FREC 408

Group Exercise 9

On page 660 there is a Statistics in the Real World problem on roaches.

- This is a study to see if the navigation of cockroaches is random or not.
- The researcher hypothesized that cockroaches do follow trails, much like bees, ants, and termites, using chemical trails.
- She used a chemical trail with pheromones as the main treatment (extract from cockroach feces), but also included a control trail using methanol (TRAIL).
- **Today, we will focus on a One-Way ANOVA of the Extract group to see if there are differences by Roach type on the trail with the pheromones. On Monday I will show the two-factor ANOVA of the data.**

Roach ANOVA

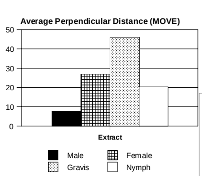
- The researcher released German cockroaches of different age, sex, and reproductive stage to see if these factors influenced trail following ability
  - **Factor** = TYPE:
  - the levels = Female; Gravid; Male; and Nymph
- She measured the movement pattern of the cockroaches and calculated an average perpendicular distance
  - **Response Variable** = MOVE

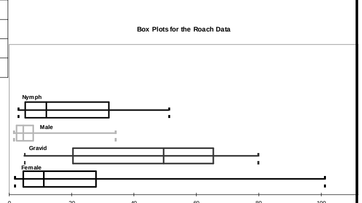
Roach Data

	Female	*Gravid*	*Male*	*Nymph*	*TOTAL*
Mean	21.07	44.03	7.38	18.73	22.80
Standard Error	5.84	5.55	1.93	3.56	2.67
Median	11.15	49.60	4.50	11.95	13.25
Mode	#N/A	#N/A	2.40	#N/A	2.40
Standard Deviation	26.13	24.84	8.61	15.92	23.89
Sample Variance	682.79	616.86	74.12	253.51	570.66
Kurtosis	4.23	-1.42	4.73	-0.47	0.89
Skewness	2.09	-0.09	2.28	0.87	1.31
Range	99.50	75.10	32.50	48.40	99.70
Minimum	1.70	4.80	1.50	2.80	1.50
Maximum	101.20	79.90	34.00	51.20	101.20
Sum	421.40	880.60	147.50	374.60	1824.10
Count	20	20	20	20	80

A lot of variation between the means, from 7.38 for males to 44.03 for Gravids.
There is also a lot of variation for all groups except the Males.

Graphs to help describe the data





ANOVA with missing values

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Female	20	421.4	21.07	682.79
Gravid	20	880.6	44.03	616.86
Male	20	147.5	7.375	74.12
Nymph	20	374.6	18.73	253.51

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	14164.0		4721.342		0.000002	4.050
Within Groups		76				
Total	45082.5	79				

$R^2 = 14164/45082.5 = .314$

ANOVA

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Female	20	421.4	21.07	682.79		
Gravid	20	880.6	44.03	616.86		
Male	20	147.5	7.375	74.12		
Nymph	20	374.6	18.73	253.51		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	14164	**3**	4721.342	**11.605**	0.000002	4.050
Within Groups	**30918.5**	76	**406.822**			
Total	45082.5	79				

Conduct a Test to see if there is a mean difference in MOVE by the levels for Type (Male, Female, Nymph, and Gravid). Use an F-test with $\alpha = .01$.

Null Hypothesis	$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
Alternative Hypothesis	H_a: At least one mean is different
Assumptions of Test	Large sample, equal variances, normal distribution
Test Statistic	$F^* = 11.605$
Rejection Region	$F_{.01,\,3 \text{ and } 76 \text{ d.f.}} = 4.05$
Comparison of Test Statistics with Rejection Region	$F^* > F_{.01,\,3 \text{ and } 76 \text{ d.f.}}$ $11.605 > 4.05$ We can reject $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ p-value < .0001

Rainfall Regression.

- An article in Geography (July 1980) used regression to predict average annual rainfall levels in California. Data on the following variables were collected for 30 meteorological weather stations scattered throughout California. For the group work we will focus on a bi-variate regression of Annual Percip on Latitude. You will have the option of examining all the variables for this problem for the last assignment

Rainfall Regression

- **Annual Percip DEPENDENT VARIABLE: Annual Precipitation in inches**
- **Altitude** The altitude of the station in feet
- **Latitude** The latitude of the station in degrees
- **Distance** Distance from the coast in miles
- Facing **I made this into a dummy variable. Stations on the Westward facing slopes of the California mountains were coded as 1, whereas stations on the leeward side were coded as 0**

Briefly Describe

	Annual Percip	*Latitude*
Mean	19.81	37.03
Standard Error	3.03	0.49
Median	15.35	36.70
Mode	18.20	33.80
Standard Deviation	16.62	2.67
Sample Variance	276.26	7.11
Kurtosis	3.05	-1.09
Skewness	1.70	0.23
Range	73.21	9.20
Minimum	1.66	32.70
Maximum	74.87	41.90
Sum	594.22	1110.8
Count	30	30

Correlation Matrix

	Annual Percip	*Altitude*	*Latitude*	*Distance*	*Facing*
Annual Percip	1.000				
Altitude	0.302	1.000			
Latitude	0.577	0.231	1.000		
Distance	-0.210	0.574	0.161	1.000	
Facing	0.598	0.050	-0.011	-0.490	1.000

As Latitude increase, average annual rainfall increases. The correlation = .577 is moderate, positive

Stations on the West side (when Facing =1) have higher average rainfall. The correlation is .598.

Stations on the west side (whne facing =1) have less distance from the ocean. The correlation is -.490.

Regression of Average Annual Precipitation on Latitude

SUMMARY OUTPUT					
Regression Statistics					
Multiple R	0.577				
R Square	0.333				
Adjusted R Square	0.309				
Standard Error	13.819				
Observations	30				
ANOVA					
	df	*SS*	*MS*	*F*	*Sig F*
Regression	1	2664.887	2664.887	13.956	0.001
Residual	28	5346.766	190.956		
Total	29	8011.654			
	Coeff	*Std Error*	*t Stat*	*P-value*	
Intercept	-113.303	35.721	-3.172	0.004	
Latitude	3.595	0.962	3.736	0.001	

Questions

- Verify that R^2 in a bivariate regression is simply the correlation (r) squared. Interpret R^2 for this model
- Run the hypothesis test to see if there is evidence that the coefficient for Latitude is different from zero.
- What does the model predict for annual precipitation when the latitude is 38 degrees?

Questions

- Verify that R^2 in a bivariate regression is simply the correlation (r) squared. Interpret R^2 for this model
  - r = .577
  - r^2 = .5772 = .3329
  - R^2 = .333

What does the model predict for annual precipitation when the latitude is 38 degrees?

- **Est Y = -113.3 +3.595(38)**
- **Est Y = -113 + 136.61**
- **Est Y = 23.3 inches**

Run the hypothesis test to see if there is evidence that the coefficient for Latitude is different from zero.

Null Hypothesis	**H_0: $\beta_1 = 0$**
Alternative Hypothesis	**H_a: $\beta_1 \neq 0$**
Assumptions of Test	**Large sample, equal variances, normal distribution**
Test Statistic	**t* = 3.736**
Rejection Region	**$t_{.05/2, 29\ d.f.}$ = 2.045**
Comparison of Test Statistics with Rejection Region	**t* > $t_{.05/2, 29\ d.f.}$ 3.736 > 2.045 We can reject H_0: $\beta_1 = 0$ p-value < .002**