Wrapping up

Objectives

- revisit the original goals of the course and check in
- suggest some strategies for final exam
- broad overview of part III

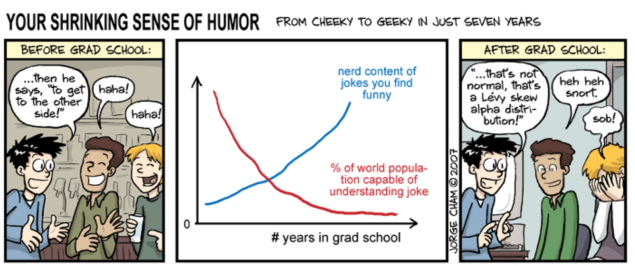
General wrap up

Goals for the semester

In addition the the learning objectives listed in your syllabus our overarching goals for the semester are to develop:

- your ability to critically assess statistical information presented to you in scientific and non-scientific fora
- your sense of how to approach answering real world questions with data
- your ability to concisely and accurately describe statistical methods and results

Goals for the semester



***footnote:** Thanks to Daragh at George Mason U. for this comic idea.*

Day 1 argument: This is a relevant class

I hoped to convince everyone here that statistics is relevant to everyone

You make many decisions during your day that are influenced by statistics

Statistics is not just relevant for public health, but also for other professions, including: policy, journalism and law

As we have tried to illustrate via the recurring "statistics is everywhere" segments, statistics is useful for understanding the news and the world around us - certainly during this pandemic we have seen a lot of public health and statistics in the news.

Non-statistical considerations

Statistics tells us about the role of chance - what would happen over many repeated samples

We want to think about the quality of the study design, what population was studied and whether the results are generalizable, sources of potential biase...

And remember that in our interpretation process we want to think about not just the statistical results.... - how meaningful is the effect - if the results suggest a relationship what are the risks and benefits of the exposure/treatment - what are the costs or implications of changing recommendations or guidelines?

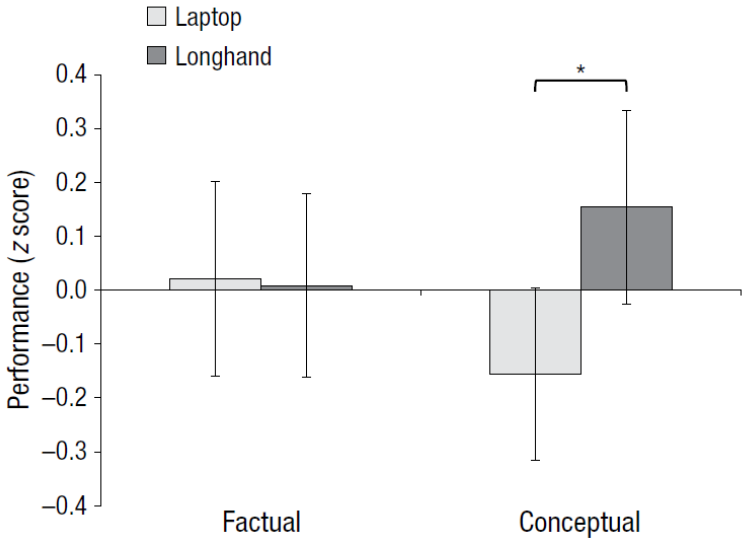
Evidence based suggestions: longhand notes

Pam A. Mueller and Dan Oppenheimer Psychological Science 2014, Vol. 25(6) 1159 -1168

The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking

Pam A. Mueller[1] and Daniel M. Oppenheimer[2]
[1]Princeton University and [2]University of California, Los Angeles

Evidence based suggestions: longhand notes



Evidence based suggestions: anxiety reduction

Virginia Clinton and Stacy Meester Teaching of Psychology 2019. vol 26(1)92-95

A Comparison of Two In-Class Anxiety Reduction Exercises Before a Final Exam

Virginia Clinton[1] and Stacy Meester[2]

Teaching of Psychology
2019, Vol. 46(1) 92-95

DOI: 10.1177/0098628318816182
journals.sagepub.com/home/top
SAGE

Evidence based suggestions: anxiety reduction

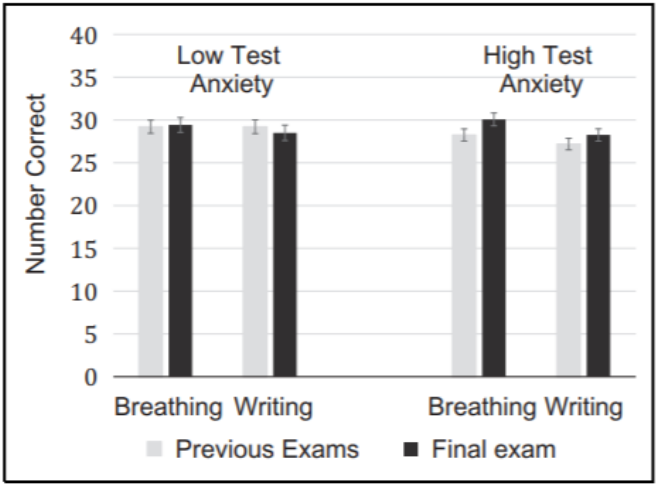


Figure 1. Previous exam and final exam performance by condition and level of trait test anxiety (means and ± 1 *SE*).

Requests from Ed

Would you be able to review the purpose and big picture application of the augment application in R?

```
##          x2       y2
## 1 1.000000 14.55482
## 2 1.152542 14.55541
## 3 1.305085 17.61083
## 4 1.457627 16.39905
## 5 1.610169 18.38384
## 6 1.762712 21.75376
```

run the model

```
lm2 <- lm(y2 ~ x2, data = dat2)
tidy(lm2)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)    -16.9     2.50      -6.74 7.98e- 9
## 2 x2              16.0     0.410     38.9  2.86e-43
```

and augment

```
augmented_dat2 <- augment(lm2)
head(augmented_dat2)
```

```
## # A tibble: 6 x 8
##      y2    x2 .fitted .resid   .hat .sigma .cooksd .std.resid
##   <dbl> <dbl>   <dbl>  <dbl>  <dbl>  <dbl>   <dbl>      <dbl>
## 1  14.6  1     -0.902  15.5  0.0650   8.20  0.126        1.91
## 2  14.6  1.15   1.53   13.0  0.0618   8.27  0.0846       1.60
## 3  17.6  1.31   3.97   13.6  0.0587   8.26  0.0876       1.68
## 4  16.4  1.46   6.40   10.0  0.0557   8.35  0.0443       1.23
## 5  18.4  1.61   8.84    9.55 0.0528   8.36  0.0381       1.17
## 6  21.8  1.76  11.3    10.5  0.0500   8.34  0.0432       1.28
```

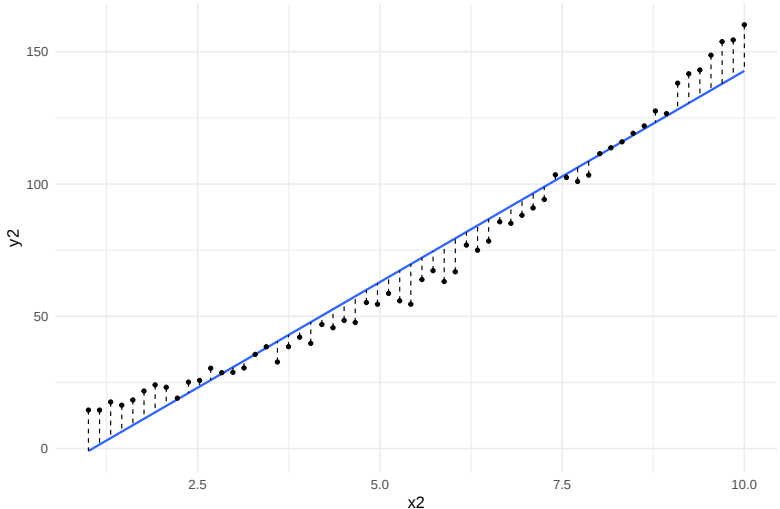
plot data and residuals

```
## Fitted model
plot1_2 <- ggplot(augmented_dat2, aes(x2, y2)) +
  geom_smooth(method = "lm", se = F) +
  geom_point() +
  geom_segment(aes(xend = x2, yend = .fitted), lty = 2) +
  theme_minimal(base_size = 15) +
  labs(title = "(a) Scatter plot")
```

plot data and residuals

```
## `geom_smooth()` using formula = 'y ~ x'
```

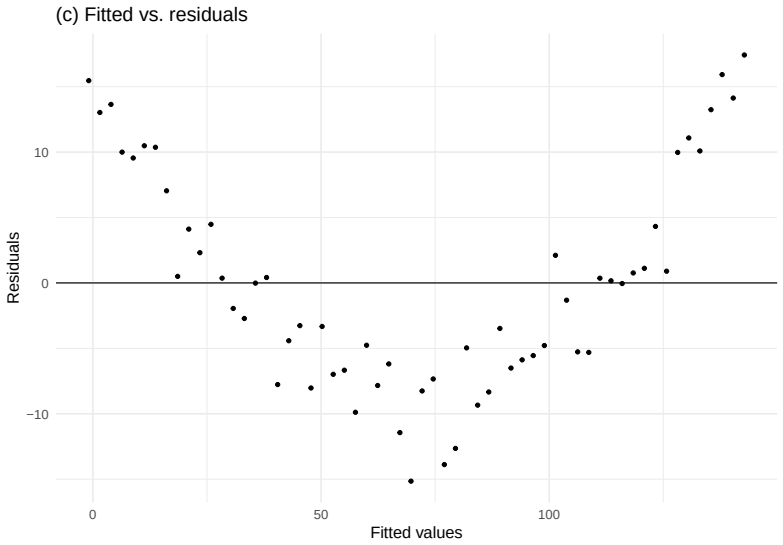
(a) Scatter plot



fitted vs residuals plotting


```
## Fitted vs. residuals
plot3_2 <-ggplot(augmented_dat2, aes(y = .resid, x = .fitted)) +
  geom_point() +
  theme_minimal(base_size = 15) +
  geom_hline(aes(yintercept = 0)) +
  labs(y = "Residuals", x = "Fitted values", title = "(c) Fitted vs. residual
```

fitted vs residuals plotting



Overview of Part III

part III: basic goal

Our overarching goal in part III is really to take our two "recipes" for statistical inference

1) Hypothesis testing
2) Confidence intervals

and figure out which ingredients to add in different situations.

choose the ingredients

We want to answer the questions:

1) What kind of an outcome variable are we working with?
2) How many groups/categories do we have data from that we want to compare?
3) Are the groups independent from each other or are observations inherently related/paired?
4) Do we meet the assumptions for a parametric test?

choose the ingredients

Based on the answers to the previous questions, we choose our "ingredients"

1) the effect/difference we want to examine
2) the variability we have in the data
3) a distribution we will be using to draw a critical value from based on:
   - our desired alpha
   - one or two tailed hypothesis

Decision tree: On the board

Recipe 1: Hypothesis testing

1) Check the assumptions for the theoretical distribution
2) Create a ratio of the effect/difference to the variability in the data
3) Generate the probability of observing that difference or greater if the null is true
4) Make a decision to reject or not reject the null hypothesis

Recipe 2: Confidence intervals

1) Generate your estimate
2) Calculate the critical value associated with your theoretical distribution
3) multiply the critical value by the variability
4) create your upper and lower bounds

For each test know:

- When to use it
- Any important assumptions that can be checked using data
- Appropriate visualization for the data
- What theoretical distribution are we using for inference
- How to construct the statistical test
  - what are the null and alternative hypotheses for the test
- How to construct the confidence interval
- relevant syntax in R
- any special notes/considerations

for example:

One sample T - used when we have 1 sample of a continuous outcome that we are comparing to a hypothesized value - assuming SRS, normality of the outcome, independence of outcomes - we might look at a histogram, density plot or qq plot - compared to a t distribution with n-1 df

for example:

One sample T - null: the mean is = hypothesized mean - alternative: could be one or two sided

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

for example:

One sample T relevant syntax:
pt() t.test(variable, alternative =“ “, mu=)

Notes: think about when the test is robust to violations of the assumptions and why

Some examples - what tests?

Today's fun fact

The length of your hand from your wrist to your elbow is the same as the length of your foot from your heel to your big toe. If I want to show that these two measures are almost perfectly correlated how might I do that?

You could do a correlation test, a linear regression, a paired t, you could show a scatterplot...

Today's fun fact

What kind of plot would I expect to see for these data?

Perfectly correlated scatterplot. Observations falling on a straight line with increasing slope.

Example 1: Staph infections

Researchers recruited 917 patients who had tested positive for staphylococcus Aureus and randomly assigned them to a staph-killing nasal ointment or placebo. They were interested in testing weather this drug was associated with a reduction in post-surgical infections. In the active treatment group 17 of 504 patients developed infections, in the placebo group 32 of 413 patients developed infections.

- What are the exposure and outcome variables?
- What kind of a test would you use for these data?
- What is the null hypothesis of this test?

Staph infections: ingredients

1) the effect/difference we want to examine
2) the variability we have in the data
3) a distribution we will be using to draw a critical value from based on:
   - our desired alpha
   - one or two tailed hypothesis

Staph infections ingredient 1: effect

Difference between two proportions:

$$\hat{p_1} - \hat{p_2} = \frac{17}{504} - \frac{32}{413} = .0337 - .0775 = -.0438$$

Staff infections ingredient 2: variability

- If the null hypothesis is true, then p_1 is truly equal to p_2. In this case, our best estimate of the underlying proportion that they are both equal to is

$$\hat{p} = \frac{\text{no. successes in both samples}}{\text{no. individuals in both samples}} = \frac{17 + 32}{504 + 413} = 0.0534$$

Staff infections ingredient 2: variability

- Our best guess at the SE for $\hat{p}$ is:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}$$

$$\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$$

This is the formula for the SE for the difference between two proportions but we have substituted $\hat{p}$ for p_1 and p_2.

$$\sqrt{0.0\hat{5}34 * (0.9466)(\frac{1}{504} + \frac{1}{413})} = 0.01492$$

Staph infections ingredient 3: distribution

Here for two sample testing, we have more than 10 successes and 10 failures in each group so we feel comfortable using a normal approximation to the binomial.

We will use a Z distribution, and an alpha of 0.05

The test is one sided because we are only interested in the left side of the distribution - decreases

Staph infections: test statistic

$$z = \frac{.0337 - .07748}{\sqrt{.0534 * 0.9466\left(\frac{1}{504} + \frac{1}{413}\right)}} = -2.936$$

Staph infections


```
## [1] 0.001662372
```

In this case we are only interested in a reduction in infections - so we will only look at the left tail of the distribution.

Staph infections

In R?

```
prop.test(x = c(17,32), # x is a vector of number of successes
          n = c(504,413), alternative="less" , correct=F) # n is a vector of
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  c(17, 32) out of c(504, 413)
## X-squared = 8.5906, df = 1, p-value = 0.00169
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.00000000 -0.01839005
## sample estimates:
##     prop 1     prop 2
## 0.03373016 0.07748184
```

Staph infections

In R?

```
binom.test(x=c(17,32),n=c(504,413), alternative="less")
```

```
##
##  Exact binomial test
##
## data:  c(17, 32)
## number of successes = 17, number of trials = 49, p-value = 0.02219
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
##  0.0000000 0.4738246
## sample estimates:
## probability of success
##              0.3469388
```

Staph infections: Confidence interval

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$-.0438 \pm 1.96 \sqrt{\frac{.0\hat{3}37(1-.0\hat{3}37)}{504} + \frac{.0\hat{7}75(1-.0775)}{413}} = -.0438 \pm 0.01542$$

95% CI is -0.0592 to - 0.04226

Example 2

August 8, 2019 Vitamin D Supplementation and Prevention of Type 2 Diabetes

BACKGROUND Observational studies support an association between a low blood 25-hydroxyvitamin D level and the risk of type 2 diabetes. However, whether vitamin D supplementation lowers the risk of diabetes is unknown.

Example 2 cont.

METHODS We randomly assigned adults who met at least two of three glycemic criteria for prediabetes (fasting plasma glucose level, 100 to 125 mg per deciliter; plasma glucose level 2 hours after a 75-g oral glucose load, 140 to 199 mg per deciliter; and glycated hemoglobin level, 5.7 to 6.4%) and no diagnostic criteria for diabetes to receive 4000 IU per day of vitamin D3 or placebo, regardless of the baseline serum 25-hydroxyvitamin D level.

What type of study is this?

What type of variable is the predictor (how many groups)?

Example 2 cont.

RESULTS By month 24, the mean serum 25-hydroxyvitamin D level in the vitamin D group was 54.3 ng per milliliter (from 27.7 ng per milliliter at baseline), as compared with 28.8 ng per milliliter in the placebo group (from 28.2 ng per milliliter at baseline). After a median follow-up of 2.5 years, the primary outcome of diabetes occurred in 293 participants in the vitamin D group and 323 in the placebo group (9.39 and 10.66 events per 100 person-years, respectively).

What kinds of tests would you use here for the vitamin D comparison? for the outcome?

Example 2

Other considerations: why might we not see the result we were expecting?

Per the discussion in the article:

"Because vitamin D supplements are used increasingly in the U.S. adult population,29 approximately 8 of 10 participants had a baseline serum 25-hydroxyvitamin D level that was considered to be sufficient according to current recommendations ($\geq$ 20 ng per milliliter) to reduce the risk of many outcomes,23,30 including diabetes.6 The high percentage of participants with adequate levels of vitamin D may have limited the ability of the trial to detect a significant effect.""

Example 3:

A study on the effects of vaping classifies people as “never vapers”, “occasional vapers”, “frequent vapers”. You interview a sample of 150 people in each group and ask a questionnaire to derive a quantitative score (between 0 and 100) on stress levels.

What kind of an outcome is this? What test is appropriate here?

Example 4:

The amygdala is a brain structure involved in the processing of memory of emotional reactions. Ten subjects were shown emotional video clips and non emotional video clips in random order. They then had their memory of the clips assessed. Recall accuracy was scored from 1 to 100.

What type of data do you have? What kind of a test is appropriate?

Example 5:

A random sample of 700 births from local records shows this distribution across the days of the week. Do these data give evidence that local births are not equally likely on all days of the week?

Day	Births
Monday	110
Tuesday	124
Wednesday	104
Thursday	94
Friday	112
Saturday	72
Sunday	84

What test would we use here?

What is the null hypothesis?

Example 5: expectation

Day	Births
Monday	100
Tuesday	100
Wednesday	100
Thursday	100
Friday	100
Saturday	100
Sunday	100

Example 6

You have heard that the grading scale is harsher at UC Berkeley than at other California universities. You want to test this rumor with data. You have data from a random sample of 100 transcripts from students at Berkeley who took PH142 and data on the letter grade distribution for undergraduate statistic courses in general from a California wide survey.

What kind of outcome? How many groups/samples?

What test would you consider here?

Example 6

You have heard that the grading scale is harsher at UC Berkeley than at other California universities. You want to test this rumor with data. You have data from a random sample of 100 transcripts from students at Berkeley who took PH142 and data on the letter grade distribution for undergraduate statistic courses in general from a California wide survey.

This is a catgorical outcome, with one sample, so we would use a chi-squred goodness of fit test

Example 6

You find a source that gives the distribution for UC intro biostat courses as: A=50%, B=30%, C=15%, Fail=5%

Grade	N	Expected?
A	224	?
B	99	?
C	17	?
F	10	?
Total	350	350

Example 6

Grade	N	Expected?
A	224	175
B	99	105
C	17	52.5
F	10	17.5
Total	350	350

statistic$= 13.72 + 0.34 + 24.00 + 3.214 = 41.28$

```
pchisq(41.28, df=3, lower.tail=FALSE)
```

```
## [1] 5.703407e-09
```

Common Issues or questions

Parametric vs non-parametric

What does it mean when I say non-parametric?

Which tests have we covered that are non-parametric?

Common Issues

method used vs P-value and CI vs. conclusion

Appropriate interpretation of a P-value and CI

Relationship between a confidence interval and a p-value