

Math 455 Class 17 February 28

Midterm 1 : Monday, March 3

In - class, 80 minutes

Not everything can be covered!

Crazy R tricks are (mostly) postponed to the next midterm.

Examples of questions (which didn't make the cut)

~~Fit~~ Design a model based on a 1-paragraph description of the problem.

Estimate R^2 from a scatter plot
(or ρ)

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Knowing β_0, β_1 , and given $X = 6$

our prediction of Y is $\hat{Y} = \beta_0 + \beta_1 \cdot 6$

the error will be $\epsilon \sim N(0, \sigma^2)$.

8. Suppose we have a data frame `df` with two variables `X` and `Y` and we create a linear model for `Y` in terms of `X` as below. Prediction and confidence intervals can be obtained from the `predict` command as shown below.

```
> lmod <- lm(Y ~ X, data = df)
> predict(lmod, newdata=data.frame(X=6), interval="prediction")
> predict(lmod, newdata=data.frame(X=6), interval="confidence")
```

- (a) (6 points) Which of the two intervals will be narrower, and why?

The confidence interval (for the mean of many observations of Y with $X=6$) will be narrower. Averaging over many observations reduces the std. err.

- (b) (2 points) Will the two intervals have the same midpoint? (Only a yes/no answer is necessary.)

YES.

- (c) (4 points) If this midpoint is m , what is the relationship between the point $(6, m)$ and the regression line?

The point $(6, m)$ is on the regression line.

9. (4 points) Fill in the blanks in the following definition of the F -distribution.

Let W_1, W_2 be independent random variables, having the χ^2 -distribution with ν_1, ν_2 degrees of freedom respectively. Then the random variable

$$F = \frac{W_1 / \nu_1}{W_2 / \nu_2}$$

is said to have the F -distribution with ν_1 numerator degrees of freedom,

and ν_2 denominator degrees of freedom.

9. Which of the following is *not* a valid null hypothesis?
- A. $H_0: \beta_1 = 0$
 - B. $H_0: \beta_1 = \beta_2$
 - C. $H_0: b_1 = b_2 = 0$ $[\hat{\beta}_1 = \hat{\beta}_2 = 0]$
 - D. All of the above *are* valid null hypotheses
10. Which of the following can never be 0 (unless the population standard deviation $\sigma = 0$)?
- A. The estimated intercept, b_0
 - B. A studentized deleted residual, t_i
 - C. The variance of the prediction error, $\sigma^2\{\text{pred}\}$
 - D. The estimate of $E\{Y_h\}$, \hat{Y}_h

Final Note: (useful for interviews!)

ATQ! Answer the question!