The t-Distribution, t-Tests, and Simulation: Part A

(William Sealy Gosset published work on the t-distribution while working at the Guinness Brewery in 1908)

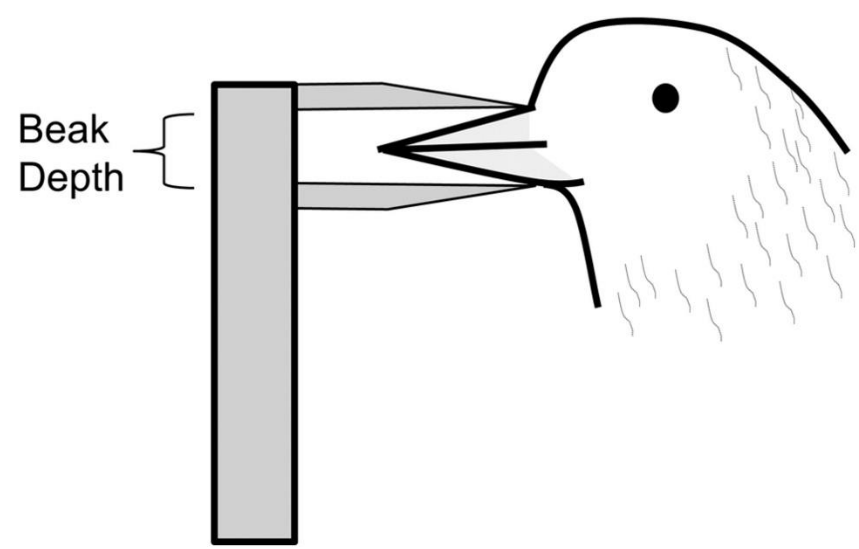
STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

Case Study: Darwin's Finches

- Darwin's Conjecture (mid 19th century): the different kinds of finches on Galapagos Islands have different kinds of beaks because they adapted to their respective environments
- Data (from 1980s): the beak depths of more than 20 generations of finches
- In 1977, there was

a draught, and only large, tough seeds were available

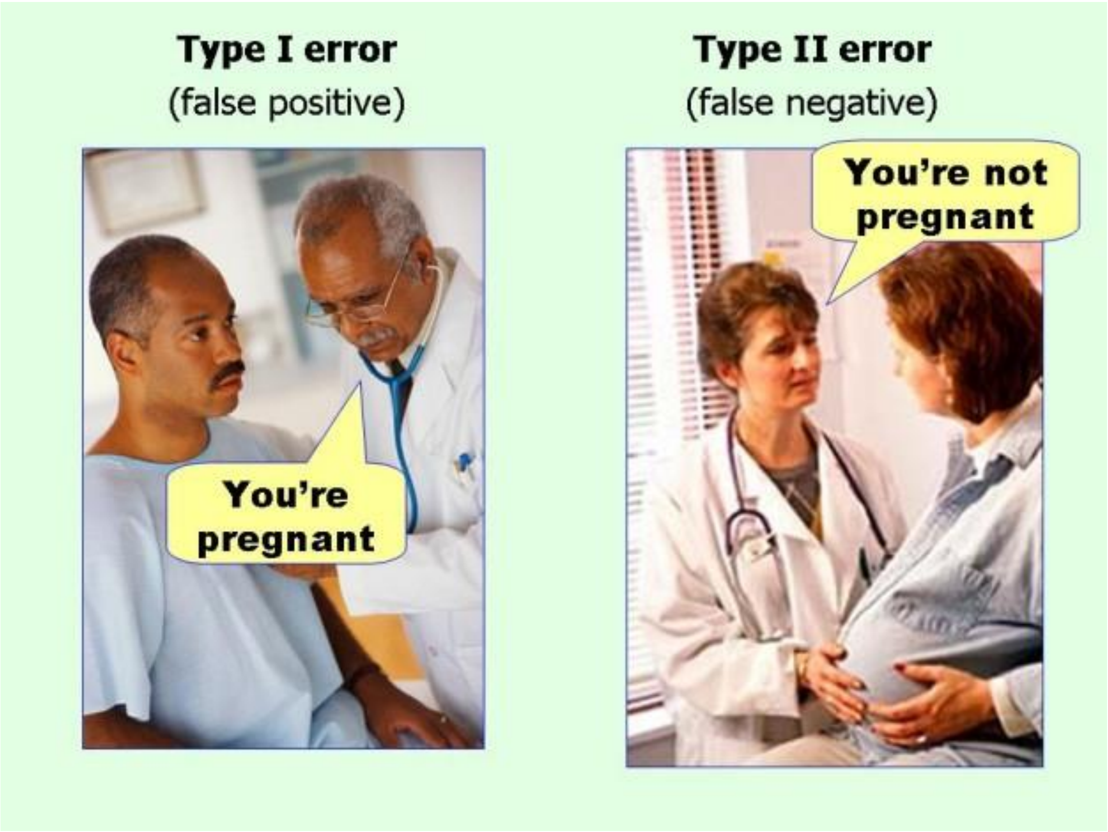


(Exploratory data analysis)

Review: The Null Hypothesis

- Null Hypothesis:
  - *Assume the beak depths are normally distributed, both pre-draught and post-draught*
  - The difference between the means of the distributions of beak depths pre-draught and post-draught is 0
- Ideally, the Null Hypothesis being true means that *nothing interesting is going on*. Ideally, rejecting the Null Hypothesis means that we learned something new. By default, we'd rather keep believing nothing interesting is going on and not believe something false
  - More or less the case here. It wouldn't be that surprising if the beak depth didn't have much to do with the toughness of the seeds

Type I errors and Type II errors



- Type I error: rejecting the Null Hypothesis even though it's true
  - (Mnemonic: this is worse, so it's type I)
- Type II error: not rejecting the Null Hypothesis even though it's false

Review: p-values

- Test statistic: some function of the observed sample
  - **The sample**: all the measurements made both pre-draught and post-draught
  - **Example of a test statistic:** the difference between the means pre-draught and post-draught
  - **A more useful example of a test statistic:** the difference between the means pre-draught and post-draught, divided by the estimate of a standard deviation
- P-value: a measure of how extreme the test statistic is, *assuming the null-hypothesis is true*

Review: p-value of a test statistic

- Assuming the Null Hypothesis is true, the probability that the test statistic will be as extreme, or more extreme, than the observed value of the test statistic, when data is repeatedly sampled from the model
  - Null Hypothesis: $X_1, X_2, X_3, \ldots, X_{89} \sim N(\mu_X, \sigma^2)$
    $Y_1, Y_2, Y_3, \ldots, Y_{89} \sim N(\mu_Y, \sigma^2)$
    $\mu_X = \mu_X$
  - Test statistic: $\mathrm{T} = \frac{\bar{X} - \bar{Y}}{SE(\bar{X} - \bar{Y})}$ ($SE(\bar{X} - \bar{Y})$ is the estimate of the SD)
  - We obtain the particular sample $x_1, \ldots, x_{89}, y_1 \ldots y_{89}$, and compute a particular t
  - (One-sided) p-value: $P(T > t)$, assuming the null hypothesis is true

Interpreting p-values

- $P(T > t) < 0.05$
  - (Small p-values in general)
  - We say "There is evidence against the Null Hypothesis"
  - If the Null Hypothesis is true, observing the value of the test statistic t that we actually observe would be unlikely
- $P(T > t) \geq 0.05$
  - (Large p-values in general)
  - We say "There is no (or weak) evidence against the Null Hypothesis"
  - If the Null Hypothesis is true, we would not be surprised to observe the value of t that we observed

ASA Statement on P-values

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

The Sampling Distribution of The Difference of Means

- Suppose $X_1, X_2, X_3, \ldots, X_n \sim N(\mu_X, \sigma^2)$
- We compute the means of the sample, $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$
- Then $\bar{X} \sim N(\mu_X, \frac{\sigma^2}{n})$ (and $\bar{Y} \sim N(\mu_Y, \frac{\sigma^2}{n})$ for the post-draught data)
- Suppose we happen to know that the beak depths are normally distributed, with an identical variance that we happen to know
- **Problem A:** What is the sampling distribution of $(\bar{X} - \bar{Y})$**?**
- **Problem B:** What relevant variable has the sampling distribution $N(0,1)$**?**

- $(\bar{X} - \bar{Y}) \sim N(\mu_X - \mu_Y, \frac{2\sigma^2}{n})$ (assuming X and Y are independent (Q: why wouldn't they be?))
- That is, if we repeatedly collect samples and measure $(\bar{X} - \bar{Y})$, the distribution will be $N(\mu_X - \mu_Y, \frac{2\sigma^2}{n})$
- $\frac{(\bar{X}-\bar{Y})}{\sqrt{2}\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

Simulation

- In R: simulate samples pre- and post-draught
- In R: the sampling distribution of the mean difference

Confidence Intervals

- **95% Confidence Interval:** the interval within which the true value of the parameter will lie 95% of the time if the Null Hypothesis is true
- Suppose $X \sim N(\mu, \sigma^2)$ (σ is known, μ is not known)
- **Problem A:** What is a 95% CI for μ**,** if we just sample the once?
- **Problem B:** What is a 95% CI for μ**,** if we sample N times?

Confidence Intervals

- $P\left(z_{.025} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{N}}} \leq z_{.975}\right) = 0.95$
- $P\left(-\bar{X} + z_{.025}\frac{\sigma}{\sqrt{N}} \leq \mu \leq -\bar{X} + z_{.975}\frac{\sigma}{\sqrt{N}}\right) = 0.95$
- $P\left(\bar{X} - z_{.975}\frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{X} - z_{.025}\frac{\sigma}{\sqrt{N}}\right) = 0.95$
- $P\left(\bar{X} - z_{.975}\frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{X} + z_{.975}\frac{\sigma}{\sqrt{N}}\right) = 0.95$

Confidence Intervals: Interpretation

- Why is it incorrect to say that the probability that the true value lies in the CI is 95%

Confidence Intervals: Interpretation

- *Glib Frequentist answer:*
  - μ is either in [1.5, 2.5] or it isn't, so the probability that $\mu \in [1.5, 2.5]$ is either 0 or 1
  - We just don't know which
- If we keep doing studies, and if our models are always correct, our parameters of interest will be in the Cis 95% of the time
- If you seriously want to assign a probability to the conjecture that $\mu \in [1.5, 2.5]$, you wouldn't automatically say it's 95% -- you would first look whether the data looks weird in some way, and adjust the probability downward if it does
  - That's the Bayesian approach

Student's t-Distribution

- $X_i \sim N(\mu, \sigma^2) \rightarrow \frac{\bar{X}}{\frac{\sigma}{\sqrt{N}}} \sim N(0,1)$
- But suppose σ^2 is not known.
- $X_i \sim N(\mu, \sigma^2) \rightarrow \frac{\bar{X}}{\frac{s}{\sqrt{N}}} \sim t_{N-1}, s^2 = \frac{\sum(X_i - \bar{X})^2}{N-1}$

(Student's t-Distribution with N-1 degrees of freedom)

- **Problem:** which is larger, $z_{.975}$ or $t_{.975}(df = 10)$?

Student's t-Distribution

- Assume that the pre-draught and post-draught populations are distributed with the same unknown SD
- s^2_{pooled} is the weighted average of s^2_X and s^2_Y.

$$s_{pooled} = \sqrt{\frac{(N_X - 1)s_X^2 + (N_Y - 1)s_Y^2}{(N_X + N_Y - 2)}}$$

Student's t-Distribution

- Suppose $Var(X_i) = Var(Y_j) = s_{pooled}$
- What's $SE(\bar{X} - \bar{Y})$?

Student's t-Distribution

- $Var(\bar{X} - \bar{Y}) = \frac{s_{pooled}^2}{N_x} + \frac{s_{pooled}^2}{N_Y} = s_{pooled}^2\,(\frac{1}{N_X} + \frac{1}{N_Y})$

- $SD(\bar{X} - \bar{Y}) = s_{pooled}\sqrt{(\frac{1}{N_X} + \frac{1}{N_Y})}$