Research Methods II

MICHAEL BERNSTEIN
CS 376

Last time

- What is a statistical test?
- Chi-square
- t-test
- Paired t-test
- ANOVA

Today

- Posthoc tests
- Two-way ANOVA
- Repeated measures ANOVA

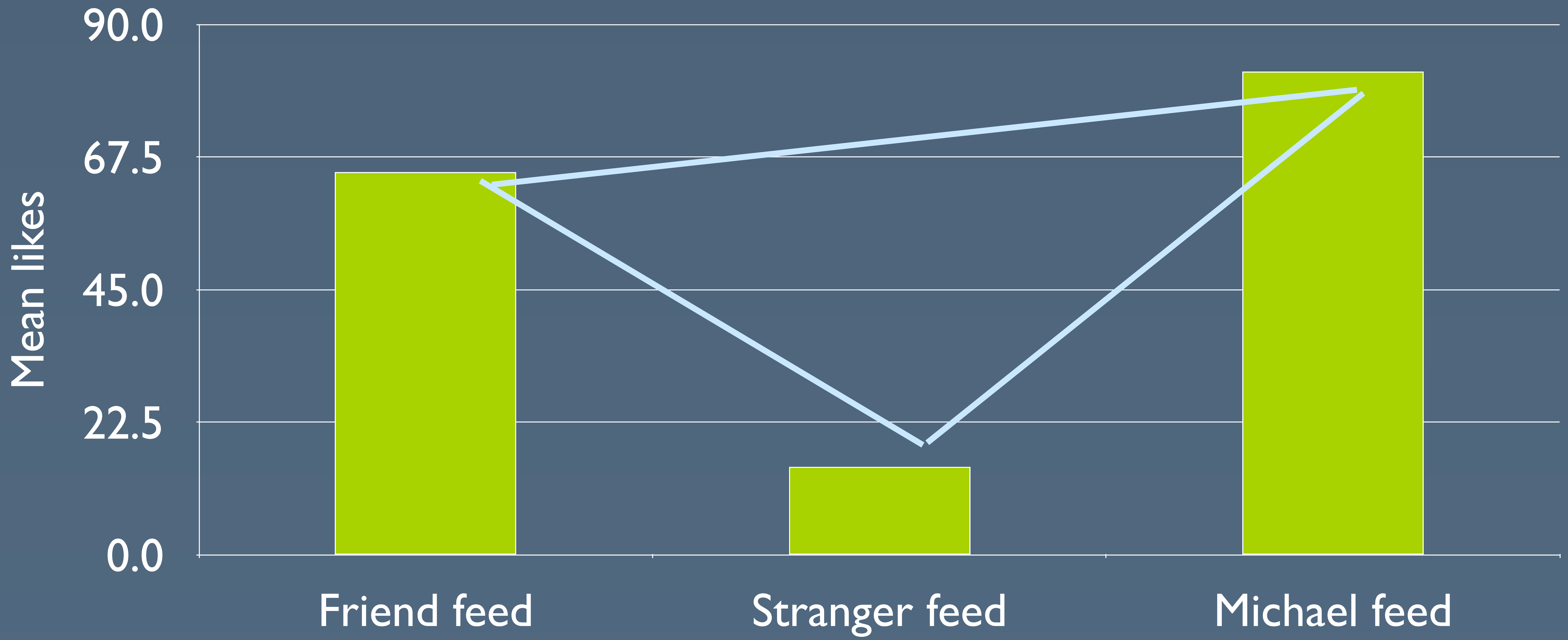
Posthoc tests

ANOVA! Are we done no

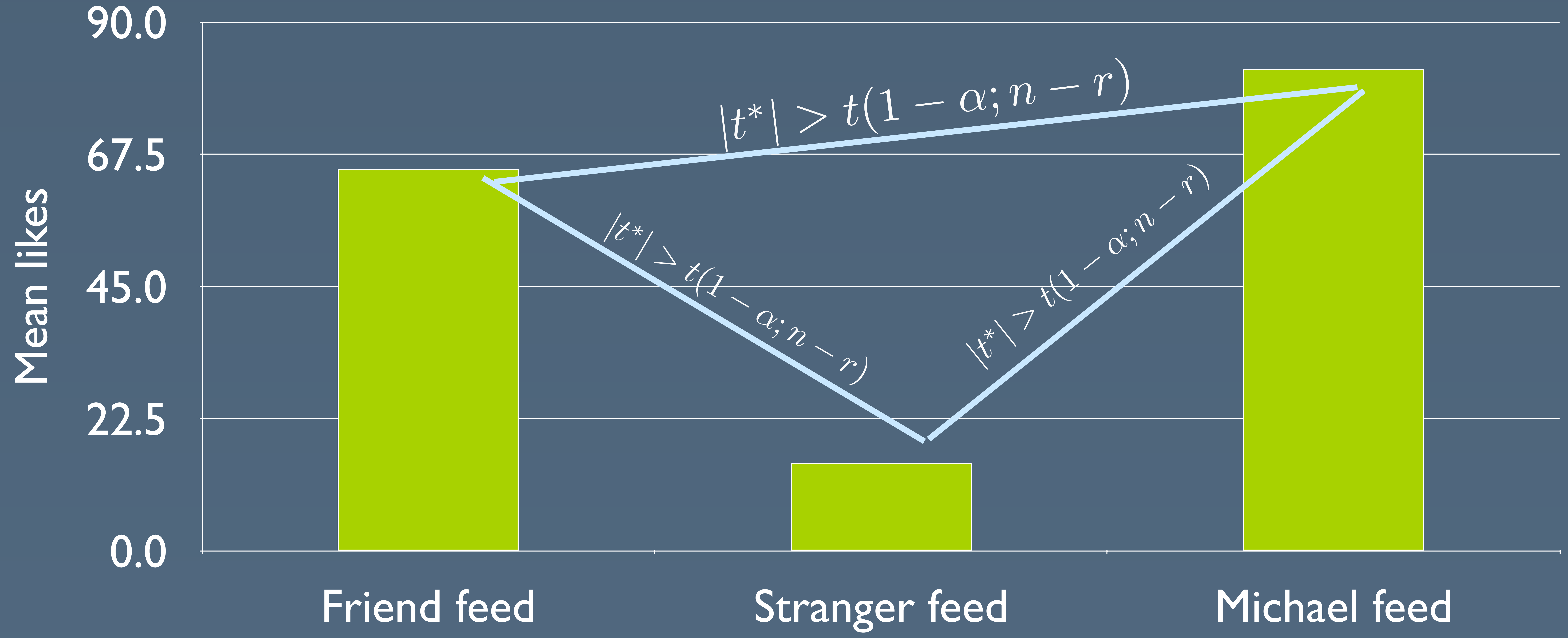
- Significant means "One of the μ_i are different."
- That's not very helpful: "There is some difference between populating the Facebook news feed with friends vs. strangers vs. only Michael's status updates"

Estimating pairwise differences

- Which pairs of factor levels are different from each other?



Roughly: we do pairwise t-tests



But…familywise error!

- $\alpha = .05$ implies a .95 probability of being correct
- If we do m tests, the actual probability of being correct is now:

$$\alpha^m = .95 \cdot .95 \cdot .95 \cdot \ldots$$

$$< .95$$

Bonferroni correction

- Avoid familywise error by adjusting α to be more conservative
- Divide α by the number of comparisons you make
  - 4 tests at $\alpha = .05$ implies using $\alpha = .0125$
- Conservative but accurate method of compensating for multiple tests

Bonferroni correction

```
> pairwise.t.test(value, group, p.adj='bonferroni')

   Pairwise comparisons using t tests with pooled SD

data:  value and group

  A       B
B 0.02971 -
C 0.00023 0.15530

P value adjustment method: bonferroni
```

Tukey test

- Less conservative than Bonferroni
- Compares all pairs of factor level means

```
> TukeyHSD(aov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = value ~ group, data = data)

$group
     diff        lwr      upr     p adj
B-A 1.375  0.1527988 2.597201 0.0257122
C-A 2.375  1.1527988 3.597201 0.0002167
C-B 1.000 -0.2222012 2.222201 0.1222307
```

Reporting

- "Posthoc tests using Bonferroni correction revealed that friend feed and Michael feed were significantly better than a stranger feed ($p<.05$), but the two were not significantly different from each other ($p=.32$)."

Two-way ANOVA

Crossed study designs

- Suppose you wanted to measure the impact of two factors on total likes on Facebook:
  - Strong ties vs. weak ties in your news feed
  - Presence of a reminder of the last time you liked each friend's content (e.g., "You last liked a story from John Hennessy in January")
- This is a 2 x 2 study: two factor levels for each factor {tie strength, reminder}

Interaction effects

- Sometimes the basic model doesn't capture subtle interactions between factors
  - Data: People who see strong ties and have a reminder are especially active
  - Result: Grand mean 8, strong tie mean 11, reminder mean 7, but mean in this cell is 20

Two-factor ANOVA test

- Test for main effects and interaction

```
> anova(lm(time ~ device * technique))
Analysis of Variance Table

Response: time
                 Df Sum Sq Mean Sq  F value    Pr(>F)    
device            1  981.0  981.02  94.5291 2.581e-12 ***
technique         2 3423.8 1711.90 164.9547 < 2.2e-16 ***
device:technique  2   75.3   37.65   3.6275   0.03522 *  
Residuals        42  435.9   10.38                       
```

factor or interaction SS MS F p

- Main effects are significant, but interaction effect is also significant

Significant interaction?

- Significant interactions mean that you can't just report the main effects — the story is more complicated
- Inspect to figure it out:

	Pen	Touch
Technique A	15.3	21.1
Technique B	23.9	33.1
Technique C	32.9	44.9

The slower techniques (B, C) harm Touch more than Pen

Repeated measures ANOVA

Within-subjects studies

- Control for individual variation using the mean response for each participant
- Before: we found the mean effect of each treatment
- Now: we find the mean effect of each participant

Repeated measures in R

repeated measures
error term

effect of subtracting
out the participant
means

remaining
main effects

```
> aov <- aov(value ~ factor(group) +
+ Error(factor(participant)/factor(group)), repeatframe)
> summary(aov)

Error: factor(participant)
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  7  5.167  0.7381

Error: factor(participant):factor(group)
              Df Sum Sq Mean Sq F value  Pr(>F)
factor(group)  2  22.75  11.375   10.92 0.00139 **
Residuals     14  14.58   1.042
```

All together now

Always follow every step!

1. Visualize the data
2. Compute descriptive statistics (e.g., mean)
3. Remove outliers >2 standard deviations from the mean
4. Check for heteroskedasticity and non-normal data
   - Try log, square root, or reciprocal transform
   - ANOVA is robust against non-normal data, but not against heteroskedasticity
5. Run statistical test
6. Run any posthoc tests if necessary