# Math 3339

TEST 2 – Fall 2017

Key

REVIEW SHEET

---

1. A headache remedy is said to be 80% effective in curing headaches caused by simple nervous tension. An investigator tests this remedy on 100 randomly selected patients suffering from nervous tension.
   a. What kind of distribution does X have?
   b. Calculate the mean and standard deviation of X.
   c. Determine the probability that exactly 80 subjects experience headache relief with this remedy.
   d. What is the probability that between 75 and 90 (inclusive) of the patients will obtain relief?
   e. Using the Normal approximation what would be the answer to the previous question?

   a. This is a binomial distribution with n = 100 and p = 0.8
   b. Mean = 100(0.8) = 80, standard deviation = $\sqrt{100(0.8)(1-0.8)} = 4$
   c. P(X = 80) = dbinom(80,100,0.8) = 0.0993
   d. P(75 ≤ X ≤ 90) = P(X ≤ 90) – P(X ≤ 74)
      = pbinom(90,100,0.8) – pbinom(74,100,0.8) = 0.9102
   e. =pnorm(90.5,80,4) – pnorm(74.5,80,4) = 0.9111

2. A restaurant serves three fixed-price dinners costing \$12, \$15, and \$20. For a randomly selected couple dining at this restaurant, let *X* = the cost of the man's dinner and *Y* = the cost of the woman's dinner. The joint pmf of *X* and *Y* is given in the following table:

| P(*x*, *y*) | | *Y* 12 | 15 | 20 |
|---|---|---|---|---|
| | 12 | 0.05 | 0.05 | 0.10 |
| *X* | 15 | 0.05 | 0.10 | 0.35 |
| | 20 | 0 | 0.20 | 0.10 |

   a. Compute the marginal pmf's of *X* and *Y*.
   b. What is the probability that the man's and the woman's dinner cost at most \$15 each?
   c. Are *X* and *Y* independent? Justify your answer.
   d. What is the expected value of the total cost of the dinner for the two people?
   e. Suppose the when a couple opens fortune cookies at the conclusion of the meal, they find the message "You will receive as a refund the difference between the cost of the more expensive and the less expensive meal that you have chosen." How much does the restaurant expect to refund?

   a. Marginal distribution of X

| X | 12 | 15 | 20 |
|---|---|---|---|
| P(X = *x*) | 0.05+0.05+0.1 = 0.2 | 0.05+0.1+0.35 = 0.5 | 0+0.2+0.1 = 0.3 |

   Marginal distribution of Y

| Y | 12 | 15 | 20 |
|---|---|---|---|
| P(Y = *y*) | 0.05 + 0.05 +0 = 0.1 | 0.05 + 0.1 + 0.2 = 0.35 | 0.1 + 0.35 + 0.1 = 0.55 |

   b. P(X ≤ 15, Y ≤ 15) = 0.05 + 0.05 + 0.05 + 0.1 = 0.25
   c. P(X=12)*P(Y=12) = 0.1*0.2 = 0.02 which is not P(X = 12, Y = 12) thus we do not have

independence.

d. $E(X) = 12(0.2) + 15(0.5) + 20(0.3) = 15.9$
$E(Y) = 12(0.1) + 15(0.35) + 20(0.55) = 17.45$
$E(X + Y) = 15.9 + 17.45 = 33.35$

e. This is $E(|X - Y|) = |E(X) - E(Y)| = |15.9 - 17.45| = 1.55$

3. On average a certain intersection results in 3 traffic accidents per month. Let X = the number of accidents at this intersection.
   a. What kind of distribution does X have?
   b. What is the probability that for any given month at this intersection exactly 5 accidents will occur?
   c. What is the probability that for any given month that at least 2 accidents will occur at this intersection?
   d. What is the probability that in a given year that less than 36 accidents will occur at this intersection?

   a. This is a Poisson distribution with $\mu = 3$.
   b. $P(X = 5)$ = dpois(5,3) = 0.1008
   c. $P(X \geq 2) = 1 - P(X \leq 1) = 1 -$ ppois(1,3) = 0.8009
   d. Mean is now 3(12) = 36; $P(X < 36) = P(X \leq 35)$ = ppois(35,36) = 0.4778

4. Consider a uniform density curve defined from x = 1 to x = 8.
   a. What is the height of the "curve"?
   Height will be the density function f(*x*) 1/(max – min)

$$f(x) = \begin{cases} \frac{1}{7}, & 1 \leq x \leq 8 \\ 0, & otherwise \end{cases}$$

   b. What percent of observations fall between x=2 and x=5?
   $P(2 < X < 5) = (1/7)(5) - (1/7)(2) = 3/7$
   c. What percent of observations fall below x = 4?
   $P(X < 4) = (1/7)(4-1) = 3/7$
   d. What percent of observations fall above x = 6?
   $P(X > 6) = 1 - P(X < 6) = 1 - (1/7)(6-1) = 2/7$
   e. What percent of observations equal 7?
   $P(X = 7) = 0$ [Since this is a continuous function the probability at exactly one value is zero]

5. A 12-inch bar that is clamped at both ends is to be subjected to an increasing amount of stress until it snaps. Let $Y$ = the distance from the left end at which the break occurs. Suppose $Y$ had pdf:

$$f(y) = \begin{cases} \left(\frac{1}{24}\right) y \left(1 - \frac{y}{12}\right), & 0 \le y \le 12 \\ 0, & otherwise \end{cases}$$

Compute the following:

a. The cdf of Y

$$F(Y) = \int_0^y \frac{1}{24} u \left(1 - \frac{u}{12}\right) du = \frac{1}{24}\int_0^y u - \frac{u^2}{12} dy = \frac{1}{24}\left[\frac{y^2}{2} - \frac{y^3}{36}\right]$$

F($y$) = 0 for $y \le 0$, F($x$) = 1 for $y \ge 12$

b. P($Y \le 4$), P($Y > 6$) and P($4 \le Y \le 6$).

P($Y \le 4$) = F(4) = $\frac{1}{24}\left[\frac{4^2}{2} - \frac{4^3}{36}\right]$ = 0.2592

P(Y > 6) = 1 – F(6) = 1 - $\frac{1}{24}\left[\frac{6^2}{2} - \frac{6^3}{36}\right\}$ = 1 – 0.5 = 0.5

P( 4 ≤ Y ≤ 6) = F(6) – F(4) = 0.5 – 0.2592 = 0.2408

c. E($Y$), E($Y^2$), and $Var$($Y$)

E(Y) = $\int_0^{12} \frac{1}{24}\left(y^2 - \frac{y^3}{12}\right) = \frac{1}{24}\left(\frac{y^3}{3} - \frac{y^4}{48}\right)\begin{matrix}12\\0\end{matrix} = 6$

E($Y^2$) = $\int_0^{12} \frac{1}{24}\left(y^3 - \frac{y^4}{12}\right) = \frac{1}{24}\left(\frac{y^4}{4} - \frac{y^5}{60}\right)\begin{matrix}12\\0\end{matrix} = 43.2$

Var(Y) = E($Y^2$) – E$(Y)^2$ = 43.2 – 36 = 7.2

d. The probability that the break point occurs more than 2 in. from the expected break point.

P(X < 4 or X > 8) = F(4) + 1 – F(8) = 0.2592 + 1 – 0.74074 = 0.51846

e. The expected length of the shorter segment when the break occurs.

Let Z = the length of the shorter segment.

Then Z(Y) = $\begin{cases} y, if\ 0 \le Y < 6 \\ 12 - y\ , if\ 6 \le Y \le 12 \end{cases}$

E(Z(Y)) = $\int_0^6 \frac{1}{24}\left(y^2 - \frac{y^3}{12}\right) + \int_6^{12} (12 - y)\left(\frac{1}{24}\right)\left(y - \frac{y^2}{12}\right)$ = 1.875+1.875 = 3.75

6. In a certain city, the daily consumption of electric power, in millions of kilowatt-hours, is a random variable $X$ having a gamma distribution with mean μ = 6 and variance = $\sigma^2$ = 12.

a. Find the value of α and β.

$\mu = \alpha\beta$; $\sigma^2 = \alpha\beta^2$

$6 = \alpha\beta \longrightarrow \alpha = 6/\beta$

$12 = \alpha\beta^2$

$12 = 6/\beta \times \beta^2$

$12 = 6\beta$

$\beta = 2$

$\alpha = 6/2 = 3$

b. Find the probability that on any given day the daily power consumption will exceed 12 million kilowatt-hours.

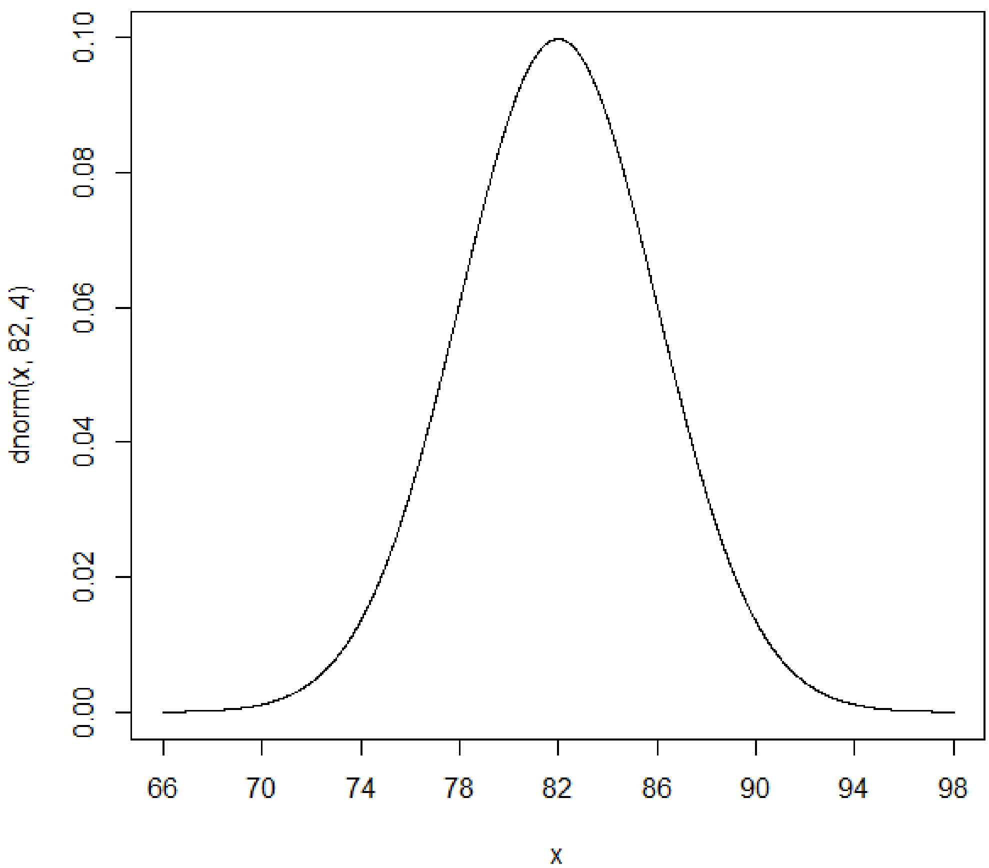
P(X > 12) = 1-pgamma(12,shape = 3, scale = 2) = 0.0619688

7. The length of time for one individual to be served at a cafeteria is a random variable having an exponential distribution with a mean of 4 minutes.
   a. Find the value of λ.
      $\mu = 1/\lambda =$ ¼

   b. What is the probability that a person waits for less than 3 minutes?
      P(X < 3) = pexp(3, ¼) = 0.5267

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

a. Sketch the distribution (This can be drawn by hand)




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

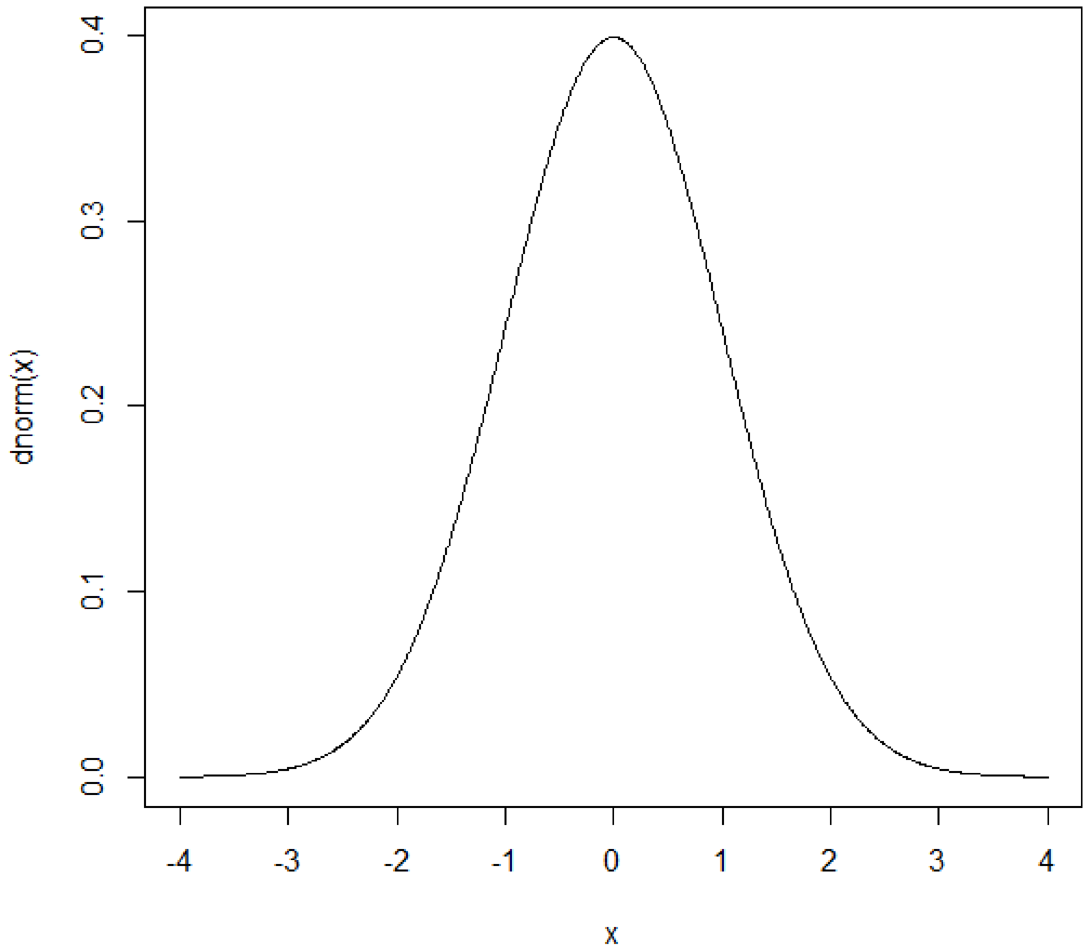
f. Find $x$ such that $P(X < x) = .97725$
   qnorm(0.97725,82,4) = 90

9. 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

10. 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

11. 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 ($n$).

12. True or False? Explain.

a. For a fixed confidence level, when the sample size increases, the length of the confidence interval for a population mean decreases.
True

b. The z score corresponding to a 98 percent confidence level is 1.96.
False, for 98% confidence z = 2.33

c. The best point estimate for the population mean is the sample mean.
True

d. The larger the level of confidence, the shorter the confidence interval.
False

e. The margin of error can be computed from $\pm z^* \cdot \frac{\sigma}{\sqrt{n}}$
True

f. A statement contradicting the claim in the null hypothesis is classified as the power.
False, the statement contradicting the claim in the null hypothesis is the alternative hypothesis.

g. If we want to claim that a population parameter is different from a specified value, this situation can be considered as a one-tailed test.
False, it is two-tail test. Alternative is “not equal to.”

h. In the P-value approach to hypothesis testing, if the P-value is less than a specified significance level, we fail to reject the null hypothesis.
False, if the p-value is less than or equal to the specified level of significance then we would reject the null hypotheis.

i. A 90% confidence interval for a population parameter means that if a large number of confidence intervals were constructed from repeated samples, then on average, 90% of these intervals would contain the true parameter.

True

j. The point estimate of a population parameter is always at the center of the confidence interval for the parameter.

True for means and proportions.

13. A certain beverage company is suspected of under filling its cans of soft drink. The company advertises that its cans contain, on the average, 12 ounces of soda with standard deviation 0.4 ounce. Compute the probability that a random sample of 50 cans produces a sample mean fill of 11.9 ounces or less.

$P(\bar{X} \leq 11.9) = P(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{11.9-12}{\frac{0.4}{\sqrt{50}}}) = P(Z \leq -1.7677) = 0.0384$ from the Z-table

Using R:

```
> pnorm(11.9,12,.4/sqrt(50))
[1] 0.03854994
```

14. A Brinell hardness test involves measuring the diameter of the indentation made when a hardened steel ball is pressed into material under a standard test load. Suppose that the Brinell hardness is determined for each specimen in a sample of size 50, resulting in a sample mean hardness of 64.3 and a sample standard deviation of 6.0. Calculate a 99% confidence interval for the true average Brinell hardness for material specimens of this type.

Point estimate = 64.3
Confidence level = 99%
Critical Value = t(df = 49) = 2.68 (In R: qt(1.99/2,49))
Standard error = $6/\sqrt{50}$ = 0.8485
Margin of error = 0.8485 * 2.68 = 2.274
Confidence interval: (64.3 – 2.274, 64.3 + 2.274) = (62.026, 66.574)
Interpret: We are 99% confident that the Brinell hardness for this type of steel ball is between 62.026 and 66.574.

```
> 64.3+c(-1,1)*qt(1.99/2,49)*6/sqrt(50)
[1] 62.02599 66.57401
```

15. The shear strength of anchor bolts has a standard deviation of 1.30. Assuming that the distribution is normal, how large a sample is needed to determine with a precision of ±0.5 the mean length of the produced anchor bolts to 99% confidence?

σ = 1.30, C = 99%, m = 0.5, z = 2.576

$$n = \left(\frac{z \times \sigma}{m}\right)^2 = \left(\frac{2.576 \times 1.3}{.5}\right)^2 = 44.85$$

We need at least a sample of **45**

16. A journal article reports that a sample of size 5 was used as a basis for calculating a 90% confidence interval for the true population mean(population variance not known). The resulting interval was [15.34, 36.66]. You decide that a confidence level of 99% is more appropriate. What are the limits of the 99% confidence interval?

Point Estimate: (15.34 + 36.66)/2 = 26
Confidence level = 0.99

Critical value = t = qt(1.99/2,4) = 4.604
Margin of error for 90% confidence: (36.66-15.34)/2 = 10.66
Critical value for 90% confidence = qt(1.9/2,4) = 2.132
To find standard error use the formula for the 90% confidence:
$M = t^* \times Se$
$10.66 = 2.132 \times SE$
$SE = 10.66/2.132 = 5$
Margin of error for 99% confidence = $4.604 \times 5 = 23.02$
Confidence interval [26 – 23.02, 26 + 23.02] = [2.98, 49.02]

17. A study of the ability of individuals to walk in a straight line reported the accompanying data on cadence (strides per second for a sample of n = 20 randomly selected men).

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| .95 | .85 | .92 | .95 | .93 | .86 | 1.00 | .92 | .85 | .81 |
| .78 | .93 | .93 | 1.05 | .93 | 1.06 | 1.06 | .96 | .81 | .96 |

a. Find a 99% confidence interval for the mean cadence of the population.
b. Test the hypothesis that the mean cadence for the population is less than 0.97 at the 5% significance level.

I did this in R:

```
> walk=c(.95,.85,.92,.95,.93,.86,1,.92,.85,.81,.78,.93,.93,1.05,.93,1.06,1.06,.96,.81,.96)
```

a.
```
> t.test(walk,conf.level = .99)

        One Sample t-test

data:  walk
t = 51.132, df = 19, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 0.8737164 0.9772836
sample estimates:
mean of x
   0.9255
```

b.
```
> t.test(walk,conf.level=.95,mu=.97,alternative = "less")

        One Sample t-test

data:  walk
t = -2.4585, df = 19, p-value = 0.01186
alternative hypothesis: true mean is less than 0.97
95 percent confidence interval:
      -Inf 0.9567977
sample estimates:
mean of x
   0.9255
```

$H_0$: $\mu = 0.97$ and $H_a$: $\mu < 0.97$
Test statistic = -2.4585
Rejection region is any $t < -1.729$ RH0
P-value = 0.01186
Decision: Since the p-value is less than 0.05, we reject the null hypothesis.
Conclusion: We say that there is significant evidence that the null hypothesis is false and that the mean cadence for men is less than 0.97.

18. Bottles of a popular cola drink are supposed to contain 300 ml of cola. There is some variation from bottle to bottle because the filling machinery is not perfectly precise. The distribution of the contents is normal with standard deviation of 3 ml. A student who suspects that the bottler is under-filling measures the contents of six bottles. The results are:

| | | | | | |
|---|---|---|---|---|---|
| 299.4 | 297.7 | 301.0 | 298.9 | 300.2 | 297.0 |

Is this convincing evidence that the mean contents of cola bottles is less than the advertised 300 ml? Test at the 5% significance level.

$H_0$: μ = 300 and $H_a$: μ < 300

Test statistic = z = $\frac{299.033-300}{3/\sqrt{6}} = -0.7895$ (used z because we know the population standard deviation)

Rejection region is any z < -1.645 FRH0

P-value = P(Z < -0.7895) = 0.2149

Decision: Since the p-value is greater than 0.05, we fail to reject the null hypothesis.

Conclusion: We say that there is not enough evidence that the null hypothesis is false. Thus there is **not** convincing evidence that the mean contents of cola bottle is less than the advertised 300 ml.

19. 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?

Type I error is rejecting the null hypothesis when in fact it is true. So this means that if the coach makes a Type I error that he went for the first down but did not get the first down.

20. 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.

$H_0$: μ = 33.5 and $H_a$: μ ≠ 33.5

Test statistic = t = $\frac{31.6-33.5}{3.4/\sqrt{12}} = -1.9358$

P-value = P(t < -1.9358 or t > 1.9358) = 2 * pt(-1.9358,11) = 0.079 FRH0

Conclusion: there is not enough evidence to conclude that the mean gas mileage of the model of car is different from 33.5

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.

$H_0$: μ = 33.5 and $H_a$: μ < 33.5

Test statistic = t = $\frac{31.6-33.5}{3.4/\sqrt{12}} = -1.9358$

P-value = P(t < -1.9358 ) = pt(-1.9358,11) = 0.0395 RH0

Conclusion: there is enough evidence to conclude that the mean gas mileage of the model of car is significantly less than 33.5

c. Explain why the answers to part a and part b are different.

The probability of getting the observed value “less than” will be smaller because we are only looking at one side.

21. Suppose that prior to conducting the 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 90% confidence interval of width of at most 0.1 for the probability of flipping a head?

For C = 90%, z* = 1.645 (Using table or qnorm(1.9/2)), p* = 0.5, m = 0.1/2 = 0.05

$$n \geq \left(\frac{1.645}{0.05}\right)^2 (0.5)(1-0.5)$$

$$n \geq 270.5543$$

Thus we need at least 271 in the sample.

22. In a sample of 539 households from a certain Midwestern city, it was found that 133 of these households owned at least one firearm. Give a 99% confidence interval for the percentage of families in this city who own firearms.

Point estimate: 133/539 = 0.2468
Confidence level = 99%
Critical value = z = 2.576
Standard Error = $\sqrt{\frac{.2468\times(1-.2468)}{539}}$ = 0.01857
Margin of error = 2.576×0.01857 = 0.0478
Confidence Interval: (0.2468 – 0.0478, 0.2468 + 0.0478) = (0.199, 0.2946)
Interpret: We are 99% confident that the percent of families in this city who own firearms is between 20% and 29.5%.

```
> 0.2468+c(-1,1)*qnorm(1.99/2)*sqrt(0.2468*0.7532/539)
[1] 0.1989645 0.2946355
```

23. 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.

$H_0$: p = 0.8 and $H_a$: p < 0.8 $\hat{p} = \frac{77}{110} = 0.7$
Test statistic = z = $\frac{0.7-0.8}{\sqrt{\frac{.8(1-0.8)}{110}}} = -2.622$
Rejection Region: P(Z < c) =0.1 qnorm(0.1) = -1.28, reject the null hypothesis if the test statistic is less than -1.28 (draw the normal curve) RH0
P-value = P(t < -2.622) = pnorm(-2.622) = 0.0044 RH0 this is less than 0.1.
Conclusion: There is strong evidence that the proportion of people who pray for world peace is significantly less than 80%.

24. A manufacturer of car batteries claims that his batteries will last, on average, 3 years with a variance of 1 year. If 5 of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5, and 4.2 years, construct a 95% confidence interval for the variance $\sigma^2$ and decide if the manufacturer's claim that $\sigma^2 = 1$ is valid. Assume the population of battery lives to be approximately normally distributed.

In R-studio:

```
> batteries = c(1.9,2.4,3,3.5,4.2)
> lcl=4*var(batteries)/qchisq(1.95/2,4)
> ucl=4*var(batteries)/qchisq(0.05/2,4)
> c(lcl,ucl)
[1] 0.2925528 6.7297174
```

The confidence interval for the variance is: (0.2926, 6.7297).

Since $\sigma^2 = 1$ is inside the interval, the manufacturer's claim is valid.