3 Resampling Methods: The Jackknife

3.1 Introduction

- In this section, much of the content is a summary of material from Efron and Tibshirani (1993) and Manly (2007). Here are several useful reference texts on resampling methods.

   1. Davison and Hinkley (1997) *Bootstrapping and its Applications*, Cambridge University Press.
   2. Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM.
   3. Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*, Chapman & Hall.
   4. Manly, Bryan F.J. (2007) *Randomization, Bootstrap, and Monte Carlo Methods in Biology*, Chapman & Hall/CRC Press.
   5. Shao, J. and Tu, T. (1995) *The Jackknife and Bootstrap*, Springer-Verlag.

- In statistical inference, the goal is often the estimation of parameters or some characteristics of a probability distribution f based on a random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of size n taken from f.

- The **empirical distribution function** $\widehat{f}$ is the discrete distribution that assigns probability $1/n$ to each value x_i $(i = 1, 2, \ldots, n)$ in the sample. Note: this is <u>not</u> a cumulative distribution function.

- Note: the x_i values do not have to be unique. That is, x_i could equal x_j for some $i \neq j$.

- If there are repeated data values, we can express $\widehat{f}$ as a vector of observed frequencies $\widehat{f}_k$ for each observed value k. That is,

$$\widehat{f}_k = \frac{\text{Number of cases when } x_i = k}{n}$$

- Therefore, $\widehat{f}$ assigns to any set A the **empirical probability**

$$\widehat{P}(\{A\}) = \frac{\text{Number of cases when } x_i \in A}{n}$$

- Table 3.1 contains data from a population of 82 pairs of average LSAT scores and average GPA of 82 American law schools who participated in a study of admission practices. The LSAT is an exam required for admittance into a US law school. The GPA is the student grade point average. A random sample of 15 pairs of LSAT scores and GPA was selected from the 82 schools. The sampled schools are in parentheses ().

- Figure 3.1 contains scatterplots of the the average LSAT scores versus the average GPA scores for (i) the population and (ii) the sample.

   - The empirical distribution $\widehat{f}$ assigns probability $1/15$ to each of the 15 data points. Consider the set $A = \{(y, z) : y < 600, z < 3.00\}$ where y is the LSAT score and z is the GPA. Then the empirical probability $\widehat{P}(\{A\}) = 5/15$ because 5 of the 15 schools are elements of the set A.

   - In statistical inference, we want to make inferences about the population shown in the left-side plot based on the sample shown in the right-side plot (Figure 3.1).

   - Specifically, we may want to make inferences about the population correlation coefficient based on the sample correlation coefficient.

school	LSAT	GPA	school	LSAT	GPA	school	LSAT	GPA	school	LSAT	GPA
1	622	3.23	22	614	3.19	43	573	2.85	63	572	3.08
2	542	2.83	23	628	3.03	44	644	3.38	64	610	3.13
3	579	3.24	24	575	3.01	(45)	545	2.76	65	562	3.01
(4)	653	3.12	25	662	3.39	46	645	3.27	66	635	3.30
5	606	3.09	26	627	3.41	(47)	651	3.36	67	614	3.15
(6)	576	3.39	27	608	3.04	48	562	3.19	68	546	2.82
7	620	3.10	28	632	3.29	49	609	3.17	69	598	3.20
8	615	3.40	29	587	3.16	(50)	555	3.00	(70)	666	3.44
9	553	2.97	30	581	3.17	51	586	3.11	71	570	3.01
10	607	2.91	(31)	605	3.13	(52)	580	3.07	72	570	2.92
11	558	3.11	32	704	3.36	(53)	594	2.96	73	605	3.45
12	596	3.24	33	477	2.57	54	594	3.05	74	565	3.15
(13)	635	3.30	34	591	3.02	55	560	2.93	75	686	3.50
14	581	3.22	(35)	578	3.03	56	641	3.28	76	608	3.16
(15)	661	3.43	(36)	572	2.88	57	512	3.01	77	595	3.19
16	547	2.91	37	615	3.37	58	631	3.21	78	590	3.15
17	599	3.23	38	606	3.20	59	597	3.32	(79)	558	2.81
18	646	3.47	39	603	3.23	60	621	3.24	80	611	3.16
19	622	3.15	40	535	2.98	61	617	3.03	81	564	3.02
20	611	3.33	41	595	3.11	62	637	3.33	(82)	575	2.74
21	546	2.99	42	575	2.92						

Sampled schools have bold-faced school numbers.

Figure 3.1

Law School Data Scatterplots: Population and Sample
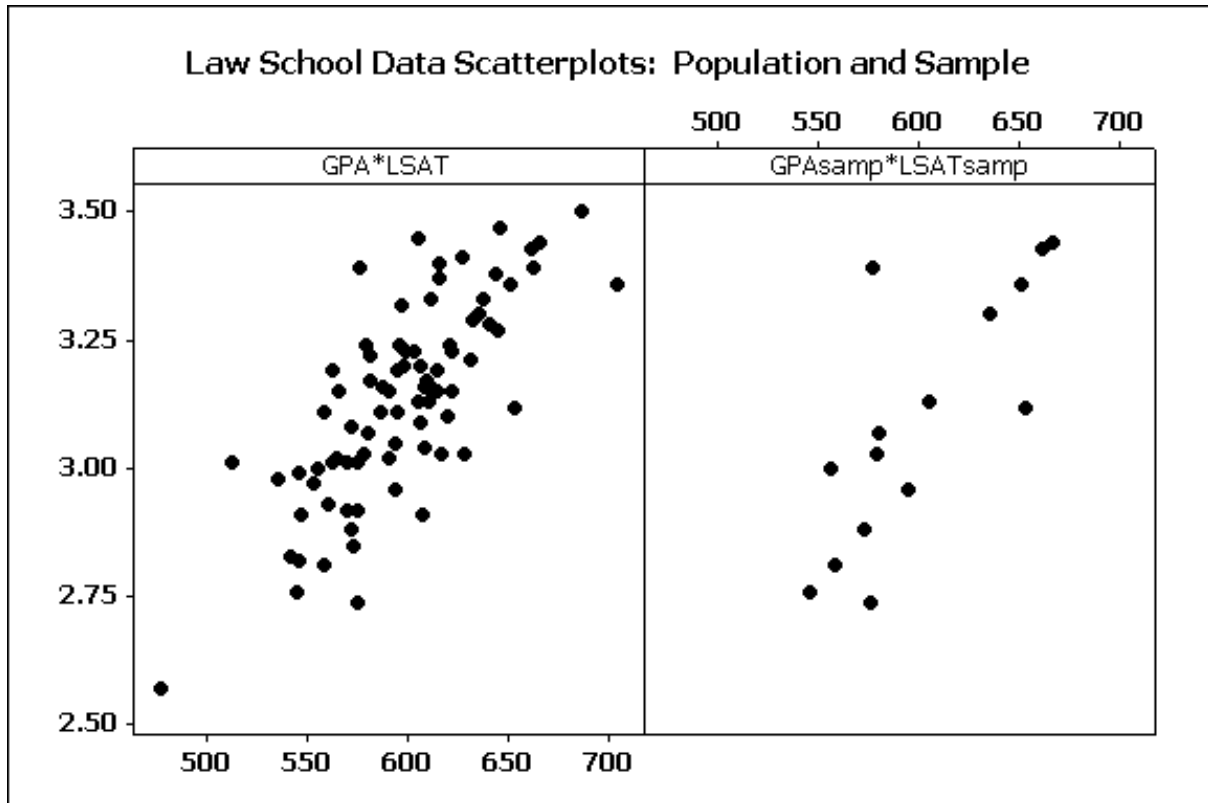
- Consider the following random sample of 50 values:

$$
\begin{array}{cccccccccccccccccccccccc}
2 & 0 & 1 & 1 & 3 & 0 & 1 & 2 & 1 & 3 & 1 & 2 & 0 & 1 & 1 & 2 & 2 & 0 & 1 & 2 & 0 & 1 & 3 & 2 & 1 \\
1 & 0 & 2 & 1 & 0 & 1 & 1 & 3 & 2 & 0 & 1 & 2 & 3 & 1 & 1 & 3 & 0 & 0 & 1 & 1 & 2 & 0 & 1 & 3 & 0
\end{array}
$$

  The empirical distribution $\widehat{f}$ would assign probability $1/50$ to each of the 50 data values. But, for a random sample, we can summarize $\widehat{f}$ by

$$
(\widehat{f}_1, \widehat{f}_2, \widehat{f}_3, \widehat{f}_4) \;=\; \left(\frac{12}{50}, \frac{20}{50}, \frac{11}{50}, \frac{7}{50} \right) \;=\; (.24, .40, .22, .14)
$$

  where $\widehat{f}_i$ = the proportion of x_i values $= 0, 1, 2,$ and 3, respectively for $i = 1, 2, 3, 4$.

- Mathematical statistics results: It can be proved that the vector of observed frequencies $\mathbf{f} = (\widehat{f}_1, \widehat{f}_2, \widehat{f}_3, \ldots)$ is a sufficient statistic for the true distribution f. That is, all of the information about f that is contained in the sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is also contained in $\mathbf{f} = (\widehat{f}_1, \widehat{f}_2, \widehat{f}_3, \ldots)$.

  - Remember that the sufficiency assumes that the data set has been generated by random sampling. For other data sets (such as time series data) this will not be true.

  - We can also write a parameter θ directly as a function of f. That is, $\theta = t(F)$. For example, μ_X = the expectation of a random variable X can be written as the parameter $\theta = \mu_X = E_f(X) = t(f)$.

- A **parameter** is a function of a probability distribution F. A **statistic** is a function of a sample $\mathbf{x}$.

  - For the law school data, the parameters $\mu_y = 597.55$, $\mu_z = 3.135$, and $\rho = .761$. Based on the sample of size $n = 15$, the corresponding statistics are $\bar{y} = 600.3$, $\bar{z} = 3.095$, and Pearson's sample correlation coefficient $r = .776$.

3.2 The Jackknife Method

- Let θ be some parameter associated with a distribution f. Suppose We have a random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and an estimator $\widehat{\theta}$ of θ, and the goal is to estimate the <u>bias</u> and <u>standard error</u> of $\widehat{\theta}$.

- Let $\mathbf{x}_{(i)}$ be the sample but with the ith observation removed:

$$
\mathbf{x}_{(i)} \;=\; (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \tag{1}
$$

  $\mathbf{x}_{(i)}$ is the ith **jackknife sample**.

- The **jackknife method** of estimation is based on the n jackknife samples $\mathbf{x}_{(i)}$ $(i = 1, 2, \ldots, n)$.

- Suppose $\widehat{\theta} = g(\mathbf{x})$ for some function g of the data. The ith **jackknife replication** of $\widehat{\theta}$ is

$$
\widehat{\theta}_{(i)} \;=\; g(\mathbf{x}_{(i)}) \tag{2}
$$

- For example, consider the following example from Manly (2007) summarized in Table 3.2. The goal is to estimate the population standard deviation. That is, $\theta = \sigma$.

  - The data set has $n = 20$ values given in the row labeled **Data**. Manly (2007) calculated the square root of the biased MLE estimator of the variance having n instead of $n - 1$ in the denominator:

$$\widehat{\theta} = \widehat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{20} (x_i - \overline{x})^2} = 1.03$$

  (and not the sample standard deviation s which has $(n - 1) = 19$ df).

  - Rows labeled 1 to 20 have one value removed, leaving 19 remaining values. The 19 values in columns 1 to 20 form the 20 *jackknife samples*.

  - The ith *jackknife replication* is

$$\sigma_i = \sqrt{\frac{1}{19} \widehat{\sum_{j \neq i} (x_j - \overline{x})^2}}$$

  which is calculated from the 19 values (without x_i) in the ith jackknife sample. The values are given in the first **SD** column.

3.2.1   Jackknife Bias Estimation

- Let $\widehat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \widehat{\theta}_{(i)}/n$. The **jackknife estimate of bias** is defined as

$$\widehat{\text{bias}}(\widehat{\theta}) \;=\; (n - 1)(\widehat{\theta}_{(\cdot)} - \widehat{\theta}) \tag{3}$$

- The **bias-corrected jackknife estimate** of θ is

$$\widehat{\theta}_{\text{jack}} \;=\; \widehat{\theta} - \widehat{\text{bias}}(\widehat{\theta}) \;=\; n\widehat{\theta} - (n - 1)\widehat{\theta}_{(\cdot)}$$

- Mathematically, it can be shown that $\widehat{\text{bias}}(\widehat{\theta})$ is an unbiased estimator of the true bias for many statistics.

- For other statistics, although $\widehat{\theta}_{\text{jack}}$ is a biased estimator of the true bias, the bias of $\widehat{\theta}_{\text{jack}}$ is reduced in comparison to the unadjusted estimate $\widehat{\theta}$.

Table 3.2. Jackknife Calculations with the Estimation of the Population Standard Deviation from a Sample of Size 20

Data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	SD	PV	Log SD	PV
	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.03 ←		0.032	
1		0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	0.879	3.959	-0.129	3.100
2	3.56		0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.056	0.586	0.055	-0.396
3	3.56	0.69		1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.036	0.971	0.035	-0.027
4	3.56	0.69	0.10		3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.043	0.840	0.042	-0.154
5	3.56	0.69	0.10	1.84		1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	0.813	5.202	-0.207	4.570
6	3.56	0.69	0.10	1.84	3.93		0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.059	0.544	0.057	-0.435
7	3.56	0.69	0.10	1.84	3.93	1.25		1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.040	0.898	0.039	-0.098
8	3.56	0.69	0.10	1.84	3.93	1.25	0.18		0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.059	0.527	0.058	-0.452
9	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13		0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.044	0.823	0.043	-0.170
10	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27		0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.052	0.671	0.051	-0.315
11	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50		0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.056	0.593	0.054	-0.389
12	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67		0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.031	1.062	0.031	0.060
13	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01		0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.055	0.617	0.053	-0.366
14	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61		1.70	0.39	0.11	1.20	1.21	0.72	1.058	0.548	0.057	-0.431
15	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82		0.39	0.11	1.20	1.21	0.72	1.048	0.738	0.047	-0.251
16	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70		0.11	1.20	1.21	0.72	1.048	0.737	0.047	-0.252
17	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39		1.20	1.21	0.72	1.037	0.962	0.036	-0.037
18	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11		1.21	0.72	1.059	0.535	0.057	-0.444
19	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20		0.72	1.059	0.537	0.057	-0.442
20	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21		1.057	0.575	0.055	-0.405

	SD	Log SD / PV
Jackknife estimate (mean)	1.096 ←	0.133
Jackknife standard error	0.273	0.290
Lower 95% confidence interval	0.526	-0.473
Upper 95% confidence interval	1.666	0.740

$$\hat{\sigma} = 1.03285$$

$$\hat{\sigma}_{(\cdot)} = 1.02952$$

$$\widehat{BIAS} = 19(1.02952 - 1.03285) = -.06327$$

$$\Rightarrow \hat{\sigma}_{JACK} = \hat{\sigma} - \widehat{BIAS} = 1.03285 - (-.06327) = 1.09612$$

3.2.2 Jackknife Standard Error Estimation

- Consider what we usually do when estimating a mean of a distribution. We use the sample mean $\overline{X} = \sum\limits_{i=1}^{n} X_i$ as our estimator.

- To see the relationship with jackknife estimation, we can write the mean with the ith observation removed as $\overline{X}_{(i)}$ as:

$$\overline{X}_{(i)} \;=\; \frac{\left(\sum_{j=1}^{n} X_j\right) - X_i}{n-1}$$

- Therefore an individual X_i can be written as

$$X_i \;=\; \left(\sum_{j=1}^{n} X_j\right) - (n-1)\overline{X}_{(i)} \;=\; n\overline{X} - (n-1)\overline{X}_{(i)}. \tag{4}$$

That is, it is possible to determine an observation from the sample mean and from the sample mean with the ith observation removed.

- We now apply this approach to statistics other than the sample mean. For example, when estimating a parameter θ with estimator $\widehat{\theta}$, if we replace the sample means in (4) with the corresponding estimators of θ, we have

$$\text{PV}(\mathbf{x}_{(i)}) \;=\; n\widehat{\theta} - (n-1)\widehat{\theta}_{(i)}$$

- $\text{PV}(\mathbf{x}_{(i)})$ is called the ith **pseudo-value**.

- Note that we expect that $\text{PV}(\mathbf{x}_{(i)}) \approx n\theta - (n-1)\theta = \theta$. Thus, each pseudo-value can be viewed as an estimate of θ, and the average of the pseudo-values is

$$\begin{aligned}
\overline{\text{PV}} \;&=\; \frac{1}{n}\sum_{i=1}^{n}\left(n\widehat{\theta} - (n-1)\widehat{\theta}_{(i)}\right) \\
&=\; \sum_{i=1}^{n}\widehat{\theta} \;+\; (n-1)\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\theta}_{(i)}\right) \;=\; n\widehat{\theta} - (n-1)\widehat{\theta}_{(\cdot)}
\end{aligned}$$

which is the **bias-corrected jackknife estimate**. That is, $\overline{\text{PV}} = \widehat{\theta}_{\text{jack}}$. This suggests the natural estimator for the variance of $\widehat{\theta}_{\text{jack}}$ is s_{jack}^2/n where s_{jack}^2 is the sample variance of the n pseudo-values:

$$\widehat{\text{var}}(\widehat{\theta}_{\text{jack}}) \;=\; \frac{s_{\text{jack}}^2}{n} \;=\; \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(\text{PV}(\mathbf{x}_{(i)}) - \overline{\text{PV}}\right)^2. \tag{5}$$

and the **jackknife standard error** of $\widehat{\theta}_{\text{jack}}$ is $\quad \text{s.e.}(\widehat{\theta}_{\text{jack}}) = \sqrt{\widehat{\text{var}}(\widehat{\theta}_{\text{jack}})}$.

- In their textbook, Efron and Tibshirani (1993) state that the jackknife standard error is

$$\text{s.e.}(\widehat{\theta}_{\text{jack}}) \;=\; \left[\frac{n-1}{n}\sum_{i=1}^{n}(\widehat{\theta}_{(i)} - \widehat{\theta}_{(\cdot)})^2\right]^{1/2} \;=\; \left[\frac{(n-1)^2}{n} \times (s^2 \text{ for the } \widehat{\theta}_{(i)} \text{ values})\right]^{1/2} \tag{6}$$

It can be shown that the standard errors in (5) and (6) are computationally equivalent.

43

- Look again at the example in Table 3.2. The goal was to estimate the population standard deviation σ. So, in this example, $\theta = \sigma$.

  - The uncorrected estimate $\widehat{\theta} = \widehat{\sigma} = 1.03285 \approx 1.03$. (in the **SD** column of **Data** row).
  - The 20 sample the jackknife replications ($\widehat{\sigma}_{(i)}$ values) appear in the **SD** column.
  - The 20 pseudo-values (the $\text{PV}(\mathbf{x}_{(i)})$ values) appear in the $\mathbf{PV}_i$ column.
  - The average of the pseudo-values $\overline{\text{PV}} = \widehat{\sigma}_{jack} = 1.09616 \approx 1.096$.
  - Therefore, $\widehat{\text{bias}}(\widehat{\sigma}) \approx (19)(1.02952 - 1.03285) = -.06327$ (slight round-off error).
  - Or, you can estimate the bias using
    $$\widehat{\text{bias}}(\widehat{\sigma}) = \widehat{\theta} - \widehat{\theta}_{\text{jack}} \approx 1.03285 - 1.09616 \approx -.06331 \text{ (more accurate)}.$$

- Suppose we want to apply the jackknife method to 3 other situations:

  1. Suppose we wanted to estimate $\theta = \mu$? What would happen if we use the sample mean $(\widehat{\theta} = \overline{X})$?

  2. Suppose we wanted to estimate $\theta = \sigma$? What would happen if we use the sample standard deviation $(\widehat{\theta} = s)$?

  3. Suppose we wanted to estimate $\theta = \sigma^2$? What would happen if we use the sample standard variance $(\widehat{\theta} = s^2)$?

  The results are summarized in the following table.

Parameter	Statistic	$\widehat{\theta}_{(\cdot)}$	$\widehat{\theta}$	$\widehat{\text{bias}}(\widehat{\theta})$	$\widehat{\theta}_{\text{jack}}$	s.e.$(\widehat{\theta}_{\text{jack}})$
μ	$\overline{x}$	1.04450	1.04450	0	1.04450	0.23695
σ	s	1.05773	1.05968	-0.03710	1.09678	0.28028
σ	$\sqrt{\widehat{\sigma}^2_{\text{MLE}}}$	1.02952	1.03285	-0.06331	1.09616	0.27280
σ^2	s^2	1.12292	1.12292	0	1.12292	0.54424

- There are several important things to notice in these results.

  1. We know $\overline{x}$ and s^2 are unbiased estimators of μ and σ^2. Therefore, applying the jackknife method to an unbiased estimator has no effect. That is, $\widehat{\theta}_{\text{jack}} = \widehat{\theta}$, and the estimated bias will then be 0.

  2. Both bias-corrected estimates of σ are still biased, but the bias is reduced compared to the original estimates (statistics). After bias correction, the two $\widehat{\theta}_{\text{jack}}$ values for estimating σ are very close.

  3. The standard errors can then be used to calculate t-based confidence intervals:

    $$\widehat{\theta}_{\text{jack}} \pm t^* \text{ s.e.}(\widehat{\theta}_{\text{jack}})$$

    where t^* is the $1 - \alpha/2$ critical value from a t-distribution with $n - 1$ degrees of freedom.

```
                    Theta(hat) values for 4 different estimates

                    mean        s        s^2        s_mle
                  1.0445    1.05968    1.12292    1.03285

        Jackknife Samples and Jackknife Replications (Table 3.2 data)

Obs  y1    y2   y3  y4   y5   y6   y7   y8   y9  y10  y11  y12  y13  y14 y15

  1  .    0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
  2 3.56   .   0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
  3 3.56 0.69   .  1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
  4 3.56 0.69 0.1   .   3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
  5 3.56 0.69 0.1 1.84   .   1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
  6 3.56 0.69 0.1 1.84 3.93   .   0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
  7 3.56 0.69 0.1 1.84 3.93 1.25   .   1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
  8 3.56 0.69 0.1 1.84 3.93 1.25 0.18   .   0.27 0.5 0.67 0.01 0.61 0.82 1.7
  9 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13   .   0.5 0.67 0.01 0.61 0.82 1.7
 10 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27  .  0.67 0.01 0.61 0.82 1.7
 11 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5   .   0.01 0.61 0.82 1.7
 12 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67   .   0.61 0.82 1.7
 13 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01   .   0.82 1.7
 14 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61   .   1.7
 15 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82  .
 16 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
 17 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
 18 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
 19 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7
 20 3.56 0.69 0.1 1.84 3.93 1.25 0.18 1.13 0.27 0.5 0.67 0.01 0.61 0.82 1.7

Obs  y16  y17 y18  y19  y20   j(mean)     j(s)     j(s^2)   j(s_mle)

  1 0.39 0.11 1.2 1.21 0.72   0.91211   0.90292   0.81526   0.87884
  2 0.39 0.11 1.2 1.21 0.72   1.06316   1.08534   1.17796   1.05639
  3 0.39 0.11 1.2 1.21 0.72   1.09421   1.06449   1.13314   1.03610
  4 0.39 0.11 1.2 1.21 0.72   1.00263   1.07159   1.14830   1.04301
  5 0.39 0.11 1.2 1.21 0.72   0.89263   0.83570   0.69840   0.81341
  6 0.39 0.11 1.2 1.21 0.72   1.03368   1.08758   1.18284   1.05858
  7 0.39 0.11 1.2 1.21 0.72   1.09000   1.06846   1.14160   1.03996
  8 0.39 0.11 1.2 1.21 0.72   1.04000   1.08852   1.18488   1.05949
  9 0.39 0.11 1.2 1.21 0.72   1.08526   1.07249   1.15023   1.04388
 10 0.39 0.11 1.2 1.21 0.72   1.07316   1.08073   1.16797   1.05190
 11 0.39 0.11 1.2 1.21 0.72   1.06421   1.08494   1.17710   1.05601
 12 0.39 0.11 1.2 1.21 0.72   1.09895   1.05959   1.12272   1.03132
 13 0.39 0.11 1.2 1.21 0.72   1.06737   1.08364   1.17426   1.05473
 14 0.39 0.11 1.2 1.21 0.72   1.05632   1.08736   1.18236   1.05836
 15 0.39 0.11 1.2 1.21 0.72   1.01000   1.07712   1.16018   1.04839
 16  .   0.11 1.2 1.21 0.72   1.07895   1.07715   1.16025   1.04842
 17 0.39  .   1.2 1.21 0.72   1.09368   1.06501   1.13424   1.03660
 18 0.39 0.11  .  1.21 0.72   1.03632   1.08807   1.18389   1.05905
 19 0.39 0.11 1.2  .   0.72   1.03579   1.08798   1.18370   1.05896
 20 0.39 0.11 1.2 1.21  .     1.06158   1.08589   1.17915   1.05692
```

R code for jackknifing the Manly data

```
library(bootstrap)
# The Manly data
x <- c(3.56, 0.69, 0.10, 1.84, 3.93, 1.25, 0.18, 1.13, 0.27, 0.50,
     0.67, 0.01, 0.61, 0.82, 1.70, 0.39, 0.11, 1.20, 1.21, 0.72)

# Sample mean
mean(x)
# Jackknife the mean
jackmean <- jackknife(x,mean)
jackmean
```

```r
# Bias-corrected jackknife estimate
meanjack = mean(x) - jackmean$jack.bias
meanjack

# Sample standard deviation
sd(x)
# Jackknife the standard deviation
jacksd <- jackknife(x,sd)
jacksd
# Bias-corrected jackknife estimate
sdjack = sd(x) - jacksd$jack.bias
sdjack

# Sample standard deviation with n in denominator (MLE)
sdmle <- function(x)(sqrt((length(x)-1)/length(x))*sd(x))
sdmle(x)
# Jackknife the MLE standard deviation (denominator with n)
jacksdmle <- jackknife(x,sdmle)
jacksdmle
# Bias-corrected jackknife estimate
sdmlejack = sdmle(x) - jacksdmle$jack.bias
sdmlejack

# Sample variance
var(x)
# Jackknife the variance
jackvar <- jackknife(x,var)
jackvar
# Bias-corrected jackknife estimate
varjack = var(x) - jackvar$jack.bias
varjack
```

R output for Jackknifing the Manly Data

```
# Jackknife the mean
----------------------
# Sample mean     1.0445
jack.se    0.2369516
jack.bias  0

jack.values
 [1]  0.9121053 1.0631579 1.0942105 1.0026316 0.8926316 1.0336842 1.0900000
 [8]  1.0400000 1.0852632 1.0731579 1.0642105 1.0989474 1.0673684 1.0563158
[15]  1.0100000 1.0789474 1.0936842 1.0363158 1.0357895 1.0615789

# Bias-corrected jackknife estimate  1.0445


# Jackknife the standard deviation
----------------------------------
# Sample standard deviation    1.059680
jack.se      0.2802791
jack.bias  -0.03710029

jack.values
 [1]  0.9029186 1.0853369 1.0644890 1.0715868 0.8357022 1.0875825 1.0684568
 [8]  1.0885209 1.0724860 1.0807253 1.0849440 1.0595853 1.0836350 1.0873628
[15]  1.0771155 1.0771511 1.0650050 1.0880677 1.0879814 1.0858855

# Bias-corrected jackknife estimate   1.096780
```

```
# Jackknife the MLE standard deviation (denominator with n)
------------------------------------------------------------
# Sample standard deviation with n in denominator (MLE)    1.032848
jack.se      0.2728037
jack.bias   -0.06330983

jack.values
 [1]  0.8788364 1.0563893 1.0360975 1.0430060 0.8134128 1.0585751 1.0399595
 [8]  1.0594885 1.0438813 1.0519008 1.0560070 1.0313246 1.0547329 1.0583612
[15]  1.0483872 1.0484218 1.0365998 1.0590473 1.0589633 1.0569234

# Bias-corrected jackknife estimate    1.096158


# Jackknife the variance
------------------------
# Sample variance       1.122921
jack.se      0.5442438
jack.bias    0

jack.values
 [1]  0.8152620 1.1779561 1.1331368 1.1482982 0.6983982 1.1828357 1.1416000
 [8]  1.1848778 1.1502263 1.1679673 1.1771035 1.1227211 1.1742649 1.1823579
[15]  1.1601778 1.1602544 1.1342357 1.1838912 1.1837035 1.1791474

# Bias-corrected jackknife estimate    1.122921
```

3.3   Jackknifing the Correlation Coefficient

- Let us consider estimation of a correlation coefficient ρ. We will calculate the jackknife estimates from the 15 (y, z) points from the following law school data (see Figure 3.1 and Table 3.1). Recall: LSAT score $= y$, GPA $= z$.

School	LSAT	GPA	School	LSAT	GPA	School	LSAT	GPA
1	576	3.39	6	580	3.07	11	653	3.12
2	635	3.30	7	555	3.00	12	575	2.74
3	558	2.81	8	661	3.43	13	545	2.76
4	578	3.03	9	651	3.36	14	572	2.88
5	666	3.44	10	605	3.13	15	594	2.96

R code for Jackknifing the law school data

```
# Jackknifing correlation coefficients for the law school data

library(bootstrap)
n=15;  # the number of design points

yzdata <- as.matrix(c(576,3.39,580,3.07,653,3.12,
635,3.30,555,3.00,575,2.74,558,2.81,661,3.43,545,2.76,
578,3.03,651,3.36,572,2.88,666,3.44,605,3.13,594,2.96))
dim(yzdata) <- c(2,n)

indata <- t(yzdata)

#
# Jackknifing Pearson's correlation coefficient
#
corr <- function(yz,indata) { cor(indata[yz,1],indata[yz,2]) }

# Pearson's correlation coefficient r
sampcorr <- cor(indata[1:n,1],indata[1:n,2])
sampcorr

# Jackknife Pearson's correlation coefficient
jacklaw <- jackknife(1:n,corr,indata)
jacklaw

# Bias-corrected jackknife estimate
corrjack = sampcorr - jacklaw$jack.bias
corrjack
```

R output for Jackknifing the law school data

```
> # Pearson's correlation coefficient r
[1] 0.7763745
>
> # Jackknife Pearson's correlation coefficient
$jack.se
[1] 0.1425186

$jack.bias
[1] -0.006473623

$jack.values
 [1] 0.8929471 0.7799687 0.8181007 0.7637068 0.7845360 0.7857184 0.7549984
 [8] 0.7361618 0.7403509 0.7760968 0.7517391 0.7670413 0.7313197 0.7761231
[15] 0.7798725

> # Bias-corrected jackknife estimate
[1] 0.7828481
```

3.4 Jackknifing Bioequivalence

- The following example was taken from Efron and Tibshirani (1993). It involves ratio estimation.

- In this study, 8 subjects wore medical patches that release a hormone into the blood stream. Each subject had blood levels of the hormone measured after wearing 3 different patches: (1) a patch containing no hormone (placebo), (2) a patch containing hormones manufactured at the old site (oldpatch), and (3) a patch containing hormones manufactured at the new site (newpatch).

- Two differences were calculated;

$$z = \text{oldpatch} - \text{placebo} \qquad\qquad y = \text{newpatch} - \text{oldpatch}$$

- The goal of the study was to establish **bioequivalence**. The patches manufactured at the new site needed to be approved by the US Food and Drug Administration (FDA). This required showing that they were bioequivalent to patches manufactured at the old site.

- The goal is to establish that patches manufactured at the new and old site "match" (bioequivalent).

- The following table summarizes the experimental data:

subject	placebo	oldpatch	newpatch	z	y
1	9243	17649	16449	8406	-1200
2	9671	12013	14614	2342	2601
3	11792	19979	17274	8187	-2705
4	13357	21816	23798	8459	1982
5	9055	13850	12560	4795	-1290
6	6290	9806	10157	3516	351
7	12412	17208	16570	4796	-638
8	18806	29044	26325	10238	-2719
				$\overline{z} =$	$\overline{y} =$
				6342	-452.3

- Statistically, the FDA criterion for bioequivalence is

$$\frac{|E(newpatch) - E(oldpatch)|}{E(oldpatch) - E(placebo)} \leq .20$$

That is, the FDA wants the patches from the new site match the patches from the old site within 20% of the amount of hormone added to the blood stream by patches from the old site compared to placebo levels.

- Let the parameter θ be the ratio $\theta = \dfrac{E(newpatch) - E(oldpatch)}{E(oldpatch) - E(placebo)}$

- The estimate of θ is

$$\widehat{\theta} = \overline{y}/\overline{z} = -452.3/6342$$

- Now we just use jackknife estimation to calculate the eight $\widehat{\theta}_{(i)}$ replications, take their mean to get $\widehat{\theta}_{(\cdot)}$, and calculate the bias and standard errors using $\widehat{\theta}$, $\widehat{\theta}_{(i)}$, and $\widehat{\theta}_{(\cdot)}$.

R code for jackknifing the bioequivlence data

```
# Jackknifing bioequivalence

library(bootstrap)
n = 8;  # the number of design points

xx= c( 9243,  9671, 11792, 13357,  9055,  6290, 12412, 18806)
zz= c(17649, 12013, 19979, 21816, 13850,  9806, 17208, 29044)
yy= c(16449, 14614, 17274, 23798, 12560, 10157, 16570, 26325)

z = zz-xx
y = yy-zz
indata <- as.matrix(cbind(y,z))

# Sample bioequivalence
sampbioeq = mean(y)/mean(z)
sampbioeq

# Define the bioequivalence function

bioeq <- function(yz,indata) { mean(indata[yz,1])/mean(indata[yz,2]) }

# Jackknife the bioequivalence

jackbioeq <- jackknife(1:n,bioeq,indata)
jackbioeq

# Bias-corrected jackknife estimate
bioeqjack = sampbioeq - jackbioeq$jack.bias
bioeqjack
```

R output for jackknifing the bioequivlence data

```
> # Sample bioequivalence
[1] -0.0713061
>
> # Jackknife the bioequivalence
$jack.se
[1] 0.1055278

$jack.bias
[1] 0.008002488

$jack.values
[1] -0.05711856 -0.12849970 -0.02145610 -0.13245033 -0.05067038 -0.08404803
[7] -0.06486298 -0.02219698

> # Bias-corrected jackknife estimate
[1] -0.07930858
```

3.5   Limitations of the Jackknife

- The jackknife method of estimation can fail if the statistic $\widehat{\theta}$ is not *smooth*. *Smoothness* implies that relatively small changes to data values will cause only a small change in the statistic.

- The sample median is an example of a statistic that is <u>not smooth</u>.

- For example, look at the Manly (2007) data. The ordered values are

0.01	0.10	0.11	0.18	0.27	0.39	0.50	0.61	0.67	0.69
0.72	0.82	1.13	1.20	1.21	1.25	1.70	1.84	3.56	3.93

  Note that there are only 2 different jackknife estimate values.

  - 1/2 of the jackknife estimates $= 0.72$ (based deleting each of the first 10 values).
  - 1/2 of the jackknife estimates $= 0.69$ (based deleting each of the last 10 values) has a value of 0.72.

- Therefore, the jackknife is not a good estimation method for estimating percentiles (such as the median), or when using any other non-smooth estimator.

- This will not be the case using the *bootstrap method of estimation.*

- An alternate the jackknife method of deleting one observation at a time is to delete d observations at a time ($d \geq 2$). This is known as the **delete-d jackknife**.

- Because we are deleting d, there will be $\binom{n}{d}$ jackknife samples of size $n - d$ and therefore $\binom{n}{d}$ jackknife replications, and the estimate of the standard error is

$$
\left(\frac{n-d}{d \binom{n}{d}} \sum (\widehat{\theta}_{(i)} - \widehat{\theta}_{(\cdot)})^2 \right)^{1/2}
$$

  where the summation is take over all $\binom{n}{d}$ subsets of points.

- One reason to use the delete-d jackknife is to handle cases of non-smooth estimators.

- In practice, if n is large and d is chosen such that $\sqrt{n} < d < n$, then the problems of non-smoothness are removed.