Lecture 11

Do redheads have a lower pain threshold?

Dr. Franco Ubaudi

The Nature of Data
Western Sydney University

Spring 2021

Do redheads have a lower pain threshold?

Previously we looked at the following scenarios:

- one qualitative variable
  e.g. Iraqi Refugees (stress category)
- two qualitative variables
  e.g. eels (species & habitat)
- one quantitative and one qualitative variable
  eg. Maternal smoking (Birth weight & smoking status)
- two quantitative variables
  e.g. Do taller people earn more (height & income)

Last week we used a *t-test* for comparing two means from two population groups.

What happens when we have more?

Do redheads have a lower pain threshold?

For the maternal smoking dataset, there could have been more than two smoking statuses, e.g.

1. Non-smoker
2. Light smoker
3. Average smoker
4. Heavy smoker

Instead we consider **hair colour** (more than two colours) versus **pain tolerance**

Do redheads have a lower pain threshold?

Anesthesiology. 2004 August ; 101(2): 279–283.

Anesthetic Requirement is Increased in Redheads

Edwin B. Liem, M.D.[*], **Chun–Ming Lin, M.D.**[†], **Mohammad–Irfan Suleman, M.D.**[‡], **Anthony G. Doufas, M.D., Ph.D.**[*], **Ronald G. Gregg, Ph.D.**[§], **Jacqueline M. Veauthier, Ph.D.**[¶], **Gary Loyd, M.D.**[#], and **Daniel I. Sessler, M.D.**[**]

J Am Dent Assoc. 2009 July ; 140(7): 896–905.

Genetic variations associated with red hair color and fear of dental pain, anxiety regarding dental care and avoidance of dental care

Anesthesiology. 2005 March ; 102(3): 509–514.

Increased Sensitivity to Thermal Pain and Reduced Subcutaneous Lidocaine Efficacy in Redheads

Edwin B. Liem, M.D.[*], **Teresa V. Joiner, B.S.N.**[†], **Kentaro Tsueda, M.D.**[‡], and **Daniel I. Sessler, M.D.**[§]

Pigment Cell Melanoma Res. 2016 March ; 29(2): 239–242. doi:10.1111/pcmr.12445.

Natural hair color and questionnaire-reported pain among women in the United States

Wen-Qing Li[1,2], **Xiang Gao**[3,4], **Shelley S. Tworoger**[4,5], **Abrar A. Qureshi**[1,2,4], and **Jiali Han**[4,6,7]

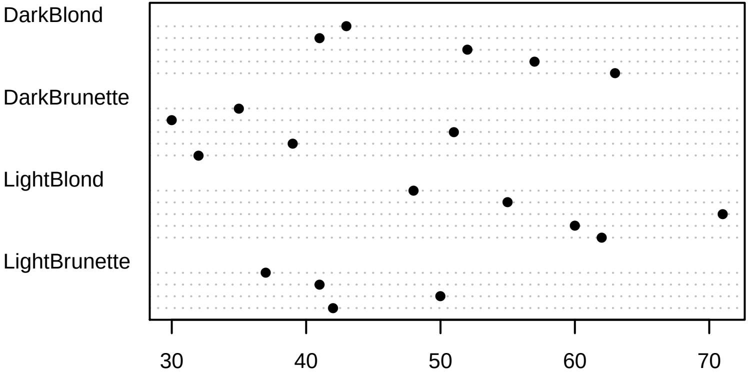
Melanocortin-1 receptor gene variants affect pain and μ-opioid analgesia in mice and humans

J S Mogil, J Ritchie, S B Smith, K Strasburg, L Kaplan, M R Wallace, R R Romberg, H Bijl, E Y Sarton, R B Fillingim, A Dahan

J Med Genet 2005;**42**:583–587. doi: 10.1136/jmg.2004.027698

Pain tolerance and hair colour

Dot chart for 19 individuals as they vary with hair colour:



Evidence for pain tolerance by hair colour?

Hypothesis

More precise:

does *average* pain tolerance vary according to hair colour?

K groups (or different hair colours) each with population mean pain tolerance μ_k

Looking for evidence that **all** μ_k are not equal

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 = \cdots = \mu_k \\ H_1 &: \quad \mu_i \neq \mu_j \text{ for at least one pair } i, j \end{aligned}$$

So even if one μ_k is sufficiently different to the others, we reject H_0

t statistic

To compare two groups (smoker vs. non-smokers) we computed a t-statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where the pooled variance is:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

How to extend this difference from equal means **for more than two groups**?

F statistic

K groups with each group k having sample size n_k.

Sample mean $\bar{x}_k$ and standard deviation s_k.

All together $n = n_1 + \cdots + n_K$ elements.

Global mean:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{K} n_k \bar{x}_k$$

The variance between groups is:

$$SS_B = \sum_{i=1}^{K} n_k (\bar{x}_k - \bar{x})^2$$

SS$_B$ is **sum of squares between** groups

If $\bar{x} = \bar{x}_k$ for all k, $SS_B = 0$

Sum of squares within groups

Measure variation within groups

$$SS_W = \sum_{i=1}^{K} (n_k - 1) s_k^2$$

This captures the variability within each group around its own mean.

$MSE = SS_W/(n-K)$ is called the **mean-square error**.

It's a pooled estimate of variance analogous to s_p^2

Let's put this together..

F statistic

$$F = \frac{SS_B/(K-1)}{SS_W/(n-K)}$$

is called the F-statistic

If $K = 2$ then $F = t^2$

Hair data F statistic

	DarkBlond	DarkBrunette	LightBlond	LightBrunette
ns	5.0	5.0	5.0	4.00
means	51.2	37.4	59.2	42.50
vars	86.2	69.3	72.7	29.67

$K = 4$, $n = 19$ measurements

Global mean $\bar{x} = 47.84$, and $SS_B = 1360.73$ and $SS_W = 1001.8$

$$F = \frac{SS_B/(K-1)}{SS_W/(n-K)} = \frac{1360.73/(4-1)}{1001.8/(19-4)} = 6.791$$

Is F large enough to reject the null hypothesis?

H_0: all category means are equal

H_1: at least one mean is significantly different

F statistic done in R

```
> oneway.test(Pain~HairColour, data=hair, var.equal=TRUE)
```

```
One-way analysis of means
data: Pain and HairColour F = 6.7914, num df = 3,
denom df = 15, p-value = 0.004114
```

Here, the p-value is based on the **F-distribution**

If $CV = 0.05$, we reject H_0 since *p-value* $< CV$

$\therefore\ \mu_i \neq \mu_j$ for at least one pair i, j

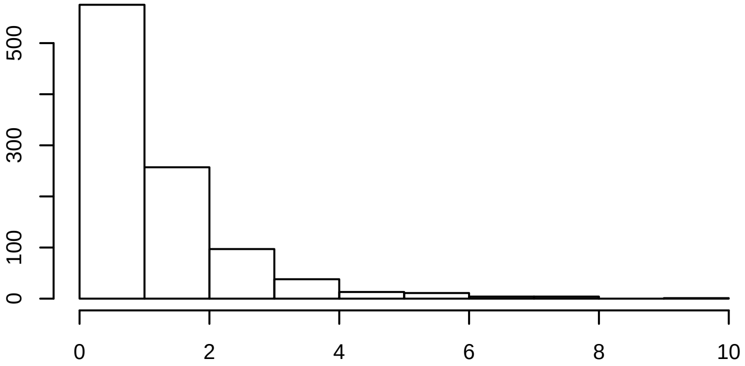
p-value using permutation simulation

```
> ## compute the data F statistic
> Fstat = oneway.test(Pain~HairColour, data=hair,
var.equal=TRUE)$statistic
>
> ## compute the F statistic when the category means are equal
> x = replicate(1000,{
+   ## shuffle the categories to force equal means
+   hair.perm = sample(hair$HairColour)
+   ## compute the F statistic using the shuffled data
+   oneway.test(Pain~hair.perm, data=hair,
var.equal=TRUE)$statistic
+ })
```

similar to the permutation version of the two group case but randomly permute all group labels.

Simulation and p-value

Recall $F = 6.7914$



In R, the p-value can be computed by
 sum($x >$ fStat)/1000

In this instance it was 0.004

Note that F-statistics are always positive and thus F-tests are in effect two-sided.

Using F dist to get a p-value

If data is approximately normally distributed or data set is large enough so CLT works, then under the null hypothesis, the F-statistic has the F-distribution.

Two parameters:

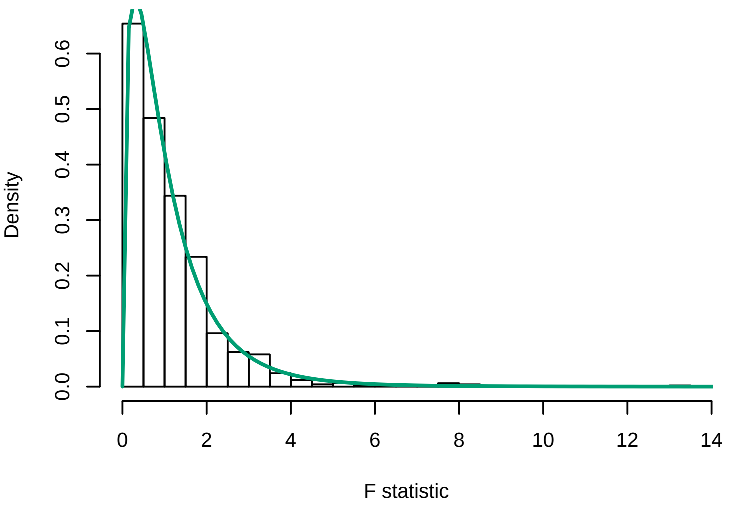
  the numerator $K - 1$

  and denominator $n - K$ degrees of freedom.

pf(fStat, 4 - 1, 19 - 4, lower = FALSE)

gives $p\text{-}value = 0.004114227$

Compare permutation simulation and F-distribution



ANOVA table

The procedure is called a *one-way Analysis of Variance* (one-way ANOVA).

ANOVA table:

	df	SSQ	Mean Sq	F stat	p-value
Between	$K-1$	SS_B	$SS_B/(K-1)$	F	
Within	$N-K$	SS_W	$SS_W/(N-K)$		

```
> fit = aov(Pain~HairColour, data=hair)
> summary(fit)

            Df Sum Sq Mean Sq F value  Pr(>F)
HairColour   3   1361   453.6   6.791 0.00411 **
Residuals   15   1002    66.8
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Post hoc tests

F-test says at least one pair of means differs but which one(s)?

First try: t-statistic for each pair

For K groups that is $K(K-1)/2$ hypotheses.

This massively increases our chances of rejecting the null hypothesis in error, (hence committing a Type I error)

How bad is it?

How often would we get an error

Time for a simulation:

```
> n = 20 # sample size per category
> K= 10 # number of categories
> grp = rep(1:K, each=n)
> x = rnorm(length(grp)) # generate sample, all with same mean
> mns = tapply(x, grp, mean) # compute category mean
> # compute pooled variance
> sp = sqrt(sum((table(grp)-1)*tapply(x, grp, var))/
+                           (length(grp)-length(mns)))
> # compute the t-statistics of the largest difference in means
> maxT = diff(range(mns))/(sp*sqrt(2/n))
> print(maxT)
```

This results in a p-value of 0.0147 – at least one pair “detected” in error.

What to do?

Set significance $p < 0.01$?

Might be too conservative or not small enough for large K

Robust way is **Tukey's range test** (also known as Tukey's Honest Significant Difference test or Tukey's HSD).

Idea is to control the *family-wise error rate*.

It allows for multiple testing by considering the distribution of the *maximum* t-statistic across all categories.

Tukey's range test

t-test for the ith vs jth group comparison:

$$t_{i,j} = \frac{\bar{x}_i - \bar{x}_j}{s_p^{ij}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Replace s_p^{ij} term by the full *pooled sample variance*:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{n_1 + n_2 + \cdots + n_k - k}$$

This is just the denominator of the F-statistic and also given in the ANOVA table by "Mean Sq".

Tukey's range test

1. For each group pair (i, j), calculate the statistic:

$$q_{ij} = \frac{|\bar{x}_i - \bar{x}_j|}{s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

2. If $q_{i,j} > q_\alpha$, then the null hypothesis $H_0 : \mu_i \neq \mu_j$, for population means μ_i, μ_j can be rejected with significance α.

q_α is the appropriate value from the *Studentized range distribution*.

Works provided either the data is normally distributed or n_j are large enough.

Family-wise error rate $<$ significance α.

Using the Studentized range distribution

In R can use `TukeyHSD`.

```
> fit = aov(Pain~HairColour, data=hair)
> TukeyHSD(fit)
```

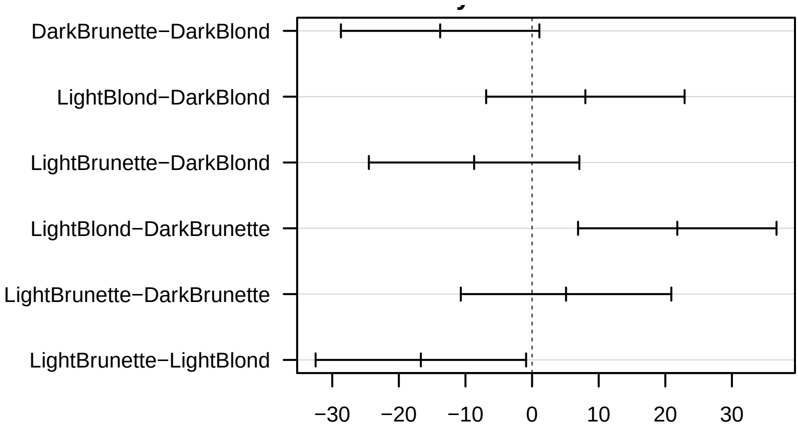
```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Pain ~ HairColour, data = hair)
```

```
                            diff        lwr        upr
DarkBrunette-DarkBlond     -13.8 -28.696741  1.0967407
LightBlond-DarkBlond         8.0  -6.896741 22.8967407
LightBrunette-DarkBlond     -8.7 -24.500380  7.1003795
LightBlond-DarkBrunette     21.8   6.903259 36.6967407
LightBrunette-DarkBrunette   5.1 -10.700380 20.9003795
LightBrunette-LightBlond   -16.7 -32.500380 -0.8996205
                               p adj
DarkBrunette-DarkBlond     0.0740679
LightBlond-DarkBlond       0.4355768
LightBrunette-DarkBlond    0.4147283
LightBlond-DarkBrunette    0.0037079
LightBrunette-DarkBrunette 0.7893211
LightBrunette-LightBlond   0.0366467
```

	DarkBlond	DarkBrunette	LightBlond	LightBrunette
DarkBlond	0.000	2.670	-1.548	1.587
DarkBrunette	-2.670	0.000	-4.218	-0.930
LightBlond	1.548	4.218	0.000	3.046
LightBrunette	-1.587	0.930	-3.046	0.000

Confidence intervals

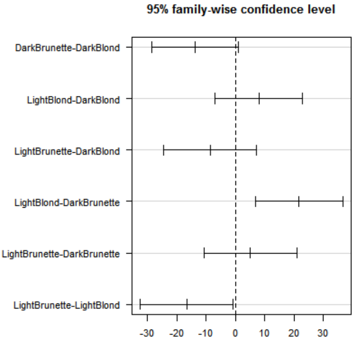


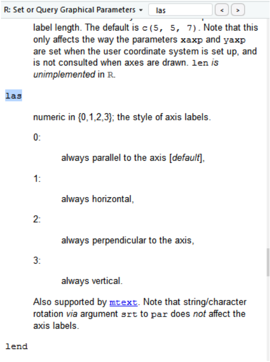
LightBlond/DarkBrunette and LightBrunette/LightBlond confidence intervals both don't overlap with 0

Generating an Interval Plot

```
par(mar = c(3, 11, 3.5, 0.5), cex = 0.7)
plot(TukeyHSD(fit), las = 1)
```

What the las parameter does (orientates axis labels) and las help, found via ?*par* in the console




R: Set or Query Graphical Parameters ▾ las

label length. The default is `c(5, 5, 7)`. Note that this only affects the way the parameters `xaxp` and `yaxp` are set when the user coordinate system is set up, and is not consulted when axes are drawn. `len` *is unimplemented* in R.

`las`

numeric in {0,1,2,3}; the style of axis labels.

0:

always parallel to the axis [*default*],

1:

always horizontal,

2:

always perpendicular to the axis,

3:

always vertical.

Also supported by `mtext`. Note that string/character rotation *via* argument `srt` to `par` does *not* affect the axis labels.

`lend`

A major study of redheads

Li, Wen-Qing et al. "Natural hair color and questionnaire-reported pain among women in the United States"

Study from around 150,000 women of varying ages, questioned multiple times.

Age corrected with multiple measurements per individual allowed for – much more complex analysis than our example above.

Results

Table 1. Mean difference in pain score according to natural hair color

	Difference in pain score					Per one unit of hair color[a]	P for trend[a]
	Black	Dark brown	Light brown	Blonde	Red		
Nurses' Health Study							
Average score							
Age-adjusted	0 (Ref)	0.92 (0.52, 1.33)	0.78 (0.37, 1.19)	0.95 (0.50, 1.40)	1.78 (1.24, 2.32)	0.17 (0.08, 0.26)	0.0002
Multivariate-adjusted[b]	0 (Ref)	1.14 (0.79, 1.48)	1.07 (0.72, 1.42)	1.28 (0.89, 1.66)	1.71 (1.25, 2.17)	0.19 (0.11, 0.26)	<0.0001
Updated score[c]							
Age-adjusted	0 (Ref)	1.05 (0.33, 1.77)	0.89 (0.17, 1.61)	0.99 (0.20, 1.79)	1.84 (0.88, 2.80)	0.15 (0.05, 0.25)	0.004
Multivariate-adjusted[b]	0 (Ref)	1.24 (0.64, 1.84)	1.16 (0.56, 1.76)	1.30 (0.64, 1.96)	1.70 (0.91, 2.49)	0.17 (0.04, 0.29)	0.009
Nurses' Health Study II							
Average score							
Age-adjusted	0 (Ref)	0.52 (0.17, 0.88)	0.59 (0.24, 0.95)	0.57 (0.19, 0.94)	1.16 (0.70, 1.62)	0.13 (0.06, 0.20)	0.0003
Multivariate-adjusted[b]	0 (Ref)	0.70 (0.35, 1.06)	0.78 (0.41, 1.14)	0.87 (0.49, 1.24)	1.19 (0.74, 1.63)	0.14 (0.08, 0.20)	<0.0001
Updated score[c]							
Age-adjusted	0 (Ref)	0.56 (−0.08, 1.20)	0.66 (0.02, 1.30)	0.71 (0.04, 1.38)	1.23 (0.41, 2.05)	0.16 (0.04, 0.28)	0.008
Multivariate-adjusted[b]	0 (Ref)	0.84 (0.20, 1.49)	0.97 (0.32, 1.63)	1.20 (0.52, 1.87)	1.38 (0.59, 2.16)	0.21 (0.10, 0.31)	0.0001
Nurses' Health Study and Nurses' Health Study II combined							
Average score							
Age-adjusted	0 (Ref)	0.71 (0.31, 1.11)	0.68 (0.41, 0.94)	0.74 (0.37, 1.10)	1.45 (0.85, 2.05)	0.14 (0.09, 0.20)	<0.0001
Multivariate-adjusted[b]	0 (Ref)	0.92 (0.49, 1.35)	0.93 (0.64, 1.22)	1.07 (0.67, 1.47)	1.45 (0.93, 1.96)	0.16 (0.11, 0.21)	<0.0001
Updated score[c]							
Age-adjusted	0 (Ref)	0.78 (0.30, 1.26)	0.76 (0.28, 1.24)	0.83 (0.31, 1.34)	1.49 (0.86, 2.11)	0.16 (0.08, 0.23)	<0.0001
Multivariate-adjusted[b]	0 (Ref)	1.05 (0.62, 1.49)	1.08 (0.63, 1.52)	1.25 (0.78, 1.72)	1.54 (0.98, 2.09)	0.19 (0.11, 0.27)	<0.0001

Reference hair colour was black hair, confidence intervals are reported for the difference.

If interval does not contain zero there is evidence for a difference.

There is evidence that redheads report more pain than women with black hair colour.

Summary

To identify if one of more categories from a set of categories have a different mean, we compute the F statistic.

The results of an F test are presented as an ANOVA table.

To identify which pairs have different means, we use Tukey's range test.