Stat310

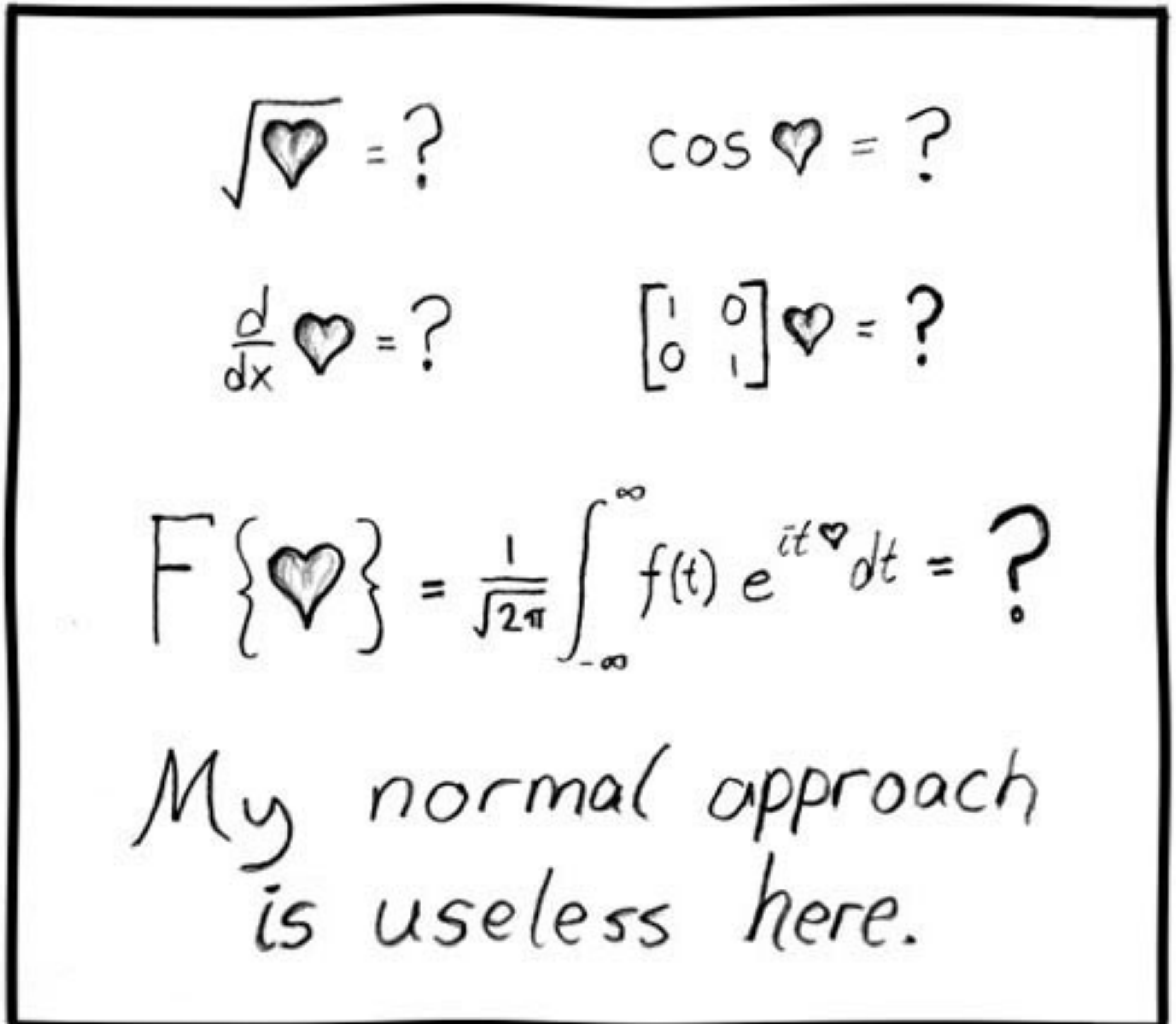
Simulation

Hadley Wickham

Test

- Don’t forget it’s on Tuesday
- All learning objectives, lectures & homework model answers up on site
- Help sessions Tuesday & Wednesday 4-5 pm

1. Valentine’s day & statistics
2. Normal distribution
3. Intro to R
4. Generating random numbers



http://xkcd.com/55/

You are perfect; I'd make no substitutions
You remind me of my favorite distributions
With a shape and a scale that I find reliable
You're as comforting as a two parameter Weibull
When I ask you a question and you answer truly
You speak as clearly as a draw from a Bernoulli
Your love of adventure is most influential
Just like the constant hazard of an exponential.
With so many moments, all full of fun,
You always integrate perfectly to one

Standard normal

For the normal distribution, we can convert any normally distributed rv to a **standard normal** by subtracting off the mean and dividing by the standard deviation.

This means we only need one table for any possible normal distribution.

Using the tables

Column + row = z

Find: $\Phi(2.94)$, $\Phi(-1)$, $\Phi(0.01)$, $\Phi(4)$

Can also use in reverse: For what value of z is $P(Z < z) = 0.90$? i.e. What is $\Phi^{-1}(0.90)$?

Find: $\Phi^{-1}(0.1)$, $\Phi^{-1}(0.5)$, $\Phi^{-1}(0.65)$, $\Phi^{-1}(1)$

Example

The time it takes me to drive to school is normally distributed with mean 20 and standard deviation 3.

What is the probability it takes me more than 20 minutes to drive to school?

What time should I leave so that I have 95% chance of getting to class by 1pm?

Example

What’s the probability I take a negative amount of time to get to school?

Can the the distribution of my driving time really be normal? Is that a problem?

$$P(Z < z) = \Phi(z)$$

$$\Phi(-z) = 1 - \Phi(z)$$

$$P(-1 < Z < 1) = 0.68$$
$$P(-2 < Z < 2) = 0.95$$
$$P(-3 < Z < 3) = 0.998$$

R

Intro to R

Open-source statistical programming environment

http://www.r-project.org

Installed on all lab computers

We're going to use it as a calculator and for some basic simulations. Extra info available on the website.

```
> 3 + 7
[1] 10

> sqrt(100) + 3 ^ 2
[1] 19

> 1:10
 [1]  1  2  3  4  5  6  7  8  9 10

> (1:10) ^ 2
 [1]   1   4   9  16  25  36  49  64  81 100
```

```
# From now on I'll just show the code
# <- means that this is a comment and will be
# ignored by R.

sin(2 * pi)
cos(pi)

exp(1:10)
log(1:10)
log(exp(1:10))

sqrt(2)^2 - 2
```

Your turn

Why does sqrt(2)^2 - 2 not equal 0?

```r
x <- c(1, 5, 3, 9, 2.4)

sum(x)
mean(x)
var(x)
sd(x)
```

Probabilities

Can use R instead of tables of probabilities. (This of course is what all modern statisticians do)

{p, d, q} x {binom, pois, unif, exp, gamma, norm}

d = pdf/pmf, p = cdf, q = inverse cdf

```r
# X ~ Binom(n = 100, p = 0.35)
# P(X = 0)
dbinom(0, size = 100, prob = 0.35)
# P(X = 35)
dbinom(35, size = 100, prob = 0.35)

# P(X < 10)
pbinom(10, size = 100, prob = 0.35)
# P(X > 10)
1 - pbinom(10, size = 100, prob = 0.35)
# P(X < 80)
pbinom(80, size = 100, prob = 0.35)
# P(30 <= X < 40)
pbinom(40, size = 100, prob = 0.35) -
  pbinom(30, size = 100, prob = 0.35)
```

```
# Probability of getting an odd number of heads
# when flipping a fair coin 100 times
sum(dbinom(1 + 2 * (0:49), size = 10, prob = 0.5))

# To figure it out, pull out the code from the
# outside in
0:49
1 + 2 * (0:49)
dbinom(1 + 2 * (0:49), size = 10, prob = 0.5)
sum(dbinom(1 + 2 * (0:49), size = 10, prob = 0.5))
```

Suffix	Distribution	Parameters
`binom`	Binomial	size, prob
`pois`	Poisson	lambda
`unif`	Uniform	min, max
`exp`	Exponential	rate
`gamma`	Gamma	shape, rate
`norm`	Normal	mean, sd

Your turn

$X \sim \text{Normal}(20, 9)$

What R code would you type to calculate $P(17 < X < 23)$?

Random number generation

Uniform to any rv

IF

$Y \sim \text{Uniform}(0, 1)$

F a cdf

THEN

$X = F^{-1}(Y)$ is a rv with cdf $F(x)$

(Assume F strictly increasing for simplicity)

Source of uniform numbers?

http://www.youtube.com/watch?v=7n8LNxGbZbs

Uses

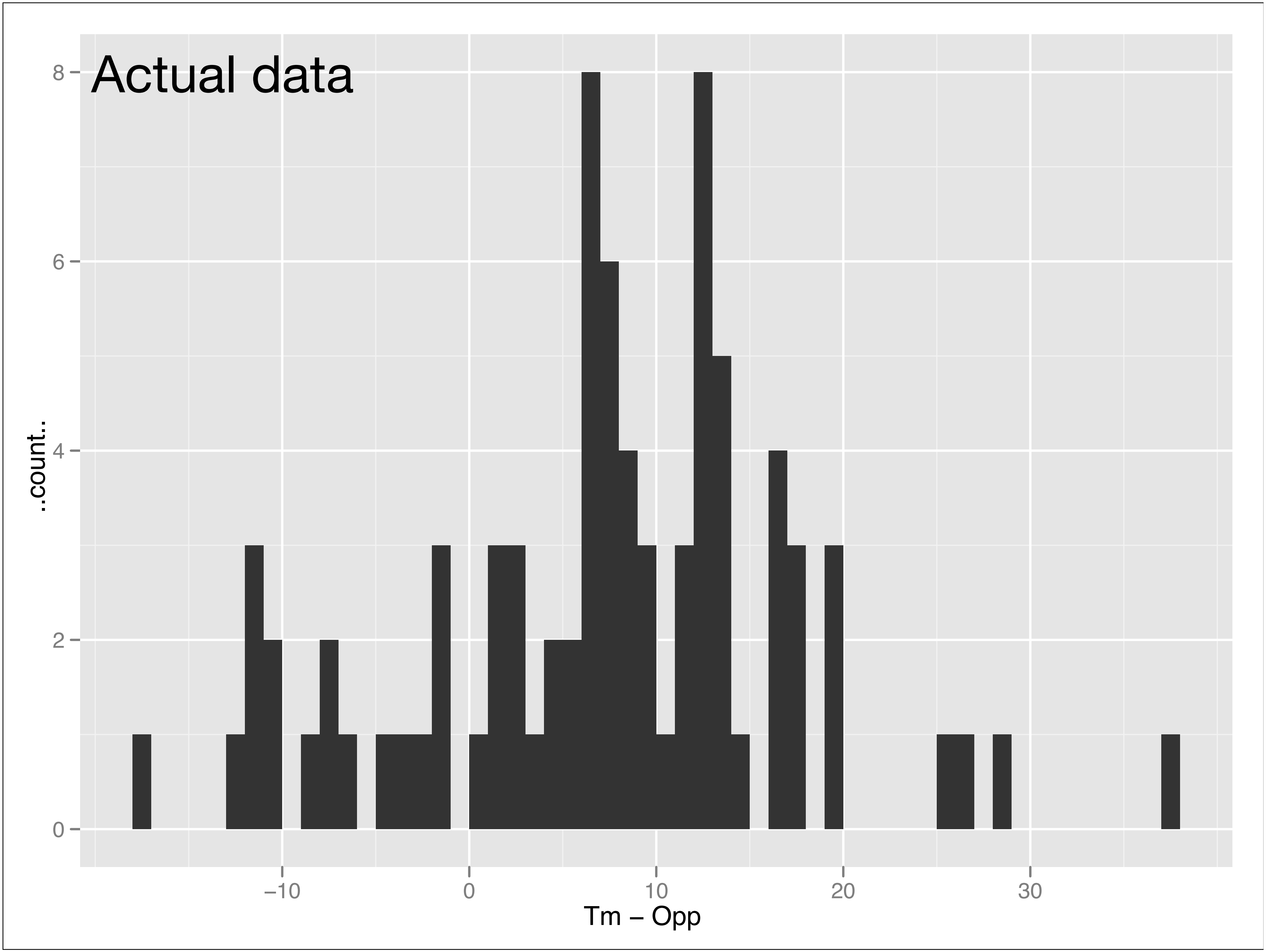
(In statistics)

Generating samples given a distribution so we can see if the data matches up

Approximating transformations numerically

```r
# same format as all other distribution related
# functions, but uses r

runif(100)
rpois(10, lambda = 10)
rbinom(100, n = 10, p = 0.5)

# useful to display graphically:
library(ggplot2)
qplot(runif(100), binwidth = 0.1)
```

```r
# Example: Lakers offence and defence
# O ~ Poisson(109.9)
# D ~ Poisson(101.8)
# What's the distribution of score differences?

o <- rpois(108, lambda = 109.9)
d <- rpois(108, lambda = 101.8)

qplot(o - d, binwidth = 1)
# If those assumptions are true, that shows a
# possible distribution of scores differences for
# one season.
```


Actual data
8
6
4
2
0
..count..
−10
0
10
20
30
Tm − Opp

```r
# If we want to see what the theoretical
# distribution looks like, we should use big
# numbers:
o <- rpois(1e5, lambda = 109.9)
d <- rpois(1e5, lambda = 101.8)

qplot(o - d, binwidth = 1)
```

```
# This is a useful way of exploring transformations
# of distributions

x <- runif(1e5)
qplot(x ^ 2, binwidth = 0.05)
qplot(sqrt(x), binwidth = 0.05)
qplot(exp(x), binwidth = 0.05)
qplot(log(x), binwidth = 0.5)

x <- rnorm(1e5)
qplot(x ^ 2, binwidth = 0.5)
qplot(exp(x), binwidth = 0.5)
```