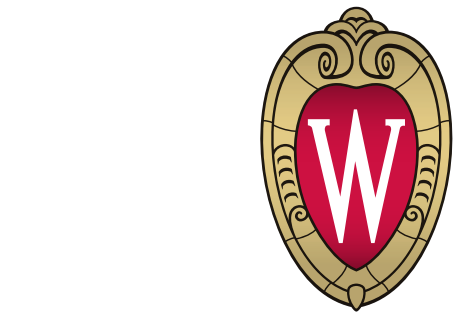
STAT324 Final Exam Review

https://dzwang91.github.io/stat324/



WISCONSIN

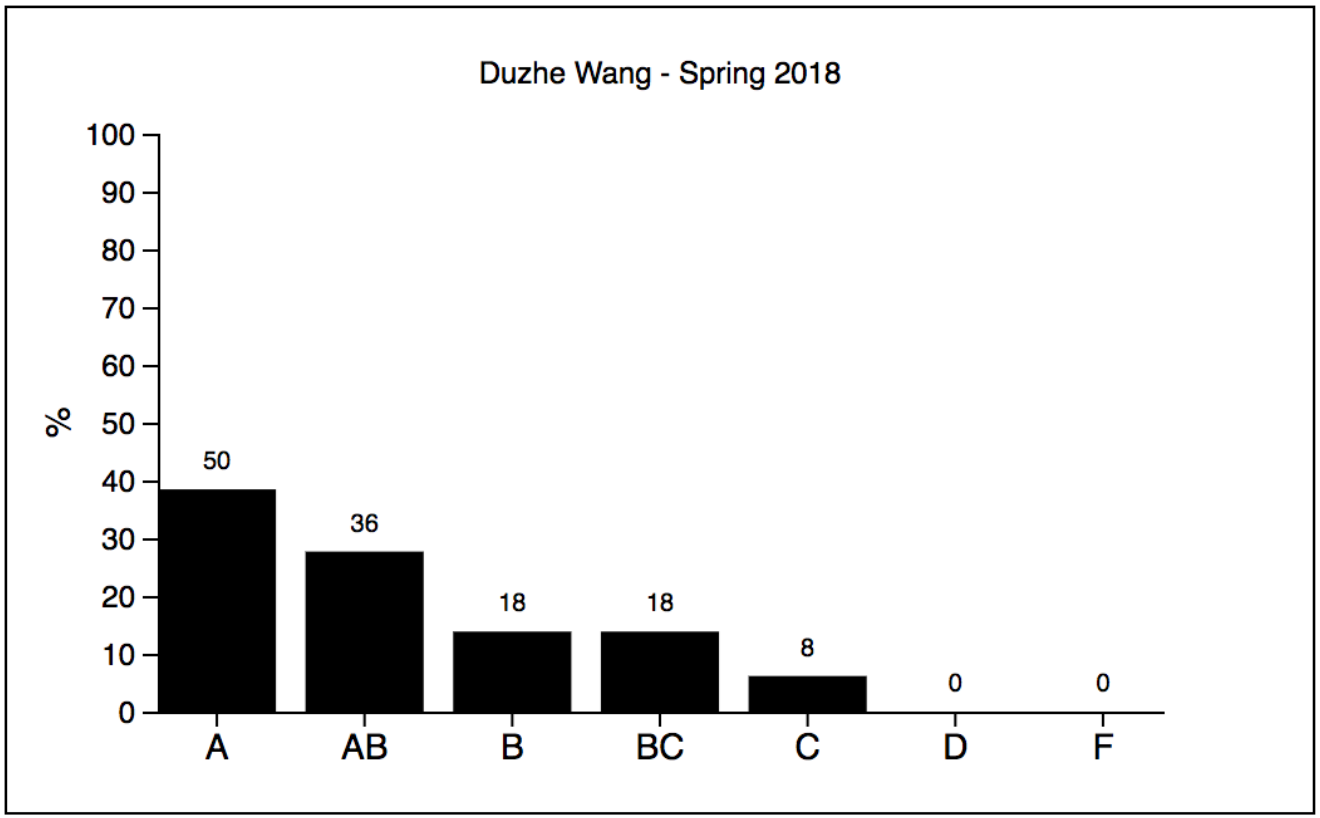
UNIVERSITY OF WISCONSIN–MADISON

Announcements

- -"Where do I take the final exam? Chamberlin 2241 or Noland 168?"
  -"**See the room assignment at Canvas.**"
- Final exam time: **7:25-9:25 pm on May 5**
- **Bring you Wiscard!**
- Bring three sheets of two-side note and a calculator. Z table, T table and F table will be provided.
- Double check your homework assignments, Midterm 1 and Midterm 2 grades at Canvas. If you find any mistake, contact your support TA ASAP. Don't wait until the last minute.
- I'll add bonus points for some previous pop-up exercises/questions in this semester to your final exam (I have an archive for that, so no need for you to remind me).

Announcements

- Previous grade distribution:



Avg GPA **3.39** | Count 130 (130)

Final exam problems

- The final exam is **cumulative**.
- 5 big problems, 120 points in total.
  - True or false problems (3 points $\times$ 12): only circle the correct answer. Include both of conceptual questions and calculation questions. No need to show your work.
  - Multiple choice problems (4 points $\times$ 13): only circle the best one answer. Include both of conceptual questions and calculation questions. No need to show your work.
  - Big calculation problems: from materials after Midterm 2 (L19-L24)
    - Two paired sample tests
    - ANOVA
    - Simple linear regression

Overview of STAT324

- Descriptive statistics
- Probability and random variables

} Fundamentals

- Point estimation
- Confidence interval
- Hypothesis testing

} Main topics of statistics

- Simple linear regression: the first model in statistics

Descriptive statistics

- Sample mean
- Sample standard deviation
- Median
- Q_1, Q_3
- Range
- IQR
- Histogram, QQ plot, Box plot...

Probability and random variables

- Basics
  - Random process, outcome, sample space, event
  - P(E)=sum of probabilities of outcomes in E
  - $0 \leq P(E) \leq 1$, P(not E)=1-P(E)
  - iid
  - A and B are independent if occurrence of one doesn't change the probability of the other, then P(A and B)=P(A)P(B).
- Mean and variance
  - Mean: $\mu = \mathbb{E}(X)$.
  - Mean properties:

$$\mathbb{E}(c) = c, \mathbb{E}(cX) = c\mathbb{E}(X),$$
$$\mathbb{E}(X + c) = \mathbb{E}(X) + c, \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

  - Variance: $\sigma^2 = \mathbb{E}[(X - \mu)^2]$.
  - Variance properties:

$$VAR(c) = 0, VAR(cX) = c^2 VAR(X), VAR(X + c) = VAR(X).$$

    For independent X and Y,

$$VAR(X + Y) = VAR(X) + VAR(Y).$$

Probability and random variables

- Discrete random variables
  - Bernoulli RV: $\mathbb{P}(Y = 1) = \pi, \mathbb{P}(Y = 0) = 1 - \pi$.
  - $\mu = \pi, \sigma^2 = \pi(1 - \pi)$.
  - Binomial RV: $X \sim Bin(n, \pi)$ is number of successes in n independent Bernoulli trials, each with P(success)=π.
  - $\mathbb{P}(X = x) = \frac{n!}{x!(n-x)!}\pi^x(1 - \pi)^{n-x}$ for $x = 0, ..., n$.
  - $\mu = n\pi$, $\sigma^2 = n\pi(1 - \pi)$.
- Continuous random variables
  - $\mathbb{P}(a \leq X \leq b)$: area under $f(x)$ between a and b, where $f(x)$ is the pdf.
  - cumulative distribution function (cdf): $F(x) = \mathbb{P}(X \leq x)$.
  - Normal distribution: $N(\mu, \sigma^2)$, μ is the mean and σ^2 is the variance.
  - If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1^2)$.
  - If $Z \sim N(0, 1^2)$, then $X = Z\sigma + \mu \sim N(\mu, \sigma^2)$.
  - $P(X < x) = P(Z = \frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma})$.
  - t distribution
  - F distribution
- CLT

Point estimation

- Simple random sample (SRS)
- Estimator, standard error, estimated standard error
- Bias, standard error, MSE
- Estimation of population mean
- Estimation of population proportion
- $X_1, ..., X_n$ are iid with mean μ and variance σ^2, then $\mathbb{E}(\bar{X}) = \mu$, $VAR(\bar{X}) = \frac{\sigma^2}{n}$.

Confidence interval

- Interpretation of a 95% confidence interval
- How can we build a confidence interval for population mean when σ is known and σ is unknown?
- $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $X_i \sim N(\mu, \sigma^2)$ i.i.d. and S is the sample standard deviation.
- How can we determine the required sample size to achieve certain confidence level?
- How can we make a confidence interval for population proportion? What is the assumption we need to check?
- How can we use Bootstrap method to build confidence interval?

Hypothesis testing

- Null hypothesis, alternative hypothesis
- Type I error, type II error, power
- What is the relationship between type I error and type II error for fixed sample size?
- How can we reduce type I error or type II error?
- How can we increase power?
- How can we calculate the power?
- What is p-value?
- What are two common methods to make a conclusion in hypothesis testing?

Hypothesis testing

- Sample mean test
- Z test
- T test
- Sign test
- Z test for population proportion
- Two sample t test
- Welch t test
- Permutation test
- Z test for comparing two independent population proportions
- Paired t test
- Paired sign test
- ANOVA

Simple linear regression

- Pearson correlation coefficient definition, interpretation
- interpretation of slope in simple linear regression
- OLS estimator
- Confidence intervals of intercept and slope
- Hypothesis testing in simple linear regression
- R squared
- Understanding results from R output


KEEP
CALM
AND
GOOD LUCK
ON
YOUR EXAMS