

Discussion 11 Solutions

1. Reading levels vary between different magazines. To avoid difficulties caused by different typefaces and sizes, an investigator just counted the number of letters and punctuation signs. Random samples of 20 sentences were selected from The New Yorker, Sports Illustrated, and National Geographic and the number of letters and punctuation signs in each were tabulated. The resulting summary statistics are:

New Yorker	Sports Illustrated	National Geographic
$\bar{y}_1 = 94.4$	$\bar{y}_2 = 92.9$	$\bar{y}_3 = 75.5$
$s_1 = 58.4$	$s_2 = 54.2$	$s_3 = 38.1$
$n_1 = 20$	$n_2 = 20$	$n_3 = 20$

Note: $s_i^2 = \sum_{i=1}^{n_i} \frac{(y_{ij} - \bar{y}_i)^2}{(n_i - 1)}$

- (a) In order to perform an ANOVA F test, what assumptions must be reasonably met? Check any of these assumptions you can. *ANSWER: We need to assume that the number of letters and punctuation signs in each sentence from the three magazines are approximately normal (we can't check this assumption with just summary measures) and that the variance of the three populations are approximately equal. We can see that the largest and smallest SDs are within a factor of 2, so this seems like a reasonable assumption. We must also assume the three groups are independent populations and the sentences within each sample are randomly chosen/independent for each.*
- (b) Construct the ANOVA table for this data and perform the relevant F test (assume that the necessary assumptions are reasonably met.)

ANSWER:

$$SSE_{Error}: (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 = 148196.4$$

$$OverallMean: \frac{n_1 * \bar{y}_1 + n_2 * \bar{y}_2 + n_3 * \bar{y}_3}{n_1 + n_2 + n_3} = 87.6$$

$$SST_{Trt}: n_1 * (\bar{y}_1 - \bar{y}_{..})^2 + n_2 * (\bar{y}_2 - \bar{y}_{..})^2 + n_3 * (\bar{y}_3 - \bar{y}_{..})^2 = 20 * (94.4 - 87.6)^2 + 20(92.9 - 87.6)^2 + 20(75.5 - 87.6)^2 = 4414.8$$

$$SST_{Tot}: 52611.2.$$

Source	SS	df	MS
Treat	4414.8	2	2207.4
Error	148196.4	57	2599.937
Tot	152611.2	59	

$F = \frac{MST_{Trt}}{MSE_{Error}} = \frac{2207.4}{2599.937} = 0.8490206$. $P(F_{2,57} > 0.8490206) = 0.4331721$. Greater than 10% can be read off F table. Fail to reject null. No evidence group means are different.

Calculations:

```
mean.ny=94.4; sd.ny=58.4; n.ny=20
mean.si=92.9; sd.si=54.2; n.si=20
mean.ng=75.5; sd.ng=38.1; n.ng=20
```

```
SSErr=(n.ny-1)*sd.ny^2+(n.si-1)*sd.si^2+(n.ng-1)*sd.ng^2
Ov.mean=(n.ny*mean.ny+n.si*mean.si+n.ng*mean.ng)/(n.ny+n.si+n.ng)
SSTrt=n.ny*(mean.ny-Ov.mean)^2+n.si*(mean.si-Ov.mean)^2+n.ng*(mean.ng-Ov.mean)^2
SSTot=SSErr+SSTrt
MSErr=SSErr/57
MSTrt=SSTrt/2
Fstat=MSTrt/MSErr
pf(Fstat, 2, 57, lower.tail=FALSE)
```

```
## [1] 0.4331721
```