

# Statistics 514: Design of Experiments

## Topic 7a Summary

### Diagnostics

- Assumptions should be
  - known
  - understood
  - disbelieved
  - checked
- Different assumptions lead to different answers
  - Want to verify assumptions or assess how far from assumptions we are
- One-way design (verifiable) main assumptions:
  - independence
  - constant variance
  - normality
- *Consequences*
  - Predictions fine.
  - Inference invalid.

### Steps

- Make assumptions.
- Check assumptions. (visual inspection/leveraged by tests)
- Fix assumptions.

### Residuals

- Estimated residuals are biased.
- Should be checked with histogram, qqplot against predicted values
- Plot against other variables to check for patterns
- Used for tests (Bartlett, Levene, Shapiro-Wilks, etc.)

- In `proc glm`, use `line means type/hovtest = levene(type = abs)` (only with one-way designs)
- For normal tests, use `proc univariate` with `normal` option.

```
proc univariate data = data
    normal; var variable
```

- Deviations from assumptions may be hard to pick apart, but more easily diagnosed with plots.

## Remedies

- Recording extra variables.
- Transformations
  - Can be calculated (if distribution is known)
  - Plot  $\mu_i$  vs.  $s_i^2$
  - General family: Box-Cox (use  $SSE$  to pick  $\lambda$ )
  - See page 20 of Topic 2 of `Linear Models Notes` for more details on Box-Cox.

## Transformations

- Comparison of means of transformed data is not same as comparison of means at original scale.

## Remedies

- Non-parametric and alternative tests
  - Rank-based methods, e.g., Kruskal-Wallis
  - Adjust df (can be fraction)
  - Brown-Forsythe – analogue of non-pooled  $t$ -test
  - Might lose power.
- More data