T-tests and comparisons of groups

Posterior Predictive Distribution

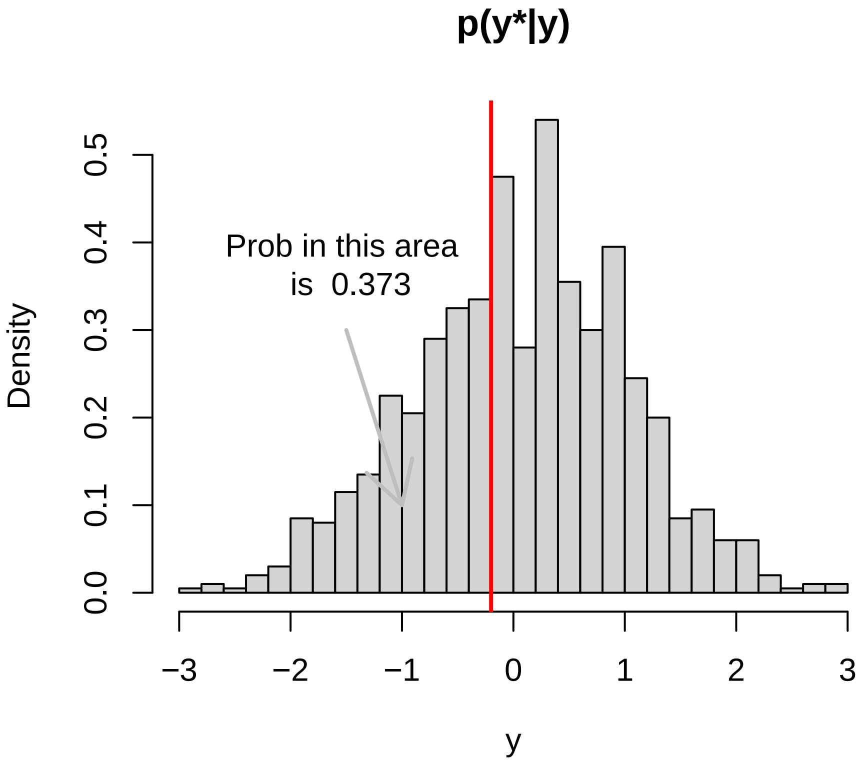
Another valuable tool in Bayesian statistics is the posterior predictive distribution which can be written as:

The posterior predictive distribution allows us to test whether our sampling model and observed data are reasonable. We will talk more about this later.

The posterior predictive distribution can also be used to make probabilistic statements about the next response, rather than the group mean. In our continuing example, we could calculate the probability of the next observed data point being greater than -0.2.

When $p(\theta|y)$ does not have a standard form,

```
posterior.mu <- codaSamples[[1]][,'mu']
posterior.sigma <- codaSamples[[1]][,'sigma']
posterior.pred <- rnorm(num.mcmc, mean = posterior.mu, sd = posterior.sigma)
prob.greater <- mean(posterior.pred > -0.2)
```

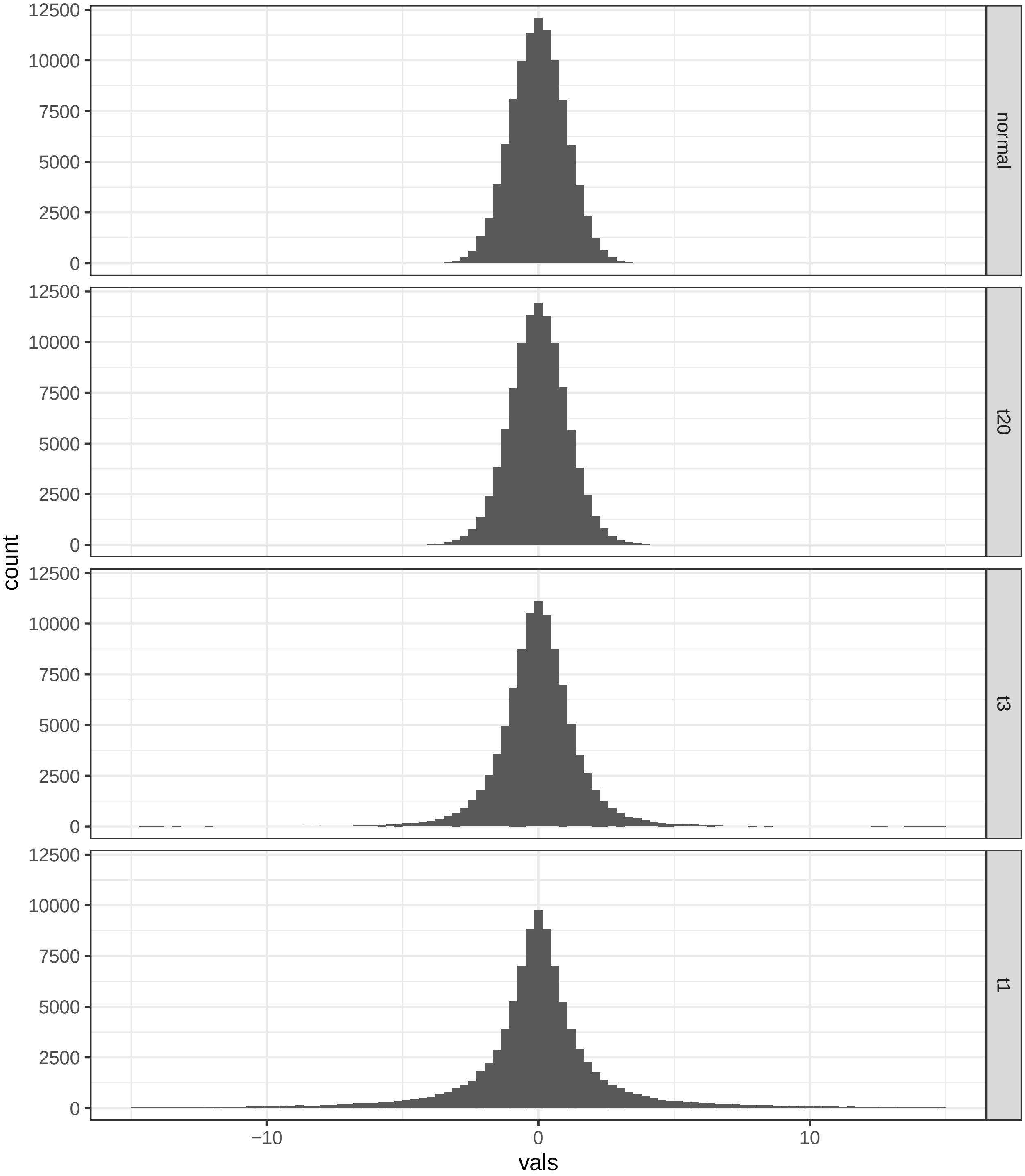


T - distribution

- While the normal distribution is often used for modeling continuous data, an alternative is the t-distribution.

- the wider tails can be illustrated thinking about the 2.5% quantile in terms of standard deviation for a specified degrees of freedom ν.
  - normal =
  - t(50) =
  - t(40) =
  - t(30) =
  - t(20) =
  - t(10) =
  - t(5) =
  - t(3) =
  - t(1) (Cauchy) =

- When the degrees of freedom gets large, the distribution approaches a normal distribution and when the degrees of freedom approach 1 the distribution becomes a Cauchy distribution



Estimation with Two Groups

A common use of the t-distribution is to make comparisons between two groups. For instance, we may be interested to determine if the mean height of two groups of OK Cupid users are different.

An aside on Null Hypothesis Significance Testing (NHST) (Ch. 11)

- What is the purpose of NHST? The goal of NHST is to decide whether a particular value of a parameter can be rejected.

- For instance, consider estimating whether a die has a fair probability of rolling a 6 ($\theta = 1/6$).
  - Then if we roll the die several times we'd expect 1/6 of the rolls to return a 6.
  - If the actual number is far greater or less than our expectation, we should reject the hypothesis that the die is fair.
  - To do this, we compute the exact probabilities of getting all outcomes. From this, we can compute the probability of getting an outcome, under the null hypothesis, as extreme or more than the observed outcome. This probability is known as a p-value.
  - The null hypothesis is commonly rejected if the p-value is less than 0.05.

- It is important to note that calculating the probability of all outcomes requires both the sampling and testing procedure.

- We are not going to get into the details, but section 11.1 in the texbook details a situation where a coin is flipped 24 times and results in 7 heads. The goal is determine if the coin is fair. Depending on the sampling procedure used, the p-value can range from .017 to .103 with this dataset.

Furthermore, The American Statistical Association also released the following 6 principles about p-values:

1. P-values can indicate how

2. P-values do not measure the probability that the studied hypothesis is true,

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,

6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

Bayesian Approach to Testing a Point Hypothesis

Consider the die rolling example. What value for (θ) would be says is meaningfully different than $\theta = 1/6 = 0.167$? If we are in a high-stakes gambling game, we might want θ to be accurate up to 0.001%, however, if we are using the dice in a friendly board game then accuracy of 2% might be sufficient.

- This range around the specified value is known as the .

- Given a ROPE and a posterior distribution, the parameter value is declared to be not credible, or rejected, if its entire ROPE lies outside of the 95% HDI of the posterior distribution of that parameter.

- A parameter value is declared to be accepted for practical purposes of that value's ROPE completely contains the 95% HDI of the posterior for that parameter.

- When the HDI and ROPE overlap, with the ROPE not completely containing the HDI, then neither of the above rules is satisfied and we withhold a decision.

- Note that the NHST regime provides no way to confirm a theory, rather just the ability to reject the null hypothesis.

Use the OK Cupid dataset and test the following claim, the mean height OK Cupid respondents reporting their body type as athletic is different than 70.5 inches (this value is arbitrary, but is approximately the mean height of all men in the sample). Interpret the results for each scenario.

```
okc <- read.csv('http://www.math.montana.edu/ahoegh/teaching/stat408/datasets/OKCupid_profiles_clean.cs

okc.athletic <- okc %>% filter(body_type == 'athletic')
```

1. Use `t.test()` with a two-sided procedure.

```
t.test(okc.athletic$height, mu = 70, alternative = 'two.sided')
```

```
##
##  One Sample t-test
##
## data:  okc.athletic$height
## t = -6.4862, df = 4710, p-value = 9.706e-11
## alternative hypothesis: true mean is not equal to 70
## 95 percent confidence interval:
##  69.56960 69.76939
## sample estimates:
## mean of x
##   69.6695
```

- Now consider whether there is a significant height difference between OK Cupid respondents self-reporting their body type as "athletic" and those self-reporting their body type as "fit"
- From the frequentist paradigm, this can be accomplished using the `t.test()` function.

```
okc.fit <- okc %>% filter(body_type == 'fit')
t.test(x= okc.athletic$height, y = okc.fit$height)
```

```
##
##  Welch Two Sample t-test
##
## data:  okc.athletic$height and okc.fit$height
## t = 15.55, df = 9702.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.9954687 1.2826521
## sample estimates:
## mean of x mean of y
##  69.66950  68.53044
```

- It is important to note there is no analog to ROPE with the t-test. If you have ever heard that statistical significance does not imply practical significance this is why.