

1. An experiment is planned to evaluate the mean number of song birds in a certain tropical forest. The researchers will sample several  $1km^2$  plots. In each plot, they will count the number of bird species present through song recording. They will test the null hypothesis that the mean number of bird species per  $km^2$  equals some fixed value. With their previous experience, they already have a good idea of the variability in species number. Therefore, they made the power curve below.
- (a)  $\mu_0 = 26$
- (b)  $\alpha = 0.10$
- (c) Approximately  $28.7 - 26 = 2.7$  birds
- (d) The dotted curve.
2. It is now well known that mothers can manipulate sex ratios at birth in many species, and researchers wanted to know if fathers can also influence offspring sex ratio. A study was conducted on red deers (*Cervus elaphus*). Sperm from 15 wild males was collected. 360 hinds (females) were kept under similar environmental conditions and provided with unlimited food supply. Each hind was artificially inseminated once. Sperm quality was measured as the percentage of morphologically normal spermatozoa. The number of sons and daughters each male had was counted in order to obtain the percentage of male offspring for each male. The data is presented here. Calculations yield  $\bar{x} = 79.4$ ,  $\bar{y} = 49.27$ ,  $\sum(x_i - \bar{x})^2 = 1105.6$ ,  $\sum(y_i - \bar{y})^2 = 2916.93$ , and  $\sum(x_i - \bar{x})(y_i - \bar{y}) = 1192.4$ . Also,  $SSErr = 1630.92$ .

Male	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sperm Quality ( $x$ )	65	72	68	73	71	77	93	77	87	85	80	83	87	78	95
Prop. of male offsprings ( $y$ )	25	33	38	42	43	41	44	42	49	57	56	58	70	74	69

(a) **Slope:**  $\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{1192.4}{1105.6} = \mathbf{1.0785}$

**Intercept:**  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x} = 49.27 - 1.0785 \times 79.4 = \mathbf{-36.3636}$

**Least Squares Regression Line:** (Proportion of male offsprings) =  $-36.3636 + 1.0785 \times$ (Sperm Quality)

- (b) We are testing the hypothesis  $H_o : \beta_1 = 0$  vs.  $H_a : \beta_1 \neq 0$

**Method 1:**

$$t = \frac{\hat{\beta}_1 - \beta_1}{\left[ \frac{\sqrt{\frac{SSErr}{n-2}}}{\sqrt{\sum(x_i - \bar{x})^2}} \right]} = \frac{1.0785 - 0}{\left[ \frac{\sqrt{\frac{1630.92}{15-2}}}{\sqrt{1105.6}} \right]} = \mathbf{3.20}$$

According to the t-table with  $n - 2 = 13$  degrees of freedom,  $0.002 < p < 0.01$ . Therefore we **reject the null hypothesis** with high confidence. We conclude that there is evidence to support the alternative hypothesis that the slope is significantly different than 0.

### Method 2:

Recall that the values of an ANOVA table can be calculated as:

Source	df	SS	MS	F
Treatment	$k - 1$	$\sum n_i(\bar{x}_i - \bar{x})^2$	$SSTr \div (k - 1)$	$MSTr \div MSErr$
Error	$n - k$	$\sum (n_i - 1)s_i^2$	$SSErr \div (n - k)$	
Total	$n - 1$	$\sum \sum (x_{ij} - \bar{x})^2$		

where  $k$  is the number of treatments and  $n$  is the total sample size. Note that  $SSTot = SSTr + SSErr$ . Therefore:

Source	df	SS	MS	F	p-value
Treatment	1	1286	1286	10.25	$0.005 < p < 0.01$
Error	13	1630.92	125.45		
Total	14	2916.93			

Again, our p-value is significantly small. Therefore we can reject the null hypothesis.

$$(c) (\hat{\beta}_0 + \hat{\beta}_1 x^*) \pm t_{\frac{\alpha}{2}, df=n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} = (-36.3636 + 1.0785 \times 65) \pm 2.160 \times 11.201 \times \sqrt{1 + \frac{1}{15} + \frac{(65 - 79.4)^2}{1105.6}} = \mathbf{(6.643, 60.834)}$$

3. Individuals from the fish species *Telmatherina sarasinorum* are of 3 different colors: females are grey, whereas males are either blue or yellow. It is hypothesized that fish with different colors may prefer to stay in different environments. Two types of environments were surveyed: shallow beach sites (yellow environment), and deeper sites with overhanging roots (blue environment). Many sites from each type were surveyed the same day, and at the same hour of the day, and fish from each color were counted. Below are the data.

Environment	Fish Color			Total
	blue	yellow	grey	
blue	14	22	36	72
yellow	25	12	31	68
Total	39	34	67	140

- (a) The degrees of freedom are calculated by taking  $(\text{number of rows} - 1) \times (\text{number of columns} - 1) = (2 - 1) \times (3 - 1) = 2$ . Therefore given a  $\chi^2$  value of 6.3077,  $0.025 <$

$p < 0.05$ . There is **moderate evidence** that the two environments tested (blue and yellow) do not have an equal number of fish colors (blue, yellow, and grey) living there.

(b) **Observed:** 36

**Expected:**  $\frac{(\text{rowtotal}) \times (\text{columntotal})}{\text{grandtotal}} = \frac{72 \times 67}{140} = 34.45$

Therefore the contribution of the cell “grey fish in blue environment” to the  $\chi^2$  value is  $\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = \frac{(36 - 34.45)^2}{34.45} = \mathbf{0.0698}$ .

(c) We must assume that each cell is independent of the others. This can be determined based on the description of the collection of the data. Use of the chi-square test must also be validated by verifying that the **expected counts** are all  $\geq 5$ .

Environment	Fish Color		
	blue	yellow	grey
blue	20.06	17.49	34.46
yellow	18.94	16.51	32.54

4. The following data represent concentration of a chemical after 5 specific treatments, with 15 data points per treatment. The raw data is shown, along with the data transformed by taking logarithms. The analysis of variance on the raw data gives an F value of 13.1, and an ANOVA on the log-transformed data yields an F value of 34.7.

We prefer the F test performed on the log-transformed data. The assumptions for normality of the residuals (normal quantile plot is a relatively straight line) and equality of variances (spread for each treatment is relatively equal) appear to be better met for the log-transformed data.

5. The following data represent the survival times (hours) of creatures created by a mad scientist using four different reanimation techniques. Seven (7) creatures were animated using each technique. Means and sample standard deviations from each method are given here.

Technique	Mean	Sample SD
A	1.1	0.84
B	2.6	1.30
C	6.8	1.15
D	10.3	1.26

(a) Refer to question 2 part (b) for a reference on how to calculate the cells in the ANOVA table. Given an initial  $MSTr = 121.66$ , the table looks like:

Source	df	SS	MS	F	p-value
Treatment	3	364.98	121.66	91.75	< 0.001
Error	24	31.83	1.326		
Total	27	396.81			

**NOTE:** If you used the given initial value  $MSTr = 92.17$ :

Source	df	SS	MS	F	p-value
Treatment	3	276.5	92.17	69.5	< 0.001
Error	24	31.83	1.326		
Total	27	308.33			

Therefore we reject the null hypothesis at the 0.01 significance level.

- (b) i. **Fisher's LSD:**  $T_{n-k, \frac{\alpha}{2}} \sqrt{\frac{2}{n} \times MSErr} = 2.797 \times \sqrt{\frac{2}{7} \times 1.326} = 1.721$
- ii. **Bonferroni:**  $T_{n-k, \frac{\alpha}{2m}} \sqrt{(\frac{1}{n} + \frac{1}{n}) \times MSErr} = 3.4 \times \sqrt{\frac{2}{7} \times 1.326} = 2.093$
- iii. **Tukey's QD:**  $Q_{k, n-k, \alpha} \sqrt{\frac{1}{n} \times MSErr} = 4.91 \times \sqrt{\frac{1}{7} \times 1.326} = 2.137$

Difference	LSD	Bonf	QD
$\mu_A - \mu_B = 1.5$	no	no	no
$\mu_D - \mu_A = 9.2$	yes	yes	yes
$\mu_C - \mu_B = 4.2$	yes	yes	yes

6. It was hypothesized that fertilizers may affect tree bark pH in agriculture areas. To assess this hypothesis, 10km by 10km regions were randomly selected in Wisconsin, and 20 plots were randomly selected in each region: 10 plots in forest areas and 10 plots in agriculture areas. In each plot, red oak trees were surveyed and their bark pH was measured. Measures from different red oak trees were averaged to yield a single measure for each plot. The data from all plots were used in a two independent sample t-test (assuming equal variances).

- (a) We are told that a two independent sample t-test was used. Therefore our t-statistic has  $n_1 + n_2 - 2$  degrees of freedom. We are given  $df = 78$ , therefore our original  $n$  must be 80. Since 20 plots in each region were chosen,  $\frac{80}{20} = 4$  regions.
- (b) The necessary assumptions are independence, normality, and equal variance. The assumption of independence is not well met due to nesting. According to the two normal quantile plots, the residuals for both the forest and agriculture samples follow a relatively straight line. Therefore the assumption of normality of the error terms is satisfied. The equal variance assumption is also well met.

## 7. True/False Questions.

- (a) False
- (b) False
- (c) False
- (d) True
- (e) True
- (f) False
- (g) False
- (h) True

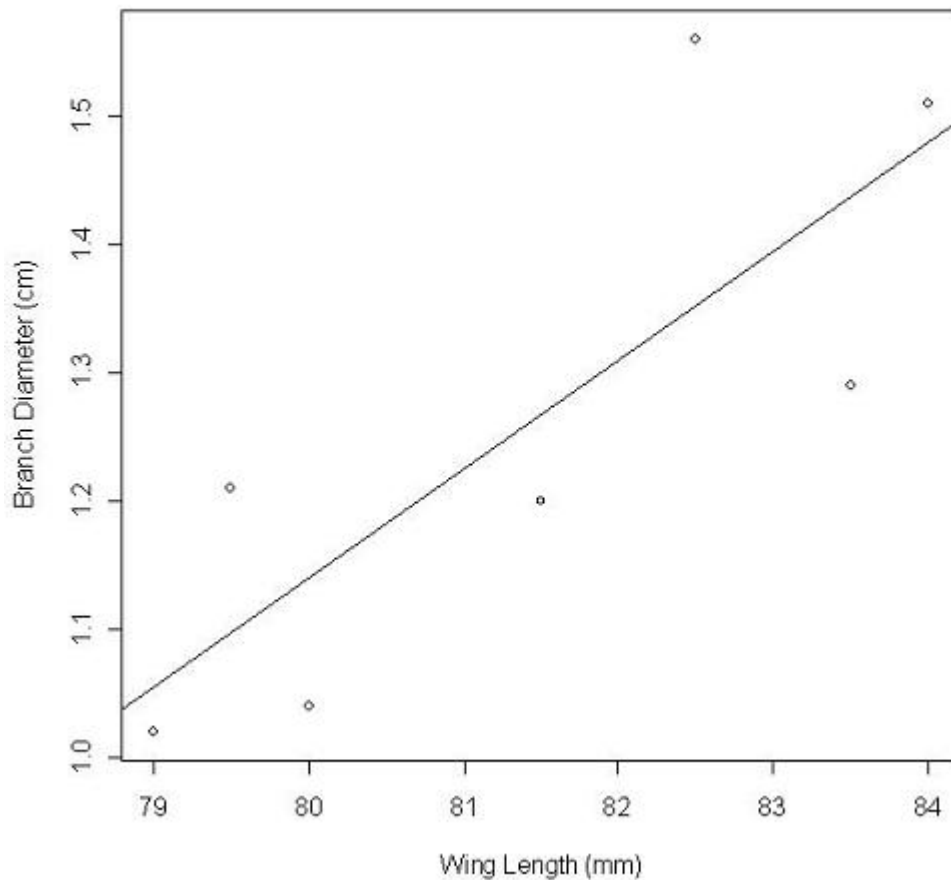
8. In a study of the tufted titmouse (*Parus bicolor*), an ecologist captured seven male birds, measured their wing lengths, and then marked and released them. During the ensuing winter, he repeatedly observed the birds as they foraged for insects and seeds on tree branches, and noted the diameters of the branches. The average branch diameter for each bird is tabulated below.

Bird	1	2	3	4	5	6	7
Wing Length (mm)	79.0	80.0	81.5	84.0	79.5	82.5	83.5
Branch Diameter (cm)	1.02	1.04	1.20	1.51	1.21	1.56	1.29

- (a) There does appear to be a relatively strong, positive, linear relationship between wing length and branch diameter. See part (b) for graph.
- (b) The equation of the regression line is

$$(\text{Branch Diameter in cm}) = -5.660 + 0.085 \times (\text{Wing length in mm}).$$

The scatterplot and corresponding regression line for the data can be seen below.

**Wing Length versus Branch Diameter in study of tufted titmouse**

(c) Using commands in R, the mean and standard deviations were found to be:

	Mean	Standard Deviation
Wing Length (mm)	81.42857	1.988060
Branch Diameter (cm)	1.261429	0.2103512

The correlation coefficient is further calculated to be:

$$r = \frac{1}{n-1} \sum \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right) = \left( \frac{1}{6} \right) 4.820084 = 0.8034.$$

This agrees with our initial observation in part (a). To verify the coefficients found in part (b):

$$\text{Slope: } \hat{\beta}_1 = r \frac{s_y}{s_x} = 0.8034 \frac{0.2103512}{1.988060} = \mathbf{0.08500557}$$

$$\text{Intercept: } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x} = 1.261429 - 0.08500557 \times 81.42857 = \mathbf{-5.660454}$$

**R Code:**

```
wing = c(79.0,80.0,81.5,84.0,79.5,82.5,83.5)
branch = c(1.02,1.04,1.20,1.51,1.21,1.56,1.29)
```

```
mw = mean(wing)
sw = sd(wing)
mb = mean(branch)
sb = sd(branch)
```

```
r = cor(wing,branch)
c(mw,sw,mb,sb,r)
```

```
bird.lm = lm(branch ~ wing)
coef(bird.lm)
```

```
b1 = r*sb/sw
b1
b0 = mb - b1*mw
b0
```