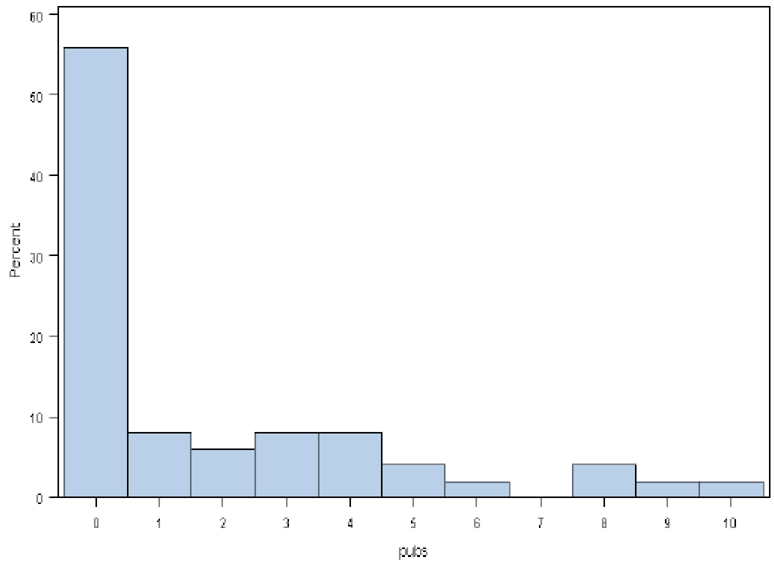
2.6 (a)



The data are quite skewed because 28 faculty have no publications.

(b) $\bar{y} = 1.78$; $s = 2.682$;

$$\text{SE}\,[\bar{y}] = \frac{2.682}{\sqrt{50}}\sqrt{1 - \frac{50}{807}} = 0.367.$$

(c) No; a sample of size 50 is probably not large enough for $\bar{y}$ to be normally distributed, because of the skewness of the original data.

The sample skewness of the data is (from SAS) 1.593. This can be calculated by hand, finding

$$\frac{1}{n}\sum_{i \in \mathcal{S}}(y_i - \bar{y})^3 = 28.9247040$$

so that the skewness is $28.9247040/(2.682^3) = 1.499314$. Note this estimate differs from SAS PROC UNIVARIATE since SAS adjusts for df using the formula skewness $= \dfrac{n}{(n-1)(n-2)}\sum_{i \in \mathcal{S}}(y_i - \bar{y})^3/s^3$. Whichever estimate is used, however, formula (2.23) says we need a minimum of

$$28 + 25(1.5)^2 = 84$$

observations to use the central limit theorem.

(d) $\hat{p} = 28/50 = 0.56$.

$$\text{SE}\,(\hat{p}) = \sqrt{\frac{(0.56)(0.44)}{49}\left(1 - \frac{50}{807}\right)} = 0.0687.$$

A 95% confidence interval is

$$0.56 \pm 1.96(0.0687) = [0.425, 0.695].$$

2.20 Sixty of the 70 samples yield confidence intervals, using this procedure, that include the true value $t = 40$. The exact confidence level is $60/70 = 0.857$.

2.22 (a) From (2.13),

$$\text{CV}(\bar{y}) = \frac{\sqrt{V(\bar{y})}}{E(\bar{y})} = \sqrt{1 - \frac{n}{N}} \frac{S}{\sqrt{n}\bar{y}_U}.$$

Substituting $\hat{p}$ for $\bar{y}$, and $\frac{N}{N-1}p(1-p)$ for S^2, we have

$$\text{CV}(\hat{p}) == \sqrt{\left(1 - \frac{n}{N}\right) \frac{Np(1-p)}{(N-1)np^2}} = \sqrt{\frac{N-n}{N-1} \frac{1-p}{np}}.$$

The CV for a sample of size 1 is $\sqrt{(1-p)/p}$. The sample size in (2.26) will be $z^2_{\alpha/2}\text{CV}^2/r^2$.

(b) I used Excel to calculate these values.

p	0.001	0.005	0.01	0.05	0.1	0.3	0.5
Fixed	4.3	21.2	42.3	202.8	384.2	896.4	1067.1
Relative	4264176	849420	422576	81100	38416	9959.7	4268.4

p	0.7	0.9	0.95	0.99	0.995	0.999
Fixed	896.4	384.2	202.8	42.3	21.2	4.3
Relative	1829.3	474.3	224.7	43.1	21.4	4.3

2.28 (a) We can think of drawing a simple random sample with replacement as performing an experiment n independent times; on each trial, outcome i (for $i \in \{1, \ldots, N\}$) occurs with probability $p_i = 1/N$. This describes a multinomial experiment.

We may then use properties of the multinomial distribution to answer parts (b) and (c):

$$E[Q_i] = np_i = \frac{n}{N},$$
$$V[Q_i] = np_i(1-p_i) = \frac{n}{N}\left(1 - \frac{1}{N}\right),$$

and

$$\text{Cov}\,[Q_i, Q_j] = -np_ip_j = -\frac{n}{N}\frac{1}{N} \quad \text{for} \quad i \neq j.$$

(b)

$$E[\hat{t}] = \frac{N}{n}E\left[\sum_{i=1}^{N} Q_i y_i\right] = \frac{N}{n}\sum_{i=1}^{N}\frac{n}{N}y_i = t.$$

(c)

$$\begin{aligned}
V[\hat{t}] &= \left(\frac{N}{n}\right)^2 V\left[\sum_{i=1}^{N} Q_i y_i\right] \\
&= \left(\frac{N}{n}\right)^2 \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \text{ Cov}\,[Q_{ij}\ \ Q_j] \\
&= \left(\frac{N}{n}\right)^2 \left\{\sum_{i=1}^{N} y_i^2 np_i(1-p_i) + \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j(-np_ip_j)\right\} \\
&= \left(\frac{N}{n}\right)^2 \left\{\frac{n}{N}\left(1-\frac{1}{N}\right)\sum_{i=1}^{N} y_i^2 - \frac{n}{N}\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i} y_i y_j\right\} \\
&= \frac{N}{n}\left\{\sum_{i=1}^{N} y_i^2 - N\bar{y}_U^2\right\} \\
&= \frac{N^2}{n}\frac{\sum\limits_{i=1}^{N}(y_i - \bar{y}_U)^2}{N}.
\end{aligned}$$