypothesis (H_0)?
- What is the alternative hypothesis (H_a)?
- What is the significance level (α)?
- What is the test statistic ($t(X)$)?
- What is the null distribution?
- What is are the critical values?
- What is the rejection region?

A running example

Pokémon is one of the best-selling video game franchises of all time.

It's changed quite a bit since the 1990s: there are now *801* of them across 7 "generations".

Each Pokémon has one or two types, as well as six key stats:

- *HP*
- *Attack*
- *Defense*
- *Special Attack*
- *Special Defense*
- *Speed*

We'll use these data[1] to test some hypotheses.

[1] Data acquired from pokémondb.

Are Electric types faster than Ground types?

Figure 1: Pikachu – quick and agile!

Are Electric types faster than Ground types?

Figure 2: Hippowdon – not going anywhere fast.

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)?
- What is the alternative hypothesis (H_a)?
- What is the significance level (α)?
- What is the test statistic ($t(X)$)?
- What is the null distribution?
- What is are the critical values?
- What is the rejection region?

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)? $\mu_d - \mu_o = 0$

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)? $\mu_d - \mu_o = 0$
- What is the alternative hypothesis (H_a)? $\mu_d - \mu_o \neq 0$

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)? $\mu_d - \mu_o = 0$
- What is the alternative hypothesis (H_a)? $\mu_d - \mu_o \neq 0$
- What is the significance level (α)? $\alpha = 0.05$

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)? $\mu_d - \mu_o = 0$
- What is the alternative hypothesis (H_a)? $\mu_d - \mu_o \neq 0$
- What is the significance level (α)? $\alpha = 0.05$
- What is the test statistic ($Z = t(X)$)? $Z = \frac{\bar{X}_d - \bar{X}_o}{\sqrt{s_d^2/n_d + s_o^2/n_o}}$

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)? $\mu_d - \mu_o = 0$
- What is the alternative hypothesis (H_a)? $\mu_d - \mu_o \neq 0$
- What is the significance level (α)? $\alpha = 0.05$
- What is the test statistic ($Z = t(X)$)? $Z = \frac{\bar{X}_d - \bar{X}_o}{\sqrt{s_d^2/n_d + s_o^2/n_o}}$
- What is the null distribution? $Z \sim \mathcal{N}(0, 1)$

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)? $\mu_d - \mu_o = 0$
- What is the alternative hypothesis (H_a)? $\mu_d - \mu_o \neq 0$
- What is the significance level (α)? $\alpha = 0.05$
- What is the test statistic ($Z = t(X)$)? $Z = \frac{\bar{X}_d - \bar{X}_o}{\sqrt{s_d^2/n_d + s_o^2/n_o}}$
- What is the null distribution? $Z \sim \mathcal{N}(0, 1)$
- What is are the critical values? -1.96 and 1.96

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)? $\mu_d - \mu_o = 0$
- What is the alternative hypothesis (H_a)? $\mu_d - \mu_o \neq 0$
- What is the significance level (α)? $\alpha = 0.05$
- What is the test statistic ($Z = t(X)$)? $Z = \frac{\bar{X}_d - \bar{X}_o}{\sqrt{s_d^2/n_d + s_o^2/n_o}}$
- What is the null distribution? $Z \sim \mathcal{N}(0, 1)$
- What is are the critical values? -1.96 and 1.96
- What is the rejection region? *We reject if* $Z > -1.96$ *or* $Z > 1.96$.

Are Electric types faster than Ground types?

- What is the null hypothesis (H_0)? $\mu_d - \mu_o = 0$
- What is the alternative hypothesis (H_a)? $\mu_d - \mu_o \neq 0$
- What is the significance level (α)? $\alpha = 0.05$
- What is the test statistic ($Z = t(X)$)? $Z = \frac{\bar{X}_d - \bar{X}_o}{\sqrt{s_d^2/n_d + s_o^2/n_o}}$
- What is the null distribution? $Z \sim \mathcal{N}(0, 1)$
- What is are the critical values? -1.96 and 1.96
- What is the rejection region? *We reject if* $Z > -1.96$ *or* $Z > 1.96$.
- What is the p-value? $p < 0.05$

Hold on, *what*?

- What is the test statistic ($Z = t(X)$)?

$$Z = \frac{\bar{X}_d - \bar{X}_o}{\sqrt{\frac{s_d^2}{n_d} + \frac{s_o^2}{n_o}}}$$

This is sometimes called *Welch's t-test*. We often use it to test the difference in means between two groups with unequal sample sizes and variances.

Saving some effort with dplyr

```
# Function to extract stats for a given type
stat_bytype <- function(poke, type, stat) {
  poke %>%
    filter(type1 == type | type2 == type) %>%
    select(stat)
}
```

Computing the test statistic

```
# Are Electric Pokemon faster than Ground Pokemon?
elec.speed <- stat_bytype(poke, "electric", "speed")
grnd.speed <- stat_bytype(poke, "ground", "speed")

# Compute difference in means + test statistic
dm <- mean(elec.speed$speed) - mean(grnd.speed$speed)
t.st <- dm / sqrt((var(elec.speed$speed)/nrow(elec.speed)) + (var(grnd.spee
t.st
```

```
## [1] 4.82403
```

Do we reject the null hypothesis on the basis of this test?

How unlikely is it that we saw this by chance?

```r
# Compute p-value
pnorm(t.st, lower.tail=F)
```

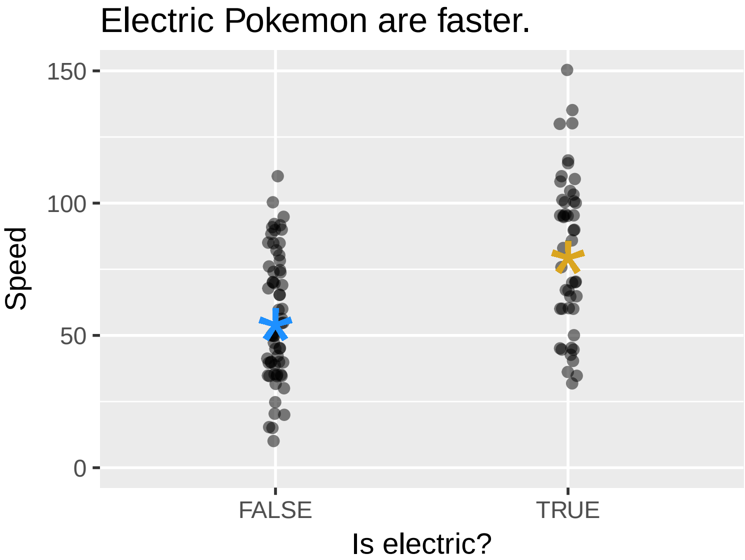
```
## [1] 7.034318e-07
```

t.test()

```r
t.test(elec.speed, grnd.speed)
```

```
##
##  Welch Two Sample t-test
##
## data:  elec.speed and grnd.speed
## t = 4.824, df = 87.101, p-value = 5.935e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  14.91619 35.82051
## sample estimates:
## mean of x mean of y
##  81.68085  56.31250
```

Another way of seeing it



A word of caution

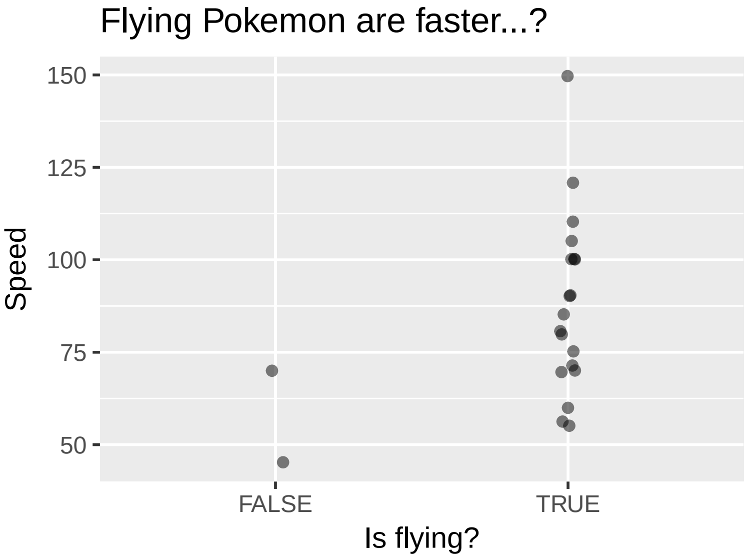
Are Flying types faster than Steel types in the first generation?

```
##
##  Welch Two Sample t-test
##
## data:  fly.speed and stl.speed
## t = 2.2231, df = 1.4184, p-value = 0.206
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -58.72798 119.41219
## sample estimates:
## mean of x mean of y
##  87.84211  57.50000
```

What would you conclude based on this result?

Do you see any problems with this comparison?

Always plot your data



Significance isn't everything

Small-sample inference lets us adjust our uncertainty estimate, so you can get the "right" standard errors

. . . but you'd be right to worry about drawing a conclusion based on 5 observations

Random permutation tests

Random permutation tests provide us a nice way of avoiding the pitfalls of parametric hypothesis testing.

The basic idea: if our data were *randomly assigned labels*, how likely is it that we would have observed this difference in means?

A random permutation test

First, we save the observed difference we want to test:

```
# Save observed difference
diff <- mean(elec.speed$speed) - mean(grnd.speed$speed)
```

One random permutation

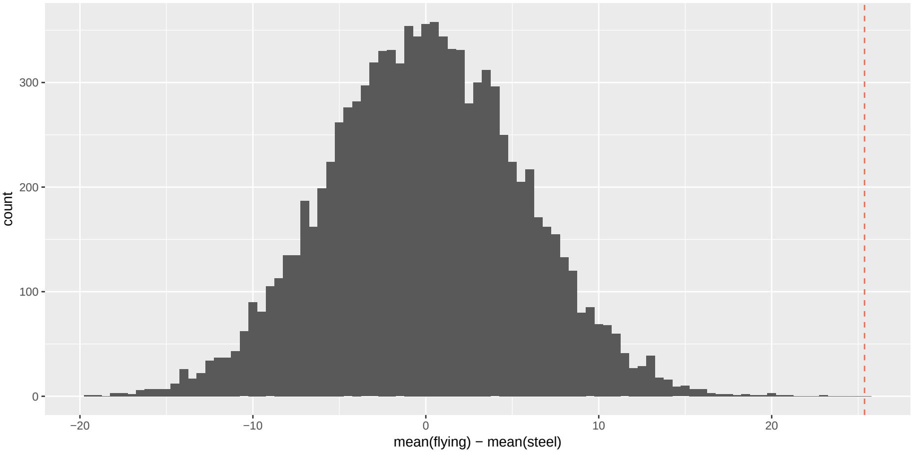
```r
# Compute one difference in means over random permutation
rp_test <- function(data, split.n) {
  shuf <- sample_n(data, nrow(data))  # Shuffle dataset

  # Compute difference in means for top vs. bottom group
  return(mean(shuf[1:split.n,]) - mean(shuf[split.n:nrow(data),]))
}
```

Then we do it lots of times!

```
# Perform 10k samples
nrep <- 10000
elec.grnd <- rbind.data.frame(elec.speed, grnd.speed)
samps <- replicate(nrep, rp_test(elec.grnd, nrow(elec.speed)))
```

How strong is our evidence?



More practice specifying hypothesis tests

In pairs: Come up with a null hypothesis and an alternative hypothesis.

Remember the recipe:

- What is the null hypothesis (H_0)?
- What is the alternative hypothesis (H_a)?
- What is the significance level (α)?
- What is the test statistic ($t(X)$)?
- What is the null distribution?
- What is are the critical values?
- What is the rejection region?

Alternatively, simulate lots of random permutations and look at how typical your observed value was.