Triola Statistics Series Review

Pearson

Basic Terms

Statistics: Methods for planning experiments, obtaining data, organizing, summarizing, analyzing, interpreting, and drawing conclusions based on data.

Population: Collection of *all* elements to be studied.

Census: Data from *every* member of a population.

Sample: Subcollection of members from a population.

Parameter: Numerical measurement of characteristic of a *population*.

Statistic: Numerical measurement of characteristic of a *sample*.

Random Sample: Every member of population has same chance of being selected.

Simple Random Sample: Every sample of same size n has the same chance of being selected.

Describing, Exploring, and Comparing Data

Measures of Center:

Population mean: μ

Sample mean: $\bar{x} = \dfrac{\sum x}{n}$

Mean from frequency dist.:

$$\bar{x} = \frac{\sum (f \cdot x)}{n}$$

Median: Middle value of data arranged in order.

Mode: Most frequent data value(s).

Midrange: $\dfrac{\text{maximum} + \text{minimum}}{2}$

Measures of Variation:

Range: maximum − minimum

Sample standard deviation:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \text{ or } \sqrt{\frac{n(\sum x^2) - (\sum x)^2}{n(n - 1)}}$$

St. dev. from frequency dist.:

$$s = \sqrt{\frac{n[\sum (f \cdot x^2)] - [\sum (f \cdot x)]^2}{n(n - 1)}}$$

Sample variance: s^2

Population st. dev.:

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

Population variance: σ^2

Distribution: Explore using frequency distribution, histogram, dotplot, stemplot, boxplot.

Outlier: Value far away from almost all other values.

Time: Consider effects of changes in data over time. (Use time-series graphs, control charts.)

Probability

Rare Event Rule: If, under a given assumption, the probability of a particular observed event is very small and the observed event occurs significantly less than or significantly greater than what we typically expect with that assumption, conclude that the assumption is probably not correct.

Relative Frequency:

$$P(A) = \frac{\text{number of times } A \text{ occurred}}{\text{number of trials}}$$

Classical Approach:

$$P(A) = \frac{s}{n} \text{ (equally likely outcomes)}$$

Probability property: $0 \le P(A) \le 1$

Complement of Event *A*:

$$P(\bar{A}) = 1 - P(A)$$

Addition Rule:

Disjoint Events: Cannot occur together.

If A, B are disjoint:

$$P(A \text{ or } B) = P(A) + P(B)$$

If A, B are not disjoint:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Multiplication Rule:

Independent Events: No event affects probability of other event.

If A, B are independent:

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

If A, B are dependent:

$$P(A \text{ and } B) = P(A) \cdot P(B|A)$$

where $P(B|A)$ is $P(B)$ assuming that event A has already occurred.

Counting

Multiplication Counting Rule: If an event can occur m ways and a second event can occur n ways, together they can occur $m \cdot n$ ways.

Factorial Rule: n different items can be arranged $n!$ different ways.

Permutations (order counts) of r items selected from n different items:

$${}_nP_r = \frac{n!}{(n - r)!}$$

Permutations when some items are identical to others:

$$\frac{n!}{n_1!n_2! \ldots n_k!}$$

Combinations (order doesn't count) of r items selected from n different items:

$${}_nC_r = \frac{n!}{(n - r)!r!}$$

Random Variables

Random Variable: Variable that has a single numerical value, determined by chance, for each outcome.

Probability Distribution: Graph, table, or formula that gives the probability for each value of the random variable.

Requirements of random variable:

1. $\sum P(x) = 1$
2. $0 \le P(x) \le 1$

Parameters of random variable:

$$\mu = \sum [x \cdot P(x)]$$

$$\sigma = \sqrt{\sum [x^2 \cdot P(x)] - \mu^2}$$

Expected value: $E = \sum [x \cdot P(x)]$

Binomial Distribution: Requires fixed number of independent trials with all outcomes in two categories, and constant probability.

n:	Fixed number of trials
x:	Number of successes in n trials
p:	Probability of success in one trial
q:	Probability of failure in one trial
$P(x)$:	Probability of x successes in n trials

$$P(x) = \frac{n!}{(n - x)!x!} \cdot p^x \cdot q^{n-x}$$

$\mu = np$ Mean (binomial)

$\sigma = \sqrt{npq}$ St. dev. (binomial)

Poisson Distribution: Discrete probability distribution that applies to occurrences of some event *over a specified interval*.

$$P(x) = \frac{\mu^x \cdot e^{-\mu}}{x!} \text{ where } e \approx 2.71828$$

ISBN-13: 978-0-13-446509-8
ISBN-10: 0-13-446509-1

EAN
9 780134 465098 90000

Normal Distribution

Continuous random variable having bell-shaped and symmetric graph and defined by specific equation.

Standard Normal Distribution:
Normal distribution with

$$\mu = 0 \text{ and } \sigma = 1.$$

Standard z score: $z = \dfrac{x - \mu}{\sigma}$

Central Limit Theorem:

As sample size increases, sample means $\bar{x}$ approach *normal* distribution;

$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

so that $z = \dfrac{\bar{x} - \mu_{\bar{x}}}{\dfrac{\sigma}{\sqrt{n}}}$

Normal Approximation to Binomial:

Requires $np \geq 5$ and $nq \geq 5$. Use $\mu = np$ and $\sigma = \sqrt{npq}$.

Determining Sample Size

Proportion:

$$n = \frac{[z_{\alpha/2}]^2 \cdot 0.25}{E^2}$$

$$n = \frac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2} \quad (\hat{p} \text{ and } \hat{q} \text{ known})$$

Mean: $n = \left[\dfrac{z_{\alpha/2}\sigma}{E}\right]^2$

Confidence Intervals (Using One Sample)

Proportion: $\hat{p} - E < p < \hat{p} + E$

where $E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$ and $\hat{p} = \dfrac{x}{n}$

Mean: $\bar{x} - E < \mu < \bar{x} + E$

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (\sigma \text{ not known})$$

Standard Deviation:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_L}}$$

Confidence Intervals (Using Two Samples)

Two Proportions:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

where $E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}$

Two Means (Independent):

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - u_2) < (\bar{x}_1 - \bar{x}_2) + E$$

where $E = t_{\alpha/2}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

df = smaller of $n_1 - 1$ and $n_2 - 1$.

Alternative Cases for Two Independent Means:

If σ_1, σ_2 unknown but assumed equal, use pooled variance s_p^2:

$$E = t_{\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Known σ_1 and σ_2: $E = z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Matched Pairs:

$\bar{d} - E < \mu_d < \bar{d} + E$

where $E = t_{\alpha/2}\dfrac{s_d}{\sqrt{n}}$ and df $= n - 1$

Common Critical z Values

CONFIDENCE INTERVAL

Confidence Level	Critical Value
0.90	1.645
0.95	1.96
0.99	2.575

HYPOTHESIS TEST: RIGHT-TAILED

Significance Level α	Critical Value
0.05	1.645
0.025	1.96
0.01	2.33
0.005	2.575

HYPOTHESIS TEST: LEFT-TAILED

Significance Level α	Critical Value
0.05	−1.645
0.025	−1.96
0.01	−2.33
0.005	−2.575

HYPOTHESIS TEST: TWO-TAILED

Significance Level α	Critical Value
0.05	±1.96
0.01	±2.575
0.10	±1.645

Hypothesis Testing

Hypothesis Test: Procedure for testing claim about a population characteristic.

Null Hypothesis H_0: Statement that value of population parameter is *equal to* some claimed value.

Alternative Hypothesis H_1: Statement that population parameter has a value that somehow differs from value in the null hypothesis.

Critical Region: All values of test statistic leading to rejection of null hypothesis.

α = Significance Level: Probability that test statistic falls in critical region, assuming null hypothesis is true.

Type I Error: Rejecting null hypothesis when it is true. Probability of type I error is significance level α.

Type II Error: Failing to reject null hypothesis when it is false. Probability of type II error is denoted by β.

Power of test: Probability of rejecting a false null hypothesis.

Procedure

1. Identify original claim, then state null hypothesis (with equality) and alternative hypothesis (without equality).
2. Select significance level α.
3. Evaluate test statistic.
4. Proceed with critical value method or P-value method:

Critical Value Method of Testing Hypotheses: Uses decision criterion of rejecting null hypothesis only if test statistic falls within critical region bounded by critical value.

Critical Value: Any value separating critical region from values of test statistic that do not lead to rejection of null hypothesis.

***P*-value Method of Testing Hypotheses:** Uses decision criterion of rejecting null hypothesis only if P-value $\leq \alpha$ (where α = significance level).

***P*-value:** Probability of getting value of test statistic *at least as extreme* as the one found from sample data, assuming that null hypothesis is true.

Left-Tailed Test: P-value = area to *left* of test statistic

Right-Tailed Test: P-value = area to *right* of test statistic

Two-Tailed Test: P-value = *twice* the area in tail beyond test statistic

Choosing Between *t* and *z* for Inferences about Mean

- σ unknown and normally distributed population: use t
- σ unknown and $n > 30$: use t
- σ known and normally distributed population: use z
- σ known and $n > 30$: use z

If none of the above apply, use nonparametric method or bootstrapping.

Wording of Conclusion

Does original claim include equality?

	Yes	No
Fail to reject H_0	"There is not sufficient evidence to warrant rejection of the claim that … [original claim]."	"There is not sufficient evidence to support the claim that … [original claim]."
Reject H_0	"There is sufficient evidence to warrant rejection of the claim that … [original claim]."	"There is sufficient evidence to *support* the claim that … [original claim]."

Hypothesis Testing (One Sample)

One Proportion: Requires simple random sample, $np \geq 5$ and $nq \geq 5$, and conditions for binomial distribution.

Test statistic: $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$ where $\hat{p} = \dfrac{x}{n}$

One Mean: Requires simple random sample and either $n > 30$ or normally distributed population.

Test statistic

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\dfrac{s}{\sqrt{n}}} \quad (\text{for } \sigma \text{ not known})$$

where df $= n - 1$

One Standard Dev. or Variance: Requires simple random sample and normally distributed population.

Test statistic: $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$

where df $= n - 1$

Hypothesis Testing (Two Proportions or Two Independent Means)

Two Proportions: Requires two independent simple random samples and $np \geq 5$ and $nq \geq 5$ for each.

Test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\bar{p}\,\bar{q}}{n_1} + \dfrac{\bar{p}\,\bar{q}}{n_2}}}$$

where $\bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$

and $\hat{p}_1 = \dfrac{x_1}{n_1}$ and $\hat{p}_2 = \dfrac{x_2}{n_2}$

Two Means (independent samples): Requires two independent simple random samples with both populations normally distributed or $n_1 > 30$ and $n_2 > 30$.

The population standard deviations σ_1 and σ_2 are usually unknown.

Recommendation: do not assume that $\sigma_1 = \sigma_2$.

Test statistic (unknown σ_1 and σ_2, and not assuming $\sigma_1 = \sigma_2$):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

df $=$ smaller of $n_1 - 1$ and $n_2 - 1$

Hypothesis Testing (Alternative Cases for Two Means with Independent Samples)

Requires two independent simple random samples and either of these two conditions:

Both populations normally distributed or $n_1 > 30$ and $n_2 > 30$.

Alternative case when σ_1 and σ_2 are not known, but it is assumed that $\sigma_1 = \sigma_2$:

Pool variances and use test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

and df $= n_1 + n_2 - 2$

Alternative case when σ_1 and σ_2 are both known values:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Matched Pairs

Requires simple random samples of matched pairs and either the number of matched pairs is $n > 30$ or the pairs have differences from a population with a distribution that is approximately normal.

d: Individual difference between values in a single matched pair

μ_d: Population mean difference for all matched pairs

$\bar{d}$: Mean of all *sample* differences d

s_d: Standard deviation of all *sample* differences d

n: Number of *pairs* of data

Test statistic: $t = \dfrac{\bar{d} - \mu_d}{\dfrac{s_d}{\sqrt{n}}}$ where df $= n - 1$

Hypothesis Testing (Two Variances or Two Standard Deviations)

Requires independent simple random samples from populations with normal distributions.

s_1^2: *larger* of the two sample variances

n_1: size of the sample with the *larger* variance

σ_1^2: variance of the population with the *larger* sample variance

Test statistic: $F = \dfrac{s_1^2}{s_2^2}$ where s_1^2 is the *larger* of the two sample variances and numerator df $= n_1 - 1$ and denominator df $= n_2 - 1$

Correlation

Scatterplot: Graph of paired (x, y) sample data.

Linear Correlation Coefficient *r*: Measures strength of *linear* association between the two variables.

Property of *r*: $-1 \leq r \leq 1$

Correlation Requirements: Bivariate normal distribution (for any fixed value of x, the values of y are normally distributed, and for any fixed value of y, the values of x are normally distributed).

Linear Correlation Coefficient:

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\,\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

$$\text{or } r = \frac{\Sigma(z_x z_y)}{n - 1}$$

Explained Variation: r^2 is the proportion of the variation in y that is explained by the linear association between x and y.

Hypothesis test

1. Using r as test statistic: If $|r| \geq$ critical value (from table), then there is sufficient evidence to support a claim of linear correlation.

   If $|r| <$ critical value, there is not sufficient evidence to support a claim of linear correlation.
2. Using t as test statistic:

   $$t = \frac{r}{\sqrt{\dfrac{1 - r^2}{n - 2}}} \text{ with df } = n - 2$$

Triola Statistics Series Review

Regression

Regression Equation: $\hat{y} = b_0 + b_1 x$

x: Independent (predictor, explanatory) variable

$\hat{y}$: Dependent (response) variable

b_1: Slope of regression line

$$b_1 = r\frac{s_y}{s_x} \text{ or } b_1 = \frac{n\Sigma(xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

b_0: y-intercept of regression line

$$b_0 = \bar{y} - b_1\bar{x}$$

Predicting value of y: If *no* linear correlation, best predicted y value is $\bar{y}$; if there *is* a linear correlation, the best predicted y value is found by substituting x value into regression equation.

Marginal Change: Amount a variable changes when the other variable changes by one unit.

Slope b_1 is the marginal change in y when x changes by one unit.

Influential Point: Strongly affects graph of regression line.

Residual: Difference between an observed sample y value and the value $\hat{y}$ that is predicted using the regression equation.

$$\text{Residual} = y - \hat{y}.$$

Least-Squares Property: The sum of squares of the residuals is the smallest sum possible.

Correlation/Regression: Variation and Prediction Intervals

Total Deviation: $y - \bar{y}$

Explained Deviation: $\hat{y} - \bar{y}$

Unexplained Deviation: $y - \hat{y}$

$$\Sigma(y - \bar{y})^2 = \Sigma(\hat{y} - \bar{y})^2 + \Sigma(y - \hat{y})^2$$

Coefficient of Determination:

$$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$

Standard Error of Estimate:

$$s_e = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}}$$

Prediction Interval for an Individual y:

$\hat{y} - E < y < \hat{y} + E$ where x_0 is given,

$$E = t_{\alpha/2}s_e\sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\Sigma x^2) - (\Sigma x)^2}}$$

and $t_{\alpha/2}$ has df $= n - 2$.

Multiple Regression

Procedure: *Obtain results using computer or calculator.*

Multiple Regression Equation:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$$

n: Sample size

k: Number of *independent* (x) variables

Important factors to consider:

P-value (from computer display) and adjusted R^2, where

$$\text{adj. } R^2 = 1 - \frac{(n - 1)}{[n - (k + 1)]}(1 - R^2)$$

Goodness-of-Fit

Goodness-of-Fit Test: Test the null hypothesis that an observed frequency distribution fits claimed distribution.

O: Observed frequency

E: Expected frequency

k: Number of different categories

n: Total number of trials

Requirements: Random data of frequency counts, and $E \geq 5$ for each category.

Test is *right-tailed* with test statistic:

$$\chi^2 = \sum\frac{(O - E)^2}{E}$$

where df $= k - 1$

Contingency Tables

Contingency Table: Two-way frequency table with row variable and column variable.

Requirements: Random data of frequency counts and $E \geq 5$ for each cell (where E is expected frequency).

Test of Independence: Test null hypothesis of no association between row variable and column variable.

Test of Homogeneity: Test that different populations have same proportions of some characteristic.

Test (independence or homogeneity) is right-tailed with test statistic:

$$\chi^2 = \sum\frac{(O - E)^2}{E}$$

where df $= (r - 1)(c - 1)$

Expected frequency (of cell):

$$E = \frac{(\text{row total})(\text{column total})}{(\text{grand total})}$$

McNemar's Test

Use for a 2×2 frequency table from matched pairs. **Requirement:** $b + c \geq 10$, where b and c are frequencies from discordant pairs.

Test is *right-tailed* with test statistic

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$$

where df $= 1$

One-Way ANOVA

Procedure: *Obtain P-value using computer or calculator.*

ANOVA (analysis of variance): Method of testing equality of three or more population means (by analyzing sample variances).

Requirements: Populations have approximately normal distributions, populations have same variance σ^2, samples are simple random samples, and samples are independent.

Decision criterion using significance level α:

P-value $\leq \alpha$: Reject null hypothesis of equal population means

P-value $> \alpha$: Fail to reject equality of population means.

Test statistic:

$$F = \frac{\text{variance between samples}}{\text{variance within samples}}$$

Two-Way ANOVA

Two-way ANOVA (analysis of variance) uses *two* factors: row factor and column factor.

Requirements: Data in each cell are from a normally distributed population; populations have same variance; sample data are from simple random samples; samples are independent; data are categorized two ways.

Balanced Design: All cells have the same number of sample values.

Procedure: Use computer or calculator to obtain P-values, then use the P-values to test for

1. **Interaction effect** between the two variables. (Stop here if there appears to be an interaction effect.)
2. **Effect from row variable**
3. **Effect from column variable**

Nonparametric Methods

Nonparametric (distribution-free) tests: Do not require assumptions about the population distributions.

Rank: Number assigned to a sample value according to its order in the *sorted* list. Lowest value has rank 1, 2nd lowest has rank 2, and so on.

Sign Test: Uses plus/minus signs instead of original data values. Used to test claims involving matched pairs, nominal data, or claims about median.

- n: Total number of (nonzero) signs
- x: Number of the less frequent sign

Test statistic if $n \le 25$: x

Test statistic if $n > 25$:

$$z = \frac{(x + 0.5) - (n/2)}{\sqrt{n}/2}$$

Reject H_0 if test statistic $\le$ critical values.

Wilcoxon Signed-Ranks Test: Uses ranks of data consisting of *matched pairs;* based on ranks of differences between pairs of values. Used to test null hypothesis that the matched pairs have differences with a median equal to zero.

- T: Smaller of two rank sums (sum of absolute values of negative ranks; sum of positive ranks)

Test statistic if $n \le 30$: T

Test statistic if $n > 30$:

$$z = \frac{T - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$$

Wilcoxon Rank-Sum Test: Uses ranks from two *independent* samples. Used to test null hypothesis that two independent samples are from populations with same median. Requires two independent random samples, each with more than 10 values.

- R: Sum of ranks for Sample 1
- n_1: Size of Sample 1
- n_2: Size of Sample 2

Test statistic: $z = \dfrac{R - \mu_R}{\sigma_R}$

where $\mu_R = \dfrac{n_1(n_1 + n_2 + 1)}{2}$

and $\sigma_R = \sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$

Kruskal-Wallis Test: Used to test null hypothesis that three or more independent samples are from populations with same median. Requires at least five observations in each independent sample of randomly selected values.

- N: Total number of observations
- R_1: Sum of ranks for Sample 1
- n_1: Number of values in Sample 1

Test statistic:

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ldots + \frac{R_k^2}{n_k}\right) - 3(N+1)$$

where test is right-tailed using χ^2 distribution with df $= k - 1$ and $k =$ number of samples

Rank Correlation: Uses ranks to test for correlation from random paired data.

- r_s: Rank correlation coefficient
- n: Number of pairs of sample data
- d: Difference between ranks for two values within a pair

Test statistic (if no ties in ranks):

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Test statistic (with ties among ranks):

$$r_s = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

Critical values for $n > 30$:

$$r_s = \frac{\pm z}{\sqrt{n - 1}}$$

Runs Test for Randomness: Used to determine whether sequence of sample data is in random order.

Run: Sequence of data having same characteristic.

- n_1: Total number of sample elements having one common characteristic
- n_2: Total number of sample elements having the other characteristic
- G: Total number of runs

If $\alpha = 0.05$, $n_1 \le 20$, and $n_2 \le 20$, test statistic is G. Otherwise,

Test statistic: $z = \dfrac{G - \mu_G}{\sigma_G}$

where $\mu_G = \dfrac{2n_1 n_2}{n_1 + n_2} + 1$ and

$$\sigma_G = \sqrt{\frac{(2n_1 n_2)(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

Statistical Process Control

Process Data: Data arranged according to some time sequence.

Run Chart: Sequential plot of *individual* data values over time

Control Chart of a process characteristic: Sequential plot over time of the characteristic values, including a centerline, lower control limit (LCL), and upper control limit (UCL).

***R* Chart (range chart):** Control chart of sample ranges.

- UCL: $D_4\bar{R}$
- Centerline: $\bar{R}$
- LCL: $D_3\bar{R}$

$\bar{x}$ Chart: Control chart of sample means.

- UCL: $\bar{\bar{x}} + A_2\bar{R}$
- Centerline: $\bar{\bar{x}}$
- LCL: $\bar{\bar{x}} - A_2\bar{R}$

***p* Chart:** Control chart to monitor the proportion p of some attribute.

- UCL: $\bar{p} + 3\sqrt{\dfrac{\bar{p}\,\bar{q}}{n}}$
- Centerline: $\bar{p}$
- LCL: $\bar{p} - 3\sqrt{\dfrac{\bar{p}\,\bar{q}}{n}}$

where $\bar{p} = \dfrac{\text{total number of defects}}{\text{total number of items}}$

Statistically Stable Process (or within statistical control): Process with only natural variation and no patterns, cycles, or unusual points

Out-of-Control Criteria:

1. There is a pattern, trend, or cycle that is not random.
2. There is a point above the upper control limit or below the lower control limit.
3. There are eight consecutive points all above or all below the centerline.

CONTROL CHART CONSTANTS

n	A_2	D_3	D_4
2	1.880	0.000	3.267
3	1.023	0.000	2.574
4	0.729	0.000	2.282
5	0.577	0.000	2.114
6	0.483	0.000	2.004
7	0.419	0.076	1.924

Negative z Scores

TABLE A-2 Standard Normal (z) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	* .0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	* .0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

NOTE: For values of z below −3.49, use 0.0001 for the area.

(continued)

*Use these common values that result from interpolation:

z Score	Area
−1.645	0.0500
−2.575	0.0050

Positive z Scores

TABLE A-2 *(continued)* Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	* .9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	* .9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.50 and up	.9999									

NOTE: For values of z above 3.49, use 0.9999 for the area.

*Use these common values that result from interpolation:

z Score	Area
1.645	0.9500
2.575	0.9950

Common Critical Values

Confidence Level	Critical Value
0.90	1.645
0.95	1.96
0.99	2.575

TABLE A-3 t Distribution: Critical t Values

	Area in One Tail 0.005	0.01	0.025	0.05	0.10
Degrees of Freedom	**Area in Two Tails** 0.01	0.02	0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.697	1.310
31	2.744	2.453	2.040	1.696	1.309
32	2.738	2.449	2.037	1.694	1.309
33	2.733	2.445	2.035	1.692	1.308
34	2.728	2.441	2.032	1.691	1.307
35	2.724	2.438	2.030	1.690	1.306
36	2.719	2.434	2.028	1.688	1.306
37	2.715	2.431	2.026	1.687	1.305
38	2.712	2.429	2.024	1.686	1.304
39	2.708	2.426	2.023	1.685	1.304
40	2.704	2.423	2.021	1.684	1.303
45	2.690	2.412	2.014	1.679	1.301
50	2.678	2.403	2.009	1.676	1.299
60	2.660	2.390	2.000	1.671	1.296
70	2.648	2.381	1.994	1.667	1.294
80	2.639	2.374	1.990	1.664	1.292
90	2.632	2.368	1.987	1.662	1.291
100	2.626	2.364	1.984	1.660	1.290
200	2.601	2.345	1.972	1.653	1.286
300	2.592	2.339	1.968	1.650	1.284
400	2.588	2.336	1.966	1.649	1.284
500	2.586	2.334	1.965	1.648	1.283
1000	2.581	2.330	1.962	1.646	1.282
2000	2.578	2.328	1.961	1.646	1.282
Large	2.576	2.326	1.960	1.645	1.282

TABLE A-4 Chi-Square (χ^2) Distribution

	Area to the *Right* of the Critical Value									
Degrees of Freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Source: Donald B. Owen, *Handbook of Statistical Tables.*

TABLE A-6 Critical Values of the Pearson Correlation Coefficient r

n	$\alpha = .05$	$\alpha = .01$
4	.950	.990
5	.878	.959
6	.811	.917
7	.754	.875
8	.707	.834
9	.666	.798
10	.632	.765
11	.602	.735
12	.576	.708
13	.553	.684
14	.532	.661
15	.514	.641
16	.497	.623
17	.482	.606
18	.468	.590
19	.456	.575
20	.444	.561
25	.396	.505
30	.361	.463
35	.335	.430
40	.312	.402
45	.294	.378
50	.279	.361
60	.254	.330
70	.236	.305
80	.220	.286
90	.207	.269
100	.196	.256

NOTE: To test $H_0: \rho = 0$ (no correlation) against $H_1: \rho \neq 0$ (correlation), reject H_0 if the absolute value of r is greater than or equal to the critical value in the table.

TABLE A-5 F Distribution ($\alpha = 0.025$ in the right tail)

Denominator degrees of freedom (df_2)	Numerator degrees of freedom (df_1): 1	2	3	4	5	6	7	8	9
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
2	38.506	39.000	39.165	39.248	39.298	39.331	39.335	39.373	39.387
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473
4	12.218	10.649	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047
5	10.007	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790