# Math 2311

## TEST 2 REVIEW SHEET

## KEY

---

**#1 – 25, Define the following:**

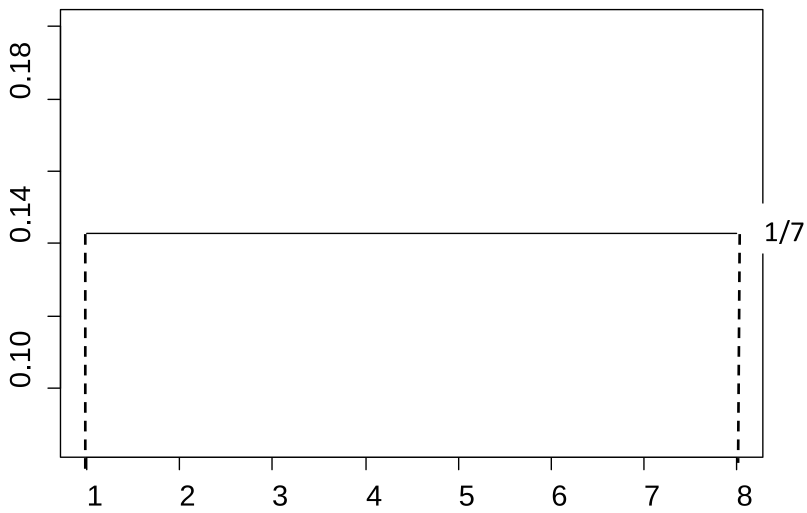
1. Continuous random variable
2. Discrete random variable
3. Density curve
4. Uniform density curve
5. Normal distribution
6. Sampling distribution (for $\bar{x}$ and $\hat{p}$)
7. Z-score
8. LSRL
9. Correlation coefficient
10. Coefficient of determination

11. Interpretation of slope of the LSRL
12. Residual
13. Sample
14. Population
15. Census
16. Simple random sample
17. Other types of sampling design
18. Experiment
19. Observational study
20. Bias
21. Subjects
22. Treatments
23. Factors
24. Control (three fundamental principles of)
25. Simulation

26. Consider a uniform density curve defined from x = 1 to x = 8.
    a. What is the height of the "curve"?

       Height = $\frac{1}{8-1} = \frac{1}{7}$

       The following is a plot of the "curve"




b. What percent of observations fall between x=2 and x=5?
Probability for uniform distribution = Area of a rectangle = length × height
$P(2 < X < 5) = (5 - 2)(1/7) = (3)(1/7) = 3/7 \approx 0.4286$

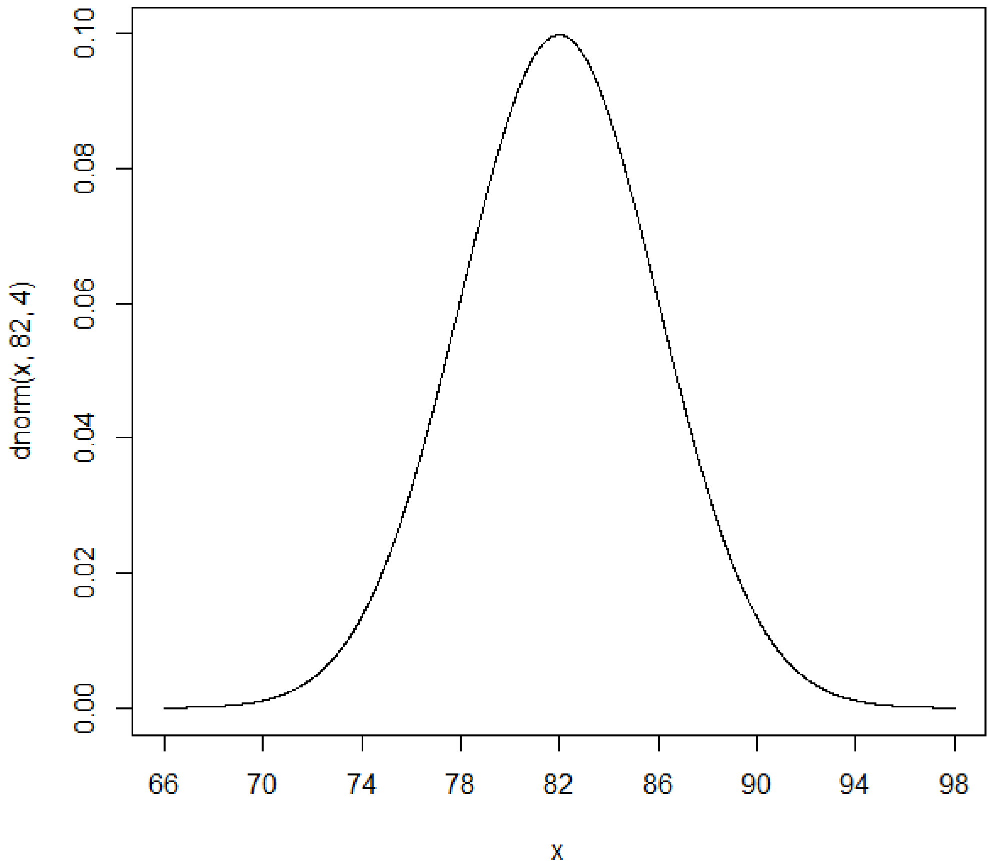
c. What percent of observations fall below x = 4?
$P(X < 4) = (4 - 1)(1/7) = 3(1/7) = 3/7 \approx 0.4286$

d. What percent of observations fall above x = 6?
$P(X > 6) = (8 - 6)(1/7) = 2(1/7) = 2/7 \approx 0.2857$

e. What percent of observations equal 7?
$P(X = 7) = 0$ (This is true for all continuous random variables)

27. Let $X$ be a normal random variable with $\mu = 82$ and $\sigma = 4$.

a. Sketch the distribution




b. According to the Empirical Rule, the middle 68% of the data falls between what values?
According to the Empirical Rule 1 standard deviation is 68% that is between 78 and 86.

c. Find $P(X < 83)$
P(X < 83) = pnorm(83,82,4) = 0.5987

d. Find $P(X > 79)$
P(X > 79) = 1 – pnorm(79,82,4) = 1 – 0.2266 = 0.7734

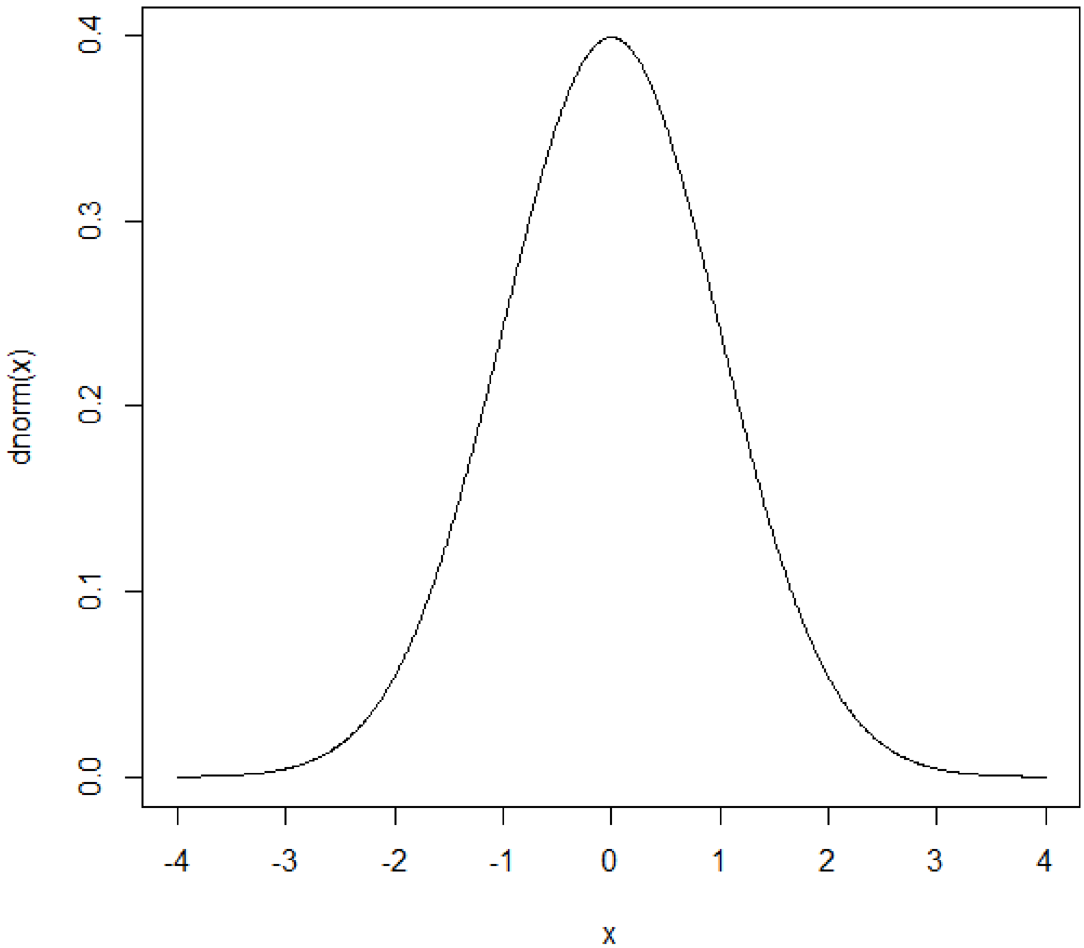
e. Find $P(73 < X < 84)$
P(73 < X < 84) = pnorm(84,82,4) – pnorm(73,82,4) = 0.6915 – 0.0122 = 0.6973

f. Find $x$ such that $P(X < x) = .97725$
qnorm(0.97725,82,4) = 90

28. Recall $Z$ is the standard normal random variable.

a. What is the mean and standard deviation for $Z$?
   $\mu = 0, \sigma = 1$

b. Sketch the distribution




c. Find $P(Z < 1.2)$
   From table: $P(Z < 1.2) = 0.8849$
   From R: P(Z < 1.2) = pnorm(1.2) = 0.8849

d. Find $P(Z < -1.64)$
   From table: $P(Z < -1.64) = 0.0505$
   From R: P(Z < -1.64) = pnorm(-1.64) = 0.0505

e. Find $P(Z > -1.39)$
   From table: $P(Z > -1.39) = 1 - 0.0823 = 0.9177$
   From R: P(Z > -1.39) = 1 – pnorm(-1.39) = 0.9177

f. Find $P(-0.45 < Z < 1.96)$
   From table: $P(-0.45 < Z < 1.96) = 0.9750 - 0.3264 = 0.6486$
   From R: pnorm(1.96) – pnorm(-0.45) = 0.6486

g. Find $c$ such that $P(Z < c) = 0.845$
   From table: c = 1.02 (area = 0.8461)

From R: qnorm(0.845) = 1.0152

h. Find $c$ such that $P(Z > c) = 0.845$

From table: use 1 – 0.845 = 0.155, c = -1.02 (area = 0.1539)

From R: qnorm(1-0.845) = -1.01522

i. Find $c$ such that $P(-c < Z < c) = 0.845$

From table : use ½ (1 – 0.845) = 0.0775, -c = -1.42, +c = +1.42

From R: qnorm((1-0.845)/2) = -1.422, c = +1.422

29. Suppose a sample of 100 subjects was taken and their scores on an exam recorded. If the population mean for the exam is 67 and population variance is 36,

a. what is the mean and standard error of the sampling distribution, $\bar{X}$?

Mean = μ = 67, Standard Error = SD( $\bar{X}$) = $\sqrt{\frac{36}{100}}$ = 0.6

b. find $P(\bar{X} < 70)$.

pnorm(70,67,0.6) ≈1

c. find $P(45 < \bar{X} < 74)$.

pnorm(74,67,0.6) – pnorm(45,67,0.6) = 1

30. What is the difference between the distributions for $X$ and $\bar{X}$?

For the distribution of $\bar{X}$, we have to take the original standard deviation and divide by the square root of the sample size ($n$).

31. The following data indicates the number of hours a swimmer practiced during a week and his best time on the 50 meter free style that week.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Hrs practicing | 2.5 | 4 | 4.5 | 6 | 7 | 7.5 | 8.5 | 9 | 11 |
| Time/sec | 29.33 | 28.76 | 28.01 | 27.96 | 27.99 | 27.35 | 27.02 | 26.85 | 26.09 |

a. Identify the explanatory and response variables for this situation.

Explanatory: Hours practiced

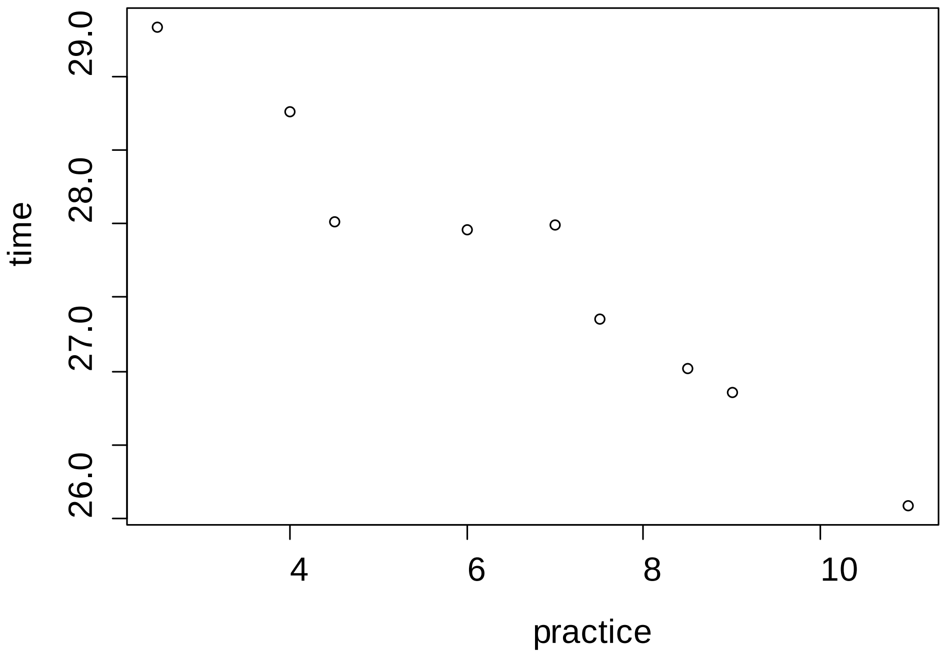
Response: Best time per seconds

b. Create a scatterplot

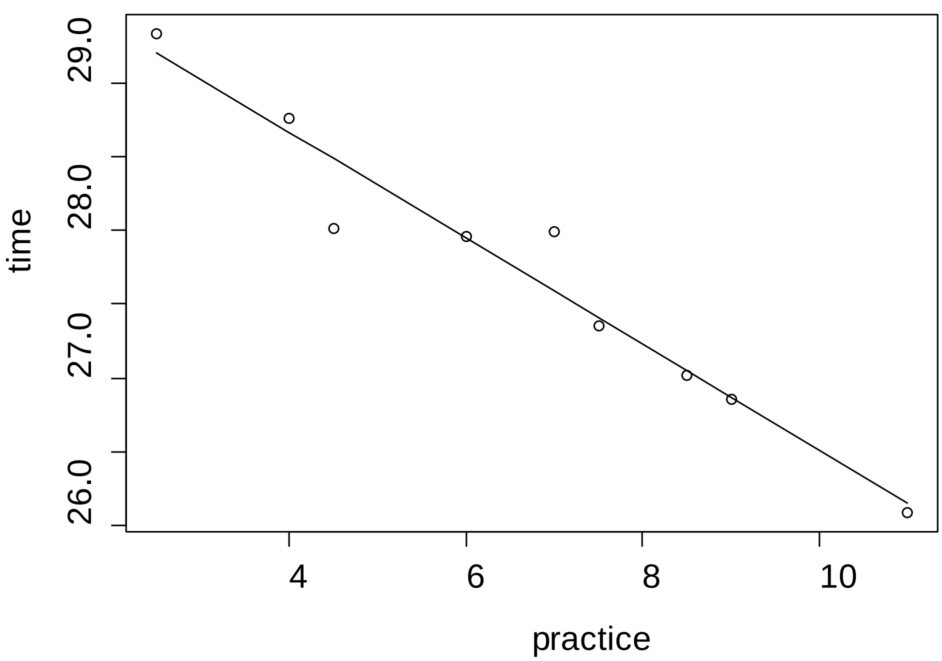
Rcode:

```
> practice<-c(2.5,4,4.5,6,7,7.5,8.5,9,11)
> time<-c(29.33,28.76,28.01,27.96,27.99,27.35,27.02,26.85,26.09)
> plot(practice,time)
```

Plot:



29.0
28.0
27.0
26.0
time
4
6
8
10
practice

29.0
28.0
27.0
26.0
time
4
6
8
10
practice

c. Give the equation for the LSRL and plot the LSRL on the scatterplot

Rcode to put line on scatterplot:

```
> lines(practice,predict(lm(time~practice)))
```

Rcode and results for LSRL equation:

```
> lm(time~practice)

Call:
lm(formula = time ~ practice)

Coefficients:
(Intercept)     practice
    30.1049      -0.3597
```

Equation: $\hat{y} = 30.1049 - 0.3597x$

d. Find the correlation coefficient and the coefficient of determination. What do each of these tell you about the data?

Rcode and results:

```
> cor(practice,time)
[1] -0.9730586
> cor(practice,time)^2
[1] 0.9468431
```

Correlation = r = -0.9731, this tells me that there is a very strong *negative* association between hours practiced and best time. (As the amount of practice increases, his best time for swimming decreases.)

Coefficient of determination = $R^2$ = 0.9468, this tells me that about 95% of the variation in the best times can be explained by this equation.

e. Based on your answers to b and d, is this a good model?

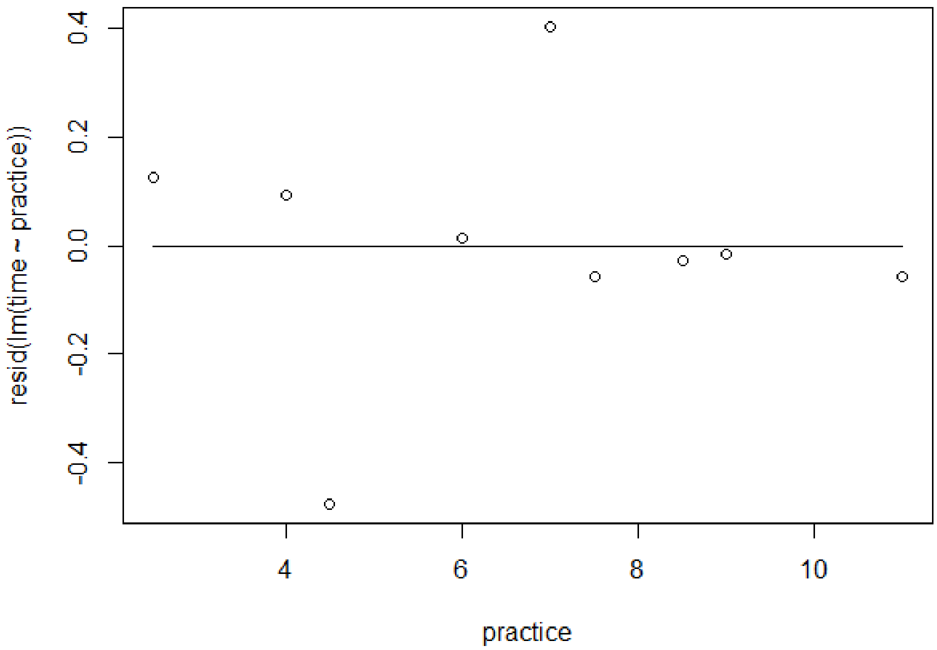
Yes, the closer $R^2$ is to 1, the better our model is a predicting the response variable.

f. Plot the residuals vs explanatory variables.

Rcode:

```
> plot(practice,resid(lm(time~practice)))
> lines(practice,rep(0,length(practice)))
```

Plot:




g. Based on your answer to f, do you still think your LSRL is the best model?
Yes the plots appear to be centered about the zero line no appearance of a form.

h. Find the residual value that corresponds to the explanatory variable value of 4.
In the observed values, when x = 4, y = 28.76. The predicted y-value is:

$$\hat{y} = 30.1049 - 0.3597(4) = 28.6661$$

Then the residual is: 28.76 – 28.6661 = 0.0939

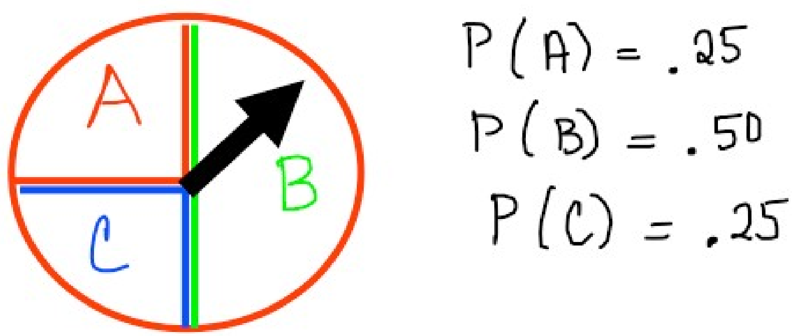
32. 1000 students were asked to give their favorite subject and favorite video game (chosen from a list). The results are recorded in this two-way table:

| | Math | Science | English | Social Studies | |
|---|---|---|---|---|---|
| Zelda | 66 | 70 | 40 | 35 | 211 |
| Final Fantasy | 54 | 75 | 60 | 30 | 219 |
| Tomb Raider | 35 | 50 | 80 | 90 | 255 |
| Assassin's Creed | 45 | 40 | 60 | 100 | 245 |
| None of these | 10 | 5 | 20 | 35 | 70 |
| | 210 | 240 | 260 | 290 | 1000 |

a. Complete the table by filling in the marginal distributions.
   See table above
b. What is the probability that someone likes Tomb Raider?
   255/1000 = 0.255
c. What percent of math students like Zelda?
   66/210 = 0.3143
d. What percent of people who like Assassin's Creed also like English?
   60/245 = 0.2449
e. What is the probability that someone likes both Science and Final Fantasy?
   75/1000 = 0.075

33. Classify each as a experiment or observational study:
    a. A professor is curious as to what the students on campus eat for lunch. He sits at the UC and watches the students between 11am and 1pm and records his findings.
       Observational study
    b. A radio station wants to know more about its listeners so a representative travels to shopping malls in the listening area to ask people what their favorite radio station is.
       Observational study
    c. A farmer wishes to know if a new feed makes a difference in how his cows behave. He gives half of his cows the new feed while the remaining cows do not change diet. He records the results after two weeks.
       Experiment

34. A game is played with the spinner below. If your spin lands on A, you win $1. If your spin lands on B, you lose $1. If the spinner lands on C, nothing happens. Ten people are playing the game.
    a. Using single digits from the random digit table, describe how you will run a simulation for the 10 players.
       Ignore 0 and 9, select 10 digits from the table from 1 – 8. If the digit is a 1 or 2 that represents that the spinner landed on A, if the digit is 3, 4, 5, or 6 that represents the spinner landed on B, if the digit is 7 or 8, that represents the spinner landed on C.




    b. Using line 120 from the random digit table, carry out the simulation 3 times.
       Starting at line 120
       Run 1: 8716 627124
       Spinner: CCABBACAAB
       Run 2: 556 5367831
       Spinner: BBBBBBCCBA
       Run 3: 8612 814287
       Spinner: CBAACABACC

c. Based on your simulation, how many people won $1 for each run?  How many lost $1?

| | Win $1 | Lose $1 | Nothing happens |
|---|---|---|---|
| Run 1 | 4 | 3 | 3 |
| Run 2 | 1 | 7 | 2 |
| Run 3 | 4 | 2 | 4 |
| Average | 9/3 = 3 | 12/3 = 4 | 9/3= 3 |