MATH 3339
Statistics for the Sciences
Live Lecture Help

James West
jdwest@uh.edu

University of Houston

Session 15

Office Hours: see schedule in the "Office Hours" channel on Teams
Course webpage: www.casa.uh.edu

Email policy

When you email me you **MUST** include the following

- MATH 3339 Section 20024 and a description of your issue in the **Subject Line**
- Your name and ID# in the **Body**
- Complete sentences, punctuation, and paragraph breaks
- Email messages to the class will be sent to your Exchange account (user@cougarnet.uh.edu)

Using R and RStudio

1. Download R from https://cran.r-project.org/
2. Download RStudio from https://www.rstudio.com/

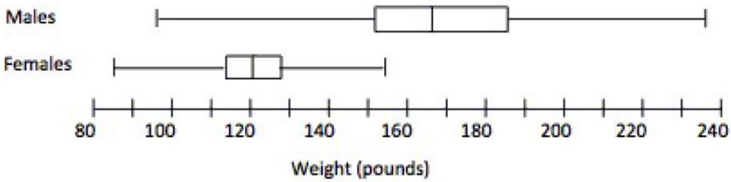
Outline

Updates and Announcements

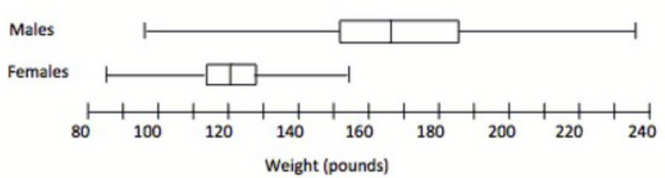
- Final Exam begins 12/06. Check your reservation!
- We will add 5% of Practice Final to Final Exam.
- Course Evaluations are open and available in CASA. These close 12/06.
- FE lasts 120 minutes, covers everything, expect ~ 20 questions

Example 1

The weights of male and female students in a class are summarized in the following boxplots:



Exan



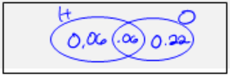
Which of the following is NOT correct?

a) The median weight of the male students is about 166 lbs.

b) The mean weight of the female students is about 120 because of symmetry.

c) The male students have less variability than the female students.

d) About 50% of the male students have weights between 150 and 185 lbs.



Hospital records show that 12% of all patients are admitted for heart disease, 28% are admitted for cancer (oncology) treatment, and 6% receive both coronary and oncology care.

1. What is the probability that a randomly selected patient is admitted for coronary care, oncology or both? (Note that heart disease is a coronary care issue.)

Let H denote patient treated for Heartdisease

Let O denote Oncology care

Want $P(H \cup O)$, we are given $P(H) = 0.12$, $P(O) = 0.28$, $P(H \cap O) = 0.06$

$$P(H \cup O) = P(H) + P(O) - P(H \cap O)$$
$$= 0.12 + 0.28 - 0.06 = \boxed{0.34}$$

Example 2

Hospital records show that 12% of all patients are admitted for heart disease, 28% are admitted for cancer (oncology) treatment, and 6% receive both coronary and oncology care.

2. What is the probability that a randomly selected patient is admitted for coronary care, given that they are a cancer patient?

$$P(H \mid O) = \frac{P(H \cap O)}{P(O)}$$

$$= \frac{0.06}{0.028}$$

$$= \frac{6}{28} = \frac{3}{14} = \boxed{0.2142}$$

Example 2

Hospital records show that 12% of all patients are admitted for heart disease, 28% are admitted for cancer (oncology) treatment, and 6% receive both coronary and oncology care.

3. Are patients that are admitted for coronary care independent of patients that are admitted for cancer?

$$P(H \cap O) = P(H) \cdot P(O) \ ?$$

or is $P(H \mid O) = P(H)$?

No ! $P(H) = 0.12$

$P(H \mid O) = 0.2142$

Example 3

A random variable X has a probability distribution as follows:

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

1. What is the value of k?

$$2k + 3k + 5k + 3k + 4k = 1$$

$$k(17) = 1 \Rightarrow k = \frac{1}{17}$$

2. What is $P(X < 2)$?

$$P(X < 2) = P(X = 0) + P(X = 1) = \frac{2}{17} + \frac{3}{17} = \boxed{\frac{5}{17}}$$

Example 4

In testing a certain kind of missile, target accuracy is measured by the average distance X (from the target) at which the missile explodes. The distance X is measured in miles and the distribution of X is given by:

X	0	10	50	100
P(X)	$\frac{1}{14}$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{2}$

Find the mean and variance for the target accuracy.

$$\mu = E[X] = \sum_{x} x \cdot P(X=x)$$

$$= 0 \cdot \frac{1}{14} + 10 \cdot \frac{1}{7} + 50 \cdot \frac{2}{7} + 100 \cdot \frac{1}{2} = 65.7429$$

$$\sigma^2 = Var[X] = E[X^2] - E[X]^2$$

Example 5

In testing a new drug, researchers found that 5% of all patients using it will have a mild side effect. A random sample of 7 patients using the drug is selected. Find the probability that:

$X \sim Binom(7, 0.05)$

1. None will have this mild side effect.

$$P(X=0) = \binom{7}{0} \cdot p^0 \cdot (1-p)^7 = dbinom(0, 7, 0.05)$$
$$= \boxed{0.69834}$$

2. Exactly 2 patients will have this mild side effect.

$$P(X=2) = \binom{7}{2} \cdot p^2 \cdot (1-p)^5$$
$$= dbinom(2, 7, 0.05) = 0.0406$$

Example 5

In testing a new drug, researchers found that 5% of all patients using it will have a mild side effect. A random sample of 7 patients using the drug is selected. Find the probability that:

3. At least one will have this mild side effect.

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$$
$$= 1 - \text{dbinom}(0, 7, 0.05)$$
$$= 0.30166$$

4. What is the expected value and variance of the number of patients that will have this mild side effect?

$$E[X] = np = 7 \cdot \frac{1}{20} = \frac{7}{20} = 0.35$$
$$Var[X] = np(1-p) = \frac{7}{20} \cdot \frac{19}{20} = \frac{133}{400} = 0.332$$

Example 6

Let X be the amount of time (in hours) the wait is to get a table at a restaurant. Suppose the cdf is represented by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}x^2 & 0 \le x \le 2 \\ 1 & x > 2 \end{cases}$$

1. Find $P(X \le 1.5)$ $F(1.5) = \frac{1}{4} \cdot \left(\frac{3}{2}\right)^2 = \frac{9}{16} = 0.5625$
2. Find $P(X \ge 1)$ $= 1 - P(X \le 1) = 1 - F(1)\ 1 - \frac{1}{4} = \frac{3}{4}$
3. Find $P(1 \le X \le 1.5)$ $= F(1.5) - F(1) = 0.3125$
4. Find the density function $f(x)$. $f(x) = F'(x) = \frac{d}{dx}\left(\frac{1}{4}x^2\right) = \frac{1}{2}x$

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Example 7

Let Z be the standard normal random variable. Calculate the following.

1. $P(|Z| \leq 2.4)$ $= P(-2.4 \leq Z \leq 2.4) = \Phi(2.4) - \Phi(-2.4)$
   $= \text{pnorm}(2.4) - \text{pnorm}(-2.4)$

2. $P(Z \leq -1.9)$ $= \text{pnorm}(-1.9)$

3. Find c such that $P(Z \geq c) = 0.98$

   $P(Z \leq c) = 1 - 0.98 = 0.02$

   $c = \text{qnorm}(0.02)$

Example 8

The weights of individual bolts produced at a manufacturing plant, X, is normally distributed. If the mean weight of the bolts is 9 grams and the standard deviation is 3.2 grams, find:

1. $P(X \leq 10.5)$ $= \text{pnorm}(10.5, 9, 3.2)$

   $= \text{pnorm}\left(\frac{10.5-9}{3.2}\right)$

2. $P(X \geq 7.1)$ $= 1 - P(X \leq 7.1)$

   $= 1 - \text{pnorm}(7.1, 9, 3.2)$

3. The value of x such that $P(X \leq x) = 0.93$

   $x = \text{qnorm}(0.93, 9, 3.2)$

Example 9

$$X \sim N(9, 3.2), \quad n = 62$$

Using the information from example 8, if we randomly sample 62 weights and determine the sample mean, what is the probability that

1. The sample mean is less than 8.5 grams. $\bar{X} \sim N(9, 3.2/\sqrt{62})$

$$P(\bar{X} \leq 8.5) = \text{pnorm}(8.5, 9, 3.2/\sqrt{62})$$

2. The sample mean is at least 10.5 grams.

$$P(\bar{X} \geq 10.5) = 1 - P(\bar{X} \leq 10.5)$$
$$= 1 - \text{pnorm}(10.5, 9, 3.2/\sqrt{62})$$

Example 10

1. A simple random sample of 100 8th graders at a large suburban middle school indicated that 86% of them are involved with some type of after school activity. Find the 98% confidence interval that estimates the proportion of them that are involved in an after school activity.

   1 proportion ⇒ use z interval

2. An SRS of 24 students at UH gave an average height of 6.1 feet and a standard deviation of .3 feet. Construct a 90% confidence interval for the mean height of students at UH.

   t interval for means

3. The average height of students at UH from an SRS of 17 students gave a standard deviation of 2.9 feet. Construct a 95% confidence interval for the standard deviation of the height of students at UH. Assume normality for the data.

   Chi-squared interval for SD

Example 11

A 98% confidence interval for the mean of a population is to be constructed and must be accurate to within 0.3 unit. A preliminary sample standard deviation is 1.7. The smallest sample size n that provides the desired accuracy is

$ME = 0.3$

$$n > \left(\frac{z^* \cdot 1.7}{0.3}\right)^2$$

$$z^* = qnorm\left(\frac{1.98}{2}\right)$$

174 works

Example 12

In a hypothesis test, if the computed P-value is less than 0.001, there is very strong evidence to

a) retest with a different sample.
b) accept the null hypothesis
c) fail to reject the null hypothesis.
d) reject the null hypothesis.

Example 13

The one-sample t statistic for a test of $H_0 : \mu = 12$ vs. $H_a : \mu < 12$ based on $n = 174$ observations has the test statistic value of $t = -1.58$. What is the p-value for this test?

p-value = pt(-1.58, 173)

```
> pt(-1.58,173)
[1] 0.05796675
```

What is $pt(-1.58, 173) = P(T \le -1.58)$

$1 - pt(t) = 1 - P(T \le t) = P(T \ge t)$

Example 14

Identify the most appropriate test to use for the following situation: A national computer retailer believes that the average sales are greater for salespersons with a college degree. A random sample of 14 salespersons with a degree had an average weekly sale of $3542 last year, while 17 salespersons without a college degree averaged $3301 in weekly sales. The standard deviations were $468 and $642 respectively. Is there evidence to support the retailer's belief?

a) One sample t test

b) Matched pairs

c) Two sample t test

d) Two sample p test

College	Non-college
$n_1 = 14$	$n_2 = 17$
$\bar{X}_1 = 3542$	$\bar{X}_2 = 3301$
$s_1 = 468$	$s_2 = 642$

$H_0: \mu_1 = \mu_2$, $H_a: \mu_1 > \mu_2$

Complete the Test

Test Stat:

$$t_1 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

```
> (3542-3301)/sqrt(468^2/14+642^2/17)
[1] 1.206668
```

p-value: $P(T \geq 1.207) = 1 - P(T \leq 1.207)$

```
> 1-pt(.Last.value,13)
[1] 0.12453
```

FTR H_0

Example 15

Data for gas mileage (in mpg) for different vehicles was entered into a software package and part of the ANOVA table is shown below:

Source	DF	SS	MS
Vehicle	2	440	220.00
Error	17	318	18.71
Total	19	758	

1. Determine the value of the test statistic F to complete the table.

$$F = \frac{MSTr}{MSE} = \frac{220}{18.71} = 11.7584$$

2. Determine the p-value.

$$\text{p-value} = P(F \geq 11.7584)$$

```
> 1-pf(220/18.71,2,17)
[1] 0.0006221769
```

RH_0

Example 16

The community hospital is studying its distribution of patients. A random sample of 330 patients presently in the hospital gave the following information:

Type of Patient	Old Rate	Number of Occurrences
Maternity Ward	20%	77
Cardiac Ward	32%	95
Burn Ward	10%	29
Children's Ward	15%	53
All Other Wards	23%	76

Test the claim at the 5% significance level that the distribution of patients in these wards has not changed.

Chi-sq goodness of fit test

Example 16 Work

```
> occur=scan()
1: 77
2: 95
3: 29
4: 53
5: 76
6:
Read 5 items
> sum(occur)
[1] 330
> props=c(20,32,10,15,23)/100
> sum(props)
[1] 1
> chisq.test(occur,p=props,correct = F)

        Chi-squared test for given probabilities

data:  occur
X-squared = 3.6298, df = 4, p-value =
0.4584
```

Example 17

The following two-way table describes the preferences in movies and fast food restaurants for a random sample of 100 people.

	McDonald's	Taco Bell	Wendy's
Iron Man	20	12	8
Dispicable Me	12	7	9
Harry Potter	6	14	12

1. What percent of the Dispicable Me lovers also like McDonald's?

2. What percent likes Harry Poter if they also like Wendy's?

Example 18

Below is the computer output for the appraised value (in thousands of dollars) and number of rooms for 20 houses in East Meadow, New York.

Predictor	Coef	Stdev	t-ratio
Constant	74.80	19.04	3.93
Rooms	19.718	2.631	7.49

$S = 29.05$ R-sq = 43.8% R-sq (adj) = 43.0%

1. What is the regression equation?

Example 18

Below is the computer output for the appraised value (in thousands of dollars) and number of rooms for 20 houses in East Meadow, New York.

Predictor	Coef	Stdev	t-ratio
Constant	74.80	19.04	3.93
Rooms	19.718	2.631	7.49

$S = 29.05$ R-sq = 43.8% R-sq (adj) = 43.0%

2. Predict the price of a 10 room house (in thousands of dollars).

Example 18

Below is the computer output for the appraised value (in thousands of dollars) and number of rooms for 20 houses in East Meadow, New York.

Predictor	Coef	Stdev	t-ratio
Constant	74.80	19.04	3.93
Rooms	19.718	2.631	7.49

$S = 29.05$ R-sq = 43.8% R-sq (adj) = 43.0%

3. Calculate the 95% confidence interval of the slope of the regression line for all homes.

Use the information provided to test whether there is a significant relationship between the price of a house and the number of rooms at the 5% level.

Example 19

The following data are for intelligence-test (IT) scores, grade-point averages (GPA), and reading rates (RR) of 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

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?

The decline of salmon fisheries along the Columbia River in Oregon has caused great concern among commercial and recreational fishermen. The paper 'Feeding of Predaceous Fishes on Out-Migrating Juvenile Salmonids in John Day Reservoir, Columbia River' (Trans. Amer. Fisheries Soc. (1991: 405-420) gave the accompanying data on 10 values for the data sets where y = maximum size of salmonids consumed by a northern squaw fish (the most abundant salmonid predator) and x = squawfish length, both in mm. Here is the computer software printout of the summary:

Coefficients:				
	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	−91.030	16.703	−5.450	0.000
Length	0.706	0.042	16.759	0.000

Using this information, compute a 95% confidence interval for the slope.

$$0.706 + c(-1,1) \cdot qt(1.95/2, 8) \cdot 0.042$$

Using R and RStudio

1. Download R from https://cran.r-project.org/
2. Download RStudio from https://www.rstudio.com/