$$F(x) = \begin{cases} 0 & x \le 0 \\ \frac{x^2}{9} & 0 < x < 3 \\ 1 & x \ge 3 \end{cases}$$

$$E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$$

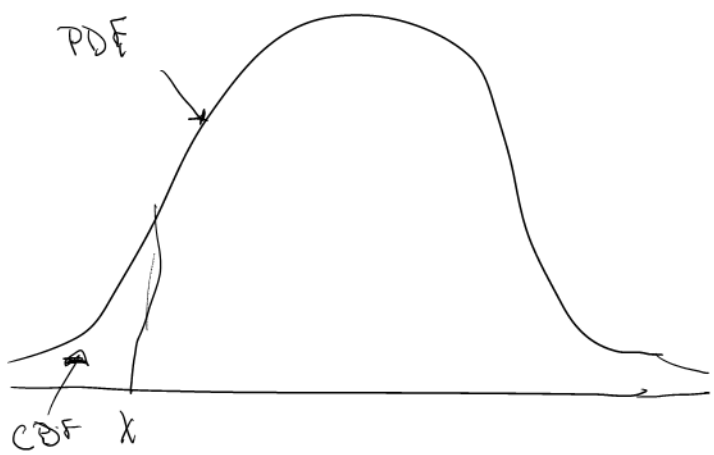
$$f(x) = F'(x) = \frac{2x}{9}\, dx$$

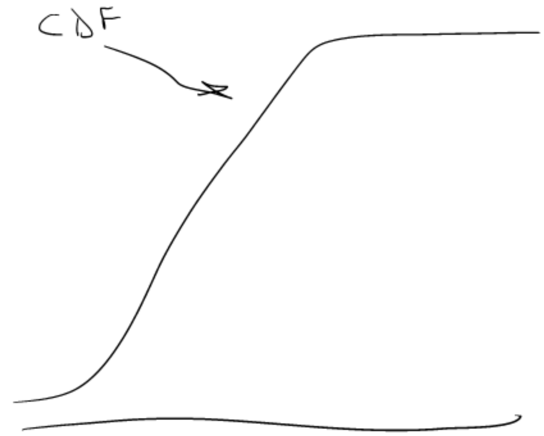
$$E(x) = \int_0^3 x\left(\frac{2x}{9}\right) dx = \frac{2}{9}\int_0^3 x^{2} dx = \frac{2}{27} x^3 \Big|_0^3 = \frac{2}{27}(3)^3 - \frac{2}{27}(0)^3$$

$+1$

$3 * 9$

$$= \underline{\underline{2}}$$





Math 3339 – Fall 2017
Review for Final Exam

---

1. (Assuming a set has all positive values) If the largest value of a data set is doubled, which of the following is <u>not</u> true?
   a. The mean increases.
   b. The standard deviation increases.
   c. The interquartile range increases.
   d. The range increases.

2. A potato chip company calculated that there is a mean of 74.1 broken potato chips in each production run with a standard deviation of 5.2. If the distribution is approximately normal, find the probability that there will be fewer than 60 broken chips in a run.

3. What does it mean if the correlation coefficient, *r*, is close to 1? is close to 0?

4. In the regression equation $y = b_0 + b_1x$, identify what b_0 and b_1 represent. Which of these will best explain the relationship between x and y?

5. Be able to identify type I and type II errors from an example.

6. It is fourth down and a yard to go for a first down in an important football game. The football coach must decide whether to go for the first down or punt the ball away. The null hypothesis is that the team will not get the first down if they go for it. The coach will make a Type I error by doing what?
   a. Deciding not to go for the first down when his team will get the first down. ⇐ correct
   b. Deciding not to go for the first down when his team will not get the first down. ⇐ Type II
   c. Deciding to go for the first down when his team will get the first down. ⇐ correct
   d. Deciding to go for the first down when his team will not get the first down. ⇐ Type I
   e. None of the above.

H_0: Not Get First down

H_A: Get First down

	H_0 is true	H_0 is false
Reject H_0 go for it	Type I error	Correct
Fail to reject Punt	Correct	Type II

7. The following table displays the results of a sample of 100 in which the subjects indicated their favorite ice cream of three listed. The data are organized by favorite ice cream and age group. What is the probability that a person chosen at random will be over 40 if he or she favors chocolate?

Age	Chocolate	Vanilla	Strawberry
Over 40	15	8	7
20 – 40	20	11	15
Under 20	8	7	9

8. A random variable X has a probability distribution as follows:

X	0	1	2	3
$P(X)$	4k	5k	8k	3k

R-Studio

```
> lcl=11*7.7^2/qchisq(1.95/2,11)
>
ucl=11*7.7^2/qchisq(0.05/2,11)
> c(lcl,ucl)
[1]  29.75313 170.92061
```

↑ var

sd = sqrt(c(lcl,ucl))

Find the probability that $P(X < 2.0)$

9. The amount of time it takes a bat to eat a frog was recorded for each bat in a random sample of 12 bats. The resulting sample mean and standard deviation were 21.9 minutes and 7.7 minutes, respectively. Assuming it is reasonable to believe that the population distribution of bat mealtimes of frogs is approximately normal,
   a. Construct a 95% confidence interval for the mean time for a bat to eat a frog.
   b. Construct a 95% confidence interval for the variance of the time for a bat to eat a frog. σ^2

$lcl = \frac{(n-1)s^2}{qchisq(1.95/2, n-1)}$

$ucl = \frac{(n-1)s^2}{qchisq(0.05/2, n-1)}$

10. Suppose that the average weekly grocery bill for a family of four is \$140 with a standard deviation of \$10. If the next 52 weeks can be viewed as a random sample from a population with this center and spread, then the approximate probability that the total grocery bill for one year is less than \$7020 is 0.0002. This problem can also be solved using the sampling distribution of the sample average. Complete probability statement.

$$P\left(Z < \underline{\qquad\qquad\qquad}\right)$$

11. One of your peers claims that boys do better in math classes than girls. Together you run two independent simple random samples and calculate the given summary statistics of the boys and the girls for comparable math classes. In Calculus, 15 boys had a mean percentage of 82.3 with standard deviation of 5.6 while 12 girls had a mean percentage of 81.2 with standard deviation of 6.7. Which of the following would be the most appropriate test for establishing whether boys do better in math classes than girls?
    a. two-sample z-test for means
    b. two-sample t-test for means
    c. chi-square test
    d. two-sample z-test for proportions
    e. none of these tests would be appropriate

X = amount of acidity

$E(X) = 4.6$

$SD(X) = 1.1$

$1.2(X + 0.1) = Y$

12. Rainwater was collected in water collectors at thirty different sites near an industrial basin and the amount of acidity (pH level) was measured. The mean and standard deviation of the values are 4.60 and 1.10 respectively. When the pH meter was recalibrated back at the laboratory, it was found to be in error. The error can be corrected by adding 0.1 pH units to all of the values and then multiply the result by 1.2. Find the mean and standard deviation of the corrected pH measurements.

$E(Y) = E[1.2(X + 0.1)]$

$E(1.2X + 0.12)$

$1.2(4.6) + 0.12$

$= 5.64$

13. What is the expected value and the variance of the discrete probability function given in the table below?

Outcome	1	2	3	4	5	6
Probability	.1	.2	.3	.3	0	.1

$SD(Y) = SD(1.2X + 0.12)$

$= 1.2(1.1)$

$= 1.32$

$E(X) = 1(0.1) + 2(0.2) + 2(0.3) + 4(0.3) + 5(0) + 6(0.1) = 1$

$var(X) = E(X^2) - [E(X)]^2$

$E(X^2) = 1^2(0.1) + 2^2(0.2) + 3^2(0.3) + 4^2(0.3) + 5^2(0) + 6^2(0.1) =$

14. The following is a stem-plot of the birth weights of male babies born to the smoking group. The stems are in units of kg.

Stems	Leaves
2	3,4,6,7,7,8,8,8,9
3	2,2,3,4,6,7,8,9
4	1,2,2,3,4
5	3,5,5,6

Find the median birth weight. Describe the shape of the distribution.

15. A sample of 100 engineers in a large consulting firm indicated that the mean amount of time they spend reading for pleasure each week is 1.4 hours. Three interns independently calculate different two-sided confidence intervals of the true mean amount of time for all of the engineers in the company. The confidence intervals of the interns were:

$\bar{x} \pm ME$

A) (.17, 2.63) B) (.554, 2.446) C) (1.167, 1.633)

a. All are calculated correctly with different levels of confidence.
b. A and C have reasonable intervals, but B does not.
c. A and B have reasonable intervals, but C does not.
d. B and C have reasonable intervals, but A does not.
e. None of these intervals is reasonable.

$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

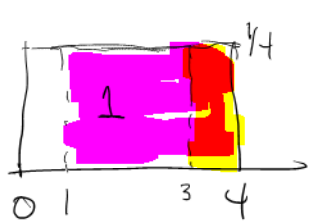
$0.7 \pm 1.96\sqrt{\frac{0.7(1-0.7)}{130}}$

$(0.62, 0.79)$

16. It has been estimated that as many as 70% of the fish caught in certain areas of the Great Lakes have liver cancer due to the pollutants present. Find an approximate 95% range for the percentage of fish with liver cancer present in a sample of 130 fish.

17. In a recent publication, it was reported that the average highway gas mileage of tested models of a new car was 33.5 mpg and approximately normally distributed. A consumer group conducts its own tests on a simple random sample of 12 cars of this model and finds that the mean gas mileage for their vehicles is 31.6 mpg with a standard deviation of 3.4 mpg.
   a. Perform a test to determine if these data support the contention that the true mean gas mileage of this model of car is different from the published value.
   b. Perform a test to determine if these data support the contention that the true mean gas mileage of this model of car is less than the published value.
   c. Explain why the answers to part a and part b are different.

18. It has been determined that the amount of time that videotapes are returned late to a certain rental store is modeled by a uniform distribution from 0 to 4 days. Answer each question showing a figure and your work.
   a. What is the probability that a randomly selected videotape will be returned between 3 and 4 days late?
   b. What is the probability that a randomly selected videotape will be returned more than 1 day late?



$A = bh$
$1 = 4h$
$h = 1/4$

$P(3 \le X \le 4) = (4-3)\frac{1}{4} = \frac{1}{4}$

$P(X > 1) = (4-1)\frac{1}{4} = 3/4$

```
> it=c(295,152,214,171,131,178,225,141,116,173,230,195,174,177,210,236,189,217,143,186)
> gpa=c(2.4,.6,.2,0,1,.6,1,.4,0,2.6,2.6,0,1.8,0,.4,1.8,.8,1,.2,2.8)
> rr=c(41,18,45,29,28,38,25,26,22,37,39,38,24,32,26,29,34,38,40,27)
> it.lm=lm(gpa~it)
> rr.lm=lm(gpa~rr)
> summary(it.lm)
Call:
lm(formula = gpa ~ it)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0889 -0.5124 -0.2380  0.3561  1.8075

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.976012   0.926602  -1.053   0.3061
it           0.010584   0.004823   2.194   0.0416 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8872 on 18 degrees of freedom
Multiple R-squared:  0.211, Adjusted R-squared:  0.1672
F-statistic: 4.815 on 1 and 18 DF,  p-value: 0.04159

> summary(rr.lm)

Call:
lm(formula = gpa ~ rr)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0950 -0.8874 -0.2305  0.8455  1.8558

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.57425    1.00480   0.572    0.575
rr           0.01370    0.03082   0.445    0.662

Residual standard error: 0.9934 on 18 degrees of freedom
Multiple R-squared:  0.01087, Adjusted R-squared:  -0.04409
F-statistic: 0.1977 on 1 and 18 DF,  p-value: 0.6619
```

19. The following data are for intelligence-test (IT) scores, grade-point averages (GPA), and reading rates (RR) of 20 at-risk students.

IT	295	152	214	171	131	178	225	141	116	173
GPA	2.4	.6	.2	0	1	.6	1	.4	0	2.6
RR	41	18	45	29	28	38	25	26	22	37
IT	230	195	174	177	210	236	198	217	143	186
GPA	2.6	0	1.8	0	.4	1.8	.8	1	.2	2.8
RR	39	38	24	32	26	29	34	38	40	27

a. Calculate the line of best fit that predicts the GPA on the basis of IT scores.
b. Calculate the line of best fit that predicts the GPA on the basis of RR scores.
c. Which of the two lines calculated in parts a and b best fits the data?

Look at p-value for testing $H_0: \beta_1 = 0$ and R^2

20. A manufacturer claims that its quality control is so effective that no more than 2% of the parts in each shipment are defective. A simple random sample of 100 parts from the last shipment contained 3 defectives.
    a. Why is a hypothesis test to determine the validity of the company's claim inappropriate? Explain your answer.
    b. What is the smallest sample size for which a test of the claim would be appropriate at a significance level of 0.05? Show your work.
    c. Suppose your answer in part b were the sample size used. Perform an appropriate test of the manufacturer's claim at the 5% level. Assume that the observed proportion is .03 for this sample size.

21. A researcher claims that 90% of people trust DNA testing. In a survey of 100 people, 91 of them said that they trusted DNA testing. Test the researcher's claim at the 1% level of significance.

$H_0: p = 0.9$

$H_A: p \neq 0.9$

$\hat{p} = \frac{91}{100} = 0.91$

22. The dean of students of a large community college claims that the average distance that commuting students travel to the campus is 32 miles. The commuting students feel otherwise. A sample of 64 students was randomly selected and yielded a mean of 35 miles and a standard deviation of 5 miles. Test the dean's claim at the 5% level.

Test statistic

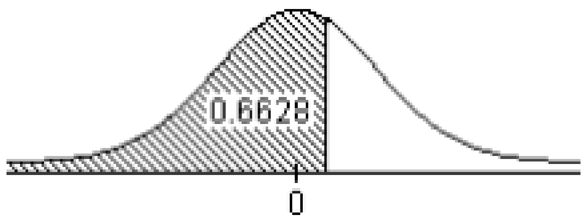
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.91 - 0.9}{\sqrt{\frac{0.9(1-0.9)}{100}}}$$

$z = 0.3333$

p-value $= P(Z \leq -0.3333 \text{ or } Z \geq 0.3333) = 2 * \text{pnorm}(-0.3333) = 0.7389$ FRH_0

23. A random sample of size 36 selected from a normal distribution with $\sigma = 4$ has $\bar{x} = 75$. A second random sample of size 25 selected from a different normal distribution with $\sigma = 6$ has $\bar{x} = 85$. Is there a significant difference between the two population means at the 5% level of significance?

24. Find the z-score that corresponds to the given area under the standard normal curve.



25. For the standard normal curve, find the z-score that corresponds to the 30th percentile.

$SE = \sqrt{\frac{P(1-P)}{n}}$

$= \sqrt{\frac{0.25(1-0.25)}{200}}$

26. In an opinion poll, 25% of 200 people sampled said they were strongly opposed to the state lottery. The standard error of the sample proportion is approximately what?

27. What is the critical value t* which satisfies the condition that the t distribution with 8 degrees of freedom has probability 0.10 to the right of t*?

28. The one-sample t statistic for a test of $\begin{matrix} H_0: \mu = 10 \\ H_a: \mu < 10 \end{matrix}$ based on n = 10 observations has the value t = -2.25. What is the p-value for this test?

$p^* = 0.5$

$C = 95\%$

$z = 1.96$

width $= 0.05$

$m = \frac{0.05}{2} = 0.025$

$n = \left(\frac{z}{m}\right)^2 p(1-p)$

$= \left(\frac{1.96}{0.025}\right)^2 (0.5)(1-0.5)$

$= 1536.64$

1537

29. Suppose that prior to conducting a coin-flipping experiment, we suspect that the coin is fair. How many times would we have to flip the coin in order to obtain a 95% confidence interval of width of at most 0.05 for the probability of flipping a head?

30. The guidance office of a school wants to test the claim of an SAT test preparation company that students who complete their course will improve their SAT Math score. Ten members of the junior class who have had no SAT preparation but have taken the SAT once were selected at random and agreed to participate in the study. All took the course and re-took the SAT at the next opportunity. The results of the testing indicated the values below. Is there significant evidence to support the claim that there is an improvement in the SAT scores after the test prep course?

Student	1	2	3	4	5	6	7	8	9	10
Before	475	512	492	465	523	560	610	477	501	420
After	500	540	512	530	533	603	691	512	489	458

31. Suppose that in a large metropolitan area, 90% of all households have a microwave oven. Consider selecting groups of six households, and let X be the number of households in a group that have microwave oven.
    a. Verify that this is a binomial distribution.
    b. For what proportion of groups will exactly four of the six households have a microwave oven?
    c. For what proportion of groups will at most two of the households have a microwave oven?
    d. What is the proportion of groups for which at least five of the six households have a microwave oven?

32. Samples of head breadths were obtained by measuring skulls of Egyptian males from three different epochs, and the measurements are listed below (based on data from Ancient Races of the Tebaid, by Thomas and Randall-Maciver). Changes in head shape over time suggest that interbreeding occurred with immigrant populations. Test the claim that the different epochs do not all have the same mean head breadth. The analysis of the data was run and the output is shown below:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
epoch	2	138.74	69.37	4.05	0.0305
Residuals	24	411.11	17.13		

33. Suppose the random variable *X* has a pdf given by $f(x) = \frac{1}{2}x^2$. If $0 \le x \le a$, find *a* so that this is a valid pdf.

34. The weight of bolts produced by a machine has standard deviation of 0.03 pounds. Assuming that the distribution is normal, how large a sample is needed to determine with a precision of ±0.005 pounds the mean weight of the produced bolts to 90% confidence?

35. The president of an all-female school stated in an interview that she was sure that the students at her school studied more, on average, than the students at a neighboring all-male school. The president of the all-male school responded that he thought the mean study time for each student body was undoubtedly about the same and suggested that a study be undertaken to clear up the controversy. Accordingly, independent samples were taken at the two schools with the following results:

School	Sample Size	Mean Study Time (hrs)	Standard deviation (hrs)
All Female	65	18.56	4.35
All Male	75	17.95	4.87

Determine, at the 2% level of significance, if there is a significant difference between the mean studying times of the students in the two schools based on these samples.

36. The data in the accompanying table resulted from an experiment run in a completely randomized design in which each of four treatments was replicated five times.

						Total	Mean
Group 1	6.9	5.4	5.8	4.6	4.0	26.70	5.34
Group 2	8.3	6.8	7.8	9.2	6.5	38.60	7.72
Group 3	8.0	10.5	8.1	6.9	9.3	42.80	8.56
Group 4	5.8	3.8	6.1	5.6	6.2	27.50	5.50
All Groups						135.60	6.78

① $W = qtukey(1-\alpha, k, N-k)\sqrt{\frac{MSE}{j}}$

k = # of groups

N = total number of observations

j = # of obs. in each group.

W = 2.0875

② Any pairs of means that has a difference of more than W = 2.0875 is significantly different

(1,2) $|5.34-7.72| = 2.38$

(1,3) $|5.34-8.56| > 2.0875$

(1,4) $|5.34-5.5| = 0.16 < 2.0875$ Not S.D.

2, 3 NO

(2,4)

(3,4)

Part of the resulting ANOVA table is

Source	SS	DF	MS	F
Treatments	38.820	3	12.940	$\frac{MSTr}{MSE} = \frac{12.94}{1.331} = 9.722$
Error	21.292	16	1.331	

P-value = 1 − pf(9.722, 3, 16) = 0.000685

a. Complete the ANOVA table.

b. Perform a significance test to see if at least two of the μ_i's are different. Use Tukey's method to determine which pairs differ significantly.

W = qtukey(1-0.05, 4, 16) * sqrt(1.331/5) = 2.0875

37. A study was conducted to determine whether remediation in basic mathematics enabled students to be more successful in an elementary statistics course. (Success here means C or better.) Here are the results of the study:

	Remedial	Non-remedial
Sample size	100	40
# of successes	70	16

Test, at the 5% level, whether the remediation helped the students to be more successful.

38. A preacher would like to establish that of people who pray, less than 80% pray for world peace. In a random sample of 110 persons who pray, 77 of them said that when they pray, they pray for world peace. Test at the 10% level.

39. Two methods were used to teach a high school algebra course. A sample of 75 scores was selected for method 1, and a sample of 60 scores was selected for method 2. The results are:

	Method 1	Method 2
Sample mean	85	83
Sample s.d.	3	2

Test whether method 1 was more successful than method 2 at the 1% level.

40. The table below displays the performance of 10 randomly selected students on the SAT Verbal and SAT Math tests taken last year.

Student	1	2	3	4	5	6	7	8	9	10
Math	475	512	492	465	523	560	610	477	501	420
Verbal	500	540	512	530	533	603	691	512	489	458

a. Calculate the least-squares regression line for this data. Report r and r-squared.
b. Compute the 90% confidence interval. Interpret this confidence interval by describing for me in words what it means in the context of this problem.
c. Is there a significant linear relationship between the variables? State the hypotheses, t-statistic, p-value, and conclusion.

41. A midterm exam in Applied Mathematics consists of problems in 8 topical areas. One of the teachers believes that the most important of these is the section on problem solving. She analyzes the scores of 36 randomly chosen students using computer software and produces the following print-out relating the total score to the problem-solving subscore, ProbSolv:

Predictor	Coef	StDev	T	p	
Constant	12.960	6.228	2.08	0.045	s = 11.09
ProbSolv	4.0162	0.5393	7.45	0.000	R-sq = 62.0%
					R-sq (adj) = 60.9%

a. What is the regression equation?
b. Interpret the slope of the regression in the context of the problem.
c. Interpret the value of R-Sq in words.
d. Calculate the 95% confidence interval of the slope of the regression line for all students.
e. Use the information provided to test whether there is a significant relationship between the problem solving subsection and the total score at the 5% level.