

---

# Estimation of Mean Contamination Using Bootstrap and PPCC

---

Thomas F. Reid  
University of South Carolina Aiken  
thomasr@usca.edu

This work was supported by a Cooperative Agreement between the South Carolina Universities Research and Education Foundation and Savannah River DOE. Contract #DE-FC09-00SR22184

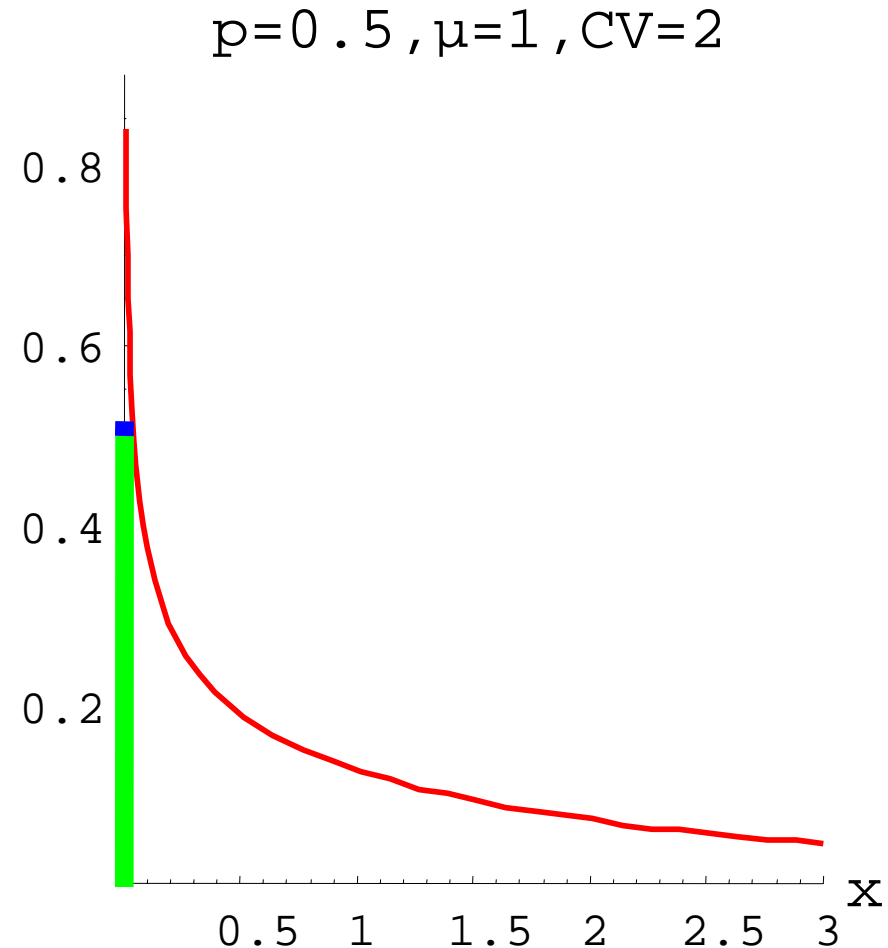
---

# Introduction

- Cleanup sites need to be proven “safe”
- Contaminated and uncontaminated regions
  - Mixed distribution, skewed to the right
- Use of “composite testing” to reduce expense
  - C “raw” samples per composite sample
  - N composite samples
- Problem:
  - due to skewness, standard *small sample* techniques do not perform well

# Mixed Contamination Distribution

- Proportion of site that is uncontaminated:  $p$
- Contaminated portion follows a Gamma distribution
- Unable to detect values below the quantization limit (QL)
  - All values below QL are recorded as  $QL/2$



---

# Max Test

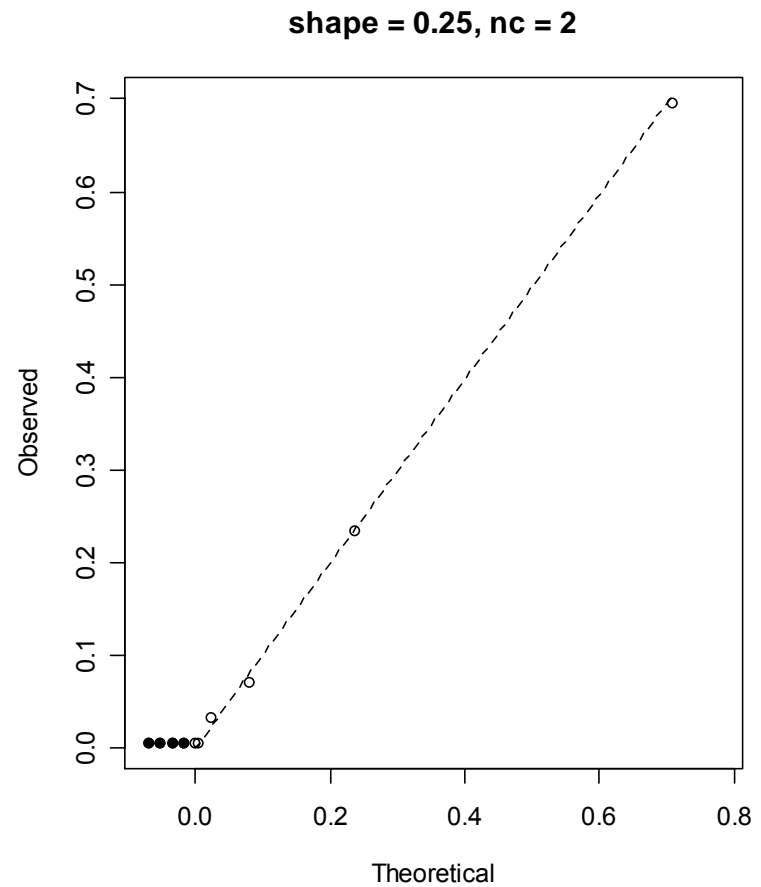
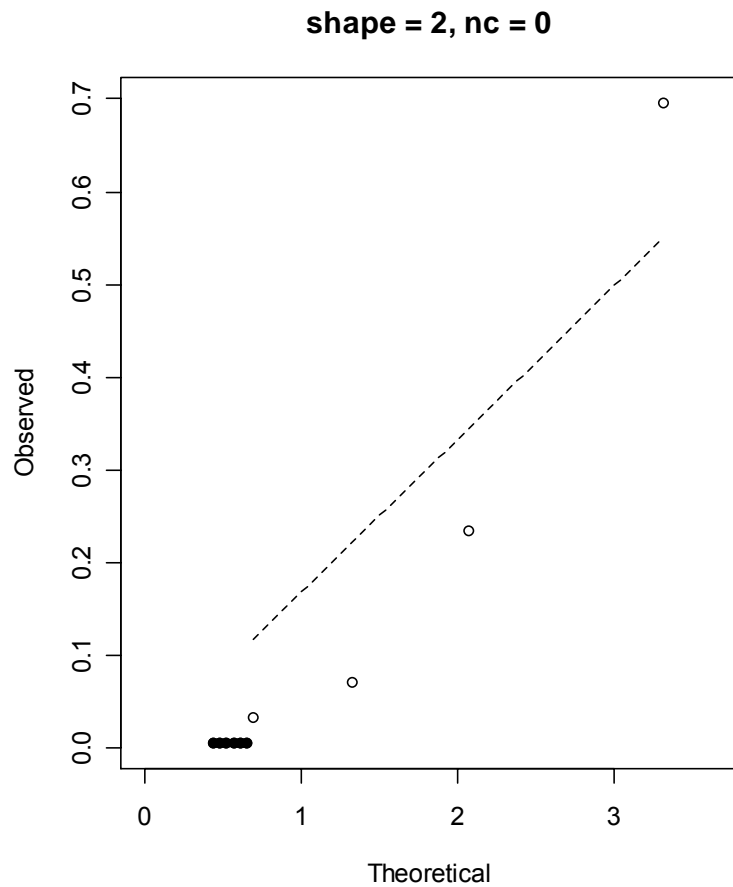
- Recommended by EPA Soil Screening Guidance for Radionuclides
- Site is “safe” if max observed value of composite samples is less than 2 times the (conservative) “soil screening level”
- Using simulation results, selected N and C so that the desired significance level is achieved (approximately)

---

# Probability Plot Correlation Coefficient (PPCC)

- Use Gamma to approximate the continuous portion of the sample distribution (i.e., contamination)
- Let  $n_{low}$  denote the number of observations below the QL. For  $n_c=0,1,\dots,n_{low}$ :
  - find shape ( $\alpha$ ) that maximizes the PPCC for fixed scale ( $\beta=1$ ) when  $n_c$  of the low readings are assumed to be from the contaminated distribution
  - choose  $n_c$  that gives maximum PPCC
- Jointly estimate  $p$ , shape, and scale

# PPCC Illustration



# Estimating Mean Contamination

## ■ Possible Estimators

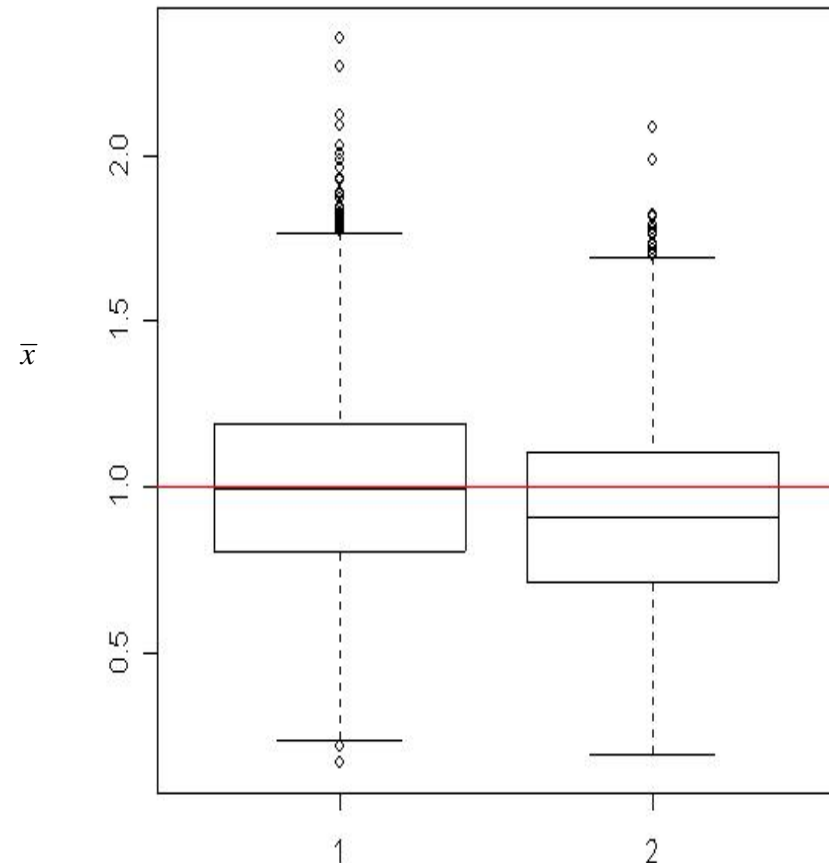
- Sample Mean ( $\bar{x}$ )
- Approx. PPCC Estimator:

$$p(QL/2) + (1-p)(\alpha\beta)$$

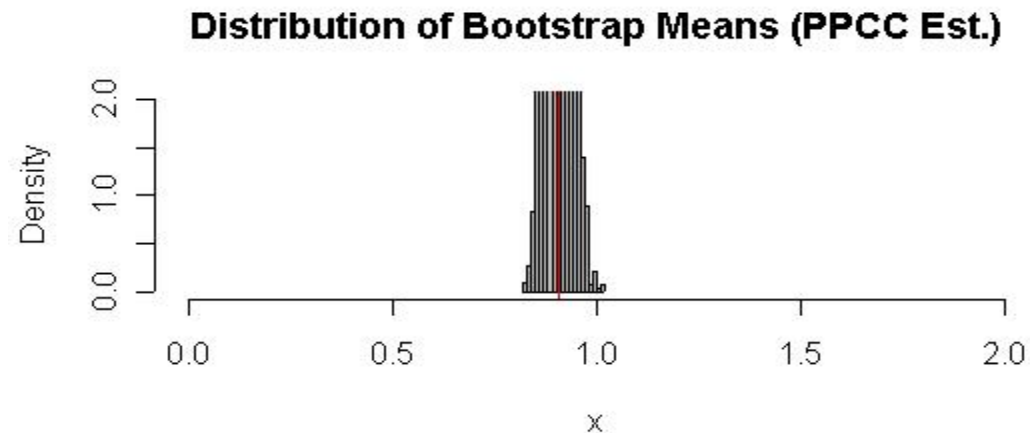
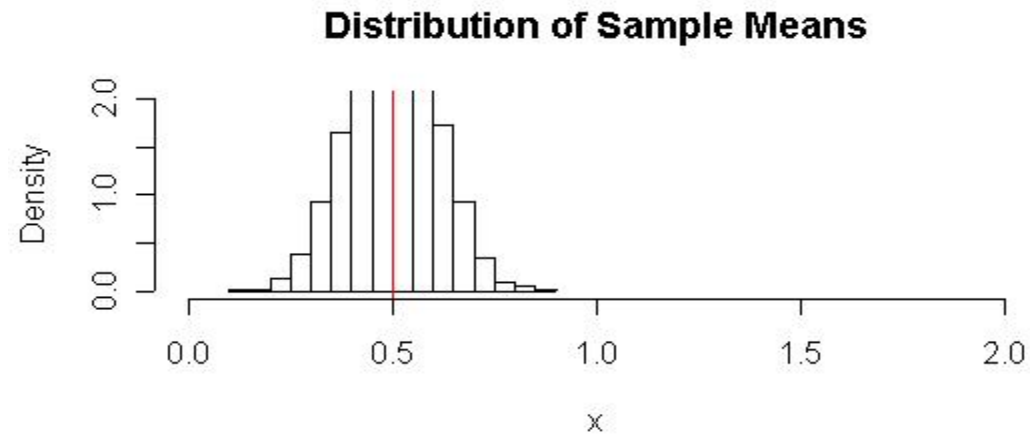
where  $p$ ,  $\alpha$ , and  $\beta$  are all estimated using PPCC

- Mixed Estimator

$$p(QL/2) + (1-p)(\bar{x}_c)$$



# CI Estimation – Not Promising!



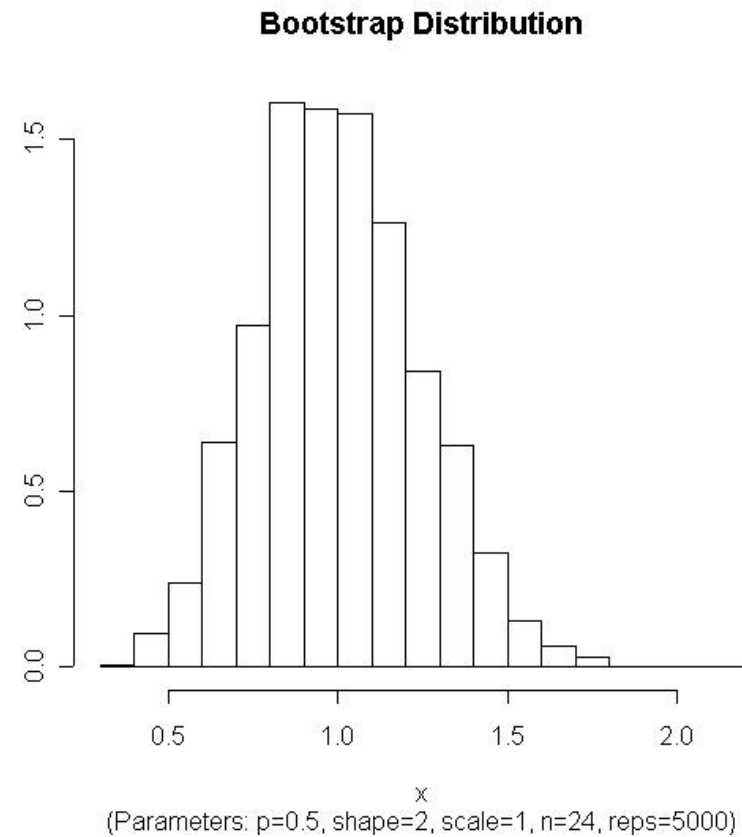
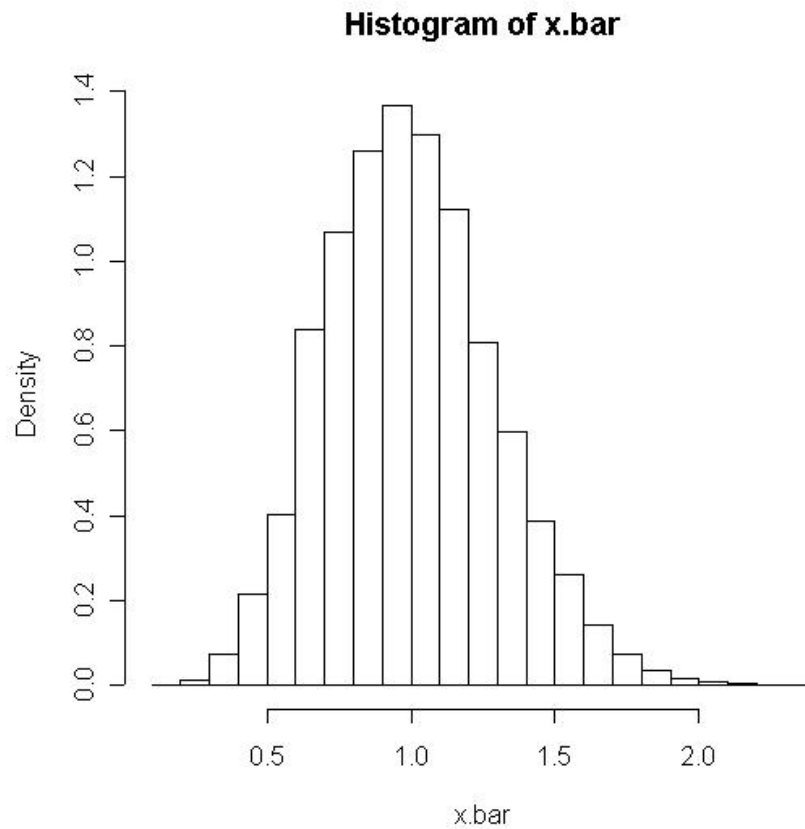
(Parameters:  $p=0.5$ ,  $\text{shape}=99.5$ ,  $\text{scale}=0.01$ ,  $n=24$ ,  $\text{reps}=5000$ ,  $\text{MU}=0.5$ )

---

# PPCC $p$ -Value Test

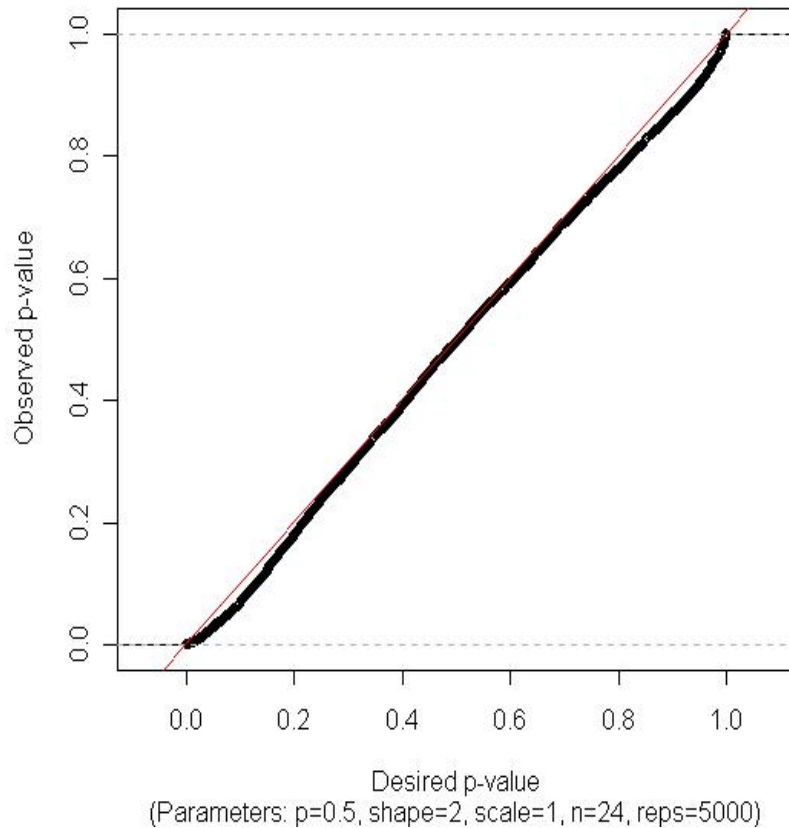
- Test of hypothesis approach
  - Use PPCC to estimate  $p$  and **shape**
  - Choose **scale** to get a mean for the mixed distribution of  $\mu_0$
  - Parametric bootstrap using **estimated parameters**
  - Empirical  $p$ -value: proportion of bootstrap replications with sample mean less than the observed sample mean
  - Reject  $H_0$  if  $p$ -value less than significance level (i.e., conclude site is “safe”)

# Approximate $p$ -value from Bootstrap

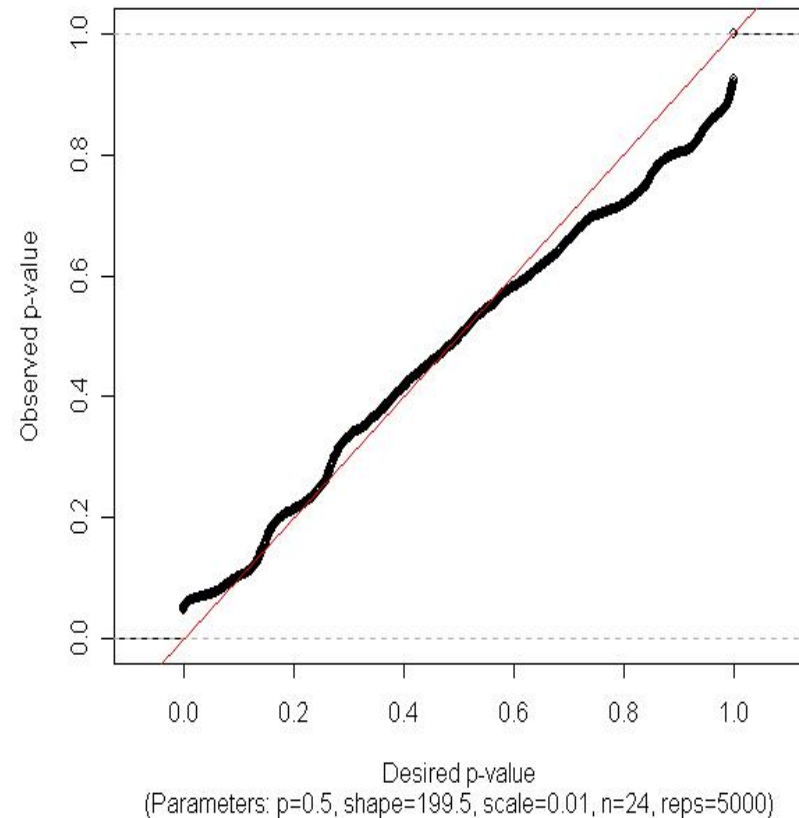


# Empirical $p$ -value Performance

Observed vs. Desired  $p$ -value



Observed vs. Desired  $p$ -value

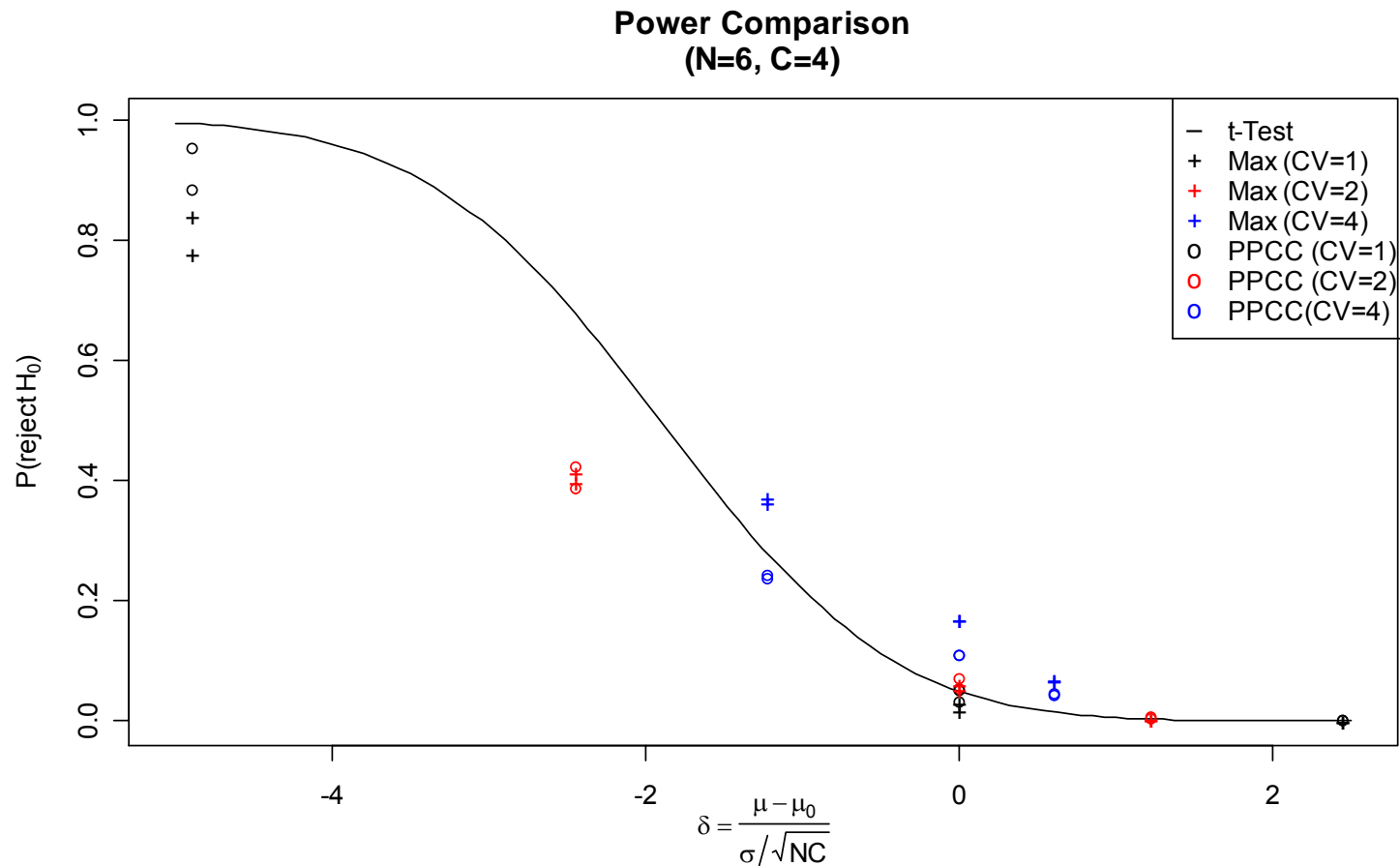


# Empirical $p$ -value Performance

( $\mu = 1$ , nominal significance=0.05)

CV	90% contaminated ( $p=0.1$ )	50% contaminated ( $p=0.5$ )
1	0.032	0.0497
2	0.0543	0.0715
4	0.1093	0.1084

# Power Comparison with Max Test



---

# Conclusions

## ■ PPCC $p$ -Value Test

- ❑ Achieved significance level is close to “desired”
- ❑ Appears to be more robust to changing CV than the Max Test
- ❑ Can be used with any sample size, not just those specified by EPA SSG

## ■ Future Research

- ❑ Contamination distributions other than Gamma
- ❑ Impact of measurement error on performance