



# STAT021 Statistical Methods II

---

## Lecture 7 ANOVA Multiple Comparisons

---

Lu Chen  
Swarthmore College  
9/25/2018

# Review

---

- ▶ ASSESS model
  - $F$  distribution  $F(df1, df2)$
  - ANOVA  $F$  test
  - One-way ANOVA table
- ▶ ASSESS error
  - Check assumptions  $\epsilon \stackrel{iid}{\sim} N(0, \sigma)$
- ▶ Deal with violation in assumptions
  - Re-CHOOSE, re-FIT and re-ASSESS
  - *Natural logarithm transformation of right skewed data.*

# Outline

---

- ▶ ASSESS:  $R^2$  statistic
  - Comparing  $R^2$  to  $F$  statistic
- ▶ Multiple pairwise comparisons
  - Problem: inflated type I error
  - Fisher's Least Significant Difference (LSD)
  - Bonferroni correction
  - Tukey's Honest Significant Difference (HSD)
  - Compare the three methods
- ▶ Example 1: ♥ Valentine's Day ♥
- ▶ Example 2: Genome-Wide Association Study (GWAS)

# Fruit fly example: Five groups

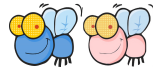
---

Would an increase in sexual activity reduce the life spans of male fruit flies? The researchers designed an experiment to answer this question. They randomly assigned 125 male fruit flies to one of the following five groups:

- ▶ **8virgin**: Each male fruit fly was assigned to live with 8 virgin female fruit flies.



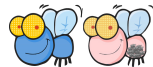
- ▶ **1virgin**: Each male fruit fly was assigned to live with 1 virgin female fruit fly.



- ▶ **8pregnant**: Each male fruit fly was assigned to live with 8 pregnant fruit flies.



- ▶ **1pregnant**: Each male fruit fly was assigned to live with 1 pregnant fruit fly.



- ▶ **None**: Each male fruit fly lived alone.



# Fruit fly example

**Model:** Denote *Longevity* as  $Y$ ,

$$Y = \mu + \alpha_k + \epsilon, \text{ where } k = 1, 2, 3, 4, 5 \text{ and } \epsilon \stackrel{iid}{\sim} N(0, \sigma)$$

**Hypotheses:**  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ ;  $H_a$ : at least one  $\alpha \neq 0$ .

## ANOVA in R

```
flymodel <- aov(Longevity ~ Treatment, data=fly)
summary(flymodel)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment      4  11939   2984.8    13.61 3.52e-09 ***
## Residuals    120   26314    219.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F = \frac{MSG_{\text{Group}}}{MSE} = \frac{2984.8}{219.3} = 13.61, P = P(F_{4,120} > 13.61) = 3.52 \times 10^{-9} < 0.05$$

# ASSESS: $R$ -squared statistic

**Coefficient of determination  $R^2$**  is the fraction of the total variability that is explained by the model.

```
summary(flymodel)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment      4  11939   2984.8    13.61 3.52e-09 ***
## Residuals    120  26314    219.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$R^2 = \frac{SS_{Group}}{SS_{Total}} = \frac{11939}{11939 + 26314} = 0.312$$

31.2% of the variability in *Longevity* is explained by *Treatment*.

# *R*-squared versus *F* statistic

---

- ▶ *F* statistic is the ratio of the average variability in the model and the average variability in the residuals

$$F = \frac{MSG_{\text{Group}}}{MSE} = \frac{\frac{SS_{\text{Group}}}{K-1}}{\frac{SSE}{n-K}}$$

It indicates whether the explanatory variable is significant in explaining the response variable (the **significance** of the model)

- ▶  $R^2$  is the fraction of the total (response) variability that is explained by the model (explanatory variables)

$$R^2 = \frac{SS_{\text{Group}}}{SS_{\text{Total}}}$$

It measures how well the explanatory variable actually explains the response variable (the **strength** of the model)

# R-squared versus F statistic

```
summary(flymodel)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment      4  11939   2984.8    13.61 3.52e-09 ***
## Residuals    120  26314     219.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

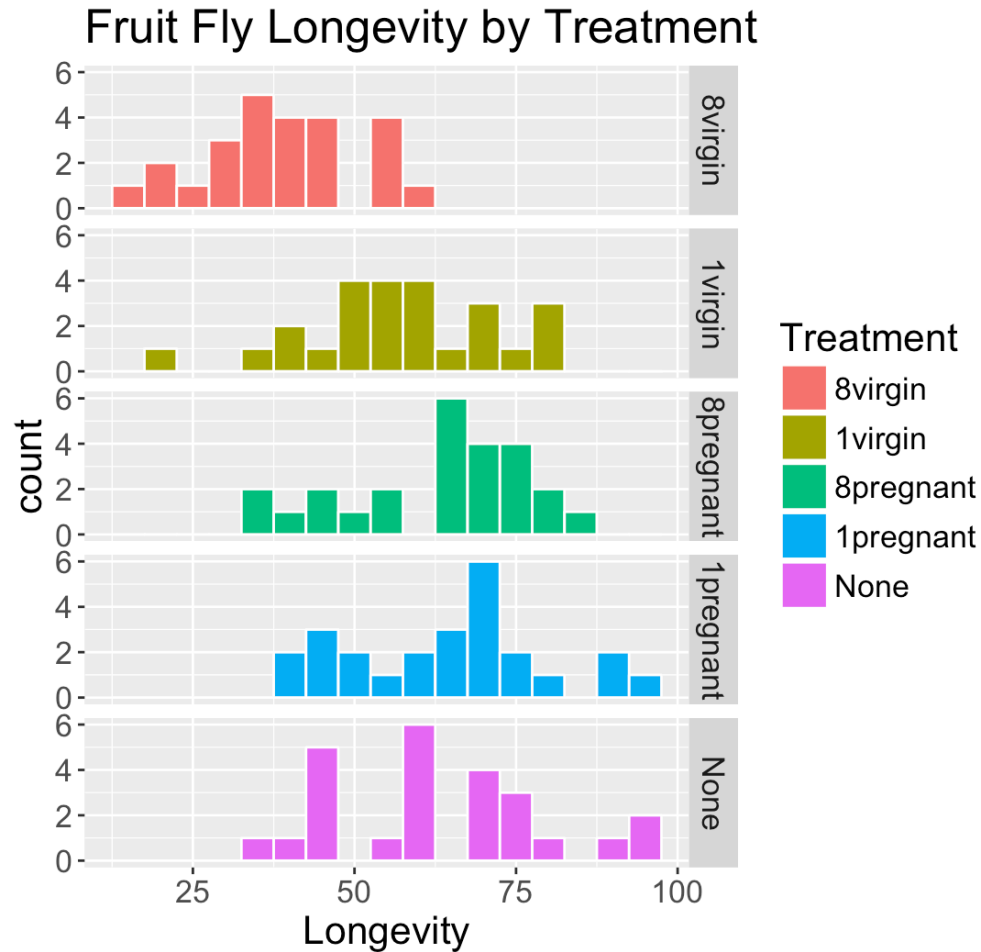
```
summary(diamodel <- aov(log(Carat) ~ Color, data=Diamonds))
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Color          3    7.62   2.5392    12.74 7.28e-08 ***
## Residuals    303   60.38   0.1993
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- ▶  $R^2_{fruitfly} = \frac{11939}{11939+26314} = 0.312$ ,  $R^2_{diamond} = \frac{7.62}{7.62+60.38} = 0.112$
- ▶ These two models have similar  $F$  statistics but very different  $R^2$  values.



# Multiple comparisons



- ▶ Based on the ANOVA  $F$  test, we reject the null hypothesis that there is no group effect (all groups have the same mean).
- ▶ At least one group has a significantly different mean from the other groups.
- ▶ Which one(s)?
- ▶ How many pairwise comparisons are there?
- ▶  $m = C_K^2 = \frac{K(K-1)}{2}$ .
- ▶ For this example, number of comparisons  $m = C_5^2 = \frac{5(5-1)}{2} = 10$ .

# Multiple comparisons

---

For comparing two population means, we may

- ▶ conduct a two-sample  $t$  test for  $H_0 : \mu_1 = \mu_2$

$$t = \frac{\bar{y}_1 - \bar{y}_2}{SE_{\bar{y}_1 - \bar{y}_2}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(df)$$

- ▶ or compute the confidence interval for  $\mu_1 - \mu_2$

$$(\bar{y}_1 - \bar{y}_2) \pm t^* SE_{\bar{y}_1 - \bar{y}_2} = (\bar{y}_1 - \bar{y}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- ▶ **Equivalence** between a level  $\alpha$  two-sided test and a level  $1 - \alpha$  confidence interval

$\mu_1$  and  $\mu_2$  are significantly different at level  $\alpha = 0.05$



95% confidence interval for  $\mu_1 - \mu_2$  does NOT contain 0

# Multiple comparisons

```
t.test(fly$Longevity[fly$Treatment=="8virgin"],
       fly$Longevity[fly$Treatment=="1virgin"])
```

```
##
## Welch Two Sample t-test
##
## data: fly$Longevity[fly$Treatment == "8virgin"] and fly$Longevity[fly$Treatment =
## t = -4.6936, df = 46.03, p-value = 2.436e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval: Interpret confidence interval: we are 95%
## -25.77647 -10.30353 confident (about the method) that the interval
## sample estimates: [-25.8, -10.3] will contain the true population
## mean of x mean of y mean.
## 38.72 56.76
```

- ▶ Level 0.05 test for  $\mu_1 = \mu_2$ :  $t = -4.7$  and  $P = 2.4 \times 10^{-5} < 0.05$ .
- ▶ 95% C.I. for  $\mu_1 - \mu_2$  is  $[-25.8, -10.3]$ , which **does NOT contain 0**.

# Multiple comparisons

```
t.test(fly$Longevity[fly$Treatment=="1pregnant"],  
       fly$Longevity[fly$Treatment=="None"])
```

```
##  
## Welch Two Sample t-test  
##  
## data: fly$Longevity[fly$Treatment == "1pregnant"] and fly$Longevity[fly$Treatment  
## t = 0.27303, df = 47.881, p-value = 0.786  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -7.892261 10.372261  
## sample estimates:  
## mean of x mean of y  
## 64.80 63.56
```

- ▶ Level 0.05 test for  $\mu_1 = \mu_2$ :  $t = 0.3$  and  $P = 0.79 > 0.05$ .
- ▶ 95% C.I. for  $\mu_1 - \mu_2$  is  $[-7.9, 10.4]$ , which **DOES contain 0**.

# Multiple comparisons: PROBLEM!

---

For the fruit fly example, we have 5 groups and therefore 10 pairwise comparisons. For each comparison,

- ▶ We reject  $H_0$  when  $P \leq 0.05$ :  $t = \frac{\bar{y}_1 - \bar{y}_2}{SE_{\bar{y}_1 - \bar{y}_2}}$  is large  $\Rightarrow$   $P$ -value is small.

If  $H_0$  is true ( $\mu_1 - \mu_2 = 0$ ),  $t = \frac{\bar{y}_1 - \bar{y}_2}{SE_{\bar{y}_1 - \bar{y}_2}}$  should be small. But if  $t = \frac{\bar{y}_1 - \bar{y}_2}{SE_{\bar{y}_1 - \bar{y}_2}}$  turns out to be large, we think the chance of getting such a large  $t$  is low and then suspect  $H_0$  is not true.

- ▶ However, although the chance is low, it is still possible. When we mistakenly reject a null hypothesis, we make a **type I error**.
- ▶ The chance of making a type I error in a test is  $\alpha = 0.05$ .
- ▶ The chance of NOT making a type I error in a test is  $1 - \alpha = 0.95$ .
- ▶ For 10 tests, the chance is of NOT making a type I error is  $(1 - \alpha)^{10} = 0.60$
- ▶ The **family-wise type I error rate** is  $1 - (1 - \alpha)^{10} = 0.40!$

# Multiple comparisons: inflated type I error

**Type I Error** is the incorrect rejection of a null hypothesis when the null hypothesis is in fact true.

- ▶ In general, the chance of making a type I error in a level  $\alpha$  test is  $\alpha$ .
- ▶ The chance of making a type I error in  $m$  level  $\alpha$  tests is  $1 - (1 - \alpha)^m$ , which is usually much larger than  $\alpha$ . We say the type I error rate is **inflated**.

## How to address the problem?

- ▶ "Ignore it " - focus only on the type I error of individual test  
Fisher's Least Significance Difference (LSD)
- ▶ Control it - reduce the type I error of individual test  
Bonferroni correction
- ▶ Or find the balance  
Tukey's Honest Significant Difference (HSD)

# Fisher's Least Significance Difference (LSD)

---

To perform pairwise comparisons of more than two means,

1. Verify that  $F$ -test is significant.
2. Compute the test statistic for  $H_0 : \mu_i = \mu_j$

$$t = \frac{(\bar{y}_i - \bar{y}_j) - 0}{\sqrt{MSE \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}} \sim t(n - K)$$

and/or the confidence interval for  $\mu_i - \mu_j$

$$(\bar{y}_i - \bar{y}_j) \pm t^* \sqrt{MSE \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

to compare means of the  $i$ th group and the  $j$ th group.

3. If the test is significant or the confidence interval does not contain 0, the two groups are significantly different.

# Fisher's Least Significance Difference (LSD)

---

**Note:** the standard error of  $\bar{y}_i - \bar{y}_j$  in ANOVA pairwise comparisons,

$$SE_{\bar{y}_i - \bar{y}_j} = \sqrt{MSE \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

is different from the one in two-sample  $t$  test:

$$SE_{\bar{y}_i - \bar{y}_j} = \sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}$$

- ▶ This is because of the model assumption of ANOVA:  $\epsilon \stackrel{iid}{\sim} N(0, \sigma)$ .
- ▶  $\sigma$  is the same for all groups.
- ▶ Therefore, we plug in  $s_i^2 = s_j^2 = MSE$ .
- ▶ Fisher's LSD does not consider the inflated type I error problem. This method is usually not recommended but sometimes used in the exploratory stage of scientific studies.



# Bonferroni correction

---

1. Choose an  $\alpha$ -level as the family-wise type I error rate, e.g.,  $\alpha = 0.05$ .
2. Compute the number of comparisons as  $m = C_K^2 = \frac{K(K-1)}{2}$  ( $K$ : # of groups), e.g.,  $m = 10$  for  $K = 5$ .
3. Compute the test statistic and/or confidence interval using **Fisher's LSD**.
4. Set individual type I error rate to  $\frac{\alpha}{m}$  or adjust individual  $P$ -value to  $m \times P$ -value to achieve family-wise type I error rate  $\alpha$  because  $1 - (1 - \frac{\alpha}{m})^m \approx \alpha$ .
  - ▶ Test:  $P\text{-value} < \frac{\alpha}{m}$ ? Or equivalently  $m \times P\text{-value} < \alpha$ ?
  - ▶ C.I.:  $t^*$  is the value for the  $t(n - K)$  density curve with area  $1 - \frac{\alpha}{m}$  between  $-t^*$  and  $t^*$ .
- ▶ Bonferroni correction is very **conservative** because it makes individual test level  $\frac{\alpha}{m}$  small and thus very hard to reject a null hypothesis.

# Tukey's Honest Significant Difference (HSD)

---

1. Choose an  $\alpha$ -level as the family-wise type I error rate, e.g.,  $\alpha = 0.05$
2. Compute the number of comparisons as  $m = C_K^2 = \frac{K(K-1)}{2}$  ( $K$ : number of groups), e.g.,  $m = 10$  for  $K = 5$ .
3. Compute the test statistic and/or confidence interval using **Fisher's LSD**.
4. Adjust the  $t(n - K)$  distribution for computing  $P$ -value and  $t^*$  value.
  - ▶ Studentized range distribution
  - ▶ Tukey's HSD controls the family-wise type I error while retaining the usefulness of the individual tests.

# Multiple comparisons in R

```
# Fisher's LSD
library(agricolae)
LSD.test(y = flymodel, trt="Treatment", group=F)$comparison
```

##	Difference	pvalue	sig.	LCL	UCL
## 1pregnant - 1virgin	8.04	0.0573	.	-0.2526583	16.332658
## 1pregnant - 8pregnant	1.44	0.7316		-6.8526583	9.732658
## 1pregnant - 8virgin	26.08	0.0000	***	17.7873417	34.372658
## 1pregnant - None	1.24	0.7677		-7.0526583	9.532658
## 1virgin - 8pregnant	-6.60	0.1177		-14.8926583	1.692658
## 1virgin - 8virgin	18.04	0.0000	***	9.7473417	26.332658
## 1virgin - None	-6.80	0.1071		-15.0926583	1.492658
## 8pregnant - 8virgin	24.64	0.0000	***	16.3473417	32.932658
## 8pregnant - None	-0.20	0.9620		-8.4926583	8.092658
## 8virgin - None	-24.84	0.0000	***	-33.1326583	-16.547342

- ▶ The *8virgin* group has significantly different life span from the other four groups. The difference between the *1pregnant* and *1virgin* group is marginally significant (when  $P$ -value is close to 0.05).

# Multiple comparisons in R

```
# Bonferroni
```

```
LSD.test(y = flymodel, trt="Treatment", group=F, p.adj="bon")$comparison
```

##	Difference	pvalue	sig.	LCL	UCL
## 1pregnant - 1virgin	8.04	0.5728		-3.938139	20.018139
## 1pregnant - 8pregnant	1.44	1.0000		-10.538139	13.418139
## 1pregnant - 8virgin	26.08	0.0000	***	14.101861	38.058139
## 1pregnant - None	1.24	1.0000		-10.738139	13.218139
## 1virgin - 8pregnant	-6.60	1.0000		-18.578139	5.378139
## 1virgin - 8virgin	18.04	0.0003	***	6.061861	30.018139
## 1virgin - None	-6.80	1.0000		-18.778139	5.178139
## 8pregnant - 8virgin	24.64	0.0000	***	12.661861	36.618139
## 8pregnant - None	-0.20	1.0000		-12.178139	11.778139
## 8virgin - None	-24.84	0.0000	***	-36.818139	-12.861861

- ▶ The output from `LSD.test()` function adjusts the  $P$ -value using  $m \times P$ -value. Directly compare `pvalue` to  $\alpha = 0.05$  or determine whether the CIs contain 0.
- ▶ The Bonferroni method has much larger  $P$ -values and wider confidence intervals than the Fisher's method.

# Multiple comparisons in R

```
TukeyHSD(flymodel) # Tukey's HSD
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Longevity ~ Treatment, data = fly)
##
## $Treatment
##
```

	diff	lwr	upr	p adj
## 1virgin-8virgin	18.04	6.439532	29.64047	0.0003240
## 8pregnant-8virgin	24.64	13.039532	36.24047	0.0000004
## 1pregnant-8virgin	26.08	14.479532	37.68047	0.0000001
## None-8virgin	24.84	13.239532	36.44047	0.0000003
## 8pregnant-1virgin	6.60	-5.000468	18.20047	0.5157692
## 1pregnant-1virgin	8.04	-3.560468	19.64047	0.3126549
## None-1virgin	6.80	-4.800468	18.40047	0.4854206
## 1pregnant-8pregnant	1.44	-10.160468	13.04047	0.9969591
## None-8pregnant	0.20	-11.400468	11.80047	0.9999988
## None-1pregnant	-1.24	-12.840468	10.36047	0.9983034

- ▶ Compare `p adj` to 0.05 or determine whether the CIs contain 0.
- ▶ The Tukey's HSD method has larger adjusted  $P$ -values and wider confidence intervals than the Fisher's method but smaller  $P$ -values and narrower CIs than the Bonferroni method.

# Multiple comparisons in R

```
LSD.test(y = diamodel, trt="Color", group=F)$comparison # Fisher's LSD
```

##		Difference	pvalue	sig.	LCL	UCL
##	D - E	0.02298611	0.7717		-0.1327410	0.17871322
##	D - F	-0.20781627	0.0083	**	-0.3617971	-0.05383549
##	D - G	-0.35757624	0.0000	***	-0.5118916	-0.20326091
##	E - F	-0.23080238	0.0009	***	-0.3660088	-0.09559599
##	E - G	-0.38056235	0.0000	***	-0.5161496	-0.24497508
##	F - G	-0.14975996	0.0281	*	-0.2833379	-0.01618206

```
LSD.test(y = diamodel, trt="Color", group=F, p.adj="bon")$comparison # Bonferroni
```

##		Difference	pvalue	sig.	LCL	UCL
##	D - E	0.02298611	1.0000		-0.1871766	2.331489e-01
##	D - F	-0.20781627	0.0500	*	-0.4156223	-1.028859e-05
##	D - G	-0.35757624	0.0000	***	-0.5658337	-1.493188e-01
##	E - F	-0.23080238	0.0053	**	-0.4132712	-4.833353e-02
##	E - G	-0.38056235	0.0000	***	-0.5635452	-1.975795e-01
##	F - G	-0.14975996	0.1687		-0.3300311	3.051115e-02

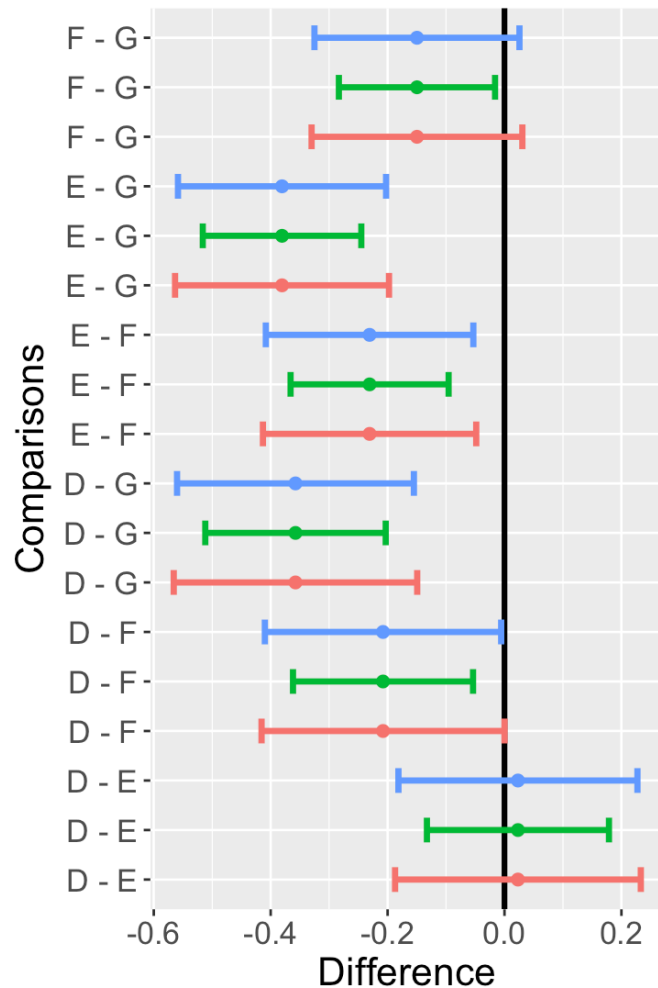
# Multiple comparisons in R

```
TukeyHSD(diamodel) # Tukey's HSD
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = log(Carat) ~ Color, data = Diamonds)
##
## $Color
##           diff           lwr           upr           p adj
## E-D -0.02298611 -0.227423474 0.1814513 0.9914530
## F-D 0.20781627 0.005671473 0.4099611 0.0412729
## G-D 0.35757624 0.154992248 0.5601602 0.0000439
## F-E 0.23080238 0.053304460 0.4083003 0.0048645
## G-E 0.38056235 0.202564412 0.5585603 0.0000004
## G-F 0.14975996 -0.025600091 0.3251200 0.1238450
```

- Based on Bonferroni correction and Tukey's HSD, diamonds with colors D and E have significantly different carats from diamonds with colors F and G, while D and E, or F and G, have no significant difference.

# Multiple comparisons



	Fisher's LSD	Bonferroni	Tukey's HSD
Overall	Liberal	Conservative	Moderate
Width of CI	Narrow	Wide	Moderate
P-value	Small	Large	Moderate

- ▶ The Bonferroni correction constrains family-wise type I error rate at a desirable level and is the most conservative.
- ▶ Tukey's HSD falls in between Fisher's LSD and Bonferroni correction and is close to Bonferroni correction.



# Valentine's Day

---

**Bloomberg**  
**View**



WORLD

## Valentine's Day and the Economics of Love



2



FEB 13, 2013 6:30 PM EST

By [Justin Wolfers](#) & [Betsey Stevenson](#)

a | [A](#)

Survey data from the [Gallup](#) Organization, where Justin works as a senior scientist, allow us to take a uniquely deep look at the [state of love](#) around the world. In 2006 and 2007, Gallup went to [136 countries](#) and asked people, “Did you experience love for a lot of the day yesterday?” It’s the largest such dataset ever collected.



# Valentine's Day


---


Money is related to love. Those with more household income are slightly more likely to experience the feeling. Roughly speaking, doubling your income is associated with being about 4 percentage points more likely to be loved. Perhaps having more money makes it easier to find time for love.

What's perhaps more striking is how little money matters on a global level. True, the populations of richer countries are, on average, slightly more likely to feel loved than those of poorer countries. But love is still abundant in the poorer countries: People in Rwanda and the Philippines enjoyed the highest love ratios, with more than 9 in 10 people providing positive responses. Armenia, Uzbekistan, Mongolia and Kyrgyzstan, with economic output per person in the middle of the range, all had love ratios of less than 4 in 10.

# Valentine's Day

 **THE WORLD BANK**  
IBRD • IDA





New to this site? [Start Here](#) 

This page in: [English](#) [Español](#) [Français](#) [العربية](#) [中文](#)

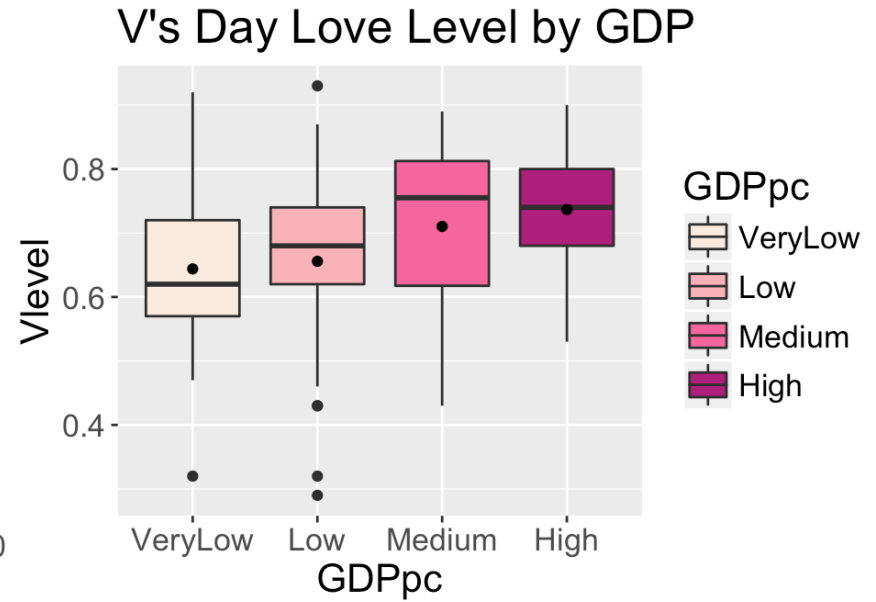
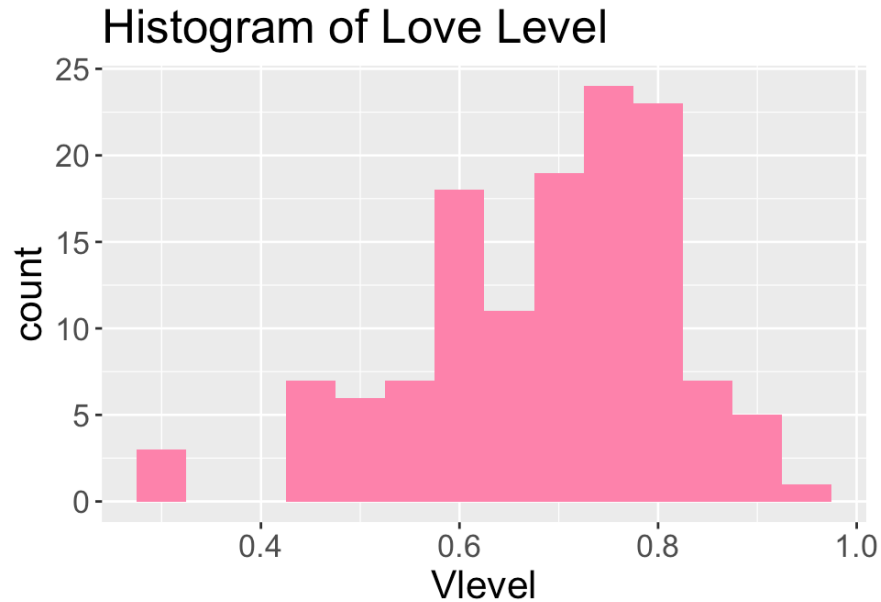
## World Bank Open Data

Free and open access to global development data



 GDP per capita (current US\$)

# Valentine's Day



##	GDPpc	Vlevel.n	Vlevel.mean	Vlevel.sd
## 1	VeryLow	33.000	0.644	0.125
## 2	Low	33.000	0.656	0.151
## 3	Medium	32.000	0.710	0.125
## 4	High	33.000	0.737	0.096

# Valentine's Day

```
summary(vdmodel <- aov(Vlevel ~ GDPpc, data=VD))
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## GDPpc          3  0.1922  0.06405     4.056 0.00862 **
## Residuals    127  2.0054  0.01579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$R^2 = 0.1922 / (0.1922 + 2.0054) = 0.087$$

```
LSD.test(y = vdmodel, trt="GDPpc", group=F, p.adj="bon")$comparison
```

```
##              Difference pvalue sig.          LCL          UCL
## High - Low          0.08090909 0.0599      . -0.002005063 0.16382324
## High - Medium        0.02635417 1.0000      -0.057205243 0.10991358
## High - VeryLow       0.09272727 0.0197      *  0.009813119 0.17564143
## Low - Medium        -0.05455492 0.4953      -0.138114334 0.02900449
## Low - VeryLow        0.01181818 1.0000      -0.071095972 0.09473234
## Medium - VeryLow     0.06637311 0.2111      -0.017186304 0.14993252
```

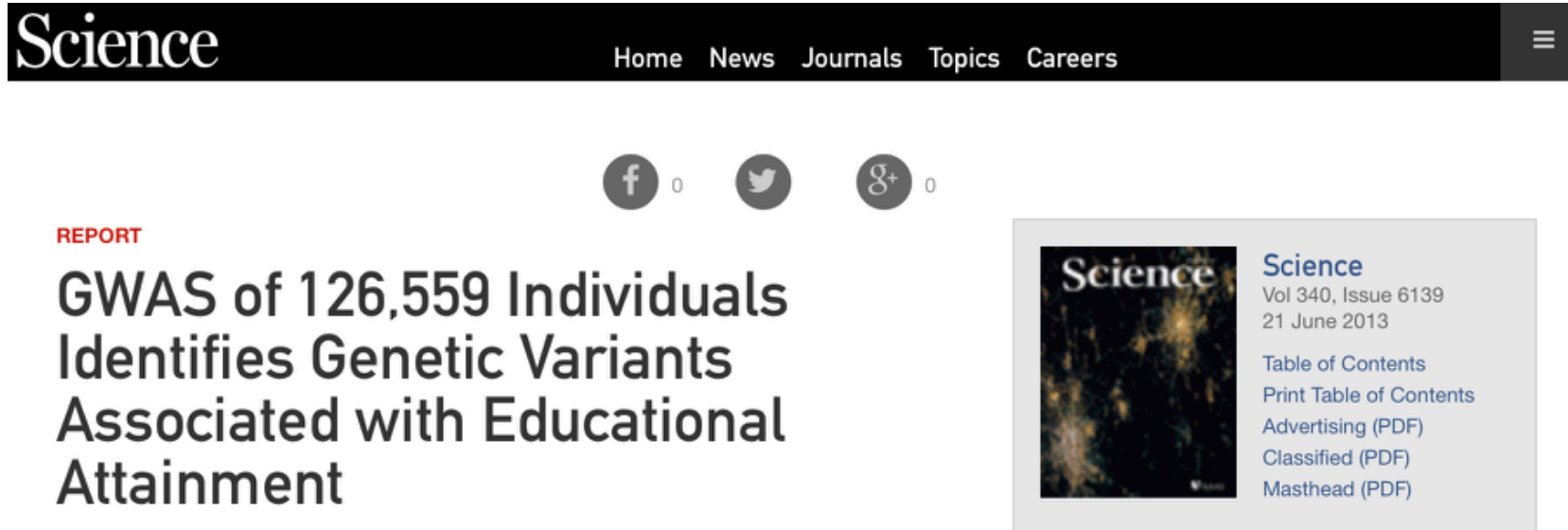
# Valentine's Day

TukeyHSD( vdmmodel ) \$GDPpc

##		diff	lwr	upr	p adj
##	Low-VeryLow	0.01181818	-0.0687188789	0.09235524	0.98094716
##	Medium-VeryLow	0.06637311	-0.0147907117	0.14753692	0.14954150
##	High-VeryLow	0.09272727	0.0121902120	0.17326433	0.01705320
##	Medium-Low	0.05455492	-0.0266088935	0.13571874	0.30251884
##	High-Low	0.08090909	0.0003720302	0.16144615	0.04848918
##	High-Medium	0.02635417	-0.0548096511	0.10751798	0.83266503

- ▶ Conclusion: There was significant difference in Valentine's Day love level among countries with different GDP per capita. Specifically, countries with very low and high GDPpc had the most significant difference, while the difference between countries with low and high GDPpc was marginally significant. The ANOVA model explained ~9% variability in the data. The rest 91% variability may be explained by other unknown variables.
- ▶ Next lecture, we will consider one additional explanatory variable using the two-way ANOVA model.

# Genome-Wide Association Study



- ▶ Sample size: 126,559
- ▶ Response variables: *EduYears* (Years of schooling) and *College* (college completion)
- ▶ Number of SNPs (explanatory variables):  $\sim 1,000,000$
- ▶ Number of tests:  $\sim 1,000,000$

# Genome-Wide Association Study

SNP	Chr	Discovery stage				Replication stage	
		Beta/OR	<i>P</i> value	<i>I</i> <sup>2</sup>	<i>P</i> <sub>het</sub>	Beta/OR	<i>P</i> value
<i>EduYears</i>							
rs9320913	6	0.106	4.19×10 <sup>−9</sup>	18.3	0.097	0.077	0.012
rs3783006	13	0.096	2.29×10 <sup>−7</sup>	0	0.982	0.056	0.055
rs8049439	16	0.090	7.12×10 <sup>−7</sup>	10.7	0.229	0.065	0.026
rs13188378	5	−0.136	7.49×10 <sup>−7</sup>	0	0.791	0.091	0.914
<i>College</i>							
rs11584700	1	0.921	2.07×10 <sup>−9</sup>	13.8	0.179	0.912	4.86×10 <sup>−4</sup>
rs4851266	2	1.050	2.20×10 <sup>−9</sup>	23.7	0.049	1.049	0.003
rs2054125	2	1.468	5.55×10 <sup>−8</sup>	7	0.325	1.098	0.225
rs3227	6	1.043	6.02×10 <sup>−8</sup>	5	0.363	1.010	0.280
rs4073894	7	1.076	4.41×10 <sup>−7</sup>	0	0.765	1.003	0.467
rs12640626	4	1.041	4.94×10 <sup>−7</sup>	10.9	0.234	1.000	0.495

- ▶ Bold rows: genome-wide-significant SNPs ( $P < 5 \times 10^{-8}$ )
- ▶ Other rows: suggestive SNPs ( $P < 10^{-6}$ )
- ▶ Replication stage
  - \* One-sided tests
  - \* For *EduYears*,  $P < \frac{0.05}{4}$
  - \* For *College*,  $P < \frac{0.05}{6}$
- ▶ Significant SNP for *EduYears*:  $R^2 \approx 0.2\%$
- ▶ All SNPs for *EduYears*:  $R^2 \approx 2\%$



# Summary

---

- ▶ ASSESS:  $R^2$  statistic  $R^2 = \frac{SS_{Group}}{SS_{Total}}$ 
  - Comparing  $R^2$  to  $F$  statistic
- ▶ Multiple pairwise comparisons
  - Problem: inflated type I error
  - Fisher's Least Significant Difference (LSD)
  - Bonferroni correction
  - Tukey's Honest Significant Difference (HSD)
  - Compare the three methods
- ▶ Example 1: ♥ Valentine's Day ♥
- ▶ Example 2: Genome-Wide Association Study (GWAS)