Formulas to Accompany

Math 1401 – Elementary Statistics

OLI/CMU

EDA: Examining Distributions

Mean

$$\bar{x} = \frac{\sum x}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Using Excel:

=AVERAGE(*select data*)

Range

$Range = Max - Min$

Using Excel:

=MAX(*select data*)-MIN(*select data*)

Quartiles

Using Excel:

=QUARTILE(*select data*, *quartile#*)

Inter-Quartile Range (IQR)

$IQR = Q3 - Q1$

Using Excel:

= QUARTILE(*select data, 3*)- QUARTILE(*select data, 1*)

Sample Standard Deviation

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

Using Excel:

=STDEV.S(*select data*)

Sample Variance

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

Using Excel:

=VAR.S(*select data*)

EDA: Examining Relationships

Correlation Coefficient

$$r = \frac{1}{n - 1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

Using Excel:

= CORREL([Data Set 1], [Data Set 2])

Least Squares Regression Line

$$\hat{y} = a + bx$$

where

$b = r\left(\frac{s_y}{s_x}\right)$

$a = \bar{y} - b\bar{x}$

Producing Data: Sampling

No relevant formulas or Excel commands

No relevant formulas or Excel commands

Probability: Introduction

Relative Frequency of Event A = $\frac{number\ of\ times\ A\ occured}{total\ number\ of\ repetitions}$

$$\boldsymbol{P(A)} = \frac{number\ of\ outcomes\ in\ A}{number\ of\ outcomes\ in\ S}$$

Probability: Finding Probability of Events

Complement Rule

$P(\text{not } A) = 1 - P(A)$

General Addition Rule

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ *for any events*

Multiplication Rule for Independent Events

$P(A \text{ and } B) = P(A) \cdot P(B)$ *for independent events*

Addition Rule for Disjoint Events

$P(A \text{ or } B) = P(A) + P(B)$ *for disjoint events*

Probability: Conditional Probability and Independence

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

General Multiplication Rule

$P(A \text{ and } B) = P(A) \cdot P(B|A)$ *for any events*

Probability: Random Variables

Mean (Expected Value):

$\mu_X = \sum_{i=1}^{n} x_i p_i$

Excel: =SUMPRODUCT(*[x's], [p's]*)

Variance:

$\sigma_X^2 = \sum_{i=1}^{n} (x_i - \mu_X)^2 p_i = \sum_{i=1}^{n} x_i^2 p_i - \mu_X^2$

Excel: =SUMPRODUCT((*[x's]* - μ_X)^2, *[p's]*)

Or =SUMPRODUCT(*[x's]*^2, *[p's]*) - μ_X^2

Standard deviation:

$\sigma_X = \sqrt{\sigma_X^2}$

Binomial Random Variables

$P(X = x) = \frac{n!}{x!(n-x)!} \cdot p^x (1-p)^{n-x}$

Excel: =BINOM.DIST($x, n, p, cumulative\ or\ not$)

Mean (Expected Value):

$\mu_X = np$

Variance:

$\sigma_X^2 = np(1-p)$

Standard deviation:

$\sigma_X = \sqrt{np(1-p)}$

z-score

$$z = \frac{x - \mu}{\sigma}$$

Excel: =STANDARDIZE(x, μ, σ)

Areas Under Standard Normal Curve:

$P(Z < z)$

Excel: =NORM.S.DIST(z, **1**)

Finding *x*-values Given *z*-scores:

$x = \mu + z \cdot \sigma$

Finding *x*-values Given Areas:

Excel: =NORM.INV(area to the left of x, μ, σ)

Areas Under Normal Curve:

$P(X < x)$

Excel: =NORM.DIST(x, μ, σ, **1**)

Probability: Sampling Distributions

Distribution of $\hat{p}$:

$\mu_{\hat{p}} = p$ *(mean)*

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ *(standard deviation)*

(Recall: $\hat{p} = \frac{\text{number of successes}}{n}$*)*

Distribution of $\bar{x}$:

$\mu_{\bar{x}} = \mu$ *(mean)*

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ *(standard deviation)*

Inference for One Sample: Estimating a Population Proportion

$z_c = 1.645$ for a 90% confidence interval

$z_c = 1.96$ for a 95% confidence interval

$z_c = 2.576$ for a 99% confidence interval

When to Use?

$n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

Margin of Error (proportion):

$$m = z_c \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence Interval for a Population Proportion:

$\hat{p} \pm z_c \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$\left(\hat{p} - z_c \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_c \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

Inference for One Sample: Estimating a Population Mean

$z_c = 1.645$ for a 90% confidence interval

$z_c = 1.96$ for a 95% confidence interval

$z_c = 2.576$ for a 99% confidence interval

Population Standard Deviation $\boldsymbol{\sigma}$ **is known.**

Margin of Error (mean):

$$m = z_c \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Interval for a Population Mean:

$\bar{x} \pm z_c \cdot \frac{\sigma}{\sqrt{n}}$

$\left(\bar{x} - z_c \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_c \cdot \frac{\sigma}{\sqrt{n}}\right)$

$t_c = T.INV.2T(1-90\%, n-1)$ for a 90% confidence interval

$t_c = T.INV.2T(1-95\%, n-1)$ for a 95% confidence interval

$t_c = T.INV.2T(1-99\%, n-1)$ for a 99% confidence interval

Population Standard Deviation $\boldsymbol{\sigma}$ **is unknown.**

Margin of Error (mean):

$$m = t_c \cdot \frac{s}{\sqrt{n}}$$

Confidence Interval for a Population Mean:

$\bar{x} \pm t_c \cdot \frac{s}{\sqrt{n}}$

$\left(\bar{x} - t_c \cdot \frac{s}{\sqrt{n}}, \quad \bar{x} + t_c \cdot \frac{s}{\sqrt{n}}\right)$

Inference for One Sample: Hypothesis Test for a Population Proportion

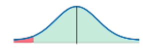
Test Statistic for Proportions

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

***p*-value**

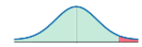
$P(Z < z)$

Excel: =NORM.S.DIST(z,**1**)



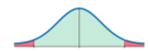
$P(Z > z)$

Excel: =1-NORM.S.DIST(z,**1**)



$P(Z < -|z|) + P(Z > |z|)$

Excel: =2*NORM.S.DIST($-|z|$,**1**)



Inference for One Sample: Hypothesis Test for a Population Mean

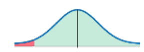
Test Statistic for Means

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

***p*-value (standard normal distribution)**

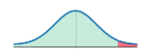
$P(Z < z)$

Excel: =NORM.S.DIST(z,**1**)



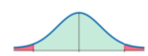
$P(Z > z)$

Excel: =1-NORM.S.DIST(z,**1**)



$P(Z < -|z|) + P(Z > |z|)$

Excel: =2*NORM.S.DIST($-|z|$,**1**)

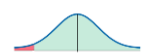


$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

***p*-value (*t*-distribution)**

$P(T < t)$

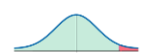
Excel: =T.DIST(t, df,**1**)



$P(T > t)$

Excel: = T.DIST.RT(t, df)

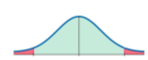
Or = 1-T.DIST(t, df, **1**)



$P(T < -|t|) + P(T > |t|)$

Excel: = T.DIST.2T($|t|, df$)

Or =2*T.DIST($-|t|, df$,**1**)



$df = n - 1$

Inference for Two Samples: Distribution of Differences in Sample Proportions

Conditions for a Normal Model:

$n_1 p_1 \geq 10, n_1(1-p_1) \geq 10, n_2 p_2 \geq 10,$ and $n_2(1-p_2) \geq 10$

$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$

$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

Inference for Two Samples: Estimate the Difference between Population Proportions

$z_c = 1.645$ for a 90% confidence interval

$z_c = 1.96$ for a 95% confidence interval

$z_c = 2.576$ for a 99% confidence interval

When to Use?

$n_1\hat{p}_1 \geq 10,\ n_1(1-\hat{p}_1) \geq 10$

$n_2\hat{p}_2 \geq 10,\ n_2(1-\hat{p}_2) \geq 10$

Standard Error *(difference in proportions):*

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Margin of Error (*difference in proportions*):

$$m = z_c\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Confidence interval (*difference in proportions*):

$$(\hat{p}_1-\hat{p}_2) \pm z_c\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$\left((\hat{p}_1-\hat{p}_2) - z_c\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}},\ (\hat{p}_1-\hat{p}_2) + z_c\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

Inference for Two Samples: Hypothesis Test for a Difference in Population Proportions

Test Statistic for Difference in Two Population Proportions

$$z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1}+\frac{\bar{p}(1-\bar{p})}{n_2}}}$$

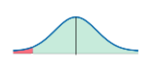
where the ***pooled proportion*** *is defined below:*

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

***p*-value**

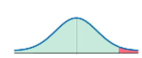
$P(Z < z)$

Excel: =NORM.S.DIST(z,**1**)



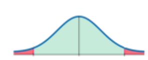
$P(Z > z)$

Excel: =1-NORM.S.DIST(z,**1**)



$P(Z < -|z|) + P(Z > |z|)$

Excel: =2*NORM.S.DIST($-|z|$,**1**)



MATH 1401 EXCEL FUNCTION LIST

Use	Syntax
Sum	=SUM([select data])
Count	=COUNT([select data])
Square Root	=SQRT(number)
Mean	=AVERAGE([select data])
Median	=MEDIAN([select data])
Range	= MAX([select data])-MIN([select data])
Sample Standard Deviation	=STDEV.S([select data]) or =STDEV([select data])
Sample Variance	=VAR.S([select data]) or =VAR([select data])
z-score	=STANDARDIZE(x, mean, standard deviation)
Quartile	=QUARTILE([select data], [# of quartile]) Example: for the 1st quartile, =QUARTILE([select data],1)
Inter-Quartile Range (IQR)	=QUARTILE([select data],3)-QUARTILE([select data],1)
Classical Definition of Probability	=(number of favorable outcomes)/(total number of possible outcomes)
Factorial	=FACT(number)
Combination	=COMBIN(Total Number, Number Chosen)
μ_X(Mean of random variable X)	=SUMPRODUCT([x's],[$P(x)'s$])
σ_X^2(Variance of random variable X)	=SUMPRODUCT(([x's]-μ_X)^2,[$P(x)'s$]) or =SUMPRODUCT([x's]^2,[$P(x)'s$])-μ_X^2
σ_X(Standard Deviation of random variable X)	=SQRT(σ_X^2)
Binomial Probabilities	=BINOM.DIST(# of successes, # of trial, probability of successes, cumulative) or =BINOMDIST(# of successes, # of trial, probability of successes, cumulative)
Standard Normal Distribution Probabilities	=NORM.S.DIST(z-score, 1) or =NORMSDIST(z-score)
Finding a z-score given a probability (area to the left)	=NORM.S.INV(probability) or =NORMSINV(probability)
Nonstandard Normal Distribution Probabilities	=NORM.DIST(x, mean, standard deviation,1) or =NORMDIST(x, mean, standard deviation,1)

MATH 1401 EXCEL FUNCTION LIST

Finding a x-value given a probability (area to the left)	=NORM.INV(probability, mean, standard deviation) or =NORMINV(probability, mean, standard deviation)
Population Standard Deviation	=STDEV.P([select data]) or =STDEVP([select data])
Population Variance	=VAR.P([select data]) or =VARP([select data])
Finding a critical t value given a probability (area to the left)	=T.INV(probability, df)
Probability from the t-distribution, Left Tail	=T.DIST(t, df, 1)
Probability from the t-distribution, Right Tail	=T.DIST.RT(t, df)
Probability from the t-distribution, Two Tails	=T.DIST.2T($\|t\|$, df)
Linear Correlation Coefficient	=CORREL([Data Set 1], [Data Set 2])
Slope	=SLOPE([$y's$],[$x's$])
y-intercept	=INTERCEPT([$y's$],[$x's$])

More Helpful Features

How to Sort Data

1. Select Data
2. Right click on selected data
3. Click Sort
4. Select the type of sorting you wish

How to Format Cells

1. Right click on selected cell
2. Select format
3. Select the desired format of the cell