# Quiz Practice Questions 3 (Attendance 6) for Statistics 503
## Introduction to Statistics
## Material Covered: Sections 4.1–4.3 Rao and Kuhn

These are practice questions for the quiz. The quiz (not the practice questions) is worth 5% and marked out of 5 points. One or more questions is closely, but not necessarily exactly, related to one or more of these questions will appear on the quiz. These practice questions are *not* to be handed in. Quizzes are to be done *using Vista* on the Internet **before** 4am of the date of the quiz. Vista will *not* allow any quiz to be done late. It is *highly* recommended that you complete this practice quiz, by hand, *before* logging onto Vista. The quiz is an **individual** one which means that each student does this quiz by themselves without help from others. Also check out *previous* quizzes given at

http://www.pnc.edu/faculty/jkuhn/courses/previous/quizzes/quizzes.html

**1.** Statistical Research Methods in the Life Sciences (Rao) Questions.

| Section | Exercise(s) | hints |
|---|---|---|
| 4.1, pages 128–129 | (4.1) | (a) interval (ratio); (b) ordinal; (c) nominal; (d) nominal (e) interval (ratio); (f) ordinal; (f) nominal |
| 4.2, pages 144–149 | (4.3) | see below |
| | (4.5) | look below |
| | (4.7) | look below |
| | (4.11) | look below |
| Summary of quiz material | | look below |

**(4.3)** rat behavior

- **(a)** draw a picture of a normal curve, with (unknown) mean $\mu$ and $\sigma = 4$
  in other words, this Exercise *assumes* a normal distribution
- **(b)** $100\pi\%$ of data is in interval $(\mu - 2, \mu + 2)$
- **(c)** $\pi = 0.85$
  since $\mu \pm \sigma/\sqrt{n} = \mu \pm 4/\sqrt{10} = \mu \pm 1.26$,
  2 units is equal to $2/1.26 = 1.58$ standard units from mean,
  so $P(-1.58 < T < 1.58) = 0.85$ (tcdf(-1.58,1.58,9))
  use t–distribution since sample size, 10, is smaller than 30
- **(d)** CI and test
  - **(i)** $\overline{Y} \pm t(n-1, \alpha/2)\sigma_{\overline{Y}} = 15.7 \pm (2.26)(0.89) = (13.44, 17.96)$
    use STATS TESTS TINTERVAL with data typed in $L_1$
  - **(ii)** statement: $H_o : \mu = 18$ vs $H_a : \mu < 18$
    test: $P(T < (15.7 - 18)/(4/\sqrt{10})) = 0.0511$ (tcdf$(-E99, (15.7 - 18)/(4/\sqrt{10}), 9)$), $\alpha = 0.05$
    conclusions: since $0.0511 > 0.05$, accept null (rats need 18 trials)
    use STATS TESTS T–Test with data typed in $L_1$

**(4.5)** social perceptiveness

- **(a)** small random sample from normal distribution
- **(b)** $\overline{Y} \pm t(n-1, \alpha/2)\sigma_{\overline{Y}} = (57.07, 64.31)$
  use STATS TESTS TINTERVAL with data typed in $L_1$
- **(c)** 90% *confident* index in (57.07, 64.31)
- **(d)** statement: $H_o : \mu = 50$ vs $H_a : \mu > 50$
  test: p–value $= 0.000056$, $\alpha = 0.10$
  conclusions: since $0.000056 < 0.10$, reject null (index greater than 50)
  use STATS TESTS T–Test with data typed in $L_1$
- **(e)** The small p–value (0.0056%) says there is essentially no chance of observing the sample average or more, assuming the index is equal to 50; that is, it implies the assumed index value of 50 must be incorrect, that is must be larger than 50.
- **(f)** Since the observed CI $(57.07, 64.31)$ does not include the assumed index value of 50, this implies 50 is incorrect. Since this 90% CI has 5% in each tail, this CI would be equivalent to performing either a *two–sided* statistical test at $\alpha = 10\%$ or a *upper–sided* statistical test at $\alpha = 5\%$.

**(4.7)** N–losses

**(a)** $(1.134, \infty)$
use STATS TESTS 2–SampTInt (yes, pool) with data typed in $L_1$ and $L_2$
use 0.90, rather than 0.95 and only use the lower bound (why?)

**(b)** 95% *confident* difference in N–losses greater than 1.134

**(c)** since $\bar{Y}_1 - \bar{Y}_2 = 2.5621$
and $\sigma_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{4.5784}{15} + \frac{5.2364}{13}} = 0.8414$ and

$$\begin{aligned}
\frac{1}{k} &= \frac{1}{n_1 - 1}\left[\frac{S_1^2/n_1}{S_1^2/n_1 + S_2^2/n_2}\right]^2 + \frac{1}{n_2 - 1}\left[\frac{S_2^2/n_2}{S_1^2/n_1 + S_2^2/n_2}\right]^2 \\
&= \frac{1}{12 - 1}\left[\frac{4.5784/15}{4.5784/15 + 5.2364/13}\right]^2 + \frac{1}{12 - 1}\left[\frac{5.2364/13}{4.5784/15 + 5.2364/13}\right]^2 \\
&= 0.0403
\end{aligned}$$

and so $1/k = 1/0.0403 = 24.83$ implies $\nu = 24$ and so

$$\begin{aligned}
\bar{Y}_1 - \bar{Y}_2 - t(\nu, \alpha/2)\sigma_{\bar{Y}_1 - \bar{Y}_2} &= 2.5621 - t(24, 0.05)(0.8414) \\
&= 2.5621 - (1.71)(0.84) \\
&= 1.1225
\end{aligned}$$

and so a 95% lower bound CI is $(1.1225, \infty)$

**(d)** the first lower bound CI, $(1.32, \infty)$, is used if the variances are known,
the second lower bound CI, $(1.134, \infty)$, is used if the variances are unknown but assumed to be equal,
the third lower bound CI, $(1.1225, \infty)$, is used if the variances are unknown and assumed to be unequal,

**(4.11)** PVR, sheep

- **(a)** statement: $H_o : \mu_d = \mu_1 - \mu_2 = 0$ vs $H_a : \mu_d = \mu_1 - \mu_2 < 0$
  test: p–value $= 0.051$, $\alpha = 0.01$
  conclusions: since $0.051 > 0.010$, accept null (PVR same, before and after)
  use STATS TESTS T–Test with *difference* data in $L_3 = L_1 - L_2$
- **(b)** $(-0.0761, 0.00859)$
  use STATS TInterval with difference data

In addition to discussing the different types of data

- nominal
- ordinal
- interval
- ratio

we looked at examples of the hypothesis test and confidence interval when $\hat{\theta}$ is an estimator of $\theta$, where the sampling distribution of $\hat{\theta}$ is normal with mean $\theta$ and *unknown* standard deviation $\hat{\sigma}_{\hat{\theta}}$ but where the random sample is *small*, typically $n < 30$, with (standardized) test statistic,

$$t = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}}$$

with calculated value $t_c$ and where,

| $H_0$ | $H_1$ | reject $H_0$ at level $\alpha$ if |
|---|---|---|
| $\theta \leq \theta_0$ | $\theta > \theta_0$ | $t_c > t(\nu, \alpha)$ |
| $\theta \geq \theta_0$ | $\theta < \theta_0$ | $t_c < -t(\nu, \alpha)$ |
| $\theta = \theta_0$ | $\theta \neq \theta_0$ | $t_c > t(\nu, \alpha/2)$ or $t_c < -t(\nu, \alpha/2)$ |

and where $\left(\hat{\theta}_L, \hat{\theta}_U\right)$ is either a/n

- (two–sided) confidence interval, CI: $\left(\overline{Y} - t(\nu, \alpha/2)\hat{\sigma}_{\hat{\theta}},\ \ \overline{Y} + t(\nu, \alpha/2)\hat{\sigma}_{\hat{\theta}}\right)$
- lower confidence interval, LCI: $\left(\overline{Y} - t(\nu, \alpha)\hat{\sigma}_{\hat{\theta}},\ \ \infty\right)$
- upper confidence interval, UCI: $\left(-\infty,\ \ \overline{Y} + t(\nu, \alpha)\hat{\sigma}_{\hat{\theta}}\right)$

Specifically, we look at tests and confidence intervals for

- $\theta = \mu$, where $\hat{\sigma}_{\hat{\theta}} = \frac{s}{\sqrt{n}}$
  and we assume normality and a small random sample
- $\theta = \mu_1 - \mu_2$, where $\hat{\sigma}_{\hat{\theta}} = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
  and we assume normality, small independent random samples
  and unknown $\sigma_1^2 = \sigma_2^2$
- $\theta = \mu_1 - \mu_2$, where $\hat{\sigma}_{\hat{\theta}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$
  and we assume normality, small independent random samples
  and unknown $\sigma_1^2 \neq \sigma_2^2$
- $\theta = \mu_D = \mu_1 - \mu_2$, where $\hat{\sigma}_{\hat{\theta}} = \frac{S_D}{\sqrt{n}}$
  and we assume normality and small *dependent* (paired) random samples