Section 10.3
“Making Sense of Statistical Significance”

(1) **Choosing a Significance Level**

For choosing a fixed level $\alpha$ of significance, a level of $\alpha = 0.05$ is generally accepted.

In deciding upon the actual level, though, two things should be considered:

(1) **The plausibility of $H_0$**

   The more difficult it will be to convince people to reject some long-standing assumption ($H_0$), the smaller $\alpha$ (the stronger evidence) will be needed.

(2) **The consequences of $H_0$**

   If rejecting $H_0$ means costly or radical change is needed (or if the well-being of individuals will be affected), then stronger evidence (smaller $\alpha$) is needed.

**Remember to ALWAYS state $\alpha$ in your concluding statements!!**
(II) **Statistical vs. Practical Significance**

When a p-value is sufficiently small (usually less than or equal to 0.05), then there exists strong evidence against \( H_0 \) (“no change” or “no effect”) in favor of \( H_a \) (some effect is present).

Even though the data may be STATISTICALLY significant and provide evidence that *some* effect is present, it does not indicate how large or how small an effect is present.

“Moral of the Story” = Do not attach too much importance to the significance test! Always plot and examine the data as well!

(III) **Validity of Statistical Inference**

Inference is only valid if data is collected properly, and surveys or experiments are conducted properly (CONTROL, RANDOMIZE, REPLICATE).

Formal inference can only correct and account for chance variation in sampling. It can NOT correct for flaws in data collection.

(Garbage in – garbage out!)
Multiple Analyses

You need more sophisticated statistical methods to handle multiple samples and multiple hypotheses. At this point in time, you should take only ONE sample at a time, testing only ONE pair of hypotheses at a time!

Why? Well ...

We know that in multiple samples taken, you can expect some variability due to chance.

In the context of significance testing ...

Test 1 hypothesis at a significance level $\alpha$, let’s say 0.05. You can be relatively confident that data with a $p$-value < 0.05 is an indication that the data are statistically significant and probably not produced just by chance.

Test 4 hypotheses at once, though, and a $p$-value of $< \alpha = 0.05$ becomes only suggestive and not conclusive. If $\alpha = 0.05$, that means that 5 in 100 samples WILL produce rare or extreme results purely by chance. By simultaneously examining multiple hypotheses and samples at the same time, you’re increasing the likelihood of getting one of those rare results.

**ASSIGNMENT:**

p. 588 (10.57)  p. 591 (10.60)
p. 589-590 (10.58, 10.59)  p. 592-593 (10.62-10.65)