ST 312: Ch. 13 Examples

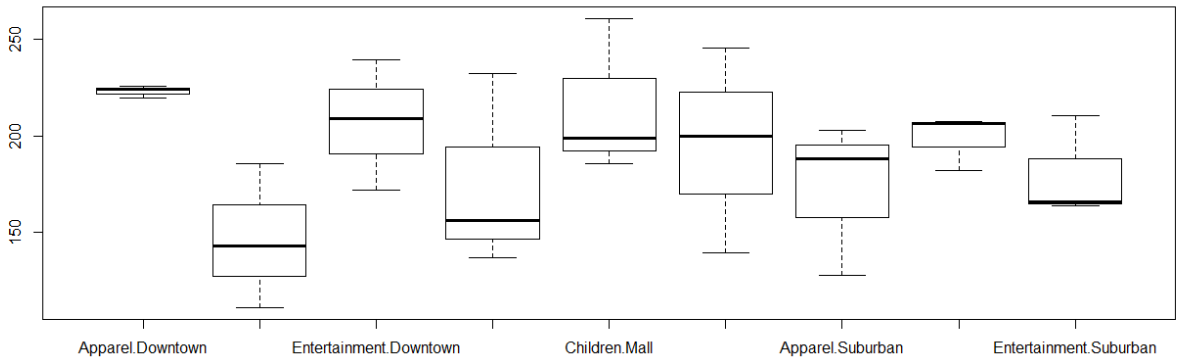
Example 1. Annual Returns (in thousands of dollars) is recorded for 27 shops in three different types of locations (Downtown, Mall, and Suburban) and for three different categories (Childrens, Apparel, and Entertainment).

(a) The table below shows the standard deviation for each cell. Are the assumptions of ANOVA satisfied?

category	location	n	SD of AnnualReturn
Apparel	Downtown	3	3.138556
Children	Downtown	3	37.324314
Entertainment	Downtown	3	33.712904
Apparel	Mall	3	50.382694
Children	Mall	3	40.114537
Entertainment	Mall	3	53.253057
Apparel	Suburban	3	39.915186
Children	Suburban	3	14.297225
Entertainment	Suburban	3	26.202422

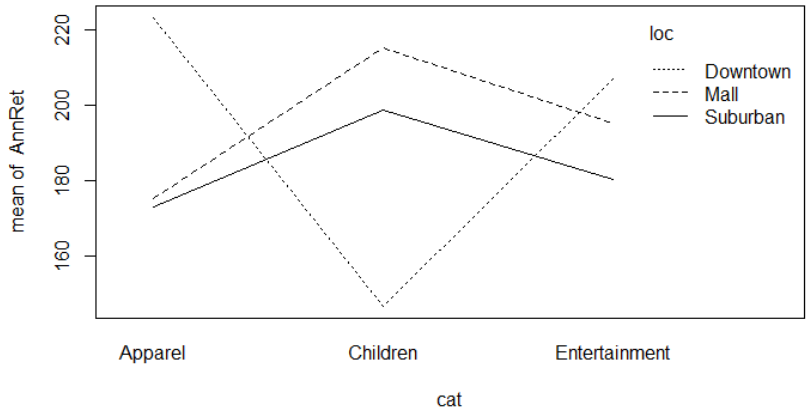
Recall: *If groups have the same size*, ANOVA performs well even if variances are different. But the standard deviation ranges from 3 to 53! There are only 3 observations per group, which could be part of the reason for the differences in variation.

(b) Below are boxplots for each cell. Does this give you any indication of what outcome might be reached?



The boxplots show the difference in variation, and that the first two groups appear to be the most different from one another. Many groups appear to be near equal. Some appear non-normal, but with only 3 observations, it's difficult to say.

(c) If we disregard the violations above (and assume ANOVA will perform well anyway), let us look now at the interaction plot below. Discuss what you see regarding an interaction and/or main effects.



The interaction plot shows that the mean Annual Return changes in the same way for Mall and Suburban locations, but seems to be different (particularly for Children) in the Downtown location. From these, I expect we may see a significant interaction between location and category.

(d) Complete the ANOVA table below. State any conclusions that may be reached using an α level of 0.10.

Source	Df	SS	MS	F Statistics
Category	**2**	237	**119**	**0.089**
Location	**2**	613	**306**	**0.230**
Category*Location	**4**	**12964**	**3241**	**2.429**
Error	**18**	24014	**1334**	
Total	**26**	37828		

Critical Values: $F_{cat} = F_{2,18} = 2.62, F_{Loc} = F_{2,18} = 2.62, F_{Inter} = F_{4,18} = 2.29$

The F statistic for the interaction is significant.

We have evidence at the 10% significance level to reject the null hypothesis of no interaction effect between Category and Location.

We cannot look at the main effects since the interaction is significant (we do not dismiss main effects as insignificant – these factors both have an effect in the form of the interaction).

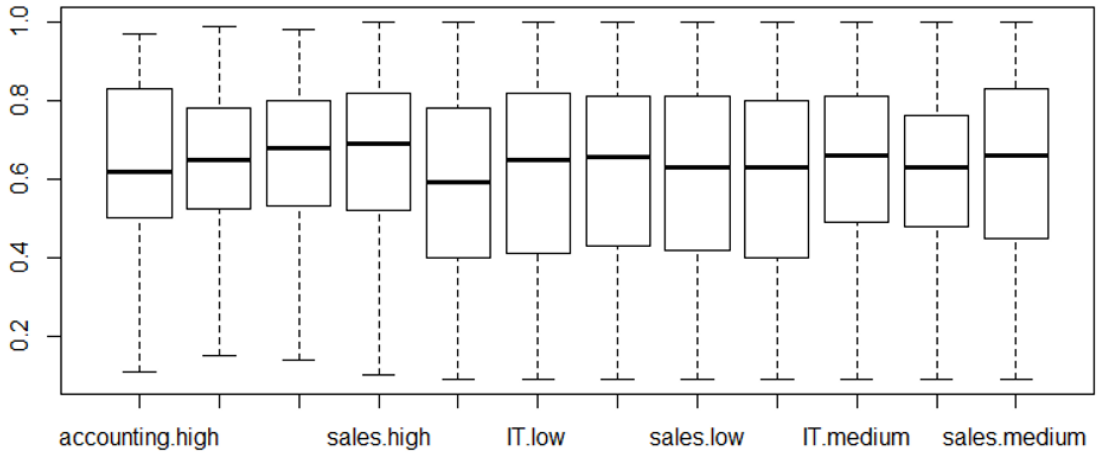
(e) Was this data from an observational study or an experiment?

This is an observational study. We cannot randomly assign location nor category to a shop.

Example 2. Job satisfaction level is recorded for 6764 employees at a large corporation in several departments (this study is *not* balanced). The employee's salary is categorized into three levels: low, medium, and high. Use two-way ANOVA to determine if department and/or salary have significant effects on job satisfaction.

(a) Below is a table of standard deviations, and side-by-side boxplots. Discuss the necessary conditions for two-way ANOVA.

	high	low	medium
accounting	0.237	0.252	0.262
IT	0.224	0.259	0.243
management	0.194	0.255	0.233
sales	0.236	0.252	0.250

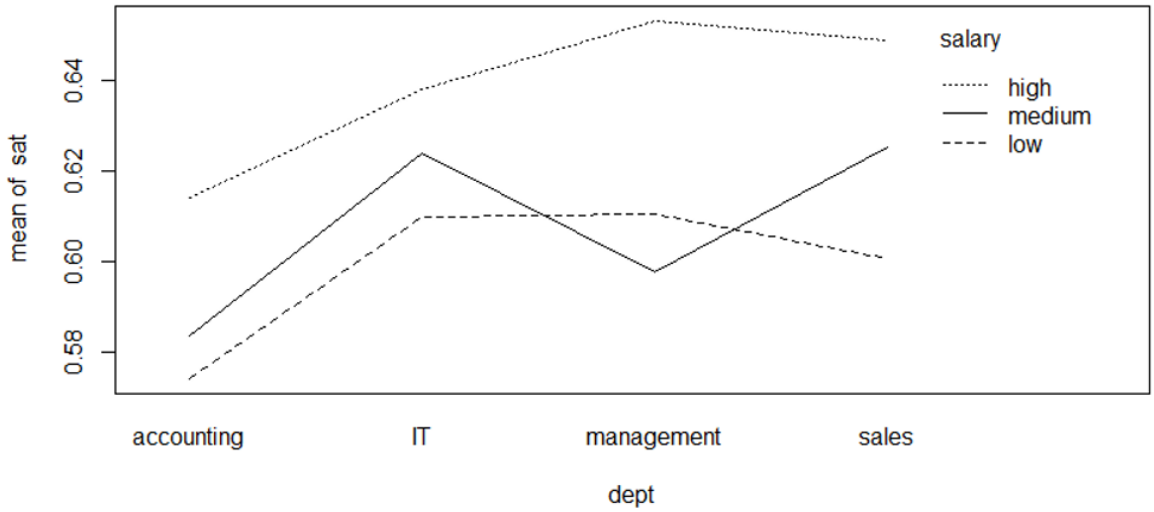


The standard deviations are all pretty close, I think it's fair to assume the condition is satisfied. The side-by-side boxplot shows that groups have near equal variance, and they appear to be symmetric (probably Normal).

(b) Based on the boxplots above, do you expect to find differences in the means? Explain.

The means do not seem to differ too significantly (the centers of each boxplot are similar, and they all overlap). However, some groups appear to be slightly different, and with *such* large sample sizes in the groups, we may find differences exist.

(c) Analyze the interaction plot shown below.



The medium salary group looks like it may cause an interaction due to the dip in mean satisfaction that can be seen in the management department. The other means seem to change in similar ways, so I'm not certain if this difference will be enough. As for main effects, there is maybe a difference between the "high" salary groups and other levels. It also looks like accounting may be lower, on average, than the other groups.

(d) Complete the two-way ANOVA to determine if department and/or salary have significant effects on job satisfaction. Use a 5% significance level.

Source	Df	SS	MS	F Statistics
Department	**3**	0.80949	**0.2698**	**4.37**
Salary	**2**	1.2846	**0.6423**	**10.41**
Interaction	**6**	0.1998	**0.0333**	**0.540**
Error	**6752**	416.598	**0.0617**	
Total	**6763**	418.8922		

Critical Values: $F_{dept} = F_{3,1000} = 2.61, F_{salary} = F_{2,1000} = 3.00, F_{Inter} = F_{6,1000} = 2.11$

We can see from the output that the interaction term is not significant. We fail to reject the null hypothesis of no interaction and move on to the main effect tests.

Both salary and department are significant at the 5% (and even lower) levels. We can conclude that department has a significant effect on mean job satisfaction rating, as does salary.

(e) The Tukey Honest Significant Differences post-hoc test can be used to test pairwise comparisons between levels of a factor. Below is the output for a Tukey HSD test for the Department variable.

Each line describes the comparison, the estimate of the difference, a confidence interval for the difference (lwr = lower limit, upr = upper limit), and an adjusted p-value (adjusted for the number of comparisons made).

Look at the confidence intervals and p-values to determine which departments have significantly different mean job satisfaction.

	diff	lwr	upr	p adj
IT-accounting	0.035990571	0.006602216	0.06537893	0.0089909
management-accounting	0.039197968	0.004868829	0.07352711	0.0176592
sales-accounting	0.032295621	0.007197406	0.05739384	0.0052495
management-IT	0.003207397	-0.028085521	0.03450031	0.9936051
sales-IT	-0.003694949	-0.024447716	0.01705782	0.9681752
sales-management	-0.006902346	-0.034206040	0.02040135	0.9157195

We can see from the Tukey test output that Accounting seems to be significantly different from the rest of the departments. We can see from the differences that Accounting appears to have the lowest mean job satisfaction ratings.